

Multilevel selection and human altruism

Alejandro Rosas

Received: 11 April 2007 / Accepted: 13 August 2007 / Published online: 1 November 2007
© Springer Science+Business Media B.V. 2007

Abstract Views on the evolution of altruism based upon multilevel selection on structured populations pay little attention to the difference between fortuitous and deliberate processes leading to assortative grouping. Altruism may evolve when assortative grouping is fortuitously produced by forces external to the organism. But when it is deliberately produced by the same proximate mechanism that controls altruistic responses, as in humans, exploitation of altruists by selfish individuals is unlikely and altruism evolves as an individually advantageous trait. Groups formed with altruists of this sort are special, because they are not affected by subversion from within. A synergistic process where altruism is selected both at the individual and at the group level can take place.

Keywords Altruism · Assortative grouping · Group selection · Free riding · Individual selection

Biological altruism: individual and group selection

In the last decades, the mainstream tradition in contemporary evolutionary biology has interpreted biologically altruistic traits as apparent, either as kin selected, reciprocally selected or sexually selected (Hamilton 1964; Williams 1966; Trivers 1971; Zahavi 1975). For this project, reconciling altruism to natural selection is equivalent to “taking the altruism out of altruism” (Trivers 1971). This enterprise has been under permanent criticism from multilevel selection theory, with a climax in Sober and Wilson’s (1998) discussion and history of the controversy. Multilevel selection theory claims that selection operates simultaneously on genes, organisms and groups of organisms. Just as individual organisms compete against each other within a group, groups of individuals compete against each other in the population. Concerning altruism, competition at the different levels exerts its selective force in opposite directions: a donation of resources to a

A. Rosas (✉)
Departamento de Filosofía, Universidad Nacional de Colombia, Ciudad Universitaria, Bogotá,
Colombia
e-mail: arosasl@unal.edu.co

neighboring individual or individuals that decreases the fitness of the donor will be favored by competition between groups, but disfavored by competition between individuals within a group. Altruistic traits will evolve when competition between groups is stronger. Some researchers believe that this process was significant in human evolution (Boehm 1997; Sober and Wilson 1998; Gintis et al. 2003).

The essential feature of this process is population structure. Against the mainstream approach, Sober & Wilson argued that all models for the evolution of altruism rely on population structure and therefore implicitly invoke the logic of group selection (1998, p. 57 and *passim*). In kin selection, for example, altruism evolves through differential reproduction of groups of siblings. This remarkable claim motivated a counter-proposal: selection for altruism can be read indifferently as operating on groups or on individuals; there being no fact of the matter as to which description is right (Sterelny 1996). Different mathematical models, with or without values for the fitness of groups, will predict the same outcomes (Kerr and Godfrey-Smith 2002). However, one may doubt whether mathematical models do capture the causal processes involved; it is in this domain that the distinction between group and individual selection is relevant (Okasha 2003).

Mainstream evolutionary biologists understand the distinction between individual and group selection processes for altruism as real or objective. Maynard Smith dedicated a couple of early papers to clarify this distinction (Maynard Smith 1964, 1976). Although he was skeptical over the empirical relevance of group selection, he clearly thought that it was objectively different from individual selection. This is not only a difference to instrumentalists, but also a difference in regard to Sober & Wilson, for whom only group selection models for altruism carve the world at its joints, whereas individual selection models are only ways of describing that do not reflect the real causal processes. The state of the art in this debate includes therefore three positions: (1) Instrumentalists hold that the distinction between group and individual selection processes is only a matter of alternative descriptions, not of alternative causal structures; (2) Multilevel selection theorists are realists in regard to group selection, but instrumentalists in regard to individual selection models for altruism; (3) Mainstream evolutionary biologists are realists both in regard to group and to individual selection. Though they doubt that nature in fact instantiates group selection for altruism, they consider it as objectively different from individual selection.

In this paper, I explore an avenue for a consensus between multilevel selection theory and mainstream evolutionary theory. I propose a distinction between individual and group selection for altruism that could be acceptable for both parties. The argument assumes a basic contention of multilevel selection, namely, that assortative grouping is causally operative whenever strong altruism (hereafter: altruism) evolves (Wilson 1979). Though some authors argue that strong altruism can evolve in randomly formed multigenerational groups, they claim that this happens only because positive assortment can arise stochastically within the multigenerational groups (Fletcher and Zwick 2004). In the end, therefore, positive assortment, whatever the causes, is present whenever strong altruism evolves. Taking this into account, I refer to ‘assortative grouping’ as the degree to which the benefits of altruists fall to other altruists, *independently of the particular way in which positive assortment arises*. Assortative grouping does not automatically legitimate talk of group selection. Rather, if we pay attention to the fact that assortative grouping arises in some cases stochastically, in others non-stochastically, this will disclose a relevant criterion for discriminating between selection processes, namely, a causal dimension that indicates whether the altruistic organism either exerts or lacks control of assortative grouping. Whether altruism is selected due to stochastic fluctuations or due to control by the altruistic organism is relevant.

Although Sober & Wilson have stated, quoting from Hamilton (1975), that the causes of assortative grouping are diverse and that they do not affect the thesis that selection for altruism involves group selection (1998, 77; 134), they do not spell out the reason for this view. They offer only the following statement as justification: whatever the causes responsible, population structure “can be modeled without a single change in the mathematical equations” (1998, 94). However, it seems reasonable to demand a place for the causes of population structure in a causally relevant discussion of group vs. individual selection. It preserves one of the concerns of multilevel selection theory namely, that selective explanations should capture the *causal* processes.

In the following section, I argue that group selection is involved when stochastic processes are responsible for assortative grouping. In the section “Non-stochastic processes leading to assortative grouping”, I present both kin altruism and human altruism between non-kin as depending on non-stochastic assortative grouping. However, only in human altruism does the altruist exert a significant control over positive assortment such that individual selection is there involved. In the section “Groups of altruists without subversion from within?”, I discuss objections against the claim that human altruism evolves through individual selection. The objections question the degree of control that human altruists exert over assortative grouping. They indirectly confirm the thesis that individual selection of altruism depends on that control. Moreover, their theoretical plausibility dissolves to some extent when one descends from abstract models into the social and psychological complexities of human action. I conclude with a brief note on the importance of synergistic selection processes in the evolution of human altruism and on the reasons for labeling traits “altruistic” even when they benefit the individual.

Stochastic processes sustaining population structure

Assortative grouping is sustained by stochastic processes in one class of cases. These can be illustrated by Maynard-Smith’s “haystack model” for group selection (Maynard-Smith 1964, 1976). The model was meant to explain the evolution of prudent restraint or “timidity” in the breeding behavior of a hypothetical species of mice. When population density is high and threatens depletion of resources, prudent restraint may be required for survival. But individual selection can hardly explain the spread of genes for timidity, for such genes lose in the individual competition against alternative genes for unrestrained breeding behavior.

The hypothetical mice live in isolated, multigenerational groups. Group structure is environmentally facilitated by food supply distributed in spatially discontinuous patches (haystacks) that last for a few generations; each haystack feeding a different group. Groups compete by exhibiting different probabilities of going extinct and of founding new colonies. At the moment of dispersal to colonize new haystacks, groups are entirely composed either of selfish or prudent individuals. In multigenerational groups, an allele coding for prudence will quickly go extinct in competition with a rival allele coding for unrestrained breeding behavior. For this reason, gene flow between colonies works against prudence. Limiting gene flow to free interbreeding after abandoning exhausted haystacks for fresh ones, Maynard Smith argued that random mating at this stage would undermine the group selection of prudence. If the initial frequency of prudent colonies is low, the establishment in the population of a gene for prudence or altruism would require some form of genetic drift, e.g., “sampling errors” in mating after dispersal. In an entirely fortuitous way, the

number of matings between altruists leading to the foundation of prudent colonies would have to exceed the expected probability.

Once the number of prudent colonies has climbed stochastically over a critical threshold, the selective advantage of prudent colonies—the fact that they flourish and contribute more migrating mice than selfish ones—will gain momentum. Altruism will then be maintained in the population even without genetic drift. Thus, group selection would be able to *maintain* the frequency of altruism once the trait is established. This is especially the case, as Maynard-Smith emphasized in 1976, when the lack of prudence in managing limited food patches causes the rapid extinction of selfish groups before new food patches are available. Drift is required only in the origination of the first altruistic groups and in the process leading to the establishment of the altruistic gene above a critical threshold (see Fig. 1). At that stage it is surely indispensable (Maynard Smith 1976, 282).

This model applies for traits that vary discretely. When traits vary continuously in the population, which plausibly applies to timidity or aggressiveness in breeding behavior, a similar model applies. If a large population with a trait varying continuously from altruism to egoism is divided randomly in groups, these will probably vary in respect to that trait. Variation between groups will be greater when sampling errors occur in the division; and the greater this variation, the greater the group selection effect that will pull the trait towards the group optimum, as confirmed in computer simulations (Wilson and Dugatkin 1997). The altruistic value will endure if a similar population structure is periodically renewed. But remarkably, between this model and the haystack model a functional similarity is apparent. In both models, variation between groups is accidentally produced: “some groups will be more altruistic than others *by chance*” (Sober and Wilson 1998, 137, *my italics*). Both here and in Maynard Smith’s model, stochastic processes are responsible for variation between groups. Group selection can take place when stochastic processes lead to the required group structure (Sober and Wilson 1998, 342, note 4).

Empirical confirmation of this theoretical possibility comes from the role of stochastic processes in the evolution of decreased virulence, where hosts provide a patchy food distribution similar to haystacks. In diseases where different virus strains compete and hosts must be alive and moving around for transmission, there is a trade-off between transmission and virulence (Frank 1996). When new hosts are infected mainly or only with tokens of the less virulent type of an original mixed population, hosts live longer or move more than hosts infected with the more virulent type, send out more propagules and spread decreased virulence. Given mixed groups of viruses within hosts and the decline of the less virulent type within them, the founding of a less virulent colony in new host is a stochastic event, a ‘sampling error’ in transmission where no special cause

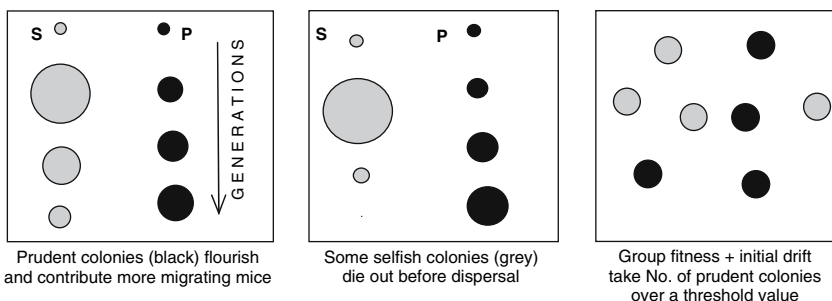


Fig. 1 John Maynard Smith’s haystack model

is responsible. When altruism is thus favored, it is reasonable to claim that assortative grouping functions as a salient factor, and that explanation ends with it. However, where non-stochastic factors are involved in the establishment of altruistic groups, explanation should point them out.

Non-stochastic processes leading to assortative grouping

When assortative grouping comes about non-stochastically, two types of cases must be distinguished. In one type, altruistic responses are regulated by proximate systems that discriminate defectors from cooperators and regulate behaviors designed to produce assortative grouping. In the other type, causal factors that are independent of the production of altruistic behavior guarantee assortative grouping, as in kin selection. Altruism as we find it in humans is a clear instance of the first case. Before arguing for this claim, it should be noted that an altruistic behavior linked by design to traits that bring about assortative grouping is able to control, by this fact, its own evolution. This process differs significantly from the stochastic process discussed above. When altruistic traits are promoted by stochastic assortative grouping, the forces that create groups as interacting units are independent of the altruistic behaviors whereby some groups out-compete other groups. But when a proximate mechanism is designed to condition altruistic responses to the presence of assortative grouping, and even to influence the production of such group structures, assortative grouping is not a fortuitous event upon which altruists stumble. A proximate mechanism in the altruistic individual is responsible for both the altruistic response and for the presence of environmental conditions that enhance the relative fitness of the altruist.

Human altruism between non-relatives is linked to detection of the appropriate environment and also to behavioral dispositions designed to bring it about. Altruistic behavior is sensitive to reputations and to other signs of altruism (Frank 1988). Moreover, it is linked to (1) a preference for altruistic partners; (2) a tendency to form long-lasting partnerships with other altruists; (3) differential generosity; (4) normative guidance and monitoring of compliance; and (5) moralistic aggression or punishment of defectors (Trivers 1971). Since Trivers' classic paper, developments in evolutionary game theory have presented these features as belonging to new strategies. But some authors recognize their presence in Trivers' original proposal (Nowak and Sigmund 1998; Sherrat and Roberts 1998; Boyd 2006). These features are consensually recognized as proper elements in human altruism and they lead directly to assortative grouping of altruists with altruists. Population structure is therefore a deliberate effect of the proximate system controlling human altruism, an effect it is designed to produce. Selfish individuals that have infected the group are discriminated against. They are either punished until they change their ways; or are excluded from interaction with altruists by some form of isolation or expulsion (Boehm 1999). By exercising control of assortative grouping, altruism enjoys a relative individual advantage over selfishness. Human altruism benefits the group without suffering from subversion from within.

The difference between cases of the type of decreased virulence and cases like human altruism occurs along a casual dimension. It depends on whether the altruistic individual is able, or not, to control assortative grouping as a necessary condition for the evolution of strong altruism, where the altruist suffers an absolute cost in the presence of free-riders. At one extreme, the causes that control assortative grouping are independent of the mechanism that controls altruistic behavior. Altruism evolves in virtue of environments such as

exhaustible food patches, which create probabilistic conditions for the survival and re-establishment of altruistic groups. At the other extreme, altruistic behavior is an adaptive response regulated by a system that both detects and produces assortative grouping. In this case, altruism is manifested only when it is individually adaptive. Therefore, individual selection is involved in its evolution.

But along this causal dimension, degrees between both extremes seem possible, depending on how well the system detects and produces the relevant environment. Kin selection seems to occupy an intermediate position. In kin selection, assortative grouping comes about non-stochastically, namely by the laws of inheritance in species with low dispersal rates. In viscous populations, individuals remain in the surroundings of their birthplace and have for this reason a high likelihood of interacting with relatives. The breeding habits of some species contribute to assortative grouping by spatially isolating the brood (birds in a nest). However, as long as kin selection is not linked to special abilities for kin recognition, the altruistic organism cannot count on mechanisms specially designed to detect and produce assortative grouping. It still shares a feature in common with assortative grouping through stochastic processes: In both cases, altruists cannot stop infection of their groups by selfish individuals nor the negative consequences on their fitness. This determines a selection process where altruism suffers from subversion from within, because altruists lose in competition with egoists within the group.

Organisms that condition the altruistic response to the operation of kin detectors make a significant move towards control of assortative grouping. Discrimination can rest on simple cues, e.g. chemical signals, and cooperative or aggressive behaviors are conditional on that discrimination. But since kin are only a proxy for altruists, kin recognition does not go far enough in engineering positive assortment. Among the aspects of human altruism mentioned above, there are two that contribute to enhance control. In humans, the altruistic response is linked to discrimination of altruists as such: altruism is conditional on the assessment of reputations, independently of genetic relatedness. The ability to signal altruistic intent and to respond to those same signals in others leads to assortative grouping (Trivers 1971; Alexander 1987; Frank 1988; Zahavi 1995; Wilson and Dugatkin 1997; Nowak and Sigmund 1998; Sherrat and Roberts 1998). Moreover, human altruists take additional measures to maintain altruistic groups free from infection through egoists. In particular, we guide our behavior by fairness norms which also motivate the punishment of deviators (Fehr and Fischbacher 2004). Both abilities are typical for human altruism. But it is still an open question how far human altruism goes in the direction of control. In the next section I review some reasons for thinking it does not go far enough.

Groups of altruists without subversion from within?

In the previous section I argued that individual selection is involved in the evolution of human altruism. But as is well known, this thesis is denied by multilevel selection theorists, for whom human altruism is a conspicuous case of how altruism evolves through group selection (Sober and Wilson 1998). In this section I will review their arguments. I do not intend to solve this disagreement, but rather show that their objections implicitly confirm a weaker conditional thesis:

(CT) If it turns out that humans control assortative grouping, individual selection is involved in the evolution of human altruism.

Its proof only requires showing that the objections against the individual selection of altruism attack the claim that humans control assortative grouping and defeat subversion from within. Both the objections and the claim they attack concern an empirical issue, to be resolved by looking at the relevant evidence. This implies that the distinction between explaining human altruism through individual selection and explaining it through group selection as overriding individual selection, is not a matter of how you choose to describe processes, but depends on an empirical fact: whether altruists exert control of positive assortment.

Let us now review the objections. Starting with the ability to discriminate between altruists and selfish types, perfect discrimination based on powerful cognitive abilities would lead to perfect assortment. But, the objection goes, selection for deceptive abilities has, very likely, taken place (Trivers 1971; Frank 1988; Sober and Wilson 1998). However, cheating must surface in behavior at some point and affect reputation (Frank 1988). In small human groups, information flows quickly and cheaters are publicly exposed. Deception becomes potentially more damaging in large societies; but humans usually confine their interaction to a reduced circle and build there long-lasting partnerships. All in all, discrimination of selfish cheaters is likely to go a long way towards securing a higher fitness for altruists within groups. The first objection is therefore not conclusive: it only suggests that the attempt to engineer assortative grouping through discrimination may fail. But if it succeeded, then altruism would be individually adaptive.

The second objection goes deeper. Cooperation in n -person prisoner's dilemmas cannot be enforced by reciprocity. As is shown both in theoretical models (Boyd and Richerson 1988, 1992) and in lab experiments with human subjects playing repeated public goods games (Fehr and Gächter 2000, 2002), it takes only few defectors to propagate defection and unravel cooperation in the absence of punishment. Punishment, however, enforces cooperation. It is a device designed to stop infection of the group by free-riders, improve the engineering of assortative grouping and thereby increase the relative fitness of altruists within groups. It raises, nonetheless, the puzzle of how such a costly behavior can evolve in the face of cooperators who avoid being punishers (second-order free-riders). Some multilevel selection theorists argue that punishment is unavoidably altruistic because it can always be undermined by higher-order free-riding (Sober and Wilson 1998, 345, note 3). The whole package sustaining human altruism cannot thus enjoy higher relative individual fitness and can evolve *only* through a net force of group selection overriding individual selection.

Some models claim to have solved the evolution of altruistic punishment without group selection. This applies to a model where the participation in a joint effort for the production of a public good is voluntary (Fowler 2005, Hauert et al. 2007). Assuming that the good produced is neither created nor consumed by non-participants and that a defector that joins the collective action to exploit non-punishing cooperators suffers an opportunity cost (Boyd and Mathew 2007), the payoff to non-participants will exceed the payoff to defectors that join in a frustrated collective effort. Non-participants will then invade a population of defectors. When the former dominate, cooperators will invade; thereafter defectors will invade cooperators starting the cycle of oscillations all over again. In these conditions, punishers can invade quickly due to stochastic fluctuations when defectors are absent. Once punishers dominate, defectors cannot invade and the advantage of non-punishing cooperators cannot make itself felt.

In our terminology, this model describes a group selection process due to the role of stochastic positive assortment. The difference between defectors and non-participants just makes the take-over by punishers through stochastic processes more probable. But

punishers remain vulnerable. It just happens that favorable conditions create a net force of group selection overriding individual selection against them. Moreover, when collective goods inevitably benefit everyone, both non-participants and defectors can tend their solitary ends while benefiting from the public good. The difference between them disappears. We return then to the same old public goods problem, where stochastic fluctuations only rarely favor selection for altruistic punishment (Boyd and Mathew 2007).

Taking altruism out of altruistic punishment

The second objection claims that altruistic punishment cannot engineer assortative grouping in the case of the n -person provision of public goods. Because punishment is vulnerable to free-riding at higher orders, it remains dependent on stochastic assortative grouping for its evolution. But strictly speaking, this objection is plausible when punishers and cooperators are thought of as strategies in theoretical models that abstract from the complexities of social institutions and psychological mechanisms. We should look closer at social human organization and motivational structure to judge whether higher-order free-riding is empirically plausible. In the negative case, humans would have control over assortative grouping and altruism would count as individually adaptive.

Consider a society with centralized political authority and taxation. The function of taxation is to simplify the citizens' contribution to public goods of all sorts, by linking their contributions to a public system based on the division of labor. Taxes provide not only for first-order public goods, such as public services and their infrastructure. It also provides for second-order public goods, such as a centralized punishing system, a policing bureaucracy that is paid with the taxes levied from citizens. Moreover, the punishing system includes monitoring and punishing those citizens that don't pay taxes. Therefore, citizens that pay their taxes not only contribute to all first-order public goods provisioned in their society; they also contribute to the second-order public good of monitoring and punishing those that do not contribute to the first-order public goods.

With this arrangement, it is impossible to higher-order free-ride without free-riding on paying taxes. But if paying taxes is first-order contribution, nobody can be a higher-order free-rider without already being a first-order free-rider. In societies with central political authority and taxation, higher-order cooperation has been turned into first-order cooperation through taxes. Punishers that monitor and deter cheating on tax payments are not making an altruistic contribution which could be free-ridden upon; punishers are paid with the same taxes the non-contribution of which it is their function to punish. The possibility of higher-order free-riding is not open and cannot undermine the system.

Consider now small-scale egalitarian societies without taxation, which have their own version of the public goods problem (Boehm 1999), namely in the provision of group defense in war or big game hunting. The temptation to free-ride is the same. Punishment of free-riders usually consists in public exposure and isolation, and ultimately in ostracism; in some cases it may go as far as the execution of offenders (Boehm 1999). Public exposure of cheaters on public goods is effective, because losing one's reputation means losing the ability to attract partners in dyadic interactions, and many benefits of cooperation in small groups run through these interactions. As shown by economic experiments and theoretical models, the need to uphold a good reputation in dyadic interactions enforces cooperation in the provision of public goods (Milinski et al. 2002; Panchanathan and Boyd 2004). Ostracism or expulsion from the community is the extreme social sanction linked to the

loss of reputation. It may be as severe as death, given that in many environments loners are not able to survive.

Do some group members consistently avoid the costs of punishing offenders? Though perhaps nobody, except perhaps for kin, will cooperate dyadically with those that have lost their reputation in a small group, it happens that exclusion from dyadic cooperation is inefficient when the offender resorts to bullying behavior. Anthropological or experimental data offer little help on the question whether some group members consistently free-ride on direct punishment of bullies. When events are serious enough to call for execution, some data suggest that the whole group participates in the decision, in particular close kin of the offender, which in some cases even carry out the group's decision (Boehm 1999). This is interesting, because it seems to be a clever device to eliminate the costs for punishers coming from acts of revenge by kin of the punished. But given the scarcity of data, I shall indulge here in an indirect argument, by asking whether human psychological architecture supports free-riding at higher orders while consistently complying at lower ones.

A realistic psychological construal of a higher-order free-rider is a character who egoistically calculates that the best strategy is to avoid punishment by cooperating, while at the same time letting others do all or most of the punishment to avoid the costs of punishing. This character bears a similarity to Hume's sensible knave, who is also a selective free-rider. Hume's sensible knave judges honesty to be a good general policy. But it is subject to many exceptions, so that the (egoistically) wisest man is he who observes the general rule and free-rides on all the exceptions (Hume 1902, §232). Let us picture a general moral rule that prescribes upholding the altruistic system, including the punishment of anyone who threatens its stability. The higher-order free-rider strategy K, standing on a psychologically egoistic perspective, treats punishment at some level (which K may decide according to the occasion) as an exception to this otherwise 'egoistically' good rule. I contend that K is not psychologically plausible, for reasons noted long ago by Hume, and recently by economist Robert Frank (1988).

Frank developed an evolutionary view about the implausibility of K, which is psychologically similar to the view held by Hume (Hume 1902, §233; Frank 1988). The human motivational mechanism includes two systems that operate independently and probably evolved independently as well. The evolutionary older mechanism is the reward mechanism, shared with most animals. It has a bias towards present rewards and devalues future rewards. This bias is a psychological barrier against reciprocal cooperation conditional on delayed returns (Stevens et al. 2005). Emotional aversion towards free-riding evolved to override the reward mechanism and facilitate cooperation in these cases (Frank 1988). The fundamental problem for a free-riding strategy like K is that it must operate without emotional aversion. Instead, K must rely on the desire to avoid punishment and the belief that it is safer to free-ride at higher-orders than at lower ones. But without the aid of the moral emotions, the reward mechanism presents the immediate rewards of free-riding as speciously attractive and urges free-riding. Assuming that the moral emotions evolved because the fear of delayed punishment proved ineffective, K will not be able to resist first-order free-riding. The desired flexibility towards free-riding at higher levels is paid with an undesired impulse control problem at the first level.

This argument is based on a particular view of emotions as props to rational decision-making. Without the moral emotions to counteract the effects of the reward mechanism, K cannot prudently avoid free-riding when punishment is likely. But with the moral emotions, the desire for free-riding disappears. The psychological system that protects from undesirably free-riding at the first level blocks the desired flexibility at higher levels. The argument suggests that K lacks psychological plausibility. The same psychological

mechanism that drives first-order cooperation will drive cooperators to uphold punishment at the relevant levels. Punishment is not altruistic in the sense of incurring a cost that first-order cooperators can consistently avoid. It requires yet an experimental approach to establish the relevant psychological facts, but the higher-order free-riding thesis is challenged here with a bit more argument than it brings in its own favor.

Conclusion

The objective distinction between individual and group selection for altruism depends on whether the altruistic organism controls positive assortment, such that altruism will be individually advantageous. I believe this distinction can be accepted both by mainstream evolutionary biologists and multilevel selection theorists. Processes of selection for altruism where individual organisms do not control assortative grouping surely exist. These shall count as group selection for altruism. There are doubts, however, regarding the existence of processes where individual organisms control assortative grouping such that they defeat subversion from within.

But assume, for a moment, that human altruism is such a case. The consequence for both mainstream evolutionary theory and multilevel selection theory is that they need to admit a new and peculiar process, in order to explain the evolution of human altruism. Human altruism is individually selected whenever a psychological mechanism, or a social system, or both, engineer assortative grouping to the extent of securing a higher relative fitness for altruists within a group. This introduces a new category of altruism and of group (Sober and Wilson 1998, 30). Groups where altruism is individually selected within them will not only outcompete groups of egoists, but also groups of altruists where altruism is individually disadvantageous. Those groups are super-organisms. They enhance selection at higher levels by defeating subversion from within. With the force of individual selection producing their inner cohesion, these groups will quickly displace other groups where the appropriate social and psychological mechanisms are not yet in place. Super-organisms may be viewed as structurally similar to genomes, where subversion from within, occasionally observable in meiotic drive, has also been suppressed by the individual selection of traits that suppress the effects of distorter genes (Leigh 1977). Though they benefit the individual, these traits deserve the name ‘altruistic’ because they bind individuals together in a common fate by suppressing the advantages of free-riders. They enhance group level selection processes and promote the evolution of higher levels of organization in nature.

Acknowledgements I thank audiences in Bogotá and Florianópolis, as well as an anonymous referee, for valuable comments.

References

- Alexander R (1987) *The biology of moral systems*. Aldine de Gruyter, New York
- Boehm C (1997) Impact of the human egalitarian syndrome on Darwinian selection mechanics. *Am Nat* 150:S100–S121
- Boehm C (1999) *Hierarchy in the forest—the evolution of egalitarian behavior*. Harvard University Press, Cambridge MA
- Boyd R (2006) Reciprocity: you have to think different. *J Evol Biol* 19(5):1380–1382
- Boyd R, Mathew S (2007) A narrow road to cooperation. *Science* 316:1858–1859
- Boyd R, Richerson PJ (1988) The evolution of reciprocity in sizable groups. *J Theor Biol* 132:337–356

- Boyd R, Richerson PJ (1992) Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol Sociobiol* 13:171–195
- Fehr E, Fischbacher U (2004) Social norms and human cooperation. *Trends Cogn Sci* 8(4):185–190
- Fehr E, Gächter S (2000) Cooperation and punishment in public goods experiments. *Am Econ Rev* 90(4):980–994
- Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415:137–140
- Fletcher JA, Zwick M (2004) Strong altruism can evolve in randomly formed groups. *J Theor Biol* 228:303–313
- Fowler JH (2005) Altruistic punishment and the origin of cooperation. *Proc Natl Acad Sci USA* 102(19):7047–7049
- Frank R (1988) *Passions within reason. The strategic role of the emotions*. W. Norton & Co, New York
- Frank SA (1996) Models of parasite virulence. *Q Rev Biol* 71:37–78
- Gintis H, Bowles S, Boyd R, Fehr E (2003) Explaining altruistic behavior in humans. *Evolution and Human Behavior* 24:153–172
- Hamilton WD (1964) The genetical evolution of social behavior. *J Theor Biol* 7:1–52
- Hauert C, Traulsen A, Brandt H, Nowak MA, Sigmund K (2007) Via freedom to coercion: the emergence of costly punishment. *Science* 316:1905–1908
- Hume D (1777, 1902) An enquiry concerning human understanding. In: Selby-Bigge L (ed) *Enquiries concerning the human understanding and concerning the principles of morals*, Oxford University Press, Oxford
- Kerr B, Godfrey-Smith P (2002) Individualist and multi-level perspectives on selection in structured populations. *Biology and Philosophy* 17(4):477–517
- Leigh EG (1977) How does selection reconcile individual advantage with the good of the group? *Proc Natl Acad Sci* 74:4542–4546
- Maynard Smith J (1964) Group selection and kin selection. *Nature* 201:1145–1146
- Maynard Smith J (1976) Group Selection. *Q Rev Biol* 51:277–283
- Milinski M, Semmann D, Krambeck HJ (2002) Reputation helps solve the ‘tragedy of the commons’. *Nature* 415:424–426
- Nowak MA, Sigmund K (1998) Evolution of indirect reciprocity by image scoring. *Nature* 393:573–577
- Okasha S (2003) Recent work on the levels of selection problem. *Hum Nature Rev* 3:349–356
- Panchanathan K, Boyd R (2004) Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* 432:499–502
- Sherrat T, Roberts G (1998) The evolution of generosity and choosiness in cooperative exchanges. *J Theor Biol* 193:167–177
- Sober E, Wilson DS (1998) *Unto others. the evolution and psychology of unselfish behavior*. Harvard U. Press, Cambridge MA
- Sterelny K (1996) The return of the group. *Philos Sci* 63:562–584
- Stevens JR, Cushman FA, Hauser MD (2005) Evolving the psychological mechanisms for cooperation. *Annual Reviews in Ecology, Evolution and Systematics* 36:499–518
- Trivers R (1971) The evolution of reciprocal altruism. *Q Rev Biol* 46:35–57
- Williams GC (1966) *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton University Press, Princeton, NJ
- Wilson DS (1979) Structured demes and trait-group variation. *Am Nat* 113:606–610
- Wilson DS, Dugatkin LA (1997) Group selection and assortative interaction. *Am Nat* 149:336–351
- Zahavi A (1975) Mate selection: a selection for handicap. *J Theor Biol* 53:205–214
- Zahavi A (1995) Altruism as a handicap—the limitations of kin selection and reciprocity. *J Avian Biol* 26:1–3