



# Rationality and social behavior

Jan Tullberg

*Stockholm School of Economics, PO Box 6501, SE-113 83 Stockholm, Sweden*

Received 22 May 2002; received in revised form 25 November 2002; accepted 7 May 2003

## Abstract

This article penetrates the relationship between social behavior and rationality. A critical analysis is made of efforts to classify some behaviors as altruistic, as they simultaneously meet criteria of rationality by not truly being self-destructive. Newcomb's paradox is one attempt to create a hybrid behavior that is both irrational and still meets some criterion of rationality. Such dubious rationality is often seen as a source of altruistic behavior.

Group selection is a controversial topic. Sober and Wilson (*Unto Others—The Evolution and Psychology of Unselfish Behavior*, Harvard University Press, Cambridge, MA, 1998) suggest that a very wide concept of group selection might be used to explain altruism. This concept also includes kin selection and reciprocity, which blurs its focus. The latter mechanisms hardly need further arguments to prove their existence. This article suggests that it is group selection in a strict sense that should be investigated to limit semantic neologism and confusion. In evaluation, the effort to muster a mechanism for altruism out of group selection has not been successful. However, this is not the end to group selection, but rather a good reason to investigate more promising possibilities. There is little reason to burden group selection with the instability of altruism caused by altruistic members of a group having lower fitness than egoistic members.

Group selection is much more likely to develop in combination with group egoism. A common project is supported by incitement against free riding, where conformist members joined in solidarity achieve a higher fitness than members pursuing more individualistic options. Group egoism is in no conflict with rationality, and the effects of group selection will be supported rather than threatened by individual selection. Empirical evidence indicates a high level of traits such as conformism and out-group antagonism in line with group egoism. These traits are also likely candidates for behavior favored by group selection since they homogenize the group and link the different individuals closer to one another and a similar fate.

© 2003 Elsevier Ltd. All rights reserved.

*Keywords:* Group egoism; Group selection; Altruism

## 1. Introduction

In many situations, individuals are asked to ignore their own rationality and instead comply with socially demanded behavior. One way of promoting such behavior is to make it rational for the individual to comply either by threats of punishment or material reward. Often, such rationality is simultaneously strived for, but also denied, in the sense that the individual behaving socially is applauded for his unselfishness rather than for his conformism. A second way of promoting social behavior is to support it with some promise of metaphysical reward: sacrifices in this life will be rewarded by God in the next life, if not by fate and luck in this life. This means that some super-natural

rationality supports the strategy of 'getting well off—by doing good'.

How is the link between social behavior that is supposed to be good for others, but implies some cost to the agent himself, to be reconciled with rationality for the agent himself? One starting point of reflection is Newcomb's paradox.

## 2. Rationality and Newcomb's paradox

Newcomb's paradox was introduced by [Robert Nozick \(1969\)](#), and it has since interested people from different disciplines. There have been supporters, but also many critics arguing that there is no real paradox, just a presentation that is confusing to some readers ([Binmore, 1994](#)). The situation is as follows: an agent

*E-mail address:* [jan.tullberg@hhs.se](mailto:jan.tullberg@hhs.se) (J. Tullberg).

has a choice between one box containing \$ $x$ , or this first box with \$ $x$  as well as a second box always containing \$1. The content  $x$  has either the value of \$0 or \$2 and this choice is made in advance by the other decision-maker, often described as a ‘Superior Being’ (SB), not just another ordinary agent. Since the two-box alternative is one dollar better than the one-box alternative for the agent, regardless of the SB’s choice of 0 or 2, the two-box alternative can be seen as dominating the one-box alternative. The two-box alternative should therefore be the choice of a rational agent. However, the paradox raises doubts whether an agent should choose two boxes to maximize his monetary reward. To many people the one-box choice is seen as modest and social, while the two-box alternative is regarded as greedy and selfish. A ‘less is more’ possibility therefore has a significant ideological appeal. Let us look closer into this paradox. The alternatives of the two decision-makers can be described in a matrix (Fig. 1).

Now the situation is complicated by the special influence of this other decision-maker, being an SB and not just an ordinary agent. The SB knows the agent very well and can make extremely good predictions about how the agent will decide. For some reasons the SB wants to be nice to the agent—turning  $x$  into \$2—if the agent chooses the one-box alternative, but being less nice—turning  $x$  into \$0—if the agent chooses the two-box alternative. Since the SB always makes good predictions regarding the agent’s behavior, she will in advance have put money or no money into the first box according to her preferences for results. Therefore the agent gets a total amount of \$2 every time he chooses the one-box alternative, but just \$1 when he chooses two-box. This higher expected value is an argument for choosing the one-box alternative. A rational agent should choose the dominating strategy and he should also choose the highest expected value. According to Nozick and his supporters the rational agent is caught in a paradox, receiving contradictory advise. Which strategy should he choose, one- or two-box?

Newcomb’s paradox is designed to make the point that it is not clear what is to be the proper advise for someone who is striving for rationality. Maybe the more modest and less greedy one-box alternative is also the best for the agent himself? The paradox, if genuine and not just confusing, does not only cause a problem regarding the position of rationality, but also elicits an element of uncertainty for generosity/altruism. If the agent prospers by abstaining from the second box, it seems dubious whether this action should properly be described as generosity/altruism. But first we need to examine the connection between the agent and his mysterious co-player. Something else than the SB is needed, not only for realism’s sake, but also for a proper understanding of the paradox itself.

		<u>The Superior Being</u>	
		gives \$2	gives \$0
<u>Agent Choice</u>	One-box	A 2	B 0
	Two-box	C 3	D 1

Fig. 1. Agent payoffs in Newcomb’s paradox.

		<u>The Superior Being</u>	
		gives \$2	gives \$0
<u>Agent Choice</u>	One-box	A 2 Probability $p_o$	B 0 Probability $1 - p_o$
	Two-box	C 3 Probability $p_t$	D 1 Probability $1 - p_t$

Fig. 2. Agent payoffs in Newcomb’s paradox.

One way of understanding the paradox is by seeing it as a game that is sequential with the Supreme Being acting as the second player. The extraordinary capacity of the SB to predict the behavior of the agent can be transformed to a more humble capacity; she observes the agent’s selection and only then, not before, makes her own choice. After the agent has chosen the one-box alternative, the SB will choose A, and after the agent makes a two-box choice, she will choose D. By these revealed preferences, we can conclude that the Supreme Being has a higher payoff herself of square A than of square B. We also know that her payoff is higher in square D than in square C. The agent can therefore make a good prediction of SB’s move. Being aware of SB’s preferences and her possibility to make these choices according to her preferences, simplifies the agent’s decision. Backward induction will tell him that one-box is the rational choice. A choice of one-box will result in square A and a payoff of \$2, while a choice of two-box will result in square D and a result of \$1. Of course any rational player will choose \$2 over \$1, posing little dilemma.

Another way of understanding the relation between the two choices is to structure it as different probabilities of the SB choice given the different alternatives of the agent’s choice. This can be described as in this revised figure (Fig. 2).

$p$  is the probability that the first box contains \$2 content and  $1 - p$  is the probability that the box contains \$0. But since interdependence exists between the agents choice and the content in the box,  $p$  is divided into two different probabilities.  $p_o$  is the probability for a \$2 content given an agent choice of one-box with  $1 - p_o$  the probability for a \$0 content given a one-box choice.  $p_t$  is the probability of a \$2 content given a two-box choice by the agent, and  $1 - p_t$  the probability of a \$0 content given a two-box choice.

The two probabilities of a \$2 content do not have the same number, but  $p_o$  approaches 1 and  $p_t$  approaches 0. When this is acknowledged, it is not problematic to recommend the modest one-box behavior. Since the probability  $p_o$  is 1, the expected value of one-box is \$2. The probability of  $p_t$  is zero and the expected value of the two-box alternative is therefore only \$1. There is no paradox, but one-box is the rational choice since it has a higher expected value.

Simply put, the mystery of Newcomb’s paradox is that the two variables are interdependent, even if that relation might be disguised. This dependence contrasts to the Prisoner’s Dilemma and many other games that stipulate that the decisions of the two decision-makers are independent.

This change can be seen as a magic interference or an illusory trick. As mentioned, a preference for just one-box can be seen as nicer and more socially beneficial. Choosing one-box can be seen as avoiding the negative outcome of mutual defection in situations with similarities to the Prisoner’s Dilemma. But the question arises as to whether such a split of  $p$  is just another deceptive presentation that masks a deviation from the rules of ordinary games to provide a more presentable solution. Or maybe there is something more in this twist? As illustrated with the transformation to a sequential game, some minor adjustment might make a one-box choice rational for the agent. This is achieved by deviating from the standard premises of most games, but following orthodoxy might not be so rewarding as investigating the effects of softening the independence assumption. There may be more interesting examples and mechanisms for interdependence than the artificial SB in Newcomb’s paradox. The understanding of the paradox might serve as a constructive starting point.

The reasoning above is also the solution to the same problem in slightly different design as to twins that separately can foresee what their sibling will do. Many religious people believe that by doing as good as they can to their fellow men, God will intervene and provide some reward. The question is as to whether there are more secular mechanisms than God that can interject making irrational social behavior rational for the agent himself. The solution to Newcomb’s paradox might carry more insight than simply being an answer to a riddle.

### 3. Social behavior

Dilemmas in games can be helpful tools for formalizing the problem between solutions with socially desired results (solutions that satisfy the minimum requirement of being Pareto efficient) and solutions that are rational for the agents (Nash equilibrium).

The game is normally between two players, but I will revise the situation to an agent acting under two sets of circumstances or environments, with the probability  $p$  to find a social environment and the probability  $(1 - p)$  for the non-social environment. The two environments need not necessarily be seen as two extreme alternatives, but rather as conditions being predominantly social or predominantly non-social.

If  $p$  is high, this is favorable to the agent and if  $p$  is low, his benefit will be lower. The average result can be expressed as different combinations of payoffs and probabilities of different environments. As mentioned, a basic assumption of most one-shot games is that the decision of the agent is independent of the other player or environment.

In a similar way as in the previous section,  $p$  may be split in  $p_{ps}$ , the probability of pro-social environment given pro-social behavior from the agent and  $p_{ns}$ , the probability of a pro-social environment given a non-social behavior by the agent. If the two variables, that is the action of the agent and the character of the environment, are independent, then  $p_{ps}$  is always equal with  $p_{ns}$ . But let us deviate from this assumption and penetrate the possibility of the agent’s behavior influencing  $p$  or the environment influencing the agent’s behavior so that a correlation exists between certain kinds of behavior and certain types of environment. Let us look into this revised game. As the reader suspected this is the same format as the previous game, but with the SB replaced by an environment and an ambition toward finding other kinds of relation to the agent than clairvoyance alone (Fig. 3).

The problem of the game is described by the two average squares. This figure does not presuppose

		The Environment		Average payoff
		Pro-social	Non-social	
Agent Behavior	Pro-social	A 2 prob. $p_{ps}$	B 0 prob. $1 - p_{ps}$	E $2 \times p_{ps} + 0 \times (1 - p_{ps})$
	Non-social	C 3 prob. $p_{ns}$	D 1 prob. $1 - p_{ns}$	F $3 \times p_{ns} + 1 \times (1 - p_{ns})$

Fig. 3. Payoffs in a PD situation.

interdependence, but when independent,  $p_{ps}$  equals  $p_{ns}$  and both can be exchanged for  $p$ . Then square F having the value of  $3p + 1(1 - p)$  will always be higher than the value of square E that is  $2p - 0(1 - p)$  because  $2p + 1 > 2p$  when  $0 \leq p \leq 1$ . This higher payoff implies that the non-social behavior is the rational behavior. Regardless of the value of  $p$ , the agent does better by choosing the non-social action and hope for encountering social favors by others. But if there is interdependence between the agent's actions and the environment, non-social behavior is not necessarily rational.

This possibility is important for situations that are high in both realism and relevance. I think there are three interesting cases for human behavior when  $p$  is split with a tendency for  $p_{ps}$  to increase and a tendency for  $p_{ns}$  to decrease. Both reason and experience indicate not only virtue by popular standards, but also rationality for pro-social behavior.

The first case is to express kin selection in the mode of this model. The question is whether it is justified to believe that  $p_{ps}$  moves towards 1, and  $p_{ns}$  moves towards 0. It seems reasonable to think that a link may exist between the social behavior of the agent and a high probability that this will be done in a kin rather than a non-kin environment. An individual being helpful to siblings provides favor to individuals with the same genetically supported behavior. The more the benefit is costly to give and valuable to receive, the more important it will be to direct the social behavior to the kin-group rather than others.

Correspondingly, the non-social agent will have limited chances of being lucky and end up in a situation of receiving help without providing any. A free rider gene is likely to be inherited by siblings, so the whole group develops towards square D—mutual non-social behavior. Behavior is likely to concentrate into the two squares A and D. If  $p_{ps}$  is large enough to compensate some investments resulting in square B, and the non-social behavior is not too much improved by benefits of square C (manifested by a  $p_{ns}$  higher than 0), the pro-social behavior can be a better strategy than the non-social strategy.

An agent's  $p_{ps}$  will never reach 1, but it might be close enough to 1 making the social strategy advantageous. As before, the difference between payoffs in squares A, B, C and D are also of prime importance. The averages in square E and F will demonstrate the crucial judgement in consolidating payoffs and probability.

A second case for a differentiation of  $p$  is reciprocity. By discriminating between reciprocal individuals and non-reciprocal individuals, an agent can increase  $p_{ps}$  so that only a fraction of the social behavior is lost on non-reciprocal individuals. In a new situation an agent might act non-socially, but succeed in being cryptic, which might generate some square C results. But with time this will be more difficult and  $p_{ns}$  will decrease while  $1 - p_{ns}$

will be a larger part of his outcome. In a small society with frequent interactions the classification of an agent as pro-social or non-social will influence the treatment he will receive. The environment will not be independent, but shaped by the agent's behavior in previous interactions. As with kin selection it seems likely that reciprocity can achieve a situation in which  $p_{ps}$  is increased and  $p_{ns}$  is decreased.

There is a third case of special interest for human beings. This case refers to a behavior with a range of different labels like group egoism, groupism and in-group/out-group differentiation (Thienpont and Cliquet, 1999). Social behavior (often labeled solidarity in this context) will be directed towards the in-group. Social markers and dicotomistic thinking, such as we-vs.-them, will help focus social interactions to a special group. Individuals that show egoistic behavior like buying at an out-group store that is more convenient or recruit an out-group employee to obtain lower cost, will risk condemnation by the own group. Not only will egoistically motivated interactions with the out-group be condemned, but also the behaviors indicating sympathy for out-group members. Ostracism might exclude the agent from the in-group and inclusion in the out-group is often not an available option. There are strong incitements to behave socially towards the in-group, and avoid social interaction with the out-group. The individuals will interact socially in square A with the in-group and receive corresponding pro-social treatment. In relation to out-group members the interactions will be mostly in square D, that of mutual non-social behavior.

#### 4. Altruism's problem with rationality

Having discussed three candidates and found good reason for differentiating between  $p_{ps}$  and  $p_{ns}$ , leaves time to discuss the most popular candidate for pro-social behavior—altruism.

If altruism is directed towards a group that treats favor with favor, we might get a sustained social behavior. But it would be a misnomer to call this altruism rather than reciprocity, kin selection or group egoism. Real altruism would not discriminate between persons returning social favors and those that do not. In the New Testament a major premise of being good according to these standards is that the benefits are directed towards people who are too poor or too antisocial to return the good deeds.

The idea of discriminating by focusing beneficial behavior towards a certain group implies selfish possibilities for the giver and undercuts the central idea of altruism. Maimonides' ladder is an altruistic scale where progress is demonstrated by a series of steps from a personal relationship with reciprocal possibilities to a

more anonymous giving with decreasing reciprocal possibilities (Singer, 1994). This article follows this tradition of using altruism in *sensu strictu*, it is different than rational social cooperation.

If altruism is to be beneficial to the altruist, it has to be obtained in some other way. As such, we have to leave rational solutions and go back to the more mysterious alternatives of Section 1. Many people are in favor of some providence taking care of a generous reward. The most widespread idea of such a solution is inherent in Christianity. Do good, sacrifice, expect nothing in return in this world—and receive eternal life in the next is the central tenet of Christian faith. Many secular thinkers are certainly influenced by this idea, but other sources might also support the idea of some indirect connection that makes altruism profitable even without divine interference.

Behavior can be analysed as strategies in which the disadvantages comply with the advantages and are offset by them. Being honest can bring a person into costly situations. But real honesty cannot only be displayed when favorable, but should be seen as a strategy carried out in a number of situations. In specific cases it would be extremely hard to make an accurate prediction, but from the experience of many situations it is possible to make a reasonable evaluation. The appreciation of many but not all situations, might very well make a rather crude honesty advantageous, even if it sometimes causes the agent negative reactions. Can altruism be such a strategy where occasional losses in square B are more than compensated by mechanisms pulling towards square A?

As pointed out, altruism is often used improperly or as a misnomer to express appreciation for behaviors that one wants to endorse and promote. Throughout society there are many actions that are socially rational for the agent, but often labeled generosity. Lots of people make sacrifices that are described as one-sided giving, but are better understood as parts of an ongoing relationship. Some of it exists simply as spillover from rational acts. When a company provides free samples or a farmer sows his seeds, some of this generates no return. But it is more accurate to view the sown seeds eaten by the birds and the samples consumed by the customers who will never buy the product as spillover from rational efforts rather than examples of altruism. Many people commit a mistake that might be termed ‘average negligence’. They highlight such non-rewarded outcomes ignoring that what really counts is the average result; some efforts will be at a loss, but others give a return that more than compensates those losses. Many times such misallocation cannot be eliminated but should be seen as an integral part of a chosen strategy. If the action of spreading seed is more rewarding than for example eating them directly, there is nothing irrational or altruistic in the farming efforts. Socially rewarded

behavior is dressed up as generosity, when you actually get some valuable reward even if it is not *comme il faut* to point at this rationality in some circumstances. Declarations about a desire to feed the birds do not alter the situation in any significant way regarding the behavior or the result. Still such declarations about intentions and speculations of possible, real, or hidden intentions are often considered most important. The central question of rationality is sometimes abandoned for the peripheral question of psychological altruism. Do you think the agent thought and felt for others as the final purpose, and not of himself, when performing his deed?

A second problem of altruism is that some real altruism is performed under the illusion of a payoff. A payoff might be highly expected even if it is not labeled as such and customs ordinate an attitude of no demands and expectations. At a birthday celebration, comments like ‘a birthday present for me, what a surprise!’ is compatible with expectations of gifts and dire disappointment if no gift is given. People that are conventional and sentimental can be influenced by social pressure to sacrifice for purposes of king, country and the poor of this world, in the expectation that such altruism is of the rational kind, that the deeds done will be rewarded. There are many confused altruists, and the question is whether or not all altruists are. What is the point of self-destructive behavior? Some writers take the opposite position and want to see virtue as well as rationale behind such a behavior.

One example of the *average negligence* is the work of Sober and Wilson (1998). Their accusation is that other scientists do commit a diametrically opposite mistake, something they call ‘the averaging fallacy’, that is, to pay too much attention to the average. They have an arithmetic example to describe their idea as illustrated in Fig. 4.

Sober and Wilson decide to ignore the interdependence between the actor’s behavior and the probability of different environments. Instead they look at a pro-social environment (like ‘Most Al’) and note that non-social behavior (‘Egoism’) has a higher payoff than social behavior (‘Altruism’). Therefore it is justified to label the social behavior as altruism. Then they proceed to other more non-social environments (like ‘Mostly

		The Environment				
		Only Al.	Most Al.	Most Eg.	Only Eg.	Average
Agent Behavior	Altruism	14	13	10	9	12.38
	Egoism	15	14	11	10	11.68

Fig. 4. Strategy and payoff.

Eg') and come to the same judgement. The point is that they do not proceed to a further comparison comparing the averages ('the averaging fallacy')—something recommended in this paper.

As seen in the figure, Sober and Wilson do calculate the average (sic) from this set of different outcomes with different probabilities. To increase correspondence with my previous matrix and reasoning the Sober and Wilson figure can be simplified to two alternative environments (as indicated by thick lines in the figure). Then we can calculate the  $p_{ps}$  and  $p_{ns}$  that tie up to the averages in the figure:

$$13p_{ps} + 10(1 - p_{ps}) = 12.38,$$

$$p_{ps} = 0.79,$$

$$14p_{ns} + 11(1 - p_{ns}) = 11.68,$$

$$p_{ns} = 0.23.$$

The superior average of the social behavior ('Altruism') shows that this behavior can be rationally chosen. But at the same time as rationality is won, altruism is lost. A behavior earning 12.38 is hardly altruistic in comparison with an alternative giving 11.68.

## 5. Group selection and rational behavior

Group selection is generally not understood as an either/or to individual selection. Few scientists, if any, doubt the existence of individual selection. Rather, group selection is seen as a possible complement, as part of some multilevel selection theory. This additional mode of selection is generally accepted as a theoretical possibility, but opinions differ widely about its significance in the real world. The controversial question is whether selection *also* exists on a group level.

A multilevel selection theory needs some stringency to accomplish a separation at different levels. In ordinary selection, heredity and variation function on the gene level, but selection acts on the organism level. Or, expressed in another way, heredity and variation influence the genotype, whereas selection works on the phenotype. When discussing higher level selection, it seems appropriate to raise all factors one level, seeing the group or super-organism, as the unit of selection, rather than the organism. Group selection is then not used as a vague metaphor, but as a homology to individual selection. For group selection the phenotype of the super-organism is the unit of selection.

Group selection should be indicated by a difference in fitness that differs from the differences caused by individual fitness. If a social animal forms groups of between 100 and 200 and they split at the upper limit, group selection might be calculated in the following way. Individuals of this animal might be distinguished by their type A and type B behavior. If individuals in

type A groups have a reproduction surplus of 5% per generation, and individuals in type B groups have a surplus of 15%, we expect to see a shift in the total population to a larger number of B's. Some groups of A's and B's will go extinct, but we would expect that the share of B groups will increase correspondingly to the 10% higher reproduction. If the share of B's in the total population increases by more or less than that number, this can be seen as some effect of advantages or disadvantages at the group level. The survival possibilities of groups differ from the reproductive pattern within groups. Group selection implies that an individual's reproductive success within the group is not the only issue of importance. A progeny of six in one group, will be outperformed by a progeny of three in another, if the latter group replaces the first.

Mayr (1997) expresses the distinction of group selection in the following way: "whether groups as cohesive wholes can serve as targets of selection. The answer is 'it depends', some do and others not... A group, the selective value of which is simply the arithmetic mean of the fitness values of the composing individuals (when in isolation) is not the target of selection... This is false or soft group selection... If due to social actions, the fitness of the group is higher or lower than the arithmetic mean of the fitness values of the composing individuals, then the group as a whole is an object of selection, hard group selection." This emphasis on the group as a unit for selection to distinguish group selection is shared by most Darwinists e.g. Maynard Smith (1976) and Dawkins (1979).

False or soft group selection tends to be nothing special, but just the actual result of the evolution of social animals. A new type of behavior, type C, might spread in the population because it is more advantageous for these individuals. There is no special group effect, just social effects primarily concerning individuals.

Some scholars are convinced that group selection is something that exists and almost by definition, offsets individual selection. If not assuming a divine intention for the mechanics of the world, there is no reason to presuppose this counter effect, but there are possibilities that group selection will increase rather than decrease the differences in individual fitness. Group selection effects are not necessarily in favor of type A in this hypothetical example. Rather, a group with the more fecund type B is more likely to achieve a benefit of numbers when in conflict with an A group.

Already Darwin (1871, p. 166) and the social-Darwinists pointed to group benefits associated with actions of social coordination. In a contest it seems likely that strong group egoism enhanced success in encounters with other groups. A precondition for this is that the rationality is well established within the group, that is, there are good reasons for members of the

in-group to show solidarity to the in-group. Group stability and strength are likely to be supported if individual and group interests are brought in harmony. If a trait would be advantageous at the group level, a precondition for having an impact is that it can be sustained within the group. This is unlikely to happen by luck or providence. Rather, it seems as if human groups have experimented extensively to nurture different concepts that are forced upon all its members in the hope that they will generate a decisive factor that will help the group to triumph over its competitors. Both numerous progeny and high group stability are likely to be an advantage in most situations of group competition. Conformity and high fertility are not the terms chosen in many descriptions. ‘Love of children’, ‘empathy with others’ and ‘strong solidarity commitment’ are more positive phrasings. An altruistic classification is often not motivated, but the shallow impression when the sacrifice elements are highlighted and the rewards neglected.

Let us discuss two kind of societies, Egoistic (*E*) and Social (*S*), that have two types of individuals, egoistic (*e*) and social (*s*). For several reasons it seems probable that under many circumstances an egoistic individual fares better than a social individual in an egoistic group; the sacrifices of the social individual will carry costs, but bring no social reward. The individual will be considered stupid or strange rather than as virtuous. This can be described as  $eE > sE$ . In the social group, on the contrary, it seems likely that social acts will bring prestige and other advantages, while egoistic pursuits will be punished. Here, the opposite relation is likely  $sS > eS$ .

One effect of such a condition is that it will imply some stability to the group. The egoistic group will stay egoistic since egoists will do better than social individuals. By the same token the social group will stay social. Such stability will increase the likelihood of a group selection effect.

Social groups and individuals might prevail over egoistic groups and individuals in a situation having the following ranking of fitness:  $sS > eS > eE > sE$ . The superiority of the social group is likely to result in egoistic individuals constituting a decreasing proportion of the total population. The social groups are sustained by social reward and punishment so the social individuals in the social group (*sS*) outperform the egoists (*eS*). In a predominantly egoistic group the possibilities for social individuals (*sE*) are poor and they gain the lowest result. More importantly, the majority of egoistic individuals in the egoistic group have a problem to adapt since they impair their situation by modifying their behavior becoming *sE*. For group selection to be an important factor, I consider it a necessity with such a lock-in effect that prevents individual adaptation and also is a brake on adaptation on an aggregated level.

It seems unlikely that group selection gains support via altruism. Altruistic individuals face some certain problems in a fixed group. Egoists that end up in square C will increase their share compared to the altruists in square A. At some point the environment will be reclassified to ‘mostly egoist’ but the advantage for egoistic behavior will prevail. Finally, we end up with only egoistic behavior in an egoistic environment.

In contrast, there are good reasons to expect group egoism to be a force for group selection. The more it can enforce the principle ‘one for all, all for one’ the more likely is such an effect. Group conformism sustained by incentives is characteristic and most important in human groups. A group that cannot maintain its distinctiveness fades away. A strong consolidation prepares the group for an encounter in which there might be group selection. It seems likely that groups with stronger conformism obtain a lower death rate than groups with lower conformistic inclination. This implying that we have a trait to some degree selected for at the individual level, but also to a substantial degree at the group level.

But the effect is not only potentially positive. Consolidation and conformism might result in a stubbornness that will increase the cost in situations of inferiority and defeat. Consolidation might inhibit cultural transfer and in that way causing group selection effects. The radical dichotomy of ‘do it our way’ or ‘die our way’ might result in die our way (Hardin, 1995). A costly religious habit that is rigorously sustained might be an example of a situation where  $eE > sE > sS > eS$ . Social is not a synonym for superiority and it is also possible to establish maladaptive social practices. Here the social group rather than the egoistic group suffers from a lock-in effect.

## 6. False or soft group selection as an explanation

In Sober and Wilson’s (1998) book group selection in a strict sense is replaced by a broad definition that includes false or soft group selection. Through enlargement, kin selection and reciprocity are included under the umbrella of group selection. (For further discussions of this book see e.g. Reeve, 2000; Trivers, 1999; Maynard Smith, 1998.) The family is considered as a group and some individuals temporarily cooperating are considered another group. This wide ‘group selection’ no doubt exists, since few if any question the practical existence or theoretical rationale with reciprocity and kin selection. But such neologism contributes little but confusion, particularly since kin selection and reciprocity are two well understood mechanisms of social behavior. Little or nothing is gained by lumping them together with undeveloped alternatives under the semblance of ‘group selection’. Group selection is not used as a homology to individual selection, but as a

loose metaphor. The possibility to distinguish different levels in a multilevel selection theory disappears. The group is not the unit of selection, but instead ‘group selection’ is used for anything that can provide a better or worse environment for some individuals compared to others. The difference between averages of groups is termed group selection and just the variation within a group is called individual selection. With the disparity and overlapping of groups, most differences between individuals can be reclassified from individual selection to group selection.

One might wonder why the proponents of group selection make their suggestion so hollow. One evident explanation is that by making group selection so wide its existence is an easy case to prove. It is hard not to suspect that the terminology to a large extent is an effort to legitimize some controversial ideas by putting them in the same black box as some well-understood processes. No one can claim the box is empty. But this method loses focus on the issue of group selection in a strict sense for false or soft group selection. If the issue is to prove group selection, then this wide concept seems to be a dubious choice. However, my impression is that the ultimate aim is to promote altruism rather than group selection (the subtitle of the book is one of several reasons for that judgement). In order to save altruism as an evolutionary possibility some new forces are needed.

Sober and Wilson search for a possibility for altruism in a population where the payoff situation can be presented following the same design as previously:  $eA > aA > eE > aE$ . The altruists are doing quite well if living in a predominantly altruistic group. The heart of the altruistic possibility is that altruists in predominantly altruistic groups do better than egoists in predominantly egoistic groups;  $aA > eE$ . However, the altruistic group is unstable since  $eA$  are doing even better. The familiar problem is that over time not only will the altruist in an egoistic group go extinct, but the altruists in the altruistic group will meet a similar fate. The expected long-term result being that the whole population will consist of only  $eE$ .

The solution tendered by Sober and Wilson is a frequent reshuffling of groups. Before decay in the A group can bring an increasing number of  $eA$  to such a high level that the A group becomes an E group, with most unfavorable conditions for the altruists, they suggest that groups might regroup so that the decaying proportion of altruists is lifted. This is the central idea of their book.

A random reshuffling will not help by itself. Some altruists might be lucky and end up in highly altruist groups, but others will end up in even more egoistic groups than average and they will suffer even more. Random reshuffling is no escape from the law of averages. What is needed is a mechanism that is non-random, but instead effectively favors the altruists by

putting them together in altruist groups and discarding the egoists in egoist groups. The mechanism looked for is supposed to be secular and regular, but a serious weakness is that Sober and Wilson have no new interesting idea of such a non-random mechanism. They promise: ‘Of course, we still must explain how, generation after generation, altruists tend to find themselves living with altruists, and selfish individuals tend to associate with other selfish individuals’ (p. 26), but they do not deliver. Repeatedly they even slip back into hoping that a random process might be sufficient (e.g. p. 64), while on occasions they confirm that a random reshuffling is insufficient. They declare openness to all possibilities. It is unclear if Sober and Wilson have any additional suggestions of practical interest.

As an example of desired social behavior Sober and Wilson often refer to the hardworking little red hen in a popular children’s story. She shares the fruits of her labor with the free riding hens, so that her hard work makes herself only a little bit better off. Ayn Rand (1957) tells a more radical story about a republic of red hens. In the book *Atlas Shrugged*, the hard working human providers get fed up with all free riders and they re-cluster in a distant valley and establish a society with new moral order to prevent the emergence of a new exploitation. The ethics of unselfishness is rejected and the valley prospers. The abandoned free riders preach the virtue of unselfishness to each other, all trying to influence the others to do the providing and hoping to get the receiving for themselves.

Sober and Wilson do not suggest such a rebellion of the red hens, but they hope the hardworking hens will be rewarded in some indirect and unintended way. That the red hen gets some benefit by working hard rather than not working is a minimal benefit, and there is an obvious threat that the little red hen will go extinct. Sober and Wilson hope that the little red hen will have the luck to have some other red hens close by, so she will not only be a provider but also a receiver. Therefore the hope is for a re-clustering mechanism. But they do not want an explicit exclusion of the free riders. It is hard not seeing the corresponding idea in most religious thinking. God is thought to take care of the rewarding so that humans should not preoccupy themselves with this issue, but trust him; a post-death super-natural selection.

Why this fascination for unintended results? It might be the case that Trait 1 is linked to the pro-social Behavior Alfa and Trait 2 to the free riding Behavior Beta. By having some mechanism that sorts Trait 1 from Trait 2 the discrimination is done, but not directly between Alfa and Beta. However such a correlated variable will change nothing of significance. Rather the significance of the traits might be that they are instrumental in separating Alfa from Beta.

Results are what counts, not a real or pretended ignorance between trait and behavior. The attraction of

Sober and Wilson to psychological altruism makes them think that non-intentional discrimination—preferring Trait 1 while not seeing the link to Behavior Alfa—is altruism, while intentional discrimination is not compatible with altruism. Then it is rather a reciprocal individual cooperating with other reciprocals to mutual benefit.

Groucho Marx said that he would like to be a member of a real sophisticated club—one that was so sophisticated it would never accept a bum like himself! Sober and Wilson appear to have a similar project. They are searching for mechanisms that support pro-social behavior. The most admirable pro-social behavior is altruism implying no discrimination against non-contributing individuals. In altruism, help is not given with reciprocal restrictions such as ‘I’ll scratch your back if you scratch mine’. Therefore these altruists need some supporting mechanisms that re-cluster them, so when they are nice to their neighbor, they will mainly be nice to other altruists. Sober and Wilson hope that their broad definition of group selection will be wide enough to provide such a mechanism. Their chances of success appear similar with those of Graucho Marx.

## 7. The probability of human group selection

Acculturation, copying and adaptation instead of survival of the fittest group are the alternatives to group selection. Of course these are frequent, but the question is whether group selection also exists.

Many other authors are inclined to see altruism as essential for sociality. Group selection is then perceived as a possibility to explain altruism and solve the problem of altruism with rationality. Edward O. Wilson (1999) considers the interesting issue with group selection being a potential explanation for altruism. Matt Ridley (1996, p 188) even defines group selection as a counter force to rational self-interest. However this focus carries the danger of losing sight of group selection linked to more promising premises. To my mind, the case for group selection in line with rationality looks much stronger than a case without. The suggestion of Sober and Wilson that some kind of group selection could materialize through frequent re-shuffling, matches poorly with anthropological data. It looks like another hypothesis leading the search efforts in the wrong direction.

More promising is to stay with the super-organism and attribute it with a potential lifespan of several human generations. Growth and fission of the group is closely related to the progeny of its individuals. Other authors have highlighted low migration as an important factor supporting a possible group selection (following Hamilton, 1975). High rates of migration reduce both genetic differences and promote acculturation diminish-

ing other differences of importance between groups. I think the most crucial factor for group selection is the possibility of different death rates of groups with different traits. Some groups increase in size and they divide or keep the spin-off groups connected in expanding societies. Other kinds of groups perish. Some females in a defeated group might be assimilated into a victorious group, but this is not fusion in proportional terms. In the following generations the frequency for the genes that differed between the two groups will be radically changed. The hypothesis proposed in this article is that winning groups have contributed to increases in conformist pre-dispositions.

Genes for non-conformist behavior from the conquered minority will not prosper in the new group, but have a continuous pressure imposed against them. In the symbols previously used  $sS > eS$ . The social system is not undermined by an individual selection that is disposed in the opposite direction—the crucial difficulty for all hypotheses of strict altruism. Instead the preferences generated by group selection will be further strengthened by individual selection within the new group.

The group egoistic hypothesis of group selection maintains that groups with individuals with a predisposition for conformism prevail better in human conflict and that such groups by a lower death rate have increased their share of the human population. Groupishness is also favored by culture, but this is hardly an either/or situation. On the contrary, a predisposition for conformism makes the task of agitation a lot easier.

This sketchy model is in line with substantial empirical evidence from social psychology. Conformism, out-group antagonism and opportunism seem to be of a magnitude that it is justified to see such behavior as characteristics of the human species (van der Dennen, 1990). I do not claim that these traits prove the necessity of group selection, but I contend that they are not only supportive but also mutually supportive. Group selection can be seen as a selection force favoring group egoism by increasing selection pressure for traits such as in-group affection and adjustment. Group egoism will in turn increase the possibility of group selection by creating stricter groups and support tendencies to cultural separation; from within group jargon to separate languages.

Crucial for the functioning of group selection is that there are mechanisms that make  $sS > eS$ . Several theoretical papers address different mechanisms, in addition to kin selection and direct reciprocity, which can evolve and maintain social behavior. Boyd and Richerson (1985) have developed the case for a transmission bias to favoring conformism. Nowak and Sigmund (1998) have contributed with stimulations to test possibilities of evolution of cooperative behavior. Indirect reciprocity can be established and maintained

when agents' decisions of giving or not giving are based on 'image scoring' of the potential receivers. The score is determined by previous behavior in the donor position vs. other group members. The cost of giving is often sufficiently compensated for by the increased possibility of receiving. Simulations by Leimar and Hammerstein (2001) indicate that a condition of 'good standing' of the receiver, and giving as a sign of extra quality of the donor, improve the theoretical possibilities of sustainable indirect reciprocity.

Other interesting contributions are generated by experimental methods. Wedekind and Milinski (2000) tested the effect of 'image scoring' corresponding to the model of Nowak and Sigmund (1998). Participants reluctant to give in the donor position were more discriminating only giving to partners with a very cooperative reputation. Giving increased when the amount at stake was relatively small as compared to the assets of the donor, so the level of risk seems to be an important factor for indirect reciprocity. Fehr and Gächter (2002) tested contribution to public goods games with and without a punishing possibility. With no punishment the contributions diminished over time. With punishment the contributions were high and increasing. 84% of participants punished someone else at least once. Because of reshuffling between groups, a player did not benefit from his own investment in 'correcting' others, but still, punishing was frequent and maintained a high level of cooperation.

In this article, I have tried to show that altruism and rationality do not have a hidden connection. All that is proven, time after time, is that for a social animal it is often not rational to act in an 'atomistic' and short-term manner, but that there is a rational potential for social behavior. This is why we became a social animal in the first place. Group egoism is a force in human mentality and behavior that is often noticed and criticized by individuals whose interests are threatened. For individuals whose interests are promoted, group egoism is cryptic and there is evidently a preference for a posture against, rather than in line with, self-interest. The proportions in this mix of hypocrisy and self-delusion are debatable. Normally group egoism of the preferred kind is described as altruism, solidarity and loyalty. Altruistic pretensions signal a commitment to contribute to the common project of the group.

Real altruism lacks an evolutionary explanation and it might be wiser to look for cultural explanations than some undiscovered mechanism. In investigating evolutionary mechanisms as group selection, it might be wiser not to burden them with assumptions and hopes of altruism. The conclusion is that group selection is not a mechanism that has initiated altruism. However group

selection and group egoism are not only a possible, but a highly probable, combination.

## Acknowledgements

I thank Olle Leimar, John Maynard Smith and Birgitta Tullberg for valuable comments and suggestions and Bohdan Sklepkovych for improving the text.

## References

- Binmore, K., 1994. *Playing Fair—Game Theory and the Social Contract*. The MIT Press, Cambridge, MA.
- Boyd, R., Richerson, P.J., 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.
- Darwin, C., 1871. *The Decent of Man, and Selection in Relation to Sex*. John Murray, London.
- Dawkins, R., 1979. Twelve misunderstandings of Kin Selection. *Tierpsychol* 51, 184–200.
- van der Dennen, J.M.G., Falger, V.S.E. (Eds.), 1990. *Sociobiology and Conflict—Evolutionary Perspectives on Competition, Cooperation, Violence and Warfare*. Chapman & Hall, London.
- Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. *Nature* 415, 137–140.
- Hamilton, W., 1975. Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox, R. (Ed.), *Biosocial Anthropology*. Malaby Press, London, pp. 133–153.
- Hardin, R., 1995. *One for All—The Logic of Group Conflict*. Princeton University Press, Princeton, NJ.
- Leimar, O., Hammerstein, P., 2001. Evolution of cooperation through indirect reciprocity. *R. Soc. 268*, 745–753 Doi:10.1098/rspb.2000.1573.
- Maynard Smith, J., 1976. Group selection. *Rev. Biol.* 51, 277–283.
- Maynard Smith, J., 1998. Review of unto other by Sober & Wilson. *Nature* 393, 639–640.
- Mayr, E., 1997. The objects of selection. *Proc. Natl. Acad. Sci. USA* 94, 2091–2094.
- Nowak, M., Sigmund, K., 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393, 573–577.
- Nozick, R., 1969. Newcomb's problem and two principles of choice. In: Rescher, N. (Ed.), *Essays in Honor of Carl G Hempel*. Reidel, Dordrecht, Netherlands.
- Rand, A., 1957. *Atlas Shrugged*. Random House, New York.
- Reeve, H.K., 2000. Book review of unto others by Sober & Sloan Wilson. *Evol. Hum. Behav.* 21 (1), 65–72.
- Ridley, M., 1996. *The Origins of Virtue. Human Instincts and the Evolution of Cooperation*. Viking, London.
- Singer, P., 1994. *How are we to Live—Ethics in an Age of Self-Interest*. Mandarin Paperbacks, London.
- Sober, E., Wilson, D., 1998. *Unto Others—The Evolution and Psychology of Unselfish Behavior*. Harvard University Press, Cambridge, MA.
- Thienpont, K., Cliquet, R., 1999. *In-Group/Out-Group Behavior in Modern Societies*. NIDI CBGS, Brussels.
- Trivers, R., 1999. As they do to you. *Sceptic Magazine*, 6, #4.
- Wedekind, C., Milinski, M., 2000. Cooperation through image scoring in humans. *Science* 288, 850–852.
- Wilson, E., 1999. Speech delivered at the annual meeting of the Association for Politics and the Life Sciences in Atlanta September.