The survival of the kindest: a theoretical review and empirical investigation of explanations to the evolution of human altruism

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln

2008

vorgelegt

von

Diplom-Kauffrau Julia Pradel

aus

Bocholt, Deutschland

Referent: Prof. Dr. Detlef Fetchenhauer

Korreferent: Prof. Harald Andreas Euler, Ph.D.

Tag der Promotion: 06.02.2009

Acknowledgments

I would like to thank Detlef Fetchenhauer, my dissertation supervisor, for broadening my mind by giving me the opportunity to explore the fascinating world of evolutionary psychology, for inspiring and advising me in many profound conversations, and for giving me so much support in mastering the diverse challenges that have been connected to this work. My great debt goes further to Harald Euler, who gave me his valuable opinion on many issues. In various sessions, Harald helped me to cultivate my linguistic style so that the phrasing of this work could be much more to the point than without his educational measures. As co-author of the second study presented in this work, he played a great part in improving the manuscript so that we could successfully publicize it. I am grateful to my wonderful colleagues Alexandra Haferkamp, Mareike Hoffmann, Thomas Schlösser, Sebastian Lotz, and Fabian Christandl who provided valuable feedback to my research and to large parts of the manuscript of this work. Putting up two “labor camps” in Vietze and Norlev, Alexandra, Mareike, and Detlef moreover spend several weeks of severe studying with me. Thanks to all the three of you—I will never forget these intensive times. Finally, I am particularly grateful to my parents, Anne and Jochen Pradel, who always supported me in accomplishing any desires I had and pepped me up whenever I was facing difficulties.

Contents

List of Figures

List of Tables

Part I: Theory

> The bravest men, who were always willing to come to the front in war, and who freely risked their lives for others, would, on an average, perish in larger number than other men. Therefore it seems scarcely possible (…) that the number of men gifted with such virtues, or that the standard of their excellence, could be increased through natural selection, that is, by the survival of the fittest.
>
> (Charles Darwin, 1871, p.163)

1. An introduction to the paradox of human altruism

In his intriguing publication “The Descent of Man,” Charles Darwin (1871) was concerned that his entire theory of evolution by natural selection might be negated by a single phenomenon prevalent in a variety of species including humans; namely altruism. If natural selection really favored the survival of the fittest, how could individuals survive who are willing to bear costs for the sake of the well-being of others?

Today, Darwin’s hypotheses on evolution by natural selection have long been supported empirically. Genetics has disclosed the physical basis and processes of heredity and it is known that traits, like prosociality, are passed directly from parents to offspring via genetic transmission. However, despite the triumph of Darwin’s theory, his early and prudent considerations of the “paradox of altruism,“ explicitly that any strategy so irrational as to sacrifice oneself for unrelated others should die out, has bothered researchers to this day.

A number of scientists have contributed valuable theories to elucidate the paradox of altruism since the time of Darwin. For example, Hamilton (1964), in his theory of *kin altruism*, explained that people behave prosocially towards those who are related by blood because such behavior serves to carry one's own genes into the next generation, not by the production of their own descendants but by aiding the reproduction of nondescendant relatives. A few years later, Trivers (1971) argued that prosocial behavior by a donor toward a beneficiary is adaptive if it is reciprocated by the beneficiary at a later time. The list of scientific insights concerning the evolution of altruism is extensive, because as of today a variety of explanations have been made out, each of them adding an important piece to solve the puzzle of human altruistic nature. However, in spite of the merits of these theories, there is still dissension about the origins of some particular oddities in the altruistic tendencies of humans, namely why humans act selflessly even when they are unobserved and when they are benefiting a stranger whom they will never meet again.

In everyday life, such behavior can be frequently observed. People return lost wallets to their owners with the cash intact, they give money to beggars, and tip waiters on the road in restaurants they will not visit for a second time. People are willing to contribute to the public good, they donate anonymously to charity, and they risk their lives to save strangers in need. All of these manners are socially desirable as they aid the functioning of our society, but displaying them implies the acceptance of harm to oneself. People neither promote their own genes by giving money to strangers, nor may they expect that the unknown beggar will recompensate them if they ever become needy themselves. So how could cooperation, especially cooperation in large and anonymous groups, survive the pressures of natural selection?

The present doctoral thesis sheds light on answers to this question. Chapter 2 sets the basis for the analysis; the subject-matter will be precisely defined and some preliminary remarks will be made. Chapters 3 to 6 provide a review of both prominent and less recognized theories on the evolution of altruism, with the different concepts compared to one another and discussed in-depth. Based on an integrative overview, I will analyze how much of the puzzle has been solved yet and which specific phenomena are still open to conjecture. Chapter 7 calls for an integration of present theories as a way to come closer to a solution of the human altruism paradox. With the aim of adding new insights to the issue, Chapter 8, 9, and 10, which form the centerpiece of this work, present three empirical studies that investigate in how far prosociality might have been favored by processes of *assortation*. Assortation denotes the selective association of individuals; in the context of this specific work, the grouping of prosocial persons. Indeed, assortation may be invoked as an explanation for the evolution of altruism, if the selfish advantage of egoistic individuals is out-competed by benefits of mutually cooperating altruists. However, to make assortation work as a driver of the evolution of altruism, two prerequisites have to be fulfilled: first, individuals have to be able to distinguish altruists from egoists, and second, altruists should, because of their perception, elect like-minded individuals for mutual cooperation in order to reap synergetic extra benefits.

The first study presented in Chapter 8 investigates whether humans are really able to identify altruists based on *first impression*. To test this, the so-called ‘thin slices paradigm’ (Ambady, Bernieri, & Richeson, 2000) was adopted. This paradigm has been used to show that people are able to identify other people’s permanent characteristics such intelligence, sociosexuality, or personality disorders by watching short videotapes of target persons. In the study presented here, judges watched 20-second silent video clips of unknown target persons and were asked to estimate the behavior of these target persons in a dictator game, which measures prosociality. Estimates were significantly better than chance indicating that humans can identify the altruistic dispositions of unknown persons.

However, people should not only be able to assess the character of strangers; people should at least be equally capable of distinguishing the good people from bad people when screening their pool of daily interaction partners. In addition, they should draw consequences from their insights by carefully choosing the right lovers, friends, and allies, and avoiding the villains.

The second study presented in Chapter 9 investigates whether individuals, in genuine groups, can identify the altruistic tendencies of their interactants. It further examines whether prosociality influences the formation of friendships in such that individuals assort themselves along the dimension of altruism. Students of six secondary school classes played an anonymous dictator game that functioned as a measure of altruism. Afterwards and unannounced, the students had to estimate their classmates' decisions and did so better than chance. Sociometry revealed that altruists were friends with more altruistic persons than were egoists.

Although the theory of assortation may explain the evolution of altruism in general, it does not explain the occurrence of inter-individual differences in altruism. However, the third study presented in Chapter 10 deals exactly with this matter. It investigates whether different levels of prosociality might have evolved as a result of different mating strategies, namely inter-individual variations in the propensity to engage in either short-term mating or long-term mating. Specifically, it assumes that prosociality is a necessity for acquiring a long-term partner, especially if an individual has to compensate for deficits in physical attractiveness. To find out whether this idea is true, the study tested whether individuals look out for different levels of prosociality depending on whether they are searching for a short-term mate or a long-term mate. Judges watched short video-clips of target persons and received additional information on the targets' prosociality. Judges were then asked to rate each of the target persons with regard to their desirability as a short-term and long-term mate. While prosociality was a significant predictor for long-term desirability, it was irrelevant when subjects chose a short-term mate. The results suggest that although altruism is costly, at least for some individuals it might be a wretched necessity to obtain access to mates and to reproduce.

The results of all three studies will be consolidated in Chapter 11. Conclusions will be drawn as to the consequences of these findings for the study of human altruism. Finally, directions for future research will be presented.

2. Specification of the subject matter

2.1. Levels of explanation: proximate versus ultimate causation

As mentioned above, the paradox of altruism is why should an individual carry out an act that is costly to perform and benefits another individual? At first sight, such behavior seems self-destructive and when focused on its direct consequences only, it has to be judged as irrational. Nevertheless, evolutionary theory suggests that there may be multiple reasons why people behave the way they do. Following the Nobel laureate Nikolas Tinbergen (1963), these reasons may be based on different levels of explanation: *proximate* and *ultimate* causes of behavior.

Proximate causation explains an individual's behavior based on trigger stimuli and internal mechanisms. It determines *how* a behavior occurs by analyzing physiological processes (e.g., hormones and neurons), psychological processes (e.g., motivations and learning), and developmental processes (e.g., gene-environment interactions). For example, it could be that prosocial behavior is due to the release of oxytocin, a hormone which is known to affect parts of the brain that trigger empathy, feelings of trust, and bonding. However, prosocial behavior could also be explained by learning; parents, by mechanisms of reward and punishment, have taught their children good manners, which include acts of sharing. Another possibility could be due to developmental processes such that people behave prosocially in particular domains like parenting because they carry certain genes that, over the course of development, endow them with the psychological machinery needed to care for potential offspring (Gaulin & McBurney, 2004).

In contrast to proximate causation, ultimate causation focuses on the evolution of the trait and its adaptive value. It asks *why* a behavior occurred the way it did (e.g., which problems altruism solved). An example of ultimate causation could be that over evolutionary time, altruistic individuals integrated better in social communities than egoistic individuals. If they received more support and shelter from other community members, they would have had better chances of survival and altruism could have survived the pressures of natural selection.

Indeed, the analysis of proximate causes of behavior is typically the domain of psychologists, while the analysis of ultimate causes is the domain of biologists (de Waal, 2008). However, which level of analysis is primarily relevant to this work? As previously mentioned, the present work takes an evolutionary viewpoint, a biological viewpoint, which focuses on ultimate causation and attempts to determine adaptiveness. For a trait to be adaptive, it has to increase an individual's fitness. According to life history theory (Stearns, 1989), this can be achieved through one of three possible goals: (1) developing to reproductive age, (2) reproducing successfully, and (3) rearing offspring to reproductive age.

The definition of these three ultimate goals is central to the present work. The stated aim is to shed light on the puzzle of human altruism. Therefore, the crucial question for the following analysis is "How far does altruism increase the fitness of individuals, that is, their survival odds and/or reproductive success?"

2.2. A definition of "altruism"

Different scientists have used the term "altruism" for distinct types of behavior (for an extensive review of social semantics see West, Griffin, and Gardner, 2007). Before starting to review theories about the evolution of altruism, the term altruism will be defined in the context of this work to avoid confusion.

Hamilton (1964, 1970) classified social behaviors depending on whether their consequences were beneficial or costly to the actor and the recipient. According to his definition, a behavior that is beneficial to the actor and costly to the recipient (+/-) is selfish, and a behavior that is costly to both persons (-/-) is spiteful. In contrast, a behavior that is costly to the actor and beneficial to the recipient (-/+) is altruistic. Finally, a behavior that is beneficial to both persons (+/+) is mutualistic. According to West et al. (2007), whether a behavior is beneficial or costly depends on its lifetime consequences for an individual. This definition implies that behavior can only be termed altruistic (-/+) if actors are never ever rewarded for their prosocial behavior. However, from an evolutionary perspective, any behavior that was favored by natural selection should have served the ultimate goal of increasing an individual's fitness (Tinbergen, 1963). Therefore, altruistic behavior in the strict sense that the donor suffers a net loss even in the long run should not exist, or, if it does, it has to be judged a maladaptation.

In contrast to West et al. (2007), Sober and Wilson (1998) use a less strict definition of altruism. They term a behavior altruistic if it is costly to perform, if it benefits another individual, and—the central concept—if the donor *does not expect* any payoff. Sober and Wilson thus allow for *unexpected* fitness benefits. The definition of altruism which is fundamental to this work, follows the idea of Sober and Wilson (1998) in that prosocial behavior that leads to an unintended and unexpected reward in the long run will nevertheless be termed "altruism." The words "prosociality" and "cooperation" will be used synonymously.

2.3. Individual-level versus group-level explanations

In recent years two camps have arisen trying to explain human altruism on different grounds: the individual-level adaptationist camp and the group-level adaptationist camp (Johnson, Prize, & Takezawa, 2008).

Advocates of individual-level theories assume that, at first sight, altruism is self-detrimental to the individual. However, when considering indirect and long-term consequences, there are several reasons why individuals benefit from displaying altruistic acts. Because these individual fitness benefits exist, altruism can evolve.

In contrast, the group-level adaptationist camp assumes that individual fitness benefits alone cannot explain the variety of altruistic behavior. In their view, there are several events in which altruistic individuals benefit the group but harm themselves and are never directly or indirectly rewarded. As these individuals are supposed to lose out in the evolutionary game when sticking to individual-level rationales, followers of this camp argue that higher-level rationales have to be considered. The central idea is that because altruism is good for the group, group-level mechanisms should develop which prevent altruism from extinction.

Both types of theories will be discussed in the following chapters. First, I will focus on a variety of individual-level theories in Chapter 3, with group-level theories investigated in Chapter 4.

3. Individual-level explanations for the evolution of altruism

3.1. Kin selection theory

Elucidating how altruism could evolve, Hamilton, in his seminal 1964 publication, explained that helping behavior towards related individuals (i.e., *kin altruism*) is beneficial to the individual who performs the altruistic act. To elaborate, Hamilton argued that because close relatives are likely to carry copies of an individual's own genes, promoting the survival and reproduction of one's relatives is a means to support the survival and reproduction of one's own genes in an *indirect* way. Thus, the theory of kin selection investigates altruism from the point of view of the gene, not the individual (Dawkins, 1976). For the gene, it does not matter which copy of itself—the own one or a copy of a copy—is passed on, it simply matters that as many copies as possible are passed on to the next generation. So taking this perspective, altruistic behavior towards relatives can actually be selfish, because it may help to carry the gene into the next generation. Therefore, kin altruism is adaptive.

The concept of kin altruism is encapsulated in a formula, which is known as "Hamilton's Rule": $b > c/r$. Here, b denotes the benefit of the recipient, c denotes the cost of the donor, and r denotes the degree of genetic relatedness between the recipient and the donor. The formula predicts that altruism is promoted if the cost for the donor is low, and if the benefit for the recipient and the relatedness between the individuals are high. If individuals display altruism only when this formula is met though, how do they assure that the relatedness with the beneficent is satisfactory? Hamilton proposed two mechanisms for this to work: *limited dispersal* and *kin discrimination.*

Limited dispersal (Hamilton, 1964, 1971), or population viscosity as it is also known, keeps related individuals together. The principle thus reflects living conditions in our evolutionary past, where families lived together in tribes. It has been suggested that altruism, which is directed indiscriminately to all neighbors, will be favored because neighbors are relatives. West, Pen, and Griffin (2002) assumed limited dispersal to be an important force to encourage high relatedness and thus cooperation through kin selection in a wide range of cases from bacteria to humans. Empirical evidence is difficult to find though; as Griffin & West (2002) stated, it is often hard to rule out alternative explanations (i.e., cooperation could be due to other mechanisms than limited dispersal). The most tangible evidence for the existence of limited dispersal stems from an experimental evolution study in bacteria. Griffin, West, and Buckling (2004) showed that siderophore production, which is costly to the individual bacterium that produces them but provides a reproductive benefit to neighbors, is higher in cultures of high relatedness than in cultures of lower relatedness.

The second mechanism that could lead to kin selection for altruism is *kin discrimination* (Hamilton, 1964). Kin discrimination predicts that altruistic behavior is preferentially directed to those individuals who are perceived as close kin. The phenomenon has been well documented in many vertebrate species (Griffin & West, 2003, Komdeur & Hatchwell, 1999). One of the best-studied cases is the long-tailed tit, a species in which individuals who failed to reproduce themselves selectively help the nest of closer relatives (Russell & Hatchwell, 2001; Sharp, McGowan, Wood, & Hatchwell, 2005). It has been shown that long-tailed tits discriminate between kin and non-kin based on vocal contact cues, which are learned from adults in the nesting period. Following these cues, long-tailed tits tend to promote relatives with whom they have been associated during the nestling phase.

However, do humans discriminate kin from non-kin, as well? A number of studies have investigated whether humans are able to identify other people’s level of relatedness, and if they do, whether they adjust their behavior accordingly (West, Gardner, & Griffin, 2008). Central areas of interest have been mother-baby interactions and mate choice. As mating with related individuals leads to offspring with homozygous recessive mutations, the discrimination of kin from non-kin would be highly adaptive. Indeed, studies on incest have shown that individuals try to avoid marrying or mating with close relatives. Lieberman, Tooby, and Cosmides (2003) found that, as in the case of the long-tailed tit, the underlying cue that humans use to assess relatedness is repeated interactions during the time of co-residence during infancy.

Going beyond kin discrimination because of repeated interactions, there is a line of work that additionally assumes a direct kin recognition mechanism such that individuals may assess relatedness based on genetic odor cues. For example, the major histocompatibility complex, the densest gene region of the human genome, influences body odors and it has been shown that these odor cues are linked to human mate choice preferences (for a review see Brown & Eklund, 1994). However, the general idea in regards to kin recognition in humans is that it usually works through learning mechanisms.

But how far has empirical research investigated the central postulate of kin selection, which states that helping relatives is beneficial and that the degree of helping should increase with relatedness? In fact, several studies have proven that humans are more inclined to promote relatives than unrelated individuals (see Barrett, Dunbar & Lycett, 2002). Some studies let predictions derived from kin selection compete against predictions based on social norms and rules for helping; for example, Burnstein, Crandall, and Kitayama (1994) found, consistent with kin selection theory but in contrast to the norm of social responsibility, that both Americans and Japanese report that in a “life-or-death” situation (e.g., saving someone from a fire) they would be more prone to help healthy relatives (who were presumably more likely to produce offspring) than nonhealthy relatives. Thus, in this study participants helped their relatives in a way that maximized their own inclusive fitness. Making even more sophisticated predictions on how helping of relatives would affect inclusive fitness, Euler and Weitzel (1996) found that a higher level of paternal certainty (the probability that a putative descendant is truly related) caused *maternal* grandparents to invest significantly more than *paternal* ones in their grandchildren. Webster (2003) replicated this result, but also showed that the effects of paternal certainty were strongest for benefactors with limited resources.

Summarizing the empirical investigations of kin selection theory, it can be concluded that consanguinity is an important predictor of helping behavior in humans. However, humans also cooperate with unrelated individuals, thus kin altruism cannot explain the whole story. But perhaps, selective interactions with individuals who are *similar*, though not related, could be likewise beneficial. This idea is considered in the next section.

3.2. Assortation: a broader concept of kin selection

3.2.1. Kin selection versus inclusive fitness

The theory of Hamilton described above is referred to in different ways. Maynard Smith (1964) called the phenomenon “kin selection,” the expression most often used. However, Hamilton himself coined the term “inclusive fitness” referring to the notion that individuals increase their fitness by helping relatives. Inclusive fitness can be divided into two components: *direct* and *indirect* fitness benefits. While direct fitness benefits denotes the component of fitness in which individuals gain through the production of their own offspring, indirect fitness benefits denotes the component of fitness in which individuals gain by helping the reproduction of related individuals (Brown & Brown, 1981).

In 1975 Hamilton pointed out that the concept of inclusive fitness is more general than that of kin selection. While kin selection is generally understood with respect to interactions between individuals who share alleles by common descent, inclusive fitness denotes the lifetime benefit from helping behavior directed towards *any* individual who is genetically similar, regardless of whether this similarity is due to genes shared by common ancestry and thus extending over most of the genome, or whether it is due to some other mechanism and concerns only a particular genetic locus for the behavior in question (West et al., 2007). Therefore, *two* possibilities exist on how to reach indirect fitness benefits: one is by helping relatives, whose genes are identical by descent (i.e., kin selection), the other is by helping those who are not related by blood but nevertheless share the same altruistic gene.

Grafen (1985) formalized the idea of consanguinity-detached relatedness by stating that the coefficient of relatedness *r* in Hamilton's Rule, is a regression coefficient, which describes how similar two individuals are over and above the average similarity of all individuals in the population. The average relatedness between two individuals who are picked randomly from the population should by definition be zero; nevertheless, there should be individuals who are more similar than average and other individuals who are less similar than average (West et al., 2008). From the viewpoint of a selfish altruistic gene, cooperation with an individual who is more similar than average, regardless of the fact that he or she does not "share blood" with the actor, can thus be beneficial.

3.2.2. The green-beard mechanism

The cooperation of individuals who share similar genes, though not being related by blood, has been termed *assortative interaction* or *assortation*. The central idea of the concept is close to kin selection, in that it is suggested that the promotion of the survival and reproduction of individuals with similar genes is a means to support the survival and reproduction of one's own genes in an *indirect* way.

Assortation mechanisms based on genes require three prerequisites: (1) the existence of a single gene, or a number of genes, which cause the behavior in question; (2) an observable mark which signals the presence of the gene(s); and (3) the preferential cooperation with those persons who carry the mark, and consequently the gene. The idea of the above-characterized gene was proposed by Hamilton (1964). However, it was Dawkins (1976) who introduced the term "green-beard gene" to associate the idea that a gene for altruism could be signaled through a well-recognizable mark like a green beard.

The concept of *green-beard altruism* implies that individuals who carry a gene for altruism can easily identify one another. Consequently, altruists are able to exclusively choose other altruists for mutual cooperation, leaving free riders no chance but to stay among themselves. As altruists may reach extraordinary benefits through mutualism, which out-competes the advantage of selfish free riders, green-beard altruism may increase the presence of altruists in a population even if genes are promoting other genes that are not exact copies of themselves in a molecular sense. In 1998, a "green-beard gene" was found in nature in the red imported fire ant (Keller & Ross, 1998); however, in humans, a green-beard gene has not yet been discovered.

3.2.3. The commitment-model

Although the existence of green-beard genes in humans is currently hypothetical, the economist Frank (1988, 2008) explained the evolution of altruism in humans on the basis of the green-beard concept (Dawkins, 1976). In his so-called "commitment model," he assumed that altruistic individuals carry a genetically coded observable mark which differentiates them from egoists: emotional commitment. Frank proposed that altruists have proximate mechanisms favoring cooperation in such that they cannot but feel ashamed if they deceive others. As they are inclined to avoid emotions of shame and guilt, they are made to behave righteously. Showing positive emotions like sympathy and responsibility, humans with an altruistic disposition naturally signal their good intentions and others are able to observe them directly. Therefore egoists can be distinguished from altruists. Because humans are free to choose their interaction partners in most situations, Frank assumed that altruists will be preferred for mutual cooperation. As altruists themselves do not want to be exploited by others, they will conditionally cooperate with other altruists only and reject egoists who lack emotional commitment. Consequently, altruism can become an advantageous strategy. Based on these considerations, Frank (1988) suggested that altruism evolved as a result of an autonomous motive system designed to retain cooperative interaction partners.

The commitment-model, as any theory of assortation, is valuable because it can explain the continuing variation in altruistic tendencies among humans (Frank, 1988). Consider a population in which there is a high proportion of altruists. In this case, altruists only face a small risk of being exploited and will be highly trusting. As altruists will naively interact with almost everybody, it will be adaptive to behave selfishly and exploit the altruists. Consequently, the number of egoists will rise; but then the altruists will become more careful and choose their interaction partners wisely. Therefore, altruism will come into favor again and the proportion of altruists will increase. This mechanism leads to frequency dependent selection and to an equilibrium of altruists and non-altruists existing side by side.

However, there is one obvious criticism of the arguments made by Dawkins (1976) and Frank (1988) that altruists could survive by signaling their good intentions and choosing one another for mutually beneficial cooperation; the appearance of cheaters could endanger the adequate assortation of individuals and lead to the extinction of altruists. To elaborate this idea, imagine a mutant who, by accident, possesses a green beard but is nevertheless selfish. As the property of a green beard enables an individual to obtain prosocial interaction partners, natural selection will favor selfish mutants with green beards who exploit others. These green-bearded free riders will quickly invade the population of green-bearded altruists and lead to their extinction. Transferring this argument on the commitment-model of Frank (1988), people, when being observed, can be assumed to display socially favorable emotions and conceal unfavorable ones. They should disguise their true character to exploit others, thereby reaping the benefits of an altruist without paying the cost of altruistic behavior (Fehr & Fischbacher, 2005).

Theoretically, the criticism has been fended off. Frank (2005) noted if natural selection is good enough to create a deceptive copy of an altruistic signal, it should also be in the position to modulate the original signal in order to prevent mimicry. In other words, an arms race between the true and the faked signal should arise. As long as the true signal is modulated before altruists are extinct, this arms race should continue indefinitely.

Nevertheless, although the commitment model may theoretically explain much about the evolution of altruism, it has attracted only little attention so far. Why is this? Johnson et al. (2008, p. 338) argue that "[t]he rejection of positive assortation as an evolutionary explanation of collective action has been based mainly on casually formulated and informally defended assumptions about the abilities of cooperators to identify one another and to ostracize free-riders." Therefore, many older models of cooperation assumed *random* encounters between individuals. However, this assumption is wobbly for two reasons.

First, it is indisputable that spatial and group structural features influence human cooperation (Johnson et al., 2008). For example, humans are more likely to interact with neighbours, individuals of the same social level, or individuals with similar hobbies than with individuals who live far away, belong to a different hierarchical level, and have divergent interests. It seems feasible to assume at least that people with similar interests have similar capabilities and might also be more similar with regard to their genetic material than individuals chosen randomly from the population are. Therefore, some researchers have started to develop models that capture such effects. For example, it has been shown that non-random encounters due to space influence the adaptiveness of cooperative strategies (Aktipis, 2004; Brauchli, Killingback, & Doebli, 1999; Killingback & Doebli, 1996).

The second reason why the assumption of non-random encounters in models of cooperation is doubtful, is, that it may be that humans are really equipped with a kind of “personality judgment instinct” (Haselton & Funder, 2006), which allows them to distinguish altruists from egoists. As mentioned at the beginning of this work, the empirical studies that are going to be presented here were designed to test this hypothesis. When examining these studies in Chapter 8 and 9, further empirical results will be reviewed which indeed have already given first hints to the existence of a limited personality judgment instinct (Sheldon, Sheldon, & Osbaldiston, 2000; Brown, Palameta, & Moore, 2003; Verplaetse, Vanneste, & Braeckman, 2007).

In summary, kin selection has been shown to explain cooperation between relatives. The broader application of the concept, positive assortation, may even explain cooperation between individuals who are unrelated in terms of genetical descent, but similar with regard to their innate altruistic tendencies. However, the relevance of assortation processes has been most widely ignored by empirical science until now.

3.3. Reciprocal altruism

3.3.1. Direct reciprocity

In 1971, only a few years after Hamilton published his theory on inclusive fitness, Trivers showed that altruism could evolve if the beneficiary rewards the beneficial act of the donor at a later time. This phenomenon is called reciprocal altruism, and in contrast to kin selection, reciprocal altruism is free of genetic self-favoritism such that it is not restricted to kin, but may account for altruism between unrelated individuals. There are several prerequisites for the occurrence of reciprocal altruism. The costs of the altruistic act for the donor have to be lower than the benefit for the recipient. Moreover, the altruist and the recipient must know each other and the interaction has to occur repeatedly. In other words, the interaction partners must have the possibility to exchange roles at a later time.

However, a rational recipient should follow the rule “Take the benefits and run!”—so why do cheaters not invade a population of reciprocal altruists over evolutionary time? Trivers solved this problem by stating that reciprocal altruism only pays off if individuals display altruism on a conditional basis, individuals should behave altruistically towards other altruists, but they should withhold altruism from cheaters because, if altruists are not paid back regularly, they end up with higher costs than benefits and this would lead to their extinction. The conditionality of altruism may, in turn, limit the occurrence of cheating behavior for the following reason; if individuals cheat in that they refrain from reciprocation, they have to anticipate that the donor will withhold help on the next encounter. Therefore, if individuals may expect to meet their interaction partner again, and if they may expect high costs from withheld help, they should definitely cooperate as long as costs of cooperation are low. In this sense, Axelrod and Dion (1988, p. 1387) noted that individuals should be inclined to reciprocate if the “shadow of the future” is long enough.

Reciprocal altruism follows the “tit-for-tat” principle (i.e., cooperation is responded to by cooperation, and non-cooperation is responded to by non-cooperation). In 1981, Axelrod and Hamilton formally supported the effective operation of reciprocal strategies by showing that people who respond according to their partner’s choice in the previous interaction gain higher payoffs than individuals who follow any other strategy. However, tit-for-tat strategies are vulnerable. First, to establish tit-for-tat, one interaction partner has to start cooperation. Second, already short disruptions of cooperation may lead to a complete collapse of the cooperative system. In small groups, a sudden withdrawal of cooperation is unlikely, because interactants consider the high probability of meeting again; thus Boyd and Richerson (1988) remarked that reciprocity may promote altruism in small groups. However, they particularly noted that reciprocity is an inadequate explanation for large-scale cooperation. As Johnson et al. (2008) noted, Boyd and Richerson’s 1988 conclusion was a prominent “fork in the road” which led many researchers to turn away from individual-level theories. However, the model that Boyd and Richerson studied was based on a disputable assumption, namely that humans execute a binary reciprocal strategy, i.e., that they either cooperate at full tilt (1.0) or not at all (0.0). Twenty years later, Johnson et al. (2008) noted that a more realistic assumption is that humans adjust their degree of cooperation steadily. Accordingly, the authors created a model in which cooperation varied continuously between 0.0 and 1.0. This slight adjustment of Boyd and Richerson’s (1988) original model—changing the strategy of reciprocity from binary to continuous—led to a result which makes it much more probable that reciprocity, besides influencing cooperation in small groups, also functioned as a driver of large-scale cooperation.

3.3.2. Indirect reciprocity

Alexander (1987) later extended the idea of Trivers’ (1971) theory of reciprocal altruism by stating that altruistic behavior advertises a tendency to cooperate and, by a process of *indirect reciprocity*, may attract cooperation from third parties in the future. Cooperation through indirect reciprocity can thus be summarized by the phrase “I help you, someone else helps me.”

Alexander pointed out that third parties may acquire information about a person’s prosociality in two ways: by carefully observing a person, or by obtaining information about an individual’s past behavior from someone else. Referring to this idea, Nowak & Sigmund (2005) identified two prerequisites that have to be fulfilled to render indirect reciprocity possible. First, individuals have to develop reputations that have to be communicated among the larger group. Second, individuals have to be equipped with cognitive abilities to identify and remember cooperative individuals beyond those with whom they have cooperated themselves—an ability that is required for *direct* reciprocity. A long line of empirical research shows that these premises are given facts, communities carefully observe their members and form a set of beliefs, perceptions and evaluations about each individual (for a review on the social psychology of reputation see Emler, 1990). Due to the human tendency to gossip, this reputation information is passed from one person to another and is often used when selecting business associates, friends, or romantic partners (Granovetter, 1985; Tinsley, O’Connor, & Sullivan, 2002).

Formal models and computer simulations support the theory of indirect reciprocity (e.g., Lotem, Fishman, & Stone, 1999; Panchanathan & Boyd, 2004). In one of these models Nowak & Sigmund (1998) showed that for the evolutionary stability of indirect reciprocity, the probability of knowing the “image score” of the recipient (i.e., his or her reputation and status) has to exceed the cost-benefit-ratio of the altruistic act. Extending insights from simulations, experimental evidence showed that the reputation of being an altruist leads to status, acceptance, and willing exchange partners (Hardy & van Vugt, 2006). Obviously, people anticipate these effects, as it has been shown that opportunities for reputation formation promote altruistic behavior in such that people cooperate more when their interaction history is made public (Nowak & Sigmund, 2005; Wedekind & Milinski, 2000).

Summarizing the ideas of reciprocity, it was shown that altruism can also be beneficial if it is directed to non-kin. Unrelated individuals have a propensity to cooperate whenever they may expect to be rewarded at a later time—either by the beneficiary himself or by third parties. Extracting the central idea of reciprocal strategies, it can be argued that individuals should be willing to cooperate whenever their actions are public, because their actions may then positively influence their reputation and increase future rewards.

3.4. Costly signaling & sexual selection

The theories of direct reciprocity (Trivers, 1971) and indirect reciprocity (Alexander, 1987) both focus on what kind of benefits individuals might retain as a reward for their prosocial behavior. In both cases, the general idea is that cooperative acts are rewarded by other cooperative acts. Amotz Zahavi (1975, 1995), however, remarked that cooperativeness does not have to be rewarded by other people's cooperativeness per se, but that there may be other rewards. It could be that cooperative behavior simply serves as a "costly signal" to advertise genetic quality such that only individuals who are particularly high in fitness may incur the cost of helping others on top of the demands of their own survival. Displaying such a costly signal could be rewarded by benefits like increased mating opportunities (Gintis, Smith, & Bowles, 2001; Zahavi, 1995).

The above thought is known as the "handicap principle" (Zahavi, 1975; Zahavi & Zahavi, 1997) and the general idea is that reliable signals must be costly to the signaler of a desirable trait, such that an individual with less of that trait could not afford to display the signal. This theory was developed in the field of behavioral ecology, although it has several parallels to classic social theories tracing back to Thorstein Veblen (1899/1994) and Marcel Mauss (1924). Costly signaling theory has gained much empirical support in studies of both animal signaling and anthropology (Gurven, Allen-Arave, Hill, & Hurtado, 2000; Lotem, Fishman, & Stone, 2002; Smith & Bird, 2000; Sosis, 2000); with regard to sexual selection it advocates that individuals of greater biological fitness signal their status through a handicapping behavior (e.g., altruism, conspicuous consumption, etc.) or a handicapping morphology (e.g., the peacock's flamboyant train).

If displaying altruism—as a handicap— indeed had positive effects on mating opportunities, it could be that individuals simply behave altruistically as a means to acquire sexual partners. If individuals differ in their opportunities to acquire a partner, they differ in their opportunities to transfer their genes into the next generation. Darwin (1871) termed this principle "sexual selection" (i.e., selection based on differential access to mates) and sharply distinguished this process from natural selection (i.e., selection based on differential survival); in fact, Darwin (1871) assumed that the evolution of many moral virtues might be explained by their sexual attractiveness. Miller (2000, 2007) extensively elaborated on the idea by stating that prosocial traits may have two kinds of signaling values for mate choice, one regarding genetic quality and another regarding a person's partnership or parenting abilities.

In regard to genetic quality, Miller (2007) argued that moral virtues, like altruism, are difficult to display if one has a high mutation load that impairs the precision of body and brain development. For example, people burdened with mutations associated with psychological disorders like autism, schizophrenia, or mental retardation tend to develop limited Theories of Mind. They consequently show deficits with regard to empathy and prosocial behaviors. Turning the argument the other way around, prosociality may therefore serve as a kind of neurogenetic warranty when judging a potential mate.

In regard to partnership or parenting abilities, the relevance of the argument is even more obvious because altruism, in addition to being a good signal of genetic qualities, is an attractive feature in its own right and individuals who display altruistic acts now are likely to care for their partner's and children's well-being at a later time (Miller, 2000; 2007). Establishing a long-term relationship with a prosocial individual can be extremely profitable as it may increase survival odds and reproductive success. Therefore, individuals who are in search of a partner should be especially open to altruistic men and women. There is empirical support for the theoretical idea that individuals who want to establish a sexual relationship look out for potential partners with prosocial traits (Buss, Shackelford, Kirkpatrick, & Larsen, 2001). Therefore, mate preferences for prosociality, whether as good genes or good partner and parent indicators, may have indeed shaped the evolution of human altruism.

However, does sexual selection explain the display of altruism in any situation? Certainly, no; altruism may indeed function as a costly signal, but natural selection should favor individuals who display this costly signal primarily in circumstances in which mating goals are prominent, that is, in situations in which potential mating partners are present. Recently, Griskevicius et al. (2007) showed that humans indeed have adaptations to adjust their behavior according to the activation of mating goals. In four experiments, the authors investigated how the triggering of mating goals increases the display of benevolence, and they showed that mating goals raise public, but not private, helping in women. In men, mating goals increase helpfulness whenever this may express heroism or dominance. These results indicate that altruism, as a costly signal, is displayed in a strategic manner.

To summarize, it has been argued that the theory of inclusive fitness may explain altruism towards related individuals. The theories of direct and indirect reciprocity may explain altruism in repeated interactions, which entail the chance of rewards. Moreover, according to the costly signaling theory, altruism is sexually attractive and helps to acquire sexual partners. It thus explains altruistic behavior in situations when potential mating partners are present. However, there is evidence for the existence of specific forms of altruism that these theories cannot explain: people cooperating with unrelated individuals in non-repeated interactions when potential mating partners are absent, such as returning lost wallets or making anonymous donations. As individual-level theories are insufficient to explain these phenomena, a second group of theories, group-level explanations, will be introduced to find out whether these theories may shed more light on the subject.

> Although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe (...) an advancement in the standard of morality will certainly give an immense advantage to one tribe over another.
>
> (Charles Darwin, 1871)

4. Group-level explanations for the evolution of altruism

As extensively debated in the preceding chapters, individuals who behave altruistically put themselves into a fitness disadvantage. So how could altruism evolve? Individual-level theories may only explain parts of this paradox. Already Charles Darwin believed that a possible solution to the puzzle might be that altruistic behavior occurs because it is advantageous at the group-level, and several scientists have seized on this idea. The details of their theories will be outlined in the following sections.

4.1. “Old” group selection

In 1962, two years before Hamilton published his seminal work about inclusive fitness, Wynne-Edwards proposed that traits could evolve through what he called “group selection.” Group selection is defined as the evolution of traits due to their positive impact on the survival odds of a population, rather than the survival odds of an individual, as is proposed for individual selection.

Wynne-Edwards (1962) thought that, through group selection, traits that were bad for the survival of a population would lead to its extinction, so that the only populations to survive would be populations of individuals who did not necessarily have traits that were best for their own survival and reproduction, but that were good for the survival and reproduction of the group. In more concrete terms, Wynne-Edwards thought that groups with selfless individuals would outcompete groups with selfish individuals. Wynne-Edwards described a number of traits for which it seemed that individuals strive for the conservation of their species in that they do things that decrease their own fitness but increase the population's chance of survival. Two of these traits are: decreased reproduction and altruism.

With regard to decreased reproduction, Wynne-Edwards (1962) noted that individuals of many species do not produce as many offspring at a time as they physiologically can, for example, birds are physiologically able to produce more eggs than they normally lay in a nest. Wynne-Edwards pointed out that this limited reproduction is not predicted by individual-level selection, as an individual's fitness increases with the reproduction rate, but that it is predicted by group selection. He argued that populations of individuals who produce too much offspring will overpopulate their habitats, use up their resources, and die out. Therefore, he concluded, individuals produce fewer offspring than physiologically possible in order to save the viability of the group.

With regard to altruism, Wynne-Edwards (1962) noted that there are many species where some individuals do not reproduce, but rather assist the breeding of other individuals. For example, worker bees do not procreate, but rather care for the offspring of the queen bee so that her reproduction rate is much higher than it could be otherwise. Wynne-Edwards argued that this kind of altruistic behavior decreases the fitness of the individual, but that it serves the group in such that individuals who help the reproduction of group members help the group to survive.

The ideas of Wynne-Edwards (1962) are similar to those of Lorenz (1973) who thought that the only meaning of life for any individual is in the conservation of the own species. But, bearing the principles of natural selection in mind, why should individuals—whether bird, bee or human—willingly make martyrs of themselves and does this make sense? Indeed, the theory of old group selection was often criticized later to commit a *naturalistic fallacy* (Moore, 1903/2008) such that Wynne-Edwards (1962) evaluated cooperative behavior on the level of the group as “good” or “desired” and concluded that thus it *had* to evolve. The naturalistic fallacy is related to the *is-ought problem* described by Hume (1739/2006) which states that many scientists make claims about what ought to be on the basis of statements about what is.

Going beyond global critics, empirical findings hardened the skepticism about Wynne-Edwards’ idea. For example, it was shown that reproductive constraint, in the way Wynne-Edwards proposed as a sacrifice to the group, does not exist. Rather, it turned out that individuals, who reproduce less than the physiological maximum at each point of reproduction, have a selfish reason to do so; reproduction and nurturing of offspring takes time and energy. If individuals had a large number of offspring at a time, it was likely that they might not have time to care for them. Consequently, not all of the progeny would survive. Moreover, caring for a great number of offspring decreases energy for survival and future reproduction and limited reproduction can thereby be advantageous in its own right. Note that the phenomenon of limited reproduction does not consequently need group selection for justification, but rather is explained by an individual’s pursuit of economizing his or her energy. As several scholars have shown, this way of prudent reproduction indeed maximizes an individual’s lifetime reproductive success (Lack, 1966; Krebs & Davies, 1993).

Now, since altruism—just like limited reproduction—should *neither* be accepted as a *purposeful* sacrificial act, could it nevertheless have evolved through group selection? As many biologists (e.g., Maynard

Smith, 1964; Williams, 1972) soon noted, this is unlikely. Even if a group with altruistic mutants existed, altruism could hardly prevail against selfishness for the following reason: group selection occurs slowly. Its effects can be seen only when groups of cooperative individuals have out-competed groups of selfish individuals such that the latter have gone extinct. This process is called "between-group selection." However, between-group selection takes time, during which inside of the groups traits can evolve through a process called "within-group selection" (i.e., individual selection). This process implies that selfish individuals out-compete altruistic individuals. Under normal conditions, the process of within-group selection for selfishness occurs much faster than between-group selection for altruism. Therefore, group selection is unlikely to prevail. Indeed, it only works under extremely restricted conditions, namely in small isolated groups. To illustrate this, Maynard Smith (1976) showed that group selection would fail as soon as one successful migrant existed. A migrant is an individual who disperses and reproduces in a foreign group and if this individual is selfish and disperses into an altruistic group, his or her selfish descendants can soon out-compete a whole group of altruistic individuals due to individual selection. Therefore, most biologists harshly rejected the idea of group selection in the 1960's and 1970's and the idea became a pariah concept, taught as an example of how not to think (Wilson, Van Vugt, & O'Gorman, in press).

4.2. "New" group selection or multilevel selection

In 1994, Wilson and Sober argued that the universal case against group selection, as it was originally formulated by Wynne-Edwards, had been overstated and was based on a misplaced emphasis on genes as replicators. The authors again theorized that social groups and other higher-level units may indeed function as "vehicles of selection" such that groups can be organisms in exactly the same sense that individuals are organisms. To illustrate their argument, they applied the following example (p. 587-588):

> Consider an imaginary population of rabbits inhabiting an island. A mutant arises that grazes more efficiently—so efficiently that a population of such mutants will overexploit their resource and go extinct. The mutation is adaptive in the limited sense of causing its bearer to have more offspring than other rabbits, but maladaptive in the larger sense of driving the population extinct. (…)
>
> However, if we imagine an archipelago of islands, only some of which contain the mutant strain, then populations driven extinct by the mutant can be replaced by other populations without the mutant. The population-level adaptation can now persist, but only because we have added a process of natural selection at that level; fit populations replace unfit populations in the same sense that fit rabbits replace unfit rabbits within populations.

At this point, a main difference between the old and the new concept of group selection is conveyed. While the old concept assumes that the driving force is selection at the level of the group only, the new concept assumes the existence of multiple levels of selection. Resurrected in this way, Sober and Wilson's new concept of group selection (Wilson & Sober, 1994; Sober & Wilson, 1998) is a unified theory of natural selection that operates on a nested hierarchy of units (West et al., 2007). Therefore, it is usually called "multilevel selection theory."

However, there are other differences between old and new group selection, where one concerns the definition of groups. The idea of new group selection is that, at certain stages of an individual's life cycle, interaction takes place with a limited number of individuals only and that, under these circumstances, cooperation can be favored (Wilson, 1977; Colwell, 1981; Wilson & Colwell, 1981). The new concept thus relies on within-population (intrademic) group selection, while the old concept focused on between-population (interdemic) group selection (Reeve & Keller, 1999). Note that the new concept of group selection thus incorporates ideas reminiscent of the concept of positive assortation. As Okasha (2005) put it, a central difference between the old and the new concept is that the old one looked at the evolution of group characters, while the new approach allowed for the evolution of individual characters within groups.

In Figure 1 (derived from West et al., 2007) the differences between old group selection (Panel A) and new group selection (Panel B) are highlighted.

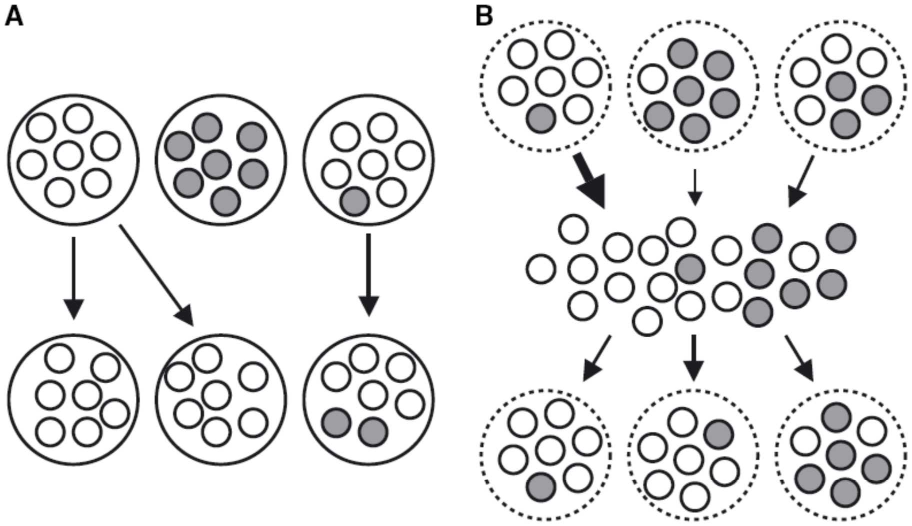


Fig. 1. The difference between old and new group selection. (West et al., 2007)

In Panel A, groups are well defined with little gene transmission between them (solid outline). The white circles symbolize cooperators, whereas the grey circles symbolize selfish individuals who do not cooperate. Competition and reproduction works between groups. The groups with more cooperators perform better, but the number of selfish individuals can increase within groups. Panel B shows the new group selection, with arbitrarily defined groups (dashed lines), and the potential for more genetic transmission between them. The different groups make distinct contributions to the same reproductive pool, from which new groups are formed. To elaborate on this idea, if there are many groups in the population that vary in the rate of altruists, the most altruistic groups will differentially contribute to the shared gene pool. Finally, although groups are arbitrarily defined in the new concept of group selection, there is the possibility of influences such as limited dispersal leading to more structuring.

To summarize, the theories of old and new group selection differ significantly from each other. The innovative way, in which new group selection (Wilson & Sober, 1994; Sober & Wilson, 1998) defines groups, as well as the fact that its multilevel approach is based on natural selection rather than self-sacrifice, has led to a much broader acceptance of the new concept of group selection than was previously achieved by the original theory of Wynne-Edwards (1962).

4.3. Individual selection and multilevel selection - mathematically the same?

Within the new concept of group selection, the evolution of cooperation depends on two factors: (1) the relationship between individual costs of cooperation and common benefits of cooperation, and (2) the genetic variance that exists within groups as opposed to between groups. The following will show that these two factors are similarly incorporated in the concept of inclusive fitness (see Chapter 3.1) and that individual and new group selection are at least “alternative ways of fitness book-keeping” (Gintis, Bowles, Boyd, & Fehr, 2008, p. 324).

Remember Hamilton’s Rule $b > c/r$ with b denoting the benefit of the recipient, c denoting the cost of the donor, and r denoting the degree of genetic relatedness between the recipient and the donor. Now consider the similarity between Hamilton’s rule and group selection. Note again that the first factor, which influences the evolution of altruism through group selection, is the cost-benefit ratio of cooperation. If the cost impact for altruists is weaker than the impact of cooperative benefits for the group (i.e., between-group selection is stronger than within-group selection), group selection predicts that cooperation will be favored. Indeed, this idea is reflected by Hamilton's Rule which implies that cooperation is favored if the benefits of the recipient (*b*) are increased and the costs of the individual (*c*) are reduced.

The second factor, which has been identified to influence the evolution of altruism in group selection models, is genetic variance within and between groups. Group selection implies that cooperation is favored when the proportion of genetic variance that exists between groups is higher than the proportion of genetic variance within groups (West et al., 2007). As Frank (1995) noted, this again is equivalent to a higher kin selection coefficient of relatedness.

Indeed, it has been known for a while that kin selection and new group selection are just different means to model the same evolutionary process; their mathematical structures are identical and both are thus theoretically valid (Hamilton, 1975; Grafen, 1984; Frank, 1986, 1998; Gardner, West, & Barton, 2007). Several scholars have shown that when both methods are applied to analyze the same problem, they lead to identical results (Frank, 1986; Wenseleers, Helantera, Hart, & Ratnieks, 2004; Gardner et al., 2007). However, does this mean that multilevel selection is as relevant as individual selection to real life?

4.4. Empirical evidence against multilevel selection

Even though kin selection and new group selection may lead to mathematically identical results in theory, there are empirical reasons why even the new concept of group selection is unlikely to be a relevant cause of evolution in nature. The crucial critique is that new group selection—as any form of natural selection—has to proceed through the differential reproduction of genetic entities and in this case the genetic entities are, besides individuals, groups.

The new concept of group selection presumes arbitrarily defined groups, but allows for processes like limited dispersal to structure groups. Now consider a situation in which limited dispersal was strong, i.e., related individuals would stick together. On the one hand, exclusive cooperation with kin was beneficial, but on the other hand, exclusive interactions with relatives would lead to incest and genetic degeneration. Consequently, a certain amount of genetic transmission between groups is needed and this is actually incorporated in the new concept of group selection. However, genetic transmission between groups also remains the flaw of the theory: a group of altruists is vulnerable to evolving toward a selfish one as soon as selfish migrants enter the group, because they will have the benefit of altruism without paying the costs. Thus, selfish individuals will spread rapidly.

To harden this argument, it should be noted that genetic transmission is not only a theoretical restriction but also an empirical fact. Migration can be observed in modern cultures and in hunting and gathering societies likewise (Richerson & Boyd, 2001; 2005). In addition, there is evidence from ancient cultures that suggests the existence of complex intermarriage systems across ethnic boundaries (Maenchen-Helfen, 1973). Above and beyond, the particulars of human warfare stand against the hypothesis of genetic group selection. As Keeley (1996) notes, anthropological evidence suggests that with the likelihood of war between societies, intermarriage increases. Additionally, some of the most brutal groups among pre-modern societies engaged in raids on neighbors for wife capture (Richerson & Boyd, 1998). Moreover, combat typically resulted in the social breakup of a conquered group, and the dispersal of its members to other groups (Soltis, Boyd, & Richerson, 1995). As defeated individuals form bonds with members of the victorious group, their genes—including selfish ones that might have contributed to their defeat—spread into the group of conquerors. Against this background of high rates of genetic transmission, it is hard to see how genetic differences between groups could have been maintained and altruism could have evolved due to group-level processes.

5. First conclusions about the powers of individual-level and group-level theories

Until now, the most important theories that have been developed to elucidate the paradox of altruism have been introduced. However, what about their practical relevance? Are these theories able to predict the altruistic behavior of humans in nature? Which theories are most important, and are some ideas wrong, while others correct? Which phenomena of prosociality can already be explained by them and which are still open to conjecture? To address these questions, this chapter aims at giving a résumé of the ideas that have been presented so far. By balancing the models against each other, first conclusions shall be drawn about the relative powers of the theories.

5.1. Why group-level theories have to be rejected

As indicated in the last chapter, the theories of group selection are lacking in practical relevance due to two reasons: (1) Old group selection has to be rejected, as individuals have never been shown to sacrifice personal fitness with the intent to serve the good of the group. The assumption that altruism evolved *because* it is good for the group is subject to a naturalistic fallacy. (2) The new concept of group selection, or multilevel selection, implies that genetic selection works, among others, at the level of the group. However, empirical evidence suggests high rates of genetic transmission between groups, so that genetic variation between them will hardly become sufficient to make group selection enforce its power.

5.2. What individual-level theories can already explain

As opposed to group-level theories, individual-level theories can explain quite a bit regarding the evolution of altruism. Kin selection postulates that helping relatives is beneficial and that the degree of helping should increase with the degree of relatedness. In Chapter 3.1, evidence was presented that indicates that humans indeed direct altruism preferentially towards kin. With regard to reciprocal altruism (see Chapter 3.3), systematic empirical investigation has been less frequent than with kin altruism. Nevertheless, there is evidence that although altruism has apparent short-term costs, it may pay in terms of reciprocal rewards in the long run (Axelrod, 1984, Nowak & Sigmund, 1998). Other indirect evidence, which supports the view that reciprocity is genetically determined, is that the norm of reciprocity seems to exist in every culture in the world (Schroeder, Penner, Dovidio, & Piliavin, 1995), although the degree of reciprocity varies across cultures (Henrich et al., 2004). However, kin altruism and reciprocal altruism do not sufficiently explain the whole dimension of altruistic behavior that can be observed nowadays in individuals.

The fact is that humans today cooperate with unrelated individuals in large and anonymous societies where the chances of repeated interactions are highly restricted. Thus, they are often willing to help others without any chance of reward. For example, by making anonymous donations, individuals reduce their own fitness and even forego benefits in terms of reputation gains. The mere anticipation of possibilities for indirect reciprocity might slightly promote this kind of large-scale altruism, but this explanation is not satisfying, as the fear of refused direct and indirect reciprocity would not suffice to discipline defectors. Individuals who preferred to behave selfishly would still gain the upper hand. To elaborate, nowadays anybody who wanted to betray others could do so and simply change his or her residence and social environment afterwards. Thus, selfish individuals would not have to bear the consequences of their misdeeds (i.e., social exclusion). They could simply start exploiting other individuals who are completely ignorant of their selfish character. Then again, selfish individuals could act up to the principle “Take the benefits and run.” To name one example, there are individuals who behave that way—marriage impostors try to get the most out of the new and anonymous world people live in, but these kinds of people are an exception. Most people show large-scale altruism even under conditions of anonymity. Why is that?

One might think that sexual selection (see Chapter 3.4) could help to clarify human’s extraordinary willingness to display altruism. Actually this reasoning is unsatisfying in almost the same manner as direct and indirect reciprocity, as it explains altruism in public—particularly in the presence of potential mates—but it does not explain altruism in large and anonymous groups.

One reason that might explain why individuals are willing to cooperate with strangers in large societies is assortation (see Chapter 3.2). As explained earlier, assortation denotes the selective association of like-minded individuals. If altruistic individuals were really able to identify one another (at first impression), they could avoid defectors. By achieving extraordinary benefits through mutual cooperation, the selfish advantage of egoistic individuals would be out-competed. There are two obstacles to this theory: first, it is questionable whether altruists are really able to identify one another; empirical evidence is needed. Second, even if such a mechanism existed, real-world observation suggests that it is far from perfect, otherwise cheaters like marriage impostors would not succeed.

Indeed, there might be another reason—apart from the possibilities of assortation—that leads people to show large-scale cooperation, and this might be related to the fear of sanctions connected to uncooperative behavior. This exact idea will be discussed in the next section.

5.3. What individual-level theories cannot explain

5.3.1. The phenomenon of strong reciprocity as some kind of ultra-sociality in humans

Gintis (2000) stated that cooperation in large groups might partly be due to an empirically identifiable form of prosocial behavior in humans, which he calls "strong reciprocity." A strong reciprocator is defined as an individual with a tendency to reward others for cooperative, norm-abiding behaviors and to impose sanctions on others for norm violations (Fehr & Fischbacher, 2003). A strong reciprocator is willing to impose these sanctions even if the norm violations did not affect the strong reciprocator himself but a third person instead. Rewards and sanctions executed by a strong reciprocator provide a fitness benefit to the group as these measures may support future cooperation and prevent future norm violations from other group members. The notion that cooperation in societies is supported by *external sanctions* is indeed far from being new, but a traditionally held view in sociology (e.g., Dahrendorf, 1958). However, performing such measures is costly to the actor because it may use up resources and time (Gintis, 2000). This exact notion, that sanctioning others is far from being selfish, is captured in an alternative label for sanctioning behavior by strong reciprocators; such behavior is also called "altruistic punishment."

It should be highlighted that *strong* reciprocity is different from *indirect* reciprocity. In the case of indirect reciprocity, which has been proposed by Alexander (1987) and formalized by Nowak and Sigmund (1998; 2005), individuals react to the behavior which other individuals have shown in pairwise interactions with third parties. They respond to cooperation with cooperation, and to non-cooperation with non-cooperation. The crucial aspect of indirect reciprocity is its connection to reputation gains. As Leimar and Hammerstein (2001) speculated, indirect reciprocity is not purely altruistic, but is favored by an individual's aiming for "good standing." In an experimental helping game, Engelmann and Fischbacher (2003) tested this assumption by giving only half of their subjects a public helping score and hence a strategic incentive to help. The authors showed that strategic incentives to build up a reputation increase an individual’s general preference for helping, but that reputation gains weaken the influence of *pure* indirect reciprocity. Bowles and Gintis (2004) consequently doubted that indirect reciprocity could be sustained in a population of self-interested individuals. Thus, they proposed strong reciprocation to be more likely to sustain cooperation, because strong reciprocators reward prosocial behavior and punish antisocial behavior despite the fact that these actions lead to net costs in terms of fitness. Where do we find strong reciprocation or altruistic punishment though?

5.3.2. Experimental evidence for strong reciprocity

In particular, strong reciprocation has been shown in many laboratory experiments that compose *ultimatum games*, *third-party punishment games*, and *public good games*.

In an ultimatum game, two players interact anonymously. They are given a certain amount of money, which has to be distributed among the two. The first player proposes how to divide the money between them, and the second player can either accept or reject this proposal. If the second player rejects, neither player receives anything. If the second player accepts, the money is split according to the proposal. The game is played only once, so that reciprocation is not an issue. A self-interested decider should accept any amount offered. Bearing this in mind, a self-interested proposer should offer the minimum possible amount. However, in a variety of studies deciders on average rejected offers below 30 %, while proposers mostly offered 50 % of the money at stake (Camerer, 2003). Obviously, the deciders were willing to impose altruistic punishment to those whom they perceived to be unfair. Anticipating this, proposers showed more cooperative behavior than predicted by a consequentialist model of self-interest.

Fehr and Fischbacher (2004) studied human's willingness to display altruistic punishment more extensively by conducting a series of third-party punishment games. Subjects were anonymously grouped into threes. While the first player decided about the split of money between himself and a second player, a third player, whose economic payoff was unaffected by the split, could impose sanctions on the first player. Although sanctioning behavior was costly, third players were willing to enforce fairness norms through punishment. The results thus showed that humans tendency for strong reciprocity extends to the sanctioning behavior of unaffected third parties.

The public good game (Andreoni, 1988) has been designed to study cooperation in groups. The general logic is that each participant receives a certain amount of private money and is grouped with three other participants under conditions of total anonymity. All participants secretly choose how much of their private money to put into a public pot. They know that the amount of money in the pot will be multiplied by the experimenter and equally distributed among the participants afterwards. Public good games are played for several rounds. While the best strategy for the group is to put all money into the pot, the best strategy for the individual is to contribute nothing but nevertheless receive his or her share from the pot. The game resembles a phenomenon known under the term "the tragedy of the commons" (Hardin, 1968). It consists of a dilemma in which multiple agents act independently in their own self-interest thereby damaging a shared resource despite the obvious fact that it is not in anyone's long term interest for this to happen. In a meta-study of twelve public goods experiments, Fehr and Schmidt (1999) showed that in the early rounds, contributions typically level between 40 % and 60 % of the endowment, but that over time contributions deteriorate until in the final round most subjects contribute nothing. The decay of contributions can be explained by strong reciprocity. Subjects who initially play cooperatively become angry with others who contribute less than they do. To retaliate against free riders, they apply the only strategy available to them and reduce their own contributions (Andreoni, 1995). Indeed, many experiments in which subjects were able to punish non-contributors directly showed that they were even willing to follow this strategy despite the fact that punishing implies a direct cost to themselves (Dawes, Orbell & Van de Kragt 1986, Sato 1987, Yamagishi 1988a, 1988b, 1992).

In summary, there is a wide range of empirical evidence from the laboratory that shows that individuals are willing to bear costs to sanction others who are unwilling to cooperate. However, why do humans show such behavior? Neither kin altruism, reciprocal altruism, nor indirect reciprocity would predict altruistic punishment. Even though costly signaling would predict heroism when potential mating partners are present, it does not predict altruistic punishment in anonymous situations. So why do people engage in altruistic punishment? If assortation functioned perfectly, costly sanctions would be unnecessary. Altruists could exclusively interact with each other. By doing this, altruists would reject defectors and ostracism would function as the ideal form of punishment because ostracism of defectors includes no costs (except screening-costs).

Highlighting the fact that standard evolutionary theories are unable to explain strong reciprocity, an alternative group of researchers argued that these findings show some kind of ultra-prosociality (i.e., true genetically unselfish behavior). The most important advocates of this idea are Bowles, Boyd, Camerer, Fehr, Gintis, Henrich, and Richerson (e.g., Henrich et al., 2004; Gintis, Bowles, Boyd, & Fehr, 2005). To distinguish these researchers from those who believe in standard individual-level theories, the terms “alternative evolutionary camp” and “standard evolutionary camp” will be used in the following. Note that these names are for reasons of identifiability, but that they do not imply any valuation—at least for this moment.

However, is the notion that strong reciprocity is evidence for “true altruism” correct? Assuming that it is correct, how could such behavior evolve?

6. Explanations to human ultra-sociality

6.1. Strong reciprocity – only an experimental artifact?

As mentioned before, followers of the alternative evolutionary camp suggest that strong reciprocity is evidence for some kind of true altruism in humans or “ultra-sociality” as it was also called (Richerson & Boyd, 1998). However, followers of the standard evolutionary camp (Price, 2008; Johnson et al., 2008) have censoriously commented on this argumentation. They dispute that even if a behavior seems individually maladaptive in the context of an economic game, it is illegitimate to assume that it would have been maladaptive in ancestral contexts, since the environments of experimental games are different from those in which cooperative behavior evolved. Therefore, strong reciprocity found in the laboratory should not be evaluated as truly altruistic (Johnson et al., 2008; Price, 2008).

To track the disagreement between the two camps, note that most evidence for strong reciprocity results from ultimatum games, third-party punishment games, and public good games. These games are characterized by anonymity and one-shot interactions (with the exception of the public good game, which is played for several rounds) and are designed with the intent to remove any incentive to engage in costly punishment or cooperation. Followers of the alternative evolutionary camp argue that according to the *selfishness axiom* of traditional economics and standard evolutionary theory, game players who are solely interested in maximizing their personal benefits should transfer the minimum sum possible and accept any amount of money offered to them. As it was seen in Chapter 5.3.2, individuals do not behave in the designated selfish way, but even engage in costly altruistic punishment. The alternative evolutionary camp thus argues that true altruism exists.

Followers of the standard evolutionary camp sharply distance themselves from this interpretation (Johnson et al., 2008; Price 2008). For one thing, they note that most evidence for strong reciprocity stems from the field of experimental economics, which has low ecological validity. In addition, they state that experimental findings of strong reciprocity are—unlike the alternative camp alleges—very much in accordance with the predictions of standard evolutionary theory. Standard evolutionary theory does *not* strictly adhere to the selfishness axiom, it does *not* view the mind as a fitness-maximizing device (Price, 2008). Instead, standard evolutionary theorists regard the psyche as an *executor of adaptations* (Tooby & Cosmides, 1992). A cognitive adaptation has to be understood as a semi-autonomous "if, then" mechanism. If a specific stimulus is present (e.g., a snake), the psychological mechanism responds in a specific way (e.g., by making the individual behave carefully). Although such mechanisms usually executed adaptive responses in the environments in which they evolved, today they can only execute their established protocols and do not know how to generate adaptive behavior per se (Price, 2008). In other words, the human psyche is incapable of optimizing behavior in evolutionarily novel situations (Johnson et al., 2008). Therefore, cognitive adaptations will produce their reactions in any environment that provides the sort of stimulus to which they were evolutionarily connected, even if a novel aspect in the environment leads the adaptation to fail its originally fitness-enhancing purpose (Burnham & Johnson, 2005).

What does this mean for the interpretation of findings from economic games? Bearing in mind that humans lived together in small groups where anonymity was rare, it might be the case that humans mistake the circumstances in economic games (which are explicitly anonymous and unrepeated) for repeated interactions (Johnson et al., 2008). This could explain why they behave as if they were in an ongoing reciprocal exchange.

Extending this argument, Hagen and Hammerstein (2006) stated that even if one-shot encounters with strangers existed in past times, an individual who met someone unfamiliar could not know whether this meeting was really a nonrecurring one. A single meeting might have caused an increased likelihood of additional encounters in the future. For example, a hunter who encountered an unfamiliar individual from another tribe when chasing a deer could not be sure whether he would meet the person a second time on another hunting occasion, so that it would be better to cooperate at the first go. Moreover, every meeting consists of many mini-reciprocal exchanges occurring over seconds or minutes (Trivers, 2004). In past times, these mini-exchanges might have been life-and-death issues in that one could either make a new friend or be killed. Therefore, humans might have evolved to exhibit a general tendency to reciprocate.

Hagen and Hammerstein (2006) highlighted another crucial aspect. In economic games, subjects are told to participate in one-shot-interactions. However, this does not guarantee that subjects interpret the situation in that way. Often, subjects in experimental economic games sit in computer labs where they participate with other subjects simultaneously. Although the identity of the players is masked so that they cannot know the specific identity of their partners, they do know that some attendants have been allocated to them (Price, 2008). The potential victims thus become identifiable, and it might be that subjects interpret their game partners as fellow citizens or fellow students (i.e., in-group members) who might be met again.

What do these arguments tell us, at the very least? All of the objections of the standard evolutionary camp are valuable, as they remind researchers to interpret their experimental findings carefully. However, they are not strong enough to eliminate the fact that strong reciprocity cannot be explained by standard individual-level theories—for three reasons.

First of all, in contrast to the doubts of Johnson et al. (2008), it seems likely that anonymity was given in ancient times. Why else would evolutionary researchers, concerned with the issue of mating for decades, preach the importance of a man being sure that a woman was trustworthy enough not to deceive by passing off another man's child as his? Even though anonymous situations, as is the case with extra-pair matings, were rare, self-interested behavior would have been advantageous in *any* situation not observed by others. The loss of resources due to gratuitous altruistic behavior would have been too large to survive the pressures of natural selection. Instead, evolution should have favored the capability of selectively adjusting the performance of self-interested and altruistic behavior according to the risk of being detected by others. Actually, humans seem to have such adaptations, as participants in economic games are apparently able to tell the difference between anonymous and non-anonymous situations and adjust their behavior accordingly (Andreoni & Petrie, 2004).

There is a second reason why experimental evidence for strong reciprocity should not be discounted due to doubts concerning the specific features of economic games. Even if economic games activated psychological adaptations for reciprocity and this was the reason why individuals showed higher altruism than predicted by the selfishness axiom, why on earth should individuals engage in costly altruistic punishment to take vengeance for unrelated third parties? This phenomenon cannot be explained by the individual-level theories.

Finally, experimental findings of strong reciprocity are confirmed by phenomena that can be observed in natural settings. In many circumstances, people hit back against others, bearing private disadvantages, although the possibility of gains through future interaction is remote or zero (Gintis, 2000). Betrayed lovers, for example, get revenge at great personal costs, although there is no future horizon of interaction with their ex-partners. Admittedly, this specific behavior might serve to restore balances of power between the two ex-partners, but there are other examples of individually maladaptive behavior. In the Nazi period, citizens were willing to save the lives of unknown Jews even though this included the risk of punishment and deportation to themselves. Today, individuals time and time again try to save others from dangers like drowning by risking their own lives. Many people are willing to associate in movements for civil rights or political democracy to improve living conditions for the society they live in, although they themselves have higher costs than benefits from these actions. All such activities include a considerable cost to the actor. Therefore, Bowles and Gintis (1986) emphasized that they cannot usually be explained in terms of self-interest or reciprocal altruism.

If Bowles and Gintis (1986) were right, from the perspective of the individual, the type of altruism exemplified above had to be judged biologically *unselfish*. It had to be judged *maladaptive*. It had to be assumed that it was driven by some kind of “conscience” leading individuals to behave in accord with social norms of morality that do not fit their own biological interest. Yet, how could this kind of conscience or “superego”, as Freud called it, evolve?

Note again that at the beginning of this work, it was stated that this analysis must solve the puzzle of altruism by using an evolutionary viewpoint, which focuses on ultimate causation and requires adaptiveness. However, at this point in the examination, it seems that we do find some kind of behavior that is individually maladaptive. Therefore, the following two alternative theories will offer reasons for why individually maladaptive altruism might have evolved nonetheless.

6.2. The parent-offspring conflict and the evolution of conscience

In 1995, Voland and Voland noted that the evolutionary approach, which explains altruistic behavior by using the perspective of the selfish gene, may explain many phenomena related to kin altruism and reciprocity, but that it obviously does not suffice to explain some other forms of human prosociality. More specifically, the authors argued that there are situations in which individuals face moral dilemmas and have to decide whether to respond in an altruistic way, which leads to individually-maladaptive solutions, or in a selfish way, which may lead to feelings of guilt that may impair a person’s life as well. If individuals try to avoid feelings of guilt, they will behave in a way that is ethically appropriate and eases their conscience, but this will often be biologically maladaptive.

Pointing out that moral conflicts are due to human conscience, Voland and Voland (1995) developed the hypothesis that the conscience does not suit the selfish-gene interests of the individual possessing a conscience, but the selfish-gene interests of this individual's parents.

To elaborate, the authors assumed that the conscience evolved due to the conflict that parents and their offspring face with regard to altruistic tendencies. As parents and their offspring are related by descent, they share 50 % of their genes. Therefore, their reproductive interests are quite similar, but not completely identical. Due to this reason, parents and children may develop divergent views on how to behave adequately; more concretely, parents will demand more altruism from their children than the latter are willing to display (Trivers, 1974, 1985).

Voland and Voland (1995) elucidated the logic of divergent interests between parents and their offspring using the example of the “weaning conflict”: Female mammals nurse their offspring as long as the cost-benefit ratio is not negative, i.e. as long as nursing increases the inclusive fitness of the mother. After some time, the baby is old enough to survive without the mother’s milk and she stops lactation to invest in another child. Although the baby at first starts crying and insists on being nursed, after a restricted transition phase during which the interest of the baby and its mother are divergent, the baby stops demanding milk. This makes sense, as the baby also shares 50 % of its genes with its full siblings. Therefore, the baby would harm its own inclusive fitness if it circumvented the production of siblings.

The weaning conflict is a typical example of a short phase of divergent interests between parents and children. However, Voland and Voland (1995) explained that such phases can be longer. In societies that install “helpers at the nest”, parents demand their offspring to assist in raising their brothers and sisters, instead of encouraging them to quickly start their own reproductive careers (Emlen, 1991). Citing case studies from European social history (Dickemann, 1979; Hager, 1992; Vernier, 1984), Voland and Voland (1995) pointed out that the insistence on such helping behavior sometimes leads to life-long role differentiations within families. In some cases, it was found that daughters especially were permanently excluded from private economic and reproductive opportunities and pushed into a helper’s role, which served the interests of their parents and siblings, but contradicted their own inclusive fitness. As Voland and Voland (1995) noted, parents’ interest in manipulating their children to behave altruistically may sometimes go beyond the wish to create helpers at the nest; parents at times gain advantages by sacrificing some of their children to the common good as a form of a “tax”. For example, they arrange marriages against their children’s interest to form political alliances (Podolefsky, 1984) or push their sons into the role of soldiers to serve the defense of the group.

As Voland and Voland (1995) stated, in all these cases, parents put their interests through much more effectively, the more disposed the deprived children are to accept the roles assigned to them. Optimally, children fulfill their parents’ wishes without direct parental control and pressure. This succeeds best by an upbringing which aims at suppressing the child’s ego (Voland & Voland, 1995, p. 406):

> A *conscience* [italics added] serves just this purpose. In case of a conflict, it regulates behavior to the child’s own disadvantage—even if powerful punishing parents are unable to exert direct influence. The conscience remains their “extended” arm; it cannot be expunged and its potency is not easily reduced. It is a “satellite” of the “selfish gene” of the parents; and whenever the situation requires, the conscience is willing to provide impulses for ethical and altruistic behavior—even long after the death of the parents.

Summing up the idea of Voland and Voland (1995), the conscience is assumed to be an “extended phenotype” of parental genes, which rules on the offspring’s behavior in a lifelong way, even when there are no longer any direct opportunities for parental manipulation. But one question arises: if the conscience may thus lead to self-damaging behavior in children—why do children tolerate this kind of exploitation executed by their parents?

Voland and Voland (1995) stated that human children are extremely helpless when they come into the world. They are dependent on their parents’ support. Highlighting the complexity of human culture, the authors argued that children need to be taught by their parents to acquire culture-specific skills because cultural conformity is necessary for humans to master their lives in reproductively flourishing ways. To become successful, children are inclined to adopt cultural norms and role models, and they fare well by trusting their parental teachers more than any other member of the group, as parents are genetically related to their children and thus have similar interests to them. The price of children’s trust is that they cannot discern in which cases parental nurture goes back to common interests and in which cases it is used as a trick to manipulate them in a way that serves their parents interests but is detrimental to themselves.

Of course, the social learning opportunities connected to parental guidance can be deemed highly important. It seems feasible that children will be interested in enjoying this support. However, should evolution not have favored children's ability to assert themselves against adverse parental manipulation? Voland and Voland (1995, p. 408) argued against this notion:

> (N)atural selection will breed "learning ability" for as long as there are net benefits to social learning despite the price of manipulability. An "ontogenetic autarchy" with immunity against the teaching influences would not have a chance of persisting in natural selection at the advanced level of primate evolution.

To highlight, the authors' statement implies that a child's obedience is a necessity for its survival if the child belongs to the group of "advanced primates". Actually, this statement slightly disagrees with the overall message of Voland and Voland (1995). In general, the authors argued that an individual's conscience contradicts his or her selfish-gene interests. However, taking the preliminary statement seriously by assuming that children would at worst die if they decided to waive the help of their parents, it can at least be argued that children's willingness to accept their parents' manipulations is ultimately in the interest of their selfish genes, as otherwise they would probably not develop to reproductive age.

Admittedly, there are individual cases in which the conscience really leads to a net cost for the individual possessing the conscience. This is true for suicide assassins, for example, who definitely do not increase their lifetime fitness by killing themselves for maniacal reasons produced by their diseased minds.

To conclude, Voland and Voland (1995) disputed that the conscience evolved as a result of an extended phenotype of parental genes. Taking the long phase of children's dependency on their parents into account, which noticeably distinguishes the human race from other animals, the authors' idea seems prudent. The extraordinary dependency of human offspring could be the reason why human altruism is much more pronounced than the altruism of other creatures that are less dependent on their parents.

However, with regard to the authors' interpretation that their theory contradicts the biological selfishness axiom, a different conclusion seems admissible (and is eventually supported by their own words). Children's willingness to adapt a conscience can be interpreted as a cost that children are willing to pay to get parental support in order to survive. Children are in need of acquiring social skills from their parents, because these skills are essential to master complex environmental and cultural challenges. In this vein, a conscience serves the survival of the genes of the individual having the conscience and thus increases his or her lifetime fitness.

Note that Voland and Voland's (1995) argument encapsulated another striking factor. The authors assumed that children's pressure to submit to their parental teachers is, amongst others, sustained by the complexity of culture to which children have to be habituated. This idea that the complexity of human culture might have contributed to human ultra-sociality will be discussed in the next section.

6.3. Gene-culture coevolution

The Greek philosopher Aristotle (384 BC – 322 BC) emphasized the excellent powers of law as a cultural instrument to underpin the polis and positively mediate relations between people with the following words: "Lawgivers make the citizen good by inculcating habits in them, and this is the aim of every lawgiver." (Aristotle, n.d./1962, p. 1103). Following the example of the ancient Greeks, who were characterized by a complex culture already hundreds of years before Christ, civilized societies today have established a variety of complex institutions, including law, in which all of them are designed to support civil obedience and cooperation. Actually, institutions are not only found in highly developed societies. Even primitive cultures show primal forms of institution-building, such as monogamy, and this suggests that humans have tended to enforce social rules and norms from early times. However, if humans have installed institutions for such a long time, it seems fair to ask whether the analysis of the evolution of human altruism may really disregard the parallel evolution of human culture.

The theory of *gene-culture coevolution*, also known as the *dual inheritance theory*, was developed in the late 1970s and early 1980s. Its central idea is to explain human behavior as a product of two interacting evolutionary processes: genetic evolution and cultural evolution (e.g., Boyd & Richerson, 1985; Cavalli-Sforza, & Feldman, 1981; Lumsden & Wilson, 1981). The concept, which will be discussed in detail in the following, can thus be understood as a symbiosis of the social sciences, which focus on culture as the primary cause of human behavioral variation, and human sociobiology and evolutionary psychology, which interpret culture as an unimportant by-product of genetic selection.

To highlight the importance of the analytical turn that is made in this chapter, note that the concept brings along two new implications that sharply distinguish this approach from standard evolutionary theories: (1) It is assumed that traits are influenced by their cultural environment inasmuch that genes, which are responsible for prosocial traits, adopt to the complexity of human institutions. Therefore, altruistic behavior, which at first glance seems biologically self-destructive, may turn out to be self-serving if it is evaluated against the cultural background in which it is displayed. (Gintis et al., 2008). (2) The theory supposes that besides the existence of genetically based traits, traits have evolved that are based on *cultural* transmission only. These traits may adapt much more rapidly to changes in the environment than genetically based traits and thus can have a great impact on human behavior (Boyd & Richerson, 1985). In the following, the basic ideas of gene-culture coevolution will be outlined and it will be discussed in how far the approach has added insights to the evolution of human altruism.

6.3.1. Basic ideas

The theory of gene-culture coevolution is strongly pushed by the group of "alternative scholars" mentioned before, who argue that the dimension of human ultra-sociality, which is, amongst others, indicated by the phenomenon of strong reciprocity in the laboratory, cannot be explained by standard evolutionary theories. To solve the puzzle of the evolution of altruistic cooperation and altruistic punishment, the authors fall back upon the idea of group selection (see Chapter 4). However, being aware of the close argument that genetic transmission renders *genetic* group selection improbable, they distinguish themselves from the ideas of genetic group selection in that they propose a selection mechanism that works on cultural rather than genetic variation between groups.

To sketch the basic idea, they argue that ultra-social behavior (i.e., biologically unselfish behavior that is not explained by standard evolutionary theories) was favored by cultural norms, which in turn evolved due to specific environmental conditions that demanded cooperation. The presence of divergent norms led to the development of divergent groups, that is, groups with many altruists as opposed to groups with many self-interested individuals. As groups of altruists fared better than groups of selfish individuals, the individual fitness losses of altruists were more than compensated by the superior performance of the groups in which they assembled. Consequently, altruism could evolve due to group selection on cultural variation.

Two problems arise: First, why should cultural group selection have a greater empirical relevance than genetic group selection? Are there specific characteristics that distinguish the two processes in a way that bolster the powers of cultural group selection? Second, how could culture, if it really opposes the genetic interests of genes and individuals, ever evolve? Both issues will be considered in the following.

6.3.2. Similarities of genetic and cultural transmission

Before reviewing the significant differences between cultural and genetic transmission, it should be highlighted that the fundamentals of the two processes are similar. Cultural evolution describes processes of passing over and adopting traditions via imitative learning. It implies that a potentially fundamental determinant of behavior is given from one individual to another. Considering this fact, Campbell (1965, 1975) as well as Baldwin (1895) argued that the principles of cultural and genetic evolution are comparable.

Richerson and Boyd (2001) highlighted the analogy between genetic and cultural evolution by stating that both concepts are historical processes, which appear at the population level and regularly result in the adaptive diversification of human manners. To follow this argument, consider how cultural evolution works over time: individuals retrieve determinants of behavior from a large population of potential parents and cultural models. Then, selection pressures operate on individuals, preferring some cultural and genetic variants against others. The new generation, from which the subsequent will be sampled, typically differs slightly from the previous one. As many generations pass, adjustments accumulate and evolution occurs. Emphasizing that genetic evolution works in the same way, Richerson and Boyd (2001, 2005) demanded that the great analytical powers of evolutionary theory, which have yielded substantive insights in the field of biology, should be likewise applied to the social sciences.

6.3.3. Differences between genetic and cultural transmission

Although genetic and cultural transmission are similar in their fundamentals, Richerson and Boyd (2001, 2005) pointed out four substantive differences between the two processes. First, when adopting a cultural trait, humans are not restricted to sampling just two "parents", as in the case for genetic transmission, but they can choose to imitate out of a great variety of people an individual whose behavior seems to be the best. In this way, cultural transmission allows single persons like teachers, leaders, or celebrities to have a great impact on great numbers of people if they all decide to imitate them. Indeed, pop stars frequently "infect" whole generations of adolescents to follow their ideas. Hence, cultural transmission may much more rapidly create behavioral variance between groups than in the case for genetic transmission.

Second, when cultural transmission takes place, individuals are not restricted to acquiring behavioral styles from individuals of their parental generation. Instead, they can choose models from completely different cohorts. For example, they may go back in time and imitate the behavior of grandparents or ancient leaders. This implies that outdated behavior is revived. However, they can also imitate behavior of peers. This implies a shortening of the life cycle of an item of culture. In addition, this difference implies that cultural transmission can enforce its powers much more quickly than genetic transmission.

Third, while humans are stuck with the genes they inherit at conception, they may acquire new habits and cultural traditions throughout their lives. In fact, culture is acquired step by step, with the possibility of early-acquired items influencing those adopted later, and for later learned manners to reject old ones. To illustrate, many people change hobbies, occupations, or political and religious beliefs repeatedly over a lifetime. Therefore, cultural transmission, in contrast to genetic transmission, allows short-dated adaptations to substitute behavior that was revealed to be disadvantageous—all within one and the same generation.

Fourth, with genetic transmission, behavioral variations can only arise once within each individual—in the very moment the embryo is formed, the genetic structure of the individual is fixed. In culture, variations can be acquired via processes of learning (i.e., imitation). Therefore, variations occur much more often. In human culture, the results of learning in one generation are passed on to the next. Thus, over the generations, cumulative improvements arise by the inheritance of acquired variation.

Summing up the argument of Richerson and Boyd (2001, 2005), cultural transmission is less restricted and quicker than genetic transmission, as people may (freely) choose the models they imitate and can vary behavioral styles frequently. They can adapt to new environmental challenges at short notice. As behavioral improvements are passed on to the next generations, younger generations can build upon improvements from earlier generations.

However, Tooby and Cosmides (1992) denied the importance of cultural transmission. In their critique, the authors distinguished between “epidemiological” and “evoked” culture. While epidemiological culture refers to the kind of cultural traditions that have been sketched above, evoked culture refers to differences that are not transmitted at all, but rather evoked by the local environment. To specify, Tooby and Cosmides (1992) argued that much of cultural variation is overestimated as being epidemiological, but to their view, all humans are endowed with the same genetically transmitted information, which simply generates divergent responses depending on the individual’s surroundings. Fending off this critique, Richerson and Boyd (2005) exemplified that if culture was only a genetically based automatic response to environmental challenges, a modern city slicker who was abandoned in the desert should as easily survive in aridity as someone native. But this is not the case. The instance thus shows that cultural transmission may generate survival benefits.

However, as it was discussed earlier, nothing initially develops *because* it leads to a sophisticated result that is advantageous at the group-level at a later point of time. Therefore, which circumstances favored the development of culture?

6.3.4. Environmental prerequisites for cultural transmission

Following Boyd and Richerson (1996), the prerequisite for establishing culture and institutions is some kind of specialized cognitive machinery, which allows individuals to imitate complex traditions. However, there are two reasons why these capacities had a hard time to evolve (Richerson & Boyd, 2001; 2005): first, the cognitive capacity to imitate traditions is a great fitness benefit, but only if there are cultural traditions to take advantage of. Obviously, there cannot be complex traditions without the cognitive tools necessary to support them, i.e., the “useless-when-rare-problem”. Second, the costs of having an elaborate culture usually exceed the benefits. Consider the behavior of most animals, which is guided by a mixture of genes and individual experience, but does not rely on culture. These animals obviously never developed culture, potentially because it was too expensive. Hence, under which circumstances could the development of cultural tradition have been beneficial?

Richerson and Boyd (2001, 2005) feasibly assumed that the genetic system is less likely to be affected by random transmission errors (mutation) than cultural tradition. They also assumed that in contrast to cultural transmission (i.e., learning via imitation), individual learning on the basis of own experiences is either costly or error prone. These two variables are closely related, because an increase in time and effort spent on learning usually leads to a decrease in mistakes.

On the basis of these assumptions, a slowly changing environment can be supposed to favor a fixed genetic rule rather than any combination of individual learning and traditional imitation. This is because selection, acting on a conventional system of genetic heritage, adapts to slow environmental changes very well, while the errors connected to learning and imitation entail considerable disadvantages.

At the other end of the scale, when the environment is rapidly changing, any form of transmission from ancestors—either genetic or cultural—is worthless, as the world of parents is simply too different form the world of their offspring. Under these circumstances, an individual does best to rely on personal experience only, because this incorporates the greatest chance to adapt to the new environmental challenges.

However, in intermediate environments, some combination of individual and social learning will probably be most adaptive. To specify, culture is particularly advantageous in environments that are shifting a lot within tens of generations, but not too fast in any one generation. Under these circumstances, culture, by accumulating individual learning, can track environmental changes more rapidly than genes, yet cut back the costs and errors associated with individual learning. Based on these considerations, Richerson and Boyd (2001) argued that the cognitive capacities for social learning developed due to the challenges of a moderately changing environment. However, what is their historical evidence that these circumstances were given at any one time?

Indeed, several authors have argued that the progressive brain enlargement found in many mammalian lineages is a result of the increasingly variable climates of the last few million years (deMenocal, 1995; Potts, 1996; Richerson & Boyd, 2000). In particular, the strongly fluctuating Ice Age climates of the Middle and Late Pleistocene represented a challenging environment for which it seems fair to assume that both individual and social learning were advantageous (Richerson & Boyd, 2001). The environment of the Pleistocene may thus explain why many species refined their learning abilities during that time. It may also account for simple forms of social learning in humans that probably served as *preadaptations* to the evolution of a capacity for complex traditions.

Yet, why did only humans, but no other animals, transcend simple forms of social learning to develop complex and cumulative cultural traditions, especially considering the intrinsic barrier to complex culture of the useless-when-rare sort? As Richerson and Boyd (2001, p. 450) note:

> We are on the horns of an explanatory dilemma. We must account for an evolutionary innovation that causes the extraordinary success of the only species to have it. Our account must explain why our species has complex culture, and why no others do, despite presumably preadapted systems for simple social learning being rather common.

The question that has to be asked is: how could social learning in humans become more efficient than in other animals? According to Richerson and Boyd (2000, 2001), a possible answer to this question is the evolution of symbols, which will be discussed in the next section.

6.3.5. A cultural innovation: symbols as in-group-markers

Symbolic markers are a striking feature of the human species in that they explicitly define groups. Typical marks are languages and dialects, or ornaments and specific clothing (e.g., a football-shirt), but also particular rituals (e.g., putting up a maypole or lighting an Easter fire). Archaeological evidence suggests that the development of symbols was a major innovation for the human race because it increased human adaptive sophistication (Bettinger, 1991; Stringer & Gamble, 1993).

Boyd and Richerson (1987) developed theoretical models in which individuals use symbolic markers to decide whom to imitate. The authors showed that in environments with migration in which the best behavioral strategy is variable from place to place, a symbolic marker serves to avoid the effect of cultural diffusion of locally maladaptive traits from neighboring environments. For example, a very simple marker, which might have been important within small groups already, would have been to imitate those individuals who were successful in terms of economic resources and reproduction.

To conclude, with markers, social learning may have functioned much more efficiently, as individuals were able to identify those individuals who displayed the locally beneficial behavior. Even the isolation of ethnic groups did not have to be complete. Foreigners could be identified easily. Suspicion probably helped to screen out mistaken foreign ideas, but after strict examination, successful strategies could be adopted (Richerson & Boyd, 2000; 2001)

Extending this argument, Richerson and Boyd (2005) assumed that cultural evolution was not only favored by the imitation of the most *successful* strategies, but also by the imitation of the most *common* strategies. In several models, the authors showed that a frequency-dependant conformist imitation strategy of the type “When in Rome, do as the Romans do” is adaptive (Boyd & Richerson, 1985; 1996). Pointing to classic experimental evidence by social psychologists Sherif and Murphy (1936) that humans are indeed inclined to adjust their behavior to that of others, Richerson and Boyd (2005) assumed conformist behavioral strategies to be another important driver of cultural evolution.

6.3.6. Cooperation due to gene-culture coevolution

Summing up the ideas of gene-culture coevolution as developed by Richerson and Boyd (1985, 2000, 2001, 2005), it was argued that (1) cultural transmission works like genetic transmission but more quickly, that (2) in many mammals, including humans, the first capacities for social learning evolved as adaptations to the fluctuating climate of the Pleistocene, and that (3) more complex forms of cultural transmission—which are found in humans only—were favored by the evolution of symbols.

The theory assumes further, that specific environmental difficulties promoted cooperation. This supposition is shored up by recently collected data from pre-modern small-scale societies (Henrich et al., 2004). To give an example, it was shown that in environments with small fishes, individuals subsist by casting for small fishes singularly. In contrast, in environments without small fishes, where groups can only live on fish if they carry out big game fishing, individuals cooperate. Generalizing this finding, it might be that in ancient times in some environments, cooperation was of vital importance, while in others cooperation could be neglected. But in how far did environments calling for cooperation favor the evolution of ultra-sociality?

The degree of environmental harshness was probably related to both the degree of social learning, as well as the degree of demands for cooperation. It seems likely that a parallel development of culture and needs for cooperation thus led to cultural norms for cooperation. As long as norms for cooperation were simple—e.g. “Men have to ascertain that their elderly parents are catered with enough food.”—no complex psychological machinery was needed to fulfill the demand. However, as soon as multiple norms existed, the challenges accrued. For example, if a second norm said “On bad fishing days, all men have to distribute their catch of fish equally among the families of the tribe”, and a third norm read “Ill members of the tribe have to be granted relief by the means of additional food supply”, individuals had to balance the pros and cons of their acts. They had to decide which norm to follow first. Putting the different rules in a permanent order would have been insufficient, because individuals had to reconsider the specific circumstances every day. They had to weigh their own needs against the hardships of others: “How hungry am I? How hungry is my wife? What about my children? Do I have to give fish to my mother today—or did she find enough berries to satisfy her hunger? How much fish did the other men catch, and are the other families replete already? And how ill is Kurt? Is he so bad off that I have to help him?” In such circumstances, individuals faced moral trade-offs because their own fitness was at risk. They had to weigh their action alternatives and had to decide whether to act selfishly, thereby increasing their own fitness, or to fulfill the cultural norms by acting altruistically. Even if they decided to follow the norms, they had to weigh which norm to follow first. All these decisions are not only reducible to facts, but also demand reflection. Individuals were in need of specific psychological equipment to fulfill these tasks.

As Richerson and Boyd (2005) argued, the distinctiveness of human psychological machinery evolved in social environments that had been shaped by cultural processes. As soon as groups established systems in which prosocial norms were enforced by reward and punishment, selection should have favored specific social instincts. To elaborate, in many situations the reward for non-cooperation can be enjoyed directly, while the cost of punishment will be suffered later on. Thus, people who overvalue the immediate payoffs of non-cooperation, may fail to collaborate although this is in their own interest to do so. If generally cooperative behavior was favored in most social environments, selection should have favored more sophisticated genetically transmitted social instincts, like feelings of guilt, which predispose individuals to cooperate, in larger and more complex social groupings.

Accepting for the moment that the conscience evolved due to culture—how could it stand the pressures of natural selection? It has to be assumed that a mutant who was able to capitalize on cultural traditions and cooperation in *observed* situations, but was *not* burdened with an unconditionally norm-abiding conscience, which would lead him to sacrifice resources for the well being of others even in anonymous situations, must have invaded a group of conscience-burdened cooperators.

As mentioned previously, gene-culture coevolutionary theorists solve this dilemma by pointing to the powers of group selection (e.g., Richerson & Boyd, 2001; Gintis, Henrich, Bowles, Boyd, & Fehr, 2008). They argue that in the years in which the modern *Homo sapiens* emerged, ultra-social behavior became fitness enhancing, but only because groups with many true altruists functioned better than groups of selfish individuals. They assume further that true altruists suffered individual fitness losses, but that these were more than compensated by the enhanced performance of the groups to which they belonged. The theorists thus conclude that the evolution of altruism is not only due to personal benefits, like the ones incorporated in kin altruism and reciprocal altruism, but that it is due to benefits realized at the level of the social group.

Computer simulations have supported the theoretical ideas outlined above. Assuming a deme-structured population (i.e., a population in which organisms interbreed with one another and share a distinct gene pool), Bowles (2001), and Bowles, Choi, and Hopfensitz (2003) tested models with culturally transmitted group differences in social institutions and genetically transmitted differences in individual behaviors. The scholars showed that intergroup conflicts could explain the evolutionary success of both (1) prosocial behaviors towards unrelated members of one's group and (2) group-level institutions, such as food sharing, even if these institutions were costly to the groups adopting them.

6.3.7. Gene-culture coevolution as explanation to strong reciprocity

In chapter 5.3.1, it might have seemed that invoking strong reciprocity as an explanation for the maintenance of altruism simply created a new evolutionary puzzle: why do people incur costs to reward cooperation and punish selfishness thereby providing benefits to third parties? However, Boyd, Gintis, Bowles, & Richerson (2003) denied that this is a new puzzle and pointed to the fact that selection pressures against altruistic punishment are not as harsh as selection pressures against altruistic cooperation. The reason for this is that the payoff disadvantage of altruistic *cooperators* relative to defectors is highly independent of the frequency of defectors in the population, whereas the payoff disadvantage for altruistic *punishers* declines as defectors become rare, because then there is no need for punishment.

For illustration, a computer simulation of Boyd et al. (2003) assumes a population with defectors, cooperators and punishers. Cooperators bear costs of cooperation, punishers bear costs of cooperation *plus* costs of the execution of punishment. Defectors bear costs of suffering punishment. However, as soon as the costs of suffering punishment exceed the costs of cooperation, defectors change their strategy and cooperate. As sanctions have to be performed less frequently then, the costs for altruistic punishers decrease. Consequently, within-group selection against punishers becomes weaker, that is, the number of punishers will decrease only very slowly and their existence will sustain cooperation within the group. Because the variation of cooperation *between* groups is thus heightened, between-group selection is enabled to make its impact: groups with altruistic punishers will out-compete groups lacking altruistic punishers.

How far does this idea coincide with the evolution of culture? As Boyd et al. (2003) note, cultural institutions moderate within-group differences in individual success. Examples for such institutions are monogamy or food sharing, which have emerged and diffused repeatedly in a wide variety of ecologies during the course of human history. For example, monogamy implies an exclusive sexual relationship with a single partner. It therefore reduces mate competition: even those individuals whose attractiveness would theoretically allow them to acquire multiple mates are restricted to one mate, because culture narrows down their opportunities. Likewise, in societies that practice food sharing, individuals who perform particularly well in hunting and gathering are unable to fully capitalize on their talent as they are forced to contribute to the common good. Consequently, cultural institutions limit within-group selection against behaviors that are individually costly but serve the group benefit. Thus, small costs, which, for example, arise through altruistic punishment, carry less weight.

As Fehr and Fischbacher (2003) pointed out, in this way culture and altruistic punishment may bolster up one another: altruistic punishment ascertains the preservation of the culturally selected norms and cultural norms protect altruistic punishers from too harsh fitness disadvantages.

However, with regard to the practical relevance of altruistic punishment in foraging societies, it has to be noted that altruistic punishment probably implied much greater costs than assumed by Boyd et al. (2003). Altruistic punishment almost certainly means to offend somebody directly for behaving non-cooperatively. In contrast to executors in modern societies with complex systems of law, altruistic punishers in foraging societies were most likely less well protected when performing such acts. At worst, an insult would have let to death. Therefore, invoking primitive institutions like monogamy or food sharing as shields of altruistic punishment seems skewed. Actually, it appears that more complex institutions, like elementary forms of law, must be considered to strengthen the idea that altruistic punishment was supported through the establishment of cultural rules.

6.4. Summary: how ultra-sociality might be explained

At the beginning of this chapter, it was questioned whether human ultra-sociality, which has not been explained by standard evolutionary theories yet, might simply be a maladaptation to novel contexts. In particular, it was debated whether psychological adaptations for prosociality, which evolved in ancient environmental contexts, are triggered in novel contexts due to mistakenly interpreted cues of reciprocity and publicity, and thus lead to self-detrimental results. However, it was noted that this interpretation is not satisfying. Rather, it seems likely that human ultra-sociality is guided by a conscience, whose moral postulations are derived from extra-individual social interests entailed in cultural norms.

Two ideas were presented for how conscience could have evolved. Starting with a résumé of the more complex theory of gene-culture coevolution, it was assumed that humans' unique competencies in the development of culture added a distinctive direction and speed to the process of genetic evolution in that humans developed proximate mechanisms, which gave way to the evolution of individually costly but group-beneficial behaviors (Boyd & Richerson, 2004; Gintis et al., 2008). Richerson and Boyd (2001) remarked that humans today are thus determined by two forces—their biological drives on the one hand and morality on the other. These forces are often collaborators, but they are antagonists whenever biologically beneficial behaviors are morally objectionable. Drawing parallels to Sigmund Freud's theory of human beings being painfully torn between an animal id and a cultural superego, the authors put forward the idea that as soon as culturally immoral behavior generates too much psychic pain, genetic benefits will be neglected and humans will obey internalized cultural norms.

Indeed, this statement is strikingly similar to Voland and Voland's (1995) idea of the conscience as a force that evolved in the parent-offspring conflict and drives humans into moral dilemmas. The authors sketched a scenario in which helpless children follow the guidance of their parents in order to obtain their support and to enjoy the advantages of socialization. As they are unable to discern which parental educational measures redound to their personal advantage and which work in the interest of their parents but to their own disadvantage, they cannot protect themselves from manipulation in the form of internalized norms.

Both theories, the theory of gene-culture coevolution as well as the theory of the extended phenotype, are valuable, as they provide innovative ideas to enlighten the paradox of the evolution of complex cooperation in humans. However, both theories have assets and drawbacks.

The theory of gene-culture coevolution makes many assumptions, starting with the causal relationship of environmental conditions in the Pleistocene and the evolution of social instincts like guilt through to the assumption of group selection on variations of culturally determined traits. These assumptions have not been supported irrefutably with the aid of empirical data. As Johnson et al. (2008) remark, the principle of Occam's razor advises scientists to explain any phenomenon with as few assumptions as possible, or, to put it in the words of Williams (1966, p. 5): "[W]hen recognized, adaptation should be attributed to no higher a level of organization than is demanded by the evidence." With the application of group selection, the theory of gene-culture coevolution indeed assumes adaptiveness on a high level of organization. Theorists have to be careful not to leapfrog important levels in between.

Compared to the theory of gene-culture coevolution, the theory of the extended phenotype is captivating due to the simplicity of its presumptions. It seems likely that the struggle between parents and their offspring influenced evolutionary processes, because the parent-offspring conflict is a prevalent phenomenon: all creatures on earth have parents. However, this is a pro and a con to the theory at the same time. Voland and Voland (1995) do not make as sophisticated suggestions as the advocates of gene-culture coevolution with regard to the question why the conscience evolved in humans only, but not in other primates.

Moreover, Voland and Voland (1995) do not state explicitly why parents should be inclined to implant a conscience into their children, which does not operate only in favor of their personal interests, but also in favor of the interests of society. Admittedly, Voland and Voland (1995) argue that parents might have personal advantages by sacrificing their children to the social group they live in. But this again implies that this time parents fulfill some kind of social norm, which serves the group but is disadvantageous to themselves. To illustrate, a son who fights for the group simultaneously fights for his parents, but, taking over a consequentialist point of view, parents would do best if they acted as free-riders in that they let other parents' sons protect the group and assigned their own son the role of the warrior in front of their own doorstep.

Therefore, even on the grounds of the theory of the extended phenotype, it seems plausible that in some way the evolution of culture has intermingled with the evolution of genes. At the very least, culture can be assumed to have given input to a conscience that evolved due to still cloudy causes.

7. A call for an integration of evolutionary theories

For a long time, evolutionary theory was strongly associated with the selfishness axiom. As Price (2008) noted, this might have been triggered by the title of Dawkins' 1976 publication "The Selfish Gene". However the evolutionary theory behind this title does not assume that selfish genes produce selfish individuals. It simply assumes that selfish genes support their own replication by increasing the fitness of their bearers. An increase in fitness may sometimes be accomplished by making individuals behave selfishly, and other times it may be accomplished by making individuals behave altruistically. Anyway, as Price (2008) notes, the central idea of present orthodox evolutionary theory is that genes replicate primarily by increasing individual fitness.

What does this approach imply? It implies that the gene-centered framework is an extremely narrow level of analysis and that it remains fair to apply only as long as measuring the fitness of a gene is independent of its surroundings. As soon as a number of genes act jointly to produce a phenotypic effect (i.e., an effect on any observable characteristic or trait of an organism), a higher level of analysis—the individual—has to be considered.

Indeed, this idea is reminiscent of the central insight of Gestalt theory, which says that the whole is different from the sum of its parts. Max Wertheimer (1925), who is credited as the founder of the theory, postulated that the "Gestalt" (engl. 'shape' or 'figure') is *primary* and *defines* the parts of which it is assembled, rather than being a secondary formation that emerges from those parts:

> Es gibt Zusammenhänge, bei denen nicht, was im Ganzen geschieht, sich daraus herleitet, wie die einzelnen Stücke sind und sich zusammensetzen,

> sondern umgekehrt, wo—im prägnanten Fall—sich das, was an einem Teil dieses Ganzen geschieht, bestimmt von inneren Strukturgesetzen dieses seines Ganzen.
>
> (There are contexts in which what is happening in the whole cannot be deduced from the characteristics of the separate pieces, but conversely; what happens to a part of the whole is, in clear-cut cases, determined by the laws of the inner structure of its whole.)

To transfer the idea of Gestalt theory to the present issue, it can be argued that only if the individual, as a higher level of analysis, is factored into the evolutionary analysis, can the meaningfulness of genetically determined traits be evaluated.

However, if it is the case that the consideration of higher levels of analysis is necessary to solve specific questions, it seems fair to ask whether the level of the individual is the highest level that has to be taken into account when trying to explain the dimension of altruism in present-day people, or whether even higher levels of analysis—like the group and the cultural environment—are useful. To elaborate, the inclusive fitness of a gene may depend on the environment within which it is situated. Yet, the relevant environment does not have to be bounded to the body of the individual. For example, the cultural conditions of ancient times might have represented an environmental force that affected individual fitness in a way that promoted prosocial traits. As Diamond (2005) noted, the leading trend in human history has been the replacement of smaller, less complex societies by larger, more complex ones. Therefore, I would like to follow theorists of gene-culture coevolution by stating that neglecting of the higher force of cultural evolution could lead to fatal mistakes when analyzing the adaptiveness of genetic variations in modern human beings (see Gintis et al., 2008). However, an application of higher levels of analysis must always proceed with caution.

What does my clamor imply? Summing up the last chapters, various theories were presented to elucidate the paradox of human altruism. Standard evolutionary theories of kin altruism, reciprocity, and sexual selection were presented to explain distinct forms of human cooperation. However, human cooperation is highly complex, and it seems that some humans show true altruism, in that they behave in biologically self-detrimental ways. However, until now, it has not been explained why this is the case. The puzzle of the evolution of altruism still remains unsolved. It seems that biologically self-detrimental behavior is indeed governed by an originally extrinsic control authority, which Sigmund Freud called the “super-ego”, that controls the natural drives of our selfish “id”. To elucidate the evolution of this kind of super-ego, alternative theories —including group-level theories—should be considered. The development and discussion of any serious theories that try to elucidate the origins of human cooperation, are useful, as all of behavioral science and eventually the whole of humankind will profit when the puzzle of altruism is finally solved. As empirical findings showed that different theories may explain distinct forms of cooperation in humans, an integration of theories, rather than a battle between ideologies, seems to be the most fruitful way to master this challenge.

Commenting on the roots of human ultra-sociality for a second time, there are still many theorists who assume that true altruism is only a maladaptation to novel interaction contexts in that individuals are unable to adjust their degree of prosocial behavior according to the degree of anonymity. If this was true, it had to be assumed that natural selection, in the long run, would raze out true morality. In this case, we were approaching the evolution of *Machiavellianism* in such that opportunistic individuals, who behave altruistically when they are observed but switch to pure selfishness as soon as this is efficient, would gain the mastery. To put it in the words of Machiavelli (1514/2008) himself:

> For a prince (...) it is not necessary to have all the [virtuous] qualities, but it is very necessary to appear to have them (...). [It] is useful, for example, to appear merciful, trustworthy, humane, blameless, religious—and to be so—yet to be in such measure prepared in mind that if you need to be not so, you can and do change to the contrary.

If present-day individuals who follow Machiavelli's standard were better off indeed, true altruism had to be judged a maladaptation and true morals should die out in the long run. While this view might be right, I consider it too discouraging to share.

Therefore, the commitment model of Frank (1988), which assumes that ultra-sociality is identifiable and is able to prevail because true altruists may avoid to be exploited by egoists, is a ray of hope on the intellectual horizon. If altruists were really able to exclusively mingle with each other, they could reach extraordinary benefits through mutualism. These synergetic benefits would out-compete the advantages of egoists, and human morality could be sustained.

As announced at the beginning, two empirical studies will be presented in the following, which were designed to test whether the hypothesis of positive assortation can be supported in reality. In Study 1, whether humans can identify the altruistic tendencies of others by first impression was investigated. In Study 2, how well individuals know the true characters of their daily interaction partners and whether individuals indeed assort themselves along the dimension of altruism was investigated.

Although theories of assortation may explain why different degrees of altruism subsist within one and the same population, they do not explain why such inter-individual differences in altruism originally evolved. Therefore, the third study in the following empirical part of this work deals exactly with this question. It is discussed whether different levels of prosociality might be explainable with the help of mate choice theories.

As keeping others unnecessarily on tenterhooks comes near to selfishness, this preliminary theoretical consideration will end now, and empiricism shall bear witness to explanations for the evolution of altruism.

Part II: Empirical investigation

8. Study 1: Not only states but traits - Humans can identify permanent altruistic dispositions in 20 seconds[1]

8.1. Introduction

The existence of altruism in humans, which puzzled already Darwin (1871), has been a niggling issue for generations of researchers, namely, “how can altruism, which by definition reduces personal fitness, possibly evolve by natural selection?” (Wilson, 1975, p. 3). A range of theories has been developed to explain the phenomenon and it seems fair to assume that various forms of selection shaped the evolution of human altruism. For example, kin altruism (Hamilton, 1964) accounts for prosocial behavior towards relatives. Reciprocal altruism (Trivers, 1971) occurs in ongoing relationships with a sufficient “shadow of the future” (Axelrod, 1984) that includes opportunities to be compensated. Recent theories of indirect reciprocity (Milinski, Semmann, & Krambeck, 2001) and altruistic punishment (Fehr & Gächter, 2002, 2003) showed that humans are willing to reward altruism and punish selfishness even if the selfish behavior is not directed to themselves but to an unrelated person instead. Thus, social norms may lead to sustained altruism in large groups (Fehr & Fischbacher, 2003).

To investigate altruistic behavior, economists, and more and more frequently also psychologists and biologists, often use the so-called dictator game (for a summary of research see Camerer, 2003). In this paradigm, two persons interact with each other only *once* and under conditions of total anonymity. The dictator is given a certain amount of money and has to divide this money between him or herself and an

---

[1] A modified version of this chapter is currently under review for publication in Evolution and Human Behavior together with Detlef Fetchenhauer and Ton Groothuis.

unknown recipient. The recipient has no means to influence or to veto the decision of the unknown dictator.

Analyzing this situation—what result would the above stated theories predict as to the decision of the dictator? According to the theory of kin altruism (Hamilton, 1964), prosocial behavior should not be observed, as it can be presumed that the recipient in the dictator game is genetically unrelated to the dictator. Likewise, the theory of reciprocity (Trivers, 1971) would not predict any cooperation, as the interaction is non-iterated and does not allow for compensation. Because the interaction is anonymous, costs in terms of losses of reputation (Milinski et al., 2001) or punishment (Fehr & Gächter, 2002, 2003) are irrelevant. The dictator game thus constitutes a "golden opportunity" (Frank, 1988, p. 73) to act self-interestedly without any negative consequences. Conclusively, the dictator should keep all the money for himself.

However, a vast number of studies, both in industrialized societies (Camerer, 2003) and hunter-gatherer societies (Henrich et al., 2004), showed that the average amount of money given to the recipient is substantially above zero, even when large amounts of money are at stake (Camerer, 2003).

The "mismatch hypothesis" (Hagen & Hammerstein, 2006, p. 341) tries to elucidate this finding by stating that humans evolved in an environment different to the setting in an economic game (Trivers, 2004). It might be the case, for example, that humans cooperate in a dictator game because their psychological mechanisms evolved for reciprocal altruism. According to this concept cooperating on the first round is often a good tactic (Hagen & Hammerstein, 2006). Indeed, anonymity was rare in ancient times and it could moreover be that evolutionarily relevant cues in a dictator game might be processed in a manner that may be cognitively impenetrable, so that humans—

even when they are unobserved—might rely on a heuristic which implies that antisocial behavior is generally identified and punished.

However, this argument neglects the potentially strong selection pressure to enhance humans' ability to modify their behavior according to the likelihood of being detected and punished by others. In fact, humans are highly sensitive to such cues and adjust their decisions accordingly (Fehr & Henrich, 2003; Gintis, Bowles, Boyd, & Fehr, 2003; Haley & Fessler, 2005).

A second phenomenon, which could not be explained yet, relates to the great variation of humans to behave altruistically under one and the same circumstance (Camerer, 2003). By far not everyone behaves altruistically in a dictator game, nor does everybody behave egoistically. Persons rather show great differences in their tendencies to cooperate when nobody is watching. Why does the behavior of individuals diverge to such an extent?

To summarize, although existing theories of the evolution of human altruism have provided great insights, they are not capable of explaining two phenomena conclusively: first, that humans show such high levels of altruism in one-shot interactions with unrelated strangers under conditions of total anonymity (e.g. in dictator games (Camerer, 2003), in trust games (Fetchenhauer & Dunning, in press), and in prisoner's dilemmas (Sally, 2000)), and second, that there are some individuals who oppositely keep all the money for themselves.

One argument, which has been developed independently in the two fields of evolutionary psychology (Miller, 2000, 2007) and economics (Frank, 1988; 2004; 2008), may explain these findings by assuming that some humans, though not all, have a stable disposition to act altruistically even in the absence of incentives to do so. According to evolutionary theory, such individuals would pursue ultimate genetic self-interest through proximate psychological mechanisms that embody a genuine concern for others (de Waal, 1997; Nesse, 2001).

From the point of view of an ancient hunter-gatherer such persons would have been predestined to be a good mates, friends, or interaction partners (Miller, 2007) and should have been preferred against those individuals who conditionally act altruistically in situations of social surveillance only.

But how could a tendency for unconditional altruism ever have evolved? Following Miller (2000, 2007), moral virtues, including altruism, evolved as fitness indicators for mate choice. Hence, altruism is sexually attractive because it advertises individual fitness in terms of parenting- and relationship-coordination abilities. Moreover it promotes good genes as altruism depends on empathetic social intelligence which is absent if one has a high mutation load that impairs the development of a complex Theory of Mind. Therefore, altruism may function as a kind of neurogenetic warranty, which, according to costly signaling theory (Gintis at al., 2001; McAndrew, 2002) is hard to fake and can therefore remain reliable over time (Zahavi & Zahavi, 1997).

The economist Frank (1988, 2004, 2008) also suggests the evolution of dispositional altruism. In his commitment-model, he postulates that unconditional altruists are proximately driven by strong moral emotions about fairness and shame, which make them behave righteously on principle. Rather than focusing on utility in terms of sexual attractiveness, he implies that altruism evolved because it is directly advantageous in every type of relationship – at least as long as altruists avoid egoists but interact exclusively with each other in order to reap extraordinary profits through collaborative synergies.

The common postulate of these two theories is that individuals with a genuine concern for others display signals of their moral intentions. If these signals can be observed so that moral intentions are identified by others, high virtue individuals may profitably team up with one another leaving lower virtue individuals no chance but to stay among themselves. Such a mechanism could also explain the continuing variation in altruistic tendencies among humans (Frank, 1988). On the one hand, the higher the percentage of altruists in a given population, the less monitoring of the altruism of others will take place, and the more adaptive it is to act non-altruistically. On the other hand, the lower the percentage of altruists in a given population, the more monitoring will take place, and the more adaptive it is to act altruistically. This would lead to frequency dependent selection and an equilibrium of altruists and non-altruists existing side by side.

There is one obvious criticism of this argument. When signaling altruism is profitable, self-interested persons should be interested in faking the relevant signal in order to reap the benefits of an altruist without paying the cost of altruistic behavior (Fehr & Fischbacher, 2005). In evolutionary terms, this means that natural selection should create a deceptive copy of the altruistic signal. Though this argument is pressing, it can be reversed, as the existence of a deceptive copy should lead to the modification of the original signal in order to prevent plagiarism (Frank, 2005). An arms race between the true signal and its fraudulent copy should arise. For the present moment of our evolutionary history, the relative strengths of the true signals of altruism are unknown and it remains an open question whether and how far humans may signal their altruistic dispositions unmistakably so that others can reliably identify them.

This study aims at disclosing whether humans can identify altruism in others via first impressions whereas it shall be highlighted that the crucial question is not whether humans can read signs to altruistic emotional *states*, but rather whether they can identify stable cues to permanent altruistic *traits*. While states are short-lived and presumed to result from direct situational factors (Spielberger, 1972), traits are highly enduring and “emanate from within individuals” (Allen & Potkay, 1981, p. 917). Being a result of distant and causal factors, like genes and upbringing (Fridhandler, 1986), traits may explain inter-individual differences in behavioral responses to a given situation (e.g. Spielberger et al., 1973). In a situation, in which individuals have the choice to either act altruistically or selfishly, individuals with an altruistic trait can thus be expected to act prosocially.

The question whether humans may identify altruistic traits by first impressions has not been answered yet. Nevertheless previous empirical research hints at the ability of humans to predict altruistic behavior of other persons. Pradel et al. (in press) showed that school students are able to estimate the dictator game decisions of their classmates better than chance. Frank et al. (1993) as well as Brosig (2002) evidenced that people can predict the altruistic behavior of others after being acquainted for only half an hour.

Brown et al. (2003) went a step further and investigated whether cooperation is predictable by first impression alone. Target persons were videotaped and asked to indicate their level of altruism in a self-report. Subsequently, judges had to rate the target's level of altruism on the basis of these video-clips. The judges' ratings correlated significantly with the self-reported altruism of target persons. However, this method incorporated an honest signaling problem in so far as self-reports of target persons might have been beautified because of the targets' interest in cultivating a positive self-concept. Therefore, it remains unsolved in how far these results evidence the recognition of true rather than self-construed altruism.

This limitation does not hold for a study by Verplaetse et al. (2007) who also investigated predictive cheater detection abilities on the basis of first impressions but used a behavioral measure to assess the degree of altruism. Using pictures of completely unknown target persons who had played a one-shot prisoner's dilemma game earlier, the authors asked participants to rate how cooperative these target persons were. Results showed that individuals could clearly differentiate between cooperative and non-cooperative target persons, but only if they responded to an event-related picture which was taken during the decision-making moment rather than when they responded to neutral pictures or event-related pictures that were taken during a practice round. The authors concluded that humans subconsciously identify cues of non-cooperativeness which consist of facial expressions evoked by significant social decisions.

Summarizing these results, it seems that humans have a limited cheater detection module. However, in all of these studies participants could observe target persons either in or directly before the very moment when those grappled with the challenge of making the social decision in question. They thus could base their estimates on stimuli that entailed cues to the emotional states of the target persons. It therefore remains unclear whether the accuracy of predictions was supported by the identification of stable altruistic traits. Though these studies provided important insights to the validity of person perception they are not able to underpin the hypotheses of Miller (2000, 2007) and Frank (1988) as these authors assumed that humans are able to identify signals which indicate permanent altruistic traits, rather than cues to emotional states. In the present study we exactly address Miller’s and Frank’s assumption by investigating whether humans are indeed able to predict the altruistic behavior of strangers via the identification of permanent cues to altruism.

To test the hypothesis, we adopted the so-called “thin slices paradigm” (Ambady et al., 2000; Ambady & Rosenthal, 1993). This paradigm has been used to show that people are able to identify permanent characteristics of others, e.g. intelligence (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004), sociosexuality (Gangestad, Simpson, DiGeronimo, & Biek, 1992), or personality disorders (Oltmanns, Friedman, Fiedler, & Turkheimer, 2004) by watching short videotapes of target persons. We created silent videoclips of target persons lasting 20 seconds and afterwards asked them to play a dictator game, thereby measuring their true level of altruism. Subsequently, judges were asked to watch and rate the targets' level of altruism. By investigating predictions that grounded on stimulus material which was on the one hand completely unrelated to altruistic behavior, and on the other hand recorded *before* the dictator game took place, we could check whether individuals were able to make out permanent signals to altruism.

8.2. Methods

In step 1, we invited N = 56 students of business administration to the laboratory, where they were videotaped sitting behind a desk and in front of a white wall. The target persons were asked to talk into the camera and briefly introduce themselves. For each target person, the zoom was calibrated in such a way that the face and the upper body could be seen on the videos.

After being videotaped, the target persons were accompanied into a different room where they were left alone to fill in a questionnaire untroubledly. Target persons initially completed three personality scales, which amongst others served as distraction tasks. Afterwards, via the instructions target persons were familiarized with the logic of the dictator game without naming the game as such. Target persons had to imagine a situation with two individuals, a dictator and a recipient, in which the dictator receives a certain amount of money. Taking the perspective of the dictator, participants had to consider the distribution of the money between themselves and an unknown recipient. Neutral vocabulary was used all throughout the instructions to avoid influencing the participants in any possible way. For example, the dictator game was referred to as a “distribution task”, the dictator was referred to as “Person A”, and the recipient was named “Person B”.

Target persons were given 60 Euro (about US $ 94) and had four alternatives for dividing the money: they could either give 30 Euro, 20

Euro, 10 Euro, or nothing to the recipient. They were told that by a random mechanism, one out of six of them was making the decision for real. They were also informed that their interaction partner was another randomly chosen individual who participated in the study, but that this recipient would never be informed about the identity of the dictator who was accountable for his or her outcome. The students were asked to write a code word on their questionnaire and put the questionnaire into a sealed envelope. A week after the experiment participants could get their money at the secretary of the department by taking the envelope with their personal code word written upon.

On average, the targets gave 20.17 Euro (± 11.67) to the recipient. Twenty-seven targets (48.2%) split the money equally (i.e., gave 30 Euro to the recipient), 14 (25.0%) targets gave 20 Euro to the recipient, and 4 (7.1%) of them gave 10 Euro to the recipient. A substantial minority of 11 participants (19. 6%) decided to transfer nothing but to keep everything for themselves.

In the second step, the logic of the dictator game was explained to a group of 34 biology students of the University of Groningen. They were then shown the videos of the 56 different target persons. Each video lasted 20 seconds and the sound was switched off during the whole experiment. After each video, the judges had to estimate which of the four alternatives the target had chosen in the dictator game. Under no circumstances did the judges receive feedback on the correctness of their estimates. Moreover, on inquiry it was assured that none of the judges knew any target persons.

8.3. Results

The judges' average estimates of the amount of money given by the dictator to the recipient differed substantially between the target persons, ranging from 4.40 Euro to 25.65 Euro. Combining the different judges' estimates into one single scale resulted in a Cronbach's alpha of .84 (the mean inter-rater correlation was .14).

Indeed, the judges' estimates were accurate as actual and estimated dictator game contributions correlated significantly (*r* = .41, *p* < .01). Giving *z*-standardized values, Figure 1 highlights this result.

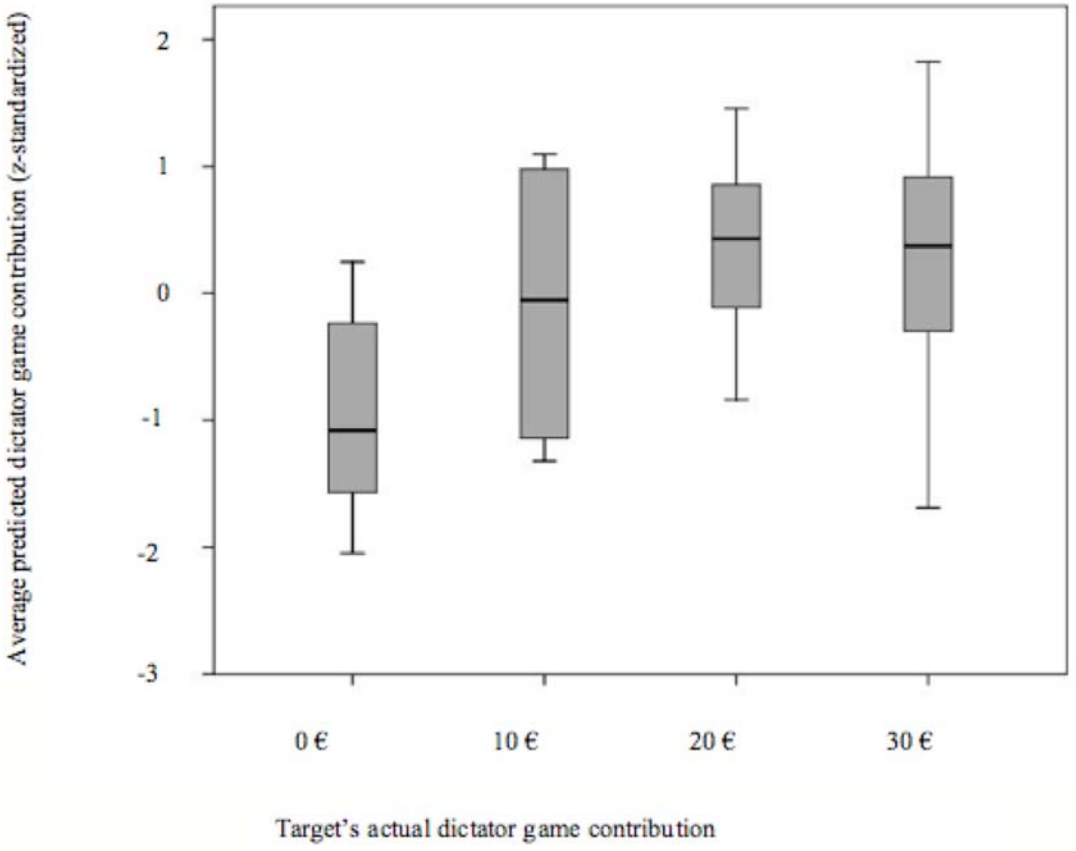


Fig. 2. Relationship between the actual dictator game contributions of target persons and the mean estimates of judges concerning the targets' contributions. Estimates for those who contributed nothing differed significantly from estimates for those who contributed 10, 20, or 30 Euro.

As can be seen, the average estimated contribution pertaining a person who kept all the money was more than 1 standard deviation lower than the average predicted contribution pertaining a person who divided the money equally (i.e. who transferred 30 Euro). Post hoc tests revealed that estimates for those who acted completely self-interestedly and transferred nothing differed significantly from the average predictions for those who belonged to the other three categories (*p* < .05). However, differences between the average predictions of the three other categories (i.e. giving either 10, 20, or 30 Euro to the recipient) were not significant. Figure 1 thus accentuates that judges performed particularly well in identifying those target persons who had kept all the money for themselves.

But which cues did the judges use for their estimates? One valid cue could have been the sex of the target as females on average gave away more money than males (M = 24.61 vs. M = 16.33 Euro; t (56) = 2.81, p = .007). In fact, judges took this sex difference into account by assuming that female targets had given away more money (M = 18.67 Euro) than males (M = 14.10 Euro) (t (56) = 3.77, p < .001). Nevertheless, the accuracy of the judges' estimates cannot be explained by the identification of this sex difference alone. A regression analysis using judges' estimates as dependent variable and both the targets' sex and their actual behavior as independent variables showed that both independent variables were significant predictors: F (2) = 9.798 (p < .01). The impact of gender was β = .34 (p < .01), whereas the impact of the targets' actual behavior amounted to β = .29 (p < .05).

8.4. Discussion

Miller (2000, 2007) and Frank (1988, 2004, 2008) hypothesize that some humans pursue ultimate genetic self-interest through psychological adaptations that embody a genuine concern for others. If such individuals carry observable cues of their altruistic traits, these cues should be used by others in order to accurately assess strangers' dispositions. Highlighting the adaptiveness of accurate person perception, Gangestad et al. (1992, p. 688) remarked, "Because interactions between unacquainted individuals regularly occur, the ability to detect certain behavioral propensities in others – sociability, trustworthiness, warmth, manipulativeness, and so forth – may be highly functional."

Our study aimed at testing whether individuals are able to estimate the level of altruism of unacquainted target persons on the basis of 20 seconds of silent video clips. As target persons had been video-taped in a setting completely unrelated to altruistic behavior, it could be checked whether individuals were able to make out permanent cues to situation-independent altruistic traits. Results showed that individuals were indeed able to estimate the altruistic behavior of target persons better than chance.

The proper identification of altruists can be advantageous in a variety of situations. In line with Miller (2000, 2007), the reliable detection of moral virtues should lead to improved assortative mating so that high value mates choose one another and low value mates have no chance but to stay among themselves. Following Frank (1988, 2004, 2008) the ability to identify altruistic tendencies might contribute to solve principal-agent problems (Grossmann & Hart, 1983) in a multiplicity of relationships. Principal-agent problems arise when individuals have to choose an interaction partner but lack information about the trustworthiness and potential future behavior of this person. This problem exists in particular when the potential interaction partner is completely unknown to the individual and can be supposed to have little commitment to behave trustworthily (Schelling, 1960).

Only recently, Verplaetse et al. (2007) investigated whether subjects are sufficiently sensitive to non-cooperative cues in strangers as accurate judgments by first impression could help to compensate the commitment information deficit. Using motionless pictures of target persons who had played a prisoner's dilemma game, the authors asked individuals to rate the cooperativeness of target persons. It was shown that humans indeed have some capability to differentiate between cooperative and non-cooperative individuals. However, this effect could only be evidenced when individuals responded to pictures that had been taken during the moment of decision-making rather than when they responded to neutral pictures unrelated to the decision task. Therefore, Verplaetse et al. concluded that humans may pick up emotional states of target persons, but are incapable of assessing permanent traits of (non-) cooperativeness.

Against the background of research on the thin slices paradigm, results of Verplaetse et al. (2007) are noteworthy; as Ambady et al. (2000) stated, short excerpts of dynamic behavior—often no longer than 30 seconds—enable accurate judgments of other persons, but a still photograph is not sufficient as it bears no dynamic information on the target person. It thus seems substantial that Verplaetse et al. (2007) could evidence any cheater detection abilities at all. However, the fact that they were not able to prove the judges' ability to read permanent cues to altruism on the basis of neutral pictures could have been due to the narrowness of cues entailed in the stimulus material.

Accordingly, discrepancies between the findings in hand and those of Verplaetse et al. (2007) might be caused by differences in the stimulus material. While Verplaetse et al. used motionless pictures, accurate assessments of permanent cues to altruism in this study were evidenced on the basis of short excerpts of dynamic behavior. To our view, the present approach was valuable in terms of external validity because our ancestors did not stick to photos either when judging a stranger on the basis of first impression.

The altruistic tendencies of our target persons were assessed via the dictator game. Bearing insights from the field of differential psychology in mind, one might ask whether the dictator game as a single behavioral measure is favorable to assess dispositional altruism, which can be defined as the general tendency to behave altruistically in a variety of situations and on different occasions. Personality psychologists tend to measure prosocial personality dispositions by the means of self-report scales (e.g. the dimension of agreeableness in the NEO-FFI by Costa & McCrae, 1998). However, from an evolutionary perspective this approach has to be judged critically. According to signaling theory, altruism is one of the characteristics where egoistic individuals should be motivated to disguise their true character in order to avoid discrimination by their social environment. Therefore, subjects in self-reports must be expected to deceive by stating a level of altruism that is above the true value. As Trivers (1991) remarked, deception may even generate self-deception as positive illusions facilitate to hide deception from others and to perform deception reasonably efficiently, that is, with modest cognitive cost. Therefore, biased self-reports may even result unintentionally.

Using a behavioral measure rather than a personality scale thus seemed essential. The dictator game represented the preferable instrument to us, not only because it assesses actual behavior, but also because the specific characteristics of the paradigm, namely non-iteration and anonymity, provide no strategic reasons to behave altruistically at all. Instead, altruistic behavior displayed in the dictator game may be attributed to a general preference of the actor to display prosocial acts. Beyond, empirical research authorizes the dictator game as an adequate instrument to assess altruistic traits, as Fetchenhauer and Huang (2004) could show that dictator game behavior correlates significantly with prosocial personality dispositions.

Although the judges' estimates were far from being perfect, our data indicate that humans are able to predict accurately and unconsciously the extent of altruistic behavior of others based on a limited set of dynamic cues in a very short time span. Judges performed especially well in identifying those target persons who kept all the money for themselves. This result strikingly replicates findings of Pradel et al. (in press). In their study, students had to assess the dictator game behavior of classmates and were also most accurate in estimating decisions of selfishly playing peers. This seems reasonable from an evolutionary point of view. It is not that important to distinguish the nice ones from the very nice ones, but it is essential to identify the real egoists as not doing so can literally be lethal.

Our results suggest that humans honestly signal altruistic dispositions through cues other than their altruistic acts and that variation in these dispositions exists. This adds a new dimension to the study of the evolution of human altruism: as altruism is identifiable it might simply be advantageous to behave altruistically due to the opportunity of altruists to carefully choose like-minded individuals for mutual cooperation.

In fact, individuals, when allowed to choose their partners freely, try to interact with those whom they expect to cooperate more (Johnson et al., 2008). In this vein, altruistic subjects in a public good game by Page, Putterman, and Unel (2005) mingled together when information about the contribution histories of players was given and proceeded to interact more productively than less cooperative coplayers did. Similarly, Sheldon et al. (2000) asked university freshmen to recruit three peers to participate in an N-person prisoner's dilemma game and found that prosocial individuals tended to stick together achieving higher group-level returns than antisocial participants. Transcending settings in the laboratory, the assortation of individuals with comparable altruistic tendencies even seems to be of importance in real life. Pradel, Euler, and Fetchenhauer (in press, see Chapter 9) could observe that students who play altruistically in a dictator game choose friends who behave similar to themselves.

As this study shows that altruistic traits are already sensed after a very short time span, it may be possible that altruists not only become choosy after a phase of shared experiences, but rather that interactions are selective from the very first moment of acquaintance.

Our findings also open a new avenue of research on the actual cues involved in non-verbal signaling of altruistic dispositions and raise the question, why these signals are reliable and thus evolutionarily stable. One cue, which judges in this study used, was sex. This cue was valid as in line with earlier studies (Eckel & Grossman, 1998), females indeed acted more altruistically than males did. However, as accuracy in judgments could not be explained by sex differences alone, other cues are likely to be relevant. One possibility is that the cues involved are involuntary facial expressions that are difficult to bring under voluntary control for strategic use. Even if a dishonest use of these cues was possible, it may be maladaptive to use them in this way because humans' willingness to engage in altruistic punishment might be especially strong towards those that signal a high level of altruism – and then do not comply with these signals.

Hence: You want to seem nice? Be nice! Humans possess the ability to recognize permanent features of altruism.

9. Study 2: Spotting altruistic dictator game players and mingling with them - The elective assortation of classmates[2]

9.1. Introduction

From birth to their grown-up years, humans daily interact with others and feel a need for social attachment (Baumeister & Leary, 1995). The dependence of humans on their social world can be risky in terms of resources, reproduction and survival, if individuals choose the wrong interaction partners. The evolutionary perspective therefore suggests that humans should have developed adequate skills to judge other persons accurately in order to find the right lovers, friends and allies, but to avoid the villains (Haselton & Funder, 2006).

One criterion for a good social interaction partner is altruism. Altruism denotes an individual's willingness to give up resources in order to benefit others. The evolution of altruism has puzzled scientists for generations and although there has been considerable progress in its theoretical explanation, the question how such self-detrimental behavior could survive the pressures of natural selection has not been answered unanimously. Most game theoretical models of the evolution of altruism assume random encounters between interaction partners (Aktipis, 2004). However, for many species, encounters with others are non-random. It has thus been supposed that the evolution of altruism may have been driven by assortation processes (Wilson & Dugatkin, 1997), i.e. the gathering of like-minded individuals (for a general overview of social selection via the cooperation between correlated characters see Frank, 1998; 2006).

---

[2] A modified version of this chapter is currently in press in Evolution and Human Behavior together with Harald Euler and Detlef Fetchenhauer.

Consider a population with egoists and altruists. If individuals were really equipped with a “personality judgment instinct” (Haselton & Funder, 2006), they should be able to distinguish altruists from egoists. As altruists should have an interest in assorting themselves, they would leave self-interested individuals no chance but to stay among each other. As altruists could consequently reap extra benefits through mutual cooperation, the selfish advantage of egoistic individuals would be out-competed. Altruism would become adaptive as it would provide a long-term benefit. (West et al., 2007, have argued that such behavior should be termed *mutualism*, rather than *altruism.*)

Several scholars (Boorman & Levitt, 1973; Eshel & Cavalli-Sforza, 1982; Frank, 1988; Peck, 1993; Wilson & Dugatkin, 1997) construed scenarios similar to the one outlined above, in which altruists prefer like-minded interaction partners against the selfish rest of the population. To make assortation, as specified in the scenario, work as a driver of the evolution of altruism, two conditions have to be fulfilled: (1) Individuals have to be able to distinguish altruists from egoists, and (2) altruists have to elect like-minded individuals for mutual cooperation in order to reap synergetic extra benefits. The present study was designed to test whether these two assumptions hold true in a natural context, as only then assortation may reliably be invoked as an additional explanation for the evolution of altruism.

As to the first condition, there is indeed evidence that humans are equipped with a psychological mechanism to detect cheaters in social interactions (Cosmides & Tooby, 1992). A few studies examined whether humans are also capable of predicting altruistic behavior reliably. Frank et al. (1993) offered groups of participants the opportunity to get acquainted with each other for 30 minutes before playing a prisoner’s dilemma game. Subjects were able to discuss the paradigm for 30 minutes and make non-binding pronouncements to their interaction partners about their game decision (defection or cooperation). Subsequently, the participants indicated their actual response covertly and were asked to predict the responses of their counterparts. Predictions were better than chance. However, Ockenfels and Selten (2000) criticized these results suggesting that the accuracy of the predictions might have been due to the explicit pronouncement of defectors. As defectors would have had little interest in deviating from their announced decision, predicting their behavior was easy. Brosig (2002), therefore, replicated the experiment under restricted conditions. Pre-communication sessions were filmed so that subjects who explicitly announced their intention to defect were excluded from the analysis. Accuracy of predictions exceeded chance level under these conditions as well.

Going a step further, Verplaetse et al. (2007) investigated whether cooperation is predictable by only minimal visual information about the person. Subjects were shown photos of unknown target persons that had been taken when the target persons made a decision in a one-shot prisoner's dilemma game, and they had to rate how cooperative these target persons behaved. The subjects were able to differentiate between cooperative and noncooperative target persons. Fetchenhauer, Groothuis, and Pradel (unpublished data) could even prove that subjects may predict the level of altruistic behavior on the basis of 20 seconds of silent video clips that were recorded in a *neutral* setting unrelated to altruistic behavior. Thus, first impressions seem to give humans a clue of what kind of person they are dealing with.

As to the second condition for the evolution of altruism by assortation, namely that altruists choose like-minded persons for mutual cooperation, assortation with respect to mating ('assortative mating') is a well established empirical phenomenon (Mascie-Taylor, 1995; Spuhler, 1968) that has been shown for a variety of somatic and psychological characteristics. The phenomenon, furthermore, exceeds mate choice as it has been evidenced as well for friendship (Berscheid, 1985). Sheldon et al. (2000) investigated assortative partner choice with particular respect to prosociality. The authors asked university freshmen to recruit three peers to participate in an N-person prisoner's dilemma game. Subjects with a prosocial value orientation tended to assort with one another. They thus achieved a group-level advantage in the game returns which counteracted the individual-level advantage of antisocial participants.

In sum, data show that persons seem to be capable to predict the level of altruism in other persons, and moreover tend to associate with those persons who show a level of prosociality similar to their own one. Yet, the empirical evidence so far is based on laboratory studies with unknown ecological validity. The present study investigates whether the prediction of altruistic behavior and the grouping of altruistically like-minded individuals holds true also in genuine social groups in a natural setting. School classes are such genuine social groups, and a study of students of different grades offers the opportunity to investigate the possible age-dependence of prosociality assessments and assortation processes.

In school classes relationships between students vary in intensity. A typical student has a few friends among classmates, several she simply likes, and others she might dislike. Additionally, the relationship to some classmates could be one of indifference. This setting, therefore, offers the opportunity to examine the relationship between social closeness and the accuracy of judgments about other students' altruistic behavior. In the present study, altruistic behavior was operationalized as the decision in a dictator game, in which each student was asked to secretly divide a sum of money between himself or herself and another anonymous classmate. Subsequently, each student was asked to predict how each classmate had divided the money in the dictator game. We assumed that individuals might be better in predicting the behavior of their friends, with whom they have shared many experiences, than the behavior of more distant persons. As to the study of assortation, the school setting furthermore allows to observe in how far students choose friends who make dictator game decisions like they themselves do.

It might be noted that when making predictions about the accuracy of character assessments, evolutionary theory takes a position very different to the one represented by mainstream social psychology. The latter would not expect adolescents to be capable of accurately assessing the behavior of their peers in such a situation as social cognition in general is regarded to be biased (Kunda, 1999). For example, Ross, Greene, and House (1977) described the so-called false consensus effect, which is due to an individual's unfounded assumption of similarity between himself and a target person. Fetchenhauer and Dunning (2006) showed the false consensus effect to be present in a variety of game-theoretical paradigms. However, the assumption of similarity does not always provoke a *false* consensus effect but may induce accurate predictions if the target person is indeed similar to the judge (Dawes & Mulford, 1966; Hoch, 1987). We shall, therefore, refer to this effect neutrally as 'consensus effect'. Because the consensus effect has been shown to be frequent within close relationships (Schul & Vinokur, 2000), it needs to be controlled for when studying social perception.

Another factor to be taken into account is the better-than-average effect, the pervasive tendency of individuals to assume that they are superior to others. This effect has been shown for a large range of socially desirable attributes, altruism included (Dunning, 2005). Therefore, persons can be expected to underestimate the trustworthiness and altruism of others (Fetchenhauer & Dunning, in press). Both effects, the consensus effect and the better-than-average effect, are independent and may co-exist as individuals may base their predictions on their own behavior (consensus effect), while at the same time using a somewhat lower anchor (better-than-average effect) to estimate the behavior of others (Fetchenhauer & Buunk, unpublished data).

The present study tests the following three hypotheses: (1) Students are able to predict the level of altruistic behavior of their classmates in a dictator game. (2) Classmates who play altruistically in the dictator game are more often labeled as likable or as friend than egoists are. (3) Classmates positively assort themselves in their friendships along the dimension of altruism, i.e. altruists have friends who play altruistically, too and egoists have friends who play egoistically, too.

9.2. Methods

9.2.1. Participants

Participants were 127 students (60 female, 67 male) from six secondary school classes of a different grade each. Age varied from 10 to 19 years: 5th grade (aged 10 to 11), 8th grade (aged 13 to 14), and 10th, 11th, 12th, and 13th grade (aged 15 to 19). Size of classes varied from 14 to 29 students. Parents had given written consent for their child's participation and were debriefed after the experiment.

9.2.2. Procedure

The experiment was conducted in two secondary schools in Cologne, Germany. Sessions, one for each class, were carried out in social science lessons. No detailed information about the experiment was given to the participants in advance. In order to guarantee experimenter-subject anonymity, a subject number was handed out to every participant. To enable participants to link subject numbers to classmates, which was necessary for a later task, the participants were seated in a circle so that they could see each other with their subject number. This procedure assured that no real names entered the data sets.

The students were first familiarized with the logic of the dictator game without naming the game as such. The participants then received the first part of the questionnaire, which was coded with their subject number, and were asked to secretly mark down their own dictator decision. This means that they had to imagine a situation with two persons, a dictator and a recipient, in which the dictator receives a certain amount of money. Taking the perspective of the dictator, students had to consider the distribution of the money between themselves and the recipient, who was an anonymous student of the class. Neutral vocabulary was used all throughout the instructions to avoid influencing the participants in any possible way. For example, the dictator game was referred to as a “distribution task” and the dictator was referred to as “Person A” while the anonymous recipient was called “Person B”.

The participants were informed that their decisions had a tangible consequence in that a payment would be made a week later. To maintain comparable incentives, the sum to be divided increased with the students’ age, ranging from 6 Euro (about US $ 9) in 5th grade to 10 Euro (about US $ 15) in 13th grade. These payments corresponded approximately to the age-specific recommendations for allowances of the German youth welfare offices (Sport-, Schul- & Jugendamt Springe, n.d.). The participants could pass money in whatever whole-numbered proportion they wanted, but could also keep the entire amount for themselves. The recipients were unknown to the participants and the participants were told that—although for the payment session recipients were secretly allotted to them later—they would never be informed about the identity of their respective classmate.

Communication was not allowed during the whole experiment. It was always made clear that information on the actual behavior of any person could never be traced back to a face or name, neither by any other subject nor the experimenter, as data management and analysis for each class were made by different persons than the experimenter.

After all participants had made their dictator decision, they received the second part of the questionnaire in which they were told that they now had to act as judges and assess how each of their classmates had decided in the preceding distribution task. The participants could not have expected this task when they had made their own dictator game decisions. For motivational reasons, it was announced that the participant with the most accurate predictions would win a ticket for the local cinema.

For a sociometric analysis, the participants were finally asked to write down which of their classmates they characterized as friend, as likable, or as dislikable. The participants were free to nominate as many classmates as they wanted for each of these three categories.

For the purpose of payment management, half of the participants were randomly assigned the role of the dictator after the experiment, the other half the role of the recipient. Each dictator was randomly matched with a recipient. Resulting pairs were paid one week later according to the distribution chosen by the respective dictator. The participants received the money in a closed envelope labeled with the subject number. No participant obtained any information about the identity of the other classmate with whom they had been paired.

9.2.3. Data preparation

As the amount of money that had to be distributed in the dictator game increased with the participants' age, payment values were transformed into percentages for reasons of comparability. Five out of 127 students (3.9 %) passed more money to the recipient than they kept for themselves. According to Camerer (2003) dictator game contributions above 50 % are rare in western cultures. The fairness principle of equality, which demands parity of resources, would suggest contributions of 50 % but nothing more. So would the equity rule, which demands a division of resources according to effort. Since both students, the dictator and the recipient, expended the same—namely none—effort in the acquisition of this money, an equal split would have been suggested. Thus it could not be ruled out that outliers who passed more than 50 % might have misunderstood the experiment, especially since all these five participants were from lower grades. For this reason, these five participants were excluded from all further analyses. In order to maintain consistency, those 3.7 % of the predictions in which students had expected their classmates to keep less than 50 % for themselves were excluded as well. However, results were only minimally influenced by these exclusions.

For the sociometric analysis the participants had been asked to classify their relations to every other classmate according to three levels of social closeness: friendship, liking, or dislike. Because students were free to make as many nominations as they wanted, some interindividual relations were not classified at all. These relations were consequently labeled as ‘indifferent’. As class size varied, relationship nominations between students of different classes were incomparable (i.e. the number of nominations was likely to depend on the number of students in the relevant class). Thus, relationship nominations were transformed into relative values taking into account the possible number of nominations that could be made in the respective class.

Two sets of data were created. Data set 1 comprised 122 cases, one for each participant, with information about the participants’ own decision in the dictator game, their average prediction concerning others, how the behavior of the participants was predicted on average by their classmates, and the relative number of nominations of friendship, liking, and dislike each individual received. In data set 2 each case represented a single prediction from one student (judge) concerning another (target). As students were asked to predict all their classmates’ behavior, $(n_i - 1)$ predictions were received from each judge, where n refers to the number of students in the respective class. A total of 2437 cases were obtained from all six classes. Moreover, data set 2 included information on the nature of the relationship between each judge and the respective object (i.e. the level of social closeness).

9.3. Results

9.3.1. Actual and predicted dictator game contributions

Participants on average contributed 37.3 % towards the other person. The largest fraction (49 % of the participants) handed half of the money to Person B. A total of 8 % of the participants kept the entire amount for themselves, while 43 % made a contribution in-between. Participants slightly underestimated the altruism of their classmates as the average predicted contribution amounted to 34.0 % (t (122) = 3.09, $p < .01$, $d = 0.24$).

Girls acted more altruistically (contribution of 42.1 %, ± 11.6) than boys (32.8 %, ± 18.1, t (122) = -3.39; $p < .001$; $d = 0.61$). This sex difference was reflected by the participants' predictions about their classmates, as girls were predicted to make higher contributions than boys (37.9 %, ± 6.4 versus 30.0 %, ± 9.4, t (122) = -5.36; $p < .001$; $d =$ 0.98).

9.3.2. The ability to estimate the altruistic behavior of others

Our first hypothesis stated that students are able to estimate the level of altruistic behavior of their classmates in the dictator game. Because every participant rated every other participant, predictions were non-independent. Moreover, the variance of individual predictions could be assumed to be partly attributable to class membership. To control for these facts, a method developed by Warner, Kenny, & Stoto (1979) was used. The average prediction of several individuals concerning a single student was adjusted so that the effects of repeatedly occurring judges and class-membership were sorted out (see Appendix A). These adjusted average predictions were z-transformed for reasons of clarity and comprehensibility.

We explored the relationship between actual contribution and adjusted average predictions by a partial correlation analysis controlling for class membership. This was necessary as the variance of actual contributions was affected by class membership, too. Five dummies were created, one for each class, with class 5 as the reference category. The average prediction of a target's behavior indeed correlated with her actual behavior ($r = .39$, $p < .001$, one-tailed) showing that students were able to predict the dictator game behavior of their classmates better than chance. This result persisted when the target's sex was controlled for (partial $r = .31$, $p < .001$, one-tailed) indicating that classmates were able to make out those variations in the degree of altruistic behavior that went beyond differences due to the sex of the target. To highlight this result, we trichotomized the actual dictator game behavior into three categories: egoistic behavior (no contribution), moderate altruistic behavior (0 % < contribution < 50 %) and altruistic behavior (contribution = 50 %). As can be seen in Figure 1, the effect of classified actual dictator contributions was highly significant [F (2, 119) = 9.27, p <.001]. The average predicted contribution pertaining a person who kept all the money was more than 1.5 standard deviations lower than the average predicted contribution pertaining a person who transferred 50 %. Post hoc tests revealed that the average predictions for those students who had contributed nothing differed significantly from the average predictions for those who belonged to the altruistic categories. However, average predictions for moderate altruistic contributions (0 % < c < 50 %) and altruistic contributions (50 %) did not differ significantly from each other. Figure 1 thus accentuates that students performed particularly well in spotting those classmates who had kept all the money for themselves.

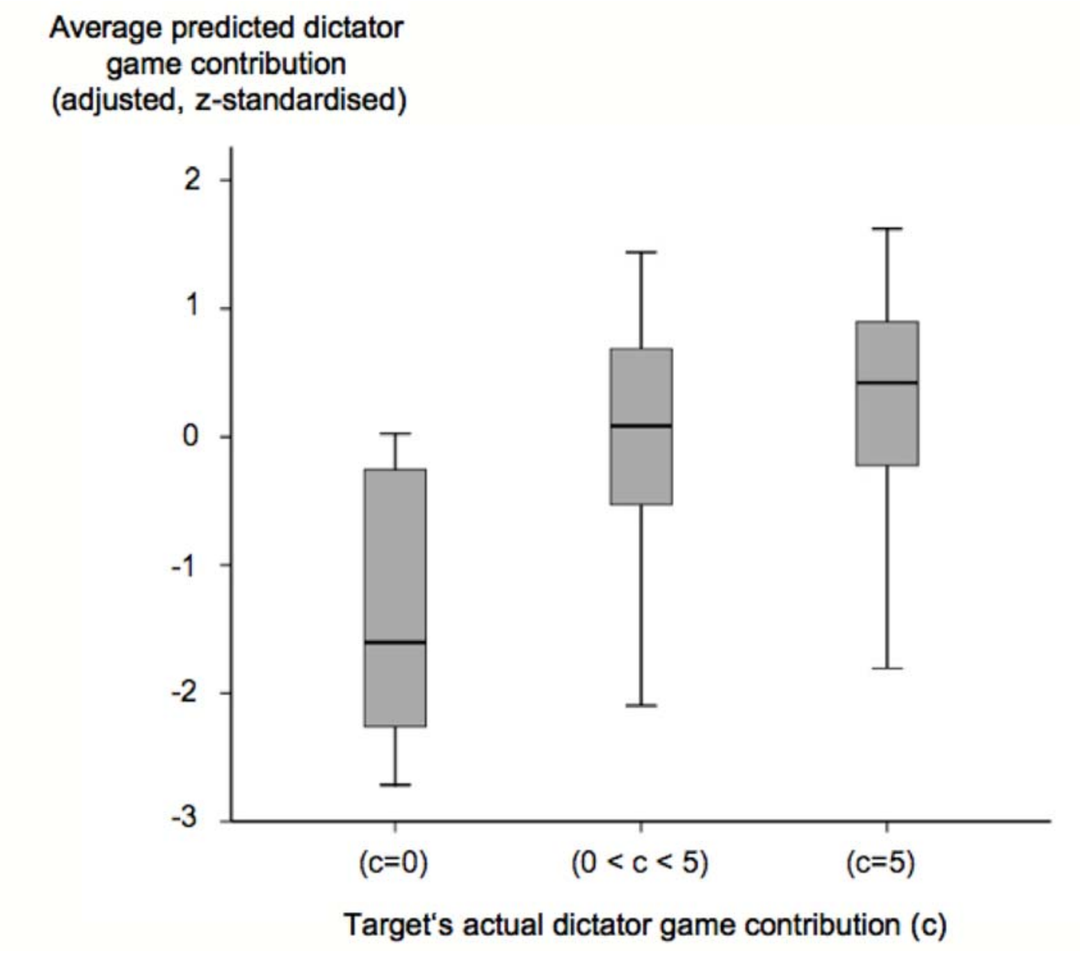


Fig. 3. The average prediction of classmates concerning the dictator game behavior of a target (n = 122) depending on the target's actual contribution in the dictator game. Prediction values are adjusted for non-independence due to class-membership and repeatedly occurring judges and targets.

Having focused on the analysis of *average* reputation so far, we now wanted to find out whether *individual* predictions concerning the behavior of an object were a good indicator for actual behavior, too. We thus repeated our analysis by using data set 2, in which a single prediction (again adjusted for non-independence of cases, see Appendix A) of one student concerning the behavior of another was treated as the unit. Students' individual predictions matched the behavior of their classmates significantly better than chance (r = .21, p < .001, one-tailed). The match was maintained when the target's sex was partialled out (partial $r = .16$, $p < .001$, one-tailed). Thus, the hypothesis that persons are able to identify altruistic disposition in classmates was confirmed.

9.3.3. The social appraisal of altruism

Our second hypothesis was that classmates who distribute the money evenly in the dictator game are more often labeled as likable or as friend than egoists are. To test this, we initially explored the relation between the number of received liking nominations and the individual's level of altruism. The correlation was slightly positive with class membership and sex partialled out (partial $r = .14$, $p = .07$, one-tailed). Likewise, the effect of altruism on received dislike nominations was slightly negative (partial $r = -.12$, $p = .10$, one-tailed). As to the question of whether altruists had more friends than egoists, the correlation between the number of received friendship nominations and altruism was not significant (partial $r = -.08$, $p = .20$, one-tailed). Summarizing these results, the hypothesis that individuals who share the money are more popular than individuals who keep everything for themselves could not be confirmed.

9.3.4. The assortation of altruists

The third hypothesis claimed that classmates positively assort themselves along the dimension of altruism. To test this, we investigated the relationship between the dictator contributions of friendship nominators and the contributions of their friendship nominees which correlated significantly with each other (partial $r = .21$, $p = .01$, one-tailed), with class membership partialled out. The correlation persisted when controlling for sex (partial $r = .19$, $p = .02$, one-tailed) which excludes the explanation of an artifact due to sex-specificity of friendship dyads, namely that altruistic girls are friends with altruistic girls and self-interested boys are friends with self-interested boys. Thus, our hypothesis that altruists choose like-minded persons as friends was confirmed.

9.3.5. Accuracy of predictions and grade

In order to investigate the relationship between the accuracy of predictions and grade, a variable was needed which indicated the quality of predictions of every individual. For this purpose, the individual accuracy rate of each participant was calculated by correlating the predictions of judges with the actual behavior of their respective targets. A total of 122 correlation coefficients resulted which can be considered to be individual accuracy rates. However, a relationship between the individual accuracy rates and grade could not be proven ($r = .10$, $p = .30$, one-tailed). Younger students performed as well as older students in predicting their classmates' decisions.

9.3.6. Accuracy of predictions and social closeness

We further aimed at finding out whether the level of social closeness between two persons (i.e. friendship, liking, dislike, indifference) influenced the accuracy of their predictions. Of all 2437 dyadic relationships, 17 % were mutual friends (both participants nominated each other as friends). Mutual liking occurred in 10 % of the dyads, mutual dislike in 8 %, and mutual indifference in 10 %. The remaining dyads (55 %) were mixed relationship declarations (e.g. Person A declared friendship and Person B declared liking) which were excluded for further analyses.

Accuracy rates (correlations between actual behavior and predictions) for the four relationship types are shown in Table 1 (complete subsamples). The accuracy of predictions varied with social closeness and was highest for mutual friendships, followed by dislike relationships and finally mutual liking relationships. The accuracy of predictions of unrelated individuals was lowest and reached chance level only. Thus, students were most accurate when they judged friends and individuals they disliked.

Dyads had been treated as independent cases until now. However, some dyads were connected such that one of the dyad members occurred repeatedly (e.g. because he or she was a mutual friend with more than one person). Hence, the present sub-samples could be construed as asymmetric Social Relations Model designs whose estimation is complex (Kenny, Cashy, & Cook, 2008). We refrained from controlling for the interconnections between cases as any corrections should have had no influence on effect sizes but on standard errors only. Yet, the stability of social closeness effects on accuracy of predictions was tested under tightened conditions so that every dyad member only appeared once. Cases were randomly selected which means that if a person was a friend with two other individuals, only one of these dyadic relationships was included in the sample of mutual friends via random sampling. As it can be seen in Table 1 (random parts of sub-samples), accuracy rates were still significant for friendships and dislike-relations. Accuracy rates for liking-relations and indifferent peers were not significant. These results were comparable to those of the complete samples: individuals judged those persons best to whom they had a decided opinion—being either positive or negative. The following analyses were thus based on the complete samples, treating predictions of individuals as independent ones.

Relationship	Complete sub-samples		Random parts of sub-samples	
	n	Accuracy of prediction	*n*	Accuracy of prediction
Mutual friendship	418	.34 $^{**}_{a}$	52	.35 $^{**}_{a}$
Mutual liking	246	.11 $^{*}_{b}$	51	.16 $_{a}$
Mutual dislike	186	.18 $^{**}_{a,b}$	41	.30 $^{*}_{a}$
Mutual indifference	274	-.07 c	35	.09 a

Table 1: Accuracy of prediction for different types of relationships (Pearson correlation coefficients). ***Note.*** **Correlation coefficients that do not share subscripts differ significantly at p < .05. n = number of predictions.**

In the analysis so far, the influence of a possible consensus effect had not been controlled for. The potential influence of a consensus effect can be seen in Table 2 which shows that individuals who disliked each other behaved differently in the dictator game, whereas mutual friends behaved alike. The high accuracy rates for friends could thus have been driven by a consensus effect.

Relationship	N	Objective similarity
Mutual friendship	209	$.16^{**}_{a}$
Mutual liking	123	$.09_{a}$
Mutual dislike	93	$-.22^{*}_{b}$
Mutual indifference	137	$.08_{a}$

Table 2: Actual similarity between dyad members (Pearson correlation coefficients). *Note.* Correlation coefficients that do not share subscripts differ significantly at *p* < .05. N = number of dyads. * *p* < .05; ** *p* < .01.

To find out how much of the reported accuracy rates in dyads was really due to the judge’s explicit ability to identify the object’s level of altruism (“true accuracy”) and how much of the prediction was due to bias in form of a consensus effect, the model of Kenny and Acitelli (2001) for the simultaneous measure of accuracy effect and consensus effect was applied (see Appendix B). Note that subsequently the term accuracy effect denotes the proportion of variance within predictions due to true accuracy and the term overall accuracy denotes the correlation between predictions and actual behavior of targets which had been given in Table 1.

For all types of relationships, Table 3 shows which proportion of the predictions was due to the observers’ distinct ability to identify altruists (accuracy effects) and which proportion was due to bias (consensus effects). Predictions were indeed inflated considerably by the consensus effect for all relationships. However, under control of the consensus effect true accuracy could be proved for friends and dislike-relations. Thus, the hypothesis that altruistic behavior can truly be predicted was at least confirmed for these two groups.

Relationship	N	Accuracy effects	Consensus effects
Mutual friendship	209	.28***$_{a}$	.58***$_{a}$
Mutual liking	123	.10 $_{b}$	.40***$_{b}$
Mutual dislike	93	.27***$_{a,b}$	.34***$_{b}$
Mutual indifference	137	.05 a,b	.53***a,b

Table 3: Accuracy and consensus effects (unstandardized regression coefficients). *Note.* Regression coefficients in the same column that do not share subscripts differ significantly at $p < .05$. N = number of dyads, *** $p < .001$.

To analyze how the concurrence of accuracy and consensus effects contributed to overall accuracy rates, the correlations between actual and predicted contributions were decomposed into true accuracy and the increment due to bias. As can be seen in Table 4, mutual friends benefited from the consensus effect as they had made identical dictator contributions. In contrast, the overall accuracy of individuals who disliked one another was reduced as they, too, assumed similarity, which did not actually exist. Predictions of individuals who liked one another were mainly driven by true accuracy, though a positive impact of the consensus effect could be shown nonetheless. For unrelated individuals a decomposition of the overall accuracy rate was impossible as overall accuracy was insignificant.

Relationship	N	*r*	True accuracy	Increment due to consensus effect
Mutual friendship	209	.34	.25	.09
Mutual liking	123	.11	.08	.03
Mutual dislike	93	.18	.25	-.07
Mutual indifference	137	-.07	n.a.	n.a.

Table 4: Partitioning of the overall accuracy correlations (r). *Note.* As the accuracy correlation for mutual indifferences was insignificant, the validity of the decomposition could not be assured. N = number of dyads.

To summarize, beyond the consensus effect a true ability to identify altruists could be demonstrated. Biased overall accuracy but also true accuracy was highest for friends and disliked persons, indicating that predictions were more valid when the judge had a firm attitude towards the object, be this attitude positive or negative, than when the attitude was one of indifference or just acceptance.

9.4. Discussion

The purpose of the present study was to test the hypotheses that (1) individuals are able to predict the level of altruistic behavior in a dictator game of individuals familiar to them, (2) individuals who play altruistically in the dictator game are judged more positively than egoists and have more friends than egoists, and (3) that individuals positively assort themselves in their friendships along the dimension of altruism. Hypotheses 1 and 3 were supported. Individuals are able to identify altruism, and altruists preferably choose other altruists for mutual cooperation. Hypothesis 2 could not be answered unambiguously as judgments of affection concerning target persons only marginally correlated with their degree of altruism. Furthermore we wanted to explore the influence of age and social closeness on the validity of social perception. While the accuracy of predictions was linked to social closeness, it was independent of age.

Accuracy of predictions was measured by asking subjects to predict the dictator contributions of their classmates. This task can be considered demanding for the participants as Dunning et al. (1990) showed that accurate predictions of someone’s behavior strongly depend on the availability of two types of information: A judge needs to have accurate information about the exact nature of the situation confronting the target and about the target’s construal or understanding of the situation at the point the target must respond. Even though participants had precise information on the objective features of the situation, as they all took part in the same anonymous dictator game, they still needed to accurately anticipate the meaning of this situation from the individual perspective of each classmate. This was likely to be a difficult task as it has to be presumed that students might have interpreted the situation in different ways.

Following Hagen and Hammerstein (2006), cues of the degree of reputational consequences in economic games are processed in a manner that may be intuitive or affective and thus cognitively impenetrable. Admittedly, our experimental setting, in which we seated participants in a circle to enable them to link classmates with subject numbers, might have created a context that students perceived as partly public, despite the explicit propositional information in the instructions of the dictator game, which announced that anonymity would always be guaranteed.

There is evidence that removing anonymity from economic games increases individuals' contributions to a public good (Andreoni & Petrie, 2004). Moreover, it has been shown that already subtle cues of observability, namely images of eyespots, affect generosity (Burnham & Hare, 2007; Haley & Fessler, 2005). It can be assumed that participants who behaved strictly egoistically when sitting face-to-face with their classmates almost surely would have behaved equally egoistically if anonymity had been even more certain. Per contra, it must be presumed that the behavior of those who transferred money could have been partly evoked by cues of publicity which activated evolutionarily relevant psychological mechanisms that manage reputation independent of explicit reasoning.

Observed altruistic behavior in the non-iterated anonymous dictator game, therefore, cannot unconditionally be interpreted as 'pure' altruism in the sense that it was independent of reputation management efforts. It rather seems plausible that there were some individuals who were strongly inclined to manage their reputations and cooperated because they perceived the experimental situation as public and thus followed rules which are rational in their everyday life (McKenzie, 2003), while there were others who were intrinsically motivated to behave altruistically regardless of the specification of the situation.

With reference to the accuracy of predictions, we do not know whether participants' predictions grounded on the exact assessment of the extent to which a participant perceived the situation as private or public and acted accordingly, but nevertheless students were able to predict with some accuracy how their classmates would behave. So whatever the level of anonymity that was perceived by an actor, it seems reasonable to assume that either the same level of anonymity factored into his or her predictions about others' behavior or—even more elaborate—that he or she made predictions about how the respective target person would interpret and thus act in this ambiguous situation. In any event, individuals were able to estimate their classmates decisions to some extent and they performed especially well in predicting egoistic decisions. This can be considered adaptive because altruistic persons should always—irrespective of their intrinsic or extrinsic motivation—be favored over egoists as cooperation partners.

The average dictator contribution of students in our study amounted to 37 %. As Camerer (2003) reported, the mean dictator contribution of adults generally ranges from 20 % to 40 %. In a study by Takezawa, Gummerum, & Keller (2006) mean dictator contributions of 11- to 14-year-olds amounted to 44 %. Thus, the level of contributions in our study is similar to that of earlier studies. Moreover, our results replicate the findings of Harbaugh, Krause, and Liday (2003) that the dictator contributions of children and adults are comparable. Benenson, Pascoe, and Radmore (2007) suggested that even young children have biologically based altruistic dispositions which are additionally fostered by socialization practices. In line with earlier studies (Eckel & Grossman, 1998), females in our study acted slightly more altruistically than males did.

Although the participants succumbed to the better-than-average effect and underestimated the absolute level of dictator contributions of their classmates, they correctly differentiated between the altruistic behavior of boys and girls in that they predicted girls to pass a higher amount than boys. As to the relative accuracy of predictions, the students performed well as the variation of average predictions matched the variation of actual behavior. Individual predictions, too, were significantly better than chance. Moreover, the students did not base their predictions on sex alone, as it could be shown that over and above the ability to differentiate between altruistic tendencies of boys and girls, students took into account the variation of altruism within both sexes. Thus, adolescents are indeed able to distinguish self-interested individuals from altruists.

The accuracy of predicting dictator contributions did not increase with grade; 11-year-olds performed as well as 19-year-olds. This result seems especially noteworthy when one considers that the 19-year-old participants had known each other for up to nine years, while the youngest students had only been acquainted for several months. Moreover, in a follow-up-study we could show that even 9-year-old children achieve comparable outcomes (Pradel & Fetchenhauer, unpublished data). As recent results indicate, infants as young as 6 months can take into account an individual's helping or hindering actions towards others by evaluating this individual as appealing or aversive (Hamlin, Wynn, & Bloom, 2007). The early development of capabilities to evaluate the prosocial behavior of others as well as their developmental stability support the proposition that the ability to evaluate altruistic dispositions in others is not primarily dependent on general learning or general socialization experiences, but is the result of biological adaptation.

Although altruistic students favored like-minded persons as friends, altruists were not generally favored as interaction partners. Liking and antipathy nominations only marginally correlated with the degree of altruism. Moreover, altruists and egoists had a comparable number of friends. This finding contradicts the hypothesis about the popularity of altruistic players. However, it should be mentioned that these German students nominated on average only three classmates as friends. The term ‘friend’ is used quite selectively in Germany and more equivalent to the term ‘best friend(s)’ in English-speaking countries. With the benefit of hindsight, one might wonder whether the hypothesis that egoists are spurned in general was justified as it is not actually generated from the models of assortation which we investigated. For example, Frank (1988) did not suggest that egoists do not have friends at all. He supposed instead that altruists preferentially cooperate with other altruists, and this supposition is indeed supported by our data. Nevertheless, it seems understandable that egoists have friends, too. They just choose or get friends who are less altruistic than friends of altruists.

Predictions about the dictator contribution of a friend were more valid than predictions about the contributions of liked and unrelated classmates, as can be expected from the intensity of communication and interaction in friendships. Interestingly, the overall accuracy for predicting the dictator contribution of disliked classmates was high as well and did not differ significantly from the accuracy of predicting the behavior of friends. When controlling for the consensus effect a similar pattern emerged. True accuracy was highest when students evaluated friends, but similar levels of true accuracy were reached for the evaluation of explicitly disliked persons. In ancestral environments, a valid evaluation of exactly these two groups must have conferred fitness benefits. Persons ought to know about the true intentions of their best friends as they are interacting with them regularly and are thus highly dependent on correct assessments of their characters. People should further be aware of the intentions of disliked individuals, that is, those with different interests, because they are potential adversaries. However, we have to admit that the direction of causation could be the opposite. It could be that individuals befriend those who they believe to have good intentions (and who truly have good intentions) whereas they dislike those who they believe to have evil intentions (and who indeed have evil intentions).

In sum, our study shows that humans—although succumbing to some well-known biases of social perception—are able to predict altruistic or self-interested behavior of others, and that they moreover tend to assort themselves in their friendships along the dimension of altruism. Speculating about the consequences of these findings, it seems reasonable to expect that altruists, who mingle electively with each other, reach extra benefits through mutual cooperation. As these benefits may counteract the individual advantage of egoists, it can be assumed that altruism may have evolved due to assortation processes between altruists who acknowledged each other's cooperative tendencies.

Our empirical results thus support the theories of assortation and complements findings on various mechanisms which have probably contributed to the evolution of altruism, like reciprocity (Trivers, 1971), social norms and punishment (Fehr et al., 2002; Fehr & Gächter, 2002, 2003), to name only a few. We therefore think that this study adds a small but important piece to solve the puzzle of human altruism.

Although the ability to spot altruists was proven, the exact nature of cues to altruism remains unknown. Frank (1988) argued that altruism is motivated by moral emotions. Drawing on the fact that emotions are linked to nonverbal behavior and involuntary facial expressions (Darwin, 1872), which are difficult to produce on command (Ekman, 1985), altruism might have been signaled through emotional displays. However, as this explanation is rather vague, future research is needed to open the black box of processes underlying the intuitive assessment of altruism. But the conclusion stands that persons can predict the altruistic behavior of acquaintances and that they draw consequences from their insights. Birds of feather flock together—so do altruists.

10. Study 3: Partner in life or one-night stand? - How reproductive strategies might have shaped the evolution of altruism

10.1. Introduction

Moral virtues have been an inspiring phenomenon for two groups of evolutionary psychologists: for those who investigate their relevance as mate choice criteria (for a cross-cultural longitudinal analysis see Buss, Shackelford, Kirkpatrick & Larsen, 2001), and for those who try to explain their phylogenetic roots (e.g. de Waal, 2008; Fehr & Fischbacher, 2003; Hamilton, 1964; Sober & Wilson, 1998; Trivers, 1971). Usually, the two lines of research rarely interact with each other as can be seen from the fact that within textbooks they tend to form completely isolated chapters. The present paper aims at integrating the two research lines by arguing that theories of mate selection might help solving the puzzle of the evolution of altruism. To specify, it is investigated whether prosocial virtues might have evolved because people look out for them when searching for a partner in life as distinguished from a one-night stand.

Research on mate preferences shows that moral traits like kindness, fidelity, sympathy, and prosociality are sexually attractive (Miller, 2007) and form the top criteria for mate choice in both sexes and across cultures (e.g. Buss, 1989). These preferences are not only reflected by interviewees' statements on research questionnaires, but can also be recognized in single's ads where lonely hearts all over the world advertise and seek moral qualities (e.g. Oda, 2001; Koziel & Pawlowski, 2003). The appreciation of morality as a partner criterion is comprehensible as high virtue individuals are likely to take interest in their partner's happiness and well-being and—considering family planning—can be expected to take care of potential off-spring (Miller, 2007).

But why do humans behave prosocially at all? Indeed, since the time of Darwin, evolutionary scholars have been harassed by this question. The principle of natural selection expects individuals to behave in ways that increase their own fitness (i.e. their chances of survival and reproduction), and not that of others. But by behaving altruistically, individuals reduce their own fitness, e.g. they share food and hazard starvation. They thus put themselves at a selective disadvantage as opposed to individuals who behave selfishly and risk their own extinction. So how could altruism evolve, and why has it not been eliminated by natural selection?

Evolutionary scientists have developed ranges of theories to elucidate parts of the puzzle including the most important ones of *kin altruism* (Hamilton, 1964), which explains altruism towards blood relatives, and *reciprocal altruism* (Trivers, 1971), which shows that prosocial behavior is useful whenever individuals can expect to be rewarded by their beneficiaries at a later time. But despite the merits of all those theories, one question has not been answered yet: if altruism is strictly beneficial in specific situations (e.g. if the relationship endures and people can expect to be rewarded), why do people considerably differ in their levels of altruism?

The present work intends to give additional insights to the evolution of altruism by drawing particular attention to the explanation of inter-individual variations in altruistic tendencies. It is supposed that prosocial personality traits might have evolved as a response to demands on the mating market. Singles who explicitly search the mating market for long-term partners, rather than short-term affairs, are likely to look out for moral virtues. It is moreover suggested that inter-individual variations in the supply of prosociality might be due to persons' inter-individual variations in other qualities. For example, people might increase their level of morality to compensate for deficits such as physical unattractiveness. As it will be suggested in the following, individuals who are seeking for a partner adjust their mate preferences for prosociality as well as their preferences for long-term as opposed to short-term relationships according to their own mate value and the environmental challenges they are confronted with (Gangestad & Simpson, 2000).

From an evolutionary point of view, mating has two basic functions: successful reproduction and child raising. The crucial point within the mating game is that men and women differ in their reproduction rates (Trivers, 1972). While for men a single sexual intercourse may suffice, the minimum effort for women includes pregnancy and, at least in pre-modern cultures, a compulsory period of breastfeeding. Moreover, women show a higher level of parental care than their male partners (Low, 1989; Kroska, 2003).

Because women cannot increase their reproduction rate significantly by obtaining more mates, they have to concentrate on quality instead of quantity to maximize their reproductive output. And the key to conceiving and raising high-quality children is to invest in a committed long-term relationship with a mate who possesses both good parenting abilities and good genes (Wilson, 1978). While good parenting abilities are signaled via general tendencies for prosocial behavior (Miller, 2007), research shows that good genes are reliably signaled via signs of physical attractiveness like facial symmetry and averageness (Rhodes, 2006). Moreover, skin quality, hair quality, or a good set of teeth are hints to a person's age and his or her current and lifetime health and thus inform about genetic quality likewise (Sugiyama, 2005). It is no wonder that the fairy tale prince of whom women dream is therefore characterized by two qualities: virtue (as good parent indicator) and handsomeness (as good genes indicator).

However, there is one dilemma: men, in contrast to women, are able to enhance their reproduction rate by inseminating large numbers of females. To maximize the number of offspring, it pays males to be "hasty, fickle, and undiscriminating" (Wilson, 1978, p.129) in that they should strive for frequent short-term mating in form of premarital sex, one-night-stands, and short affairs. And indeed, men report desiring four times as many sexual partners as women do (Buss & Schmitt, 1993). But if aggressiveness and assertiveness in short-term mating is the best strategy for men to maximize the number of mates, why do only some men invest in fleeting sexual encounters?

By far not every man is blessed with the natural gift of attracting numerous women as sexual partners without the need of committing himself to a long-term relationship. The *theory of strategic pluralism* by Gangestad & Simpson (2000) thus states that less attractive men, forced by the mate preferences of choosy women, have to show high commitment and prosocial behavior in order to obtain one long-term partner and raise joint children. At this point, the argumentation conveys how mate choice might have contributed to the evolution of altruism: although prosocial behavior is costly, for some men it might have been a wretched necessity to obtain access to females and to reproduce.

But, “misery loves company” and women are likewise restricted in mate choice, as they face difficulties to acquire a partner who is both highly attractive and highly prosocial, i.e. kind, understanding, caring, generous, trusty, etc. (Gangestad & Simpson, 2000). Empirical findings show that men’s physical attractiveness is positively correlated with unrestricted sociosexuality (i.e. the willingness and ability to engage in multiple short-term affairs), and that unrestricted sociosexuality again is negatively correlated with capacity for close relationships, warmth, responsibility, and trustworthiness (Simpson et al., 2004). Women are thus forced to trade off between a man’s attractiveness and his prosocial virtues. According to Gangestad and Simpson (2000) women should cope with this conflict by choosing their mates in accord with their own qualities as well as the environmental challenges they are confronted with. If the environment is harsh and/or a woman is weak and in particular need of a partner’s support, she should neglect physical attractiveness but should focus on men's prosociality and engage in a long-term relationship. Conversely, if the environment is temperate and/or the woman is a competent provider herself, she should focus on attractiveness (i.e. genetic quality) and should be willing to engage in (extra-pair) short-term relationships with highly attractive men, although this includes the risk of loosing her primary less attractive mate as a long-term benefactor.

This extra-pair short-term mating with an attractive man can be especially profitable when women reach the fertile days of their menstrual cycle, as they might be able to lay a high-quality cuckoo's egg in their long-term partner's high-quality nest (Benshoof & Thornhill, 1979). Empirical results show that women's sexual interest near ovulation reflects this opportunity. In their fertile days, women have a higher tendency to dress up (Haselton, Mortezaie, & Pillsworth, 2007) and to engage in extra-pair flirtation (Haselton & Gangestad, 2006). Moreover, ovulating women against non-ovulating women prefer more competitive men (Gangestad, Garver-Apgar, & Simpson, & Cousins 2007) and the scent of men with more masculine and symmetrical faces (Thornhill & Gangestad, 1999). All these attributes are cues to good genes (Rhodes, 2006; Scheyd, Garver-Apgar, & Gangestad, 2008) and it seems that women tend use them in their fertile days to get the best semen available.

Summing up, the theory of strategic pluralism suggests that men just like women are likely to possess psychological mechanisms which make them susceptible to both short-term and long-term mating.

When looking for a long-term partner, women should be highly interested in finding a good and prosocial provider. This suggestion is supported by empirical findings of Brase (2006) who showed that cues of a positive disposition for parental investment (e.g. caring behavior towards babies or an elderly person) increase women's evaluations of male attractiveness for romantic relationships. By contrast, when looking for a short-term partner, women should be primarily interested in genetic quality, which is signaled through physical attractiveness; prosociality should be a nonsignificant criterion for short-term mate choice.

Indeed, research of Kelly and Dunbar (2001) supports the idea of varying female mate preferences as a function of intends for either short-term or long-term relationships. The scholars asked female subjects to rate personality profiles of men and found that altruism was deemed essential in long-term relationships, while in short-term relationships bravery was much more important. In a recent study of Gangestad et al. (2007), women watched video-clips of men who had been interviewed for a potential lunch date. The women rated the desirability of each man both as a long-term and as a short-term mate. A second sample of women rated each man on ten broad dimensions preferred in long-term and/or short-term mates. Results showed that good partnership/parenting indicators (i.e., faithfulness, warmth, intelligence, potential to be a good father, potential for financial success) predicted men's long-term desirability, while good genes indicators (i.e., arrogance, confrontativeness, muscularity, and physically attractiveness) predicted short-term desirability. Providing additional evidence, Kruger, Fisher and Jobling (2003) showed that in hypothetical scenarios, which were based on passages of British romantic literature, females preferred proper (i.e. law-abiding, compassionate, kind, and monogamous) heroes for long-term relationships while they were more likely to choose dark (i.e. dominant, rebellious, frequently a criminal, and often promiscuous) heroes as partners the shorter the relationship under consideration.

Turning the focus to long-term preferences of men, we suppose that they will try to find someone loyal and prosocial just like women do. Indeed, Li and Kenrick (2006) evidenced that one of men's main criteria for choosing long-term mates is kindness. This preference seems to be adaptive, as men who commit themselves to a long-term partner, should be interested in avoiding women who are likely to conceive by another man and might even pass of a bastard child as theirs (Buss & Schmitt, 1993)

With regard to short-term preferences, empirical research proofs that men prioritize physical attractiveness against all other values (e.g. Li and Kenrick, 2006). But does this mean that men neglect prosociality in potential short-term mates? Until now, this question has not been studied, but indeed, we hypothesize that this is the case. Men who invest in short-term mating intend to quit the relationship soon. However, they should be interested in the survival of potential offspring resulting from the sexual intercourse. One might assume that just because of that, men look out for prosocial women who promise to personify caring single mothers. But we deny this option. However caring a woman is, her prosociality will be useless if she gives birth to unhealthy children and is to weak to pull her progeny through without male support. We therefore think that men are only looking out for attractiveness in short-term mates. As far as we can see, this supposition is braced by literature and art, as god's gift to men has traditionally been illustrated as a slinky vamp – a creature with the genetic quality to survive solitarily but with no cue to warm-heartedness.

One might recognize that our hypothesis that men trade of prosociality with attractiveness implies that women correspond to men in that their prosociality is negatively correlated with their physical attractiveness as well as their willingness to invest in a short-term relationship. This assumption is feasible as women who view themselves as physically attractive were evidenced to be higher in unrestricted sociosexuality and women who are unrestricted in sociosexuality were shown to be less agreeable, not moralistic, and ethically inconsistent (Simpson et al., 2004).

To summarize, the present study explores the connection between mating strategies (i.e. long-term versus short-term mating) and mate preferences (attractiveness versus prosociality). By investigating both sexes and experimentally varying information on prosociality within potential sexual partners, we sought to extend previous research on mate choice criteria, which largely focused on females' preferences and mainly consisted of surveys or scenario studies. We presented short video-clips of target persons to a group of judges of the opposite sex and additionally gave information on the targets' prosociality by announcing how each of the target persons had behaved in a money-sharing task (a 'dictator game'). We asked judges to rate each of the target persons with regard to their desirability as short-term and long-term mates. In doing so we tested the following hypotheses: (1) when evaluating a prospective long-term mate, both men and women are interested in prosociality (besides physical attractiveness). (2) When evaluating a prospective short-term mate, prosociality does not matter, but men and women are inclined to focus on physical attractiveness only.

10.2. Methods

10.2.1. Stimulus material

151 students (77 female, 74 male) of different disciplines were videotaped at the University of Groningen (The Netherlands). Target persons sat in front of a white wall, and introduced themselves into the camera. The video-material was cut into small clips, with each target person being shown for 20 seconds. On the basis of these silent video-clips, 7 male judges rated the attractiveness of female targets (Cronbach's Alpha: .83, Intraclass-coefficient: .42) and 5 female judges rated the attractiveness of male targets (Cronbach's Alpha: .81, Intraclass-coefficient: .46) on a 7-point-scale. All judges were psychology students from the University of Cologne (Germany). Aver-

age ratings of physical attractiveness for target persons ranged from 1.1 for men and 2.0 for women (least attractive targets) to 5.9 for men and 6.6 for women (most attractive targets) with a mean value of 3.19 for men (SD = 1.16) and 4.31 for women (SD = 1.03).

10.2.2. Participants and procedure

Participants were 28 biology students from the University of Groningen, 19 females with a mean age of 21.88 years (SD = 1.66) and 9 males with a mean age of 22.76 years (SD = 2.17). Participants were invited to the laboratory where they received a questionnaire.

In the first part of the questionnaire, participants were familiarized with the logic of the so-called dictator game. Participants had to imagine a situation with two individuals, a dictator and a recipient, who only interact once: The dictator receives 10 Euro and has to decide whether to keep all the money for him/herself or whether to transfer half of the money to an unknown recipient. The dictator knows that his/her identity and his/her decision will always stay anonymous. To avoid influencing the participants, neutral vocabulary was used to explain the dictator game (e.g. the dictator was called Person A, the recipient was called Person B and the dictator game was referred to as ‘distribution task’).

In the second part of the questionnaire, participants were informed that they were now going to watch video-clips of target persons who had taken part in the dictator game in the role of the dictator. Participants were asked to rate each of the target persons with regard to his/her desirability as a prospective mate on a 7-point scale. Next to the scales, the questionnaire entailed information on how prosocially each of the target persons had behaved in the dictator game (i.e. whether they had kept all of money for themselves or whether they had split the money equally). Unknown to the participants, this information had been randomly assigned to the target persons before the experiment took place.

The experiment started when participants had finished reading the instructions. Male participants were shown silent video-clips of female targets. Female participants were shown silent video-clips of male targets. To control for sequence effects in the presentation of target persons, video-clips were randomly shown in two directions (i.e. forwards and backwards). The sequence of target persons did not affect results.

One half of the participants rated the desirability of target persons as long-term mates; the other half rated the desirability of target persons as short-term mates. For a long-term relationship, participants were asked to imagine high emotional involvement and the potential of a life-long bond. For a short-term relationship, participants were asked to imagine a fleeting sexual affair without considerable emotions. Participants were randomly allocated to the two groups. With a time gap of one week, participants were shown the video-clips again. This time, participants had to accomplish the respective opposite task (i.e. participants who had rated short-term desirability in week 1, rated long-term desirability in week 2 and vice versa). The sequence of rating long-term and short-term desirability in either week 1 or week 2 did not affect results.

10.3. Results

Desirability as long-term and short-term mates had been measured on a 7-point scale. When judging long-term desirability, participants ascribed target persons a mean value of 2.21 (SD = .97). When judging short-term desirability, a mean value of 2.76 (SD = 1.36) was ascribed.

To explore the associations between short-term/long-term desirability of targets and their prosociality, physical attractiveness and gender, we initially conducted intercorrelations between the variables (see Table 5).

The desirability of individuals as long-term mates correlated with their desirability as short-term mates ($r = .66$, $p < .01$), but was not identical. As predicted, people preferred prosocial individuals against selfish individuals when evaluating a potential long-term mate ($r = .32$, $p < .01$). However, when evaluating a potential short-term mate, people preferred selfish individuals against prosocial individuals ($r = -.19$, $p < .01$). Physical attractiveness was stronger correlated with short-term desirability ($r = .77$, $p < .01$) than with long-term desirability ($r = .52$, $p < .01$; Fisher's $z = 3.94$, $p < .01$). Gender was associated with desirability as a short-term mate ($r = .38$, $p < .01$), indicating that men cherished women as short-term mates more than women cherished men as short-term mates. Finally, gender was related to physical attractiveness ($r = .48$, $p < .01$), indicating that men perceived women as more attractive than the other way around. Because the information whether an individual had behaved prosocially or selfishly in the dictator game had been randomly assigned to the target persons, as required, prosociality was uncorrelated to gender and physical attractiveness.

	1. Desirability as long-term mate	2. Desirability as short-term mate	3. Prosociality: Selfish = 0, Prosocial = 1	4. Physical attractiveness	5. Targets' gender: Male = 0, Female =1
1.	—	.66**	.32**	.52**	.11
2.		—	-.19**	.77**	.38**
3.			—	-.14	-.01
4.				—	.48**
5.					—

Table 5: Intercorrelations between the criteria variables desirability as long-term and short term-mate, and the predictor variables prosociality, physical attractiveness, and the gender of the target person. *Note.* ** $p < .01$

Summing up bivariate results, our hypothesis that long-term desirability is positively associated with prosociality was strongly supported. In contrast, short-term desirability was negatively associated with prosociality. Physical attractiveness was relevant for both short-term and long-term desirability.

To investigate whether the same relations held true on a multivariate level, we conducted two regression analyses. First, we regressed long-term desirability on the predictor variables prosociality, physical attractiveness, and gender (R^2 = .45, F = 40.485; see Table 6).

Predictor variable	B	SE B	*ß*
Selfish = 0; Prosocial = 1	.82	.12	.41**
Physical attractiveness	.58	.06	.67**
Male = 0; Female = 1	-.42	.14	-.21**

Table 6: Association between the desirability of individuals as long-term mates, their prosociality, their physical attractiveness, and their gender. *Note.* R^2 = .45; ** p < .01

As hypothesized, prosociality significantly influenced the desirability as a long-term mate (*ß* = .41, p < .01): individuals who reportedly had split the money equally in the dictator game received .82 points more on the 7-point desirability scale than individuals who reportedly had kept everything for themselves. As was expected, physical attractiveness also predicted long-term desirability (*ß* = .67, p < .01). Long-term desirability of female targets was lower than long-term desirability of male targets (*ß* = -.21, p <.01). This finding contrasts bivariate results, which showed that gender and long-term desirability were uncorrelated (r = .11, n.s.). Indeed, the negative beta weight of gender in the regression analysis results from the high intercorrelation between gender and the second predictor physical attractiveness (r = .48, p < .01). Instead of explaining valid variance within the regression, gender accounts for error variance in attractiveness. Thus, gender functions as a suppressor and its negative beta weight has to be judged a methodological artifact. In sum, our findings show that both males and females judge the opposite sex the more desirable as a long-term mate, the more prosocial and the more physically attractive they are.

Table 7 gives parameter estimates for the regression of short-term desirability on the predictors prosociality, physical attractiveness, and gender ($R^2 = .61$, $F = 75.023$).

Predictor variable	B	SE B	ß
Selfish = 0; Prosocial = 1	-.17	.11	-.08
Physical attractiveness	.65	.05	.75**
Male = 0; Female = 1	.04	.12	.02

Table 7: Association between the desirability of individuals as short-term mates, their prosociality, their physical attractiveness, and their gender. *Note.* $R^2 = .61$; ** $p < .01$

Only physical attractiveness significantly influenced short-term desirability (ß = .75, $p < .01$). Neither prosociality nor gender influenced multivariate results, although they had been moderately correlated with short-term desirability on a bivariate level. In sum, the regression shows that for both males and females the single significant criterion for an individual's desirability as a short-term mate is his/her physical attractiveness.

10.4. Discussion

The purpose of the present study was to test the hypotheses derived from the theory of strategic pluralism (Gangestad & Simpson, 2000) that individuals are focusing on prosociality when searching for a long-term mate, but that they do not care for prosociality when looking for a short-term mate. Both hypotheses were supported for both sexes.

Our empirical findings back up recent theoretical considerations of Miller (2007) who argued that mate choice might have shaped the evolution of moral virtues. Indeed, the existence of prosocial behavior is engaging evolutionary theorists for a long time, because it is costly to perform, and why should such self-detrimental behavior survive the pressures of natural selection? This study shows that prosociality may serve as a means to obtain sexual partners. It thus goes with findings of Farrelly (2007) who showed that people intuitively seem to be aware of the magnetism of altruism as they preferentially direct prosocial behavior towards more desirable members of the opposite sex. It seems fair to conclude that especially those individuals who are unable to attract short-term mates due to a lack of physical attractiveness are well advised to invest in prosocial acts, as they serve as good partnership and parenting indicators, and may thus pave the way to reproduction. Indeed, the notion that an individual's disposition for prosociality should be contingent on his or her other qualities may account for the great inter-individual variance in prosociality which is still troubling evolutionary theorists.

Complementing research on mate choice, our results show that both men and women may contemplate long-term and short-term mating as sexual strategies but that the application of each of these strategies is related to distinct preferences, namely that prosociality is valued in long-term relationships, but that it is irrelevant when choosing a short-term mate. Basically, these results are in accord with most evolutionary theories of mate choice. However, our findings are identical for both sexes. To our impression, this result strikes, as in mainstream evolutionary theory the topic of mate choice has been ruled by the question how hard-to-please women make their careful choice. Preferences of men have been taken note of rarely, and if researchers did, they often assumed them to be simple and archaic: a "real hottie" should be very young, and highly attractive. However, in this study, men corresponded to women in appreciating prosociality as a core asset in a long-term mate.

Prosociality was operationalized as altruistic behavior in a dictator game. We can only speculate about the way participants might have interpreted this cue, but it seems likely that they took prosocial behavior in the dictator game as an indicator for general altruistic tendencies. This heuristic makes sense as empirical research shows that dictator game behavior correlates significantly with prosocial personality dispositions (Fetchenhauer & Huang, 2004).

As theorized at the outset, women's strive for prosocial partners is motivated by their need for long-term commitment and parental investment. However, men's strive for prosociality in long-term relationships can be interpreted in two ways. First, men might be looking for a deeply devoted wife who is unlikely to conceive by another man. But second, men might be searching for a woman who cuts a warm and loving figure as mother of their children. However, if the second explanation was the driving force, men should have applied the same standards of prosociality to short-term mates, as a woman with caring abilities should have been ever more valuable if men anticipated that they themselves would omit parental care. As men did not care for prosociality in short-term partnerships, their preference for prosociality in long-term partnerships seems to reflect their primary interest in finding a loyal wife rather than their interest in finding a loving mother for potential off-spring.

Commenting on the signaling value of prosocial traits, our results are substantive with regard to a hypothesis formulated by Miller (2007). The author argued that moral virtues advertise two classes of traits: good partnership/parenting abilities and good genes. With regard to prosociality as good genes indicator, Miller referred to costly signaling theory (Zahavi & Zahavi, 1997) and argued that prosocial virtues are difficult to display if one has a high mutation load that impairs the precision of brain development. For example, many mutations associated with autism, schizophrenia and other disorders disturb the development of a complex Theory of Mind, which is essential to display empathy and considerate behavior. Therefore, Miller states, moral virtues may serve as a “neurogenetic warranty” for good genes (Miller, 2007, p. 101). However, our data suggest that people do not use prosocial behavior as cue to genetic quality. If prosocial behavior was used as a good gene indicator, it should have been a significant criterion for choosing a short-term mate. But this was not the case—neither in our study, nor in the study conducted by Gangestad et al. (2007) who investigated short-term and long-term preferences of women. Therefore, it seems, that individuals interpret prosociality—at least prosocial dictator game behavior—as a cue to good partnership/parenting abilities but not as a cue to good genes.

To conclude, our results suggest that the evolution of human prosociality as well as its amplitude, which sharply distinguishes the human race from the rest of the animal kingdom, has been underpinned by sexual selection. More than any other species in the world, humans are in need of biparental care to achieve reproductive success. This need is reflected by women’s mate choice preferences for prosociality when they are in search of a long-term partner. But going beyond child-raising, a long-term, perhaps lifetime bond may yield considerable pay-off in various other domains for both partners, as a good couple can function more efficiently than either individual alone (Buss & Schmitt, 1993). Therefore, men, just like women, value prosocial personality dispositions in potential long-term partners. As individuals seem to trade off between physical attractiveness and prosociality depending on whether they are looking for a short-term or long-term partner, it seems fair to assume that inter-individual variations in prosocial personality dispositions are a result of inter-individual variations in other mating relevant criteria, particularly physical attractiveness. In other words: If you are equipped with a skew nose and a belly like a bear, you might not be an amorist but as long as you have a great heart you can nevertheless win the partner in life you dream of.

11. General discussion

11.1. Conclusions

This work aimed at shedding light on the evolution of human altruism. In the theoretical part, a number of explanations to the phenomenon were reviewed. On the one hand, it was concluded that genetic group selection for altruism has to be rejected. On the other hand, it was noted that the most influential individual-level theories of kin altruism (Hamilton, 1964) and reciprocity (Trivers, 1971; Alexander, 1987) may explain a lot. However, these theories are insufficient to justify the evolution of ultra-social behavior in humans, i.e. behavior which traces back to human conscience.

Human conscience was conceived of as an internal moral authority that, superficially analyzed, contradicts human biological desires in that it provokes maladaptive behavior. However, daring a rather profound view, it was argued that as to the complexity of culture, the evolution of human conscience might yet again have been an adaptation to master compound social challenges.

Modern humans face multifaceted social decision problems everyday. They are able to solve these problems because they have distinctive psychological features. First of all, their cognitive capacity, which can be perceived as their psychological *hardware*, enables them to reflect on the characteristics of complex situations. Moreover, the internalization of cultural norms, that is human psychological *software*, allows them to evaluate the pros and cons of action alternatives and to respond to the situation in a culturally appropriate way. Cultural conformity is to the best advantage of most present-day individuals in the majority of cases. In this way, it was argued that the conscience functions as a superior control authority, which enables humans to deduce adaptive behavior patterns.

Two theories were introduced which offer explanations of how human conscience might have evolved, the theory of the extended phenotype and the theory of gene-culture coevolution. As it was discussed, both theories provide valuable ideas; nevertheless they bear certain weaknesses that are particularly linked to the fact that they make non-proven assumptions. However, although the reasons for the evolution of conscience have not yet been explained satisfactorily, the statement rests that humans have a moral conscience. But does this mean that individuals whose conscience leads them to unconditionally comply to moral standards win in the evolutionary game?

As discussed, many theorists doubt this assumption. They rather assume that true altruism is a maladaptation. If this was right, opportunistic individuals with Machiavellian intelligence (Byrne & White, 1988; 1997), who display altruism when they are observed, but switch to strict egoism when facing an anonymous situation, should gain the upper hand in the long run.

Offering an alternative to this desolate scenario of the future of mankind, the theory of assortation, in particular the commitment-model of Frank (1988, 2008) was raised which assumes that true altruism may stand the pressures of natural selection, because true altruists are able to identify other true altruists and consequently choose one another for mutual cooperation. Because cooperating altruists reach extraordinary benefits through mutualism, they may out-compete egoists (i.e., strict egoists as well as opportunistic individuals with Machiavellian intelligence).

The commitment model makes two assumptions: (1) true altruism is identifiable, and (2) individuals assort themselves along the dimension of altruism. We conducted two studies to test whether these hypotheses are supportable.

In Study 1 we applied the thin slices paradigm and investigated whether individuals are able to estimate the level of altruism of unacquainted target persons on the basis of 20 seconds of silent video clips. As target persons had been videotaped in a setting completely unrelated to altruistic behavior, it could be checked whether individuals were able to make out permanent cues to situation-independent altruistic traits. The results were striking as individuals estimated with notable accuracy how altruistically *complete strangers* had behaved when these strangers had faced a situation in which they had been unobserved. Hypothesis 1 that true altruism is identifiable was thus supported.

Study 2 tested whether individuals in genuine groups can identify the altruistic tendencies of their daily interaction partners. It was further observed whether altruism influences the formation of friendships in such that individuals assort themselves along the dimension of altruism. Students of six secondary school classes played an anonymous dictator game. Afterwards and unannounced, the students had to estimate their classmates' decisions. Again, subjects' estimations of the altruistic behavior of others were better than chance. Thus, our second study secured the validation of hypothesis 1 that altruism is identifiable. Moreover, hypothesis 2 was supported, as altruistic subjects preferably chose other altruists as friends.

In Study 2 we additionally investigated in how far social closeness (i.e., friendship, liking, dislike, and indifference) influenced the accuracy of predictions. Estimates concerning the dictator contribution of a friend were more valid than predictions about the contributions of liked classmates, as it was expected from the intensity of communication and contact in friendships. Predictions about the contributions of unrelated classmates were nothing but random. However, the accuracy for predicting the dictator contributions of disliked classmates was high and did not differ significantly from the accuracy of predicting the behavior of friends. It was argued that a valid evaluation of exactly these two groups—friends and disliked persons—is adaptive. Persons ought to know about the true intentions of their best friends, because they are interacting with them frequently and are thus very dependent on correct assessments of their characters. People should further be aware of the intentions of disliked individuals, that is, those with different interests, because they are potential adversaries.

However, balancing the results of Study 1 and 2 against each other, why were people astonishingly accurate when predicting the behavior of complete strangers (Study 1), but unable to predict the altruistic behavior of classmates whom they met with indifference (Study 2)?

The disparity of these results is probably linked to the fact that subjects in Study 1 and Study 2 processed divergent information to predict the behavior of their respective target persons. In Study 1, subjects almost certainly relied on their intuitive personality judgment abilities. It has been shown that judgmental accuracy in thin slices studies is generally linked to intuitive impressions of nonverbal behavior (Ambady et al., 2000). Especially global molar impressions, although fuzzier and vaguer than codings of specific behaviors, were shown to yield most accurate personality judgments (Ambady & Rosenthal, 1993). Probably, such global and intuitive character assessments are made only if one encounters someone else for the first time in life.

In contrast to the thin slices paradigm that explicitly called for the activation of human intuitive powers of person perception, the design of Study 2 probably motivated subjects to recall the *reputations* of their classmates rather than making spontaneous judgments about their altruistic tendencies. Obviously, reputations of classmates met with indifference were simply less clearly figured and thus less valid than reputations of classmates towards whom participants had a firm attitude, be this attitude positive or negative. Research by Anderson and Shirako (2008) supports the view that the validity of reputations is dependant on the social connectedness of the person who possesses the reputation. The authors found that on average, individuals' reputations are only moderately related to their record of behavior, but that the link between reputation and actual behavior is more intense for people who are more well-known and receive more social interest in the community than for less well-known people.

Therefore, results of Study 1 and 2, which at first glance seem to be contradictory, can be assumed to simply reflect different kinds of personality judgment abilities in humans: on the one hand, it seems that people are quite accurate in making rapid and intuitive judgments on the altruism of unfamiliar persons. On the other hand, people use reputations to evaluate the altruism of known persons; these reputations are the more valid the more socially visible the target person is.

In sum, both studies confirm the existence of the two prerequisites for the evolution of altruism through assortation: the predictability of altruistic behavior and the association of altruists. Both mechanisms were simultaneously evidenced in the natural context of school classes. This suggests that our altruistic ancestors were likewise able to identify other altruists within their groups and to consolidate alliances. Going further, it seems feasible to assume that our ancestors, when meeting a stranger, were able to decide at short notice whether it was useful to establish a friendship with this person or whether it was more appropriate to face this person with caution. Therefore, it seems likely that positive assortation supported the evolution of altruism.

Although the theory of assortation may explain the evolution of altruism in general, it does not explain the existence of inter-individual differences in altruism. In Study 3, the focus was therefore turned to a different phenomenon that was supposed to shed light on this issue: sexual selection. Based on the theory of strategic pluralism (Gangestad & Simpson, 2000), it was hypothesized that men and women, when looking for a partner, should be interested in both altruism (as indicator of good partnership and parenting abilities) and physical attractiveness (as indicator of genetic quality), but that these traits are difficult to get simultaneously: highly attractive people are able to acquire mates without being nice; they will tend to engage in frequent short-term mating. Nice people are likely to show altruism *because* they have to compensate for deficits in attractiveness. For them, altruism is a necessity to obtain at least one long-term partner.

In view of the ensuing trade-off between attractiveness and prosociality, which lonely-hearts face, we investigated whether the desirability of prosociality varies depending on whether individuals are looking for a short-term or a long-term mate. As moral virtues are the more important the closer the relationship to a partner, Study 3 assumed that (1) altruism is highly valued in potential long-term mates, but that (2) altruism is insignificant for the judgment of potential short-term mates. Judges rated the short-term and long-term desirability of target persons on the basis of short video-clips as well as on the basis of information on the level of altruism of each target. Both hypotheses were supported: altruism pushed a target's long-term desirability, but it was insignificant for a target's short-term desirability.

The results suggest that although altruism is costly, at least for some individuals, especially for those who lack in physical attractiveness, it is a wretched necessity to obtain access to mates and to reproduce. Hence, it seems fair to assume that the evolution of inter-individual variations in altruistic personality dispositions has been effected by inter-individual variations in the propensity to engage in either short-term mating or long-term mating which in turn is influenced by an individual's variations in other mating relevant criteria.

The three studies presented in this work provided small, but certainly valuable pieces for solving the puzzle of human altruism. In the following section, propositions for future research will be made that are designed to address both limitations of the previous studies as well as novel research questions.

11.2. Future Research

11.2.1. A further investigation of assortation processes: Do prosocial university students mingle with each other and are they more successful than egoists?

Present-day humans are willing to cooperate in anonymous and large groups, although there are high risks of exploitation. In the theoretical part of this work, it was argued that this phenomenon might be due to assortation processes. In Study 2 of the empirical part, it was shown that assortation is a real-life fact: school students were able to estimate the altruism of their classmates and assembled in friendships along this dimension. However, the classes that we investigated consisted of 14 to 29 students only. If anything, our design thus resembled the structural features of an ancient small-scale society (e.g., a small band), but it did not grasp the characteristics of present-day large, and anonymous societies. The question whether large-scale cooperation exists due to assortation of altruistic individuals is thus still open to conjecture.

Study 1 indicated that assortation might be possible under quasi-anonymity, because it proved that humans judge the altruistic tendencies of complete strangers better than chance. But even if altruism is identifiable in unknown persons, do altruists, who meet in more or less anonymous contexts, assemble over time? And above all, are altruists in the end really more successful than egoists?

The University of Cologne, which is a quite anonymous association of people as it consists of over 44.000 students, seems to be the perfect place for conducting a longitudinal field study concerned with the investigation of these questions. To specify, it could be observed whether altruistic university freshmen over time mingle with other altruistic students and whether they are more successful in their studies, because, for example, they achieve better results whenever group work is demanded and cooperate in exchanging protocols of courses.

To research the issue, university students should be recruited at the beginning of their studies (e.g., in the introductory lecture on economic and social psychology at our department). Over the time of their studies, students should be tested with respect to different characteristics. There should be a specific pattern on the basis of which students are able to construe an identification code (id). Each time they participate in a survey, they should be asked to indicate the id so that their data could be accumulated on an individual basis.

In the first surveys, besides demographics, prosociality should be measured. To grasp different dimensions of prosociality, scales could be used, but a variety of economic games should be applied likewise. For example, the dictator game could be used to identify true altruism free of strategic concerns (for a description see Chapter 8.2). In addition, a trust game could be played to identify trustfulness and/or trustworthiness. In this game an investor receives an endowment e and can transfer a part of that endowment p via the experimenter to a trustee. The investor knows that the experimenter will multiply the transferred sum (e.g., times 3). The trustee receives the multiplied amount of money and decides how much to send back to the investor. If the trustee is completely trustworthy in that he/she sends 50 % of the received amount back to the investor, the best strategy for the investor is to initially transfer the whole endowment ($p = e$). The use of a trust game thus allows for measuring both trustfulness, if the participant is in the role of the investor, and trustworthiness, if the participant is in the role of the receiver.

In the following surveys, diverse control variables should be measured. For example, students should be asked to fill out intelligence tests, and personality scales (e.g., the “Big Five”). Finally, cooperative behavior in the context of studying should be surveyed with specific items (e.g., “How often do you lend out written notes to fellow students?” “Do you enjoy working together in study-groups?” etc.).

To investigate the relationship between prosociality and studying success, each individual’s data would have to be associated with his or her grades. To accomplish this, in the following semesters whenever grades are announced, students should be asked to inform the department about their respective id-grade-linkage. To assure anonymity, we would optimally construe an online-tool that students could use to make their announcement. This tool could become an integral part of the department’s website, so that students could use it whenever they visit the website to find out about their grades.

To investigate in how far students associate along the dimension of altruism, the grouping of individuals in small courses could be analyzed. Especially the experimental course, which economic and social psychology students in Cologne enroll at the end of their duration of studies, provides a good opportunity to identify cooperative alliances. At the beginning of this course, students should be asked to indicate their degree of acquaintance with every course member. It would be interesting to see whether some individuals are already acquainted with each other and whether these acquaintanceships have to do with similar levels of prosociality between the course members. The students should moreover be asked to estimate the level of altruism of each of their fellow students, and to indicate their willingness to cooperate with each of them.

Students in this experimental course usually have to engage in a lot of teamwork during the semester. It would thus be interesting to see in how far additional interaction and communication influences the perceived level of acquaintance between individuals, the validity of their judgments about each other, as well as their choice preferences concerning the cooperation with specific course members. Therefore, at the end of the course, the questions that were asked at the beginning of the semester, should be asked for a second time. To ensure experimenter-subject-anonymity (i.e. docent-student-anonymity) in this specific part of the study, subject numbers should be used when students make their evaluation of each other. In addition, the docent should leave the classroom during the experiment.

How could students be convinced to participate in this study? Students should be told that (1) we are trying to assess whether there are any correlations between personality characteristics and success in studying our subject, and (2) that their participation would, in the long-term, give us the opportunity to coach our students in the most effective way. However, this announcement should only be made after students had played the economic games, as otherwise, this announcement could increase demand characteristics that might influence the level of prosocial behavior of students displayed in economic games—and this would be fatal, as prosocial behavior is the most crucial variable of this study.

To my view, conducting such a study would be valuable for several reasons: (1) it allowed for the investigation of the prosociality-success-linkage as well as for the investigation of assortation processes in a large and more or less anonymous group; (2) it allowed for the collection of data in a real-life setting; (3) it provided high-quality data at moderate surveying costs; and (4) the results of this study might moreover provide a basis to optimize the coaching of subsequent generations of students of our subject.

Having discussed possible future research connected to the theory of assortation, in the next sections the focus will be drawn to research on sexual selection.

11.2.2. Preferences for prosociality depending on own providing capabilities and situational challenges: Are jobless women particularly prone to “good guys”?

The theory of strategic pluralism (Gangestad & Simpson, 2000; see Study 3 in Chapter 10) suggests that men and women were selected to use both long-term and short-term mating tactics. With regard to women, Gangestad and Simpson (2000) predicted that they evolved to trade off between the genetic fitness of men and their partnership/ parenting abilities, whereas the specific mating strategies and preferences which women adopted depended on the nature of their local environment. If the local environment was complicated and demanded biparental care, women deemed the investment potential of prospective mates more valuable than their genetic fitness and consequently adopted long-term mating tactics almost exclusively. If, in contrast, pathogens were common in the local environment, or the environment signaled the weight of genetic fitness of offspring in other ways, women deemed indicators of genetic fitness of prospective mates more valuable than their partnership/parenting abilities. In such environments, a higher number of women should have engaged in short-term, extra-pair mating to acquire genetic benefits from men who offered less parental investment, although this strategy included the risk of loosing their primary long-term mates. Based on these suggestions, Gangestad and Simpson (2000) concluded that the mating tactics and preferences of women should have varied *between* populations as to the inter-population differences of environmental challenges.

However, Gangestad and Simpson (2000) additionally assumed that preferences of women should have varied *within* populations, as women should have inter-individually differed in the extent to which they could have profited from obtaining genetic versus material benefits. If some women were good providers themselves, they could have primarily enhanced their fitness by mating with men who provided better genes. These women should either have preferred long-term mates with particularly good genes (if they were to get), or they should have engaged in opportunistic extra-pair mating with such men. In contrast, if other women could have enhanced their fitness by obtaining mates who offered more material benefits, these women should have preferred long-term mates with superior partnership/parenting abilities and they should have tended to renounce extra-pair mating completely to circumvent the loss of their primary mate by all means.

The connection between women's sociosexuality and their personality traits provide indirect support for these assumptions. Unrestricted women, who are prone to short-term mating, tend to be more socially dominant, more extroverted, and less harm-avoidant (Gangestad & Simpson 1990). As Gangestad and Simpson (2000) note, each of these traits should facilitate the gaining and maintenance of independent resources and accordingly increase women's willingness to waive long-term relationships.

However, as Gangestad and Simpson (2000) themselves noted, direct support for the assumed correlations is mixed. On the one hand, differences in mate preferences *between* populations were supported. Across the cultures surveyed by Buss (1989), Eagly and Wood (1999) found that an increase in women's access to resources and power is linked to a lower mate preference for financial success, which can be assumed as an indicator of good partnership/parenting abilities. But, on the other hand, within cultures this relationship could not be evidenced: Women with high-paying jobs tend to value resources as much as or even more than women with lower paying jobs do (e.g., Buss 1989; Townsend 1989; Wiederman & Allgeier 1992). Gangestad and Simpson (2000) offered an explanation for this contradiction. They assumed that the effects of women's access to resources across cultures might be especially high because what women learn about the value of a mate's resources is shared within a culture.

However, if mate preferences exist for central biological reasons, to my view, women—also *within* populations—should differ with regard to their mate preferences for prosociality, as eventually every culture consists of subcultures. Thus, women who lack financial resources should be more interested in finding prosocial men than women who are good providers themselves. Actually, it seems likely that results of Buss (1989), Townsend (1989), and Wiederman and Allgeier (1992), which showed a positive link between women's own wealth and their interest in resources of a potential partner, were not so much triggered by their pursuit of prosociality (i.e., an indicator of partnership/parenting abilities), but rather driven by their pursuit of male status (i.e., a good genes indicator). And indeed, the theory of assortative mating would expect that high-status women look out for high-status men.

To find out whether a lack of financial autonomy in women is nevertheless linked to a higher preference for good male providers, a study should be conducted, in which providing abilities are operationalized as prosocial behavioral tendencies, rather than the absolute amount of financial resources a man has. The design of Study 3 of this work (see Chapter 10) could be used as a model for such a study, as it met this demand. As a reminder, we investigated whether partnership/parenting abilities are an important factor for long-term mate choice, whereas the key feature of our study was to experimentally manipulate partnership/parenting abilities by giving information on the degree of prosocial behavior of the potential mate (instead of the height of financial resources).

To explore whether women's mate preferences for prosociality indeed vary as a function of their own providing abilities, two groups of women could be studied: working women and unemployed women. Unemployment is connected to a decreased access to material resources and a reduction in psychological well-being with an increased rate of depression and anxiety (Wilson & Walker, 1993). It thus can be hypothesized that unemployed women show a greater preference for prosocial men than working women, because prosocial men signal a higher sense of responsibility and willingness to care for their partner's well-being and supply.

To survey unemployed women, one part of the study could be conducted in employment offices. To survey working women, interviewees could recruit participants on the street or, preferentially, in public buildings. Demographic characteristics other than occupational status should be held constant (i.e., samples should be similar in terms of age, residence, and marital status, etc.).

As noted, the design of the study should be based on the design of Study 3. For reasons of practicability, photos of target persons, rather than video-clips should be used. This approach would allow surveying participants on the street and in public buildings without the need of additional technical equipment. Women should be given questionnaires with photos of target persons and information on their prosociality (i.e., information on the dictator game behavior of each target). To control for methodological effects, photos and prosociality-information should be varied (see Chapter 10.2.2). Women should be asked to rate the desirability of men either as short-term mates, or as long-term mates.

The overall-design of the experiment would be a 2 (occupational status: employed vs. unemployed) X 2 (mating context: short-term vs. long-term) X 2 (sequence of photos: forwards vs. backwards) X 2 (sequence of information on prosociality of target: forwards vs. backwards) – between-subject-design.

According to the theory of strategic pluralism, unemployed women, as opposed to working women, should (1) place more weight on prosociality in long-term partners, and (2) should have a decreased desire for short-term mates, especially if they were in a relationship at the moment of the interview so that engaging in an extra-pair short-term relationship included the risk of loosing their primary mate.

To my view, the conduction of this study would be valuable, because it allowed for a more precise investigation of hypotheses derived from the theory of strategic pluralism, namely that individuals adjust their mating strategy, and accordingly their mating preferences, depending on the environmental challenges they are confronted with, as well as their personal capabilities. Exactly these two variables differ between working and unemployed people, as unemployment may implicate both situational problems as well as personal deficiencies. In particular, an examination of unemployed women allowed for the investigation of concrete hypotheses under conditions of high ecological validity. Moreover, as unemployment is a severe problem nowadays, a deeper understanding of the dimension of its effects could be valuable for our society. If jobless individuals indeed changed their mating habits, this again would be evidence for the severe impacts of unemployment on all areas of life. Such results should once again sensitize society for its responsibility towards its weakened members.

In fact, an investigation of the effects of unemployment on *male* mating strategies would be likewise interesting. While strategic pluralism theory does not predict changes in male mating strategies as a result of environmental challenges, it does predict that men’s propensity to engage in short-term mating is contingent on their ability to satisfy the short-term mate preferences of women. Thus, their tendency to perform short-term strategies should be a direct function of their genetic qualities, whereas men’s tendency to invest in single, committed long-term relationships should be inversely related to their genetic fitness. If unemployment was a signal to low genetic fitness—or was at least perceived as such a signal in the eyes of women—unemployed men should be more inclined to look out for long-term mates. However, unemployment is also a signal for decreased providing potential. Thus, although unemployed men might be inclined to signal more sympathy and warmth, they will be less able to signal providing capabilities in terms of resources. Whether and how men change their mating strategies in consequence of unemployment is less predictable than in the case of women. However, it would be likewise interesting to examine.

Finally, besides studying the effects of unemployment on mate preferences for altruism, it would be likewise interesting to investigate in how far unemployment alters an individual's propensity to display altruism. As it was discussed, altruism is a costly signal. It might be that unemployed individuals are thus less prone to demonstrate this virtue.

Although the topic of unemployment will be left aside in the following, the general idea of variations in the display of altruism will be highlighted in the next section. To specify, it will be asked whether different mating targets elicit divergent efforts to display altruism as a costly signal.

11.2.3. Conspicuous courtship display: Does "Meg Ryan" provoke more altruism than "Angelina Jolie"?

As discussed in the theoretical part of this work, costly signaling theory may help to explain altruistic behavior (Zahavi, 1975; Zahavi & Zahavi, 1997). It suggests that individuals frequently engage in costly behaviors as a means of signaling to others positive information about themselves. As altruism signals an individual's ability to incur costs without the need of reaping benefits, such displays enhance an individual's status and prestige (e.g. Hardy & Van Vugt, 2006), and, as a consequence may improve the individual's ability to attract and keep desirable mates.

Based on hypotheses derived from costly signaling theory, Griskevicius et al. (2007) recently investigated the idea that self-sacrifice in the context of mating might actually be self-presentation. In four experiments, the authors investigated whether the triggering of mating motives influences the display of conspicuous benevolence. In women, mating goals increased public, but not private, helping. In men, mating motives did not increase helping in general (e.g., mailing a letter that someone had dropped on the way to the post office), but mating goals increased helping in contexts in which men could display heroism or dominance (e.g., diving into icy water after a stranger falls from a boat in a storm).

However, against the background of strategic pluralism theory (Gangestad & Simpson, 2000), men should adjust the specific kinds of conspicuous courtship displays according to the characteristics of the mating context. They should signal good genes virtues selectively in short-term mating, while they should display good partnership/parenting abilities primarily in long-term mating contexts (Miller, 2007). To test whether this hypothesis holds true, a study could be conducted which examines the effect of short-term versus long-term mating motives on men's willingness to show prosocial behavior. Men could be invited to the laboratory. Different mating motives (i.e., either short-term or long-term mating) could be induced. Afterwards, it could be measured whether the level of prosociality differs as a function of mating motives.

To induce mating motives, subjects could be exposed to photos of attractive females and they could be asked to imagine and write down stories about potential dates with these persons.

To measure the prosociality of men, a dictator game could be used. If long-term mating motives indeed increased dictator game contributions of men, this would retroactively justify the approach that we employed in Study 3. As a reminder, in that study we assumed that women interpreted prosocial dictator game behavior of target persons as valid cues to men's willingness to engage in a committed long-term relationship.

However, if results of Griskevicius et al. (2007) are representative, measuring prosociality with a dictator game is probably not sufficient, because in their study conspicuous benevolence was only evidenced for situations in which men could display heroism and dominance. The dictator game is characterized by anonymity. Thus, heroism, which implies publicity, can hardly be constituted in this game. Therefore, a second measurement of prosociality should be applied. Following the example of Griskevicius et al. (2007), specific items could be developed to measure different forms of helpfulness, that is, non-heroic as well as heroic helpfulness.

The overall design of the experiment would be a 2 (participant sex) X 3 (mating motive: short-term vs. long-term vs. control) X 3 (Prosociality: dictator game vs. non-heroic helpfulness-items vs. heroic helpfulness-items) mixed-factorial design. Sex and mating motive would be between-participants factors, while prosociality would be a within-participant factor, meaning that everyone would have to play a dictator game and would have to answer all questions about non-heroic and heroic helping.

Subjects would be invited to the laboratory. The study should preferably be accomplished in separate parts. To avoid potential suspicion and biased responses, a cover story should be devised to separate the different parts of the study. For example, participants could be told that they were going to take part in three unrelated studies in one session (i.e., participating in a money-sharing-task, story-writing to examine their visualization talent, and answering survey questions about diverse behavioral preferences).

Part I: Triggering of mating goals through story-writing

Subjects would be accompanied in a room, where they were left alone to fill out a questionnaire. In the control condition, a poster of a person of the same sex should hang on the wall of the lab room. In the test conditions, posters with attractive persons of the opposite sex should hang on the wall to induce a mating state. These posters should either show individuals desirable as short-term mates, or individuals desirable as long-term mates. To illustrate, if a man was in a short-term condition, he should be exposed to a vamp-like woman (e.g., someone looking like Angelina Jolie). If he was in a long-term condition, he should be exposed to a warm-hearted woman of the "girl next door"-type (e.g., someone looking like Meg Ryan). To meet different tastes, it might be considerable to hang up two photos in each category, (e.g., an attractive vamp with blond hair and another one with dark hair). Prior to the study, posters should have been evaluated to assure that equal "overall attractiveness" of short-term and long-term stimuli is given and that the subjects shown on the posters effectively match the stereotypic category of either short-term or long-term mates.

Subjects should be offered a seat in front of the posters. They should be asked to look at the posters and to select the person whom they thought was the most desirable sexual partner. Then, participants should be asked to imagine a date with this individual and to write down their ideas about this date. In the short-term condition, subjects should be asked to imagine a typical short-term date with the stimulus person (e.g., a holiday flirtation or a one-night-stand). In the long-term condition, subjects should be asked to imagine a date connected to a typical long-term relationship (e.g., a first romantic dinner with pleasant conversation during which they developed the wish to romantically pursue the stimulus person). In the control condition, participants should imagine to be friend with the same-sex person shown on the poster. They should, for example, imagine going out clubbing with that person. As a manipulation check, a content analysis should be made, classifying the vocabulary which participants used in their stories into two groups: words typically related to (1) *short-term mating* (e.g., unrestrained sex) or (2) *long-term mating* (e.g., an affectionate glance). The amount of words connected to these categories should significantly differ between the three conditions.

Part II: Measuring dictator game behavior

Afterwards, subjects should be given a new questionnaire in which the logic of the dictator game should be explained (see Chapter 8.2). The experimenter should leave the room. Participants should be asked to play the dictator game in the role of Person A with real money (e.g., 10 Euro in coins). Money should be given to them in an envelope. Subjects should be asked to freely decide how to split the money between themselves and an unknown Person B. Subjects should be asked to directly keep the money that they ascribed to themselves, and to put the money for Person B back into the envelope and to through it into an urn.

Part III: Measuring helping behavior

After the dictator game, participants should be given another questionnaire in which they had to indicate in how far they were willing to help others. (It might be useful to place a filler task between the dictator game and the helping items to prevent an intensification of demand characteristics and biased responses.) Similar to the study of Griskevicius et. al. (2007) items should reflect non-heroic as well as heroic helping, (e.g., *non-heroic helping*: helping elderly persons to do their shopping; *heroic helping*: running into a burning building in which someone is trapped).

Hypotheses and further considerations

According to the theory of strategic pluralism, prosociality as a cue to good partnership/parental abilities should be displayed primarily in the long-term mating condition. Assuming that both dictator game contributions and non-heroic helping behavior indicate good partnership/parenting abilities, these two displays should be higher in the long-term condition as opposed to the control condition and the short-term condition, although, as discussed, it could be that effects for *anonymous* dictator game contributions are weaker than effects for items designed to measure *public* helping behavior.

A second objection to the hypothesis is that Griskevicius et al. (2007) showed that mating goals only increase *heroic* helping in men. If this result was replicated, it would be interesting to observe whether long-term mating motives *and* short-term mating motives increased heroic helping, as indeed heroic helping might on the one hand be an indicator of good partnership/parenting abilities, but on the other hand it might likewise be an indicator of genetic fitness. Of course, it would also be interesting to investigate the same relation for women.

11.2.4. Condition-dependant costs: Are “good-gene-altruists” nobler than “good-parent-altruists”?

In the preceding section, one problem loomed: in some cases it is hard to differentiate whether a specific prosocial trait is either an indicator of good partnership/parenting abilities, or an indicator of genetic quality, or even both (e.g., heroic helping behavior).

Actually, Miller (2007) exactly assumed these *three* types of signaling functions of moral virtues. As discussed in the theoretical part of this work, with regard to genetic quality Miller (2007) argued that moral virtues, like altruism, are difficult to display if one has a high mutation load that impairs the precision of body and brain development. For example, people burdened with mutations associated with psychological disorders like autism tend to develop limited Theories of Mind. They consequently show deficits with regard to empathy and prosocial behaviors. Turning the argument the other way around, prosociality may therefore serve as a kind of neurogenetic warranty when judging a potential mate. With regard to partnership/parenting abilities, the relevance of the argument was even more obvious because altruism is an attractive feature in its own right and individuals who display altruistic acts now are likely to care for their partner's and children's well-being at a later time.

However, to differentiate the signaling functions of prosociality (i.e., whether a prosocial behavior is a good genes or a good partner/parent indicator), Miller (2007) made specific predictions. These predictions were based on varying costs of prosociality as a function of an individual's genetic fitness (i.e. his/her "condition"). Miller (2007) stated that individuals with higher genetic quality should be easier able to suffer the costs of prosociality and should consequently display moral virtues more steadily. Thus, stable prosociality would be an indicator of good genes, and "good gene prosociality" should correlate positively with other fitness indicators, such as health, fertility, intelligence, or physical attractiveness. On the contrary, Miller (2007) assumed moral virtues as good partnership/parenting indicators to obtain their temporal reliability (from the beginning of a relationship to a long-term relationship) not so much from condition-dependence, but, amongst others, from the social-reputational costs connected to moral backsliding.

If Miller's assumptions are true, individuals should vary their prosocial behavior depending on (1) their level of genetic quality, and (2) the reputational costs connected to a misdeed. To test this, a donation study could be conducted.

Assuming that donating behavior is a prosocial virtue that may signal both good genes as well as good partnership/parenting qualities, specific falsifiable predictions can be made on the basis of Miller's (2007) considerations, such that individuals with good genes show donation behavior often and steadily, while individuals with good partnership/parenting abilities show donation behavior more selectively in contexts with reputational consequences in relation to a prospective mate, because for them the display of donation behavior is more costly in relation to their overall fitness.

Participants should be invited to the laboratory. To trigger mating goals, they should be welcomed by an attractive experimenter of the opposite sex. First of all, the genetic fitness of participants should be measured via two proxi-variables: physical attractiveness and intelligence. To measure physical attractiveness, participants should be asked to present themselves into the video camera so that based on these video-clips, a group of judges could rate the physical attractiveness of participants later. Afterwards, participants should be left alone to fill out an intelligence test. (In addition, it might be considerable to hand out a scale to measure sociosexuality, as indeed high genetic quality should also be related to unrestricted sociosexuality.)

Subsequently, donation behavior should be measured. In the *public* condition with high reputational consequences, the attractive experimenter of the opposite sex should personally hand out the financial reward connected to this study (e.g., 10 Euro in coins). Presenting an official donation box (e.g., a box of doctors without borders) to the participants, the experimenter should ask them whether they are willing to spend a portion of their money for a charitable purpose.

By contrast, in the *anonymous* condition without reputational consequences, participants should receive their financial reward in an envelope together with the questionnaire in which intelligence is tested. The experimenter should leave the room. An info leaflet inside the envelope should call attention to the donation box, which should stand on a table in the lab room. Having accomplished the intelligence test, subjects could make their donation choice unobserved and free of reputational consequences.

In the public condition with high reputational consequences, social demand characteristics would be high so that individuals with good genes (i.e., highly attractive, intelligent people (unrestricted in sociosexuality)), as well as individuals with good partnership/parenting abilities should show high levels of donation. However, in the anonymous condition, donation behavior of individuals with good genes should be higher than donation behavior of individuals with good partnership/parenting abilities.

The conduction of this study seems useful as it offers a possibility to test whether the assumption that prosocial behavior primarily indicates good partnership/parenting abilities is really true. This supposition, indeed, is incorporated in many studies on evolutionary theories of mate choice (like the one that was presented in Chapter 10). However, before proceeding to conduct studies that are based on this assumption, the assumption should at first be challenged empirically.

12. Closing words

The theories and the research presented in this work indicate that the study of the evolution of altruism has been extensive but that it is still expanding. The research proposals that were outlined in the last section may have been slightly suggestive of how much delicate work still has to be accomplished to compose the human altruism puzzle with all its tiny pieces. However, any science that deals with a complex phenomenon like human cooperation should not only focus on the investigation of tiny elements—it should instead always keep the big picture in view by asking the “big questions.”

One of these big questions certainly is in how far human behavior is really determined by genes and how much of behavioral variation is due to culture. For a long time, biology and the social sciences have been opposing camps in that they tried to explain the same behavioral tendencies with divergent means. It seems that slowly but surely a more integrative approach gains ground on the basis of which biology and culture may best be understood as forces that are mutually restrictive, but likewise constructive. In this regard, the combination of cross-cultural research and evolutionary thinking has to be pressed ahead as it may bring about the most sophisticated conclusions as to the relative importance of genes and culture for the shaping of the human race.

Altogether, it is invigorating to see how evolutionary theory allows for the collaboration of scholars from divergent scientific fields and how their disputes lead to ever more interesting questions that open new avenues for investigation. The Nobel laureate William Lawrence Bragg once said “The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.” In this spirit, the study of the evolution of human altruism promises to create still a few viewpoints that will force us to rethink our beliefs about the world and our place in it.

13. References

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

Alexander, R. D. (1987). *The biology of moral systems.* Aldine, NY: Hawthorne.

Allen, B. P., & Potkay, C. R. (1981). On the arbitrary distinction between states and traits. *Journal of Personality and Social Psychology, 41*, 916 - 928.

Ambady, N., & Rosenthal, R. (1993). Half a minute: predicting teacher evaluations from thin slices of nonverbal behaviour and physical attractiveness. *Journal of Personality and Social Psychology, 64*, 431 - 441.

Ambady, N., Bernieri, F.J., & Richeson, J.A. (2000). Toward a histology of social behaviour: judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology, 32*, 201 - 271.

Anderson, C., & Shirako, A. (2008). Are individual's reputations related to their history of behavior? *Journal of Personality and Social Psychology, 94*, 320 – 333.

Andreoni J., & Petrie, R. (2004). Public goods experiment without confidentiality: A glimpse into fund-rainsing. *Journal of Public Economics, 88*, 1605 - 1623.

Andreoni, J. (1988). Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics, 37*, 291 - 304.

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

Aristotle (n.d./1962). *Nicomachean ethics.* Indianapolis, IN: Bobbs-Merrill.

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

Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science, 242*, 1385 - 1390.

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

Baldwin, J. M. (1895). *Mental development in the child and the race: Method and Processes.* New York: MacMillan.

Barrett, L., Dunbar, R., & Lycett, J. (2002). *Human evolutionary psychology.* Princeton, NJ: Princeton University Press.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497 - 529.

Benenson, J. F., Pascoe, J., & Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and Human Behavior, 28*, 168 - 175.

Benshoof, L., & Thornhill, R. (1979). The evolution of monogamy and loss of estrus in humans. *Journal of Social and Bioloogical Structures, 2*, 95 - 106.

Berscheid, E. (1985). Interpersonal attraction. In G. Lindzey, & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., pp. 110 - 168). New York, NY: Random House.

Bettinger, R. L. (1991). *Hunter-gatherers: archaeological and evolutionary theory.* New York, NY: Plenum.

Boorman, S. A., & Levitt, P. A. (1973). A frequency-dependent natural selection model for the evolution of social cooperation networks. *Proceedings of the Natural Academy of Sciences of the USA, 70*, 187 – 189.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behaviour as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*, 599 - 614.

Bowles, S. (2001). Individual interactions, group conflicts, and the evolution of preferences. In S. N. Durlauf & H. P. Young (Eds.), *Social dynamics* (pp. 155 - 190). Cambridge, MA: MIT Press.

Bowles, S., Choi, J., & Hopfensitz, A. (2003). The co-evolution of individual behaviours and social institutions. *Journal of Theoretcial Biology, 223*, 135 - 147.

Bowles, S., & Gintis, H. (1986). *Democracy and capitalism: property, community, and the contradictions of modern social thought.* New York, NY: Basic Books.

Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical Population Biology, 65*, 17 - 28.

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

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

Boyd, R., & Richerson, P. J. (1987). The evolution of ethnic markers. *Cultural Anthropology, 2*, 65 - 79.

Boyd, R., & Richerson, P. J. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology, 132*, 337 - 356.

Boyd, R., & Richerson, P. J. (1996). Why culture is common but cultural evolution is rare. *Proceedings of the British Academy, 88,* 77 - 93.

Boyd, R., & Richerson, P. J. (2004). *The nature of cultures.* Chicago, IL: University of Chicago Press.

Brase, G. L. (2006). Cues of parental investment as a factor in attractiveness. *Evolution and Human Behavior, 27*, 145 - 157.

Brauchli, K., Killingback, T., & Doebeli, M. (1999). Evolution of cooperation in spatially structured populations. *Journal of Theoretical Biology, 200*, 405 - 417.

Brosig, J. (2002). Identifying cooperative behaviour: some experimental results in a prisoner's dilemma game. *Journal of Economic Behaviour & Organization, 47*, 275 - 290.

Brown, J. L., & Brown, E. R. (1981). Kin selection and individual selection in babblers. In R. D. Alexander, & D. W. Tinkle (Eds.), *Natural Selection and Social Behavior: Recent Research and New Theory* (pp. 244 - 256). New York, NY: Chiron Press.

Brown, J. L., & Eklund, A. (1994). Kin recognition and the major histocompatability complex: an integrative review. *American Naturalist, 143*, 435 - 461.

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

Burnham, T. C., & Hare, B. (2007). Engineering human cooperation. Does involuntary neural activation increase public goods contributions? *Human Nature, 18*, 88 - 108.

Burnham, T. C., & Johnson, D. D. P. (2005). The biological and evolutionary logic of human cooperation. *Analyse & Kritik, 27*, 113 - 135.

Burnstein E., Crandall C., & Kitayama S. (1994). Some neo-Darwinian decision rules for altruism: weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology, 67*, 773 - 789.

Buss, D. M. (1989). Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12*, 1 - 49.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: a contextual evolutionary analysis of human mating. *Psychological Review, 100*, 204 - 232.

Buss, D. M., Shackelford, T. K., Kirkpatrick, L. A., & Larsen, R. J. (2001). A half century of mate preferences: the cultural evolution of values. *Journal of Marriage and Families, 63*, 492 - 503.

Byrne, R. W., & Whiten, A. (1988). *Machiavellian intelligence: social expertise and the evolution of intellect in monkeys, apes and humans.* Oxford, UK: University Press.

Byrne, R. W., & Whiten, A. (1997). Machiavellian intelligence. In A. Whiten, & R.W. Byrne (Eds.), *Machiavellian intelligence II: extensions and evaluations* (pp. 1 - 23). Cambridge, UK: University Press.

Camerer, C. (2003). *Behavioral game theory: experiments on strategic interaction.* Princeton, NJ: Princeton University Press.

Campbell, D. (1969). Variation and selective retention in socio-cultural evolution. *General Systems, 16*, 69 - 85.

Cavalli-Sfornza, L., & Feldman, M. (1981). *Cultural transmission and evolution: a quantitative approach.* Princeton, NJ: Princeton University Press.

Colwell, R. K. (1981). Group selection is implicated in the evolution of female-biased sex ratios. *Nature, 290*, 401 - 404.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptation for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163 - 228). New York, NY: Oxford University Press.

Costa P. T., Jr., & McCrae R. R. (1998). *The NEO-PI/NEO-FFI manual supplement.* Odessa, FL: Psychological Assessment Resources.

Dahrendorf, R. (1958). Homo sociologicus: Ein Versuch zur Geschichte, Bedeutung und Kritik der sozialen Rolle, 16. Auflage. Wiesbaden, Germany: VS-Verlag.

Darwin, C. R. (1871). *The descent of man, and selection in relation to sex.* London, UK: John Murray.

Darwin, C. R. (1872). *The expression of the emotions in man and animals.* London, UK: John Murray.

Dawes, R., & Mulford, M. (1966). The false consenus effect and overconfidence: flaws in judgment or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes, 65*, 201 - 211.

Dawes, R. M., Orbell, J. M., & Van de Kragt, J. C. (1986). Organizing groups for collective action. *American Political Science Review, 80*, 1171 - 1185.

Dawkins, R. (1976). *The selfish gene.* Oxford, UK: Oxford University Press.

de Menocal, P. B. (1995). Plio-pleistocene african climate. *Science, 270*, 53 - 59.

de Waal, F. B. M. (1997). *Good natured: the origins of right and wrong in humans and other animals.* Cambridge, MA: Harvard University Press.

de Waal, F. B. M. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology, 59*, 279 - 300.

Diamond, J. (2005). *Guns, germs and steel: The fates of human societies.* New York: Norton & Company.

Dickemann, M. (1979). Female infanticide, reproductive strategies, and social stratification: a preliminary model. In N. A. Chagnon, & W. Irons (Eds.), *Evolutionary biology and human social behavior – an anthropological perspective* (pp. 321 – 367). North Scituate, MA: Duxbury.

Dunning, D. (2005). *Self-insight: roadblocks and detours on the path to knowing thyself.* New York, NY: Psychology Press.

Dunning, D., Griffin, D., Milojkovic, J., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology, 58*, 568 - 81.

Eckel, C., & Grossman, P. (1998). Are women less selfish than men? Evidence from dictator games. *The Economic Journal, 108*, 726 - 735.

Ekman, P. (1985). *Telling lies.* New York: W. W. Norton.

Emlen, S. T. (1991). Evolution of cooperative breeding in birds and mammals. In J. F. Krebs, & N. B. Davies (Eds.), *Behavioral ecology – an evolutionary approach,* 3rd Edition (pp. 301 - 337). Oxford, UK: Blackwell.

Emler, N. (1990). A social psychology of reputation. *European Review of Social Psychology, 1*, 171–193.

Engelmann, D., & Fischbacher, U. (2003). Indirect reciprocity and strategic reputation building in an experimental helping game. *Working Paper Series Charles University Center for Economic Research and Graduate Education Academy of Sciences of the Czech Republic Economics Institute, 215.* Prague, Czech Republic.

Eshel, I., & Cavalli-Sforza, L. L. (1982). Assortment of encounters and evolution of cooperativeness. *Proceedings of the National Academy of Sciences of the USA, 79*, 1331 - 1335.

Euler, H. A., & Weitzel, B. (1996). Discriminative grandparental solicitude as reproductive strategy. *Human Nature, 7*, 39 - 59.

Farrelly, D., Lazarus, J., Roberts, G. (2007). Altruists attract. *Evolutionary Psychology, 5*, 313 - 329.

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

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

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*, 63 - 87.

Fehr, E., & Fischbacher, U. (2005). Altruists with green beards. *Analyse & Kritik, 27*, 73 - 84.

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

Fehr, E., & Gächter, S. (2003). Egalitarian motive and altruistic punishment. *Nature, 425*, 392 - 393.

Fehr, E., & Henrich, J. (2003). Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In: Hammerstein, P. (Ed.), *Genetic and cultural evolution of cooperation* (pp. 55 - 82). New York, NY: MIT Press.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114*, 817 - 868.

Fetchenhauer, D., & Dunning, D. (2006). Perceptions of prosociality and solidarity in self and others. In D. Fetchenhauer, A. Flache, A. P. Buunk, & S. Lindenberg (Eds.), *Solidarity and prosocial behavior* (pp. 225 - 242). New York, NY: Kluwer Academic/Plenum Publishers.

Fetchenhauer, D., & Dunning, D. (in press). Do people trust too much or too little? *Journal of Economic Psychology.*

Fetchenhauer, D., & Huang, X. (2004). Justice sensitivity and distributive decisions in experimental games. *Personality and Individual Differences, 36*, 1015 - 1029.

Frank, R. H. (1988). Passions within reasons: The strategic role of the emotions. New York, NY: W.W. Norton & Company.

Frank, R. H. (2004). *What price the moral high ground?* Princeton, NJ: Princeton University Press.

Frank, R. H. (2005). Altruists with green beards: Still kicking? *Analyse & Kritik, 27*, 85 - 96.

Frank, R. H. (2008). On the evolution of moral sentiments. In C. Crawford & D. Krebs (Eds.), *Foundations of evolutionary psychology* (pp.371-379). New York, NY: Lawrence Erlbaum Associates.

Frank, R. H., Gilovich, T., & Regan D. T. (1993). The evolution of one-shot cooperation: an experiment. *Ethology and Sociobiology, 14*, 247 - 256.

Frank, S. A. (1986). Hierarchical selection theory and sex ratios. I. General solutions for structured populations. *Theoretical Population Biology, 29*, 312 - 342.

Frank, S. A. (1995). George Price’s contributions to evolutionary genetics. *Journal of Theoretical Biology, 175*, 373 - 388.

Frank, S.A. (1998). *Foundations of social evolution.* Princeton, NJ: Princeton University Press.

Frank. S. A. (2006). Social selection. In C.W. Fox, & J.B. Wolf (Eds.), *Evolutionary genetics: Concepts and case studies* (pp. 350 – 363). New York, NY: Oxford University Press.

Fridhandler, B. M. (1986). Conceptual note on state, trait, and the state-trait distinction. *Journal of Personality and Social Psychology, 50*, 169 - 174.

Gangestad, S. W., Garver-Apgar, C. E., Simpson, J. A., & Cousins, A. J. (2007). Changes in women's mate preferences across the ovulatory cycle. *Journal of Personality and Social Psychology, 92*, 151 - 163.

Gangestad, S. W., Simpson, J. A., DiGeronimo, K., & Biek, M. (1992). Differential accuracy in person perception across traits: examination of a functional hypothesis. *Journal of Personality and Social Psychology, 62*, 688 - 698.

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

Gardner, A., West, S. A., & Barton, N. H. (2007). The relation between multilocus population genetics and social evolution theory. *The American Naturalist*, 169, 207 - 226.

Gaulin, S., & McBurney, D. (2004). *Evolutionary Psychology*, 2nd Ed. Upper Saddle River, NJ: Pearson.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology, 206*, 169 - 179.

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

Gintis, H., Bowles, S., Boyd, R. T., & Fehr, E. (2005). *Moral sentiments and material interests: the foundations of cooperation in economic life (Economic learning and social evolution).* Cambridge, MA: MIT Press.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2008). Gene-culture coevolution and the emergence of altruistic behavior in humans. In C. Crawford, & D. Krebs (Eds.), *Foundations of evolutionary psychology* (pp. 313 – 329). New York, NY: Lawrence Erlbaum Associates.

Gintis, H., Henrich, J. Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of human morality. *Social Justice Research, 21*, 241 - 253.

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

Griffin, A. S., & West, S. A. (2002). Kin selection: fact and fiction. *Trends in Ecology and Evolution*, 17, 15 - 21.

Griffin, A. S., & West, S. A, (2003). Kin discrimination and the benefit of helping cooperatively breeding vertebrates. *Science, 302*, 634 – 636.

Griffin, A. S., West, S. A., & Buckling, A. (2004). Cooperation and competition in pathogenic bacteria. *Nature, 430*, 1024 - 1027.

Grafen, A. (1984). Natural selection, kin selection, and group selection. In J.R. Krebs & N.B. Davies (Eds.), *Behavioral ecology: an evolutionary approach* (pp. 62 - 84). Oxford, UK: Blackwell Publications.

Grafen, A. (1985). A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology, 2*, 28 - 89.

Granovetter, M. S. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology, 91*, 481 - 510.

Griskevicius, V., Tybur, J. M., Sundie, J. M., Cialdini, R. B., Miller, G. F., Kenrick, D. T. (2007). Blatant benevolence and conspicuous consumption: when romantic motives elicit strategic costly signals. *Journal of Personality and Social Psychology, 93*, 85 – 102.

Grossman, S. J., & Hart, O. (1983). An analysis of the principal agent problem. *Econometrica, 51*, 7 - 46.

Gurven, M., Allen-Arave, W., Hill, K., & Hurtado, M. (2000). It's a wonderful life: signaling generosity among the Ache of Paraguay. *Evolution and Human Behavior, 21*, 263–282.

Hagen, E. H., & Hammerstein, P. (2006). Game theory and human evolution: a critique of some recent interpretations of experimental games. *Theoretical Population Biology, 69*, 339 - 348.

Hager, B. J. (1992). Get thee to a nunnery: female religious claustration in medieval Europe. *Ethology and Sociobiology, 13*, 385 - 407.

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

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

Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature, 228,* 1218 - 1220.

Hamilton, W. D. (1971). Selection of selfish and altruistic behaviour in some extreme models. In J. F. Eisenberg, & W. S. Dillon (Eds.), *Man and beast: comparative social behavior* (pp. 57 - 91). Washington, DC: Smithsonian Press.

Hamilton, W.D. (1975). Innate social aptitudes of man: an approach of evolutionary genetics. In R. Fox (Ed.), *Biosocial anthropology* (pp. 133 - 155). New York, NY: Wiley.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450,* 557 - 559.

Harbaugh, W. T., Krause, K. S., & Liday, S. J. (2003). Bargaining by children. *University of Oregon Economics Working Paper* No. 2002-4. Available at SSRN: DOI: 10.2139/ssrn.436504.

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

Hardy, C. L., & Van Vugt, M. (2006). Nice guys finish first: the competitive altruism hypothesis. *Personality and Social Psychology Bulletin, 32*, 1402 - 1413.

Haselton, M. G., & Funder, D. C. (2006). The evolution of accuracy and bias in social judgment. In M. Schaller, J.A. Simpson, & D.T. Kenrick (Eds.), *Evolution and Social Psychology* (pp. 16 - 37). New York, NY: Psychology Press.

Haselton, M. G., Gangestad, S. W. (2006). Conditional expression of women's desires and men's mate guarding across the ovulatory cycle. *Hormones and Behavior, 49*, 509 - 518.

Haselton, M. G., Mortezaie, M., & Pillsworth, E. G. (2007). Ovulatory shifts in human ornamentation: near ovulation, women dress to impress. *Hormones and Behavior, 51*, 40 - 45.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2004). Overview and Synthesis. In J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, & H. Gintis (Eds.), *Foundations of human sociality. Economic experiments and ethnographic evidence from fifteen small-scale societies.* Oxford, UK: University Press.

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: the pros and cons of projection. *Journal of Personality and Social Psychology, 53*, 221 - 234.

Hume, D. (1739/2006). *A Treatise of Human Nature. Being an Attempt to introduce the experimental Method of Reasoning into Moral Subjects.* Retrieved November, 17, 2008, from http://etext.library.adelaide.edu.au/h/hume/david/h92t/.

Johnson, D. D. P., Price, E., & Takezawa, M. (2008). Renaissance of the individual. Reciprocity, positive assortment, and the puzzle of human cooperation. In C. Crawford, & D. Krebs (Eds.), *Foundations of evolutionary psychology* (pp. 331 – 352). New York: Lawrence Erlbaum Associates.

Keeley, L. H. (1996). *War before civilization: the myth of the peaceful savage.* New York, NY: Oxford University Press.

Keller, L., & Ross, K. G. (1998). Selfish genes: a green beard in the red fire ant. *Nature, 394,* 573-575.

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

Kenny, D. A., & Acitelli, L. K. (2001). Accuracy and bias in the perception of the partner in a close relationship. *Journal of Personality and Social Psychology, 80*, 439 - 448.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis.* New York: The Guilford Press.

Killingback, T., & Doebeli, M. (1996). Spatial evolutionary game theory: hawks and doves revisited. *Proceedings: Biological Sciences, 263*, 1135 - 1144.

Komdeur, J., & Hatchwell, B. J. (1999). Kin recognition: function and mechanism in avian societies. *Trends in Ecology and Evolution, 14*, 237 - 241.

Koziel S, & Pawłowski B. (2003). Comparison between primary and secondary mate markets: an analysis of data from lonely hearts columns. *Personality and Individual Differences, 35*, 1849 - 1857.

Krebs, J. R., & Davies, N. B. (1993). *An introduction to behavioral ecology,* 3rd Edition. Oxford: Blackwell.

Kroska, A. (2003). Investigating gender differences in the meaning of household chores and child care. *Journal of Marriage and Family, 65*, 456 - 473.

Kruger, D. J., Fisher, M., & Jobling, I. (2003). Proper and dark heroes as dads and cads. Alternative mating strategies in british romantic literature. *Human Nature, 14*, 305 - 317.

Kunda, Z. (1999). *Social cognition. Making sense of people.* Cambridge MA: MIT University Press.

Lack, D. (1966). *Population studies of birds*. Oxford: Clarendon Press.

Leimar, O., & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society London, Series B, 268*, 745 - 753.

Li, N., & Kenrick, D. (2006). Sex similarities and differences in preferences for short-term mates: what, whether, and why. *Journal of Personality and Social Psychology, 90*, 468-489.

Lieberman, D., Tooby, J., & Cosmides, L. (2003). Does morality have a biological basis? An empirical test of the factors governing moral sentiments relating to incest. *Proceedings of the Royal Society of London, Series B, 270*, 819 - 826.

Lorenz, K. (1973). *Die Rückseite des Spiegels - Versuch einer Naturgeschichte menschlichen Erkennens*. München, Germany: Piper.

Lotem, A., Fishman, M. A., & Stone, L. (1999). Evolution of cooperation between individuals. *Nature, 400*, 226 - 227.

Lotem, A., Fishman, M. A., & Stone, L. (2002). From reciprocity to unconditional altruism through signaling benefits. *Proceedings of the Royal Society of London, Series B: Biological Sciences, 270*, 199–205.

Low, B. S. (1989). Cross-cultural patterns in the training of children: an evolutionary perspective. *Journal of Comparative Psychology, 103*, 313 - 319.

Lumsden C., & Wilson, E. (1981). *Genes, mind and culture: the coevolutionary process*. Cambridge, MA: Harvard University Press.

Machiavelli, N. (1514/2008). *The Prince*. Retrieved December, 4, 2008 from http://www.uvawise.edu/history/wciv1/prince.html

Maenchen-Helfen, O. J. (1973). *The world of the Huns.* Berkeley, CA: University of California Press,.

Mascie-Taylor, C. G. N. (1995). Human assortative mating: evidence and genetic implication. In A. J. Boyce, & V. Reynolds (Eds.), *Human populations. Diversity and adaptations* (pp. 86-105). Oxford, UK: Oxford University Press.

Mauss, M. (1924). *The gift: forms and functions of exchange in archaic societies*. London: Cohen and West.

Maynard Smith, J. (1964). Group selection and kin selection. *Nature, 201*, 1145 - 1147.

Maynard Smith, J. (1976). Group selection. *Quarterly Review of Biology, 51*, 277 - 283.

McAndrew, F. T. (2002). New evolutionary perspectives on altruism: multilevel-selection and costly-signaling theories. *Current Directions in Psychological Science, 11*, 79 - 82.

McKenzie, C. R. M. (2003). Rational models as theories – not standards – of behavior. *Trends in Cognitive Sciences, 7*, 403 - 406.

Milinski, M., Semmann, D., & Krambeck, H.-J. (2001). Reputation helps solve the 'tragedy of the commons'. *Nature, 415*, 424 - 426.

Miller, G. F. (2000). *The mating mind.* London: Vintage.

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

Moore, G. E. (1903/2008). *Principia Ethica.* Retrieved November, 17, 2008 from http://fair-use.org/g-e-moore/principia-ethica.

Nesse, R. M. (2001). *Evolution and the capacity for commitment.* New York: Russell Sage Foundation.

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

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

Ockenfels, A., & Selten, R. (2000). An experiment on the hypothesis of involuntary truth signaling in bargaining. *Games and Economic Behavior, 33*, 90 - 116.

Oda, R. (2001). Sexually dimorphic mate preference in Japan: an analysis of lonely hearts advertisements. *Human Nature, 12*, 191 - 206.

Okasha, S. (2005). Maynard Smith on the levels of selection question. *Biology and Philosophy, 20*, 989 - 1010.

Oltmanns, T. F., Friedman, J., Fiedler, E. R., & Turkheimer, E. (2004). Perceptions of people with personality disorders based on thin slices of behaviour. *Journal of Research in Personality, 38*, 216 - 229.

Page, T., Putterman, L., & Unel, B. (2005). Voluntary association in public good experiments: Reciprocity, mimicry, and efficiency. *The Economic Journal, 115*, 1032 - 1053.

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

Peck, J. R. (1992). Group selection, individual selection, and the evolution of genetic drift. *Journal of Theoretical Biology, 159*, 163 - 187.

Podolefsky, A. (1984). Contemporary warfare in the New Guinea highlands. *Ethnology, 23*, 73 - 87.

Potts, R. (1996). *Humanity's descent: the consequences of ecological instability.* New York, NY: William Morrow.

Pradel, J., Euler, H., & Fetchenhauer, D. (in press). Spotting altruistic dictator game players and mingling with them: The elective assortation of classmates. *Evolution and Human Behavior.*

Price, M. E. (2008). The resurrection of group selection as a theory of human cooperation. *Social Justice Reserach, 21*, 228 - 240.

Reeve, H. K., & Keller, L. (1999). Levels of selection: burying the units-of-selection debate and unearthing the crucial new issues. In L. Keller (Ed.), *Levels of selection in evolution* (pp. 3 – 14). Princeton, NJ: Princeton University Press.

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

Richerson, P. J., & Boyd, R. (1998). The evolution of human ultrasociality. In I. Eibl-Eibisfeldt, & F. Salter (Eds.), *Ideology, warfare, and indoctrinability*. Oxford, UK: Berghahn Books.

Richerson, P. J., & Boyd, R. (2000). Built for speed: pleistocene climate variation and the origin of human culture. In F. Tonneau, & N. S. Thompson (Eds.), *Perspectives in ethology, 13: evolution, culture, and behavior* (pp. 1 – 45).

Richerson, P. J., & Boyd, R. (2001). Built for speed, not for comfort: Darwinian theory and human culture. *History and Philosophy of the Life Sciences, 23,* 423-463.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone*. Chicago, IL: University of Chicago Press.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": an egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279 - 301.

Russell, A. F., & Hatchwell, B. J. (2001). Experimental evidence for kin-biased helping in a cooperatively breeding vertebrate. *Proceedings of the Royal Society of London, Series B, 268*, 2169 - 2174.

Sally, D. (2000). A general theory of sympathy, mind-reading, and social interaction, with an application to the prisoner's dilemma. *Social Science Information, 39*, 567 - 634.

Sato, K. (1987). Distribution and the cost of maintaining common property resources. *Journal of Experimental Social Psychology, 23,* 19 - 31.

Schelling, T. (1960). *The strategy of conflict.* Cambridge, MA: Harvard University Press.

Scheyd, G. J., Garver-Apgar, C. E., Gangestad, S. W. (2008). In C. Crawford & D. Krebs (Eds.), *Foundations of evolutionary psychology* (pp. 239 - 259). New York, NY: Erlbaum.

Schroeder, D. A., Penner, L. A., Dovidio, J. F., Piliavin, J. A. (1995). *The psychology of helping and altruism.* New York, NY: McGraw-Hill.

Schul, Y., & Vinokur, A. (2000). Projection in person perception among spouses as a function of the similarity in their shared experiences. *Personality and Social Psychology Bulletin, 26*, 987 - 1001.

Sharp, S. P., McGowan, A., Wood, M. J., & Hatchwell, B. J. (2005). Learned kin recognition cues in a social bird. *Nature, 434*, 1127 - 1130.

Sheldon, K. M., Sheldon, M. S., & Osbaldiston, R. (2000). Prosocial values and group-assortation within an N-person prisoner's dilemma. *Human Nature, 11*, 387 - 404.

Sherif, M., & Murphy, G. (1936). *The psychology of social norms*. New York, NY: Harper & Brothers.

Simpson, J. A., Wilson, C. L., & Winterheld, H. A. (2004). Sociosexuality and romantic relationships. In J. H. Harvey, A. Wenzel, & S. Sprecher (Eds.), *Handbook of sexuality in close relationships* (pp. 87 - 112). Mahwah, NJ: Erlbaum.

Smith, E. A., & Bird, R. L. (2000). Turtle hunting and tombstone opening: public generosity as costly signaling. *Evolution and Human Behavior, 21,* 245–261.

Sober, E. & Wilson, D. S. (1998). *Unto others: the evolution and psychology of unselfish behavior.* Boston, MA: Harvard University Press.

Soltis, J., R. Boyd, R., & Richerson, P. J. (1995). Can group-functional behaviors evolve by cultural group selection? An empirical test. *Current Anthropology, 36*, 437-94.

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

Spielberger, C. D. (1972). Anxiety as an emotional state. In C.D. Spielberger (Ed.), *Anxiety and behavior* (pp. 3 - 20). New York, NY: Academic Press.

Spielberger, C. D., Auerbach, S. M., Wadsworth, A. P., Dunn, T. M., & Taulbee, E. S. (1973). Emotional reactions to surgery. *Journal of Consulting and Clinical Psychology, 40*, 33 - 38.

Sport-, Schul- & Jugendamt Springe (n.d.). *Infoblatt: Taschengeld für Kinder und Jugendliche* [Info: Pocket money for children and adolescents]. Retrieved August, 18, 2008, from http://www.springe.de/Download/40-Info%20Taschengeld-neu.pdf.

Spuhler, J. N. (1968). Assortative mating with respect to physical characteristics. *Eugenics Quarterly, 15*, 128-140.

Stearns, S. C. (1989). Trade-offs in life-history evolution. *Functional Ecology, 3*, 259 - 268.

Stringer, C., & Gamble, C. (1993). *In search of the Neanderthals: solving the puzzle of human origins*. New York, NY: Thames and Hudson.

Sugiyama, L. S. (2005). Physical attractiveness in adaptationist perspective. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 292 - 343). Hoboken, NJ: Wiley.

Takezawa, M., Gummerum, M., & Keller, M. (2006). A stage of the rational tail of the emotional dog: roles of moral reasoning in group decision making. *Journal of Economic Psychology, 27*, 117 – 139.

Thornhill, R., & Gangestad, S. W. (1999). The scent of symmetry: a human pheromone that signals fitness? *Evolution and Human Behavior, 20*, 175 - 201.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie, 20*, 1410 - 433.

Tinsley, C. H., O'Connor, K. M., & Sullivan, B. A. (2002). Tough guys finish last: the perils of a distributive reputation. *Organizational Behavior and Human Decision Processes, 88*, 621–645.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York, NY: Oxford University Press.

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

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of men: 1871 – 1971* (pp. 136 - 179). Chicago, IL: Aldine.

Trivers, R. L. (1974). Parent-offspring conflict. *American Zoologist, 14*, 249 - 264.

Trivers, R. L. (1985). *Social evolution.* Menlo Park, CA: Benjamin Cummings.

Trivers, R. L. (1991). Deceit and self-deception: the relationship between communication and consciousness. In M. Robinson, & L. Tiger (Eds.), *Man and beast revisited* (pp. 175 - 191). Washington, DC: Smithsonian.

Trivers, R. L. (2004). Mutual benefits at all levels of life. *Science, 304*, 964 - 965.

Veblen, T. (1899/1994). *The theory of the leisure class.* New York: Dover.

Vernier, B. (1984). Putting kin and kinship to good use: the circulation of goods, labor, and names on Karpathos (Greece). In H. Medick, & Sabean, D.W. (Eds.), *Interest and emotion-essays on the study of family and kinship.* Cambridge, UK: Cambridge University Press.

Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: the sequel. A kernel of truth in predictive cheating detection. *Evolution and Human Behavior, 28* , 260 - 271.

Voland, E., & Voland, R. (1995). Parent-offspring conflict, the extended phenotype, and the evolution of conscience, *Journal of Social and Evolutionary Systems, 18*, 397 – 412.

Warner, R. M., Kenny, D. A., & Stoto, M. (1979). A new round robin analysis of variance for social interaction data. *Journal of Personality and Social Psychology, 37*, 1742 - 1757.

Webster G. (2003). Prosocial behavior in families: moderators of resource sharing. *Journal of Experimental Social Psychology, 39*, 653 - 660.

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

Wenseleers, T., Helantera, H., Hart, A., & Ratnieks, F. L. W. (2004). Worker reproduction and policing in insect societies: an ESS analysis. *Journal of Evolutionary Biology, 17*, 1035 - 1047.

Wertheimer, M. (1925). Über Gestalttheorie. *Philosophische Zeitschrift für Forschung und Aussprache, 1*, 39-60.

West, S. A., Gardner, A., & Griffin, A. S. (2008). Kinship and social behavior. In C. Crawford, & D. Krebs (Eds.), *Foundations of evolutionary psychology* (pp. 91 – 114). New York, NY: Lawrence Erlbaum Associates.

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

West, S. A., Pen, I., & Griffin, A. S. (2002). Cooperation and competition between relatives. *Science, 296*, 72 - 75.

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

Williams, G. C. (1972). *Adaptation and natural selection: a critique of some current evolutionary thought.* Princeton, NJ: Princeton University Press.

Wilson, D. S. (1977). Structured demes and the evolution of group advantageous traits. *American Naturalist, 111*, 157 - 185.

Wilson, D. S., & Colwell, R. K. (1981). The evolution of sex ratio in structured demes. *Evolution, 35*, 882 - 897.

Wilson, D. S., & Dugatkin, L. A. (1997). Group selection and assortative interactions. *American Naturalist, 149*, 336 – 351.

Wilson, D. S. & Sober, E. (1994). Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences, 17*, 585 - 654.

Wilson, D. S., Van Vugt, M., & O'Gorman, R. (in press). Multilevel selection theory and major evolutionary transitions: implications for psychological science. *Current Directions in Psychological Science.*

Wilson, E. O. (1975). *Sociobiology.* The Abridged Edition. Cambridge, MA: Harvard University Press.

Wilson, E. O. (1978). *On human nature.* Cambridge, MA: Harvard University Press.

Wynne-Edwards, V. C. (1962). *Animal dispersion in relation to social behaviour*. Edinburgh, UK: Oliver and Boyd.

Yamagishi, T. (1988a). The provision of a sanctioning system in the United States and Japan. *Social Psychology Quarterly, 51*, 265 - 271.

Yamagishi, T. (1988b). Seriousness of social dilemmas and the provision of a sanctioning system. *Social Psychology Quarterly, 51*, 32–42.

Yamagishi, T. (1992). Group size and the provision of a sanctioning system in a social dilemma. In W. Liebrand, D. M. Messick, & H. Wilke (Eds.), *Social dilemmas: theoretical issues and research findings* (pp. 267–287). Oxford, UK: Pergamon.

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

Zahavi, A. (1995). Altruism as a handicap — the limitations of kin selection and reciprocity. *Journal of avian biology, 26*, 1 - 3.

Zahavi, A., & Zahavi, A. (1997). *The handicap principle: a missing peace of Darwin's puzzle.* New York, NY: Oxford University Press.

14. Appendices

Appendix A: Controlling for non-independence of prediction scores within school classes

Warner et al. (1979) showed that in a so-called *Social Relations Model* in which every individual of a certain group rates every other individual of this group, a prediction X of an actor i concerning partner j can be expressed by the following equation:

$$X_{ij} = m + a_i + b_j + g_{ij}$$

where

- m denotes the mean prediction within the group,
- a_i is the actor effect for person i, which is his tendency to exhibit a consistent level of predictions across interaction-partners
- b_j is the partner effect of person j, which again measures the tendency of group members to judge a person in a consistent way
- g_{ij} is the relationship-effect which is set at the dyad level. It measures the prediction of actor i concerning partner j after removing their individual level tendencies (i.e. the actor- and the partner.

The estimation of the actor-effect equals:

$$a_i = \frac{(n-1)^2}{n(n-2)} M_{i.} + \frac{n-1}{n(n-2)} M_{.i} - \frac{n-1}{n-2} M_{..}$$

where

- n is the group size,
- $M_{i.}$ is the mean of predictions exhibited by person i,

- $M_{.i}$ is the mean of predictions received by person I, and
- $M_{..}$ is the mean of all the predictions.

The estimation of the partner-effect equals:

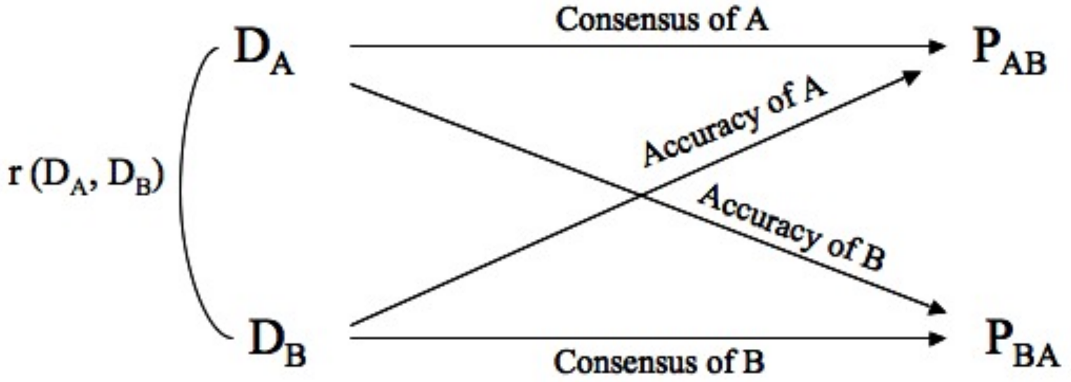
$$b_i = \frac{(n-1)^2}{n(n-2)} M_{.i} + \frac{n-1}{n(n-2)} M_{i.} - \frac{n-1}{n-2} M_{..}$$

As the partner-effects are precisely those values, which signify the tendencies of objects to be predicted as behaving altruistic by all their classmates, we used partner-effects as adjusted prediction scores when testing hypothesis 1 on an aggregated level (e.g. the question whether the average reputation of a student matched his actual behavior). This means that actual dictator contributions of the students were correlated with their partner-effects (n = 122).

When testing hypothesis 1 on the individual level (e.g. whether individual predictions, too, matched the behavior of their classmates), we additionally integrated the unique perception of an actor concerning a partner into our analysis. That is, the sum of partner-effects and relationship-effects belonging to a single prediction ($b_i + g_{ij}$) were correlated with the actual dictator contributions of the student (n = 2437).

Appendix B: Measuring true accuracy and the consensus-effect in mutual predictions of two persons

Kenny and Acitelli (2001) depicted a paradigm for the simultaneous measure of accuracy and consensus effect in 2-person relationships: the Actor-Partner Interdependence Model (APIM). Our replica is presented in the following figure:



It consisted of four variables: (1) the dictator contribution of Person A “D_A”, (2) the dictator contribution of Person B “D_B”, (3) the prediction of person A concerning the dictator contribution of person B “P_{AB}”, (4) the prediction of person B concerning the dictator contribution of person A “P_{BA}”. While the actual dictator contributions of the two persons D_A and D_B were treated as determinants, the predictions of the dictator contributions P_{AB} and P_{BA} (in which the first subscript refers to the judge and the second subscript refers to the object) were treated as outcome variables. The four paths represent accuracy and consensus effects in the predictions of the two persons: diagonal paths from D_A to P_{BA} and D_B to P_{AB} denote accuracy effects, whereas horizontal paths from D_A to P_{AB} and D_B to P_{BA} denote consensus effects as they refer to the extent to which the judge implicitly assumes that the object will make a dictator contribution like he himself did. Whether the consensus effect lowers or betters predictions depends on whether person A and person B are actually similar (correlation between D_A and D_B).

We estimated the paths of the APIM using multilevel modeling with SPSS. The dyad was treated as the unit. As person A and person B were indistinguishable (i.e. both were friends or both nominated each other as likable etc.) accuracy and consensus effects were assumed to be similar for both persons.

Having calculated accuracy and consensus effects, regression coefficients were standardized. Subsequently overall accuracy rates (i.e. the correlation of actual behavior and predictions) could be split into true accuracy and the increment due to bias, with the latter being the product of actual similarity and the consensus effect. The equation reads as follows:

overall accuracy = *true accuracy* + *increment due to bias*

$r\ (D, P) = \text{accuracy effect} + r\ (D_A,\ D_B) * \text{consensus effect}$.