

Group egoism; investigating collective action and individual rationality

Jan Tullberg*

Stockholm School of Economics, P.O. Box 6501, SE 113 83 Stockholm, Sweden

Received 30 September 2003; accepted 21 November 2005

Abstract

The paper discusses rationality at two levels. According to a common opinion the key to rationality at the group level is a willingness to contribute unselfishly at the individual level. I claim that even when rationality at the individual level is denied, it is often achieved by social punishment. The difficulty to organize a social project is exaggerated. The relative ease of implementing individual rationality has the disadvantage of frequently establishing and maintaining social order that, despite some enthusiastic proponents, fails to deliver rationality at the group level.

© 2006 Elsevier Inc. All rights reserved.

JEL classification: A13

Keywords: Group egoism; Punishment; Social rationality; Altruism

1. Introduction

Neo-Darwinism has shed significant light on two types of social behavior; kin selection and reciprocity. This article focuses on a third kind of social behavior herein labeled “group egoism”. It is linked to methods and mechanisms such as badging, in-group cooperation and out-group competition. In humans, there exists a considerable pre-disposition to act in coalition with others under the devise “one for all, and all for one”. This type of behavior is often admired as solidarity, but also frequently considered a serious problem for instance in fueling ethnic conflict. Tribalism (Hardin, 1977) and groupism (Allot, 1999) are other suggested terms similar in concept to group egoism.

The general picture can be described in the following way. There is a benefit (B) for the group that can be obtained as collective good; higher wages, a sustainable environment and an effective

* Fax: +46 8 307225.

E-mail address: jan.tullberg@hhs.se.

		Group level	
		RAT	IRR
Individual level	rat	section 4	section 5
	irr	section 6	section 7

Fig. 1. Four combinations of rationality.

defense. The individual cannot achieve this benefit by himself but if obtained by the group, the share of the benefit for an agent (b_a) in a group of n members can be calculated as $B/n = b_a$. When all members contribute to the realization of this social benefit the aggregated cost (C) can be distributed $C/n = c_a$. If $b_a > c_a$, the aggregation for the group becomes $B > C$ and the project can then be considered rational at the group level (RAT).

The problem for obtaining solidarity is that it is even better to free-ride because $b_a > b_a - c_a$. The agent then gains the benefit B/n just marginally impaired without his contribution, by avoiding to make the more substantial sacrifice c_a . To prevent defection the free-rider is threatened by punishment (P). If the expected value of punishment, P , exceeds the cost c , it is rational for the individuals to join the efforts. The second condition of rationality (rat) is: $p_a > c_a$.

However, rationality at the individual level (rat) does not by necessity result in rationality at the group level, but the behavior might be irrational at the group level (IRR). Even if it is rational for the agent to contribute, because of the threat of P , and the common good B is obtained, the aggregated cost C might exceed B . Both the delivery of P and the suffering of this punishment implies significant costs for individuals in the group.

An often suggested way to reach group rationality is to focus on the social good and ignore rationality at the individual level. The individual is asked to perform c_a even if there is no direct benefit for him and no threat of punishment. One reason for this is to avoid the double cost of P . The individual is asked to act unselfishly (irr) for a greater good (RAT). However, irrationality at the individual level is neither any guarantee for rationality at the group level, but the aggregated result of benefits and costs might turn out to $B < C$ that is the situation of (IRR). This gives four different possibilities (Fig. 1).

Before discussing each of these four combinations in the matrix, there will be a first section penetrating the complexity of rationality and another section bringing up some experiments that contribute to the understanding of group egoism. After four sections discussing the different combinations of rationality, the paper ends with an eighth conclusive section.

2. The anatomy of rationality

The analysis will start by discussing two central questions for the reasoning in the article and for the relevance of its conclusions. First, how to draw the demarcation line between rationality and irrationality on the individual level? Second, how to attain aggregation of individuals with heterogeneous preferences?

Strong assumptions will often make the central line of argument more convincing. But if these assumptions for the reasoning are very strong and demanding, they might exclude most real world situations and the whole exercise is just becoming a thought experiment. It is important not to make the simplification of the context in a model at the price of relevance.

2.1. The demarcation line between rationality and irrationality

Some economists prefer to define rationality on the basis of very strong assumptions. The rational agent is myopic, has complete information and unlimited calculation capability and is always selecting the optimal solution. The risk with this strict definition of rationality is that most humans will empirically be classified as non-rational or irrational according to such a demanding standard. Since long, psychologists as [Kahneman and Tversky \(1979\)](#) have made not only efforts, but also taken some pleasure, in demonstrating such an economic man a rare species.

I think it is reasonable to see perfect rationality as one end of a spectrum and an inept confusion as the other extreme. However, instead of excluding everything but perfect rationality from the rationality label, I think it is preferable to include more in rationality when drawing the demarcation line between rationality and irrationality. Following Adam Smith, illustrated in the following quote, I would describe man as acting for his long-term self-interest rather than being myopic: “. . . the passion for present enjoyment; which, though sometimes violent and very difficult to be restrained, is in general only momentary and occasional. But the principle which prompts to save is the desire of bettering our condition, a desire which, though generally calm and dispassionate, comes with us from the womb, and never leaves us till we go into the grave.” ([Smith, 1776](#), p. 341).

Experimental economists have abandoned demands on economic man to optimize without regarding the costs of doing so ([Harrison, 1992](#)). It is often hardly irrational to follow a heuristic rule instead of making a complex calculation when only a dismal gain is at stake. The central demand of rationality is to make an adequate evaluation when it is economically motivated. The bounded rationality school ([Simon, 1955](#)) is often seen as an attack on rationality, but I think it should rather be seen as a proponent for this wider view of rationality. “Satisficing”, rather than optimizing, helps the agent to turn his focus on more pressing tasks. In my opinion, a long-term view should be seen as the rational, and the myopic and optimizing as a deviation that might be irrational; sometimes the best is the enemy of the good.

Rationality in the treatment of this article is strongly linked to personal self-interest. A person that ignores risk, neglects common knowledge, abstains from vital information in an important issue and decides by a whim might be seen as behaving according to preferences, but not to self-interest. Such a person should therefore, be considered irrational. Rationality does not require to find the best of all possible solutions, but rather to take the better of two alternatives. To be irrational it is not enough with a negligence, but some serious neglect or willful sacrifice of self-interest has to occur.

I think some readers object to classifying willful sacrifices as irrational, so there is a need to elaborate more on that issue. The opinion that a certain action is a moral duty might be a decisive factor to act against self-interest. How should such considered preferences against self-interest be classified? I think it is constructive to treat virtue and rationality as two separate factors. Some behavior perceived by the agent as a moral duty are also very rational, e.g. he gains esteem by others and avoids punishment. Other moral acts imply self-sacrifice and can therefore, be classified as irrational. All immoral acts are not in line with self-interest, but many are self-destructive.

There are problems with definitions of rationality that are including more into rationality than suggested here. One alternative is to accept any consistent personal choice as rational; since the person performed according to that person's preferences. Such a use does not only imply the triumph of rationality, but also of psychological egoism. The classical example is Hobbes explaining his gift to a beggar as egoism since the action was motivated by feeling good about himself. If all behaviors are seen as manifested preferences, there are no irrational individuals, just some individuals with odd preferences. The very fact that a person is doing something indicates that he is satisfying some personal preference that dominates over the drawbacks. Every apparent self-sacrifice – perceived as altruistic or destructive – then becomes psychologically selfish and rational (for an effort to refute the hegemony of psychological egoism see [Batson, 1991](#)). Interest is, therefore, often more useful than preferences implying an objective element.

In an evolutionary perspective it is motivated to understand feelings and psychological inclinations as assistant mechanisms, to self-interest (then not only for the organism, but including reproductive considerations). In a changing environment some such first order preferences like emotional inclinations might be contrary to second order preferences like self-interest. Some observers honour such rigid first-order preferences because they meet the consistency criteria. Here, I want to make the point that a permanence of preferences might be a source to irrationality.

Is it then possible for an observer of an agent to make a more objective evaluation or is this just an exchange of one set of subjective preferences for another? The observer's judgment whether the action of an agent is rational or not should avoid giving his own preferences center stage, but consider the position, interest and preferences of the agent. Different observers will often differ in their judgments of how to act rationally in a situation, and therefore, also make different judgments of whether the actual actor is behaving rational or irrational. Some fuzziness around the demarcation line is not a result of poor definitions, but that the cognitive problems are significant. There are tough borderline cases between rational and irrational, but this does not result in an impossible mission.

In a dialog of Longinus the proper action is discussed. "Were I Alexander," said Parmenio, "I would accept of these offers made by Darius." – "So would I too," replied Alexander "were I Parmenio." Disagreeing with Alexander the Great, I think the issue at stake is not temperament, personality or preferences, but how to reach a rational decision for Alexander and Macedonia concerning peace proposals of the Persian ruler.

The human capacity for empathy opens for the possibility to understand another person's position. In the situation of a terminally ill person, suffering substantial pain, it is easy to understand a preference for euthanasia and to consider such a choice rational. It is also possible to understand the suicidal teenager suffering a devastating romantic disappointment. However, the empathy in that case is not likely to lead to the conclusion that suicide in this situation is rational.

The empathy exercise does not imply that you abandon your own judgement to become a clone of the other person. You can perform an empathic reflection even without sympathy for the other. Fierce opponents in war and peace often reflect over the best move of their opponent and try to imagine the situation from his perspective. This is normally done without losing either their own identity or their adequate judgment of different actions.

2.2. *Aggregation and heterogeneous preferences*

Also for aggregated rationality it is an observer that makes the judgement considering the agents' preferences and interests. In the introduction I have described a base case where individuals in the group have homogeneous preferences and interests. This is a simplifying assumption under

which I will try to make some points of interest. If these conclusions are only applicable under very special circumstances, they would be of very limited interest. I will therefore, present some reasons why it is motivated to expect that the conclusions in the simplified context also will be valid in a situation with more heterogeneous and complex preferences.

Even if preferences are not identical, there are often some general similarities. A commodity such as money gives the possibility to fulfill many different preferences. Life and health is a precondition for many of the good things in life. Rawls (1971) talked about “primary goods”, as those people wanted to have more of, because they increased the chances of realizing other goals and wants. This is much in line with the tradition, since Mill, to always prefer more to less with only two exceptions. Man also tries to limit his input of work and has an inclination for spending that sometimes end up with having eaten the cake despite also wanting to have it uneaten (Mill, 1836). Even if preferences and interests are not identical, they can show strong overlap among group members and that is sufficient for, e.g. a trade union. A strike is possible even if all members have different spending plans for the demanded increase in pay.

Many social dilemmas assume a substantial synergy effect making the aggregated benefit larger than the investments in the public project and larger than investment in personal projects. Here there is no such assumption, but the social project might be inefficient providing less benefits than the aggregated input. In the RAT-qualification no utilitarian ethics is implied; there is no assumption that all individuals have a duty to contribute to the common good. Nor is it assumed that collective action should be restricted to public goods. The gains are substantial for the group members if they can obtain transfer of private goods from other groups.

Sometimes preferences are not only varying in magnitude, but stand in direct opposition. Some citizens might be not only less enthusiastic than others about the transformation of a junkyard to a park, but consider it an impairment since such a change is closing the possibility of a transformation to public parking. Military defense is a public good in some opinion, but a public bad according to others. The policy to be implemented is most often determined by power, making the number of people in support versus the number in opposition an important factor. Sometimes the in-group (often the majority group) can be seen as morally wrong in exploiting the out-group, and this might be immoral but not irrational. Even if the cost to the out-group is higher than the gains of the in-group it is not irrational from an in-group perspective (but of course a mistake from a utilitarian perspective). The classification as irrational is spared to situations where the behavior becomes destructive for the in-group.

One way to make a more ambitious examination of costs and benefits for different members of the group is to quantify the benefits. Money is rarely used explicitly in this article, but one can see costs and benefits as measured in money as well as in utility, a more ultimate but less transparent currency. One method for sorting out rational projects using money is the willingness-to-pay method. The benefit B is calculated as the sum of all individuals' estimate of the maximum price they should be willing to pay to obtain e.g. a potential new public good. This sum is a demand estimate including the consumer surplus. The cost estimate in money, C , is often more easy to quantify. If C is less than B , it is rational for the group to introduce this public good. How the cost is to be shared is a separate issue that can be controversial. I will not discuss the influence of strategic considerations on the willingness-to-pay declarations and other complications, but just notice that, if honest, this method is a way to obtain a grounded answer to the question whether something has aggregated value over costs or not.

Aggregated preferences are not only the result of individual preferences. The individual preferences are often very influenced by the common evaluation of different issues. Herd effects are likely to increase homogeneity in preferences (Scharfstein and Stein, 1990, 2000). This might

increase rationality as when an enlightening idea wins general acceptance. But also irrational behavior can become popular and then grow further to a dominant position. There are several other mechanisms for a dubious herd behavior. Nietzsche made a judgement that has not been refuted by another century of human experience: “Madness is something rare in individuals—but in groups, parties, peoples, ages, it is the rule” (Nietzsche, 1886, p. 85).

3. Some experiments relevant to group egoism

Two components of group egoism have been investigated in several experiments. Some have penetrated the conditions for in-group preference. The other subject of interest looks into the mechanism of punishment. Behavioral experimentation might be helpful in illustrating and understanding the phenomenon. What environmental conditions initiate in-group preference and lead to punishment of disloyal group members?

Silverman and Case (1998) tested in-group preference. The participants in the test were given the choice of allocating some monetary support to unknown needy individuals of their own ethnic group or of other groups. In some experimental settings contributions to the own group was impaired by “waste”. Out-group members received \$1 for every given dollar, whereas the own group members received just 0.8 or 0.5. The more ethnocentric, the stronger inclination to favor the own group, even when the reduction by waste made such help less substantial. At no extra cost the own group was preferred by about 90% of the participants, at medium cost around 30–40% of participants preferred to give to their in-group, and at a high cost 10–20% gave to their in-group. Men were more ethnocentric than women. Blacks and Jews were more ethnocentric than white Americans.

Often competition between groups is entrenched in history, with feelings of superiority, victimization and solidarity that enhance in-group preference. An experiment might indicate the threshold of history and conflicting interests needed to initiate a strong in-group preference. The classical Robin Cave experiment (Sherif et al., 1961) showed how 11-year-old boys at a summer camp could be turned into two antagonistic groups with very strong in-group preferences and out-group hostility. No ethnic difference existed between the two groups, but several group-reinforcing factors were used. The two groups were kept apart 1 week prior to introduction and a multi-step competition with a desired prize awarded to the boys in the winning group.

Perhaps fewer incitements are needed to create group polarization? Tajfel et al. (1971) suggested and tested the minimal group situation by just splitting participants into two groups and assigning membership. The experiment was devised as a first base case experiment to show no in-group preference, and the plan was to add environmental factors to achieve an effect. But just a split was sufficient to create an in-group preference. Even when participants were split into groups by a neutral and arbitrary method, the preferences for the own group were manifested. In one experiment, the participants looked shortly at pictures with a large number of dots and then made a guess at their number. Afterwards two groups were formed consisting of “over-estimators” and “under-estimators” based on their prediction of the number of dots. Another differentiation was made according to the preference for pairs of abstract paintings creating Klee and Kandinsky groups. Still the group preference was demonstrated in such minimal groups, when participants attributed points to two anonymous members, one from each group. Participants often selected low point alternatives that favored the in-group receiver rather than alternatives giving more points to both receivers but favoring the out-group member. These experiments verify a quick formation of alliances rather than a necessity for a long indoctrination (Brown, 1986).

Another experiment sheds light on the tendency to punish people that try to free-ride in a group project. [Fehr and Gächter \(2002\)](#) had groups of four persons playing a public goods game in six rounds. The money contributed to the common pool was increased by the experimenter with the factor 1.6, and then distributed equally at 0.4 per person. Contributions from others in the group bring benefit to a player, but his own contribution to the pool provides only a 0.4 return. To make cooperation even more difficult, people were shifted between groups so that there were always three new people involved in each interaction. In a setting with no punishment, contribution fell from round to round to almost nothing in the sixth round. This was not a surprise; even in permanent groups contributions fall in a similar manner. How can this behavior be understood? First, people are not altruists. They do not believe in setting a good example and in forgiving the less social individual in the spirit of the New Testament. Nor do they decide to be rational according to economic theory and give nothing right away in round one. They seem to show a will to contribute, but a bit under the average rather than over the average.

In a second version of the game the participants of the group could fine any of the other three players at a cost of one for a punishment of three. Also in this setting people shifted between groups. In comparison with the previous setting contributions started at a higher level and then increased during the six periods. People who contributed little attracted more punishment and ended up at a disadvantage to those contributing at an average. Contributing much more than average was not to a person's advantage. Here, a social order of effective contribution is maintained.

Each player benefits if other players contribute more due to the fear of being punished by acting too selfish. But punishing others brings a direct cost and no benefit to the punisher (not even the fractious return when giving). Furthermore, this kind of free-riding might not just be seen as shirking for personal gains, but also as Christian forgiveness by turning the other cheek and setting a good example. Since there is no explicit social agreement or some mandate to sustain that agreement the punishing can be seen as vigilantism. But punishment proved an effective way to defend a common interest. Furthermore, it was popular since 84% of the players acted in punishing others on some occasions. Despite a higher average contribution over time the punishing did not diminish, but there was permanently a substantial pressure being placed on participants giving the least.

[Fehr and Gächter \(2002\)](#) stress the point that economic theory is at odds with this behavior. It is worth noting that this behavior is even more disturbing for conventional morality. The free-riding from contribution in setting one could not be prevented despite support from conventional morality. However, in the second setting, participants abstained from the possibility to free-ride from punishing despite less sanction from conventional morality. It seems that another behavior is manifested—other than that suggested by rationality of economic theory and altruism by moral philosophy.

Why did people punish instead of free-ride from the punishing in the hope that others will do the police work? Contributions to punishment were sustained a lot better than direct contributions. Altruism is positive action but nondiscriminatory in helping both loyal and disloyal. Punishment is negative, but it is focused on those that are disloyal and in that way, supports group cohesion. In many situations conformism to group loyalty brings advantages, and these advantages are sustained by punishing free-riders. The higher starting level of contributions in setting two indicates that players expected to be punished if contributing too little. Furthermore, punishing the deviator is often cheap and the act of not participating in punishment may be viewed as badly as committing the deviation itself.

Looking at human behavior there seems to be a prudent case for conformism. Punishment of in-group defectors is frequent and hard. Heathens might be killed, but preferably converted to the

true belief. Heretics are condemned by bulls of excommunication and wanted preferably dead. During war, spies and deserters are killed while prisoners of war often are spared for exchange or as slaves. Out-group relations are often hostile, but there are also truces and alliances. Nonetheless, keeping an in-group conformist discipline seems to be a standard move to secure a power base.

Those tendencies to arbitrary group formations and costly punishment without personal benefit might be interpreted as irrational emotions or actions. I think a more reasonable understanding is that the act of joining groups and punishing defectors has been rational in many situations for a long time. This has generated a predisposition for group egoistic behavior. A predisposition is but a rule of thumb. It can be a good general strategy, but by neglecting rare situations it is non-optimal (Dawes, 2001). Group formation and punishment in these experiments were designed not to benefit the individual but this special case of irrationality may say something about the general rationality. The situations when they are rational are so frequent that the conditions themselves are not scrutinized and only the general rational behavior is applied.

4. rat/RAT

As mentioned in Section 1 the focus here will be on group egoism, but some comments might be useful regarding other mechanisms used in obtaining rationality on individual and social levels.

4.1. “Atomic” egoism

In many circumstances, the individual agent (a) receives the benefits (b_a) and carries the cost (c_a) of his action. A rational criterion for doing an action is $b_a > c_a$. In many situations the externalities for others will be marginal compared to the surplus for the agent; $b_a - c_a$. Such “atomic” activities might not be praised in conventional religion and philosophy, but a lot of deeds will be done according to this criterion and they are very likely to increase the total social welfare $B > C$. People buy one product rather than another and they spend time on one activity rather than another they would enjoy less.

4.2. Asocial egoism

The egoist cares for his personal benefits, but little of the costs carried by others, exemplified by the individual other (o). When the agent gets the benefits b_a and the other the costs c_o , there is no restriction of $b_a > c_o$. It can well be that $b_a < c_o$. Even if the agent’s benefit is larger than the cost of the other, it is disputable whether this transaction will be socially good. The car thief needs the parked car more than the legal owner does at the moment of the theft, but limited conclusions should be drawn out of that. Is this theft not an action that will threaten the social fabric?

Generally, there is no mechanism for asocial egoism to result in social rationality. When costs for others are ignored, when making a decision with effects for others, it is likely that aggregated costs will exceed aggregated benefits.

4.3. Reciprocity

A sequence of an action 1 followed by an action 2 might look as follows. The agent makes the other a favor at a cost in the first action; $1c_a$ and $1b_o$. The second action consists of the other reciprocating by doing the agent a favor; $2c_o$ and $2b_a$. For this reciprocal relation to continue it is necessary that both agent and other are satisfied with the outcome, which can be formalized to

$2b_a > 1c_a$ and $1b_o > 2c_o$. These conditions imply that the social benefit is larger than zero (because $2b_a + 1b_o > 1c_a + 2c_o$). Both actors have a veto in this quality control, so the filter is strong for discarding socially irrational behavior.

It can be said that if $1b_o \gg 1c_a$, the agent should have an obligation to do it, but if the other get such a huge advantage, it would be strange if he cannot make a minor contribution, $2c_o$, that compensates the agent for his contribution $1c_a$. Reciprocity looks most suitable for being rational both at the individual level and on the aggregated level.

4.4. *Group egoism*

In ethnic conflicts, a cleansing project might provide rationality for the group (RAT) by meeting the condition $B > C$. But it might not meet a more utilitarian condition $B > C + C$ (out-group). In-group rationality is the more limited purpose of group egoism. Pretension is almost always that this also implies a total calculation of benefit and this can be supported by reasons for excluding “improper interests” of other people. Such talk might cause some confusion about effects and intentions.

Rather, the benefit for group members is exactly to impose severe costs on the out-group by taking their property or by discriminating them. Such advantages for members of the own group are often considered immoral by observers, but such a criticism is not undermining the rational foundation—rather it might constitute it. The RAT classification focuses the group and does not say anything about the overall desirability, just that group members succeed in obtaining $B > C$.

In the Prisoner’s Dilemma, in the original form, the problem is to help the crooks improving from the Nash equilibrium to a Pareto superior solution. *Omerta*, the Sicilian law of silence, is one such solution. Discussing this example of possible cooperation is instructive in showing that the common good of the sub-group might not be the common good of a larger group. In the Prisoner’s Dilemma case the desired result from society is that the crooks remain in the Nash equilibrium unable to cooperate. This is also true, e.g. for price cartels.

As mentioned there is a second condition for group egoism to be rational. Individuals are asked to contribute c_a and if they do not comply, they risk the punishment P . The probability of being charged and sentenced, s , provides together with P the expected punishment for the agent if defecting; $P \times s = p_a$. If $p_a > c_a$, group members such as the agent better provide efforts for the common benefit.

If an effective incitement policy is established this implies a high level of punishment for members who fail in solidarity and a high probability of detection. This system might be self-supporting and with little dependence of $B > C$. Often, there is an increased need for high in-group pressure if the benefit for the group is dubious. In-group members then get problems to correct the chosen policy. Stubbornness gives strength in a group conflict, but is also serves as a source for Concorde fallacies. The social rationality is gone, but the individual who stops being loyal to the failing project will be punished. For him it is rational to continue to go along.

Most group egoistic projects will be less extreme than ethnic conflict. A substantial part of the benefit is obtained from out-groups, but the situation is less antagonistic. Representatives of special interest, unions and nations will, if honest, acknowledge that a positive balance of benefit to cost for the own group is of prime interest, but they will also claim that their suggestion is positive when costs and benefits for out-groups are taken into consideration. These wider effects might not be crucial for mustering support from the own group or for the RAT classification, but will influence the possibilities to reach a solution with other groups.

Some generalizations seem motivated to be made about group interests in modern western societies. Different special groups overlap so the identification of members with one specific group affiliation cannot be taken for granted by its leaders. Group representation plays a major role in society, but often more as some defense system against unfairness, towards the group. The goal seems aimed at not losing position rather than to obtain new privileges for the own group. This is to some extent an effect of a desired moderate image, but it might also mirror reality.

In the slow political process, the influence of mere individuals is hard to absorb, so to obtain any influence there is a need for coordinated action. It does not only say something about the number of people holding an opinion or interest, but also about the degree of commitment. An interest that cannot muster a vocal interest group is probably so insignificant that it can be considered negligible.

There is often a general praise for unity and “seeing the whole picture”. Special interest groups are often criticized for causing sub-optimal solutions. However, in modern societies it seems that the egoism of one group can be tamed by the egoism of other groups. Galbraith (1952) wrote about a system of countervailing powers. But such a neutralization of group power by other groups is not the only moderating factor. Furthermore, there is ambivalence in commitment from members because they also belong to other groups. This gives the leaders a weak grip upon the members. The possibility is limited to mobilize hard punishment against defectors. In most situations the groups have to deliver a larger B than C for the members, meeting the aggregated criterion. Also on the meta-group level – society – the outcome is often quite satisfying.

5. rat/IRR

A classical example of a rat/IRR outcome is the “tragedy of the commons” (Hardin, 1968). For each farmer it is rational to put an extra animal on the common grazing fields. It gives him a significant benefit at a marginal cost of over-exploiting the fields. However, he will also suffer from negative externalities of other surplus cows. If the farmers cannot limit each other’s herds, they will all suffer. Over-grazing will cause more harm than the extra animals bring in benefits. The individual incentives direct actions in a way that causes an irrational solution for the group. A cost split with others will be dwarfed if combined with a personal undiluted benefit. Expressing this dilemma in another version, a personal cost will be avoided if the benefit will be shared and thereby diluted. These are not proper comparisons between costs and benefits, and an agent’s acting on such incitements will not bring a rational solution. The solution is to arrange a social order in which aggregated costs and benefits are compared in one calculus, and individualized costs and benefits in another. A failure to organize such an order will keep the misallocation going even if the negative outcome is understood. It is necessary to take the step from understanding to a change in social order.

The most pressing problem with collective action is not necessarily that of obtaining obedience; to succeed in getting unity in support of a project deemed to be in the group’s interest. Another type of problem that might be more severe is “runaway projects”. The goals of the groups are sustained by incitements, but the goals are irrational on a group level. Collective punishment is often cheap and when in place it is difficult to withdraw. Just a suggestion of modification or being soft on offenders of a rule might be seen as high treason leading to ostracism. Military discipline often has rules of punishing soldiers not taking part in punishing others. In the Roman army not only were soldiers that attempted to escape from the fighting line punished by death, but also negligence by the closest soldier of performing an execution was punished by a penalty of death (Dixit and Nalebuff, 1991). In addition to in-group pressure come out-group counter pressures

that not only generate fear, but also sustain loyalty. These conditions make the existence of the rule rather independent of $B > C$.

Leaders often have strong incentives to pursue an irrational group interest, because they are entangled in their own vows, holy promises, etc. For leaders of groups that seem locked in an endless struggle, there is often a personal more safe behavior to keep on fighting the holy war than to reach a settlement that might disrupt the loyalty of their own core group. The fate of Sadat and Rabin might illustrate the danger of other solutions than to keep on striving for total victory. One of the dangers for group rationality is the influence of leader interest. The dragging on of conflict might carry limited personal sacrifices for the leader personally while securing his position, but this does not eliminate the possibility of the aggregated benefit being $B \ll C$. This situation can be maintained because $p_a > c_a$ for the ordinary group member. In a hierarchical society a large influence for leader benefits will be a source to irrational social solutions.

In a negotiation, stubbornness will be an advantage if the opponent is less stubborn, but two stubborn parties will result in an unproductive standoff. Many group conflicts are pursued with the determination to out-stubborn the other party. Suggestions for a compromise to end the conflict might be condemned as defeatism and seriously undermine a pose of accepting nothing short of surrender from the other group. “We have nothing to fear, but fear itself”.

A general treatment is to be even more aggressive toward heretics than heathens. Christian bulls of communication and Muslim *fatwas* are for those that deviate from their proper line. The unionists talk about strikebreakers as scabs. Alliances and specific projects will change, but maintaining the group and avoiding fraction often becomes a permanent part of group strategy.

Ostracism as a way to maintain the group is often demonstrated as a force strong enough to prevail over external influences as well as personal preferences. Female mutilation, to mention an extreme example, can hardly qualify as group rational, but the social system can enforce the rule with such force that it can become individually rational to comply. Spending life as an outcast might be even worse than being crippled. Ostracism is a factor found also in western democracies, for example, whistleblowers in companies. Such individuals often receive lavish praise and most positive media treatment. However, in their prime group, they get a more negative appraisal and when media attention diminishes, the negative is what prevails. This result has such a strong impact that initially admired people often get severe depressions and even commit suicide (Velesquez, 1992, p. 402).

For a group living primate the rational policy might be to follow the devise “my country—right or wrong.” The wisdom, moral and effect of public policy are often difficult to appreciate. Even if later generally considered wrong, there is the highly excusable element that you have been wrong with lots of others. The record of being right when most were wrong implies some pretension of moral or intellectual superiority that can feel most irritating for the majority. A more popular analysis would be that “at that time this position seemed to be right to most sensible people, alas we were wrong”.

If there is a group rationality this might result in an advantage compared to other groups and group selection might accentuate this advantage. But the solution might also be maladaptive and if the pattern is hard to break there might be a negative group selection effect (Tullberg, 2003). Punishment provides the rationality to carry the cost, but when looking at the aggregated numbers, $B < C$ show that the behavior is irrational at the group level. Ideology that makes the interest of the group to something more than the aggregated interests of its individuals makes it harder to have an enlightened discussion about the group’s interest.

6. irr/RAT

All inhabitants of a town might appreciate a clean park more than a littered one. The benefit exceeds the cost, the condition for RAT. Still on some occasions, it is convenient to dispose of some trash by throwing it on the ground, and this marginal action will not noticeably diminish the common good. However, a lax attitude generally shared by others will, and so there exists a problem in succeeding to uphold the proper behavior. Even with such uncontroversial questions the introduction of strong incitements are controversial. Few societies have embarked on a policy as in Singapore where compliance is obtained by imposing, and effectuating, high fines. Most societies hope for a social behavior by agitating for social virtues despite getting mixed results.

For altruism the important thing is often the readiness to help others rather than to scrutinize the reasons for their pleas and to control that there is no inverse relation, that $c_a > b_o$. If so, not only is there a lack of individual rationality, but probably also social rationality. Still many people are attracted to such morality. Altruistic proponents often claim that $B > C$ even if $c_a > b_o$. The power of example might substantiate a more caring society and generate positive external effects.

The altruistic agitation entertains a dubious view upon instrumentality. Often there are promises of over-compensation such as eternal life for the followers of the script. Simultaneously it is asked that the prescribed deeds be made for their intrinsic value with little or no respect taken to the promised rewards.

On a personal level, many people do not want to be just prudent citizens avoiding prison and fines, and striving for social respect and money. Instead they claim to act socially for virtue's own sake. Most religions mirror and support such a split view. The good is to be done for its own sake and then consistently, but still surprisingly, the divinity itself and the religious system show up with rewards and punishment. Social deeds should be done in an altruistic intention and the reward officially unexpected.

Such social display is generally viewed as social virtue not only in religious circumstances. Most everybody in an organization knows that a person has worked long and hard to get a promotion, but still a comment of total surprise is seen as modest and proper. The rewarded should claim thankfulness and exhibit a disinterest in social prestige, even if this will not convince anyone. We should promote good deeds, do good deeds and be in favor of rewarding good deeds, but simultaneously pretend to be surprised that good deeds get rewarded. It might be considered cynical, but I think the intended message is something like the following: "semi social individuals might need rewards and punishments to keep in line, but more responsible and faithful individuals will do right from pure virtue. I am such a good person driven by virtue, not incitements."

Long traditions have maintained the assertion that the interest of the group requires a readiness for individual self-sacrifice. Incentives can be provided, but the voluntary sacrifice has been hailed even by very disparate creeds. Also totalitarians have a strange attraction to the volunteer, even when discarding free choices as liberal hogwash. The Third Reich and the Soviet Union had few obstacles from ordering people to work or reduce their net pay. They certainly ordered high demands, but instead of commanding all they wanted, emphasis was also placed on volunteer work and contributions. A possible explanation may be that the ideological target was to make a separate private sphere immoral. The group could claim all of an individual's resources. Accepting collectivist demands might be a device not only to fool others, but internalized it serves as protection from what Orwell termed "thought crimes" (Orwell, 1949).

6.1. *Overestimating irrational virtue*

Most societies understand the importance of supporting social norms, so there are generally some agitation that “crime does not pay”. Social virtues such as honesty and hard work often pay at least in the long run. Moralists normally argue that this is not sufficiently effective, indicating that they would provide a more effective motivation if given more influence. They claim to uphold norms without the restriction of making it rational to comply. To use the Christian metaphor, they can make people “walk the second mile”. In Roman times the Israelites had a duty to assist Roman soldiers with transportation for a mile. Jesus proposed his listeners to walk a second mile as a virtue, as a volunteer sacrifice, one that was not sanctioned by the authorities.

Irrationality on the individual level is often overstated. A factor confusing many observers is that even in states with a most effective police force, there are some successful criminals. The ordinary citizen can then point at the successful criminal and brag about his own virtue in abstaining from these criminal rewards. But abstaining from being a successful criminal is not a relevant choice. The choice is about *trying* to be a successful criminal or not committing a crime. The non-criminal might be less well off than the successful criminal, but he is better off than the criminal that is found guilty and sentenced. The central comparison is between the two criminal groups combined on one hand and the non-criminal citizen on the other. If crime, free-riding, etc. actually pay, then there is irrationality in following the law which would indicate a serious problem.

Political scientists have been interested in how collective goods problems are solved in societies where the state does not play a prominent role as judge and law-enforcer. It was commonly thought that volunteer agreement was of the category irr/RAT. A fishing community was not expected to have a system capable of punishing people that over-fished, so there was no force of $p_a > c_a$ counteracting $b_a > b_a - c_a$. However, research now reveals that more incitements exist than previously realized. Two important findings came out in a major study by Ostrom (1990) covering different geographic regions and resources threatened by overexploitation. There was always organized monitoring and there was always punishment of offenders. This result suggests that human behavior is not as irrational as often claimed.

6.2. *The rationality with preaching irrationality*

Egoism often needs no motivation for the egoist himself. The main target even for egoistic messages such as l’Oréal’s advertising campaign slogan “because I am worth it”, is to make other people think they should agree. The problem is not to motivate the ego to get, but to motivate the other to give. Some kind of obligation must be introduced. Free-riding is not the only egoistic way—another more advanced way is in being the broker of unselfish ethics. No other morality than altruism gives the egoist such good arguments for an uneven exchange “You give and I get”. To preach altruism to others seem to be the logically most promising strategy for a manipulative egoist. An often-rewarding advice in detective stories is to follow the money. Undeniably many religions and their priesthoods have become rich by preaching the virtue of being poor and unselfish.

To the agent, being an egoist, b_a is primary and c_o is secondary.

To the other, being an altruist, b_a is primary and c_o is secondary.

If the agent can convince the other to be altruistic, he can obtain an unbalanced interaction to his own favor. Don’t the two different views fit like hand in glove? When a thief takes a cloak this is egoism, but if he then is given the coat this act is altruism (Matt 5:40). Is this not describing and

promoting the same transfer from two different angels; two sides of the same coin? An illustrative point is that even the evangelists got confused in the separation. St Luke (6:29) describing the same event uses the transfer of the coat as the egoistic act and the cloak as the act of altruism. There is a difference in action of a thief stealing and an altruist giving, but there are many hybrids of threat, promises, fraud, manipulation and guilt. The similarity in result deserves attention.

6.3. Kin selection

A different way to link rationality at the social level with irrationality at the personal level is the mechanism of kin selection (Hamilton, 1964). An egoist will only carry out an action if the benefit exceeds the cost, that is $b_a > c_a$. Kin selection can modify this restriction and motivate a behavior even if this condition is not met. If some kin (k) with the genetic degree of relatedness (r) to the agent gain the advantage b_k , the modified restriction is $b_a + b_k \times r > c_a$.

A critique of nepotistic norms is that they might promote some sacrifices that are not socially optimal, but just nepotistically rational. An action can be quantified as follows: b_a is 0, c_a is 3, r is 0.5 and b_k is 2. It is rational for the kin to promote such an action by the agent, because his benefit is 2 whereas the costs are just 3×0.5 . Socially it is not rational and even less so from the agent's perspective because the cost is 3, but the benefit reduced to 2×0.5 . If kin rationality is reflective rather than standardized in rules which puts one part at a disadvantage, the outcome is likely to be socially positive. The crucial question is whether it is rational for the doer, the agent, not the receiver, the kin. Kin selection motivates some actions, which are not in the interest of the agent's organism, but in the interest of his genes. Such "irrationality" at the individual level is therefore not unstable, but supported by preferences generated by natural selection.

The social surplus $b_a + b_k - c_a$ is larger than the calculation from the doer's perspective presented above, since the reduction of benefits from b_k to $b_k \times r$ is eliminated. Considering these possibilities it seems likely that quality control is too harsh rather than too soft. A possible critique is that kin selection contains too demanding restrictions and does not provide sufficient social interaction. However, the strict demands make it likely that the actions taking place motivated by kin selection will be socially beneficial (RAT).

7. irr/IRR

This category is one that is seldom promoted, but often used to critically describe the behavior of certain individuals and groups. Because of irrationality on both individual and social levels, the performers of this strange behavior have every reason to change. Since few argue in favor of such irrationality, it seems motivated to discuss whether this behavior is real or just a propagandistic attack upon a mistake purposely done by very few. One explanation for a deliberate irrationality in these two senses is that the purpose to strive for is perceived to be on an even higher level. Not only the individual but also the group is supposed to sacrifice for this greater good. Heroism, self-sacrifice and the likes, can be promoted for a higher rationality such as serving the wishes and commands of God.

One illustration of an irrational stubborn attitude is the fate of Ashkenazi Jews. During the First Crusade 1096 in Germany they were given the choice of converting and being set free or being executed (Chazan, 1987, p. 245). This convert alternative was not a sham meant to humiliate people that were doomed to be killed anyway, but a real choice between life and death. The Jewish believers demonstrated the strength in their beliefs by putting their necks below the axe

and were thus willing to see their children die in the true faith rather than survive as heretics. Both as individuals and as a group it is apparently irrational to die in the name of God or some other symbolic action of little material significance.

As a long-term strategy such a rigid affiliation to group rules might do well (MacDonald, 1998). Between pogroms that hurt both on individual and group levels, there have been periods when tight group affiliations in the Jewish in-group gave significant competitive advantages compared to individuals belonging to less restrictive majority groups (MacDonald, 1998).

However, it is possible to suspect irrationality also in the long-term. Religious creeds that strive to convert members of out-groups can compensate biological disadvantages with recruiting advantages. Creeds bent on celibacy and suicide have not gone extinct.

The sect “Peoples Temple” committed a mass suicide of 900 people in Guyana in 1978. One might speculate as to what motivated their leader Jim Jones. There seems to be some megalomaniac attraction in convincing one’s followers to make the ultimate sacrifice for no apparent constructive purpose. Political leaders previously preaching their personal commitment to the cause and interest of the group seem ready to make a turnaround when at a loss. The group itself can be sacrificed as a vindication of the leader’s ideas. At the end of the war Hitler had the conviction that the defeat of the Third Reich should also mean an end to the German people. Castro preaches a theme about “socialism or death” for the Cubans. A totalitarian society gives extra possibilities for such self-centered and stubborn irrationality.

A less ideological and more social example might illustrate how “memes” (Dawkins, 1976) can prosper, even when genes suffer. Until recent times cities have been ecological sinks where death rates exceed birth rates. Conversion of country folk generated not only a city’s growth, but migrants also helped compensate for a consistent reproduction in the city that was below replacement levels (Söderberg et al., 1991).

It seems likely that strong group affiliations rational on the group level for small human groups in the ancestral environment might have evolved an emotional underpinning used also by new groups for which the advantages are insufficient to cover the significant costs. The discussion above focus on disadvantages for the in-groups themselves. More destructive effects occur when not only the in-group is hurt, but some out-group as well. The destructive potential of collective action is most significant (Olson, 1971).

8. Conclusions

For people already performing some behavior with the cost c , it makes sense in trying to make it a civic duty and to gain some points, versus the non-performers for whom performing c is revolting or at least a nuisance. Conformity easily becomes a virtue because lots of people are conformist and have an interest to make their behavior into something of significance.

What constitutes a collective benefit is often analytically unclear, even if there is a strong common belief that the conformity of the here and now is of a superior quality than the conformity of other times and places. Often it is hard to see the possibility of a qualitative step in expanding some behavior from most people to all people. Sometimes it seems that the majority is unsure to such a degree, that they are unwilling to allow others to make another choice simply because that other choice might be a better one.

To me it seems that man is very far from being an ultra social being asking what is the common good. Rather his social affection seems to be a forced necessity of life. Other people irritate and threaten us, and one way to cope with these issues is to unite and fight. This frequently functions as a good tactic both in defense and offence. To support loyalty is the prime issue and it creates

the precondition for successful collective action. The central premise for loyalty is to maintain the rational motivation for others to comply (rat). In contrast, a worthy goal for the common project is in practice often a secondary goal. Traditions prevail or new leadership choices are made of projects expected to sustain a group spirit of loyalty to the leadership. The benefit for the group is most often not analyzed in a qualified way.

Utilitarianism has the condition $c_a < b_o$. Without this social gain of the action there is no virtue in helping others, but if many actions with social gain are done the aggregated result is likely to be $B > C$. But there is no rational argument on the micro-level for being a utilitarian. Why should an agent carry the cost c_a without a personal connection to the benefit b_o ?

Many comparisons basically confront egoism as $b_a < c_o$ versus altruism as $c_a < b_o$. Then the conclusion is that socially rational altruism is better than socially irrational egoism. Maybe they really just say that the social rational is socially better than the social irrational, a thesis to which it is easy to agree. But how to make the choice between the other two possibilities – social rational egoism $b_a > c_o$ versus social irrational altruism $c_a > b_o$? I suggest the correct conclusion to be that, in the long-term, all of these four unbalanced relations will be socially destructive as models for human behavior.

Psychologists and philosophers often focus on intention and motivation. Economists and biologists will often discard talk like “intend self-less action, get rewards as a surprise” and point at the primacy of the statistical expected value in the real world.

8.1. *Human behavior in different sections*

I think that the behavior in Fig. 1 of Sections 4 and 5 is significant, but small in Sections 6 and 7. In agitating against ideas we disagree with, it is easy to overstate the influence of irrationality.

It is hard to judge overall rationality so it seems that individual adjustment to the dominant opinion is the rule. However, this is not so much flexibility and rational optimizing as it is obedience to political correct conformity. Life is not so much flexible tactics, as it is choosing a rather good strategy and then sticking to it rather rigidly. It seems that increased social behavior in Section 4 has the positive potential. The instrumentality of being “rat” provides the internal stability, and “RAT” might provide the basis for a comparative advantage contra other groups. Furthermore, such an advantage might inspire imitation.

Reaching a social rationality by downgrading rationality on the individual level attracts many authors and preachers of different ideologies. The incitement of social pressure is used but this tactics is denied, ignored, or delegated to superior beings. Such moralistic agitation hails irrationality on a personal level as the key to rationality at an aggregated level. According to this view, selfishness brings disaster and self-sacrifice brings advantages, so the goal of this agitation is to convince people to make sacrifices and these sacrifices will bring benefits to others. However, this agitation imposes restrictions and people will be forced to obey if avoiding to be punished. Appearance of unselfishness should rather be seen as opportunism to social pressure than as useful efforts to find solutions to problems at a group level.

For reformist ambitions, it seems that rationality on the group level has to get more attention. The social benefits of these projects are often dubious. Very general and lofty principles on a universal level with a very loose connection to real human behavior are poor substitutes. Also in our more secularized time, there is a dominance of a romantic view upon concepts such as altruism and solidarity. At the aggregated level rationality is seen as a natural consequence of good intentions and willingness of sacrifices. Our society and our times are often criticized for

being too focused on rationality. In the field of norms for social behavior, rationality is hardly getting enough attention.

According to this analysis rationality at the individual level is an underrated factor. A suggested social project often implies serious costs to these individuals and therefore, group benefits need to be both substantial and realistic. Different projects are presented as irr/RAT solutions but it is often more justified to classify them as rat/IRR solutions. Social pressure makes it rational for the individual to comply with conformist behavior. The goal is praised, but the social benefit is most disputable if examined by a rational analysis.

Acknowledgement

I thank Bohdan Sklepkovych for improving the text.

References

- Allot, R., 1999. Religion and science—sex and society: forms and processes of cohesion. In: Thienpont, K., Cliquet, R. (Eds.), *In-Group/Out-Group Behavior in Modern Societies*. NIDI CBGS, Brussels.
- Batson, C.D., 1991. *The Altruism Question: Toward a Social-Psychological Answer*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Brown, R., 1986. *Social Psychology*, second ed. The Free Press, New York.
- Chazan, R., 1987. *European Jewry and the First Crusade*. University of California Press, Berkeley.
- Dawes, R., 2001. *Everyday Irrationality: How Pseudo-Scientists, Lunatics and the Rest of us Systematically Fail to Think Rationally*. Westview Press, Boulder, Colo.
- Dawkins, R., 1976. *The Selfish Gene*. Oxford University Press, Oxford.
- Dixit, A., Nalebuff, B., 1991. *Thinking Strategically—The Competitive Edge in Business, Politics and Everyday life*. W.W. Norton & Company, New York.
- Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. *Nature* 415, 10 January.
- Galbraith, J.K., 1952/1993. *American Capitalism*. Transaction Publishers, New Brunswick, NJ, USA.
- Hamilton, W.D., 1964. The genetical evolution of social behaviour. *Journal of Theoretical Biology* 7, 1–52.
- Hardin, G., 1968. The tragedy of the commons. *Science* 162, 1243–1248.
- Hardin, G., 1977. *The limits of altruism*. Bloomington.
- Harrison, G., 1992. Theory and misbehavior of first-price auctions: reply. *The American Economic Review* 82, 5.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* (2), 263–291.
- MacDonald, K., 1998. Indoctrination and group evolutionary strategies: the case of judaism. In: Eibl-Eibesfeldt, I., Salter, F. (Eds.), *Ethnic Conflict and Indoctrination*. Berghahn Books, New York.
- Mill, J.S., 1836/1988. On the definition and method on political economy, in: Haus, D.M. (Ed.), *The Philosophy of Economy. An Anthology*. Cambridge University Press, Cambridge, pp. 52–69.
- Nietzsche, F., 1886/1983. *Beyond Good and Evil*. Penguin Books, Harmondsworth.
- Rawls, J., 1971. *A Theory of Justice*. Oxford University Press, Oxford.
- Olson, M., 1971. *The Logic of Collective Action—Public Goods and the Theory of Groups*. Schocken Books, New York.
- Orwell, G., 1949/1983. 1984. *Atlantis*, Stockholm.
- Ostrom, E., 1990. *Governing the Commons—The Evolution of institutions for Collective Actions*. Cambridge University Press.
- Scharfstein, D.S., Stein, J.C., 1990. Herd behavior and investment. *American Economic Review* 80 (3), 465–479.
- Scharfstein, D.S., Stein, J.C., 2000. Herd behavior and investment: reply. *American Economic Review* 90 (3), 705–706.
- Sherif, M.H., White, H., Sherif, C.W., 1961. *Intergroup Conflict and Cooperation: The Robin Cave Experiment*. University of Oklahoma Book Exchange, Norman.
- Silverman, I., Case, D., 1998. Ethnocentrism vs. pragmatism in the conduct of human affairs. In: Eibl-Eibesfeldt, I., Salter, F. (Eds.), *Ethnic Conflict and Indoctrination*. Berghahn Books, New York.
- Simon, H., 1955. A behavioral model of rational choice. *Quarterly Journal of Economics* 69, 99–118.
- Smith, A., 1776/1981. *An Inquiry into the Nature and Causes of the Wealth of Nations* Liberty Classics, Indianapolis.

- Söderberg, J., Jonsson, U., Persson, C., 1991. *A Stagnating Metropolis—The economy and Demography of Stockholm 1750–1850*. Cambridge University Press, Cambridge.
- Tajfel, H., Billig, Bundy, R., Flament, C., 1971. Social categorization and intergroup behavior. *European Journal of Social Psychology* 1, 149–178.
- Tullberg, J., 2003. Rationality and social behavior. *Journal of Theoretical Biology* 224, 469–478.
- Velesquez, M., 1992. *Business Ethics—Concepts and Cases*. Prentice Hall, New Jersey.