doi:10.1111/j.1420-9101.2006.01146.x

COMMENTARY

# How altruism evolves: assortment and synergy

J. A. FLETCHER* & M. DOEBELI†

*Department of Zoology, University of British Columbia, Vancouver, BC, Canada
†Departments of Zoology and Mathematics, University of British Columbia, Vancouver, BC, Canada

If one defines altruism strictly at the population level such that carriers of the altruistic genotype are required to experience, on average, a net fitness cost relative to average population members, then altruism can never evolve. This is simply because a genetically encoded trait can only increase in a population (relative to alternative traits) if the mean fitness of individuals carrying this genotype is higher than the population average fitness. This is true whether the genotype of interest encodes a self-serving behaviour such as enhanced predator avoidance, or an altruistic behaviour in which the actor enhances the fitness of those it interacts with more than its own. The paradox in the evolution of altruism is that carriers that are, on average, at a local disadvantage (i.e. compared to those they interact with) can still have higher fitness than the population average and hence can increase overall.

The most fundamental explanation for how altruism (defined by local interactions) increases in a population requires that there be assortment in the population such that the benefit from others falls sufficiently often to carriers (and at the same time nonaltruists are stuck interacting more with each other). Nonadditivity if present can play a similar role: when collective cooperation yields synergistic benefits (positive nonadditivity) altruistic behaviour can evolve even in the absence of positive assortment, and when there are diminishing returns for cooperation (negative nonadditivity) the evolution of altruism is hindered (Queller, 1985; Hauert *et al.*, 2006).

In their target article Lehmann & Keller (2006) use a form of Hamilton's rule (1964, 1975) to classify different mechanisms by which helping behaviours can evolve. However, the version they develop tends to obscure the fundamental roles that assortment and nonadditivity play. Their framework also confuses local and population-wide definitions of altruism in making distinctions between nonrelatives and relatives, and what they label as mere 'cooperation' vs. true 'altruism'. We argue that a previous generalization of Hamilton's rule developed by Queller (1985) makes clear the roles played by assortment and nonadditivity and therefore serves as a better starting point for classifying various proposed models and mechanism of how altruistic traits can evolve.

*Correspondence:* Jeffrey A. Fletcher, Department of Zoology, The University of British Columbia, 2370-6270 University Blvd., Vancouver, V6T 1Z4 BC, Canada.
Tel.: +604 225 0251; fax: +604 822 2416; e-mail: fletcher@zoology.ubc.ca

## Queller's generalization of Hamilton's rule

Hamilton's rule predicts that the genotype frequency for an altruistic trait will increase in the next generation if the inequality $rB > C$ is satisfied. Here $C$ represents the fitness cost paid by an average individual for exhibiting the helping behaviour, $B$ is the average benefit provided by this help, and $r$, while originally thought of narrowly as a measure of relatedness between helpers and recipients, can more generally be thought of as a measure of assortment between individuals with the helping *genotype* on the one hand, and the helping *phenotypes* (behaviours) of others with which helpers interact (Queller, 1985) on the other hand. Queller's $r$ term, which measures this assortment using covariance, is given by:

$$r = \frac{\mathrm{cov}(G_A, P_0)}{\mathrm{cov}(G_A, P_A)}, \qquad (1)$$

where $G_A$ measures the genotype or breeding value in each individual in the population (subscript A for actor), $P_A$ is the phenotypic value of each actor (e.g. 0 for defection and 1 for cooperation), and $P_0$ is the average phenotype of others interacting with each individual actor (subscript O for others).

This ratio of covariance terms (eqn 1) is less intuitive than $r$ as a measure of relatedness by descent (the original meaning), but rearranging terms yields a straightforward interpretation of Queller's version of Hamilton's rule. Assuming (as Hamilton's version does) that benefits and costs are additive, Queller's version can be written:

$$\mathrm{cov}(G_A, P_0)B > \mathrm{cov}(G_A, P_A)C. \qquad (2)$$

This says simply that the altruistic genotype represented by $G_A$ increases in frequency if those with the genotype on average get more benefit from the behaviour of others than they pay in cost for their own behaviour. The LHS term measures the assortment (covariance) between those with this focal genotype and the helping behaviours of others, scaled by the value of these behaviours ($B$); the RHS term measures the assortment (covariance) between those with the genotype and their own helping behaviours, scaled by the cost of these behaviours ($C$). Taking the covariance over the whole population ensures that if this inequality holds for the helping genotype, it cannot simultaneously hold for the alternative nonaltruistic genotype. Therefore, when Hamilton's rule is satisfied, carriers on average have higher direct fitness than the population average. This form of Hamilton's rule has the advantage of working equally well for interactions among relatives, nonrelatives, and even across species (Fletcher & Zwick, 2006), as well as accommodating the genotype/phenotype differences that result from

 

conditional behaviour, e.g. in iterated interactions (Queller, 1985; Fletcher & Zwick, 2006). Note that Hamilton's rule (including Queller's version) only applies when selection is directional, but not when selection is disruptive (Doebeli & Hauert 2006).

Queller (1985) also showed how to incorporate nonadditivity into Hamilton's rule with an additional covariance term that measures the assortment between those with the focal genotype and the degree to which cooperative behaviours are mutual, scaled by the amount of deviation from additivity:

$$\mathrm{cov}(G_A, P_0)B + \mathrm{cov}(G_A, P_A P_0)D > \mathrm{cov}(G_A, P_A)C. \quad (3)$$

This deviation value ($D$) can be positive (representing synergy), negative (representing diminishing returns), or zero (representing additivity). This inequality shows that there are two fundamental ways to compensate for an average carrier's local sacrifice: (i) sufficient help from others, and/or (ii) sufficient synergistic fitness rewards for mutual cooperation.

## Lehmann and Keller's proposed framework

Lehmann and Keller's (L&K) proposed classification framework relies on a form of Hamilton's rule summarized in their Table 2. In contrast to viewing the $r$ term in Hamilton's rule as a measure of assortment, L&K's approach is to use the most basic meaning of $r$ (relatedness by descent as measured by whole-genome similarity), but then modify the meaning of the benefit ($B$) and cost ($C$) terms to highlight what they believe are the most fundamental mechanisms by which helping behaviours evolve. They designate the modified benefit and cost terms as $b$ and $c$*.

### Examples of mechanisms not included in proposed framework

Other than when helpers give themselves a direct benefit, L&K's framework includes just one mechanism for the evolution of cooperation among nonrelatives: iterative reciprocal behaviour. Yet many other specific mechanisms have been proposed to account for cooperation among nonrelatives. Queller's version of Hamilton's rule tells us that if there are no conditional behaviours (such as reciprocity) then there must be alternative mechanisms for creating positive assortment and/or synergistic fitness benefits. We now consider a couple of models of the evolution of altruism among nonrelatives that do not seem to fit into L&K's classification framework.

Pepper & Smuts (2000, 2002) provide a model of a mechanism they call 'environmental feedback.' Here assortment is mediated by a clumped distribution of resources, and nonadditivity is inherent in the fact that, after consumption, resources grow back logistically. Aggregations around resource clumps that are dominated by cooperators (restrained eaters in this model) get much more per capita benefit than aggregations dominated by defectors (unrestrained eaters). Unrestrained eaters tend to deplete local resources down to a level where they grow back very slowly. The logistic nonlinear payoff from the environment tends to disproportionately favour cooperator-dominated aggregations. This synergy, together with the assortment generated by clumped resources, explains why altruists (that eat less and leave more for their neighbours) can have higher fitness than *average defectors* in the population, despite having lower fitness than *local defectors*.

In the Avilés (2002) model there are co-evolving traits for the tendency to join groups and the tendency to cooperate. Synergy is explicit in the model's fitness function (Avilés, 1999) and nonadditivity disproportionately rewards individuals in groups dominated by high levels of cooperation. As L&K point out, defectors exist in this model in the form of group joiners with low cooperative tendencies, but this does not necessarily lead to defector domination as they claim. Whereas these cheaters have the highest fitness within their groups, synergistic fitness payoffs to groups dominated by cooperative group joiners disproportionately rewards these altruists, again despite their local disadvantage. Thus even when fitness functions are synergistic, the *local* paradox of altruism remains, while *globally* altruism can be stable over time.

Both of the models above focus on interactions among nonrelatives where cooperative behaviours are unconditional and synergy plays a critical role. Yet L&K classify them both (see their Table 3) as being strictly in their 'greenbeard' category. This category is for interactions among relatives where there is 'a linkage disequilibrium between the gene encoding a phenotypic trait used for recognition and the gene(s) responsible for helping' (Lehmann & Keller, 2006 p.). However, this category does not apply to these models as there are no mechanisms for recognizing phenotypic traits in others. In fact none of L&K's categories seem to apply to these models, which have interactions among unrelated individuals and no conditional behaviour. However, in a classification framework based on Queller's version of Hamilton's rule the mechanisms involved in these models (both assortment among unrelated unconditional cooperators and synergism) are made explicit.

While in these papers neither Pepper & Smuts (2002) or Avilés (2002) emphasize synergy as a fundamental explanation for the success of altruists, it plays a crucial role that is not captured explicitly in L&K's proposed framework. L&K do briefly address the issue of synergy, but suggest that it should be accounted for in their term for what helpers directly give themselves. This ignores that nonadditive effects depend on collective action and

*(Note that in L&K the symbols $B$ and $C$ have related but different definitions than those used here)

are not decomposable and attributable to individual action alone. The lack of an explicit classification involving nonadditivity seems to lead to some misclassifications using the L&K framework.

Even in a situation where fitness functions are additive, there is no conditional behaviour, and altruists give nothing to relatives (or to themselves), altruism can increase if randomly formed groups last for more than one generation, as is true in classic 'haystack' models (Maynard Smith, 1964; Wilson, 1987). There can be a sufficient increase in assortment among cooperative nonrelatives such that altruism is selected for, even if groups last only two generations (Fletcher & Zwick, 2004). This possibility, as well as others that cause unrelated and unconditional cooperators to interact more often than random, is also not accounted for in any of L&K's classifications. Overall, it seems that their framework might be too narrow, especially with regard to interactions among nonrelatives, and with regard to the role of nonadditivity. In contrast, Queller's version of Hamilton's rule easily accommodates both assortment among nonrelatives and nonadditivity.

## A bias towards relatedness

As mentioned, reciprocal altruism with conditional iterative behaviours is an explicit mechanism for the evolution of cooperation in L&K's framework. Yet, here too there are problems. In their framework interactions among nonrelatives are classified very differently from interactions among relatives – not only as being due to different mechanisms (e.g. reciprocity vs. kin selection), but also as being due, respectively, to self-interest vs. sacrifice. Interactions among nonrelatives are designated *a priori* as merely 'cooperative' (in L&K's terminology), whereas interaction among relatives are classified as truly altruistic (L&K Table 2).

In L&K's framework, when unrelated conditional cooperators repeatedly interact, all the benefit they receive from others is used to reduce their cost in the $c$ term (L&K eqn 5 and Table 2). This implements a population-wide definition of altruism: if on average conditional helpers receive more benefit than their cost in interactions with nonrelatives (which must be true for the trait to increase), then they are classified as nonaltruistic (merely cooperative). On the other hand, a conditional cooperator that conditions its behaviour on whether a partner is related or not, has all the benefit it receives from others added to the $b$ term (and is considered to have experienced a true altruistic cost because the $c$ term is unmodified and therefore greater than zero). This leads to a very unsatisfactory result in this classification framework: among nonrelatives, discriminating about who to cooperate with is considered completely self-serving, whereas among relatives the same situation is considered altruistic. Note that, in any case, conditional cooperators are locally altruistic in that on average they cause their interaction partners to do both better than themselves, and better than they would do interacting with defectors.

As L&K point out, the product $rB$ in Hamilton's rule can have two interpretations. Traditionally it is interpreted as a measure of indirect fitness: the degree to which, on average, the benefit *given* by a focal altruist increases the direct fitness of others carrying the same altruistic trait. The product $rB$ can alternatively be interpreted as a measure of direct fitness benefits: the degree to which an average carrier *receives* benefit from others (whether related or not). It is a common misconception when using the indirect fitness approach to assume that altruists' fitness can be enhanced by higher fitness in relatives (e.g. more offspring), while carriers themselves have less direct fitness. For example, a recent review article on the evolution of cooperation states that inclusive fitness differs from other models of how helping behaviours evolve 'in that the cooperative individual need not benefit from its act' (Sachs *et al.*, 2004, p. 143). This perspective confounds an accounting technique with a mechanism.

In order to avoid double counting one can either record the average amount of benefit provided by altruists to other carriers ($rB$) at its source or at its destination, but not both. The indirect fitness approach records what is sent out from carriers towards related helpers; the direct fitness approach records what is received by carriers from others. Of course in reality, carriers can both give and receive benefit and for the helping genotype to increase overall, on average, what is received must exceed cost (such that the average direct fitness of carriers exceeds the population average). In this sense, the evolution of helping behaviours (even nonconditional ones) involves reciprocity (i.e. assortment between the helping genotype and help from others to which they give benefit).

The direct and indirect fitness approaches yield the same result if the value of altruists' behaviours equals the value of others' behaviours, but if this assumption does not hold (for instance if interactions are heterospecific and symbionts exchange different levels of benefit) then only the direct fitness approach to Hamilton's rule works correctly (Fletcher & Zwick, 2006). This suggests that the direct fitness approach is more general – whether a particular helping genotype increases fundamentally does not depend on how much benefit carriers *provide* to relatives, but only on whether the amount carriers *receive* from others (related or not) allows them to have above average fitness. Note that in Queller's version of Hamilton's rule, because there is no $G_O$ term, a direct fitness perspective is required where $B$ is a measure of the benefit coming from others.

Whereas L&K state in their introduction that they will use the direct fitness approach to Hamilton's rule, it tends to confuse their presentation that they switch back and forth between the two interpretations. For instance, they

switch to the indirect fitness interpretation to explain the meaning of the *b* term in their eqn 4. This interpretation supports their emphasis on relatedness as it requires genetic similarity between actors and recipients, whereas the more general direct fitness approach used by Queller highlights the common mechanisms that select for altruism among both relatives and nonrelatives.

### Types of help

Models vary in how the help from altruists is distributed. The two most common scenarios are that benefit is given only to others, or that benefit is given to the helper's whole local group and then distributed among all group members. Pepper (2000) distinguishes these types as 'other-only' and 'whole-group', respectively. In the latter case some of the benefit comes back to the giver, but helpers are still at a relative disadvantage to those they interact with in that only helpers pay a cost, but locally all share in the benefit. This is true even if an altruist's share of the benefit it provides is greater than its cost – what Wilson (1979, 1990) has called 'weak' (as opposed to 'strong') altruism.

In L&K's framework what altruists give themselves directly is subtracted from the cost term in *c*. If *c* becomes negative, L&K call this 'cooperation' rather than (weak) altruism. Note, however, that the distinction between weak and strong altruism (or between cooperation and altruism as in L&K) is not fundamental: the local paradox of cooperation exists for both weak and strong altruists. In both scenarios, cooperators, on average, cause those they interact with to do both better than themselves, and better than they would do interacting with defectors. The distinction between weak and strong altruism has been emphasized (e.g. Nunney, 1985) because globally it is the boundary condition for what can evolve given random interactions each generation. Under this idealization, weak altruism evolves, but strong altruism cannot. On the other hand, if interactions are not determined completely at random, either result is possible. If there is negative assortment (i.e. interactions among cooperators are less frequent than expected), then the local dilemma is more exposed and even weak altruists can be selected against. Conversely, if there is positive assortment, even strong altruism can evolve. As mentioned above, a distinction based on whether *c* is more or less than zero is also muddied in the L&K framework because they subtract benefit *received from others* in calculating their *c* term, not just the benefit that altruists *give themselves* directly.

## Conclusion

Lehmann & Keller's goal of providing a framework for classifying different mechanisms by which helping behaviours can evolve is a very worthy one. Two fundamental mechanism involved in the evolution of cooperation are: (i) an assortment among the helping genotype of interest and the helping behaviours of others with which carriers interact (related or not) and (ii) nonadditive fitness effects for mutual cooperation. Because L&K's framework does not emphasize either of these mechanisms, we suggest that Queller's version of Hamilton's rule forms a better basis upon which to build a classification system. It would be very helpful if each description of proposed models made clear the following points:

- What is the nature of the local dilemma? In what ways do local interactions put altruists at a relative disadvantage and those they interact with at an advantage? Of less import, but useful in comparing models: is benefit other-only or whole-group; and in the case of random interaction models, is altruism strong or weak?
- What features of the model, if any, affect assortment between carriers of the altruistic genotype and the helping behaviours of others (e.g. population viscosity among relatives, kin recognition, conditional behaviour among nonrelatives, etc.), thus helping carriers overcome the local dilemma?
- What features of the model, if any, lead (explicitly or implicitly) to nonadditive fitness consequences?

Model descriptions that are explicit about these points would help in addressing L&K's goal of reducing the confusion about the mechanisms promoting the evolution of altruism.

## Acknowledgments

We are grateful to L. Avilés and F. Guillaume for helpful comments on the manuscript. JAF was supported by the NSF International Fellowship Program (USA) and MD was supported by NSERC (Canada) and by the James S. McDonnell Foundation (USA).

## References

Avilés, L. 1999. Cooperation and non-linear dynamics: an ecological perspective on the evolution of sociality. *Evol. Ecol. Res.* **1**: 459–477.

Avilés, L. 2002. Solving the freeloaders paradox: genetic associations and frequency-dependent selection in the evolution of cooperation among nonrelatives. *Proc. Natl. Acad. Sci.* **99**: 14268–14273.

Doebeli, M. & Hauert, C. 2006. Limit's of Hamilton's rule. *J. Evol. Biol.* **19**: 1386–1388.

Fletcher, J.A. & Zwick, M. 2004. Strong altruism can evolve in randomly formed groups. *J. Theor. Biol.* **228**: 303–313.

Fletcher, J.A. & Zwick, M. 2006. Unifying the theories of inclusive fitness and reciprocal altruism. *Am. Nat.* **168**: 252–262.

Hamilton, W.D. 1964. The genetical evolution of social behavior I and II. *J. Theor. Biol.* **7**: 1–52.

Hamilton, W.D. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. In: *Biosocial Anthropology* (R. Fox, ed.), pp. 133–155. John Wiley and Sons, New York.

Hauert, C., Michor, F., Nowak, M.A. & Doebeli, M. 2006. Synergy and discounting of cooperation in social dilemmas. *J. Theor. Biol.* **239**: 195–202.

Lehmann, L. & Keller, L. 2006. The evolution of cooperation. A general framework and a classification of models. *J. Evol. Biol.* **19**:.

Maynard Smith, J. 1964. Groups selection and kin selection. *Nature* **201**: 1145–1147.

Nunney, L. 1985. Group selection, altruism, and structured-deme models. *Am. Nat.* **126**: 212–230.

Pepper, J.W. 2000. Relatedness in trait group models of social evolution. *J. Theor. Biol.* **206**: 355–368.

Pepper, J.W. & Smuts, B.B. 2000. The evolution of cooperation in an ecological context: an agent-based model. In: *Dynamics in Human and Primate Societies: Agent-Based Modeling of Social and Spatial Processes* (T.A. Kohler & G.A. Gumerman, eds), pp. 45–76. Oxford University Press, Oxford.

Pepper, J.W. & Smuts, B.B. 2002. Assortment through environmental feedback. *Am. Nat.* **160**: 205–213.

Queller, D.C. 1985. Kinship, reciprocity and synergism in the evolution of social behavior. *Nature* **318**: 366–367.

Sachs, J.L., Mueller, U.G., Wilcox, T.P. & Bull, J.J. 2004. The evolution of cooperation. *Q. Rev. Biol.* **79**: 135–160.

Wilson, D.S. 1979. Structured demes and trait-group variation. *Am. Nat.* **113**: 606–610.

Wilson, D.S. 1987. Altruism in mendelian populations derived from sibling groups: the haystack model revisited. *Evolution* **41**: 1059–1070.

Wilson, D.S. 1990. Weak altruism, strong group selection. *Oikos* **59**: 135–140.