JOURNAL OF Evolutionary Biology
eseb European Society for Evolutionary Biology

doi: 10.1111/jeb.12253

SHORT COMMUNICATION

How to distinguish altruism from spite (and why we should bother)

D. B. KRUPP*†‡

*Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada
†Department of Psychology, Queen's University, Kingston, ON, Canada
‡Program in Evolution of Governance, One Earth Future, Broomfield, CO, USA

Keywords:

adaptation;
fitness partitions;
function;
inclusive fitness;
social evolution.

Abstract

Social behaviour is often described as altruistic, spiteful, selfish or mutually beneficial. These terms are appealing, but it has not always been clear how they are defined and what purpose they serve. Here, I show that the distinctions among them arise from the ways in which fitness is partitioned: none can be drawn when the fitness consequences of an action are wholly aggregated, but they manifest clearly when the consequences are partitioned into primary and secondary (neighbourhood) effects. I argue that the primary interaction is the principal source of adaptive design, because (i) it is this interaction that determines the fit of an adaptation and (ii) it is the actor and primary recipients whom an adaptation foremost affects. The categories of social action are thus instrumental to any account of evolved function.

The concepts of altruism, spite, selfishness and mutual benefit have served as signposts to adaptive design since the development of inclusive fitness theory (Hamilton, 1963, 1964, 1970). Throughout this time, however, they have been mired in confusion. Altruism has alternately been defined by the action's effects on lifetime fitness or by its immediate effects on fecundity and survival, at the level of the population or of the group, and with or without regard to the actor's costs (reviewed in Grafen, 1984; Lehmann *et al.*, 2006; West *et al.*, 2007). The definition of spite has faced similar complications (Foster *et al.*, 2001; Gardner *et al.*, 2007). It is not my aim here to rehash these arguments, which I regard as well described (West *et al.*, 2007). Instead, I mean to deliberate on the influence of another facet of methodology: the ways in which fitness is partitioned. I suggest that, although there are numerous ways to carve up the effects of social behaviour on fitness, some bear more fruit than others.

As a recent example, Lehmann *et al.* (2006) have shown that altruism and spite can be conceived of as close counterparts; that is, the same action can be fairly described as altruistic in one sense and spiteful in another. Their point is well taken: an increase in the fitness of some individuals usually entails a decrease in the fitness of others, because resources are finite. As I show below, however, their conclusion rests on a choice of fitness partitions. Indeed, there are many equivalent ways of organizing the effects of social behaviour on fitness, each leading to the same result when an action is successful: an increase in the representation of the allele causing it. Nevertheless, there is an historical, widely practiced approach whose logic is especially relevant to biological problems. It is my purpose here to illuminate this logic.

The relevance of fitness partitions

Consider the quintessential case of an allele that causes an action in an inelastic, haploid, asexual, homogeneous population and assume that all fitness effects are linear and small. Inelastic populations – ones that cannot shrink or grow – closely resemble those facing strict ecological limitations. For current purposes, however, they also pose the most difficult condition for the argument that social actions can be distinguished. This is because the restrictions they impose on population fitness imply that benefits accruing to some individuals must be paid by levying costs on others (*cf.* Lehmann *et al.*, 2006; see also Grafen, 1984; Taylor, 1992a,b; Grafen & Archetti, 2008). I work with the inclusive fitness

Correspondence: Daniel B. Krupp, One Earth Future, 525 Zang Street, Suite C, Broomfield, CO 80021, USA.
Tel.: +1 303 533 1715; fax: 613 533 2964;
e-mail: dbkrupp@oneearthfuture.org

approach and track the lifetime fitness effects of an action across the entire population, a practice that coincides with Hamilton's (1964, 1970), for reasons outlined by Rousset (2004), Lehmann *et al.* (2006) and West *et al.* (2007), among others. In the analysis, a focal actor bearing a causal allele performs an action that affects the direct fitness of K individuals, each by a total (i.e. counting *all* direct fitness effects) amount a_k, where k denotes the kth affected individual. Importantly, the actor, too, belongs to the set of K individuals.[1]

Under these conditions, an action will be favoured by selection when $\sum_k r_k a_k > 0$, where r_k is the coefficient of consanguinity between the actor and the kth affected individual. The population size is fixed, so any action that raises the fitness of some individuals must lower the fitness of others by a commensurate amount, or $\sum_k a_k = 0$. Thus, it is impossible to characterize the nature of the act; it is neither helpful nor harmful, because it is equally both. Even the actor is lost in this analysis, hidden as it is among the K individuals affected by the action. By consequence, no social action can be distinguished from any other: the terms altruism, spite, selfishness and mutual benefit have no meaning here.

The set of K affected individuals can, however, be partitioned into subsets. Lehmann *et al.* (2006) used an actor-recipient partition, in which the actor's total effects on itself are treated separately from its total effects on the $J = K–1$ recipients. Indexing the actor as individual 0 and the jth recipient of the action by j, the condition for which selection favours an action is $\sum_j r_j a_j + a_0 > 0$ and the corresponding restriction on total fitness effects imposed by an inelastic population is $\sum_j a_j = -a_0$. Because the actor has been identified, a distinction can at this stage be drawn between the 'self-serving' actions that increase the actor's total direct fitness ($a_0 > 0$) and the 'sacrificial' actions that do not ($a_0 < 0$). Nevertheless, they cannot be described as helpful or harmful to others, because their effects on the recipients are again of both kinds: in a population of constant size, help to the jth recipient ($a_j > 0$) that surpasses the total amount spent by the actor ($a_j > –a_0$) must be compensated by harm to another class and *vice versa* (Grafen, 1984; Taylor, 1992a,b; Queller, 1994; Lehmann *et al.*, 2006; Grafen & Archetti, 2008). As Lehmann *et al.* (2006) found, altruism folds into spite and, by the same token, so too does selfishness into mutual benefit.

Still, there is a more traditional, widely practiced partition (e.g. Grafen, 1984; Taylor, 1992a,b, 2010; West *et al.*, 2002; Taylor *et al.*, 2007; Grafen & Archetti, 2008; West & Gardner, 2010; Taylor & Maciejewski, 2012), one that further differentiates 'primary' from 'secondary' recipients by way of the causal sequence of an action's effects. In it, the actor and primary recipients engage in an interaction that has direct fitness consequences for them. If the positive and negative consequences of this primary interaction do not cancel exactly, there will result a change in competition among the secondary recipients (which may include some or all of the individuals involved in the primary interaction), driven by a surfeit or deficit in the number of surviving offspring or breeding vacancies. The secondary recipients thus comprise the 'economic neighbourhood' (Grafen, 1984; Queller, 1994) in which the downstream consequences of the primary interaction take hold. For instance, when reproduction is continuous and births 'drive' mortality (a Moran process with a birth-death protocol; Taylor, 2010), an actor may incur a primary fecundity cost to provide a larger primary fecundity benefit to a set of recipients. Subsequently, the excess fecundity resulting from the primary interaction causes a secondary decrease in both the fecundity and survivorship of the neighbourhood. The actor and primary recipients thereby feel the effects of the action, whereas the secondary recipients feel the effects of the effects.

Hence, it is possible to unpack a partial, primary effect $–c$ from the total effect of an action on the actor's direct fitness a_0. Likewise, a partial, primary effect of an action b_i on the ith primary recipient can be freed from the total effect a_i for each of the $I \leq J$ primary recipients. As is well known (e.g. Grafen, 1984; Queller, 1994; West *et al.*, 2002), this partition recasts the condition for which selection favours an action as

$$\sum_i r_i b_i - c - \bar{r}\left(\sum_i b_i - c\right) > 0, \qquad (1)$$

where $\bar{r}$ is the actor's average coefficient of consanguinity to the neighbourhood. On the left hand side of inequality (1), the first two terms ($\sum_i r_i b_i - c$) capture the primary effects and the last term ($\bar{r}\big(\sum_i b_i - c\big)$) captures the secondary effects of the action on the actor's inclusive fitness. Following Queller (1994), inequality (1) simplifies to $\sum_i R_i b_i - c > 0$, where $R_i = \frac{r_i - \bar{r}}{1 - \bar{r}}$ is the genetic relatedness of the actor to the ith primary recipient, recovering a common expression of Hamilton's rule.

Critically, nothing in the preceding analysis implies that $\sum_i b_i = c$, because it is the total effects that must balance in an inelastic population, not the primary effects. Consequently, an action that provides a primary benefit to the ith primary recipient ($b_i > 0$) that exceeds the primary cost spent by the actor ($b_i > c$) can be described as helpful without implying that it is harmful to other primary recipients. So it is possible to distin-

---

[1] Note that the analysis pertains, nevertheless, to the individual rather than to the 'trait group' level (Pepper, 2000).

Table 1 The four categories of social behaviour.

		Primary effect on recipients ($\sum_i b_i$)	
		+	–
Primary effect on actor (–*c*)	+	Mutual benefit	Selfishness
	–	Altruism	Spite

guish an action in terms of its joint primary effects on the actor ($-c > 0$ or $-c < 0$) and on the primary recipients ($\sum_i b_i > 0$ or $\sum_i b_i < 0$). This obtains the familiar categorization scheme of social behaviour found in Table 1. It is perhaps worth noting that the net primary effect $\sum_i b_i$ could be replaced by the average primary effect $\bar{b}_i = \frac{\sum_i b_i}{I}$ without altering the scheme in any meaningful way.

The significance of the primary interaction

Table 1 should be understood strictly in the context of primary effects: the effects of the primary interaction on the direct fitness of the actor and primary recipients; hence, the appearance of $-c$ and $\sum_i b_i$. Nowhere does it make reference to secondary effects, total effects, all recipients, consanguinity, relatedness, or inclusive fitness. Although the total effects must be counted in any inclusive fitness analysis, only that subset of direct fitness resulting from the primary interaction on those who partake in it merits a behavioural distinction. This may seem arbitrary, but it is not.

The aim of the biological enterprise is to understand the living world – not merely the evolutionary history of an adaptation but also the functional purpose for which it was designed (Gardner, 2009). The individuals involved in the primary interaction take centre stage because it is to this audience that an adaptation performs. The primary effects of an action depend on the degree of fit between the actor's behaviour and the primary recipients' responses, whereas the secondary effects do not. Rather, the secondary effects depend on the residual changes in competition that result from the primary effects. When the total recipients outnumber the primary recipients ($J > I$), these residual competitive effects are distributed more broadly and therefore more thinly over the neighbourhood. Moreover, the actor and primary recipients feel the effects of the action twice if they are included among the secondary recipients. Therefore, the actor and the average primary recipient feel the effects of the action more strongly than does the average secondary recipient. Secondary recipients are thus bit players in the design of an adaptation, although they remain necessary to any account of its evolution.

As Gardner *et al.* (2007) and West & Gardner (2010) have pointed out, there are obvious distinctions between helping and harming behaviours: the literature is replete with reports of organisms equipped with adaptations to feed, protect, teach and shelter on the one hand, and to poison, battle, undermine and evict on the other. For instance, the pathogenic bacterium *Pseudomonas aeruginosa* produces siderophores (iron-scavenging molecules) and bacteriocins (toxins), the former to the benefit and the latter to the detriment of the primary recipients. If we assume that the production of these compounds imposes a lifetime fitness cost on the producer, then – categorizing behaviour by the primary effects – siderophore production is altruistic and bacteriocin production is spiteful. Production of these compounds is consequently expected to evolve under different population regulation regimes: siderophore production rising as competition becomes increasingly global and neighbours increasingly related (Griffin *et al.*, 2004), bacteriocin production rising as competition becomes increasingly local and neighbours intermediately related (Inglis *et al.*, 2009, 2011). Total aggregation of the primary and secondary effects, however, cannot differentiate siderophore from bacteriocin production because, in an inelastic population, both result in help to some and harm to others. Accordingly, this approach is not suited to predicting the respective population regulation regimes responsible for their evolution. With respect to questions of adaptive design, the practical benefits to partitioning the primary and secondary effects are ample (Gardner *et al.*, 2007; West & Gardner, 2010).

The significance of the primary-secondary partition is easily overlooked when the mechanisms of population regulation and structure are not fully described. Hidden assumptions of population elasticity may be commonplace, for instance, and in such cases, the categorization of a social action will seem evident: helpful actions increase recipient fitness and harmful actions impede it. However, even here, actions are implicitly categorized by their primary effects; secondary effects may obtain in elastic populations, but they are generally set aside in descriptions of social action. In any event, limited elasticity will often be apposite, because populations cannot grow indefinitely and because selection favours one allele over others only when there is competition among them (Lehmann *et al.*, 2006).

The defining criteria of the four categories of social action suggested above (the joint deviations of $-c$ and $\sum_i b_i$ from 0) closely resemble but nonetheless depart from Hamilton's own. Grafen & Archetti (2008) interpret Hamilton's definitions as turning on total rather than primary fitness effects, although these effects are still restricted to the individuals involved in the primary interaction. As much is clear from Hamilton's (1964, p. 15) discussion of the 'conservation' of fitness between actor and primary recipient, as his categorization of social actions was indifferent to secondary effects. The difference between the current scheme and Hamilton's is minor, however, and may be thought of as a slight practical improvement: as Grafen & Archetti (2008) point out, the primary effects take the same sign as the total effects but are simpler to track theoretically and to measure empirically.

The effects of (evolutionarily successful) social actions are always the same: an increase in the representation of the allele causing the action as a result of an increase in the fitness of positive relatives and a concomitant decrease in the fitness of negative ones (Hamilton, 1970; Gardner *et al.*, 2007; West & Gardner, 2010). On this basis, it could be argued that describing a particular action as altruistic or spiteful is a matter of personal taste and does not reflect the actual evolutionary processes involved. However, this confuses function with explanation. Despite sharing the same fundamental cause, the distinctions among altruism, spite, selfishness and mutual benefit are sharp because they reflect adaptive design. Partitioning fitness into primary and secondary effects reveals the distinctions among the social actions precisely and so exposes this design more fully.

Acknowledgments

I am indebted to Peter Taylor for sharing his thoughts on this problem with me over the course of several years. I thank Clay Cressler, Samir Okasha, Carly Rozins, Peter Taylor and Stuart West for many helpful suggestions and comments on the manuscript, and I acknowledge financial support from NSERC and SSHRC.

References

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

Gardner, A. 2009. Adaptation as organism design. *Biol. Lett.* **5**: 861–864.

Gardner, A., Hardy, I.C.W., Taylor, P.D. & West, S.A. 2007. Spiteful soldiers and sex ratio conflict in polyembryonic parasitoid wasps. *Am. Nat.* **169**: 519–533.

Grafen, A.. 1984. Natural selection, kin selection and group selection. In: *Behavioural Ecology* (J.R. Krebs & N.B. Davies, eds), pp. 62–84. Blackwell Scientific, Oxford.

Grafen, A. & Archetti, M. 2008. Natural selection of altruism in inelastic viscous homogeneous populations. *J. Theor. Biol.* **252**: 694–710.

Griffin, A.S., West, S.A. & Buckling, A. 2004. Cooperation and competition in pathogenic bacteria. *Nature* **430**: 1024–1027.

Hamilton, W.D. 1963. The evolution of altruistic behavior. *Am. Nat.* **97**: 354–356.

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

Hamilton, W.D. 1970. Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**: 1218–1220.

Inglis, R.F., Gardner, A., Cornelis, P. & Buckling, A. 2009. Spite and virulence in the bacterium *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci.* **106**: 5703–5707.

Inglis, R.F., Roberts, P.G., Gardner, A. & Buckling, A. 2011. Spite and the scale of competition in *Pseudomonas aeruginosa*. *Am. Nat.* **178**: 276–285.

Lehmann, L., Bargum, K. & Reuter, M. 2006. An evolutionary analysis of the relationship between spite and altruism. *J. Evol. Biol.* **19**: 1507–1516.

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

Queller, D.C. 1994. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**: 70–73.

Rousset, F.. 2004. *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton.

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. Ser. B.* **249**: 299–302.

Taylor, P. 2010. Birth-death symmetry in the evolution of a social trait. *J. Evol. Biol.* **23**: 2569–2578.

Taylor, P. & Maciejewski, W. 2012. An inclusive fitness analysis of synergistic interactions in structured populations. *Proc. R. Soc. Ser. B.* **279**: 4596–4603.

Taylor, P.D., Day, T. & Wild, G. 2007. Evolution of cooperation in a finite homogeneous graph. *Nature* **447**: 469–472.

West, S.A. & Gardner, A. 2010. Altruism, spite, and greenbeards. *Science* **327**: 1341–1344.

West, S.A., Pen, I. & Griffin, A.S. 2002. Cooperation and competition between relatives. *Science* **296**: 72–75.

West, S.A., Griffin, A.S. & Gardner, A. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**: 415–432.

Received 2 July 2013; accepted 23 August 2013