doi:10.1111/j.1420-9101.2006.01128.x

An evolutionary analysis of the relationship between spite and altruism

L. LEHMANN,* K. BARGUM† & M. REUTER‡

*Laboratory of Ecological and Evolutionary Dynamic, Department of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland
†Department of Biological and Environmental Science, University of Helsinki, Helsinki, Finland
‡The Galton Laboratory, University College London, London, UK

Keywords:

altruism;
game theory;
Hamilton's rule;
harming;
helping;
inclusive fitness;
social evolution;
spite.

Abstract

We investigate the selective pressures on a social trait when evolution occurs in a population of constant size. We show that any social trait that is spiteful simultaneously qualifies as altruistic. In other words, any trait that reduces the fitness of less related individuals necessarily increases that of related ones. Our analysis demonstrates that the distinction between 'Hamiltonian spite' and 'Wilsonian spite' is not justified on the basis of fitness effects. We illustrate this general result with an explicit model for the evolution of a social act that reduces the recipient's survival ('harming trait'). This model shows that the evolution of harming is favoured if local demes are of small size and migration is low (philopatry). Further, deme size and migration rate determine whether harming evolves as a selfish strategy by increasing the fitness of the actor, or as a spiteful/altruistic strategy through its positive effect on the fitness of close kin.

Introduction

Many traits not only affect the fitness of their bearer but also that of other individuals and are hence considered as social traits. Selection on such traits can be analysed using the concept of inclusive fitness (or kin selection), introduced by Hamilton (1964a) more than 40 years ago. Based on his theory, Hamilton derived a condition for the evolution of these traits. In its most common form, now known as 'Hamilton's rule', it states that a trait evolves if $Rb - c > 0$, where $-c$ is the effect of the trait on the fitness of the actor, b is its effect on the fitness of the recipient and R is the coefficient of relatedness measuring the tendency of the recipient to bear the same genes as the actor.

Correspondence: Laurent Lehmann, Laboratory of Ecological and Evolutionary Dynamic, Department of Biological and Environmental Sciences, PO Box 65, 00014 University of Helsinki, Helsinki, Finland.
Present address: Department of Genetics, University of Cambridge, Cambridge, UK.
Tel.: +44 1223 332584; fax: +44 1223 333992;
e-mail: ll316@cam.ac.uk

Inclusive fitness theory represented a breakthrough because it provided an explanation for the evolution of social traits that actually reduce the fitness of the actor ($c > 0$). Starting with Hamilton, the literature distinguishes two ways how traits resulting in a fitness cost for the actor can be favoured by kin selection. First, a trait can be altruistic and spread because the actor increases the fitness of his relatives (R and b positive). Secondly, the trait can be spiteful and spread because it reduces the fitness of a recipient who is less likely to bear the genes of the actor than is an individual taken at random from the population (both R and b negative).

While the concept of altruism has been applied to many traits (including obvious examples such as sterile worker castes in social insects (Hamilton, 1964b) and less obvious ones such as natal dispersal (Hamilton & May, 1977; Taylor, 1988) or mate choice (Taylor & Getz, 1994; Lehmann & Perrin, 2003), the topic of spite has been haunted by lasting debates over what should be considered spiteful both theoretically and empirically (Hamilton, 1970; Gadagkar, 1993; Keller *et al.*, 1994; Foster *et al.*, 2000, 2001). One point of debate is how

spite is defined with respect to the behaviour's effect on the actor. Hamilton's analysis is based on fitness effects and specifies that spiteful traits have a negative effect on the actor's personal fitness ($c > 0$). A review by Foster *et al.* (2001) therefore correctly dismissed definitions based on the effect of the behaviour on the actor's inclusive fitness (rather than personal fitness) or those in which an immediate fitness cost is compensated by a delayed fitness benefit (resulting in a net benefit for the actor's personal fitness). However, discrepancies remains in that some authors (e.g. Vickery *et al.*, 2003; Gardner & West, 2004b; Johnstone & Bshary, 2004) do not distinguish between effects on fitness and effects on fecundity, and define spiteful traits as those that decrease the actor's fecundity. This may lead to wrong classifications of traits as fecundity is not necessarily a correlate of fitness, which also takes competition into account (Rousset, 2004). Hence, a behaviour can reduce an actor's number of juvenile produced while increasing its number of juveniles actually reaching adulthood due to reduced competition. Such a behaviour must consequently be considered selfish rather than spiteful.

The distinction between fecundity and fitness is also relevant when quantifying the effect of the behaviour on recipients. Sociobiologists have traditionally viewed traits as spiteful, if the social act resulted in a negative effect on the fecundity of its immediate recipient, whereas acts that result in a positive effect on the fecundity of its recipient are regarded as altruistic. Again, basing the definition on fitness can lead to a different classification, because under certain conditions traits can increase fecundity, but not fitness (Rousset, 2004, p. 113). In addition, using fitness as a criterion opens the possibility to apply the concepts of altruism and spite to traits that do not involve direct interactions between individuals, but which nevertheless evolve due to the same kin-selective logic as other social acts. This has done very successfully with traits such as dispersal [resulting in reduced competition for relatives (Taylor, 1988; Frank, 1998, Gandon & Michalakis, 1999) or prudent resource use (leaving more resources for relatives (Frank, 1995; Foster, 2004)].

Importantly, focussing on fitness rather than fecundity also allows to assess the effects of the trait on individuals other than the immediate recipients of the social act. Although the fecundity of these third parties is unaffected by the act, their fitness will change due to the act's impact on average population (or deme) fecundity and the intensity of competition. Taking these indirect fitness effects into account can improve our understanding of the ultimate causes that drive the evolution of social traits. In this article, we will show that traits that are spiteful with respect to their direct recipients have a positive fitness effect on other individuals in the population. This means that spite against negatively related individuals leads to a fitness increase – and hence altruism – towards the (necessarily positively related) remainder of the population. Inversely, an altruistic trait directed against relatives results in a fitness decrease (spite) in the negatively related remainder of the population. Altruism and spite hence represent two sides of the same coin.

Our analysis will also shed light on the distinction between 'Hamiltonian spite' and 'Wilsonian spite' (Foster *et al.*, 2000). The definition of the former is based on Hamilton's original proposition of the concept which focused on the effect of a harming behaviour on the fitness of unrelated individuals (Hamilton, 1970). He argued that with negative signs on both the behaviour's fitness effect on the recipient and on relatedness, Hamilton's rule could be fulfilled and the behaviour favoured by selection. Wilson (1975) provided an alternative interpretation of the inclusive fitness effect of a spiteful act. He stated that 'if the spiteful trait causes a relative to prosper to a compensatory degree, the gene favouring spite will increase in the population at large' (Wilson, 1975, p. 57). Some authors (Foster *et al.*, 2000, 2001; Gardner & West, 2004b) have perceived Wilson's account as an evolutionary mechanisms distinct from Hamilton's original concept. However, we will show that due to the inter-relationship between altruism and spite, Wilson's definition of conditions favouring spite coincides with Hamilton's. The only difference is that Wilson concentrates on the positive fitness effect on relatives, whereas Hamilton focusses on the negative effect on the negatively related recipients of the act.

Analysis

Inclusive fitness effect

We start our analysis with a description of the selective pressure acting on a social trait evolving in a population that is of constant size. This trait can affect the fecundity and/or survival of actors and that of various recipients. The strength of both effects is a function of the phenotype of an individual which in turn is determined by its genotype. Selection on such a trait can be analysed by considering a mutant allele coding for a trait value deviating from that expressed by individuals bearing a resident genotype fixed in the population. The direction of selection on such a mutant is given by the inclusive fitness effect (Hamilton, 1964a), which is the sum of the effects of a focal individual (FI) bearing the mutant trait value on the fitness of all individuals in the population (including the FI), weighted by their genetic similarity with the FI. Alternatively, the inclusive fitness effect can be evaluated in the direct fitness manner as a relatedness weighted sum of the effects of all individuals in the population on the fitness of a FI bearing the mutant trait value (Taylor & Frank, 1996; Rousset & Billiard, 2000). Both ways of evaluating selection are strictly equivalent (Rousset, 2004, p. 108).

When some individuals in the population are affected equally by the FI's behaviour, it is not necessary to

 doi:10.1111/j.1420-9101.2006.01128.x

consider them individually and convenient to group such individuals into classes. If we group the individuals of the population into the FI and j classes and assume additive gene action and weak selection, we can write the inclusive fitness effect as

$$\Delta W_{\mathrm{IF}} = -cQ_{\bullet} + \sum_j b_j Q_j. \qquad (1)$$

In this equation, $-c$ denotes the effect of the FI's trait on its own fitness, which is defined here as its expected number of offspring that reach adulthood (Hamilton, 1964a; Grafen, 1984; Rousset, 2004). The effect of the FI on the fitness of members of class j is denoted b_j, which represents a total effect, summed up over all recipients in class j. These effects are weighted by probabilities of genetic identity, where $Q_{\bullet}$ gives the probability of genetic identity between two homologous genes randomly sampled from the FI and Q_j the probability that an individual of class j bears a copy of a homologous gene randomly drawn in the FI. Such probabilities of genetic identity are calculated under a neutral model based on the life cycle used to evaluate the fitness effects c and b_j. Evaluated in this way, S is a consistent measure of the effect of selection on gene frequency change (Taylor, 1996; Rousset, 2003, 2004) and describes the effect of all individuals of a gene lineage on the fitness of individuals bearing that lineage, thus, it is a measure of the effect of a gene lineage on its fitness.

The effects of the FI's trait on recipients ($-c$ and the b_j's) can be positive or negative, resulting in either fitness costs or benefits. For a pair of interacting individuals including the FI and a member of class j, the fitness effects can be used to categorize behaviours (Grafen, 1985). Following (Hamilton, 1964a, 1970) and Rousset (2004, p. 114), four types of behaviours can be distinguished: altruism is defined by $c > 0$ and $b_j > 0$, cooperation by $c < 0$ and $b_j > 0$, spite by $c > 0$ and $b_j < 0$, and finally selfishness by $c < 0$ and $b_j < 0$. This categorization is usually carried out for the focal actor and the recipient that are in direct physical interaction. However, it can be carried out for all classes of recipients in the population, including those that are indirectly affected by the behaviour of the FI (e.g. through the reduction of competition or the increase in the availability of some resource). The sign of the fitness effect then indicates whether the indirect effect of a costly act is altruistic $b_j > 0$, neutral $b_j = 0$ or spiteful $b_j < 0$ towards a third party not involved in physical interactions with the actor.

The inclusive fitness effect given in eqn (1) is based on the assumptions that the population is of constant size and has no stage structure. But, for any population of constant size or fluctuating between various sizes in a stationary demographic regime, the effects of the FI on the fitness of all classes of recipients sum to zero. Indeed, as the mean number of reproducing individuals is fixed (ultimately due to some limiting resource) and fitness measures the number of offspring produced relative to other individuals, any increase in the number of adult offspring left by one class necessarily goes along with a decrease in the number of adult offspring left by another class. This property also holds in a situation where the trait under selection affects the demography of the population itself (Appendix 3). Here, this property implies that

$$c = \sum_j b_j, \qquad (2)$$

(Rousset & Billiard, 2000; Rousset, 2004). This equality shows how the change in the fitness of the individuals is distributed among the FI itself (on the left-hand side) and the various classes of recipients (on the right-hand side). This equality also has a direct consequence for the categorization of social traits. In particular, it shows that if a trait is considered spiteful towards a class $j (b_j < 0)$, it must necessarily also qualify as altruistic with respect to one or more other classes in order for eqn (2) to be fulfilled. Vice versa, traits that are altruistic towards a class of recipients must at the same time be spiteful to one or more other classes.

The inter-relationship between altruism and spite also shows that if the fitness effects of the trait on all members of the population are taken into account, these two categories are not necessarily linked to a social act's effect on the fecundity of the recipient. A trait can be altruistic to a class of recipients either because it directly increases the fecundity of its members (the FI is expressing a helping trait towards members of that class) or because it decreases the fecundity of members of another class (the FI is expressing a harming trait towards members of this other class). Inversely, spite towards a class can result either from a harming trait decreasing the fecundity of that class or from a helping trait increasing the fecundity of any other class. This relationship is illustrated in Fig. 1, which shows the effect of a harming and a helping trait on the fecundity and fitness of the members of a population. Although harming and helping differ in their effects on fecundity, both lead to an increased fitness of the focal family and a decrease in the fitness of the remainder of the population. By taking into account all effects of the trait on fitness (direct and indirect resulting from changes in competition) our analysis differ from those concentrating exclusively on one stage of the life-cycle (Foster *et al.*, 2000, 2001; Gardner & West, 2004b; Johnstone & Bshary, 2004) and thus lead to an interpretation of altruism and spite in terms of ultimate (vs. proximate) categories of behaviours. The figure also illustrates the logic behind eqn (2): conditional on the size of the population in the offspring generation, any increase in a family's share among new breeders necessarily results in a decrease in the share of other members of the population.

An important consequence of the relationship between altruism and spite concerns the distinction between 'Hamiltonian spite' and 'Wilsonian spite'. The distinction

 doi:10.1111/j.1420-9101.2006.01128.x

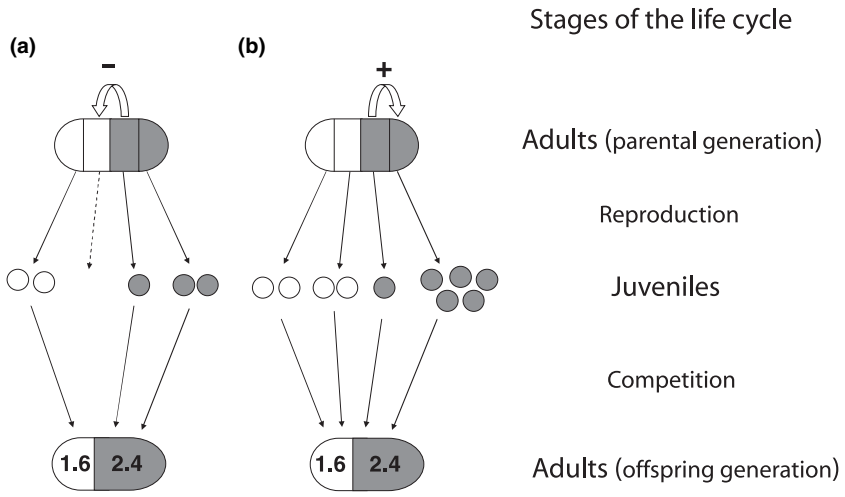

Fig. 1 Two populations, each with four adults initially producing two juveniles, are experiencing a social action where (a) a white individual is harmed and (b) a grey individual is helped. The act of harming or helping result in a cost of one offspring to the grey actor. The harmed individual produces no offspring, whereas the helped individual produces five offspring instead of two. From each parental population, four juveniles survive competition to form the offspring generation. Different interpretations of altruism and spite stem from emphasizing different stages of the life-cycle. Defined from the physical interactions at the adult or juvenile stage (Foster *et al.*, 2000, 2001; Gardner & West, 2004b); or from the effect on fecundity (Vickery *et al.*, 2003; Gardner & West, 2004b; Johnstone & Bshary, 2004), spite and altruism are interpreted as being different evolutionary phenomena. However, seen from the perspective of fitness, both types of interactions result in exactly the same evolutionary consequences. That is, both harming the white group and helping the grey group leads to the grey group increasing in frequency and the white group decreasing in frequency in the population. Accordingly, the total effect of the actor on the fitness of all individuals in the population sums up to zero.

between the two has been based on the two authors' differing interpretation of the selective forces favouring the evolution of spiteful traits. Hamilton argued that such traits were favoured because they decrease the fitness of a class of individuals genetically dissimilar to the FI, whereas Wilson stated that they were favoured because they increase the fitness of individuals genetically similar to the FI. By concentrating on different elements of the sum on the right-hand side of equation eqn (2), it becomes clear that the fitness consequences of a same trait are perfectly compatible both with Hamilton's statement that spite is favoured due to its negative effect on nonkin and with Wilson's interpretation that spite spreads due to its positive effect on kin. Accordingly, 'Hamiltonian spite' and 'Wilsonian spite' can thus be regarded as mutually compatible descriptions of an identical evolutionary process.

Hamilton's rule

The condition for the evolution of a costly social trait is often expressed in the form of Hamilton's rule (Hamilton, 1964a). An (almost always implicit) assumption behind the use of Hamilton's rule is that the recipients of the FI's trait (excluding himself) can be divided into two categories: those more and those less likely to share genes with the FI. These two groups could for instance be the offspring of the FI and offspring of other parents, sibs and nonsibs, or colony-mates and others. In this situation the inclusive fitness effect is given by

$$\Delta W_{\mathrm{IF}} = -cQ_{\bullet} + b_0 Q_0 + b_1 Q_1, \qquad (3)$$

where Q_0 and Q_1 are the probabilities of genetic identity between the FI and an individual of the more and less related class, respectively. In a population that is panmictic and of infinite size or that follows Wright's infinite island model of dispersal we have $Q_1 = 0$.

Hamilton's rule can be derived from eqn (3) by taking advantage of eqn (2) to eliminate one class of recipients. We can eliminate the effect of less related actors (b_1) by using $c = b_0 + b_1$ in order to get an expression for the inclusive fitness effect that involves only the more related class. After rearrangements, the inclusive fitness effect of the behaviour is positive, ($S > 0$), when

$$R_0 b_0 - c > 0 \qquad (4)$$

is satisfied. The coefficient of relatedness is given by ratio of differences of probabilities of genetic identities $R_0 \equiv \frac{Q_0 - Q_1}{Q_{\bullet} - Q_1}$. It measures how similar an individual of the more related class is to the FI, relative to an individual of the less related class, and thus describes the standardized value of an individual of the more related class for the mutant gene lineage. Condition (4) is the form of Hamilton's rule commonly used to express the conditions for the evolution of costly helping traits. It is also the condition that Wilson refers to when pointing

out that a spiteful trait will spread if 'the spiteful trait causes a relative to prosper to a compensatory degree' (Wilson, 1975, p. 57), whereby the compensation implies that $R_0 b_0 > c$.

In exactly the same way we can eliminate b_0 from S. Then, the condition for invasion ($S > 0$) implies that

$$R_1 b_1 - c > 0. \tag{5}$$

This is the form of Hamilton's rule usually used to describe the conditions under which a costly harming behaviour would invade a population. The coefficient of relatedness is now given by $R_1 \equiv \frac{Q_1 - Q_0}{Q_\bullet - Q_0}$ which measures how dissimilar less related individuals are from the FI relative to more closely related individuals. Since by definition $Q_1 < Q_0$, this coefficient of relatedness takes only negative values.

Mathematically, conditions, eqns (4) and (5) are strictly equivalent statements about the direction of selection. Therefore, focusing on negative or positive relatedness for understanding the evolution of social mutant trait is merely a matter of taste. It is worthwhile to point out that the equivalence of eqns (4) and (5) does not imply that $R_0 b_0 = R_1 b_1$. Indeed, for a behaviour to be positively selected it is necessary that $R_0 b_0 > R_1 b_1$, which amounts to saying that the increment in the number of genes of the focal gene lineage transmitted to the next generation, through the effects of actors on the fitness of the various classes of recipients, exceeds the number of genes lost. When $R_0 b_0 = R_1 b_1$ the trait is selectively neutral ($S = 0$, see Appendix 1).

Example

In this section we illustrate our analysis with a concrete example of the selective pressure on a harming trait. We consider a mutation for such a trait appearing in a population fixed for a resident allele that does not express harming. We assume that the population is haploid ($Q_\bullet = 1$) and consists of an infinite number of demes of finite size N (Wright's infinite island model). We also assume that a kin recognition mechanism enables mutant individuals within demes to discriminate between individuals born in their natal deme and those born in different demes. Individuals are thus able to selectively inflict damage upon individuals from different demes (immigrants). The life cycle consists of the following consecutive events. (1) Breeding occurs. A large number (say infinite) of juveniles are produced by each of the N adult individuals of a deme. All adults die. (2) Juveniles disperse independently from each other with probability m to another deme. (3) Juveniles interact socially. Individuals bearing the mutant allele express the harming trait, resulting in a decrease in survival for the actors and recipients. No other selection pressure applies. (4) Regulation occurs. Among all surviving juveniles, N are sampled to form the next generation.

To determine whether harming spreads in such a population, we proceed along the lines of Taylor (1992a,b) and derive the inclusive fitness effects directly. These effects could otherwise be evaluated following the direct fitness approach by constructing an explicit fitness function and taking partial derivatives with respect to the phenotype of the different classes of individuals (Frank, 1998; Rousset & Billiard, 2000). For the interpretation of these effects we adopt a 'gene centred' point of view, concentrating on a copy of the mutant allele residing in a focal adult. We then evaluate the fitness of this adult as the expected number of its offspring reaching adulthood (Rousset, 2004, pp. 138–140). While doing so, we consider the adult's offspring (the focal actors) and the fitness effects arising from their harming acts as an expression of the parental genotype (hence 'gene-centred'), rather than considering the offspring as separate individuals. The approach thus treats as 'focal actors' those individuals bearing the mutation present in the focal adult and affecting its transmission over one round of the life cycle (from focal adult to adults in the next generation), which in the case of our model is mediated through the effect of the mutation on the behaviour of the juveniles carrying it.

We assume that the acts of harming performed by all offspring of the focal parent reduce its fitness by C units and that of parents from other demes (the recipients of harming) by D units. These effects are assumed to be sufficiently weak to be assumed constant and independent of the proportion of immigrants among the juveniles of a deme.

The selective pressure on the harming behaviour just described is given by Hamilton's rule (eqn 4 or 5) with $-c$ being the effect of the juveniles' trait on the fitness of the focal adult, b_0 the effect of juveniles' trait on the fitness of $N - 1$ other adults from the same deme, and b_1 the effect of juveniles' trait on the fitness of adults from other demes. All three effects are functions of the parameters C, D and m. The effect of the mutant trait on the focal adult's fitness is given by

$$-c = -C + \frac{(1-m)^2(D+C)}{N}. \tag{6}$$

The first term of this equation expresses the loss in fitness resulting from the survival cost to juveniles of expressing the harming trait. This cost is offset by a positive second term, the benefit due to reduced competition. This benefit is proportional to the fraction of members of the focal family that remain philopatric ($1 - m$), the total reduction in survival due to harmning acts ($D + C$), and the probability that reduced competition will benefit members of the focal family, given by their proportional representation among the juveniles present in the deme ($(1 - m)/N$)

The effect of the mutant trait on the fitness of other parents from the focal deme is

 doi:10.1111/j.1420-9101.2006.01128.x

$$b_0 = (1-m)^2 \frac{(D+C)(N-1)}{N}. \quad (7)$$

This effect is equal to the share of the benefit of reduced competition caused by the harming behaviour of philopatric members of the focal family $(D + C)(1 - m)$ which goes to the philopatric members of other families, occurring with frequency $(1 - m)(N - 1)/N$ among the juveniles of the deme. Finally, the effect of the trait on adults from other demes is given by $b_1 = c - b_0$, which is

$$b_1 = (1-m)(-D + m(D+C)) + m(-D + (D+C)). \quad (8)$$

This equation separates the fitness effect caused by focal actors that do and do not migrate into two term, both which comprise the negative effect of harming as well as the associated positive effect of reduced competition.

Because of our assumption of the infinite island model dispersal, the probability of genetic identity between two individuals sampled from the same deme after dispersal is given by Wright's measure of population structure, hence $Q_0 = F_{ST} = (1 - m)^2/[1 + (2 - m)m(N - 1)]$. The probability of identity between individuals from different demes is zero ($Q_1 = 0$). Substituting these probabilities of genetic identity and the fitness effects into Hamilton's rule (eqn 4 or 5) and re-arranging we obtain the condition under which the harming trait spreads as

$$\frac{C}{D} < \frac{(1-m)^2}{N(1-(1-m)^2)}. \quad (9)$$

Equation (9) indicates that the selective pressure on harming varies inversely with deme size and migration rate. More precisely, conditions for invasion of a harming trait become more stringent whenever migration rate is high and/or deme size large, requiring a greater cost-to-effect ratio for the trait to be favoured. Migration rate and deme size further determine as what type of trait the harming behaviour must be considered. For instance, for low migration rate (Fig. 2), harming evolves as a selfish trait because it increases the fitness of the focal parent ($-c > 0$) at the expense of other parents ($b_0 < 0$, $b_1 < 0$). With increasing migration rate or deme size, however, harming turns into an altruistic/spiteful trait. It spreads despite a fitness cost to the focal parent and due to the positive effect on the fitness of related parents in the focal deme ($b_0 > 0$), which is obtained at the expense of the fitness of parents of other demes ($b_1 < 0$).

Discussion

In our analysis of the selective pressure on social traits, we focused on the effect of a mutant allele on the actor and all potential recipients in the population. Our analysis emphasizes that whenever a mutant allele resulting in a cost for the actor spreads through kin selection, it does so both by increasing the fitness of relatives and decreasing that of less or unrelated individuals. This implies that any trait that is altruistic towards one or more classes of relatives, is also necessarily spiteful towards one or more classes of less related individuals. Consequently, altruism and spite do not represent different evolutionary forces but are two facets of the same selection pressure acting on a social trait, arising from the trait's direct and indirect effects on the fitness of different members of the population.

The relationship between altruism and spite emerges because we assess the effect of the social act in terms of fitness, not fecundity, and consequently must include all members of the population in our analysis that are potentially affected. We therefore consider any costly trait for the actor that has a positive (negative) effect on a class of individuals as altruistic (spiteful) towards this class. This type of definition is common in the theoretical literature dealing with social traits that do not involve direct interactions between actors and recipients, such as dispersal and resource exploitation (Hamilton & May, 1977; Taylor, 1988; Frank, 1995; Gandon & Michalakis, 1999; Foster, 2004). Different interpretations of altruism and spite stem from looking at different stages of the life-cycle (Fig. 1). Defined from the physical interaction at the adult or juvenile stage (Foster *et al.*, 2000, 2001) or from effects on fecundity (Vickery *et al.*, 2003; Gardner & West, 2004b; Johnstone & Bshary, 2004), spite and altruism are interpreted as being distinct evolutionary phenomena. Although this distinction is probably useful when describing the actual mechanistic effect of the traits

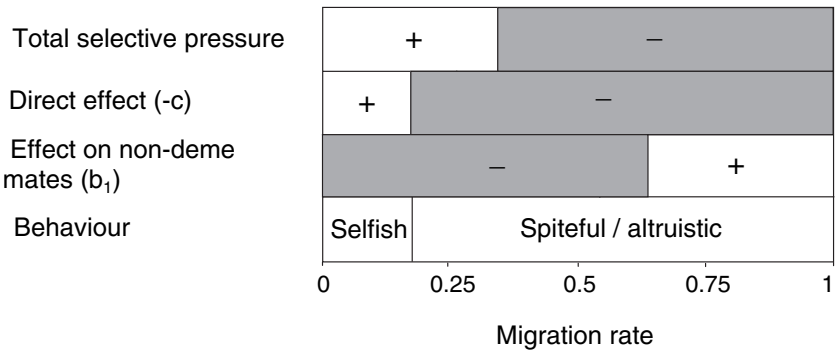

Fig. 2 Signs of effects on fitness of a harming act as a function of the migration rate. The figure shows the signs of the total selective pressure (given by $R_0b_0 - c$ or $R_1b_1 - c$), of the effect of juveniles of the focal parent on its fitness ($-c$) and of the effect of juveniles stemming from parents breeding in different demes on the fitness of the focal parent (b_1). The cost-to-harm ratio is $|C/D| = 0.15$ and the patch size is $N = 5$. Grey shading indicate negative signs, white shading positive signs.

involved, it is far restricted in its application and probably misses out the ecological processes underlying the evolution of both helping and harming traits.

A direct consequence of the relationship between altruism and spite as demonstrated here is that Hamiltonian and Wilsonian spite need not represent different processes. As we have shown above and illustrated in Fig. 1, a harming trait directed against nonkin necessarily results in a positive effect on the fitness of kin. These effects are perfectly compatible both with Hamilton's statement that spite is favoured due to its negative effect on nonkin and with Wilson's interpretation that spite spreads due to its positive effect on kin. In order to be 'Wilsonian', traits are therefore not required to encompass aspects of helping (by directly increasing the fecundity of close relatives) in addition to harming (by directly decreasing the fecundity of nonrelatives) as is sometimes assumed (Gardner & West, 2004b). This of course does not mean that traits involving both harming and helping does not exist. Indeed, conditions for their invasion are likely to be less stringent than those for pure harming traits, due to their dual benefit.

The modelling approach used here (Rousset & Billiard, 2000; Rousset, 2003, 2004) also demonstrates that the sign of the coefficient of relatedness is a natural consequence of the differences of the probabilities of genetic identity. Further, Hamilton's rule usually used for the evolution of altruism (i.e. with positive relatedness, eqn 4) and Hamilton's rule usually used for the evolution of spite (i.e. with negative relatedness, eqn 5) are in fact extracted from the same selective pressure and can be used interchangeably. In particular, negative relatedness arises as the difference between the FI's genetic similarity with the less and the more related class of individuals. Since by definition more related individuals are genetically more similar, this difference is negative.

Our example illustrates how parameters of population structure shape the way harming is selected for. Through their effect on the inclusive fitness effects and the measure of relatedness, the values of migration rate and deme size determine both the exact nature of a trait (selfish vs. altruistic/spiteful) and the direction of selection (Rousset, 2004). For instance, at low migration rate, a harming behaviour qualifies as selfish, because the reduction in competition resulting from harming has a positive effect on the fitness of the focal parent (Fig. 2). At intermediate migration rate or deme size, harming provides a smaller direct benefit to the focal parent which no longer balances the cost of performing the behaviour. The trait is therefore counter-selected at the level of the focal parent alone but can spread due to its positive effect on the fitness of close kin. This result is in line with the view that some degree of population structure is required for harming traits to spread (Gardner & West, 2004b). Increasing migration rate or deme size even further dilutes the indirect benefit that harming provided through its positive effect on the fitness of related parents and the trait is no longer selectively favoured.

The results outlined above illustrate the advantage of constructing models based on clearly defined life cycles. By linking the selective pressure to explicit parameters of the life cycle it is possible to relate the strength and direction of kin selection to the basic life history of an organism, as well as to derive predictions that can be tested in the wild. This assures that the value of relatedness perfectly matches the assumed life-cycle. It thereby takes into account all factors potentially influencing gene frequency change, such as migration and the level of local and/or global drift. Taking these factors into account is crucial when modelling social evolution in finite or geographically structured populations, because their effect on relatedness means that they in part determine the strength and direction of selection (Taylor, 1988, 1992a; Gandon & Michalakis, 1999; Rousset, 2004). By contrast, relatedness can be safely treated as a parameter in family structured populations where the population is assumed infinite and panmictic, a common practice for example in models of sex ratio evolution (e.g. Trivers & Hare, 1975; Boomsma & Grafen, 1991; Reuter & Keller, 2001). In this case, the coefficient of relatedness represents a fixed quantity determined by pedigree relationships.

Our model takes a 'gene-centred' approach, which considers all offspring of a mutant female as a cohesive unit rather than isolated individuals. In doing so, our model analyses a situation in which interactions occur between members of different gene lineages. Our model then predicts that harming will spread whenever few gene lineages compete. This is exactly the situation prevalent in two biological systems recently put forward as examples of spite. The first is the production of soldier larvae in polyembryonic wasps (Gardner & West, 2004a). In these parasitoids, single eggs laid into host caterpillars multiply clonally. Part of the larvae develop into sterile soldiers, the role of which is to damage larvae of other clones present in the same host. The second example is the production of bacteriocins (Gardner *et al.*, 2004), toxins released into the environment by many bacterial species. Although bacterial lineages carrying the bacteriocin gene are immune to the toxin, others are susceptible and will be killed. Both systems are similar in that a harming behaviour has evolved in a situation in which competition occurs between a potentially large number of individuals derived from few founder genotypes. A major difference between these systems and our model, however, lies in the use of kin discrimination. Whereas our model assumes that harm is directed against immigrants but not juveniles born locally, both soldiers larvae and bacteriocins harm any nonkin, that is, all individuals from other gene lineages. Adjusting our model to this assumption (Appendix 2) demonstrates that harming traits directed against any nonkin can be selectively favoured in small, structured populations. However, such harming always increases the fitness of the focal gene lineage. In the gene centred approach, it is consequently considered as selfish.

 doi:10.1111/j.1420-9101.2006.01128.x

Further examples of harming strategies are likely to be found in systems in which similar conditions as described above prevail. Pathogen systems might prove particularly interesting in this respect. First, the evolution of harming in these cases is of medical importance due to its interaction with virulence (Gardner *et al.*, 2004). Secondly, populations of small effective size might be relatively common in these systems. Due to either virulence or host immunity, pathogens live in temporary habitats which, taken together with low rates of successful horizontal transmission, will lead to hosts being infected with only few different pathogen lineages at a time.

The examples of soldier larvae and bacteriocins also illustrate the importance of competition for the evolution of social traits. In parasitoid wasps, the amount of food available is limited by the biomass of the infected host (Giron *et al.*, 2004). Along with cannibalism and siblicide, soldier larvae are only one of many strategies parasitoids have evolved in order to prevail in the face of strong resource competition. Similarly, the evolution of bacteriocins is driven by competition. Accordingly, bacteriocins are only produced when populations enter the stationary phase, i.e. when bacteria are sufficiently dense for competition to occur (Riley & Wertz, 2002). In many cases, the production of toxins relies on a mechanism of quorum-sensing and is induced by factors released by other bacteria, hence indicating the presence of potential competitors (Miller & Bassler, 2001; Riley & Wertz, 2002).

Another behaviour cited as an example of spite is the elimination of males by social insect workers. Colonies of social Hymenoptera (ants, bees, wasps) with a simple family structure (single, once-mated queens) are characterized by sex ratio conflict between the queen, who favours an even sex allocation, and workers, who favour a sex allocation biased towards closely related females. In order to adjust colony investment, workers of some species have been shown to eliminate male brood (e.g. Sundström *et al.*, 1996), an act proposed to be Wilsonian spite (Foster *et al.*, 2001). Here again, competition is crucial for the behaviour to be favoured. Eliminating males only then increases the workers' fitness if food is the resource limiting the colony's productivity. In this situation, eliminating a less related male increases the chance of a more related female to reach adulthood. Indeed, worker manipulation of investment is not favoured if colony productivity is limited by a resource other than food, for example by the queen limiting the number of eggs laid (Bulmer & Taylor, 1981; Reuter & Keller, 2001). The example of workers eliminating males also illustrates the relationship between altruism and spite. Worker sex ratio manipulation has both served as an example of workers acting in favour of their more closely related females (Sundström *et al.*, 1996), as well as an example of worker spite against less related males (Foster *et al.*, 2001).

In conclusion, considering the fitness consequences on all individuals in a population that are affected by a social trait highlights that altruism and spite are two sides of the same coin. Altruism and spite are both implicated in the evolution of social acts resulting in a fitness cost for the actor. Thus, a fundamental difference between behaviours is not whether they are spiteful or altruistic, but whether a social act promotes a gene lineage by augmenting reproduction or reducing competition (helping or harming). Focussing the attention on the demographic and/or ecological conditions favouring either helping or harming has the potential to significantly bolster our understanding of the evolution of social interactions. The insight might even go beyond the field of evolutionary biology, in that what evolutionists call harming bears a close resemblance to what ecologists call interference competition. The occurrence of interference rather than exploitation competition might hence be explained by population structure and size just as selection in favour of harming.

Acknowledgments

We thank Laurent Keller, Michel Chapuisat and three anonymous reviewers for helpful comments on this manuscript. LL acknowledges financial support from the Academy of Finland in the form of a grant to Hanna Kokko. MR acknowledges financial support from the European Commission in form of a Marie Curie Intra-European fellowship. KB was funded by the Finnish School in Wildlife Biology, Conservation and Management.

References

Boomsma, J. & Grafen, A. 1991. Colony-level sex-ratio selection in the eusocial hymenoptera. *J. Evol. Biol.* **4**: 383–407.

Bulmer, M. & Taylor, P.D. 1981. Worker-queen conflict and sex ratio theory in social Hymenoptera. *Heredity* **47**: 197–207.

Foster, K.R. 2004. Diminishing returns in social evolution: the not-so-tragic commons. *J. Evol. Biol.* **17**: 1058–1072.

Foster, K.R., Ratnieks, F.L.W. & Wenseleers, T. 2000. Spite in social insects. *Trends Ecol. Evol.* **15**: 469–470.

Foster, K.R., Wenseleers, T. & Ratnieks, F.L.W. 2001. Spite: Hamilton's unproven theory. *Ann. Zool. Fenn.* **38**: 229–238.

Frank, S.A. 1995. Mutual policing and repression of competition in the evolution of cooperative units. *Nature* **377**: 520–522.

Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press, Princeton, NJ.

Gadagkar, R. 1993. Can animals be spiteful? *Trends Ecol. Evol.* **8**: 232–234.

Gandon, S. & Michalakis, Y. 1999. Evolutionary stable dispersal rate in a metapopulation with extinction and kin competition. *J. Theor. Biol.* **199**: 275–290.

Gardner, A. & West, S.A. 2004a. Spite among siblings. *Science* **305**: 1413–1414.

Gardner, A. & West, S.A. 2004b. Spite and the scale of competition. *J. Evol. Biol.* **17**: 1195–1203.

Gardner, A., West, S.A. & Buckling, A. 2004. Bacteriocins, spite and viru1lence. *Proc. R. Soc. Lond., B, Biol. Sci.* **271**: 1529–1535.
Giron, D., Dunn, D.W., Hardy, I.C.W. & Strand, M.R. 2004. Aggression by polyembryonic wasp soldiers correlates with kinship but not resource competition. *Nature* **430**: 676–679.
Grafen, A. 1984. Natural selection, kin selection and group selection. In: *Behavioral Ecology. An Evolutionary Approach* (J. R. Krebs & N. Davies, eds), pp. 62–84. Blackwell Scientific Publications, Oxford.
Grafen, A. 1985. A geometric view of relatedness. In: *Oxford Survey in Evolutionary Biology* (R. Dawkins & M. Ridley, eds), pp. 28–90. Oxford university press, Oxford.
Hamilton, W.D. 1964a. The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**: 1–16.
Hamilton, W.D. 1964b. The genetical evolution of social behaviour. II. *J. Theor. Biol.* **7**: 17–52.
Hamilton, W.D. 1970. Selfish and spiteful behavior in an evolutionary model. *Nature* **228**: 1218–1220.
Hamilton, W.D. & May, R.M. 1977. Dispersal in stable habitats. *Nature* **269**: 578–581.
Johnstone, R.A. & Bshary, R. 2004. Evolution of spite through indirect reciprocity. *Proc. R. Soc. Lond., B, Biol. Sci.* **271**: 1917–1922.
Keller, L., Milinski, M., Frischknecht, M., Perrin, N., Richner, H. & Tripet, F. 1994. Spiteful animals still to be discovered. *Trends Ecol. Evol.* **9**: 103.
Lehmann, L. & Perrin, N. 2003. Inbreeding avoidance through kin recognition: choosy females boost male dispersal. *Am. Nat.* **162**: 638–652.
Miller, M.B. & Bassler, B.L. 2001. Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**: 165–199.
Reuter, M. & Keller, L. 2001. Sex ratio conflict and worker production in eusocial Hymenoptera. *Am. Nat.* **158**: 166–177.
Riley, M.A. & Wertz, J.E. 2002. Bacteriocins: evolution, ecology, and appli1cation. *Annu. Rev. Microbiol.* **56**: 117–137.
Rousset, F. 2003. A minimal derivation of convergence stability measures. *J. Theor. Biol.* **221**: 665–668.
Rousset, F. 2004. *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton, NJ.
Rousset, F. & Billiard, S. 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J. Evol. Biol.* **13**: 814–825.
Rousset, F. & Ronce, O. 2004. Inclusive fitness for traits affecting metapopulation demography. *Theor. Popul. Biol.* **142**: 1357–1362.
Sundström, L., Chapuisat, M. & Keller, L. 1996. Conditional manipulation of sex ratios by ant workers – a test of kin selection theory. *Science* **274**: 993–995.
Taylor, P.D. 1988. An inclusive fitness model for dispersal of offspring. *J. Theor. Biol.* **130**: 363–378.
Taylor, P.D. 1992a. Altruism in viscous populations – an inclusive fitness model. *Evol. Ecol.* **6**: 352–356.
Taylor, P.D. 1992b. Inclusive fitness in a homogeneous environment. *Proc. R. Soc. Lond., B, Biol. Sci.* **240**: 299–302.
Taylor, P.D. 1996. Inclusive fitness arguments in genetic models of behaviour. *J. Math. Biol.* **34**: 654–674.
Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *J. Theor. Biol.* **180**: 27–37.
Taylor, P.D. & Getz, W.M. 1994. An Inclusive fitness model for the evolutionary advantage of sibmating. *Evol. Ecol.* **8**: 61–69.
Trivers, R. & Hare, H. 1975. Haplodiploidy and evolution of social insects. *Science* **191**: 249–263.
Vickery, W.L., Brown, W.S. & FitzGerald, G.J. 2003. Spite: altruism's evil twin. *Oikos* **102**: 413–416.
Wilson, E.O. 1975. *Sociobiology: The New Synthesis*. Harvard Press, Cambridge, MA.

Appendix 1

Eliminating the effect of the FI on its fitness ($-c$) from the inclusive fitness effect eqn (3) by using $c = b_0 + b_1$, we get after rearrangements

$$\Delta W_{IF} = (Q_0 - Q_\bullet)b_0 + (Q_1 - Q_\bullet)b_1. \qquad (10)$$

Substituting into this equation $b_1 = -b_0(Q_0 - Q_\bullet)/(Q_1 - Q_\bullet)$ solved from $b_0R_0 = b_1R_1$ we find that $S = 0$. A trait yielding such an inclusive fitness effect is neutral.

Appendix 2

We can modify our model by assuming that the offspring of the focal parent use their discrimination capacity in order to direct harm against all nonsiblings (i.e. juveniles born from a parent other than their own). In this case, the effect of the mutant trait on the fitness of the focal parent is unaffected by the change in the recipients of harm and still given by eqn (6). In contrast, the effect of the mutant trait on the fitness of other parents in the focal deme (eqn 7) now has to incorporate the cost of being harmed by the focal family of magnitude $-D(1 - m)^2$ and the effect on parents in other demes changes accordingly. The conditions under which harming juveniles from other parents spreads is

$$\frac{C}{D} < \frac{(1-m)^2}{N}. \qquad (11)$$

This inequality can be satisfied only for parameter values that results in a positive effect on the fitness of the focal parent ($-c > 0$). Accordingly, in a 'gene-centred' perspective, such a trait is to be considered as selfish.

Appendix 3

In their original demonstration of eqn (2), Rousset & Billiard (2000) showed that the fitness effects sum up to zero when fitness is assessed in a population of constant size. In this appendix, this assumption is relaxed in order to include variations in the size of the population resulting from stochasticity (be it intrinsic or environmental). The fluctuations in population size may themselves depend on the trait under selection. For simplicity, we consider only a panmictic population but the argument is general. We assume that conditional on nonextinction, the demography of such a population follows a Markov chain with state space (1, 2, 3, …), which determines the transition between the number of adults in the parental generation (N) and the number of adults in the offspring generation (N'). In such a

 doi:10.1111/j.1420-9101.2006.01128.x

population, the fitness of an individual can be decomposed into elements $w(N',N)$ measuring its expected number of juveniles reaching adulthood in a population of given size N' when he is breeding in a population of size N (Rousset, 2004; Rousset & Ronce, 2004). These elements thus measure the fitness of an individual conditional on the realization of the pair of demographic states N and N'. In order to get the total fitness of an individual, the elements $w(N',N)$ are weighted by the probabilities of the occurrence of the demographic states and the reproductive values of the offspring in such states. Accordingly, the inclusive fitness effect (eqn 1) will be a complex function and accordingly, it will be cumbersome to directly show that the property of eqn (2) holds. However, we can get around this by showing that conditional on the pair of demographic states (N',N), the effects of all actors in the population on the fitness component $w(N',N)$ of an individual sum up to zero. Consider a monomorphic population, that is, a population where all individuals bear the same trait value. Then, the fitness component $w(N',N)$ is a constant and thus independent of trait values. This can be seen by noting that for any trait value, the probability that a gene taken at random in the offspring generation of size N' descends from any of the N individuals in the parental generation must be equal to one, and this can be expressed as $1 = \frac{N}{N'} w(N', N)$, where $1/N'$ is the probability of picking the gene and $w(N',N)$ is the fitness of any of the N parents. It must then be true that

$$w(N', N) = \frac{N'}{N}. \tag{12}$$

The total derivative of this function with the respect to trait value is thus equal to 0. This derivative can be expanded with the chain rule to sum up the effects of each class of actors (under the form of partial derivatives) on the fitness of a FI (see also Chapter 6 of Rousset, 2004 and Rousset & Billiard, 2000). Accordingly, conditional on the realization of a transition from states N to N', the effects of all actors on the fitness component $w(N',N)$ sum up to zero. As this holds for all demographic states (whose probabilities of occurrence might depend on trait value), the total effect of actors on the fitness of a FI sum up to zero.

Received 25 January 2006; revised 1 March 2006; accepted 3 March 2006