

## COMMENTARY

## Modelling social evolution: the relative merits and limitations of a Hamilton's rule-based approach

T. WENSELEERS

*Zoological Institute, Catholic University of Leuven, Leuven, Belgium*

Lehmann & Keller (2006) convincingly argue that cooperation or altruism can evolve only when at least one of the following conditions are met: (i) cooperation has direct personal benefits, (ii) individuals are genetically related, (iii) individuals have information on the likely behaviour of others or (iv) individuals recognize other cooperators through a phenotypic label (green beard recognition). The first condition is quite obvious. The three others are also intuitive, given that they provide various conditions under which interacting individuals are positively related, meaning more likely than chance to interact with fellow cooperators (Frank, 1998; Pepper & Smuts, 2002; Foster & Wenseleers, 2006; Foster *et al.*, 2006). By focusing on costs and benefits and relatedness, Lehmann and Keller (2006) are adopting an inclusive fitness or Hamilton's rule-based approach to classify social evolution models (Hamilton, 1964). Personally, I believe a Hamiltonian perspective is indeed a very intuitive one, and should be adopted more widely. This is true particularly in the area of game theory, where currently little effort is made to interpret results this way, and models are frequently simulation-based, making general, intuitive interpretation of results difficult. In fact, in my own work, Hamilton's rule has always taken a central place in the derivation and interpretation of model results (e.g. Wenseleers *et al.*, 2003, 2004a, b; Ratnieks *et al.*, 2006).

At the same time, however, one may also wonder whether it is possible to translate every single social evolution model into the form of Hamilton's rule. Here, I am skeptical, the reason being that Hamilton's rule has well-known limitations. In particular, Hamilton's rule only works correctly under weak selection and additive gene action (Cavalli-Sforza & Feldman, 1978; Michod, 1982; Grafen, 1985; Bulmer, 1994; Hamilton, 1964, 1995; Frank, 1997; Roze & Rousset, 2003; Rousset, 2004). In addition, evolution is normally assumed to be close to equilibrium, with most of the population fixed for a single type (Frank, 1998). Finally, with Hamilton's rule being a deterministic equation, it entirely neglects stochasticity,

and thus ignores the possibility for deleterious altruistic genotypes to go to fixation as a result of drift (Frank, 1997), which models have shown to be quite possible in small populations (Eshel, 1972; Rousset, 2004).

In the section below I will give two worked-out examples where some of these limitations come into play and where, due to violations of assumptions, Hamilton's rule does not correctly predict evolutionary change. The implication is that partitioning selection into direct and indirect fitness components is not always as straightforward as Lehmann and Keller make out. In many cases, the conditions under which a cooperative genotype would spread relative to a noncooperative one simply cannot be written in the form of Hamilton's rule. In such situations, the inclusive fitness framework of Lehmann and Keller cannot be used to accurately classify models.

### Pure strategy games

Hamilton (1964, 1995) thought that evolution is a gradual process and that mutants would usually behave only slightly differently from the wild type. This assumption of weak selection was also explicitly made in the derivation of his theory of inclusive fitness (Hamilton, 1964; Michod, 1985; Grafen, 1985; Bulmer, 1994, p. 194) and is still made even in more recent extensions of the theory (Frank, 1998). It is clear, however, that in practice, the assumption of weak selection must frequently be violated. Indeed, strong selection occurs in all models where individuals are considered to employ pure (discrete) strategies, as opposed to either continuous or mixed (probabilistic) strategies. Examples abound (Grafen, 1979; Queller, 1984), but for the purpose of illustration I will just cite two examples. Godfray (1987) analysed when a cooperative, nonsiblicidal type of parasitoid wasp would be able to invade in a population of siblicidal ones. Wenseleers *et al.* (2001) analysed Mendelian segregation as an evolutionary game, and calculated the equilibrium ratio of cooperative, fairly inherited chromosomes and noncooperative, driving chromosomes. In both cases, it was shown that Hamilton's rule does not correctly predict the conditions for spread of the cooperative type (Bulmer, 1994, p. 194; Wenseleers, 2001; Wenseleers *et al.*, 2001).

In the Appendix, I give a formal proof that Hamilton's rule fails for pure strategy games. The reason for its failure, however, can also be seen intuitively. Consider an evolutionary game in which the fitness of an individual  $w$  is dependent both on its own level of altruism  $p$  and the expected level of altruism of social interactants  $q$ , as in  $w = w_0 - cp + b \cdot q + d \cdot p \cdot q$ , where  $w_0$  is the baseline fitness,  $c$  and  $b$  are the costs and benefits of altruism in an all-defecting population and  $d$  is the synergistic benefit that arises when both players cooperate (Queller, 1984, 1985). Using this payoff structure, it can be shown that altruism can invade in an all-defecting population when  $b \cdot r > c$  if individuals play mixed strat-

*Correspondence:* T. Wenseleers, Zoological Institute, Catholic University of Leuven, Naamsestraat 59, B-3000 Leuven, Belgium.  
Tel.: +32 16 323964; fax: +32 16 324575;  
e-mail: tom.wenseleers@bio.kuleuven.be

egies, but when  $b \cdot r + d \cdot r > c$  if individuals play pure strategies (see Appendix and Grafen, 1979; Queller, 1984). These conditions show that the circumstances under which altruism can evolve are more liberal when individuals adopt pure as opposed to mixed strategies, and that the original version of Hamilton's rule,  $b \cdot r > c$ , only holds for the mixed (or continuous) strategy case. Why is this the case? The reason is quite simple. Imagine that an individual would play pure strategies and that it would interact with a clonemate with probability  $r$ , where  $r$  is the coefficient of relatedness. In that case, the individual could be certain that if it cooperated itself, the other would cooperate as well with probability  $r$ , and this would result in an additional synergistic benefit  $d$ . This is not true, however, if individuals play probabilistic, mixed strategies, since in that case the other individual might still end up either cooperating or defecting. Hence, when individuals adopt pure strategies, the condition for the evolution of altruism is not given by Hamilton's rule, but is actually more liberal. The only exception, when Hamilton's rule would still correctly predict evolutionary change even with pure strategies is when there is no synergistic benefit of both interactants cooperating ( $d = 0$ ; Queller, 1984), or if the probability of expression or 'penetrance' of the cooperator allele is low, as this would again result in weak selection (Godfray, 1987; Wenseleers, 2001). Under diploid genetics, an added complication is that genes can have nonadditive effects as a result of dominance or interactions with other loci (epistasis). This can again preclude an analysis in terms of Hamilton's rule (Cavalli-Sforza & Feldman, 1978; Roze & Rousset, 2004). Incidentally, it should be noted that pure- and mixed-strategy evolutionarily stable strategies (ESSs) in general are not the same (see Appendix and Grafen, 1979; Queller, 1984), and that the classic result in game theory (Bishop & Cannings, 1978; Bulmer, 1994, p. 155) that shows that the two are equivalent only applies when relatedness is zero.

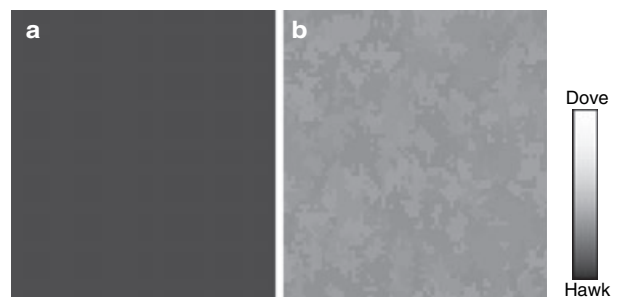
### Nonlinear dynamics

Another common assumption in inclusive fitness and ESS models is that the population is nearly monomorphic for a single type (e.g. Frank, 1998). By calculating the resident strategy that is stable against invasion by a rare mutant, the ESS can then easily be determined (Maynard Smith, 1982; Frank, 1998). In this method, relatedness is easily taken into account. For example, if the fitness  $w$  of an individual is a function of its own level of altruism  $p$  and the level of altruism  $q$  of the social interactant, then a rare, slightly more altruistic variant will be able to invade when  $\partial w / \partial p + \partial w / \partial q \cdot r > 0$  (Frank, 1998). This is a generalized version of Hamilton's rule, with  $r$  equal to relatedness and with  $\partial w / \partial p$  and  $\partial w / \partial q$  describing the costs and benefits of altruism (Frank, 1998).

But can populations ever be assumed to be monomorphic? Adaptive dynamics theory has shown that even if

one starts off with a monomorphic population, there can be diversifying selection. This can lead to bifurcations and to a stochastic equilibrium whereby two or more types stably coexist (e.g. Doebeli *et al.*, 2004). In such situations, Hamilton's rule again cannot predict the equilibrium level of cooperation. To give just one example of this phenomenon, consider the spatial hawk-dove game (Hauert & Doebeli, 2004), where individuals interact on a grid, and the payoff to an individual of interacting with another one is  $E(p, q) = (V/2) \cdot p \cdot q + v \cdot (1-p) \cdot q + (V-C) / 2 \cdot (1-p) \cdot (1-q)$ , where  $p$  and  $q$  are the level of dove behaviour the actor and recipient express, and  $V$  and  $C$  are the value of the contested resource and fighting cost, respectively (Fig. 1). If individuals play this game on a square grid and interact with their four nearest neighbours, then the relative fitness of a single rare mutant that expresses a level of dove of  $p$  in a population expressing a level of dove  $q$  is  $w = 4 \cdot E(p, q) / (E(q, p) + 3 \cdot E(q, q))$ . In this equation, the numerator is the payoff of the mutant playing against four wild type neighbours, and the denominator is the payoff of the neighbours playing against one mutant and three wild types.

The Hamilton's rule condition above would predict that in the absence of relatedness, an equilibrium occurs when  $\partial w / \partial p = 0$ , which gives an ESS level of dove of  $q^* = 1 - (5/3)(V/C)$  (for a formal treatment of spatial ESSs see Nowak & May, 1993). Simulation shows that if one starts off from an initially low level of cooperation



**Fig. 1** Nonlinear dynamics in the spatial continuous hawk-dove game. In both (a) and (b) we assume a hawk-dove payoff structure with a fighting cost  $C$  that is twice the value  $V$  of the contested resource [payoff values:  $E(C, C) = 0.25$ ,  $E(C, D) = 0$ ,  $E(D, C) = 0.5$ ,  $E(D, D) = -0.25$ ]. (a) If the population is allowed to evolve from an initially low level of cooperation, the expected ESS level of dove of  $1 - (5/3)(V/C) = 0.17$  is reached. (b) However, if one starts off from an initially high level of cooperation, the population gets stuck at a much higher average level of cooperation whereby several types stably coexist. At this stochastic equilibrium, the average level of cooperation is 0.53. Results were obtained using the simulator of C. Hauert (<http://www.univie.ac.at/virtuallabs/Snowdrift/struct.mixed.html>) using a square  $70 \times 70$  grid wrapped around a torus, a Von Neumann neighbourhood (interaction with four nearest-neighbours), synchronous updates, a thermal update rule with no noise, a mutation rate of 0.00001 and a standard deviation of gaussian mutations of 0.01. Results are shown after 20 000 generations. Shades of grey represent the level of cooperation at each site of the lattice.

( $q \equiv 0$ ), the population does indeed converge to this equilibrium (Fig. 1). However, if one starts off from an initially high level of cooperation ( $q \equiv 1$ ), the population gets stuck at a much higher average level of cooperation (Fig. 1). The equilibrium that is reached is a stochastic equilibrium characterized by the stable coexistence of several types (Fig. 1). At intermediate starting values, with initial levels of dove between  $1-(5/3)(V/C)$  and  $1-(V/C)$ , the population also turns out to be in equilibrium, because even though lone, slightly more selfish mutants would be able to invade, pairs of such mutants would die out.

Again, it is quite clear that the nonlinear dynamics seen in spatial games are hard to capture using a simple condition such as Hamilton's rule. Although van Baalen & Rand (1998) claimed that the invasion condition in spatially explicit games can be reduced to some form of Hamilton's rule, this conclusion was still based on a simplified pair-approximation technique, which neglects higher-order spatial structure. In addition, Di Paolo (1999) has shown that the high levels of cooperation observed in spatial games cannot be explained in terms of kin selection. Instead, discreteness, stochasticity and out-of-equilibrium dynamics all contribute to the final result.

## Conclusion

Lehmann & Keller (2006) have done an excellent job of classifying different models of cooperation into a single inclusive fitness scheme. As I have shown, however, inclusive fitness theory also has inherent limitations, and due to violations of assumptions often does not accurately predict the equilibrium level of cooperation. In such cases, the scheme of Lehmann and Keller cannot be used to accurately classify models. Generalizing inclusive fitness theory to alleviate some of its current limitations remains an important challenge for the future.

## Acknowledgments

I am most grateful to Steve Frank for helping me derive the pure strategy game results. I also thank Kevin Foster, Andy Gardner and Hanna Kokko for comments and the late John Maynard Smith for explaining me the intuitive reason for why pure and mixed strategy games differ. This work was supported by a fellowship from the Fund for Scientific Research – Flanders.

## References

- van Baalen, M. & Rand, D.A. 1998. The unit of selection in viscous populations and the evolution of altruism. *J. Theor. Biol.* **193**: 631–648.
- Bishop, D.T. & Cannings, C. 1978. A generalised war of attrition. *J. Theor. Biol.* **70**: 85–124.
- Bulmer, M. 1994. *Theoretical Evolutionary Ecology*. Sinauer Associates Publishers, Sunderland, MA.
- Cavalli-Sforza, L.L. & Feldman, M.W. 1978. Darwinian selection and altruism. *Theor. Popul. Biol.* **14**: 268–280.
- Di Paolo, E.A. 1999. A little more than kind and less than kin: the unwarranted use of kin selection in spatial models of communication. *Adv. Artif. Life* **1674**: 504–513.
- Doebeli, M., Hauert, C. & Killingback, T. 2004. The evolutionary origin of cooperators and defectors. *Science* **306**: 859–862.
- Eshel, I. 1972. On the neighbor effect and the evolution of altruistic traits. *Theor. Popul. Biol.* **3**: 258–277.
- Foster, K.R. & Wenseleers, T. 2006. A general model for the evolution of mutualisms. *J. Evol. Biol.* **19**: 1275–1282.
- Foster, K.R., Wenseleers, T. & Ratnieks, F.L.W. 2006. Kin selection is the key to altruism. *Trends Ecol. Evol.* **21**: 57–60.
- Frank, S.A. 1997. The Price Equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* **51**: 1712–1729.
- Frank, S.A. 1998. *The Foundations of Social Evolution*. Monographs in Behavior and Ecology, Princeton University Press, Princeton, NJ.
- Godfray, H.C.J. 1987. The evolution of clutch size in parasitic wasps. *Am. Nat.* **129**: 221–233.
- Grafen, A. 1979. The hawk-dove game played between relatives. *Anim. Behav.* **27**: 905–907.
- Grafen, A. 1985. A geometric view of relatedness. *Oxf. Surv. Evol. Biol.* **2**: 28–29.
- Hamilton, W.D. 1964. The genetical evolution of social behaviour. I & II. *J. Theor. Biol.* **7**: 1–52.
- Hamilton, W.D. 1995. *Narrow Roads of Gene Land. Vol. 1. Evolution of Social Behaviour*. W.H. Freeman, New York.
- Hauert, C. & Doebeli, M. 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* **428**: 643–646.
- Lehmann, L. & Keller, L. 2006. The evolution of cooperation and altruism - a general framework and a classification of models. *J. Evol. Biol.* **19**: 1365–1376.
- Maynard Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge University Press, New York.
- Michod, R.E. 1982. The theory of kin selection. *Annu. Rev. Ecol. Syst.* **13**: 23–55.
- Nowak, M.A. & May, R.M. 1993. The spatial dilemmas of evolution. *Int. J. Bifurcat. Chaos* **3**: 35–78.
- Pepper, J.W. & Smuts, B.B. 2002. A mechanism for the evolution of altruism among nonkin: positive assortment through environmental feedback. *Am. Nat.* **160**: 205–213.
- Price, G.R. 1970. Selection and covariance. *Nature* **227**: 520–521.
- Queller, D.C. 1984. Kin selection and frequency-dependence – a game theoretic approach. *Biol. J. Linn. Soc.* **23**: 133–143.
- Queller, D.C. 1985. Kinship, reciprocity and synergism in the evolution of social behavior. *Nature* **318**: 366–367.
- Ratnieks, F.L.W., Foster, K.R. & Wenseleers, T. 2006. Conflict resolution in insect societies. *Annu. Rev. Entomol.* **51**: 581–608.
- Rousset, F. 2004. *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press Monographs in Population Biology, Princeton, New Jersey.
- Roze, D. & Rousset, F. 2003. Selection and drift in subdivided populations: a straightforward method for deriving diffusion approximations and applications involving dominance, selfing and local extinctions. *Genetics* **165**: 2153–2166.
- Roze, D. & Rousset, F. 2004. The robustness of Hamilton's rule with inbreeding and dominance: kin selection and fixation probabilities under partial sib mating. *Am. Nat.* **164**: 214–231.
- Wenseleers, T. 2001. An introduction to the evolutionary modelling of conflict. In: *Proceedings of the EU 'INSECTS' Training*

Workshop on the Mathematical Modelling of Social Behaviour, pp. 21–53. Great Hucklow, UK, 14–18 September 2001.

Wenseleers, T., Ratnieks, F.L.W. & Billen, J. 2001. Towards a General Theory of Conflict: The Sociobiology of Mendelian Segregation. In: *Chapter 9. Conflict from Cell to Colony*, pp. 163–173. PhD thesis, University of Leuven, Belgium.

Wenseleers, T., Ratnieks, F.L.W. & Billen, J. 2003. Caste fate conflict in swarm-founding social Hymenoptera: an inclusive fitness analysis. *J. Evol. Biol.* **16**: 647–658.

Wenseleers, T., Hart, A.G. & Ratnieks, F.L.W. 2004a. When resistance is useless: policing and the evolution of reproductive acquiescence in insect societies. *Am. Nat.* **164**: E154–E167.

Wenseleers, T., Helanterä, H., Hart, A.G. & Ratnieks, F.L.W. 2004b. Worker reproduction and policing in insect societies: an ESS analysis. *J. Evol. Biol.* **17**: 1035–1047.

**Appendix: proof that Hamilton’s rule fails for pure strategy games**

Below I will prove that the original version of Hamilton’s rule does not work for pure strategy games. To start, write reproductive success  $w$  as a function of one’s own level of altruism ( $p$ ) and of the expected degree of altruism of social interactants ( $q$ ), as in

$$w = p \cdot q \cdot E(C, C) + p \cdot (1 - q) \cdot E(C, D) + (1 - p) \cdot q \cdot E(D, C) + (1 - p) \cdot (1 - q) \cdot E(D, D), \tag{A.1}$$

where the  $E$ ’s are the payoffs of the different strategy pairs (cooperate,  $C$ , or defect,  $D$ ). For example,  $E(D, C)$  is the payoff when the actor defects and the partner cooperates. This equation can be rewritten as follows

$$w = -c \cdot p + b \cdot q + d \cdot p \cdot q + E(D, D), \tag{A.2}$$

where  $c = E(D, D) - E(C, D)$  and  $b = E(C, D) - E(C, C)$  are the cost and benefit of altruism (measured in an all-defecting population),  $d = E(C, C) - E(C, D) + E(D, D) - E(D, C)$  is any synergistic benefit that arises when both players cooperate (Queller, 1984, 1985) and  $E(D, D)$  is the baseline fitness  $w_0$ . A gene for altruism spreads when the slope of reproductive success  $w$  on genotype  $g$  is positive (Price, 1970; Frank, 1998), which is when

$$\beta_{wg} = -c \cdot \beta_{pg} + b \cdot \beta_{qg} + d \cdot \beta_{pqg} > 0 \tag{A.3}$$

The slope of actor phenotype  $p$  on actor genotype  $g$ ,  $\beta_{pg}$ , is by definition one. The slope of partner phenotype  $q$  on actor genotype  $g$ ,  $\beta_{qg}$ , is the definition of relatedness ( $r$ ) (Frank, 1998). The only regression coefficient left to calculate is  $\beta_{pqg}$ , the slope of the product of actor and partner phenotype  $p \cdot q$  on actor genotype  $g$ . This regression can be expanded as

$$\begin{aligned} \frac{\Delta p q}{\Delta g} &= \frac{(p + \Delta p)(q + \Delta q) - p q}{\Delta g} \\ &= \frac{p \cdot \Delta q + q \cdot \Delta p + \Delta p \cdot \Delta q}{\Delta g} \\ &= p \cdot (\Delta q / \Delta g) + q \cdot (\Delta p / \Delta g) + \Delta q \cdot (\Delta p / \Delta g). \end{aligned} \tag{A.4}$$

Since  $\Delta q / \Delta g$  is the slope of partner phenotype on actor genotype, the definition of relatedness, and  $\Delta q = \Delta p \cdot r$  (Frank, 1998),  $\beta_{pqg} = p \cdot r + q \cdot 1 + \Delta p \cdot r$ . In this equation,  $\Delta p$  is the phenotypic deviation of a random individual that changes from one strategy to another. In typical kin selection models, one assumes that only behavioural variants of small effect occur (Bulmer, 1994; Hamilton, 1995), in which case  $\Delta p$  approaches zero. This assumption of weak selection, however, is only appropriate when players can select their behaviour from a continuous set of alternatives and when no player behaves very differently from the population average. The optimum favoured under such circumstances is termed a continuous or mixed ESS (Maynard Smith, 1982). For the pure strategy case, however, the assumption of weak selection no longer holds and  $\Delta p$  equals  $1 - 2p$  because if  $p = 0$ , then the switch to the opposite strategy  $p = 1$  means that  $\Delta p = 1$ , whereas if  $p = 1$ , the switch to the opposite strategy of  $p = 0$  means that  $\Delta p = -1$ . Substituting the regression above into eqn A.3 yields the condition that altruism is selectively favoured when

$$-c + b \cdot r + d \cdot (p \cdot r + q + \Delta p \cdot r) > 0, \tag{A.5}$$

which can be rearranged to

$$-C + B \cdot r + d \cdot \Delta p \cdot r > 0, \tag{A.6}$$

where, following Frank (1998), the costs and benefits of altruism are defined as  $-C = \partial w / \partial p = -c + q \cdot d$  and  $B = \partial w / \partial q = b + p \cdot d$  (Foster & Wenseleers, 2006). The first two terms of this equation match Hamilton’s rule (Frank, 1998). Importantly, however, the equation also contains a third correction term, which is needed when strategies are pure and selection is strong.

Observing that, for symmetrical games, the expected behaviour of a partner ( $q$ ) is equal to that of a random individual, we can set  $p = q$  and solve eqn A.6 for the optimal level of altruism  $p$ . With mixed strategies  $\Delta p = 0$ , and we obtain the mixed strategy ESS

$$\hat{p} = \frac{c - b \cdot r}{d \cdot (1 + r)}. \tag{A.7}$$

With pure strategies  $\Delta p = 1 - 2p$ , and we obtain the pure strategy ESS

$$\hat{p} = \frac{c - (b + d) \cdot r}{d \cdot (1 + r)}. \tag{A.8}$$

In addition, by setting  $p$  and  $q$  to zero in eqn A.6, we can check that altruism can invade in an all-defecting population when  $b \cdot r > c$  if individuals play mixed strategies, but when  $b \cdot r + d \cdot r > c$  if individuals play pure strategies. This again demonstrates that Hamilton’s rule, at least in its original form, only holds for continuous or mixed strategy games. For earlier discussion of the differences between pure and mixed strategy games see Grafen (1979) and Queller (1984).