doi:10.1002/evl3.150

Kin discrimination, negative relatedness, and how to distinguish between selfishness and spite

Matishalin Patel,[1,2] Stuart A. West,[1] and Jay M. Biernaskie[3]

[1]*Department of Zoology, University of Oxford, Oxford OX1 3SZ, United Kingdom*

[2]*E-mail: matishalin.patel@sjc.ox.ac.uk*

[3]*Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, United Kingdom*

Received August 24, 2019
Accepted November 18, 2019

Spiteful behaviors occur when an actor harms its own fitness to inflict harm on the fitness of others. Several papers have predicted that spite can be favored in sufficiently small populations, even when the harming behavior is directed indiscriminately at others. However, it is not clear that truly spiteful behavior could be favored without the harm being directed at a subset of social partners with relatively low genetic similarity to the actor (kin discrimination, causing a negative relatedness between actor and harmed recipient). Using mathematical models, we show that (1) the evolution of spite requires kin discrimination; (2) previous models suggesting indiscriminate spite involve scenarios where the actor gains a direct feedback benefit from harming others, and so the harming is selfish rather than spiteful; (3) extreme selfishness can be favored in small populations (or, more generally, under local competition) because this is where the direct feedback benefit of harming is greatest.

KEY WORDS: **Competition, harming, inclusive fitness, kin selection, social evolution, super-territory, territory size.**

Impact summary

Spite is the hardest type of social trait to explain because it involves an individual harming itself (reducing its own Darwinian fitness) to inflict harm on others. It has always been thought that spite should be rare because organisms will usually harm others for some feedback benefit for themselves or their offspring (e.g., easier access to food or mates)—in other words, most harming traits are selfish rather than spiteful. It has been argued that truly spiteful harming can be favored if it is directed specifically at less-genetically related group members (nonkin) and ultimately benefits more-related group members (kin). However, there is also a persistent idea that spite directed indiscriminately at others could evolve in sufficiently small populations. For example, some have predicted that animals should hold "super-territories" to spitefully exclude others from resources. Using mathematical models, we show that (1) the evolution of spite requires kin discrimination; (2) previous models suggesting indiscriminate spite involve scenarios where the harming individual gains a direct feedback benefit, and so the harming is selfish rather than spiteful; (3) extreme selfishness, like holding super-territories, can be favored in small populations (and in small groups with local competition) because this is where the feedback benefit of harming is greatest. Overall, we examine how to model natural selection acting on harming traits in order to distinguish between selfishness and spite.

Spite is the hardest type of social trait to explain. Spiteful behavior reduces the lifetime fitness of both the recipient and the performer (actor) of that behavior (Hamilton 1970). In terms of Hamilton's rule, $-C + RB > 0$, spite represents the case where there is a fitness cost to the actor (positive C) and a fitness cost to the harmed recipient (negative B), which can only be favored if the genetic relatedness term, R, is negative. Understanding the meaning of negative relatedness is therefore crucial for explaining how and why spite evolves.

It has been argued that the evolution of spite requires kin discrimination, allowing the actor to direct harm toward a

Evolution Letters

subset of individuals with whom they share relatively low genetic similarity (Wilson 1975; Foster et al. 2000, 2001; Gardner et al., 2004, 2007; Gardner and West 2004a,b, 2006; Lehmann et al. 2006; West and Gardner 2010). Specifically, spite can be favored when harming the less-similar individuals in a social group (primary recipients) reduces competition and therefore benefits the unharmed individuals (secondary recipients). In this case, negative relatedness arises because the actor's genetic similarity to primary recipients is less than its genetic similarity to secondary recipients (Gardner and West 2004a,b; Lehmann et al. 2006; Gardner et al. 2007). In contrast, without kin discrimination, harming behaviors could not be directed at individuals to whom the actor is negatively related, so indiscriminate spite should be impossible.

However, a number of theoretical studies have suggested the possibility for indiscriminate spite. Hamilton (1970) originally suggested that if genetic similarity is measured relative to the entire population (including the actor), then there will be a negative relatedness between the actor and all others in the population, especially when the population is small. Consequently, several papers have predicted that spiteful harming, directed indiscriminately at others, could be favored in sufficiently small populations (Hamilton 1970, 1971; Grafen 1985; Vickery et al. 2003; Taylor 2010; Smead and Forber 2012). As a specific example, Verner (1977) and Knowlton and Parker (1979; Parker and Knowlton 1980) suggested that individuals could be favored to hold territories that are larger than needed for their own interest ("super-territories") in order to spitefully exclude others from resources. It is not clear, though, whether such indiscriminate harming traits are truly spiteful.

Here, we resolve this disagreement over indiscriminate spite. Many harming traits will be costly to primary recipients ($B < 0$) but provide a direct fitness benefit to the actor, because they reduce competition for the actor or its offspring. Consequently, the traits are selfish ($-C > 0$) rather than spiteful ($-C < 0$) (Hamilton 1970; Keller et al. 1994; Foster et al. 2001; West and Gardner 2010). We address the possibility that indiscriminate harming traits like territory size have been misclassified as spiteful when they are actually selfish (Colgan 1979; Tullock 1979). Our specific aims are to: (1) determine generally whether indiscriminate harming evolves as a spiteful or a selfish trait; (2) examine how different modeling approaches can change the meaning of negative relatedness and lead to misclassification of harming traits; (3) re-analyze Knowlton and Parker (1979) as an example to illustrate the different modeling approaches and to resolve whether super-territories are truly spiteful.

Harming Traits

We first modeled natural selection acting on a harming trait, following the approach of Lehmann et al. (2006). The trait has a fitness effect on a focal actor ($-C$) and on two categories of recipients: the harmed primary recipients and the unharmed secondary recipients who benefit from reduced competition (fitness effects B_1 and B_2, respectively). We define an individual's fitness as its number of offspring that survive to adulthood (not simply the number of offspring produced), which is consistent with other definitions used for classifying social traits (Hamilton 1964; Rousset 2004; Lehmann et al. 2006; West et al. 2007). We assume that fitness effects on the actor, primary recipients, and secondary recipients must sum to zero because of competition for finite resources (Rousset and Billiard 2000):

$$-C + B_1 + B_2 = 0, \tag{1}$$

implying that any decrease in fitness for one category necessarily means an increase in fitness for another. Our model could apply to any finite population of constant size or to a local "economic neighborhood" (Queller 1994) in which there is a zero-sum competition for access to the next generation. Key examples of such local competition include polyembryonic wasps competing for resources inside a host (Gardner and West 2004a; Gardner et al. 2007), male fig wasps competing for females inside a fig (West et al. 2001), or bacteria competing for local resources (Gardner et al. 2004).

To predict the direction of natural selection acting on the harming trait, we considered the fate of a mutant harming allele in a population of individuals with a fixed, resident genotype. The success of the mutant allele depends on its "inclusive fitness effect" (Hamilton 1964): the sum of effects from a focal actor's mutant trait on its own fitness and on the total fitness of each recipient category, weighted by their genetic similarity with the actor. Under the usual assumptions of weak selection and additive gene action, the inclusive fitness effect for our model is

$$\Delta W_{\mathrm{IF}} = -C + B_1 Q_1 + B_2 Q_2, \tag{2}$$

where Q_1 and Q_2 are probabilities of sharing identical genes between the focal actor and a random individual from the primary and secondary recipients, respectively. We note that the fitness effects in equation (2) could alternatively be weighted by relatedness coefficients, where genetic similarity is measured with respect to a reference population (e.g., $R_i = \frac{Q_i - \bar{Q}}{1-\bar{Q}}$, where $\bar{Q}$ is the average genetic similarity in the entire population, including the actor; Hamilton 1970). However, doing this would not change any of the results given below, so we prefer the simpler approach that follows from equation (2).

In the following sections, we examine two different ways of defining the category of secondary recipients and therefore partitioning the fitness effects of harming. Both methods correctly predict the direction of selection (they give the same sum as in eq. (2)). The first partitioning also maintains complete separation of direct and indirect fitness effects ($-C$ and RB,

respectively), making it appropriate for classifying harming traits as selfish ($-C > 0$) or spiteful ($-C < 0$). We therefore propose that the first partitioning presented below—which may at first seem unconventional—is best for the purpose of classifying harming traits. In contrast, the second partitioning—which may be seen as the more conventional approach—actually obscures the separation of direct and indirect fitness effects, making it inappropriate for classifying harming traits.

IS INDISCRIMINATE HARMING SPITEFUL OR SELFISH?

We determined the conditions for a harming trait to be classified as spiteful or selfish. For this purpose, we assume that the focal actor, primary recipients, and secondary recipients are mutually exclusive categories. This ensures that the actor is not a recipient of its own behavior, and so the $-C$ term in the inclusive fitness effect (eq. (2)) captures all effects of the actor's harming behavior on its own fitness. From equation (2), we derived the typical two-party version of Hamilton's rule by eliminating the fitness effect on secondary recipients, using $B_2 = C - \mathrm{B}_1$ (from eq. (1)). After rearrangement, the inclusive fitness effect is positive, and the harming trait is favored, when

$$-C + \frac{Q_1 - Q_2}{1 - Q_2} B_1 > 0, \tag{3}$$

which is Hamilton's rule with the relatedness between actor and primary recipients given by $\frac{Q_1 - Q_2}{1 - Q_2} \equiv R_1$. This is the genetic similarity between the actor and an individual from the potential primary recipients, measured relative to an individual from the potential secondary recipients.

Equation (3) implies that indiscriminate spite cannot evolve. This is because negative relatedness (and hence an indirect fitness benefit of harming) will arise only if harm can be directed at primary recipients who are less genetically similar to the actor than secondary recipients are ($Q_1 < Q_2$). In contrast, if the actor were harming others indiscriminately—for example, harming a random subset of a population or local economic neighbourhood—then its expected similarity to these primary recipients would be the same as to the set of potential secondary recipients ($Q_1 = Q_2$), and relatedness would be zero ($R_1 = 0$). This implies that indiscriminate harming will be favored when it is a selfish trait with a positive direct fitness benefit ($-C > 0$).

WHY DOES MISCLASSIFICATION OCCUR?

Misclassification of harming traits can occur because the fitness effects of social traits can be partitioned in different ways (Frank 1998). An alternative way of partitioning the effects of harming is to include the actor in the set of secondary recipients who may benefit from reduced competition. In fact, it is often implicitly assumed that the set of potential secondary recipients is the entire population (or economic neighborhood), including the focal actor (Hamilton, 1970, 1971; Grafen 1985; Vickery et al. 2003; Taylor 2010; Smead and Forber 2012). To make this explicit, we re-write the inclusive fitness effect as

$$\Delta W_{\mathrm{IF}} = -c + b_1 Q_1 + b_2 \bar{Q}, \tag{4}$$

using lowercase letters to indicate that the fitness effects no longer match those from equation (2). In particular, b_2 is now the benefit of reduced competition that may be experienced by all individuals in population (including the actor), and $\bar{Q}$ is the probability of genetic identity between the focal actor and a random individual from the entire population (including itself). It follows that $-c$ is not a total direct fitness effect because it excludes the secondary benefit of harming that feeds back to the focal actor (increased direct fitness due to reduced competition; Fig. 1).

We used equation (4) to derive an analogue of Hamilton's rule, which reveals a different version of negative relatedness. For example, in a population (or economic neighborhood) of N individuals, an actor could indiscriminately harm a random subset of individuals with genetic similarity Q_1 to the actor. If the entire population is in the set of secondary recipients, then the expected genetic similarity between the actor and these recipients is $\bar{Q} = \frac{1}{N} 1 + \frac{N-1}{N} Q_1$ (where the first term accounts for the actor's similarity to itself). Eliminating the fitness effect on secondary recipients (using $b_2 = c - b_1$) shows that indiscriminate harming is favored when

$$-c + \frac{-1}{N - 1} b_1 > 0, \tag{5}$$

where $-1/(N - 1)$ is the relatedness between actor and primary recipients, measured with respect to the entire population ($\frac{Q_1 - \bar{Q}}{1 - \bar{Q}} \equiv R_{1,\mathrm{p}}$). This is the version of negative relatedness that has led to predictions of indiscriminate spite in small populations (e.g., Hamilton 1971; Grafen 1985).

However, although the term $\frac{-1}{N-1} b_1$ resembles an indirect fitness benefit ($RB > 0$), it also incorporates the secondary fitness benefit of harming that feeds back to the focal actor. This can be made more explicit by deriving an analogue of Hamilton's rule from equation (4), this time eliminating the fitness effect on primary recipients (using $b_1 = c - b_2$). For example, in a well-mixed population of N individuals, indiscriminate harming is favored when

$$-c + \frac{1}{N} b_2 > 0, \tag{6}$$

where $1/N$ is the relatedness between the actor and the entire population (including itself), measured with respect to primary recipients ($\frac{\bar{Q} - Q_1}{1 - Q_1} \equiv R_{2,\mathrm{p}}$). The term $(1/N)b_2$ accounts for the fraction of the secondary benefit (reduced competition) that feeds back to the focal actor, which gets larger as the actor makes up a larger fraction of the population (as N declines).

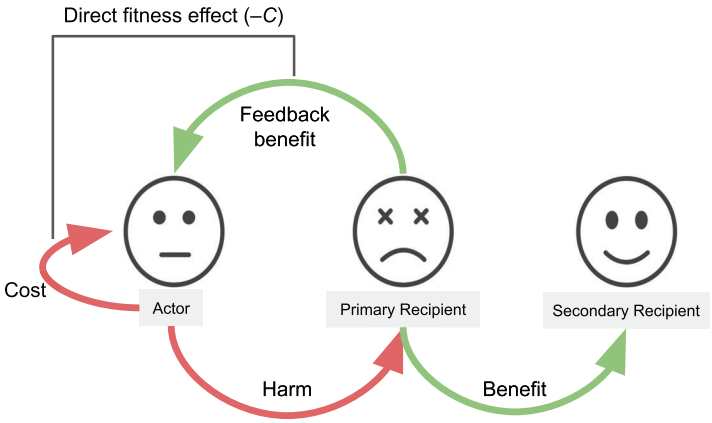

Figure 1. **Partitioning the fitness effects of a harming trait. When a focal actor harms a primary recipient, this reduces competition and may therefore benefit the unharmed secondary recipients and the actor itself ("feedback benefit"). Some modeling approaches include the actor in the set of secondary recipients of the harming trait. However, the total direct fitness effect (–*C* in Hamilton's rule) includes the fecundity cost of expressing the harming trait plus the feedback benefit.**

Our key distinction here is that harming behaviors can be either beneficial or costly to the actor ($-C > 0$ or $-C < 0$), whereas spiteful behaviors are strictly costly to the actor ($-C < 0$). We showed that indiscriminate harming is always favored because it is beneficial to the actor—it has a positive effect on the actor's number of surviving offspring ($-C > 0$). Moreover, indiscriminate harming can be favored most in small populations (or small economic neighborhoods) because this is where the focal actor can benefit most from the reduced competition that results from its harming behavior.

Revisiting "Super-Territories"

We next re-examined the territory size model from Knowlton and Parker (1979) and Parker and Knowlton (1980). We first analyzed the model to fully separate direct and indirect fitness effects (applying eq. (2)), asking whether the model predicts selfish behavior, as expected. We then used the alternative approach (applying eq. (4)) to illustrate why previous studies have interpreted territory size as a spiteful trait.

We considered a finite, deme-structured population ("island model") with d demes (assuming $d > 1$) and n individuals competing for territory in each deme (total population size is $N = dn$). Individuals that secure a territory have offspring and then die before a fraction m of their offspring disperse independently to a random deme in the entire population. All individuals have a genetically determined strategy for the size of territory that they try to obtain. Taking over a larger territory has three effects: (1) it incurs a fecundity cost for the actor (we assume a linear cost with increasing trait size, with slope $-a$ and $a \in [0,1]$; Parker and Knowlton (1980) consider more complex cost functions, with no change to qualitative predictions); (2) it harms the actor's deme mates by taking resources away and reducing their fecundity; (3) it reduces the competition faced by all remaining offspring in the population to secure a territory in the next generation.

We first assumed that the actor, primary recipients, and secondary recipients are mutually exclusive categories (as in eq. (2)). In the Appendix, we derive an expression for the fitness, W, of a focal actor. This is a function of the focal actor's strategy, x (a continuous number of territory units that it attempts to gain; $x > 0$); the average strategy of the actor's deme mates (primary recipients), y; and the average strategy in all other demes (secondary recipients), z. We used this "neighbor-modulated" fitness function to derive the inclusive fitness effect, by taking partial derivatives with respect to the strategies of the different categories of individuals (Taylor and Frank 1996; Rousset and Billiard 2000):

$$\begin{aligned}\Delta W_{\mathrm{IF}} &= \frac{\partial W}{\partial x} + \frac{\partial W}{\partial y} Q_1 + \frac{\partial W}{\partial z} Q_2, \\ &= -C + B_1 Q_1 + B_2 Q_2 \end{aligned} \tag{7}$$

where all partial derivatives are evaluated in a monomorphic population ($x = y = z$). We derive expressions for Q_1 and Q_2 in the Appendix, and with these we determined the equilibrium of the model ($\hat{z}$, where directional selection stops) by solving $\Delta W_{\mathrm{IF}} = 0$. We also checked that the equilibrium is a convergence-stable strategy, denoted z^*, meaning that if the population is perturbed from the equilibrium then natural selection will push it back ($\frac{\mathrm{d}\Delta W_{\mathrm{IF}}}{\mathrm{d}z}\big|_{z=\hat{z}=z^*} < 0$).

We found that the equilibrium of our model, $z^* = 1/(aN)$, is identical to that originally predicted by Parker and Knowlton (1980); however, our analysis shows that the optimal territory-size strategy is selfish rather than spiteful. Territory size cannot be spiteful in this model because the actor's genetic

similarity to individuals in other demes is always equal to or less than the similarity to deme mates ($Q_1 \geq Q_2$). Accordingly, the relatedness to primary recipients (measured relative to secondary recipients) is never negative ($R_1 \geq 0$), and so there is no indirect benefit of larger territory size. Moreover, when offspring dispersal is limited ($m < 1$) and deme mates are positively related ($R_1 > 0$), there is no indirect benefit of smaller territory size (as a form of helping). This is because limited dispersal increases competition among offspring within the deme, which promotes harming and exactly cancels the effect of positive relatedness (Taylor 1992; Queller 1994). Territory size therefore evolves for its direct benefit only, with larger territories promoted by a smaller fecundity cost to the actor (smaller a) and smaller population size (smaller N). Specifically, the direct fitness effect at equilibrium ($z = z^*$) is

$$-C = \frac{a\,(d-1)\,d(m-1)^2}{N-1}, \tag{8}$$

which is either positive (when $m < 1$) or zero (when $m = 1$). In the case of full offspring dispersal ($m = 1$), the equilibrium is the point where the fecundity cost to the actor is exactly balanced by the feedback benefit experienced by its offspring (reduced competition for space in the next generation). As the population approaches this equilibrium, however, direct fitness is always positive ($-C > 0$), confirming that territory size evolves as a selfish trait (Fig. 2).

We next assumed that the set of secondary recipients is the entire population, including the focal actor (as in eq. (4)). In this case, the inclusive fitness effect is

$$\begin{aligned}\Delta W_{\mathrm{IF}} &= \frac{\partial W}{\partial x} + \frac{\partial W}{\partial y} Q_1 + \frac{\partial W}{\partial z_{\mathrm{p}}} \bar{Q}, \\ &= -c + b_1 Q_1 + b_2 \bar{Q}\end{aligned} \tag{9}$$

where z_{p} is the average territory size strategy in the entire population (including the focal actor), and all partial derivatives are evaluated at $x = y = z_{\mathrm{p}}$. As expected, solving for the equilibrium of equation (9) gives the same answer as before, $z^* = 1/(aN)$.

This version of the model shows, however, how territory size could be misclassified as spiteful. For example, in a fully mixing population at the equilibrium ($m = 1$; $z_{\mathrm{p}} = z^*$), the first term in equation (9) is

$$-c = -\frac{aN}{N-1}, \tag{10}$$

which is always negative. This term reflects the fecundity cost of the focal actor's territory size strategy; however, it is not the total direct fitness effect because it excludes the feedback benefit experienced by the actor's offspring (reduced competition). As noted above, when $m = 1$ this feedback benefit should exactly balance the fecundity cost at equilibrium. Following equations (5) and (6), we can calculate the feedback benefit as $(-1/[N-1])b_1$ or $(1/N)b_2$ (both evaluated at $z_{\mathrm{p}} = z^*$), which gives the expected result, $aN/(N-1)$. The partitioning in equation (9) therefore splits the total direct fitness effect of territory size into two separate terms, $-c + (-1/[N-1])b_1$ or $-c + (1/N)b_2$, which could be misinterpreted as a direct fitness cost ($-C < 0$) and an indirect fitness benefit ($RB > 0$).

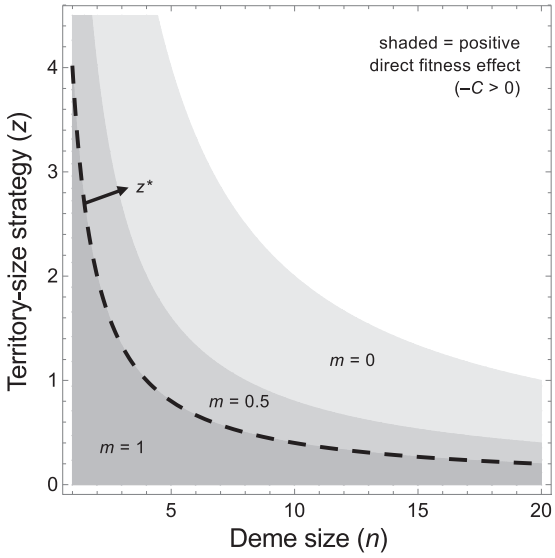

Figure 2. Territory size and direct fitness. Larger territory size is promoted by smaller population size (smaller *dn*) and reduced offspring migration from the deme (smaller *m*), both of which increase the direct benefit to an actor for harming its deme mates. However, reduced migration also increases the relatedness among deme mates, which inhibits larger territory size. Ultimately, the optimal territory size strategy (*z, dashed line) is independent of migration rate and evolves as if the population were fully mixed (*m* = 1). Other parameters used were as follows: *d* = 5, *a* = 0.05.**

Discussion

We examined both an illustrative model of harming traits and a specific scenario for territory size. In both models, we found that (1) the evolution of spite requires kin discrimination, where the actor harms only a subset of other individuals (those with relatively low genetic similarity); (2) without kin discrimination, harming can be favored but only when there is a sufficient direct, feedback benefit to the actor (reduced competition for the actor or its offspring); (3) indiscriminate harming is more favored in small populations (or small economic neighborhoods), where the direct feedback benefit to the actor is greatest; (4) previous studies have misclassified indiscriminate harming as spite, partly because they misinterpret the direct feedback benefit as an indirect (kin-selected) benefit ($R_1 B_1 > 0$). Overall, we illustrate why indiscriminate harming traits are selfish rather than spiteful, and how to model harming traits to distinguish between selfishness and spite.

CLASSIFYING HARMING TRAITS

For the purposes of classifying harming traits, we found that it is easiest to treat the actor, primary recipients, and secondary recipients as separate categories. This makes it straightforward to separate the total direct and indirect fitness effects of harming ($-C$ and R_1B_1, respectively) and ensures that non-zero relatedness will always be associated with an indirect fitness effect. For example, spiteful harming ($-C < 0$, $B_1 < 0$) requires that harm is directed at primary recipients to whom the actor is negatively related (with respect to secondary recipients; $Q_1 < Q_2$ and $R_1 < 0$), resulting in a positive indirect fitness effect ($R_1B_1 > 0$) (Lehmann et al. 2006; Gardner et al. 2007). In contrast, when harming is indiscriminate, the actor has zero relatedness to primary recipients (with respect to secondary recipients; $Q_1 = Q_2$ and $R_1 = 0$), and so harming can be favored as a selfish trait only ($-C > 0$, $B_1 < 0$).

We showed that misclassification of indiscriminate harming is due to an assumption that the secondary benefit of harming that returns to the focal actor (feedback benefit) is an indirect rather than direct benefit (Hamilton, 1970, 1971; Grafen 1985; Vickery et al. 2003; Taylor 2010; Smead and Forber 2012). This means that some of the actor's direct benefit of harming has been accounted for by a fraction of the fitness effects on recipients, giving the appearance of an indirect benefit. For example, in a group of N individuals, where all individuals (including the actor) are considered secondary recipients, a fraction of the fitness effect on primary recipients ($-1/[N-1]B_1$) actually accounts for the direct feedback benefit of indiscriminate harming.

Others have suggested that harming traits should be classified based on their primary effects only, rather than their total fitness effects (Krupp 2013). This means that indiscriminate harming traits like larger territory size, which may be associated with a survival or fecundity cost ($-c < 0$ in the terms of our model), would be classified as spiteful, despite the feedback benefit to the focal actor. We argue, however, that a classification based on total effects to the actor and primary recipients ($-C$ and B_1) is more useful (Hamilton 1964; West et al. 2007). This is because it emphasizes the fundamental distinction between spiteful harming, which is favored by indirect fitness benefits and requires kin discrimination, versus selfish harming, which is favored by direct fitness benefits and does not require kin discrimination (West and Gardner 2010). Similar arguments have been made for maintaining the distinction between helping traits that may be altruistic ($-C < 0$, $B_1 > 0$) or mutually beneficial ($-C > 0$, $B_1 > 0$) (West et al. 2007).

INDISCRIMINATE HARMING IN NATURE

We found that selfish indiscriminate harming can be favored most under local competition (e.g., small populations or small economic neighborhoods). This is because harming primary recipients leads to reduced competition for all individuals in the population or group, and a focal actor receives a larger fraction of this secondary benefit when it makes up a larger fraction of the population or group. Indiscriminate harming can therefore be thought of as producing a type of public good for secondary recipients (Tullock 1979), analogous to indiscriminate helping, which is often thought of as a public good for primary recipients. A key difference is that indiscriminate helping is inhibited by local competition (Taylor 1992; Griffin et al. 2004); in contrast, indiscriminate harming requires local competition so that the focal actor can actually benefit from the reduced competition that results from its harming (Gardner and West 2004b).

So where can we expect to find the most extreme examples of selfish harming? As recognized by Hamilton (1970), very small populations will tend to extinction, so harming traits in these populations are unlikely to be observed. But examples of extreme selfishness should also be found in small groups with relatively local competition, such that harming other individuals significantly reduces competition for the actor. One potential example is in fig wasps, where males fight for access to females, and—as our model predicts—the intensity of fighting increases sharply as the number of males in the fig declines (Murray 1989; West et al. 2001; Reinhold 2003). Fig wasp fighting has been used as a potential example of spite, but if kin discrimination is absent, then it fits better as an example of extreme selfishness, which is similarly promoted by localized competition (Gardner and West 2004b). Other potential examples include competition among female honey bees for a colony and situations where males engage in local competition for mates (e.g., *Melittobia* parasitoids; Griffin and West 2002). Our analyses suggest that, for all of these cases, it will be crucial to distinguish between the direct and indirect benefits of harming others.

ACKNOWLEDGMENTS

We thank Guy Cooper, Daniel Krupp, Asher Leeks, Alan Grafen, and Tom Scott for comments on the manuscript. MP was supported by a studentship from the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/M011224/1), and JMB was supported by a fellowship from the BBSRC (BB/M013995/1).

AUTHOR CONTRIBUTIONS

All authors conceived and designed the study, MP drafted the initial version of the manuscript, and all authors contributed to later versions of the manuscript.

LITERATURE CITED

Colgan, P. 1979. Is a super-territory strategy stable? Am. Nat. 144:604–605.

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

Foster, K. R., T. Wenseleers, and F. Ratnieks. 2001. Spite: Hamilton's unproven theory. Anna. Zool. Fennici 38:229–238.

Frank, S. A. 1998. Foundations of social evolution. Princeton Univ. Press, Princeton, NJ.

Gardner, A., and S. A. West. 2004a. Spite among Siblings. Science 305:1413–1414.
———. 2004b. Spite and the scale of competition. J. Evol. Biol. 17:1195–1203.
———. 2006. Spite. Curr. Biol. 16:R662–R664.
Gardner, A., S. A. West, and A. Buckling. 2004. Bacteriocins, spite and virulence. Proc. Roy. Soc. Lond. B 271:1529–1535.
Gardner, A., I. C. W. Hardy, P. D. Taylor, and S. A. West. 2007. Spiteful soldiers and sex ratio conflict in polyembryonic parasitoid wasps. Am. Nat. 169:519–533.
Grafen, A. 1985. A geometric view of relatedness. Oxford Surv. Evol. Biol. 2:28–90.
Griffin, A. S., and S. A. West. 2002. Kin selection: fact and fiction. Trends Ecol. Evol. 17:15–21.
Griffin, A. S., S. A. West, and A. Buckling. 2004. Cooperation and competition in pathogenic bacteria. Nature 430:1024–1027.
Hamilton, W. D. 1964. The genetical evolution of social behaviour. I and II. J. Theor. Biol. 7:1–52.
———. 1970. Selfish and spiteful behaviour in an evolutionary model. Nature 228:1218–1220.
———. 1971. Selection of selfish and altruistic behaviour in some extreme models. Pp. 57–91 *in* J. F. Eisenberg and W. S. Dillon, eds. Man and beast: comparative social behavior. Smithsonian Press, Washington, DC.
Keller, L., M. Milinski, M. Frischknecht, N. Perrin, H. Richner, and F. Tripet. 1994. Spiteful animals still to be discovered. Trends Ecol. Evol. 9:103–113.
Knowlton, N., and G. A. Parker. 1979. An evolutionarily stable strategy approach to indiscriminate spite. Nature 279:419–421.
Krupp, D. B. 2013. How to distinguish altruism from spite (and why we should bother). J. Evol. Biol. 26:2746–2749.
Lehmann, L., K. Bargum, and M. Reuter. 2006. An evolutionary analysis of the relationship between spite and altruism. J. Evol. Biol. 19:1507–1516.
Murray, M. G. 1989. Environmental constraints on fighting in flightless male fig wasps. Anim. Behav. 38:186–193.
Parker, G. A., and N. Knowlton. 1980. The evolution of territory size—some ESS models. J. Theor. Biol. 84:445–476.
Queller, D. C. 1994. Genetic relatedness in viscous populations. Evol. Ecol. 8:70–73.
Reinhold, K. 2003. Influence of male relatedness on lethal combat in fig wasps: a theoretical analysis. Proc. Roy. Soc. Lond. B 270:1171–1175.
Rousset, F. 2004. Genetic structure and selection in subdivided populations. Princeton Univ. Press, Princeton NJ.
Rousset, F., and S. Billiard. 2000. A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. J. Evol. Biol. 13:814–825.
Smead, R., and P. Forber. 2012. The evolutionary dynamics of spite in finite populations. Evolution 67:698–707.
Taylor, P. D. 1992. Altruism in viscous populations—an inclusive fitness model. Evol. Ecol. 6:352–356.
———. 2010. Birth-death symmetry in the evolution of a social trait. J. Evol. Biol. 23:2569–2578.
Taylor, P. D., and S. A. Frank. 1996. How to make a kin selection model. J. Theor. Biol. 180:27–37.
Taylor, P. D., A. J. Irwin, and T. Day. 2000. Inclusive fitness in finite deme-structured and stepping-stone populations. Selection 1:153–164.
Tullock, G. 1979. On the adaptive significance of territoriality: comment. Am. Nat. 113:772–775.
Verner, J. 1977. On the adaptive significance of territoriality. Am. Nat. 111:769–775.
Vickery, W. L., J. S. Brown, and G. J. FitzGerald. 2003. Spite: altruism's evil twin. Oikos 102:413–416.
West, S. A., and A. Gardner. 2010. Altruism, spite, and greenbeards. Science 327:1341–1344.
West, S. A., M. G. Murray, C. A. Machado, A. S. Griffin, and E. A. Herre. 2001. Testing Hamilton's rule with competition between relatives. Nature 409:510–513.
West, S. A., A. Griffin, and A. Gardner. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. 20:415–432.
Wilson, E. O. 1975. Sociobiology. Harvard Univ. Press, Cambridge, MA.

Associate Editor: A. Gardner

Appendix: Territory-size model

Deriving the fitness function

Here, we derive an expression for the fitness of a focal actor with a mutant territory size strategy, based on the models of Knowlton and Parker (1979; Parker and Knowlton 1980). We consider a population that is structured into d demes of n individuals competing for territories, where each deme has A units of available territory. The focal actor's strategy, x, represents a continuous number of territory units that it attempts to gain ($x > 0$). The average strategy of the actor's deme mates is y, and the average strategy in all other demes is z.

We first calculate the expected offspring production (expected fecundity, F) for the focal actor, an individual in the actor's deme, and an individual in another deme. These expected values depend on: (1) the probability of an individual acquiring a territory (assuming that available spaces are acquired completely randomly); (2) the cost associated with the individual's strategy (assuming fecundity declines linearly with increasing territory size strategy; $f(x) = 1 - ax$, where $0 < a < 1$). For the focal actor, there are A/y spaces available in the deme, and we use the simplifying assumption that a mutant individual has priority to claim the territory units denoted by its strategy (Knowlton and Parker 1979). Therefore, the focal actor has a $1/n$ probability of acquiring a territory, and its expected fecundity is

$$F_x = \frac{1}{n}\frac{A}{y}f(x). \tag{A1}$$

The space available for others in the patch depends on whether or not the focal actor claims a territory. The actor gains access to the patch with probability A/ny, and in this case $(A - x)/y$ spaces remain; otherwise, A/y spaces are available. The expected fecundity for one of the $n - 1$ deme mates of the focal actor is therefore

$$F_y = \frac{1}{n-1}\left(\frac{A}{ny}\frac{A-x}{y}f(y) + \left(1 - \frac{A}{ny}\right)\frac{A}{y}f(y)\right). \tag{A2}$$

Finally, for an individual in another deme in the population, there are A/z spaces available, and so the expected fecundity for one of these individuals is

$$F_z = \frac{1}{n}\frac{A}{z} f(z). \tag{A3}$$

We next calculate the focal actor's fitness, $W(x, y, z)$, which is the number of its offspring that survive to compete for a territory in the next generation. This can be partitioned into two terms, the first term accounting for offspring that compete on the focal actor's natal deme (those that did not disperse, with probability $1 - m$, and those that dispersed but landed on the natal deme, with probability m/d) and the second term accounting for offspring that disperse with probability m to compete in the $d - 1$ non-natal demes:

$$W = \frac{\left(1-m+\frac{m}{d}\right)F_x}{(1-m)F_x+(n-1)(1-m)F_y+\frac{1}{d}(mF_x+(n-1)mF_y)+\frac{d-1}{d}nmF_z} + \frac{\frac{d-1}{d}mF_x}{(1-m)nF_z+\frac{1}{d}\left(mF_x+(n-1)mF_y\right)+\frac{d-1}{d}nmF_z}, \tag{A4}$$

where the denominator of the first and second terms account for, respectively, all offspring competing in the focal actor's natal deme and all offspring competing in any other deme in the population. Equation (A4) is the fitness function used to calculate the inclusive fitness effect in Equation 7 of the main text. To express the focal individual's fitness in terms of x, y, and z_p (the average territory size strategy in the entire population, including the focal individual), we substituted $(x + (n - 1)y - dnz_p)/(n - nd)$ for z in Equation (A4). This gives the fitness function used to calculate the inclusive fitness effect in Equation 9 of the main text.

Deriving probabilities of genetic identity

Next, we derive probabilities of genetic identity by descent in a finite deme-structured population, following the approach of Taylor et al. (2000). In particular, we needed the probability of identity between the focal actor and a randomly selected deme mate (Q_1), between the actor and a randomly selected individual in another deme (Q_2), and between the actor and a randomly selected individual in the entire population (including itself), defined as

$$\bar{Q} = \frac{1}{d}\left(\frac{1}{n} + \frac{n-1}{n}Q_1\right) + \frac{d-1}{d}Q_2. \tag{A5}$$

The remaining probabilities of identity are given by the following recursive equations:

$$Q_1 = \left((1-m)^2\left(\frac{1}{n} + \frac{n-1}{n}Q_1\right) + \left(1-(1-m)^2\right)Q_2^{\mathrm{p}}\right)(1-u)^2 \tag{A6}$$

$$Q_2 = \left((1-m)^2 Q_2 + \left(1-(1-m)^2\right)Q_2^{\mathrm{p}}\right)(1-u)^2, \tag{A7}$$

where u is the "contrived mutation rate" from Taylor et al. (2000). We solved Equations (A5)–(A7) simultaneously and linearised to the first order around the point $u = 0$, giving:

$$Q_1 = 1 - 2dnu \tag{A8}$$

$$Q_2 = 1 + \left(\frac{2d(m-1)^2}{(m-2)m} - 2dn\right)u \tag{A9}$$

$$\bar{Q} = 1 + \frac{2\left(d\left(1-(m-2)m(n-1)\right)-1\right)}{(m-2)m}u. \tag{A10}$$

These are the probabilities of genetic identity used in Equations 7 and 9 of the main text.