

Joint evolution of multiple social traits: a kin selection analysis

Sam P. Brown^{1,*} and Peter D. Taylor²

¹Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

²Department of Mathematics and Statistics, Queen's University, Kingston, Ontario K7L 3N6, Canada

General models of the evolution of cooperation, altruism and other social behaviours have focused almost entirely on single traits, whereas it is clear that social traits commonly interact. We develop a general kin-selection framework for the evolution of social behaviours in multiple dimensions. We show that whenever there are interactions among social traits new behaviours can emerge that are not predicted by one-dimensional analyses. For example, a prohibitively costly cooperative trait can ultimately be favoured owing to initial evolution in other (cheaper) social traits that in turn change the cost–benefit ratio of the original trait. To understand these behaviours, we use a two-dimensional stability criterion that can be viewed as an extension of Hamilton's rule. Our principal example is the social dilemma posed by, first, the construction and, second, the exploitation of a shared public good. We find that, contrary to the separate one-dimensional analyses, evolutionary feedback between the two traits can cause an increase in the equilibrium level of selfish exploitation with increasing relatedness, while both social (production plus exploitation) and asocial (neither) strategies can be locally stable. Our results demonstrate the importance of emergent stability properties of multidimensional social dilemmas, as one-dimensional stability in all component dimensions can conceal multidimensional instability.

Keywords: inclusive fitness; public goods; exploitation; altruism; cooperation; selfishness

1. INTRODUCTION

From humans to microbes, social dilemmas (Sachs *et al.* 2004; Lehmann & Keller 2006; West *et al.* 2007) interact and intertwine, with individuals simultaneously constructing, defending and exploiting a diverse array of shared public goods. The most well-known areas of interacting social traits concern investments in cooperation and in the enforcement of cooperation (via various mechanisms, termed punishment, policing or sanctions; Frank 2003; Gardner & West 2004; Fehr & Gächter 2002). For example, in many species of eusocial hymenoptera, worker insects must navigate interacting investment decisions over whether to forgo direct reproduction (a potentially altruistic act, yielding reproductive costs to the actor and benefits to the hive) and whether to eat the 'illegal' eggs of other workers (a policing act, reducing the cost/benefits of worker egg-laying). A number of theoretical studies predict that policing is favoured under conditions of low relatedness, acting to maintain cooperation via enforcement; apparently altruistic traits become favoured in the presence of policing, as the alternative selfish actions are associated with punitive sanctions. As a result, cooperation can even increase with decreasing relatedness (Frank 1995, 2003; Wenseleers *et al.* 2004). Comparative studies of worker laying and policing in ants, bees and wasps support these conclusions (Wenseleers & Ratnieks 2006).

Another established field of study of multiple social trait evolution concerns altruism and dispersal. A large

body of theory highlights the dependence of altruism on levels of dispersal, mediated both by relatedness and local competition (Hamilton 1970; Taylor 1992; West *et al.* 2002; Gardner & West 2006; Platt & Bever 2009). In return, dispersal can be shaped by levels of altruism, mediated in particular by habitat saturation (Hamilton & May 1977; Le Galliard *et al.* 2003). Theoretical models allowing for the joint evolution of both altruism and dispersal have predicted a range of outcomes, spanning both positive and negative correlations among the two traits (Perrin & Lehmann 2001; Le Galliard *et al.* 2005) and even the emergence of multicellularity (Pfeiffer & Bonhoeffer 2003; Hochberg *et al.* 2008). Finally, the evolution of discriminatory altruism has also been approached in a multi-trait context, with conditional helping behaviours co-evolving with potentially discriminatory markers or 'tags' (Axelrod *et al.* 2004; Jansen & van Baalen 2006; Rousset & Roze 2007).

Whereas cooperation/enforcement, altruism/dispersal and altruism/discrimination have received significant attention and generated intriguing results (and further specific examples of social trait pairings have been studied, e.g. cooperation and signalling, Brown & Johnstone 2001; belligerence and bravery, Lehmann & Feldman 2008), there has been to our knowledge no general study of social evolution in multiple dimensions. Here we treat multidimensional dilemmas as a general social phenomenon, and ask in the most general terms, what happens if social evolution occurs in a multi-dimensional trait space? We present a flexible model framework for the study of any multidimensional social dilemma, among relatives or non-relatives.

Our approach is to construct an adaptive dynamics system (Dieckmann & Law 1996) based on the inclusive

* Author for correspondence (sam.brown@zoo.ox.ac.uk).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2009.1480> or via <http://rsob.royalsocietypublishing.org>.

fitness effects of the different traits. With the simplest possible assumptions (for example, traits determined at a single locus, linkage equilibrium, weak selection, additive gene action and small mutation rate) the inclusive fitness effect provides a measure of allele frequency change (Rousset & Billiard 2000; Taylor *et al.* 2007), and the system can reasonably provide the evolutionary trajectories of the set of traits under the action of selection. When these assumptions fail (as is almost always the case), the dynamical equations become complex and intractable (see Roze & Rousset 2008; for a multilocus model of social behaviour that incorporates recombination) so that our system is put forward as a simple approximation to actual evolutionary trajectories.

We illustrate our framework with a number of specific examples (the production and exploitation of a public good; the production of multiple public goods; investment in multiple selfish traits), revealing a number of surprising results owing to multi-trait feedbacks (e.g. more selfishness at higher relatedness; positive runaway public goods production at low relatedness; ‘either-or’ bistable interactions among selfish traits at low relatedness). A general result is that in two-dimensional models, Hamilton’s classic rule (Hamilton 1964) for selection on each trait separately (appendix A, equation A 4) requires a further ‘interaction’ condition (appendix A, equation A 6) to adequately describe the two-dimensional system dynamics.

2. RESULTS

(a) *Production and exploitation*

We first illustrate our approach using a generic model of the joint evolution of the production and exploitation of a public good. For example, humans can share pastures, simultaneously enhancing them with the addition of fertilizer and exploiting them with the addition of livestock (the exploitation trait being the focus of the classic ‘tragedy of the commons’ of Hardin 1968). Analogously, microbes can enhance supplies of accessible extracellular iron through the secretion of scavenging siderophore molecules to bind insoluble ferric Fe(III), and then exploit this newly accessible iron with the expression of appropriate surface receptors to uptake iron bound to siderophores (Crosa *et al.* 2004; Griffin *et al.* 2004).

Consider two social traits: a cooperative, constructive trait x that builds a shared public good G at a direct cost to the actor (e.g. investment in fertilizer or siderophores), and a selfish, exploitative trait y that diminishes the public good and provides a direct benefit to the recipient (e.g. investment in livestock or siderophore-binding receptors). By looking at the joint evolution of x and y , we consider the simultaneous navigation of a tragedy of the commons (driven by exploitation, y) and a public goods or collective action dilemma (driven by contributions, x). A fitness function of the form $W(x_0, y_0, X, Y) = y_0 G(X, Y) - C(x_0, y_0)$ captures these relationships, where x_0 and y_0 are the focal individual’s level of production and exploitation of the public good, and X and Y are the same traits averaged across the focal individual’s group or neighbourhood. The direct benefits of exploitation are proportional to the density G of the public good, which in turn depends on the prevailing local levels of production and exploitation, while C captures

the direct costs of investment in x and y (the assumption that exploitation is an active, costly process is necessary to generate intermediate levels of investment in exploitation).

Hamilton’s rule (appendix A, equation A 4) offers a simple heuristic for understanding when selection would favour an increase in either trait, based on the inclusive fitness effects H and K of increases in x and y , respectively. Specifically, given within-group relatedness R and no genetic correlation between x and y , the inclusive fitness effects of an increase in x and in y can be calculated as:

$$H(x, y) = \frac{\partial W}{\partial x_0} + R \frac{\partial W}{\partial X}, \quad (2.1a)$$

and

$$K(x, y) = \frac{\partial W}{\partial y_0} + R \frac{\partial W}{\partial Y}, \quad (2.1b)$$

with all derivatives evaluated at $x_0 = X = x$ and $y_0 = Y = y$ (inclusive of fitness effects calculated via the neighbour-modulated fitness technique; see Taylor & Frank 1996; Wenseleers *et al.* in press; appendix A, equations A 1–A 2). Hamilton’s rule predicts that, considered as independent traits (holding other traits invariant), high relatedness should favour both cooperative production (high x) and restraint (low y); conversely, low relatedness should favour both free-riding (low x) and rapaciousness (high y), reflecting two distinct ways to be a social cheat (Brown *et al.* 2002). However, this independence assumption fails under many realistic scenarios.

Figure 1 illustrates the joint dynamics of public goods investment x and exploitation y for an explicit form of the above model, chosen to demonstrate the qualitative features that can emerge owing to simple ecologically mediated fitness interactions among social traits. Figure 1a illustrates two joint equilibria (where the nullclines intersect), one stable and the other unstable owing to the strong trait interdependence at this point (changes in x have too large an impact on the equilibrium value of y , and *vice versa*). Condition A 6 in appendix A formalizes and generalizes this distinction into a threshold condition for the stability of a social equilibrium in two dimensions. This condition has the form $H_x K_y > H_y K_x$ (subscripts denote partial derivatives: $H_x = \partial H / \partial x$, etc.). Here, H_x and K_y measure the strength of selection on each trait at equilibrium. In a sense they measure the restoring force for departures from equilibrium in x and y . On the other hand the cross-terms H_y and K_x measure the ‘disturbing’ effect of each variable on the equilibrium of the other. More precisely, the quotient $-H_y/H_x$ measures the effect of changes in y on the equilibrium value of x (x^*). For example, if $-H_y/H_x = 2$, a small increase in y will cause an increase in x^* of twice the amount. Similarly, the quotient $-K_x/K_y$ measures the effect of changes in x on y^* . Condition A6 in appendix A can be interpreted as requiring that the joint restoring force exceed the joint disturbing effect. Worked examples are provided in the supporting online materials.

The existence of the unstable equilibrium in figure 1a introduces bistability into the system dynamics. Depending on the initial conditions of x and y , evolution proceeds towards either a stable intermediate level of both production and exploitation, or towards no social

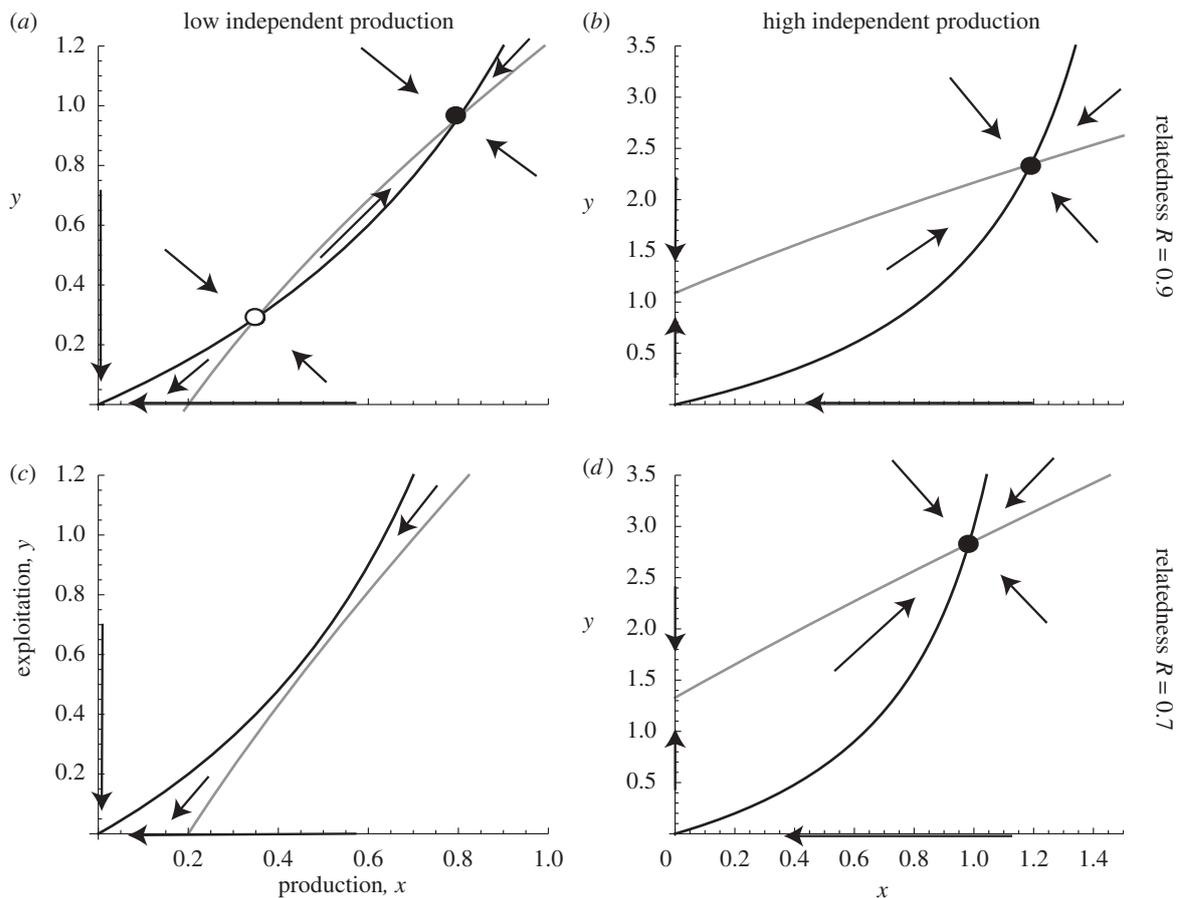


Figure 1. Joint evolution of production (x) and exploitation (y) of a shared public good, G . The generic fitness function $W(x_0, y_0, X, Y) = y_0 G(X, Y) - C(x_0, y_0)$ is specified with the public goods function $G(X, Y) = (\beta + X)/(kY + \alpha)$ and the cost function $C(x_0, y_0) = k_x x_0^2 - k_y y_0$. We assume that the ecological dynamic of the public good G can be approximated by $dG/dt = \beta + X - (kY + \alpha)G$, where β and α are the rates of independent public goods production and degradation, respectively, and k weighs the deleterious impact on G of exploitation. The equilibrium value of G is the formula we have used above. Grey curves are exploitation nullclines ($K = 0$, appendix A), revealing y -equilibria as a function of fixed production x . Black curves are production nullclines ($H = 0$), x -equilibria as a function of fixed y . All nullclines are convergence stable with the other variable fixed. Nullcline intersections mark points of joint evolutionary equilibria; filled circles are convergence stable, open circles unstable. Parameters: $\alpha = 0.6$, $k = k_x = k_y = 0.5$. (a,c) Low independent public goods production, $\beta = 0.1$, (b,d) high independent production $\beta = 1$. (a,b) Relatedness $R = 0.9$, (c,d) $R = 0.7$.

interaction with both x and y driven to zero (a similar result is implicit in Gardner & West 2004, where the condition for maintenance of punishment and cooperation is less strict than the condition for invasion of punishment and cooperation). This ‘asocial’ boundary outcome is locally stable, since when a public good is not exploited ($y = 0$), there is no benefit to its production, and when there is no production of the public good (and it is not sufficiently generated by any independent process), there is no reward for investment in the mechanisms of exploitation. The more ‘social’ outcome of intermediate x and y is also locally stable because of the same synergistic interactions—when x is large, there are significant rewards to positive y , and *vice versa*. As a consequence of the bistability in figure 1a (and more generally, whenever any social equilibrium fails the stability test in condition A6), there will be regions of parameter space in which a small change in initial conditions (for example in either siderophore production or reception) can lead to a large change in the final evolutionary outcome—a highly social endpoint on the one hand, and an asocial one on the other.

Siderophore-mediated social interactions can have a particularly strong synergy owing to the obligate

co-dependence of siderophore molecules and siderophore receptors (each are liabilities in the absence of the other—receptors are a particular liability in the absence of siderophores, as they then serve only to uptake toxic bacteriocins; Crosa *et al.* 2004). If, however, the public good is also generated in significant quantities by other independent forces, then the basin of attraction for the asocial endpoint can disappear (figure 1b)—for instance it may pay to place sheep on a pasture (positive exploitation), even if it has never been fertilized (zero production).

The stable and unstable evolutionary equilibria given high relatedness ($R = 0.9$) are illustrated in figure 1a by filled and open circles, respectively. In figure 2a we track these equilibria (x^* , y^*) as a function of relatedness R , varied from 0 to 1. As expected from an analysis of the x trait in isolation, increasing R leads to an increase in the stable equilibrium x^* . However, contrary to the analysis of y in isolation, increasing R also leads to an increase in the stable equilibrium level y^* of selfish exploitation; that is, we see maximal selfishness at high relatedness. The reason for this effect is again the synergistic coupling between x and y —less production at low R means less

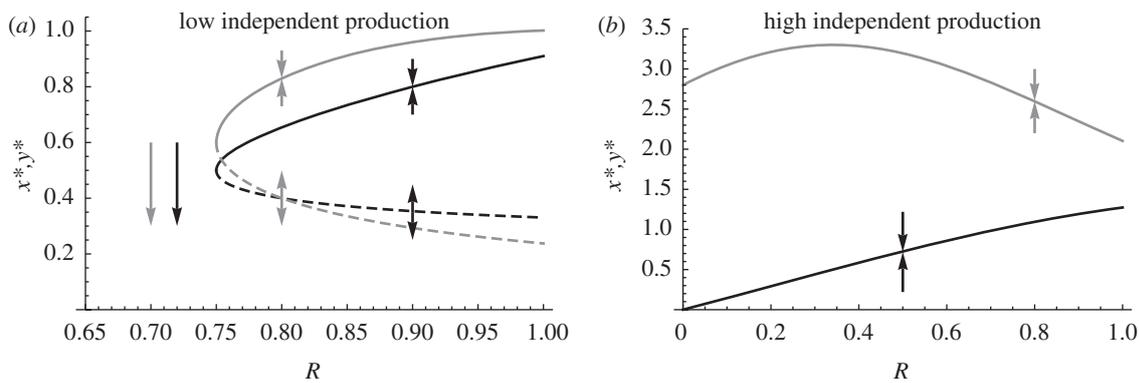


Figure 2. Equilibrium production (x^* , black) and exploitation (y^* , grey) of a shared public good, G , as a function of relatedness $R_x = R_y = R$. Solid lines are attractors (equivalent to filled circles in figure 1), dotted lines are repellers (equivalent to open circle in figure 1). Focal fitness is defined in figure 1. Except for R , figure 2a has the parameter set of figure 1a,c and figure 2b has the parameter set of figure 1b,d. (a) Low independent production and (b) high independent production.

public goods G and therefore less benefit from investments in exploitation. As R decreases from 1, the stable intermediate equilibrium declines, the threshold separating asocial and social outcomes (dotted lines in figure 2a) rises, and the zone of attraction for the asocial outcome increases. At a critical value of R (near $R = 0.75$ in figure 2a), a catastrophic transition occurs as the attracting and repelling equilibria merge, and for lower R only the asocial outcome remains (e.g. figure 1c). Thus, under conditions of low relatedness we see no selfish behaviour ($y^* = 0$), as there is no constructive behaviour ($x^* = 0$) to socially parasitize. If independent sources of public goods provision are sufficiently high (figure 1b, e.g. the pasture can grow by itself), then the joint-stable selfish exploitation y^* can peak at intermediate or low relatedness (thus approaching the classical one-dimensional prediction), reflecting the weakening of the evolutionary feedback between x and y (figures 1d and 2b).

(b) Synergistic and antagonistic interactions among social traits

In general, one-dimensional social analyses fail whenever there are fitness interactions among social traits. For social traits x and y , this is the case whenever the inclusive fitness effect of one trait is a function of the other (i.e. $H_y \neq 0$ or $K_x \neq 0$; see appendix A). In this case, Hamilton's cost and benefit terms (appendix A, equation A 3) become co-evolving dynamical variables, and thus neither of the one-trait Hamilton's rules (appendix A, equation A 4) can be properly assessed except in the context of the other. When an increase in one variable leads to an increase in the inclusive fitness effect of the other, we say we have evolutionary synergy (H_y and K_x are positive), and this is seen to be the case in figure 1. In contrast, when an increase in one variable leads to a decrease in the inclusive fitness effect of the other, we say we have evolutionary interference (H_y and K_x are negative).

Figure 3 provides an illustration of neutral, synergistic and interference forms of interaction among social traits, with the example of investment in two separate public goods. When the two public goods are entirely independent in both their costs and benefits (H_y and $K_x = 0$; figure 3a,d), we see that the two-dimensional evolutionary equilibrium remains stable, and behaves entirely as

expected from separate, one-dimensional treatments (the ESS level of public goods investment increases with relatedness). In contrast, when H_y and K_x are non-zero, more interesting dynamics occur. For instance, consider a scenario where the benefits of the public goods only accrue through their synergistic interaction (H_y and $K_x > 0$); when produced in isolation they confer only further costs. This scenario is consistent with the extracellular production of co-enzymes and is illustrated in figure 3b,e. As for figure 1, we see that synergistic interactions among social traits can produce bistability between an asocial boundary condition (investment in neither trait) and the mutual production of both traits (in this example, an unbounded runaway), illustrated in figure 3b for $R = 0.8$. In figure 3e we then show how this unstable equilibrium (separating runaway processes to joint-zero or joint-maximal public goods provision) varies with relatedness. At low relatedness, the positive runaway remains attainable, but the threshold to entering this runaway becomes increasingly high as relatedness decreases. Finally, we illustrate a case of interference (H_y and $K_x < 0$; figure 3c,f), where each public good is separately useful yet costly in combination. The resulting dynamics mirror the following treatment of interfering selfish traits (figure 4); however, here we see that for interfering altruistic traits, the effects of interference are increasing with relatedness, so that the bistability (convergence to investment only in x , or only in y , with the outcome dependent on initial conditions) emerges at high relatedness.

Our illustrative models so far have been 'open' in that we have chosen the value of relatedness independent of underlying demographic processes. Models in which relatedness emerges from demographic parameters such as deme size or dispersal rates (Taylor 1992; Rousset 2004; Lion & van Baalen 2008) are 'closed', and our final example is of this type. The open approach has allowed us a measure of flexibility in observing model behaviour with, for example, variable relatedness, but it is important to see the analysis at work in a particular population structure, and our final example provides a discussion of this. While the 'open model' approach sacrifices some ecological realism, it gains in terms of generality—allowing comparisons across diverse demographies—and in terms of tractability (Taylor & Frank 1996;

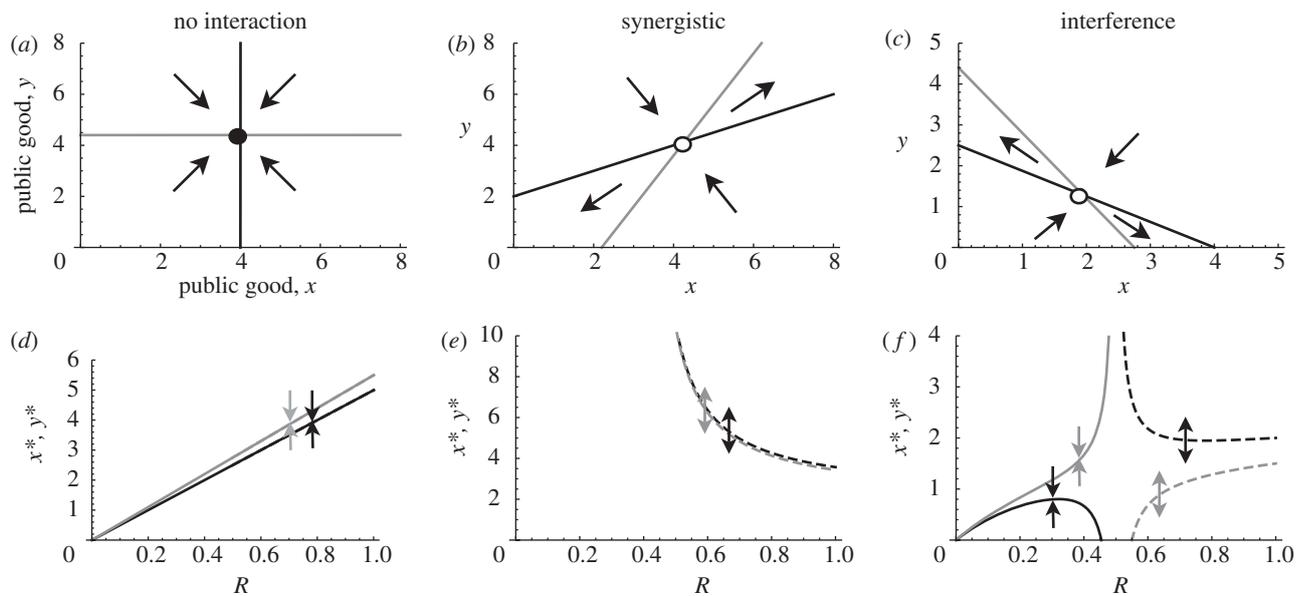


Figure 3. Joint evolution of two cooperative public goods traits (x and y), for neutral (a, d), synergistic (b, e) and interference (c, f) interactions among the public goods. Fitness is defined by $W(x_0, y_0, X, Y) = pX + qY + sXY - cx^2 - ky^2$ representing the costs and benefits of investment into two costly public goods traits x and y (with independent social benefits p and q , and direct costs c and k). The presence of both public goods together can lead to increased (synergy, $s > 0$) or decreased (interference, $s < 0$) rewards. Model presentation mirrors figures 1 and 2. ($a-c$) Plot x and y nullclines (black and grey curves) with stable and unstable intersections marked by filled and open circles, respectively (for $R = 0.8$). ($d-f$) Plot x and y equilibria (black and grey curves) with stable and unstable equilibria marked by solid and dotted lines, respectively. (a, d) No public goods interaction: $p = 1$, $q = 1.1$, $c = k = 0.1$, $s = 0$. (b, e) Synergistic interaction: $p = -1$, $q = -1.1$, $c = k = 0.1$, $s = 0.5$. (c, f) Interference interaction: $p = 1$, $q = 1.1$, $c = k = 0.1$, $s = -0.4$.

Frank 1998; Gardner & West 2006). Clearly, the two approaches are highly complementary, and it remains to be seen what further insights can be generated via closed models of multiple social trait evolution under specific demographic regimes.

Our final worked example presents a two-dimensional generalization of Frank's (1994) model of the evolution of selfishness in an infinite island model with demes of size n and non-overlapping generations (Wright-Fisher demography). This well-known model traditionally works directly with fitness, assumed to have the form $W(x_0, X) = (x_0/X)(1 - X)$, where x_0 is focal selfishness and X is average selfishness in the focal deme. In our structured version of this model it is more appropriate to suppose that behaviour affects only fecundity and let the population structure determine mortality through offspring recruitment. Thus, in a one-trait model, we would take fecundity F to have the classic form $F(x_0, X) = (x_0/X)(1 - X)$, with other components of fitness W determined by the population structure. In fact, because of the internal symmetry of the island model, a simple general result (Taylor 1992; P. D. Taylor, D. Cownden & T. Lillicrap 2009, personal communication) shows that the inclusive fitness effect $H(x)$ of the behaviour x is independent of offspring dispersal patterns and relatedness, and has the same sign as the direct effect of focal behaviour on focal fecundity:

$$H(x) \sim \frac{dF}{dx_0} = \frac{\partial F}{\partial x_0} + \frac{\partial F}{\partial X} \frac{1}{n}, \quad (2.2)$$

the second term providing the focal individual's share of the effect of average deme behaviour on focal fecundity. Although relatedness in this model is *not* equal to $1/n$,

but depends both on deme size n and offspring dispersal rate d , we see that $1/n$, in formula (2.2), plays a role analogous to relatedness.

In our two-dimensional generalization of this model (figure 4), a common resource (e.g. space) is exploited in different ways for two distinct selfish purposes (e.g. acquisition of food and shelter). The most interesting behaviour occurs when cross-inhibition is more severe than self-inhibition (i.e. the social cost of x is more inhibiting to direct benefits of y than is the cost of y itself, and *vice versa*; figure 4a,c). Under these conditions of maximal trait interaction (here, interference) we again see the destabilization of equilibria (appendix A, equation A 6) that are independently stable in each constituent dimension.

However, now the destabilization occurs under conditions of low relatedness, which in this explicit demography can correspond to larger deme sizes. Specifically, we find that in very small demes both traits can coexist at low levels (figure 4c), but in larger demes one trait dominates (excluding the rival trait), with the possibility of bistability (e.g. $n = 5$; figure 4a). As the importance of cross-inhibition is reduced (figure 4b,d), we see greater equilibrium stability (e.g. $n = 5$; figure 4b), yet interference is still manifested in the maximization of selfish trait y at intermediate deme sizes (figure 4d) and intermediate relatedness (electronic supplementary material, fig. S2).

3. DISCUSSION

Our analysis presents a simple and general framework for the study of the joint evolution of multiple social traits. We demonstrate the importance of analysing a complete set of stability conditions for multidimensional social dilemmas, as separate stability in each component

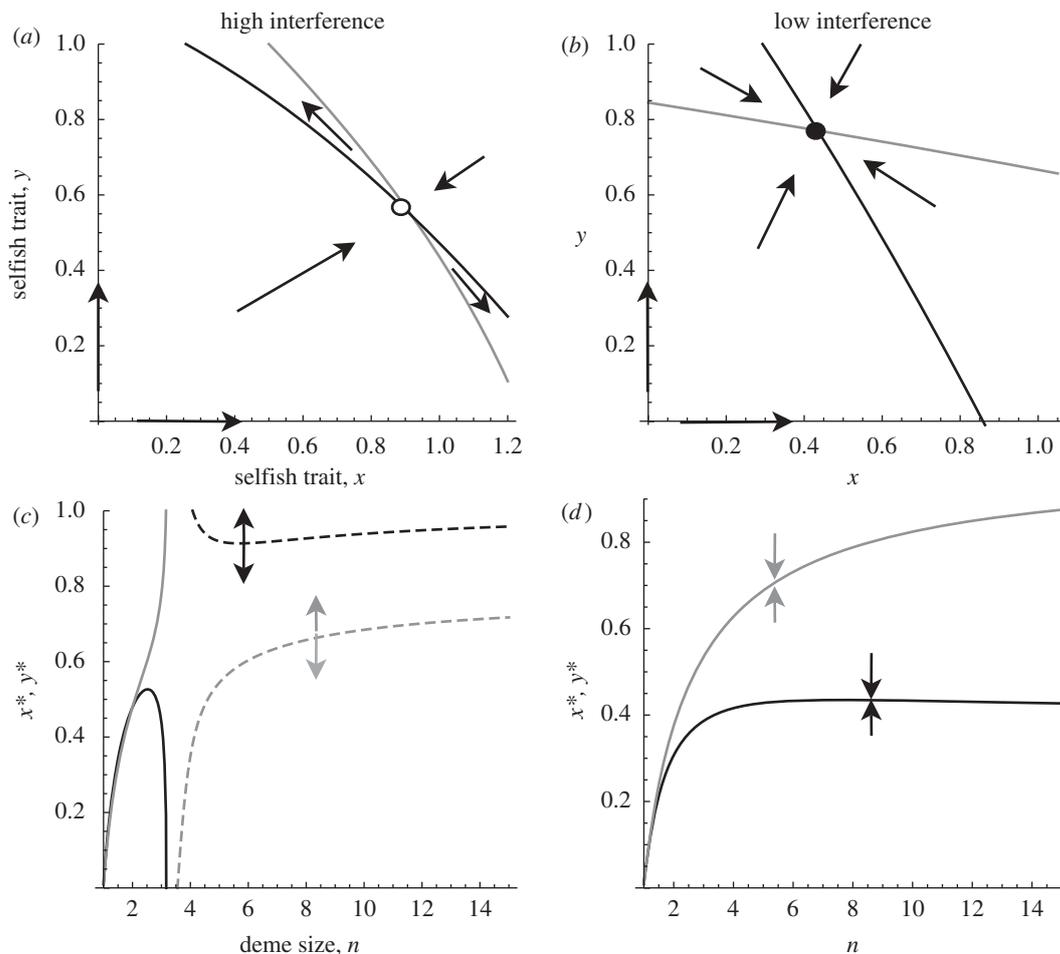


Figure 4. Joint evolution of two competitive traits (x and y), for high (a, c) and low (b, d) levels of mutual interference in an island model. We set this model in an infinite island population with demes of size n . Focal fecundity is $F = (x_0/X)(1 - k_1X - k_2Y) + (y_0/Y)(1 - k_3X - k_4Y) - k_5x_0^2 - k_6y_0^2$, and the inclusive fitness effects of an increase in x and y are $H(x, y) = \partial F/\partial x_0 + (\partial F/\partial X)(1/n)$ and $K(x, y) = \partial F/\partial y_0 + (\partial F/\partial Y)(1/n)$, respectively. Model presentation mirrors figure 3. (a, b) x and y nullclines (black and grey curves) with stable and unstable intersections marked by filled and open circles, respectively. (c, d) x and y equilibria (black and grey curves), with stable and unstable equilibria marked by solid and dotted lines, respectively. (a, c) Strong interference: $k_1 = 0.1, k_2 = 0.9, k_3 = 0.8, k_4 = 0.1, k_5 = 0.1, k_6 = 0.1$. (b, d) Weaker interference: $k_1 = 0.7, k_2 = 0.7, k_3 = 0.3, k_4 = 0.3, k_5 = 0.1, k_6 = 0.3$. (a, b) Deme size n is fixed at $n = 5$.

dimension can conceal multidimensional instability. For the case of two interacting traits, we derive a two-dimensional stability criterion dependent on relatedness (appendix A, conditions A 6–A 7) that can be viewed as a multidimensional extension of Hamilton's rule. We offer a number of examples to illustrate the general stability condition. Our general theoretical framework now raises the vital empirical challenge of understanding the nature of interactions among multiple social traits.

The synergistic interaction between matching production and exploitation traits (figures 1 and 2) is likely to be mirrored across many interacting social traits. In the case of two synergistic productive traits, we can see an analogous threshold between mutual restraint and mutual production, illustrating again the potential for bistable dynamics given multiple social dimensions (figure 3b). The bistability between the presence of multiple social traits and the presence of none is consistent with the co-inheritance of multiple virulence (exoproduct) genes in pathogenicity islands (Hacker & Kaper 2000; Nogueira *et al.* in press) and the co-regulation of multiple public goods traits in quorum-sensing bacteria (Miller & Bassler 2001), highlighting the potential

biomedical relevance of understanding multiple social trait evolution in bacteria (Brown *et al.* in press).

Our focal example of the joint evolution of a production and exploitation trait could be readily extended to consider further interacting traits. For instance, investments in the production and subsequent exploitation of a shared public good (figures 1 and 2) are likely to interact with investment in defence of the public good (in a microbial setting, investment in allelopathic traits such as bacteriocins; Brown *et al.* 2009). When there are more than two interacting traits, the system and the analysis becomes more complicated. Given a stable asocial boundary (all traits constrained by selection to zero), the addition of further social dimensions trivially diminishes the likelihood that the asocial limit remains stable (simply as there are more ways to be social). More subtly, as dimensions are added, the likelihood that any particular dimension becomes social (i.e. non-zero) increases, as evolution in an additional dimension can potentially favour the elaboration of the initially constrained focal trait (for instance, in figure 1b, the focal dimension x is constrained to zero when $y = 0$; however, unilateral elaboration of y from zero subsequently favours

elaboration of x). This illustration echoes and generalizes Sober & Wilson's (1998) argument that any mechanism that leads to the emergence of affordable punishment will allow the elaboration of a more costly, and therefore previously constrained, cooperative trait (Sober & Wilson 1998, p. 144).

Following escape from the asocial boundary, predicting the ultimate system fate becomes more complex with increasing dimensionality. As the number of traits increases, the conditions for stability of any equilibrium become more severe, but the number of equilibrium points also generally increases. If the costs of large values of all traits are prohibitive, the dynamical system will stay bounded and has to go somewhere (converging to a stable point or cycle, or exhibiting chaotic behaviour) and the ecological systems that survive today are generally those that have arrived at a stable convergence. It is likely that such convergence has often wound up at the boundary on which one or more potential traits have been lost (although synergistic trait interactions will tend to favour trait coexistence), so that the interesting question becomes which behaviours are maintained in a particular system. In this study we have begun to build a framework in which this question might be studied.

We thank Angus Buckling, Andy Gardner, Sébastien Lion, Craig Maclean, Daniel Rankin, François Taddei and Stuart West for helpful discussions and/or comments on an earlier draft. We thank the Wellcome Trust (S.P.B.) and the Leverhulme Foundation (P.D.T.) for financial support.

APPENDIX A. GENERAL METHODOLOGY

Let $W(x_0, y_0, X, Y)$ be the fitness of an individual (x_0, y_0) in a neighbourhood (X, Y) . Then the inclusive fitness effects of an x -increase and of a y -increase have the form:

$$H(x, y) = R_x B_x(x, y) - C_x(x, y), \quad (\text{A } 1)$$

and

$$K(x, y) = R_y B_y(x, y) - C_y(x, y), \quad (\text{A } 2)$$

where

$$\begin{aligned} C_x &= -\frac{\partial W}{\partial x_0} & B_x &= \frac{\partial W}{\partial X} \\ C_y &= -\frac{\partial W}{\partial y_0} & B_y &= \frac{\partial W}{\partial Y}, \end{aligned} \quad (\text{A } 3)$$

all derivatives evaluated at $x_0 = X = x$, and $y_0 = Y = y$ (Taylor & Frank 1996). The notation is chosen because, in models of cooperative behaviour, C and B can be regarded as a cost and a benefit. We allow the possibility that the x - and y -relatedness are different. For example, if x is a cooperative and y a selfish trait, the x -neighbourhood might be smaller than the y -neighbourhood. If $H(x, y)$ (or $K(x, y)$) is positive, selection will favour higher values of x (or y). We write this as

$$\begin{aligned} R_x B_x(x, y) &> C_x(x, y) \\ R_y B_y(x, y) &> C_y(x, y), \end{aligned} \quad (\text{A } 4)$$

and we obtain the form of Hamilton's (1964) classical rule for each trait separately. There is an assumption here of independence in x and y ; that is, no genetic correlation between them (individuals with high or low values of one trait are neither more nor less likely to carry high or

low values of the other). Given this assumption, the condition for a point (x^*, y^*) to be an *evolutionary equilibrium* is that $H(x, y)$ and $K(x, y)$ be zero.

We now turn to the question of the stability of such an equilibrium: what will happen to small departures from (x^*, y^*) ? The classical condition of convergence stability (Christiansen 1991) is that $H_x < 0$ and $K_y < 0$ at (x^*, y^*) , where the subscript denotes partial derivative. In this case, both x and y will be separately stable with the other variable held fixed; that is, if the population-wide values of x or y depart from equilibrium, selection will favour those values closer to the equilibrium.

To account for the joint effect of departures in both traits, we consider the dynamical system

$$\begin{aligned} \frac{dx}{dt} &= H(x, y) \\ \frac{dy}{dt} &= K(x, y), \end{aligned} \quad (\text{A } 5)$$

modelled after the standard adaptive dynamics (Leimar in press). Note that a stationary point of the system ($H = K = 0$) will be a point of evolutionary equilibrium. For such a point to be dynamically stable (Otto & Day 2007), we first require that $H_x < 0$ and $K_y < 0$, and these are the conditions (above) of separate convergence stability, but we also require the additional condition $H_x K_y > H_y K_x$ that the linearized system have positive determinant. If this fails, departures from equilibrium in one variable can destabilize the other. This condition can be written

$$aR_x R_y + b_x R_x + b_y R_y + c > 0, \quad (\text{A } 6)$$

where

$$\begin{aligned} a &= B_{xx} B_{yy} - B_{xy} B_{yx}, \\ b_x &= B_{xy} C_{yx} - B_{xx} C_{yy}, \\ b_y &= B_{yx} C_{xy} - B_{yy} C_{xx}, \\ c &= C_{xx} C_{yy} - C_{xy} C_{yx}, \end{aligned}$$

and (as above) the second subscript denotes partial derivative. When the two interaction neighbourhoods are the same, $R_x = R_y = R$, condition (A 5) becomes

$$aR^2 + bR + c > 0, \quad (\text{A } 7)$$

where $b = b_x + b_y$.

REFERENCES

- Axelrod, R., Hammond, R. A. & Grafen, A. 2004 Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. *Evolution* **58**, 1833–1838.
- Brown, S. P. & Johnstone, R. A. 2001 Cooperation in the dark: signalling and collective action in quorum-sensing bacteria. *Proc. R. Soc. Lond. B* **268**, 961–965. (doi:10.1098/rspb.2001.1609)
- Brown, S. P., Hochberg, M. E. & Grenfell, B. T. 2002 Does multiple infection select for raised virulence? *Trends Microbiol.* **10**, 401–405. (doi:10.1016/S0966-842X(02)02413-7)
- Brown, S. P., Inglis, R. F. & Taddei, F. 2009 Evolutionary ecology of microbial wars: within-host competition and (incidental) virulence. *Evol. Appl.* **2**, 32–39. (doi:10.1111/j.1752-4571.2008.00059.x)
- Brown, S. P., West, S. A., Diggle, S. P. & Griffin, A. S. In press. Social evolution in microorganisms and a

- Trojan horse approach to medical intervention strategies. *Phil. Trans. R. Soc. B*.
- Christiansen, F. B. 1991 On conditions for evolutionary stability for a continuously varying character. *Am. Nat.* **138**, 37–50. (doi:10.1086/285203)
- Crosa, J. H., Mey, A. R. & Payne, S. M. 2004 *Iron transport in bacteria: molecular genetics, biochemistry and role in pathogenicity and ecology*. Washington, DC: ASM Press.
- Dieckmann, U. & Law, R. 1996 The dynamical theory of coevolution: a derivation from stochastic ecological processes. *J. Math. Biol.* **34**, 579–612. (doi:10.1007/BF02409751)
- Fehr, E. & Gächter, S. 2002 Altruistic punishment in humans. *Nature* **415**, 137–140. (doi:10.1038/415137a)
- Frank, S. A. 1994 Kin selection and virulence in the evolution of protocells and parasites. *Proc. R. Soc. Lond. B* **258**, 153–161. (doi:10.1098/rspb.1994.0156)
- Frank, S. A. 1995 Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* **377**, 520–522. (doi:10.1038/377520a0)
- Frank, S. A. 1998 *Foundations of social evolution*. Princeton, NJ: Princeton University Press.
- Frank, S. A. 2003 Repression of competition and the evolution of cooperation. *Evolution* **57**, 693–705.
- Gardner, A. & West, S. A. 2004 Cooperation and punishment, especially in humans. *Am. Nat.* **164**, 753–764. (doi:10.1086/425623)
- Gardner, A. & West, S. A. 2006 Demography, altruism, and the benefits of budding. *J. Evol. Biol.* **19**, 1707–1716. (doi:10.1111/j.1420-9101.2006.01104.x)
- Griffin, A. S., West, S. A. & Buckling, A. 2004 Cooperation and competition in pathogenic bacteria. *Nature* **430**, 1024–1027. (doi:10.1038/nature02744)
- Hacker, J. & Kaper, J. B. 2000 Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* **54**, 641–679. (doi:10.1146/annurev.micro.54.1.641)
- Hamilton, W. D. 1964 The genetical evolution of social behaviour, I and II. *J. Theor. Biol.* **7**, 1–52. (doi:10.1016/0022-5193(64)90038-4)
- Hamilton, W. D. 1970 Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**, 1218–1220. (doi:10.1038/2281218a0)
- Hamilton, W. D. & May, R. M. 1977 Dispersal in stable habitats. *Nature* **269**, 578–581. (doi:10.1038/269578a0)
- Hardin, G. 1968 The tragedy of the commons. *Science* **162**, 1243–1248.
- Hochberg, M. E., Rankin, D. J. & Taborsky, M. 2008 The coevolution of cooperation and dispersal in social groups and its implications for the emergence of multicellularity. *BMC Evol. Biol.* **8**, 238. (doi:10.1186/1471-2148-8-238)
- Jansen, V. A. A. & van Baalen, M. 2006 Altruism through beard chromodynamics. *Nature* **440**, 663–666. (doi:10.1038/nature04387)
- Le Galliard, J.-F., Ferriere, R. & Dieckmann, U. 2003 The adaptive dynamics of altruism in spatially heterogeneous populations. *Evolution* **57**, 1–17.
- Le Galliard, J.-F., Ferriere, R. & Dieckmann, U. 2005 Adaptive evolution of social traits: origin, trajectories, and correlations of altruism and mobility. *Am. Nat.* **165**, 206–224. (doi:10.1086/427090)
- Lehmann, L. & Feldman, M. W. 2008 War and the evolution of belligerence and bravery. *Proc. R. Soc. B* **275**, 2877–2885. (doi:10.1098/rspb.2008.0842)
- Lehmann, L. & Keller, L. 2006 The evolution of cooperation and altruism: a general framework and a classification of models. *J. Evol. Biol.* **19**, 1365–1376. (doi:10.1111/j.1420-9101.2006.01119.x)
- Leimar, O. In press. Multidimensional convergence stability and the canonical adaptive dynamics. In *Elements of adaptive dynamics* (eds U. Dieckmann & J. A. J. Metz). Cambridge, UK: Cambridge University Press.
- Lion, S. & van Baalen, M. 2008 Self-structuring in spatial evolutionary ecology. *Ecol. Lett.* **11**, 277–295. (doi:10.1111/j.1461-0248.2007.01132.x)
- Miller, M. B. & Bassler, B. L. 2001 Quorum sensing in bacteria. *Annu. Rev. Microbiol.* **55**, 165–199. (doi:10.1146/annurev.micro.55.1.165)
- Nogueira, T., Rankin, D. J., Touchon, M., Taddei, F., Brown, S. P. & Rocha, E. P. C. In press. Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Curr. Biol.*
- Otto, S. P. & Day, T. 2007 *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton, NJ: Princeton University Press.
- Perrin, N. & Lehmann, L. 2001 Is sociality driven by the costs of dispersal or the benefits of philopatry? A role for kin-discrimination mechanisms. *Am. Nat.* **158**, 471–483. (doi:10.1086/323114)
- Pfeiffer, T. & Bonhoeffer, S. 2003 An evolutionary scenario for the transition to undifferentiated multicellularity. *PNAS* **100**, 1095–1198. (doi:10.1073/pnas.0335420100)
- Platt, T. G. & Bever, J. D. 2009 Kin competition and kin cooperation. *Trends Ecol. Evol.* **24**, 370–377. (doi:10.1016/j.tree.2009.02.009)
- Rousset, F. 2004 *Genetic structure and selection in subdivided populations*. Princeton, NJ: Princeton University Press.
- 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. (doi:10.1046/j.1420-9101.2000.00219.x)
- Rousset, F. & Roze, D. 2007 Constraints on the origin and maintenance of genetic kin recognition. *Evolution* **61**, 2320–2330. (doi:10.1111/j.1558-5646.2007.00191.x)
- Roze, D. & Rousset, F. 2008 Multilocus models in the infinite island model of population structure. *Theor. Popul. Biol.* **73**, 529–542. (doi:10.1016/j.tpb.2008.03.002)
- Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. 2004 The evolution of cooperation. *Q. Rev. Biol.* **79**, 136–160.
- Sober, E. & Wilson, D. S. 1998 *Unto others: the evolution and psychology of unselfish behavior*. Cambridge, MA: Harvard University Press.
- Taylor, P. D. 1992 Altruism in viscous populations: an inclusive fitness model. *Evol. Ecol.* **6**, 352–356. (doi:10.1007/BF02270971)
- Taylor, P. D. & Frank, S. 1996 How to make a kin selection argument. *J. Theor. Biol.* **180**, 27–37. (doi:10.1006/jtbi.1996.0075)
- Taylor, P. D., Day, T. & Wild, G. 2007 From inclusive fitness to fixation probability in homogeneous structured populations. *J. Theor. Biol.* **249**, 101–110. (doi:10.1016/j.jtbi.2007.07.006)
- Wenseleers, T. & Ratnieks, F. L. W. 2006 Enforced altruism in insect societies. *Nature* **444**, 50. (doi:10.1038/444050a)
- Wenseleers, T., Hart & Ratnieks, F. L. W. 2004 When resistance is useless: policing and the evolution of reproductive acquiescence in insect societies. *Am. Nat.* **164**, E154–E167. (doi:10.1086/425223)
- Wenseleers, T., Gardner, A. & Foster, K. R. In press. Social evolution theory: a review of methods and approaches. In *Social behaviour: genes, ecology and evolution* (eds T. Székely, J. Komdeur & A. J. Moore). Cambridge, UK: Cambridge University Press.
- West, S. A., Pen, I. & Griffin, A. S. 2002 Cooperation and competition between relatives. *Science* **296**, 72–75. (doi:10.1126/science.1065507)
- 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. (doi:10.1111/j.1420-9101.2006.01258.x)