

Developments in the theory of social evolution

Andy Gardner

PhD

University of Edinburgh

2004

For Susan

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated in the text. This work has not been submitted for any degree or professional qualification except as specified.

Andy Gardner, 27th October 2004

Acknowledgements

The work presented in this thesis benefited from the comments, discussion, collaboration and encouragement of many people, including: Dave Allsop, Nick Barton, Angus Buckling, Angeles de Cara, Brian Charlesworth, Ric Charnov, Tim Clutton-Brock, Troy Day, Francisco Dionysio, Steve Frank, Alan Grafen, Ashleigh Griffin, Meghan Guinnee, Toby Johnson, Alex Kalinka, Mark Kirkpatrick, Natasha Mehdiabadi, Arcadi Navarro, Sean Nee, Ric Paul, Andrew Read, Sarah Reece, Tim Sands, Dave Shuker, Jay Taylor, Peter Taylor, Tim Vines, Stu West, Susan Williamson and Jelle Zuidema.

In particular, I thank my supervisors, Nick and Stu, for giving me complete freedom in my research, yet remaining firmly supportive and offering expert supervision throughout. They have been, and continue to be, an inspiration.

I thank the Biotechnology & Biological Sciences Research Council for providing me with my studentship.

Special thanks are due to Meghan, for printing, binding and submitting for me in the UK. And Alex, for dropping everything and helping out when it all went wrong.

Finally, I thank my parents, John and June, and my partner, Susan, for their encouragement and support. I couldn't have done this without them.

Publications

The following published papers have arisen from this thesis, and are included in the Appendix:

- Gardner, A. & West, S.A. (in press) Spite and the scale of competition. *Journal of Evolutionary Biology*.
- Gardner, A. & West, S.A. (in press) Cooperation and punishment, especially in humans. *American Naturalist*.
- Gardner, A. & West, S.A. 2004. Spite among siblings. *Science* 305, 1413-1414.
- Gardner, A., West, S.A. & Buckling, A. 2004. Bacteriocins, spite and virulence. *Proceedings of the Royal Society of London Series B – Biological Sciences* 271, 1529-1535.
- Gardner, A., Reece, S.E. & West, S.A. 2003. Even more extreme fertility insurance and the sex ratios of protozoan blood parasites. *Journal of Theoretical Biology* 223, 515-521.
- Gardner, A. & Zuidema, W. 2003. Is evolvability involved in the origin of modular variation? *Evolution* 57, 1448-1450.

The unpublished chapters 3, 7 and 8 are collaborative efforts, though in each case the majority of the work is my own. In chapter 3, D. Allsop gathered the relevant data, and D. Allsop, E.L.Charnov and S.A.West contributed to the interpretation of results and discussion. Chapter 7 is my own work, and was supervised by S.A.West and N.H.Barton. The motivation for chapter 8 arose from discussions with A.T.Kalinka, who devised the simulations to complement my analytical treatment.

Abstract

The study of social evolution is concerned with fitness consequences of interactions between individuals. It has proven to be an excellent area for relating theoretical predictions to empirical observations. I develop social evolution theory in several ways. (1) I demonstrate that limited male fecundity and small mating groups can select for extreme fertility insurance, curbing female biased sex allocation under local mate competition, which explains puzzling sex ratios in protozoan blood parasites. (2) I examine the underlying causes of an observed statistical invariant in the relative size at sex change in animals, revealing that it does not imply as much conservation of biology across taxa as previously imagined. (3) I extend recent theory regarding how local competition impedes the evolution of altruism to show that it also promotes the evolution of spite. This allows me to re-interpret several behaviours in terms of spitefulness, and predict where spite will occur in nature. (4) I apply spite theory to the evolution of chemical (bacteriocin) warfare in bacteria, and derive novel predictions for the evolution of virulence caused by bacterial parasites. (5) I formalize a verbal model for the evolution of costly punishment as a mechanism of promoting cooperation, revealing a logical flaw and the true source of its (potential) selective benefit. (6) I develop a multi-locus methodology for arbitrary social interactions, and apply this to a dynamically-sufficient co-evolutionary analysis of cooperation and costly punishment, revealing when punishment is favoured by selection. (7) I apply this methodology to the evolution of mutation robustness for a simple two-locus model with recombination and inbreeding.

Contents

1. Introduction	1
2. Even more extreme fertility insurance and the sex ratios of protozoan blood parasites	9
3. A dimensionless invariant for relative size at sex change in animals: explanations and implications	25
4. Spite and the scale of competition	55
5. Bacteriocins, spite and virulence	73
6. Cooperation and punishment, especially in humans	90
7. Social evolutionary multi-locus methodology	111
8. Recombination and the evolution of mutational robustness: a two-locus model	144
9. Discussion	160
Bibliography	172
Appendix	201

1. Introduction

Social evolution: definitions and classification

Social evolution theory is concerned with the fitness consequences of interactions between individuals. Classically, social behaviours are categorised according to their impact on the reproductive success of the ‘actor’ and any ‘recipients’ (figure 1.1; Hamilton 1964, Trivers 1985). The categories are: (1) *mutualism*, where both actor and recipient directly benefit from the behaviour; (2) *selfishness*, where the actor gains at the expense of the recipient; (3) *altruism*, where the behaviour is detrimental to the actor but beneficial for the recipient; and (4) *spite*, where the behaviour is harmful for both actor and recipient. Mutualism and selfishness, which enhance the reproductive success of the actor, are easily explained. Less easy to account for are instances where an individual acts to its own detriment. In particular, much attention has focused on the problem of altruism (for example, Hamilton 1963, 1964, Wilson 1975).

		Effect on recipient	
		+	-
Effect on actor	+	mutualism	selfishness
	-	altruism	spite

Figure 1.1. A classification of social behaviours.

Altruism and Hamilton's rule

The answer lies in the possibility of statistical associations between individuals. Hamilton (1963, 1964, 1970) developed two equivalent ways of thinking about fitness when social partners are correlated. If a positive correlation exists between the behaviours of social partners then altruists will tend to associate with altruists. Thus, altruists suffer a direct cost through their behaviour but benefit from the altruistic behaviour of their social partners. This is the *neighbour-modulated* view of fitness. Alternatively, if there are genetical correlations between social partners then an altruistic gene will reduce the actor's number of direct descendants, but it will also enhance the transmission of altruistic genes via the recipients of the altruism. This may result in a net benefit for the altruistic gene, in which case the actor is said to have maximized its *inclusive fitness*. Whichever of these two views of fitness are taken (Frank 1997a, 1998), the result is the following statement, known as *Hamilton's rule*: selection will act to favour the social trait when $RB > C$, where B is the direct benefit to the recipient, C is the direct cost to the actor, and R is the relatedness of the recipient to the actor. Positive relatedness might result from, for example, genealogical closeness. Since this is generally the cause of such associations, the special form of selection has often been referred to as *kin selection* (Maynard Smith 1964).

Individual versus group

The evolution of altruism has often been framed in terms of the tension between *individual selection* and *group selection*, with the former favouring selfishness and the latter favouring selflessness. A persistent problem which arises at all levels of biological organisation is the 'tragedy of the commons' (Hardin 1968, Maynard Smith & Szathmary 1995, Frank 1998), in which individuals would all benefit from existing in a cooperative group, yet there is an immediate incentive for each individual to behave in a less than cooperative way. Given that one's selflessness is liable to be exploited by another's selfishness, everyone behaves selfishly, and hence the group as a whole does badly. Self-restraint can be favoured when there is a positive correlation between the social partners,

so that when an individual agrees to act selflessly it can be sure that, to a certain degree, the rest of the group will behave accordingly (Frank 1998). Again, the rule $RB > C$ applies, and group selection is found to be mathematically equivalent to kin selection (Price 1972a, Hamilton 1975, Grafen 1984, Wade 1985, Frank 1986, 1998, Queller 1992).

Spite

Spite, the flip-side of altruism, has received very little attention. If an individual pays a cost ($C > 0$) in order to inflict harm ($B < 0$) on a negative relation ($R < 0$), then its behaviour may satisfy Hamilton's rule ($RB > C$), and be favoured by selection (Hamilton 1970, Grafen 1985a, Foster et al. 2001). There are a number of reasons why spite has been neglected. There is a pervasive belief that relatedness is a probability measure, and logically rules out the possibility of negative relations. Within social evolution theory, negative relatedness is acknowledged as a possibility, but strong negative relatedness is regarded as biologically implausible (Hamilton 1970). Generally, behaviours which might be regarded as spiteful are usually re-expressed in terms of altruism or selfishness.

Sex allocation

Social evolution has proven to be an excellent field for relating theoretical predictions to empirical observations. Indeed, in some cases the strong empirical support is quantitative as well as qualitative (Werren 1980, Charnov 1982a, Herre 1987, Godfray 1994). The astounding success of social evolutionary biology is partly because the field has firm but conceptually simple theoretical foundations, but also because there has been much focus on social traits which are well characterized biologically, are easy to measure, and are strongly tied to fitness (Seeger & Stubblefield 1996). The model trait for social evolution theory has been sex allocation (Charnov 1982a, Frank 1998, 2002). This is well understood in terms of the fundamental trade-off between investment of resources into sons and into daughters. Sex allocation is easily defined and measured. It is the proportion of resources invested into male as opposed to female function – for example, in a gonochoristic species it is the fraction of investment into sons rather than daughters –

and often this can be measured simply by counting the number of progeny of each sex. In some contexts, sex allocation theory works far better than it is expected to, with simple models generating accurate predictions for sex ratios across taxa despite obvious variation in what are expected to be important biological details (Allsop & West 2003a). However, there are also situations where the theory is relatively poor at explaining the variation in sex allocation, and such examples are worthy of further investigation, as it seems that here key biological details have been missed.

Although sex allocation does not immediately appear to fit into the framework of figure 1.1, it can be phrased in terms of kin selection and group selection (Hamilton 1967). For example, male fecundity is often effectively limitless, and a group's productivity will be related to the number of fertile females. An individual who contributes more daughters than sons improves the success of the group. Given that the total success of all the males in the mating group is equal to the total success of the females, individual males will be more successful than individual females in groups with an excess of females, so a mother who allocates more resources to sons will capitalise on the altruism of her female-producing counterparts. Female biased sex ratios are predicted, and observed, within species where there is relatedness between mating partners, and when this is absent (i.e. random mating) then resources are allocated equally to the two sexes. Hamilton (1967) described female bias due to such *local mate competition* in terms of kin selection, but noted the group selection interpretation. Sex allocation theory, with its precise and readily testable predictions, has played a major role in cutting through the verbiage to show that the kin selection versus group selection debate is empirically empty.

Evolution of virulence

In recent years, there has been growing interest in applying social evolution theory to explaining the evolution of virulence, which is the damage done by a parasite to its host (Frank 1996a). Clearly, there are major economic and humanitarian gains from understanding how parasite population structure impacts on virulence. Again, the general principle is that there is a tension between the individual and the group, such that less

aggressive exploitation of the host will lead to a longer-lived infection and hence enhanced transmission for all the resident parasites, but those individuals who are more virulent will enjoy immediate transmission benefits. Thus, when there is high parasite relatedness (due to a small number of lineages constituting the infection) self-restraint and hence less virulence is favoured, but when the parasites are less related (a larger number of lineages in the host) then exploitation and high virulence is predicted (Hamilton 1972, Bremermann & Pickering 1983). Unfortunately, virulence theory has enjoyed less success than the theory of sex allocation, suggesting that important aspects of infection biology are overlooked by current models (Herre 1993, 1995, Chao et al. 2000, Read & Taylor 2001, Read et al. 2002, Griffin & West 2002, Davies et al. 2002).

Scale of competition

A topic that has recently been the focus of much discussion in social evolution research concerns the impact of competition between social partners as a major obstacle in the evolution of altruism (reviewed by West et al. 2002a). Motivation for work on this topic stems from earlier suggestions that indiscriminate altruism should be looked for in low diffusive (or 'viscous') populations, such that social partners will tend to be genealogical kin (Hamilton 1964). Simulation studies were unable to verify this prediction, and beautifully simple mathematics (Taylor 1992a, b) emerged to show that the effects of enhanced kinship were exactly countered by competition between kin over the whole range of dispersal rates in standard models. Essentially, with low dispersal there is high kinship between social partners but an inability for altruistic groups to export the enhanced productivity their altruism brings, thus leading to a competitive strain within such groups. With a higher dispersal rate there is the opportunity to inflict the increased competition on the population as a whole, but there is also lower kinship. This has spawned a variety of theoretical developments to account for the competition effect, either by adding extra terms into Hamilton's rule (Grafen 1984, Frank 1998) or by redefining relatedness itself (Queller 1994).

Repression of competition

A growing body of literature now concerns itself with how altruism can be favoured at all, given the inhibitory effects of competition. One solution is that groups might evolve mechanisms of repressing competition between individuals, thus facilitating altruism, and hence the emergence of higher cooperative units (reviewed by Frank 2003a). Such repressive mechanisms – for example, worker policing in eusocial insects, and punishment in human societies – can often provide much stronger motivation for curtailing selfish behaviour than any pressure for kin selected self restraint. However, it is often unclear how these mechanisms could themselves be favoured. A rich literature on such themes has arisen in parallel in the social sciences (for example, Ostrom 1990), and this could fruitfully be related to the foundations of social evolution theory.

Developments in the theory of social evolution

In this thesis I will touch on and develop the above themes. Specifically:

Chapter 2: The application of local mate competition to sex allocation in protozoan blood parasites represents one of the instances where the theory does not perform so well. Specifically, empirical observations give much less female-bias than is predicted by existing models (West et al. 2001a). I examine the interaction between two causes of reduced female bias – limited male fecundity and finite mating groups (collectively, “fertility insurance”) – to show that adding these biological details brings the theory more closely in line with the observations.

Chapter 3: I then move to the other extreme to examine a context where biological details do not seem to matter at all – the relative timing of sex change in sequential hermaphrodites, which is remarkably invariant (>90% of the variance explained by the prediction that sex change occurs at 72% of maximum body size) across phyla and despite diverse biology and orders of magnitudes of body sizes (Allsop & West 2003a). I formalize the dimensionless theory underlying this observation, answer some recent

criticism as to whether this really constitutes an invariant, obtain estimates of the mean and variance for key dimensionless life history parameters underlying the timing of sex change. This sheds some light on which aspects of the biology need to be conserved and which are less constrained.

Chapter 4: I extend recent theory on the impact of local competition in the evolution of social behaviours. I show that since increased competition between social partners reduces relatedness, relatedness may plausibly take negative values. As well as inhibiting the evolution of altruism, local competition favours the evolution of spite. I use this theory to show that spite is a general evolutionary phenomenon, to reinterpret several behaviours in terms of spite, and to suggest where spite is likely to occur in nature.

Chapter 5: Spite theory is applied to the evolution of chemical (bacteriocin) warfare in bacteria. Bacteriocin production is modeled quantitatively as a function of bacterial kinship and scale of competition. The theory is then applied to bacterial parasites, generating novel predictions for the evolution of virulence, and highlighting how introducing some biological details can dramatically alter the predictions of virulence theory.

Chapter 6: I formalize a verbal argument for the evolution of costly punishment, with special attention to humans (Sober & Wilson 1998). This involves extending standard social evolutionary methodology to encompass multiple non-independent co-evolving traits, namely cooperation and punishment. The formalism reveals a logical error in the verbal argument, and suggests how costly punishment might be favoured.

Chapter 7: The extension of social evolution methodology to include multiple co-evolving traits is pursued in more general terms. A multi-locus methodology is borrowed from theoretical population genetics (Barton & Turelli 1991, Kirkpatrick et al. 2002) and its applications to social evolution are highlighted. The multi-locus methodology is integrated with the foundations of social evolution.

Chapter 8: The multi-locus methodology is applied to the co-evolution of a mutational robustness gene and a linked mutating locus, for a range of recombination and inbreeding rates. I show that increased recombination and reduced inbreeding facilitates the evolution of costly robustness. Because robustness has no long term benefit, this process is detrimental to the mean fitness of the population.

2. Even more extreme fertility insurance and the sex ratios of protozoan blood parasites[†]

Abstract

Theory developed for malaria and other protozoan parasites predicts that the evolutionarily stable gametocyte sex ratio (z^* ; proportion of gametocytes that are male) should be related to the inbreeding rate (f) by the equation $z^* = (1-f)/2$. Although this equation has been applied with some success, it has been suggested that in some cases a less female biased sex ratio can be favoured to ensure female gametes are fertilised. Such fertility insurance can arise in response to two factors: (i) low numbers of gametes produced per gametocyte and (ii) the gametes of only a limited number of gametocytes being able to interact. However, previous theoretical studies have considered the influence of these two forms of fertility insurance separately. We use a stochastic analytical model to address this problem, and examine the consequences of when these two types of fertility insurance are allowed to occur simultaneously. Our results show that interactions between the two types of fertility insurance reduce the extent of female bias predicted in the sex ratio, suggesting that fertility insurance may be more important than has previously been assumed.

Introduction

One of the many successful applications of sex allocation theory has been the study of how competition for mates between related males can favour the evolution of female biased sex ratios (Charnov, 1982a; Godfray, 1994; Hamilton, 1967; West *et al.*, 2000a). Recent years have seen an increasing interest in applying this theory (local mate

[†] Published as: Gardner A., Reece S.E., & West S.A. 2003 Even more extreme fertility insurance and the sex ratios of protozoan blood parasites. *Journal of Theoretical Biology* **223**, 515-521 (see Appendix).

competition; LMC) to malaria and related protozoan parasites (Read *et al.*, 2002a; West *et al.*, 2001a). Here, the appropriate prediction is that the evolutionarily stable strategy (ESS; Maynard Smith, 1982) gametocyte sex ratio (z^* ; proportion of gametocytes that are male) should be related to the inbreeding rate (f) by the equation $z^* = (1-f)/2$ (Hamilton, 1967; Nee *et al.*, 2002; Read *et al.*, 1992). When there is complete inbreeding ($f=1$; i.e. a single lineage or clone is selfing), the ESS is to produce the minimum number of males required to fertilise the available female gametes and thus, maximise the number of zygotes. Conversely, when gametes in the mating pool are of a mixture of lineages, f decreases and the sex ratio increases in order for each lineage to maximise its genetic representation in the zygote population. The relationship between the inbreeding rate and sex ratio has been able to explain a number of sex ratio patterns in Apicomplexan parasite populations (reviewed by West *et al.*, 2001a; Read *et al.*, 2002a). However, there are a number of observations that cannot be explained by this equation. In particular: (1) across *Haemoproteus* populations in birds the sex ratio does not correlate with an expected correlate of the inbreeding rate (prevalence; Shutler *et al.*, 1995; Shutler & Read, 1998); (2) in malaria parasites, sex ratios within and between infections can be extremely variable (Osgood *et al.*, 2002; Paul *et al.*, 2002; Paul *et al.*, 2000; Paul *et al.*, 1999; Pickering *et al.* 2000; Schall, 1989; Taylor, 1997), and less female biased sex ratios can lead to greater transmission success (Robert *et al.*, 1996).

A potential explanation for these contradictory observations is “fertility insurance” – the production of a less female biased sex ratio to ensure that all female gametes are fertilised (West *et al.*, 2002b). Before describing how fertility insurance can influence the ESS sex ratio it is necessary to describe the background biology. In malaria and related Haemosporin parasites, haploid sexual stages (gametocytes) are taken up from the host in the blood meal of a vector. Once inside the midgut, the haploid gametocytes differentiate into haploid gametes and fuse to form zygotes. These resulting diploid zygotes undergo meiosis and asexual proliferation before migrating to the vector’s salivary glands where they wait to enter a new vertebrate host. Each female gametocyte (macro-gametocyte) will differentiate into one female gamete, whereas each male gametocyte (micro-gametocyte) will produce several motile male gametes. The number

of viable gametes produced per male gametocyte varies enormously across species: 4-8 in mammalian malaria parasites (Read *et al.*, 1992); ~2 in some lizard malaras (Schall, 2000); 5-1000 in *Eimeriorin* intestinal parasites (West *et al.*, 2000a).

Fertility insurance can occur for two broad reasons – which are summarised here but discussed more fully in West *et al.* (2002b). First, the number of male gametes produced per gametocyte (c) may be a limiting factor (Read *et al.*, 1992). If the mean number of viable gametes produced per male gametocyte is c , then the ESS sex ratio must be $z^* \geq 1/(c+1)$, otherwise there will not be enough male gametes to fertilise the female gametes (fig 2.1A; Read *et al.*, 1992). Second, the ability of gametes to interact may be a limiting factor. West *et al.* (2002b) investigated this possibility by assuming that the number of gametocytes whose gametes can interact (q) is restricted. In this case a less female biased sex ratio is favoured to avoid the stochastic absence of males in a mating group of q gametocytes (figure 2.1B; West *et al.*, 2002b). A low q could occur for a number of reasons including low male gamete motility, high gametocyte or gamete mortality, low gametocyte density, or small blood meals (Shutler & Read, 1998; Paul *et al.*, 1999, 2000, 2002; Reece & Read, 2000; West *et al.*, 2001a, 2002b). Recent attention has focused on how the host immune response may influence and vary the importance of these factors (Paul *et al.*, 1999, 2000, 2002; Reece & Read, 2000).

In order to make their analyses mathematically tractable, previous studies have considered the influence of these two forms of fertility insurance separately. When examining the influence of male gametocyte fecundity (c), Read *et al.* (1992) assumed that the gametes from a large of gametocytes can interact ($q \rightarrow \infty$), and when examining the influence of the number of gametocytes whose gametes can interact (q), West *et al.* (2002b) assumed that male gamete fecundity was not a limiting factor ($c \rightarrow \infty$; i.e. one male gametocyte is able to provide enough gametes to fertilise all of the female gametes in its mating group arising from q gametocytes). It has subsequently been assumed that the overall effect of these two factors can be examined by seeing which is more constraining, and favours the least female biased sex ratio (West *et al.*, 2002b). However, there is the possibility that these factors may interact – when both c and q are

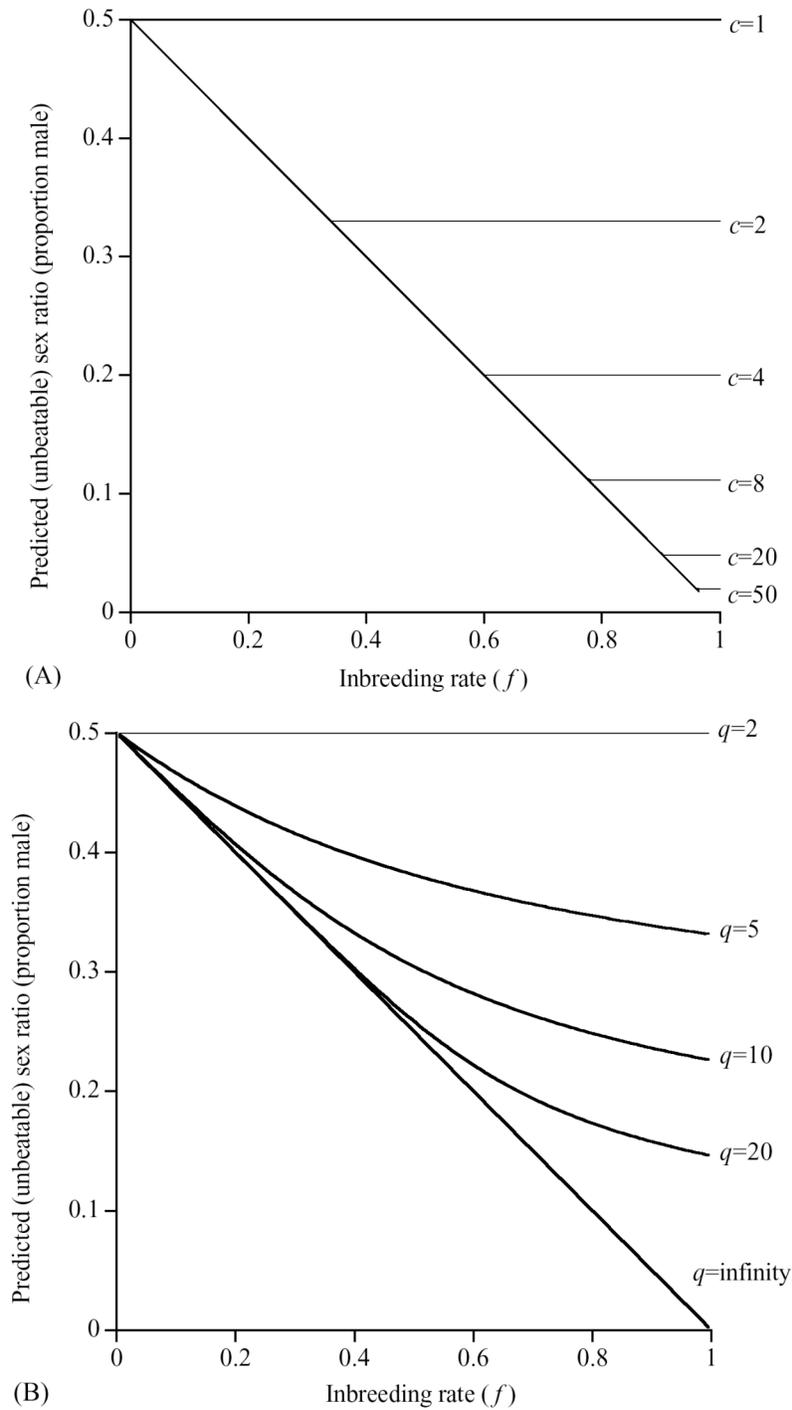


Figure 2.1. The relationship between the predicted unbeatable sex ratio (proportion of gametocytes that are male; z^*) and the inbreeding rate (f). (A) shows the unbeatable sex ratio when the number of gametes produced by each male gametocyte (c) varies and gametes from all gametocytes in a very large group can interact ($q \rightarrow \infty$; Read *et al.* 1992). (B) shows the unbeatable sex ratio when the number of gametocytes whose gametes can interact (q) is limited and the number of gametes produced by each male gametocyte (c) is not limiting (West *et al.* 2002b).

low, even if there are males in a mating group, these males may not be able to provide enough gametes to fertilise all the female gametes. Although this scenario could logically occur, it is not clear whether this interaction will significantly influence the ESS sex ratio. We use a stochastic analytical model to address this problem and consider how the unbeatable sex ratio is influenced by the interaction of finite values for both c and q . We use life history terminology associated with malaria parasites, but our results are applicable to any Apicomplexan parasite with dimorphic sexual stages.

Methods

We consider a large population of vertebrates harbouring malaria parasites and supporting a large number of blood-feeding dipteran vectors (effects due to small numbers of vertebrate hosts is negligible unless the number of hosts is extremely small; Taylor & Bulmer, 1980). Every host contains a large pool of haploid gametocytes circulating in the peripheral blood, comprising n independent lineages (all notation is given in table 2.1). Within a lineage, all gametocytes are clonally derived from a single sporozoite founder individual. Each lineage produces a proportion z of male gametocytes and $1-z$ of female gametocytes, where z is determined by a single biallelic nuclear gene. A common 'Null' allele exists at frequency $1-m$ and has $z = z_N$, and a vanishingly rare 'Mutant' allele exists at frequency m and has $z = z_M$. We may assign each host individual to one of $n+1$ classes on the basis of the number of Mutant lineages carried. Each host is fed upon by a large number of vectors, transmitting q gametocytes to each vector in the process. Once in the midgut of the vector, each male gametocyte gives rise to c male gametes and female gametocytes each give rise to a single female gamete. Random syngamy ensues, and the resulting next generation of zygotes are, following Read *et al.* (1992), assumed to reflect the genetic composition of the next generation of infections. It is worth noting that although each vector contains a single mating group of size q the predictions of this analysis will hold for any number of such groups, provided that there is no exchange of gametes between mating groups.

Symbol	Definition
$Bi(k, \square)$	Binomial distribution: k trials and probability of success \square
c	Number of viable male gametes per male gametocyte
f	Inbreeding coefficient; $f = n^{-1}$
g_x	Number of X-allele male gametes remaining viable
$HypGeo(\square, \square, \square)$	Hypergeometric distribution: \square trials, and \square potential successes out of \square
M	The Mutant allele
m	Population frequency of the mutant
N	The Null allele
n	Number of independent lineages per vertebrate host
p	Probability of male gamete survival
q	Number of gametocytes whose gametes can interact in the vector
$S_{x,y}$	Success of the X-allele in a host containing y Mutant infections
w_x	Absolute fitness of the X-allele
z	Sex ratio (proportion male gametocytes per lineage)
z^*	Evolutionarily stable (ES) sex ratio
z_x	Sex ratio employed by the X-allele
\square	Species-specific number of gametes released per male gametocyte
\square_k	Number of X-allele females in a mating group
\square_k	Number of X-allele males in a mating group
\square_k	Total number of X-allele gametocytes in a mating group
$\square_{x,y}$	Frequency of X-alleles in successful male ($y=1$) or female ($y=0$) gametes
\square	Relative fitness of the Null, w_N / w_M ; Mutant invades if $\square < 1$
\square	Number of zygotes produced by the mating group

Table 2.1. Definition of parameters, variables and distributions referred to in chapter 2.

The fitness of the Null is the mean success of a Null lineage from each host-class weighted by the number of Null lineages in the host-class and the frequency of that host-class. As the mutant is vanishingly rare, so that $m \ll 0$, the fitness of the Null is dominated by its success in vectors feeding upon hosts containing no Mutant lineages.

$$w_N \ll \frac{1}{n} S_{N,0} = f S_{N,0} \quad [2.1]$$

where $S_{N,0}$ is the mean number of zygotic Null alleles produced per vector feeding on a host harbouring zero Mutant lineages, and f is the degree of inbreeding. The Mutant never occurs in such hosts, and almost never occurs in hosts with other Mutant lineages,

so its fitness is dominated by its success in vectors feeding upon hosts with one Mutant lineage and $n-1$ Null lineages.

$$w_M \leq S_{M,1} \quad [2.2]$$

where $S_{M,1}$ is the mean number of zygotic Mutant alleles derived from a vector feeding on a host containing one Mutant infection only. The Mutant invades if $w_M > w_N$ and so the ESS sex ratio z^* is the value of z_N , such that $\bar{z} = w_N / w_M$ is not less than unity for all $0 \leq z_M \leq 1$. Exact solutions for $S_{N,0}$ and $S_{M,1}$ will be determined, so that for known q , c and f pairs of sex ratio strategies may be compared.

A vector feeding on a Null-only host is assured of obtaining q Null gametocytes in its bloodmeal. $\bar{z}_N \sim Bi(q, z_N)$ are male, where $Bi(k, \pi)$ represents the binomial distribution with k trials and probability of success π , and the remaining $\bar{z}_N = q - \bar{z}_N$ are female, so that there are $c \bar{z}_N$ male gametes and \bar{z}_N female gametes able to interact in the midgut. The number of zygotes, \bar{z} is the smaller of these two values, and since zygotes are diploid the number of Null alleles formed in that vector is $2 \bar{z}$.

$$S_{N,0} = \sum_{\bar{z}_N=0}^q \binom{q}{\bar{z}_N} z_N^{\bar{z}_N} (1 - z_N)^{q - \bar{z}_N} 2 \min\{c \bar{z}_N, \bar{z}_N\} \quad [2.3]$$

A vector feeding on a host containing one Mutant and $n-1$ Null lineage will obtain q gametocytes of which $\bar{z}_M \sim Bi(q, f)$ are Mutant and $\bar{z}_N = q - \bar{z}_M$ are Null. These will comprise $\bar{z}_M \sim Bi(\bar{z}_M, z_M)$ Mutant males and $\bar{z}_M = \bar{z}_M - \bar{z}_M$ Mutant females, and $\bar{z}_N \sim Bi(\bar{z}_N, z_N)$ Null males and $\bar{z}_N = \bar{z}_N - \bar{z}_N$ Null females. The number of zygotes, \bar{z} is then the lower of the two values $c (\bar{z}_M + \bar{z}_N)$ and $\bar{z}_M + \bar{z}_N$, meaning that there are \bar{z} successful male gametes and \bar{z} successful female gametes. Of the former, a proportion $\bar{z}_{M,1} \sim HypGeo(\bar{z}, c \bar{z}_M, c (\bar{z}_M + \bar{z}_N) / \bar{z})$ will be Mutant, where $HypGeo(\bar{z}, \bar{z}_M, \bar{z})$ represents the hypergeometric distribution with \bar{z} trials and \bar{z}_M potential successes out of $\bar{z}_M + \bar{z}_N$ and of the latter a proportion

$\bar{m}_{M,0} \sim \text{HypGeo}(\bar{m}_M, \bar{m}_M + \bar{m}_N) / \bar{m}$ will be Mutant. The success of the Mutant is simply $\bar{m} / (\bar{m}_M + \bar{m}_{M,0})$ (Taylor, 1981; Charnov, 1982a).

$$S_{M,1} = \sum_{\bar{m}_M=0}^q \sum_{\bar{m}_M=0}^{\bar{m}_M} \sum_{\bar{m}_N=0}^{q-\bar{m}_M} \frac{q!}{\bar{m}_M! \bar{m}_N!} c^{\bar{m}_M} (1-f)^{q-\bar{m}_M} \frac{\bar{m}_M!}{\bar{m}_M!} z_M^{\bar{m}_M} (1-z_M)^{\bar{m}_M} \frac{q!}{\bar{m}_M! \bar{m}_N!} z_N^{\bar{m}_M} (1-z_N)^{q-\bar{m}_M} \quad [2.4A]$$

$$\min\{c(\bar{m}_M + \bar{m}_N), q \bar{m}_M \bar{m}_N\} (E[\bar{m}_{M,1}] + E[\bar{m}_{M,0}])$$

where

$$E[\bar{m}_{M,1}] = \begin{cases} \frac{\bar{m}_M}{\bar{m}_M + \bar{m}_N} & \text{if } \bar{m}_M + \bar{m}_N > 0 \\ 0 & \bar{m}_M + \bar{m}_N = 0 \end{cases} \quad [2.4B]$$

$$E[\bar{m}_{M,0}] = \begin{cases} \frac{\bar{m}_M \bar{m}_M}{q \bar{m}_M \bar{m}_N} & \text{if } q \bar{m}_M \bar{m}_N > 0 \\ 0 & q \bar{m}_M \bar{m}_N = 0 \end{cases} \quad [2.4C]$$

These expressions reveal whether the Mutant allele can invade a population fixed for the Null. We determined the ESS sex ratio iteratively, such that the value of z_N in each round is the sex ratio of the successfully invading Mutant or successfully defending Null of the previous round, and z_M is a randomly assigned value. After an indefinite number of rounds the Null will assume and subsequently retain the value of z^* , so that at any time the currently unbeaten z can be tested for evolutionary stability by plotting \bar{m} for z_N equal to the putative z^* against all $0 \leq z_M \leq 1$ and rejecting if $\bar{m} < 1$ for any z_M .

To check our expressions, we will now derive expressions [2.3] and [2.4] for the special cases where q or c are infinite, i.e. corresponding to the analyses of Read *et al.* (1992) and West *et al.* (2002b). In both cases, we find that the results agree with these previous analyses.

In West *et al.* (2002b) the implications of finite mating group size for fertility insurance were made amenable for mathematical treatment by assuming limitless male fecundity.

This represents a special case of our model, such that $c \rightarrow \infty$ and equations [2.3] and [2.4] reduce to:

$$S_{N,0} = \sum_{\substack{q=0 \\ \Delta_N=0}}^q \binom{q}{\Delta_N} \binom{q}{\Delta_N} z_N^{\Delta_N} (1 - z_N)^{q - \Delta_N} \quad [2.5A]$$

where

$$\Delta_N = \begin{cases} q - \Delta_N & \text{if } \Delta_N > 0 \\ 0 & \Delta_N = 0 \end{cases} \quad [2.5B]$$

and

$$S_{M,0} = \sum_{\Delta_M=0}^q \sum_{\Delta_M=0}^{\Delta_M} \sum_{\Delta_N=0}^{q - \Delta_M} \binom{q}{\Delta_M} \binom{q - \Delta_M}{\Delta_N} f^{\Delta_M} (1 - f)^{q - \Delta_M} \binom{\Delta_M}{\Delta_M} \binom{\Delta_M}{\Delta_M} z_M^{\Delta_M} (1 - z_M)^{\Delta_M - \Delta_M} \binom{q - \Delta_M}{\Delta_N} \binom{q - \Delta_M}{\Delta_N} z_N^{\Delta_N} (1 - z_N)^{q - \Delta_M - \Delta_N} \quad [2.6A]$$

$$\Delta (E[\Delta_{M,1}] + E[\Delta_{M,0}])$$

where

$$\Delta = \begin{cases} q - \Delta_M - \Delta_N & \text{if } \Delta_M + \Delta_N > 0 \\ 0 & \Delta_M + \Delta_N = 0 \end{cases} \quad [2.6B]$$

$$E[\Delta_{M,1}] = \begin{cases} \frac{\Delta_M}{\Delta_M + \Delta_N} & \text{if } \Delta_M + \Delta_N > 0 \\ 0 & \Delta_M + \Delta_N = 0 \end{cases} \quad [2.6C]$$

$$E[\Delta_{M,0}] = \begin{cases} \frac{\Delta_M \Delta_M}{q \Delta_M \Delta_N} & \text{if } q \Delta_M \Delta_N > 0 \\ 0 & q \Delta_M \Delta_N = 0 \end{cases} \quad [2.6D]$$

Conversely, in the deterministic analysis of Read *et al.* (1992), the fertility insurance consequences of limited male fecundity were investigated under the assumption of large mating group size. This special case, $q \rightarrow \infty$, reduces equations [2.3] and [2.4] to give

$$S_{N,0} = 2q \min\{c z_N, (1 - z_N)\} \quad [2.7]$$

and

$$S_{M,1} = q \min\{c(z_M f + z_N(1 - f)), (1 - z_M)f + (1 - z_N)(1 - f)\} \\ \frac{z_M f}{z_M f + z_N(1 - f)} + \frac{(1 - z_M)f}{(1 - z_M)f + (1 - z_N)(1 - f)} \quad [2.8]$$

Although both $S_{N,0}$ and $S_{M,1}$ are linear functions of q , and therefore have infinite solutions, the relative fitness of the Null allele may still be evaluated as \bar{r} is the ratio of the two and hence is finite. The predictions converge with those of Read *et al.* (1992) for $c \geq 1$, but being more general, are able to predict the male biased ESS sex ratio when males fecundity is more limiting than that of females, so that $c < 1$.

We considered the possibility of stochastic male fecundity, specifically, how accurately do expressions [2.3] and [2.4] predict the ESS sex ratio when the value of c represents the expectation of a random variable? Assuming that males all produce the same species-specific number (\bar{g}_N) of gametes of which a proportion p will be viable for fertilization, [2.3] and [2.4] become

$$S_{N,0} = \frac{q}{\bar{g}_N} \frac{\bar{g}_N}{g_N} \frac{q}{\bar{g}_N} \frac{z_N}{(1 - z_N)^{q/\bar{g}_N}} \frac{\bar{g}_N}{g_N} p^{g_N} (1 - p)^{\bar{g}_N/g_N} 2 \min\{g_N, q/\bar{g}_N\} \quad [2.9]$$

and

$$S_{M,1} = \prod_{\substack{q \\ \square_M=0 \\ \square_M=0 \\ \square_N=0 \\ g_M=0 \\ g_N=0}} \prod_{\substack{\square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M}} \prod_{\substack{q \\ \square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M}} \prod_{\substack{\square_N \\ \square_N \\ \square_N \\ \square_N \\ \square_N \\ \square_N}} \prod_{\substack{g \\ \square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M}} (1-f)^{q \square_M} \prod_{\substack{\square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M}} \prod_{\substack{z \\ \square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M}} (1-z_M)^{\square_M} \prod_{\substack{\square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M}} \prod_{\substack{q \\ \square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M}} \prod_{\substack{\square_N \\ \square_N \\ \square_N \\ \square_N \\ \square_N \\ \square_N}} (1-z_N)^{q \square_M \square_N} \prod_{\substack{g \\ \square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M}} \prod_{\substack{g \\ \square_M \\ \square_M \\ \square_M \\ \square_M \\ \square_M}} p^{g_M+g_N} (1-p)^{\square(\square_M+\square_N)\square g_M \square g_N} \min\{g_M + g_N, q \square \square_M \square \square_N\} (E[\square_{M,1}] + E[\square_{M,0}]) \quad [2.10A]$$

where

$$E[\square_{M,1}] = \begin{cases} \frac{g_M}{g_M + g_N} & \text{if } g_M + g_N > 0 \\ 0 & g_M + g_N = 0 \end{cases} \quad [2.10B]$$

$$E[\square_{M,0}] = \begin{cases} \frac{\square_M \square \square_M}{q \square \square_M \square \square_N} & \text{if } q \square \square_M \square \square_N > 0 \\ 0 & q \square \square_M \square \square_N = 0 \end{cases} \quad [2.10C]$$

Results and Discussion

We have discriminated between two types of fertility insurance, in response to (i) low male gamete fertility (low c), and (ii) the ability of gametes to interact (low q). Previous theoretical work has examined the effect of these two types of fertility insurance separately. Specifically, West *et al.* (2002b) assumed that when both of these factors are operating, the effect for sex ratio evolution can be determined by seeing which leads to a greater reduction in the predicted female bias (i.e. which of figures 2.1A and 2.1B predicts the least female biased sex ratio). In contrast, our model explicitly allows for both types of fertility insurance to act simultaneously, and hence allows for any interactions. In figures 2.2-2.4 we give example predictions when the two types of fertility insurance are allowed to act separately as previously assumed by West *et al.* (2002b) (part A of the figures) or simultaneously in our model (part B of the figures). Our results show that when both c and q are low, the ESS sex ratio may be higher than predicted when considering these two effects separately.

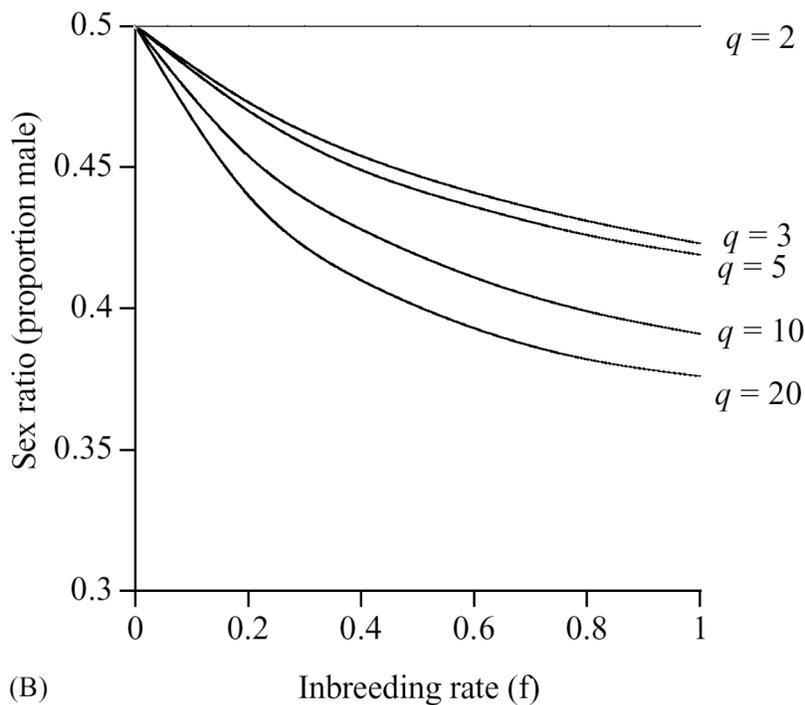
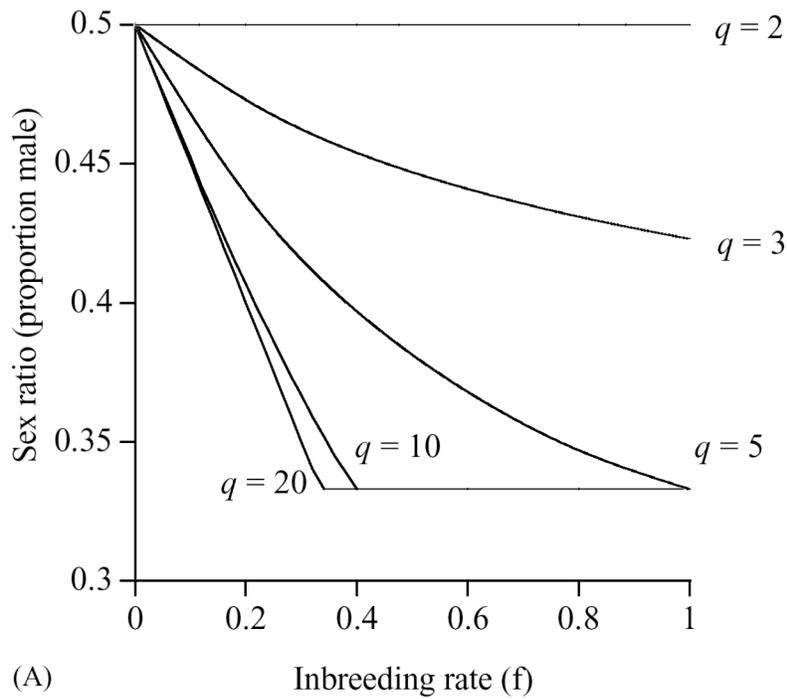


Figure 2.2. (A) shows the relationship between predicted sex ratio and inbreeding rate, for given values of q when $c = 2$ assuming no interaction between the two types of fertility insurance and (B) shows the relationship between ESS sex ratio and inbreeding rate arising from equations 2.1-2.4, for given values of q when $c = 2$.

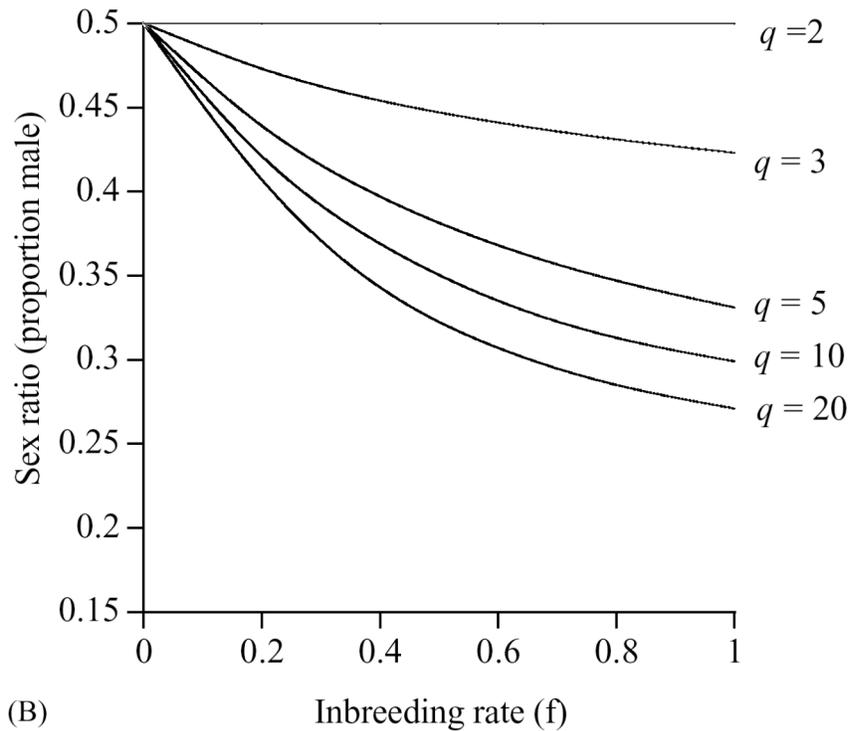
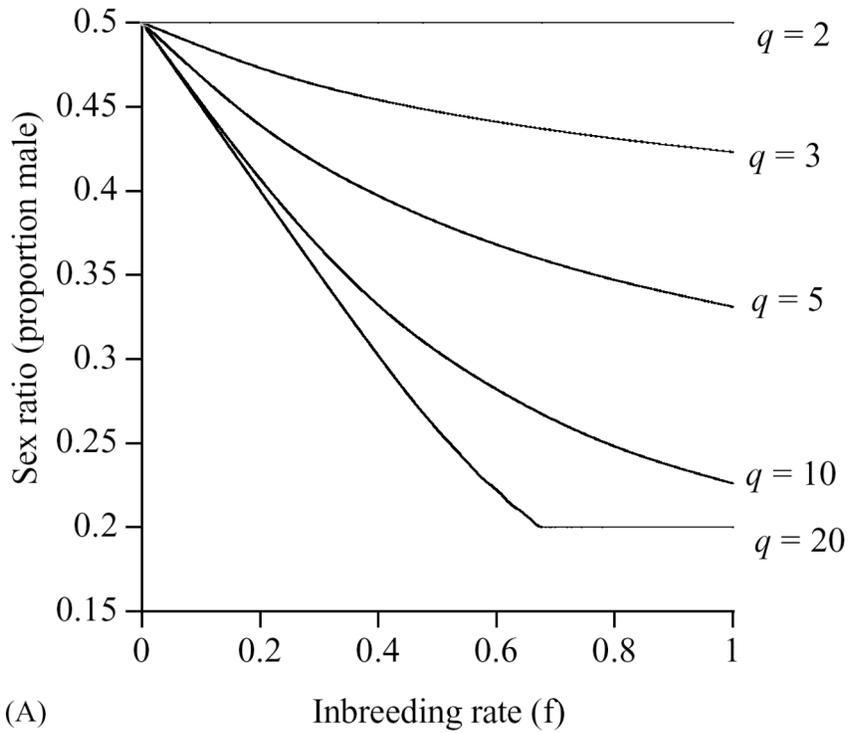


Figure 2.3. (A) shows the relationship between predicted sex ratio and inbreeding rate, for given values of q when $c = 4$ assuming no interaction between the two types of fertility insurance and (B) shows the relationship between ESS sex ratio and inbreeding rate arising from equations 2.1-2.4, for given values of q when $c = 4$.

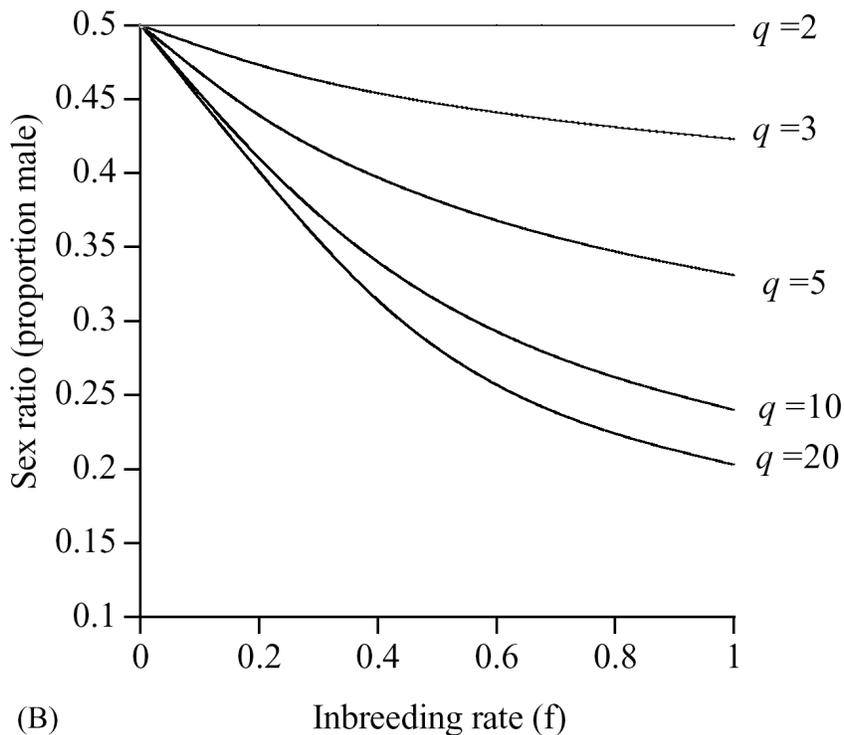
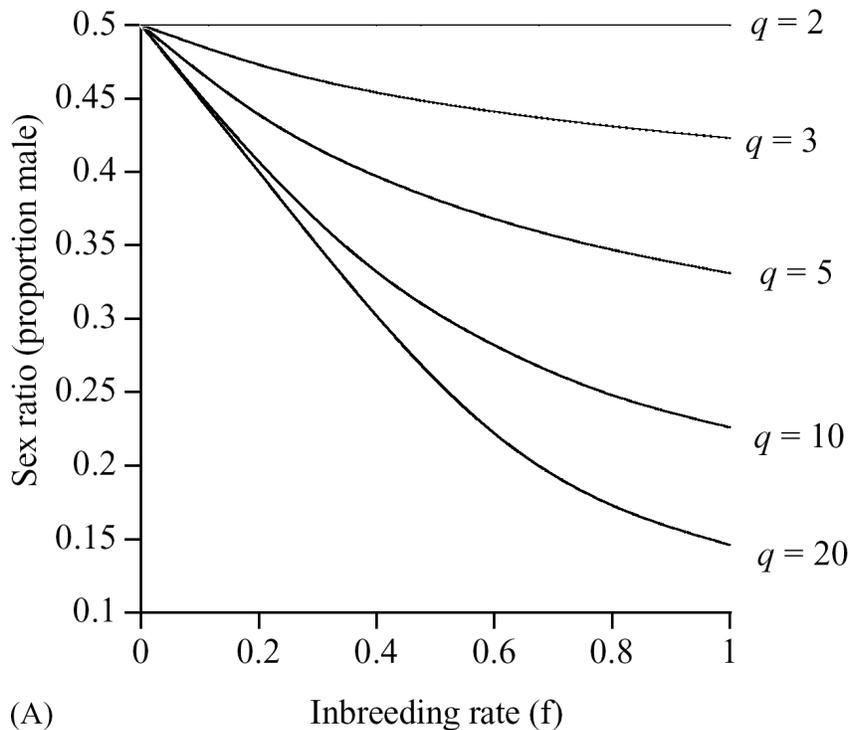


Figure 2.4. (A) shows the relationship between predicted sex ratio and inbreeding rate, for given values of q when $c = 8$ assuming no interaction between the two types of fertility insurance and (B) shows the relationship between ESS sex ratio and inbreeding rate arising from equations 2.1-2.4, for given values of q when $c = 8$.

Why does our model predict a less female biased sex ratio? It has been assumed that one male gametocyte will be able to provide enough gametes to fertilise all the female gametes in the mating group that arises from q gametocytes. This is not the case if $(q-1) > c$. More generally, the male gametocytes will not be able to fertilise all the female gametes when $(q-\bar{q}) > c\bar{q}$, where \bar{q} is the number of male gametocytes in a mating group. This risk of not having enough males to fertilise the females in the group leads to less female biased sex ratios being favoured. Another way of conceptualising this is that a finite q increases the potential for low c to be a problem – when gametes can not interact as successfully (finite q), a mating group may contain only a single or small number of male gametocytes, and so the gamete fecundity (c) of these males is more likely to be a limiting factor.

Our model shows that the interaction between the two types of fertility insurance can have a surprisingly large influence on the ESS sex ratio. In the examples that we give, the predicted sex ratio can be up to 0.1 higher (figure 2.2, when $c=2$, $q=10$ and $f=0.3$). In this instance the sex ratio deviates from equality (0.5) by approximately half the amount inferred by West *et al.* (2002b). Although increasing c proportionally reduces the degree of female bias, the complex interplay between male fecundity and size of mating groups makes it difficult to relate the magnitude of this effect to q . In the limit, as q increases towards infinity, the effect dissipates as the predictions converge with those of Read *et al.* (1992). However, as q rises it increases the propensity for c to become limiting. The effect is therefore a dome-shaped function of q , although the exact relationship crucially depends upon the particular parameter values.

We also extended our model to allow stochastic variability in the number of viable gametes per gametocyte (c) – see expressions [2.9] and [2.10]. This could occur through variation in the number of gametes produced per gametocyte, or through mortality. Adding in this stochasticity (for invariant $E[c]$) gives further reduction in the female bias predicted, although this effect is negligible in all but the smallest of mating groups. However, a novel prediction arises from this form of stochasticity, as it allows the investigation of the mean value of $c < 1$, so that male fecundity is lower than that of

females. In this case, a male biased sex ratio is favoured. For the case of $q \ll \infty$ equations [2.7] and [2.8] remain valid even for $c < 1$, and male biased ESS sex ratios are easily demonstrated. Switching the roles of males and females in the classic LMC relation, the result of Read et al. (1992) can be extended so that, as before, for $c \geq 1$ $z^* = \max\{(1-f)/2, 1/(c+1)\}$, yet now for $c \leq 1$ $z^* = \min\{(1+f)/2, 1/(c+1)\}$. This prediction contrasts with standard LMC models constructed for insects (e.g. Nagelkerke & Hardy, 1998; West & Herre, 1998), where male biased sex ratios are never predicted, due to the assumption that one male can mate any number of females (analogous to assuming $c \ll \infty$). Male biased sex ratios have been observed in some samples of lizard malaria (Paperna & Landau, 1991), although the necessarily small sample sizes mean that these observations should be treated with caution.

To conclude, our analysis has revealed that fertility insurance can be a more potent evolutionary buffer to female biased sex ratios in malaria and related parasites than previously suggested. Clearly, the outstanding problem is to obtain empirical estimates of c and q , and how their values are influenced by factors such as host immune responses. The existing literature has recently been reviewed (West *et al.*, 2002b), and sadly very little is known.

3. A dimensionless invariant for relative size at sex change in animals: explanations and implications

Abstract

Recent comparative studies across sex changing animals have found that the relative size and age at sex change are strikingly invariant. In particular, 91-97% of the variation in size at sex change across species can be explained by the simple rule that individuals change sex when they reach 72% of their maximum body size. However, this degree of invariance is surprising, and has proved controversial. In particular, it is not clear why this result should hold given that there is considerable biological variation across species in factors that can influence the evolutionarily stable timing of sex change. Our overall aim here is to explain this result, and determine the implications for other life history variables. Specifically, we use a combination of approaches to: (a) formalise previous analytical theory in this area; (b) examine the robustness of the empirical invariance result; (c) carry out sensitivity analyses to determine what the empirical data imply about the mean value and variation in several key life history variables. Our results demonstrate the usefulness of the dimensionless approach for explaining empirical patterns, and making novel testable predictions.

Introduction

Sex change occurs in a variety of animals, including fish, echinoderms, crustaceans, molluscs and polychaete worms (Charnov 1982a; Policansky 1982; Allsop & West 2004a). Evolutionary theory suggests that sex change is favoured when: (i) the reproductive success of an individual varies with their age or size, and (ii) the relationship differs between the sexes. In this case, natural selection favours a strategy where individuals start as the sex whose fitness increases more slowly with age, and then change to the other sex at a later stage (Ghiselin 1969; Leigh et al. 1976; Charnov 1982a;

Warner 1988a, b). Although there are considerable differences in details across species, this basic idea is well established and has been widely applied to numerous animals and plants.

A recent development in the field of sex change has been the use of a dimensionless approach. Charnov and Skúladóttir (2000) pointed out that predictions for the evolutionarily stable (Maynard Smith & Price 1973) size or age at sex change could be expressed in terms of several dimensionless quantities. Specifically, their results depend upon k/M , $\frac{1}{M}$ and $\frac{1}{\alpha}$, which are the relative growth rates (k , the Bertalanffy coefficient; Bertalanffy 1938), the adult instantaneous mortality rate (M), the age at first breeding ($\frac{1}{\alpha}$), and the coefficient in the equation relating male fertility to size ($\frac{1}{\alpha}$; where male fertility is proportional to L^α , and L is length (size)). Charnov and Skúladóttir (2000) showed that populations/species with the same values of these dimensionless quantities are predicted to have the same: (1) relative size at sex change, given by size at sex change / maximum size (L_{50}/L_{Max}); (2) relative age at sex change, given by age at sex change / age at first breeding ($\frac{1}{\alpha}$); (3) breeding sex ratio, defined as the proportion of breeders that are male. Several studies have suggested that $\frac{1}{M}$ and k/M can be invariant within and even across taxa (Charnov 1993; Gemmill et al. 1999). Consequently, an invariant relative size and age at sex change are predicted whenever $\frac{1}{\alpha}$ is also invariant.

It could be expected that $\frac{1}{\alpha}$ would be roughly invariant within species, or across closely related species sharing a similar life history. Consistent with this, Charnov and Skúladóttir (2000; see also Skúladóttir & Petursson 1999) found an invariant relative size at sex change across populations of the shrimp *Pandalus borealis* in the waters off Iceland. However, the invariant relationship also holds across species. Allsop and West (2003b) found that the relative size and age at sex change were invariant across fish species. They then extended their study to include data from all sex changing animals, including data from 77 species of fish, echinoderm, crustacean and mollusc, and found that the relative size at sex change was invariant across all species (Allsop & West 2003a). Specifically, that 91-97% of the variation in size at sex change across species can be explained by the rule that species change sex when they reach 72% of their maximum

size. This result holds despite huge variation in size and biology across species, from a 2 mm long crustacean, to a 1.5 m long fish (figure 3.1).

This extent of invariance in the relative size at sex change across species was relatively surprising, and has proved controversial (Millius 2004; Allsop & West 2004c; Buston et al. 2004). The main reason for this is that there is considerable biological variation across species in factors that could influence the evolutionarily stable timing of sex change. In particular: (1) Sex change is from male to female (protandrous) in some species, female to male (protogynous) in others, and even in both directions in some (Nakashima et al. 1995). Given that female fertility is likely to be proportional to body size cubed (L^3 ; Charnov 1979, 1993), this means that for sex change to be favoured there must be variation in λ between protandrous ($\lambda < 3$) and protogynous ($\lambda > 3$) species. (2) In some species there is a fraction of individuals who mature early as the second sex, termed early maturers, which is likely to correlate with or cause variation in λ (Charnov 1982a; Charnov 1982b; Warner 1984; Allsop & West 2004a, b). (3) There is a huge variation in mating system across species. For example, even within just the fish species, the range of mating systems includes harems, leks, monandry, and aggregation spawners (Warner 1984; Allsop & West 2003b). (4) The cues and mechanisms involved in sex change vary across species, dependent upon factors such social environment, size and parasitism (Shapiro 1980; Shapiro & Lubbock 1980; Shapiro 1981; Scharer & Vizoso 2003). (5) While λM is likely to be invariant across all sex changing species, with a value of approximately 2, it is not clear that k/M is similarly invariant (Charnov & Berrigan 1990; Charnov 1993; Gemmill et al. 1999).

Our overall aim here is to determine how Allsop and West's (2003a,b) invariant results can hold in the face of so much biological variation across species, and the implications for other life history variables. Our paper is divided into three sections. In the first section, we test the underlying generality of the sex change invariant predictions. Specifically, we: (a) provide a more formal analytical proof of Charnov and Skúladóttir's (2000) invariant predictions; (b) show that the sex change invariants can also be predicted by an alternative modeling approach, assuming that sex change occurs in response to

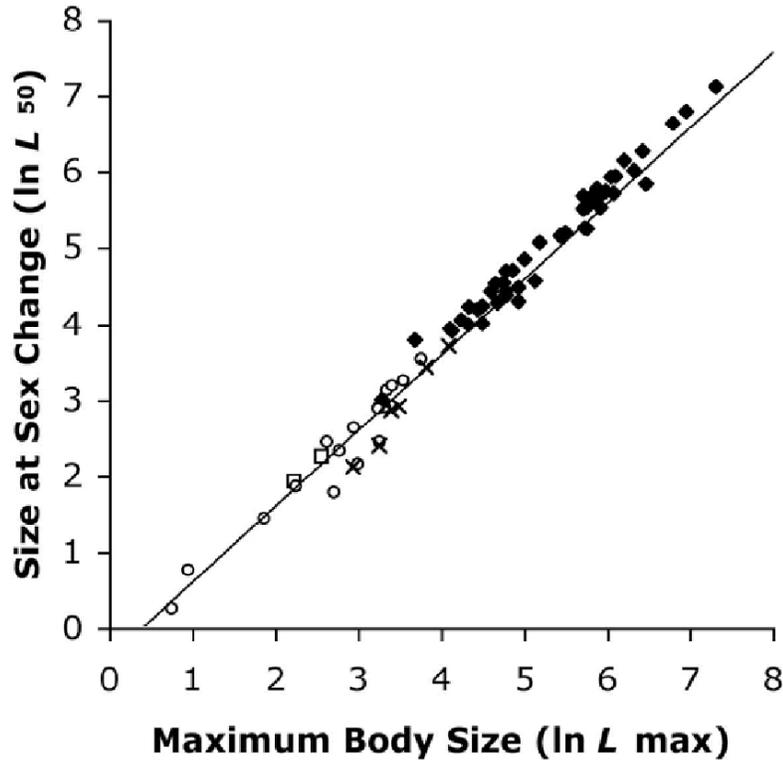


Figure 3.1. Log-Log plot of L_{50} vs L_{Max} for 77 species of sex changing animals with species as independent data points. Data split by taxa: \square = Echinodermata, \circ = Crustacea, \blacklozenge = Chordata, \square = Mollusca. The regression has a slope fixed at 1 giving an intercept of -0.32 ± 0.05 (95% CI), ($r^2 = 0.97$, $n = 77$ species). The OLS slope is 1.05 ± 0.03 (95% CI) with an intercept of -0.55 ± 0.07 (95% CI) ($r^2 = 0.98$). The mean relative size at sex change (L_{50}/L_{Max}) is 0.72 (95% CI; $0.67-0.77$), implying that individuals change sex when they reach 72% of their maximum size. Size (L_{50} & L_{Max}) is measured in mm prior to logarithmic transformation.

social environment, as is known to occur in some fish (Shapiro 1981; Warner & Swearer 1991; Allsop & West 2004b), rather than in response to age or size, as assumed by Charnov and Skúladóttir's (2000) model.

In the second section, we examine the statistical significance of Allsop and West's (2003a,b) empirical result showing that the relative size at sex change is invariant. Buston et al. (2004; Milius 2004) have criticised Allsop and West's analysis, and argued that an alternative null model approach based upon randomization techniques does not support the invariant prediction. We: (c) assess the validity of Buston et al's analysis; (d) develop

and test a more appropriate null model. This section quantifies and expands upon some issues we have raised in a previous short comment (Allsop & West 2004c).

In the third section, we carry out a sensitivity analysis based on Charnov and Skúladóttir's model to test how variation in the model's key life history parameters (\bar{M} , k/M and \bar{L}) influences the predicted relationship between size at sex change and maximum size, and hence the relative size at sex change invariant. Specifically, we: (e) use the existing information on \bar{M} , k/M and the relative size at sex change to estimate \bar{L} ; (f) estimate how variation in \bar{M} , k/M and \bar{L} influences the extent and nature of the relative size at sex change invariant that would be expected; (g) estimate the amount of variation in \bar{M} and k/M that is consistent with the empirical data.

When is an invariant relative size at sex change expected?

Formalising Charnov & Skulladotir's Invariant Predictions

In this section we determine the conditions for invariance in relative age and size at sex change, as predicted by Charnov and Skúladóttir (2000). Assuming a stationary age structure, the evolutionarily stable strategy (ESS; Maynard Smith & Price 1973) for timing of sex change maximizes the fitness function:

$$w \mu S_1 \bar{L} S_2 \quad [3.1]$$

where S_1 is the total reproductive success as the first sex and S_2 is the total reproductive success as the second sex (Leigh et al. 1976, Charnov 1979, 1982). The reproductive success gained through the function of the first sex is given by integrating over all ages (x) from maturity (at age \bar{L}) to sex change (at age \bar{L}) the product of instantaneous reproductive rate ($Q_1(x)$) and the probability ($p(x)$) of surviving to age x :

$$S_1 = \int_{\bar{L}}^{\infty} p(x)Q_1(x)dx \quad [3.2]$$

Following Charnov and Skúladóttir (2000), we assume: (1) a constant instantaneous mortality rate (M), such that the probability of survival takes the form $p(x) = A \exp(-Mx)$, where A is a constant; (2) that the instantaneous reproductive rate takes the form of the power law $Q_1(x) = B_1 L(x)^{\alpha_1}$, where $L(x)$ is the size at age x , B_1 is a constant, and α_1 is the fitness exponent relating reproductive rate for the first sex; and (3) that size is related to age according to the Bertalanffy function $L(x) = L_{\text{Max}} (1 - \exp(-k_1 x))$, where L_{Max} is the asymptotic (maximum) size of the species and k is the Bertalanffy growth coefficient. This equation has been used successfully to model growth of indeterminate growers (Beverton 1963; Pauly 1980; Beverton 1992). Hence we can write

$$S_1 = AB_1 L_{\text{Max}}^{\alpha_1} \int_0^{\infty} e^{-Mx} (1 - e^{-k_1 x})^{\alpha_1} dx \quad [3.3]$$

and for the second sex, with fecundity given by the power law $Q_2(x) = B_2 L(x)^{\alpha_2}$, we have

$$S_2 = AB_2 L_{\text{Max}}^{\alpha_2} \int_0^{\infty} e^{-Mx} (1 - e^{-k_1 x})^{\alpha_2} dx \quad [3.4]$$

so that fitness may be written as a function of the age at sex change (x):

$$w \mu \left[\int_0^x e^{-Mx} (1 - e^{-k_1 x})^{\alpha_1} dx \right] \left[\int_x^{\infty} e^{-Mx} (1 - e^{-k_1 x})^{\alpha_2} dx \right] \quad [3.5]$$

(Charnov & Skúladóttir 2000). Since the exponents of the power laws relating reproductive rate to size may be different for the two sexes (e.g. $\alpha_1 \neq \alpha_2$), fitness is a function of the direction, as well as the timing, of sex change. The sex with the larger power law exponent has the greatest relative improvement in reproductive rate as the individual grows, and so fitness is maximised by reproducing as this sex when large, and as the other sex when small. Sex change is therefore predicted to be in the direction of increasing power law exponent, i.e. $\alpha_1 \leq \alpha_2$ (Leigh et al. 1976, Warner et al. 1975).

It is generally true that for indeterminate growers female fecundity scales approximately with the cube of size ($\lambda_{\text{female}} \approx 3$; references in Charnov 1979, 1993), so we may drop subscripts and, following the notation of Charnov and Skúladóttir (2000), set $\lambda_{\text{female}} = 3$ and $\lambda_{\text{male}} = \lambda$. From the above argument, we know that $\lambda_i = \text{Min}\{3, \lambda\}$ and $\lambda_b = \text{Max}\{3, \lambda\}$, i.e. sex change is predicted to be from male to female (protandry) when $\lambda < 3$, and from female to male (protogyny) when $\lambda > 3$. Obviously when $\lambda = 3$ sex change is not favoured (Leigh et al. 1976, Charnov 1982a). Hence, fitness can be expressed as

$$w \mu = \begin{cases} \int_{\lambda}^{\infty} e^{\lambda M x} (1 - e^{\lambda k x})^{\lambda} dx & \int_{\lambda}^{\infty} e^{\lambda M x} (1 - e^{\lambda k x})^3 dx \\ \int_{\lambda}^{\infty} e^{\lambda M x} (1 - e^{\lambda k x})^3 dx & \int_{\lambda}^{\infty} e^{\lambda M x} (1 - e^{\lambda k x})^{\lambda} dx \end{cases} \quad \text{if} \quad \begin{matrix} \lambda < 3 \\ \lambda > 3 \end{matrix} \quad [3.6]$$

Charnov and Skúladóttir (2000) applied Buckingham's π theorem (Buckingham 1914, Stephens & Dunbar 1993), which suggests that the fitness function [3.6] could, in principle, be rewritten as a function of constants and the dimensionless life-history parameters λ/λ , λM , k/M and λ . Since λ and λ have units of time, k and M have units of inverse time, and λ is an exponent in a power function, then the values λ/λ , λM , k/M and λ are all unitless. Being able to express fitness in these terms indicates that the fitness function [3.6] is invariant for circumstances where λ/λ , λM , k/M and λ are invariant, and hence fitness is predicted to be maximized (for example) at the same relative age at sex change (λ/λ) in all contexts where the other values (λM , k/M and λ) are invariant.

We will now derive an explicit fitness function in terms of the aforementioned dimensionless quantities, using the standard technique of switching variables. The key to this derivation is to understand that since the units of time are arbitrary (they cancel to give a dimensionless fitness value), we can rescale time by introducing an arbitrary scaling constant ($c; x = c y$). For instance, if relating x seconds to y minutes, the scaling constant is $c = 60$. Hence, where we see x in the fitness function, we can substitute in $c y$. The limits of the integrals also need to be rescaled, for instance the lower bound of the first integral is $x = \lambda$ time units, hence rescaling gives $y = \lambda/c$ time units. Finally, since x

= $c y$, we can write $dx/dy = c$, and hence $dx = c dy$, so that we may replace dx with $c dy$.

This rescaling gives a new fitness function

$$w \mu \begin{cases} \int_0^{\tau/c} e^{Mcy} (1 - e^{kcy})^{\beta} dy \int_0^{\tau/c} e^{Mcy} (1 - e^{kcy})^3 dy & \beta < 3 \\ \int_0^{\tau/c} e^{Mcy} (1 - e^{kcy})^3 dy \int_0^{\tau/c} e^{Mcy} (1 - e^{kcy})^{\beta} dy & \beta > 3 \end{cases} \quad \text{if} \quad [3.7]$$

The next step is to rescale so that τ units on the x -scale (the time to maturity) is equal to 1 unit on the y -scale, i.e. rescaled time is measured in units of ‘maturation time’ ($y = x/\tau$, and hence $c = \tau$). Substituting this into [3.7] yields

$$w \mu \begin{cases} \int_0^1 e^{M\tau y} (1 - e^{k\tau y})^{\beta} dy \int_0^1 e^{M\tau y} (1 - e^{k\tau y})^3 dy & \beta < 3 \\ \int_0^1 e^{M\tau y} (1 - e^{k\tau y})^3 dy \int_0^1 e^{M\tau y} (1 - e^{k\tau y})^{\beta} dy & \beta > 3 \end{cases} \quad \text{if} \quad [3.8]$$

Finally, we can rearrange in terms of our dimensionless quantities:

$$w \mu \begin{cases} \int_0^1 e^{\tau M y} (1 - e^{\tau(k/M)y})^{\beta} dy \int_0^1 e^{\tau M y} (1 - e^{\tau(k/M)y})^3 dy & \beta < 3 \\ \int_0^1 e^{\tau M y} (1 - e^{\tau(k/M)y})^3 dy \int_0^1 e^{\tau M y} (1 - e^{\tau(k/M)y})^{\beta} dy & \beta > 3 \end{cases} \quad \text{if} \quad [3.9]$$

This is the fitness function that Charnov and Skúladóttir (2000) predicted but did not find an explicit expression for. We have confirmed that fitness can be written in terms of the dimensionless quantities τ/τ , τM , k/M and β , and have shown that it takes the form of expression [3.9]. We can check that these dimensionless quantities constitute a full set. Where there are v variables and u orthogonal units, the expression can be rewritten in terms of $v-u$ dimensionless quantities (Buckingham 1914, Stephens & Dunbar 1993). Here, we have five variables (τ , M , k & β ; $v=5$) and one unit (time; $u=1$) and hence a set of four dimensionless numbers.

Note that the marginal fitness $dw/d(\bar{x})$ will have the same sign for all circumstances where \bar{x} , \bar{M} , k/M and \bar{c} are invariant. This means that for all situations in which \bar{M} , k/M and \bar{c} are invariant, the ESS relative age at sex change (\bar{x}^*/\bar{c} , such that $dw/d(\bar{x}) = 0$ and $d^2w/d(\bar{x})^2 < 0$ when evaluated at $\bar{x} = \bar{x}^*/\bar{c}$) will be an invariant. Under these circumstances we also predict further invariants. The size at sex change (L_{50}) is given by the Bertalanffy equation $L(\bar{x}^*) = L_{\text{Max}} (1 - \exp(-k \bar{x}^*))$. Since $k \bar{x}^* = \bar{M} \bar{c} k/M \bar{c} \bar{x}^*/\bar{c}$ is invariant, the size at sex change relative to asymptotic size (L_{50}/L_{Max}) is predicted to be invariant (Charnov & Skúladóttir 2000). The relative size at maturity ($L_{\text{Mat}}/L_{\text{Max}} = 1 - \exp(-\bar{c}k)$) is also expected to be an invariant, since $\bar{c}k = \bar{M} \bar{c} k/M$ is an invariant. Finally, the ratio of the number of breeders of the first sex (N_1) and the number of breeders of the second sex (N_2) is

$$\frac{N_1}{N_2} = \frac{\int_0^{\bar{x}^*} e^{-\bar{M}x} dx}{\int_{\bar{x}^*}^{\infty} e^{-\bar{M}x} dx} = \frac{e^{-\bar{M}\bar{x}^*}}{e^{-\bar{M}\bar{c}}} \bar{c} = \frac{e^{-(k/M)(\bar{x}^*/\bar{c})}}{e^{-\bar{c}k/M}} \bar{c}, \quad [3.10]$$

which, for invariant \bar{M} , k/M and \bar{c} (and hence also invariant \bar{x}^*/\bar{c}), is predicted to be an invariant quantity.

Sex Change Invariant Predicted by an Alternative Approach

Charnov and Skúladóttir's (2000) model assumes that timing of sex change is determined by size or age. However, it has been shown in numerous fish species that the timing of sex change can be stimulated by the social environment (Robertson 1972; Shapiro 1981; Warner & Swearer 1991; Allsop & West 2004b). For example, in the cleaner fish *Labroides dimidiatus*, the largest females change sex to become male harem holders upon removal of the male from the social group (Robertson 1972). Here we consider the situation where social environment is assumed to be the primary determinant of when sex change occurs. Specifically, we model a protogynous species in which females change sex to males to maintain a constant sex ratio, following Shapiro and Lubbock (1980).

This means that our model has Charnov and Skúladóttir's (2000) third invariant prediction (that of a constant breeding sex ratio) as its underlying starting assumption. Despite this representing essentially the extreme opposite mechanism underlying sex change to that assumed by Charnov and Skúladóttir (2000), we are also able to predict the first two invariant predictions of Charnov and Skúladóttir (2000), those concerning the relative size and age at sex change.

Consider a protogynous species in which the largest (oldest) males each have harems of F females. The largest female is selected to change sex when the ratio of breeding females to breeding males (N_1/N_2) in the population is greater than F , and so the breeding sex ratio, defined as the ratio of mature females to males in the population will be invariant and equal to F . This breeding sex ratio will be given as

$$\frac{N_1}{N_2} = \frac{\int_0^{\bar{x}^*} e^{-\bar{x}^* M x} dx}{\int_0^{\bar{x}^*} e^{-\bar{x}^* M x} dx} = \frac{e^{-\bar{x}^* M} - 1}{-\bar{x}^* M} \bar{x}^* M = F, \quad [3.11]$$

$\bar{x}^* M$ is known to be invariant within taxa (Charnov & Berrigan 1990, 1991; Gemmill et al. 1999), and so for equation [3.11] to hold $\bar{x}^* M$ must also be invariant. Since k/M is also known to be invariant within taxa (Charnov 1993), the product $\bar{x}^* M \times k/M = k\bar{x}^*$ is an invariant. Applying the Bertalanffy growth equation, the relative size at sex change ($L_{50}/L_{\text{Max}} = 1 - \exp(-k\bar{x}^*)$) is predicted to be invariant, giving the first of Charnov and Skúladóttir's (2000) invariance predictions. Dividing $\bar{x}^* M$ by $\bar{x}^* M$ yields an invariant relative age at sex change \bar{x}^*/\bar{x}^* , which is the second of Charnov and Skúladóttir's (2000) invariance predictions.

Is the Relative Size at Sex Change Invariant Across Species?

Allsop and West (2003a,b) tested for invariant relationships by using the standard methodology of whether a log-log plot gave a slope not significantly different from 1.0 (Harvey & Pagel 1991; Charnov 1993; Brown et al. 2000). In particular, they tested for:

(a) an invariant relative size at sex change by examining the relationship between log mean size at sex change and log maximum size across all sex changing taxa (Allsop & West 2003a, b; figure 3.1); (b) an invariant relative age at sex change by examining the relationship between log mean age at sex change and log mean age at maturity for sex changing fish (Allsop & West 2003b). In these analyses a significant positive relationship between maximum size (age at maturity) and size (age) at sex change is not surprising, and merely reflects that larger species change sex when bigger – the crucial point is determining the extent of variance in the relative size at sex change (Allsop & West 2003a, b).

Buston et al. (2004; Milius 2004) criticised this approach, and instead suggested the use of a null model based upon randomisation techniques. Specifically, they generated data for each species by randomly assigned a maximum body size, assumed the size at maturity to be 50% of maximum body size, and then randomly assigned a size at sex change between 50% and 100% of maximum body size ($L_{Mat}/L_{Max} = 0.5$, $L_{50}/L_{Max} \sim U[0.5, 1]$). This analysis generated data that gave similar slopes to the real data when examining the slope between log size at sex change and log maximum size. Consequently, Buston et al. suggested that Allsop and West's invariant result was in fact non-significant.

However, Buston et al.'s model can be rejected empirically, and also because it is not a true null model. Empirically: (1) Buston et al.'s model cannot produce the observed sex change data, as 5 of the 77 species in the dataset (Allsop & West 2003a) change sex below their lower limit of 50% of the maximum body size (the crustaceans *Acontionostoma marionis*, *Ichthyoxenus fushanenensis*, *Emerita analoga*, and the gastropods *Crepidula adunca* and *C. linulata*). (2) The distribution of relative size at sex change in the actual data is significantly different from the uniform distribution Buston et al. assume (Allsop & West 2004c). (3) Buston et al. arbitrarily assigned the size at maturity a value which forces a good fit between the model and the data (Allsop & West 2004c). Since the size at maturity is set to 50% and size at sex change is uniformly distributed over the range between maturity and maximum size, the model predicts an average relative size at sex change of 75% which is very close to the observed 72%. However, previous work has

suggested that a more accurate average size at maturity is 65% (Charnov 1993), which would give a mean size of sex change of 83%, which is far from the observed data. (4) The assumption of a uniform distribution in relative size at sex change assumes no selection on size at sex change, which has been shown to not be the case in numerous studies over the last 35 years (Warner et al. 1975; Charnov et al. 1978; Charnov 1982a).

Furthermore, and more fundamentally, Buston et al.'s model is not null because it assumes an invariant relative size at maturity, which is intimately linked to an invariant relative size at sex change in the model of Charnov and Skúladóttir (2000). An invariant relative size at maturity follows from two of the three dimensionless invariants required by the invariant sex change predictions, $\square M$ and k/M . If these are invariant then their product $\square k$ is invariant, and so the relative size at maturity ($L_{\text{Mat}}/L_{\text{Max}} = 1 - \exp(-\square k)$) is also an invariant. We show in the third section of this paper that these are the crucial invariants for the Charnov- Skúladóttir model (and that variation in \square is less important), so we would expect Buston et al.'s null model to produce an invariant relative size at sex change, and hence fit the empirical data.

If an invariant relative size at maturity is not assumed, then more appropriate null models can be developed and the predictions of this differ significantly from the observed data. We examine a model in which maturation size is uniformly distributed from size zero to maximum size, and size at sex change uniformly distributed from size at maturity to maximum size (i.e. $L_{\text{Mat}}/L_{\text{Max}} \sim U[0,1]$, $L_{50}/L_{\text{Max}} \sim U[L_{\text{Mat}}/L_{\text{Max}},1]$), and find that this more appropriate null model predicts significantly more variation in L_{50}/L_{Max} than is observed in the dataset, and the r^2 statistic for the observed data is significantly higher than predicted by the null model – see figure 3.2.

Buston et al. (2004) have criticised this null model on the grounds that maturation at size zero is implausible. They suggest that altering the model so that maturation is bounded by 40% and 80% of maximum size is more appropriate, and find that the associated variance in the relative size at sex change is not significantly different from that observed in the dataset at the 5% level. However, this approach is equally arbitrary and *ad hoc*. What is

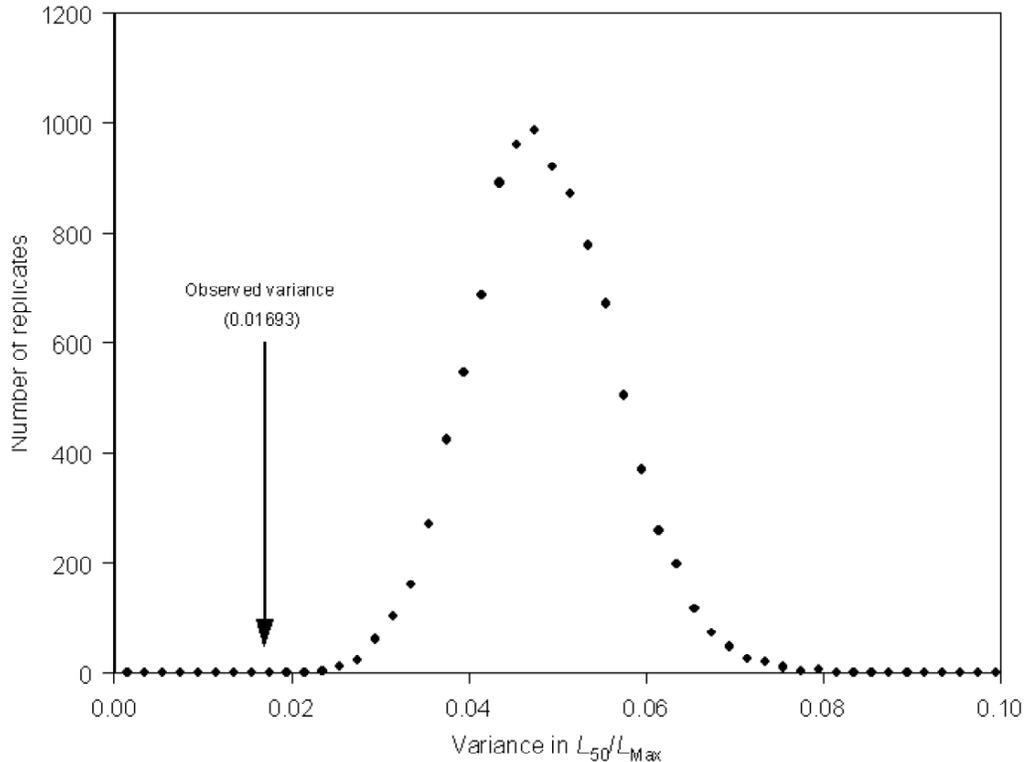


Figure 3.2A. Testing the more appropriate null model: size at maturity (L_{Mat}) is a uniformly distributed random variable bounded by $[0, L_{Max}]$, and size at sex change is a uniformly distributed random variable bounded by $[L_{Mat}, L_{Max}]$, where L_{Max} is asymptotic size. The dots denote the distribution of variance in L_{50} / L_{Max} for 10,000 replicates of a simulated dataset of 77 species of sex changers. The arrow indicates that the variance observed in the real dataset (0.017) is significantly lower (estimated $P < 0.0001$) than predicted by the null model.

the basis for 40-80%, and why not any of infinite different possibilities? How much variation is required in the relative size at maturity before $\square M$ and k/M are not statistically invariant? What would be a suitable minimum size at maturity? These points are particularly important because a true null model should exclude any related factors, and this is not the case here, because theory predicts that the size at sex change should depend upon the size at maturity. Furthermore, invariance is statistical, and so the more appropriate question should be how much variance could there be in the different parameters to explain the data. We explore this approach in the next section.

Before moving on to this next section, figure 3.2B also illustrates an important caveat about r^2 values when testing for invariant relationships. An r^2 value gives the amount of

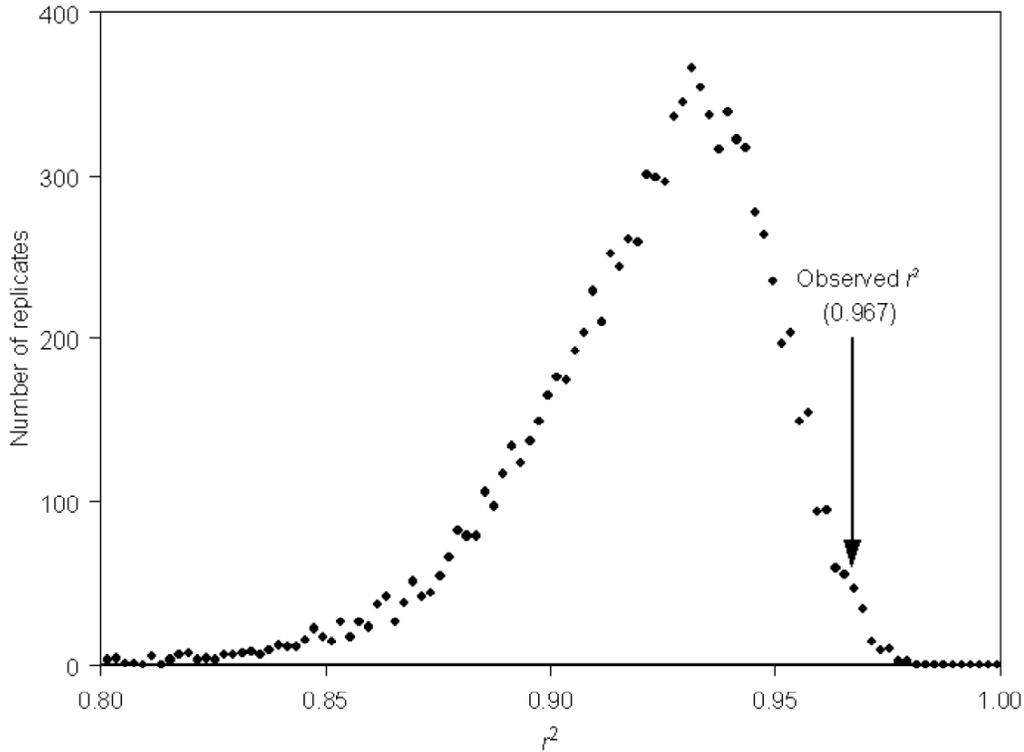


Figure 3.2B. Testing the more appropriate null model: size at maturity (L_{Mat}) is a uniformly distributed random variable bounded by $[0, L_{Max}]$, and size at sex change is a uniformly distributed random variable bounded by $[L_{Mat}, L_{Max}]$, where L_{Max} is asymptotic size. The dots denote the distribution of the r^2 statistic for the best fit invariant relative size at sex change, for 10,000 replicates of a simulated dataset of 77 species of sex changers. The arrow indicates that the r^2 observed in the real dataset (0.967) is significantly higher (estimated $P = 0.0102$) than predicted by the null model.

variance explained when comparing against the null model of no relationship between the two variables. However, as mentioned above, we expect the mean size at maturity and mean size at sex change to be positively correlated, because both will be greater in larger species. Indeed our null model shows that this alone can produce an average r^2 value of 92.1% (figure 3.2B). Assuming an invariant relative size at sex change explains 96.7% of the variation in the actual data (figure 3.1), suggesting that the invariant relationship explains 58.2% of the variation in the data not explained by our null relationship between size at maturity and sex change ($0.582 = (96.7 - 92.1) / (100 - 92.1)$). This value is still very large compared to the average of 2.5-5.4% from evolutionary and ecological studies (Moller & Jennions 2002).

Taxa	Direction of sex change					
	Both directions Intercept (+/- 95% CI)	Both directions RSSC (+/- 95% CI)	Male first Intercept (+/- 95% CI)	Male first RSSC (+/- 95% CI)	Female first Intercept (+/- 95% CI)	Female first RSSC (+/- 95% CI)
Arthropoda (Crustacea)	-0.42 (0.12)	0.66 (0.08)	-0.44 (0.12)	0.64 (0.09)	-	-
Chordata (Fish)	-0.26 (0.04)	0.77 (0.03)	-0.29 (0.11)	0.75 (0.09)	-0.25 (0.04)	0.78 (0.03)
Mollusca	-0.57 (0.16)	0.57 (0.10)	-0.57 (0.16)	0.57 (0.10)	-	-
Echinodermata	-0.32 (0.05)	0.73 (0.03)	-0.32 (0.05)	0.73 (0.03)	-	-
Annelida	-	-	-	-	-	-

Table 3.1. Empirical values for the Relative Size at Sex Change (RSSC = L50/LMax) derived from Log-Log regression of the size at sex change (L50) against the maximum size (LMax) with the slope fixed at proportionality (i.e. 1) The intercept for the regression is also given in the table. Data are split by Taxa and by direction of sex change. Empty cells represent instances where there are too few data points to perform the regression for these categories alone. Data obtained from Allsop & West (2003a,b)

Taxa	k/M	Source	$\square M$	Source
Chordata	0.56	Charnov 1993 (from Beverton & Holt 1959, Beverton 1963)	≈ 2	Charnov & Berrigan 1990 (from Beverton & Holt 1959, Beverton 1963)
Arthropoda	0.39	Charnov 1993 (from Charnov 1979)	≈ 2	Charnov 1979, Charnov 1989
Echinodermata	0.3	Charnov 1993 (from Ebert 1975)	No data	Charnov 1993
Annelida	No data	-	1.45 – 2.5	Gemmill et al 1999

Table 3.2. Summary of estimates for the key life history parameters k/M and $\square M$ for the major taxonomic groups containing sex-changing animals. Note there is no such data for the Mollusca.

Species	First Sex	Size at Maturity (L_{Mat}) in mm	Maximum Size (L_{Max}) in mm	Reference
<i>Acanthopagrus berda</i>	M	125	310	Tobin et al. 1997
<i>Bodianus rufus</i>	F	100	230	Warner & Robertson 1978
<i>Clepticus parrae</i>	F	90	180	Warner & Robertson 1978
<i>Cryptotomus roseus</i>	F	20	70	Robertson & Warner 1978
<i>Epinephelus marginatus</i>	F	438	1050	Marino et al. 2001
<i>Epinephelus morio</i>	F	509	895	Brule et al. 1999
<i>Epinephelus rivulatus</i>	F	194	350	Mackie 2000
<i>Labroides dimidatus</i>	F	15	90	Robertson & Choat 1974
<i>Plectropomus leopardus</i>	F	340	600	Ferreira & Russ 1995
<i>Sarpa salpa</i>	M	185.5	375	Villamil et al. 2002
<i>Lithognathus mormyrus</i>	M	225	370	Lorenzo et al. 2002
<i>Achoerodus viridis</i>	F	245	620	Gillanders 1995
<i>Mycteroperca bonaci</i>	F	826	1500	Crabtree & Bullock 1998
<i>Epinephelus chlorostigma</i>	F	275	500	Grandcourt 2002
<i>Lethrinus mahsena</i>	F	220	360	Grandcourt 2002
<i>Scarus ghobban</i>	F	150	370	Grandcourt 2002

Table 3.3 Empirical values for the size at maturity (L_{Mat}) and the maximum size (L_{Max}) for 17 species of sex changing fish. Also shown is the direction of sex change, with the first sex either male (M) or female (F)

Sensitivity Analysis: Consequences of Variation in $\square M$, k/M and \square

Allsop and West (2004a) observed that the relative size at sex change is statistically invariant ($L_{50}/L_{\text{Max}} \approx 0.72$) over sex changing animals ranging in size from 2mm to 1.5m. This result is predicted by Charnov and Skúladóttir's (2000) model if $\square M$, k/M and \square are also invariant. It has been shown that $\square M$ is likely to be invariant across sex changing species (Charnov 1993, Gemmill et al. 1999), and so Allsop and West's result suggests that k/M and \square are also invariant, or have relatively little influence on the ESS relative size at sex change. Determining the answer to this requires a sensitivity (elasticity) analysis of Charnov and Skúladóttir's (2000) model, and the examination of how variation in these different life history variables influences the evolutionarily stable size at sex change and the degree of invariance in the relative size at sex change.

We will: (1) use published values for $\square M$ and k/M to obtain an estimate of \square from the Charnov- Skúladóttir model, given the observed data for sex changing fish (Allsop and West 2003b); (2) estimate $\square M$ and k/M directly from the sex changing fish data; and (3) introduce variation into each of the dimensionless quantities $\square M$, k/M and \square in turn, to see how much variation in each corresponds to the observed variation in L_{50}/L_{Max} . We restrict our attention to fish only, since there is much more data on the relevant life history variables, than for other sex changers.

Variation in \square

Assuming $\square M \approx 2$ (Charnov and Berrigan 1990, 1991; Charnov 1993) and $k/M \approx 0.6$ (Charnov 1993), the optimal relative age (and hence size) at sex change can be determined numerically from equation [3.9]. We do this for a range of \square in figure 3.3. These results predict that: (1) the relative size at sex change (L_{50}/L_{Max}) is positively correlated with the male fitness exponent (\square), and (2) this positive correlation is very

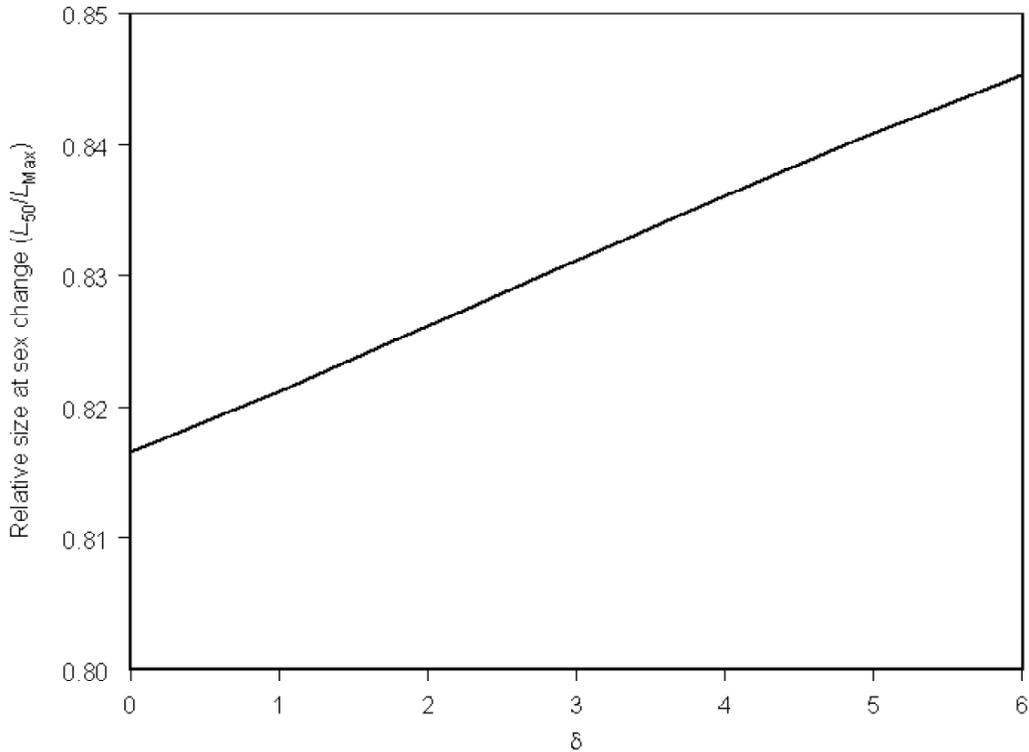


Figure 3.3 The ESS relative size at sex change (L_{50}/L_{Max}) predicted by the Charnov-Skuladottir model assuming the published estimates of $\bar{\rho}/M = 2$ and $k/M = 0.6$ and a range of ρ . When male fecundity increases with size ($\rho > 0$) the model and published estimates predict a relative size at sex change which is higher than that observed for sex changing fish ($L_{50}/L_{Max} = 0.77$).

weak. The weakness of the correlation between the male fitness exponent (ρ) and the timing of sex change suggests that this life history parameter need not be particularly invariant in order for the relative size at sex change to show great invariance.

The positive correlation between and the relative timing of sex change can be explained as follows. For protandrous species ($\rho < 3$), an increase in ρ means that the relative success of the small males is reduced, and so the individual is selected to increase their reproduction as a male in order to make up this quota of their total reproduction, hence longer time is spent as the first sex. Conversely, for protogynous species ($\rho > 3$), an increase in ρ means that the relative fitness of the large male is increased, so that less time need be spent reproducing as a male, and hence the individual spends longer

reproducing as the first sex. In both instances, an increase in β is associated with delayed sex change.

With regards to this prediction that the relative size at sex change should increase weakly with β , the data do reveal a slight tendency for protogynous fish ($\beta > 3$, $E[L_{50}/L_{Max}] = 0.79$) to have a higher relative size at sex change than protandrous fish ($\beta < 3$, $E[L_{50}/L_{Max}] = 0.74$). However, as might be expected given the weak predicted effect, this difference is not significant ($P = 0.21$).

The results given in figure 3.3 also predict that the relative size at sex change is too high for it to explain the empirical observation $L_{50}/L_{Max} = 0.77$ (data described by Allsop & West 2003b; note that this is the best-fit invariant for the data, rather than the expectation of the data – the latter was employed in the cited paper, leading to a slightly different value, 0.79) for sex changing fish. This suggests that either the model is incorrect or that the values of βM and k/M published for fish in general do not correspond to those in our dataset.

We investigate this possibility using the data sets compiled by Allsop & West (2003a,b) and find support for the suggestion that sex changing fish may have βM and/or k/M values different from those published for other species. The product βk determines the relative size at maturity ($L_{Mat}/L_{Max} = 1 - \exp(-\beta k)$), so we may use size at maturity data (compiled in table 3.3) to estimate βk . Assuming an invariant relative size at maturity (i.e. a slope of unity on a plot of $\text{Log}[L_{Mat}]$ against $\text{Log}[L_{Max}]$, figure 3.4) we find that the least squares regression gives $L_{Mat}/L_{Max} \approx 0.46$, and hence $\beta k \approx 0.62$. This value is approximately half of that published for other fish ($\beta M \beta k/M = \beta k \approx 1.2$), indicating that sex changing fish mature at a relatively small age and size than other fish.

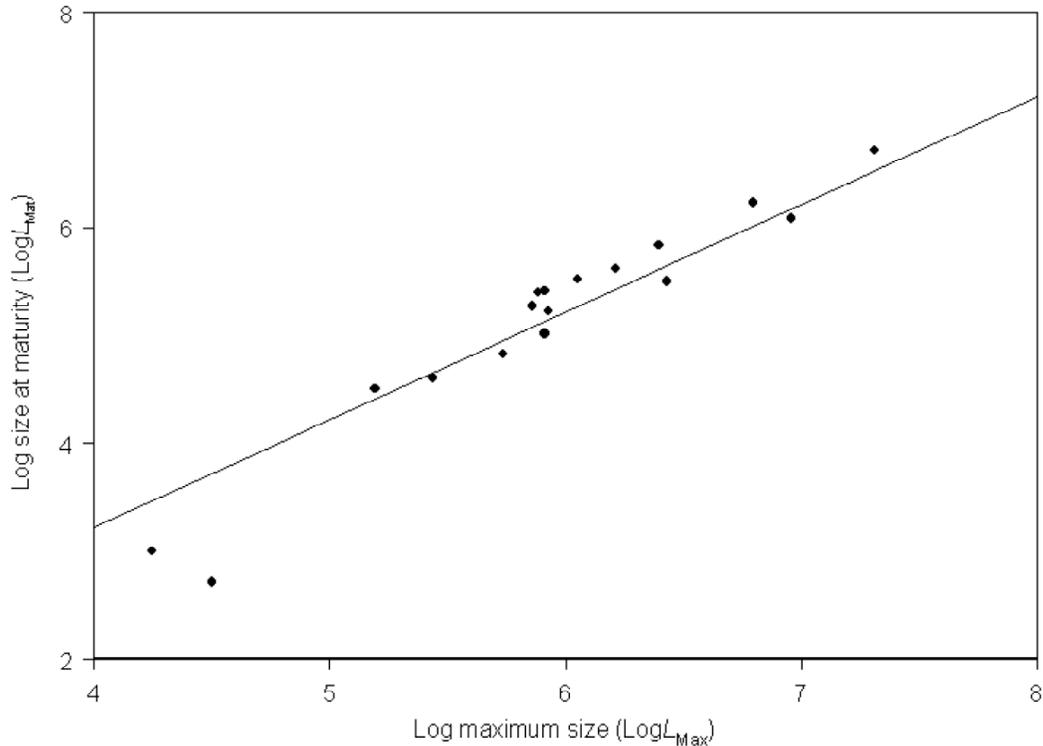


Figure 3.4 Log-Log plot of L_{Mat}/L_{Max} for 17 species of sex changing fish. Fixing the regression slope at 1, it has an intercept of $-0.78 \pm (0.16 \text{ 95\% CI})$ ($r^2=0.90$, $n=17$). The relative size at maturity invariant (L_{Mat}/L_{Max}) is 0.46, showing that sex changing fish mature at approximately 46% of their maximum body size, on average. Size (L_{Mat} and L_{Max}) is measured in mm prior to logarithmic transformation.

Day and Taylor (1997) have warned against the use of von Bertalanffy's equation in relation to size at maturity. Since immature organisms do not reproduce and hence can allocate more of their energy budget into growth, the growth rate coefficient (k) may be somewhat higher pre-maturation than post-maturation. However, the associated bias in the estimation of the post-maturation growth rate from age at maturity data is in the wrong direction to explain the discrepancy between the observed $\square k$ and that published for fish in general. In the next section we show that this reduced estimate of $\square k$ is more consistent with the observed timing of sex change.

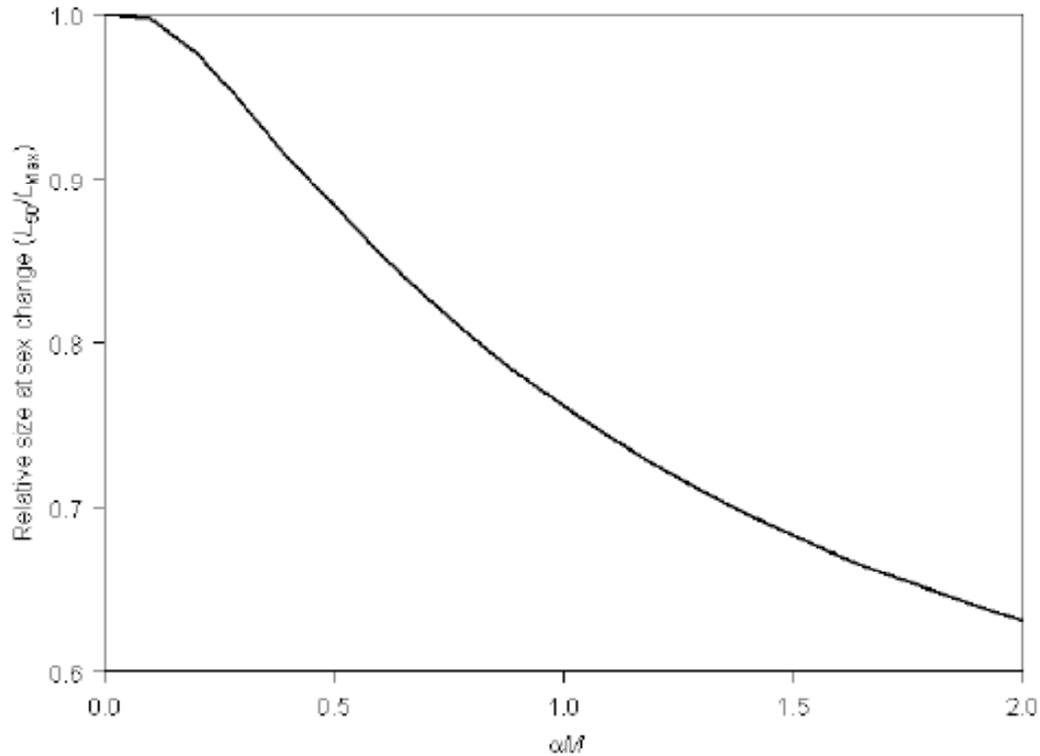


Figure 3.5 The ESS relative size at sex change (L_{50}/L_{Max}) predicted by the Charnov-Skuladottir model assuming $\alpha \approx 3$ and $\alpha M \approx k/M = \alpha k = 0.616$, as estimated from the size at maturity data. By plotting for a range of αM to see where the model predicts the observed size at sex change invariant ($L_{50}/L_{Max} = 0.77$), we can obtain an estimate of αM , and hence k/M .

Estimating αM and k/M

We now estimate values for αM and k/M for sex changing fish. As outlined before, we can numerically solve expression [3.9] to give a relative age (and hence size) at sex change given values for α , αM and k/M . By exploring a range of these three parameters, we can determine which triplets give the observed $L_{50}/L_{Max} = 0.77$. As we have seen, the male fitness exponent (α) impacts very little on the relative size at sex change – we can essentially ignore this parameter, and restrict our attention to the two parameters αM and k/M . Recalling that some species in the dataset will have $\alpha < 3$ while others have $\alpha > 3$ (since there is a mixture of protandry and protogyny), we will proceed by assuming $\alpha \approx 3$ for the purposes of estimating the other two life history parameters. We estimated that the

invariant product of $\bar{\phi}M$ and k/M is approximately 0.62, and so the parameter set is effectively reduced to a single dimension: e.g. given $\bar{\phi}M$, k/M will be given by $0.62/\bar{\phi}M$.

In figure 3.5 we determine the impact on the relative size at sex change of variation in $\bar{\phi}M$, by allowing it to take a range of values while satisfying the estimate of $\bar{\phi}k = 0.62$. We find that the model predicts the observed invariant relative size at sex change ($L_{50}/L_{\text{Max}} = 0.77$) when $\bar{\phi}M \approx 0.96$. From this we also estimate k/M to be $\bar{\phi}k/\bar{\phi}M \approx 0.64$.

Assessing variation in $\bar{\phi}$, $\bar{\phi}M$ and k/M

We have obtained estimates for the average values of $\bar{\phi}$, $\bar{\phi}M$ and k/M in sex changing fish, but it is unclear what degree of invariance in each of these is required to give the result that 97.5% of the variation in relative size at sex change is explained by the simple rule that they change sex at 77% of their maximum size. We may make a qualitative assessment of how variation in these life-history parameters translates into variation in the relative size at sex change by varying the value of each parameter in turn while holding the other two constant (at their estimated values; $\bar{\phi} = 3$, $\bar{\phi}M = 0.96$, $k/M = 0.64$). Our results are given in figure 3.6). As before we find that the value of the male fitness exponent ($\bar{\phi}$) has little impact on the relative size at sex change (figure 3.6A), while variation in $\bar{\phi}M$ (figure 3.6B) and especially k/M (figure 3.6C) have a more dramatic effect.

We find that the value of $\bar{\phi}M$ correlates positively with the relative size at sex change (figure 3.6B), when the values of k/M and $\bar{\phi}$ are held constant. One way to visualize this is to hold k , M and $\bar{\phi}$ constant, and to vary the age at maturity, $\bar{\phi}$. It makes sense for species that mature later to change sex later, in order to make up their quota of reproduction as the first sex. This means that species with a higher $\bar{\phi}M$ will, all else being equal, change sex at a greater size.

We have also found that increasing k/M increases the relative size at sex change (figure 3.6C), when $\bar{\phi}M$ and $\bar{\phi}$ are held constant. To see why, hold $\bar{\phi}$, M and $\bar{\phi}$ constant, and

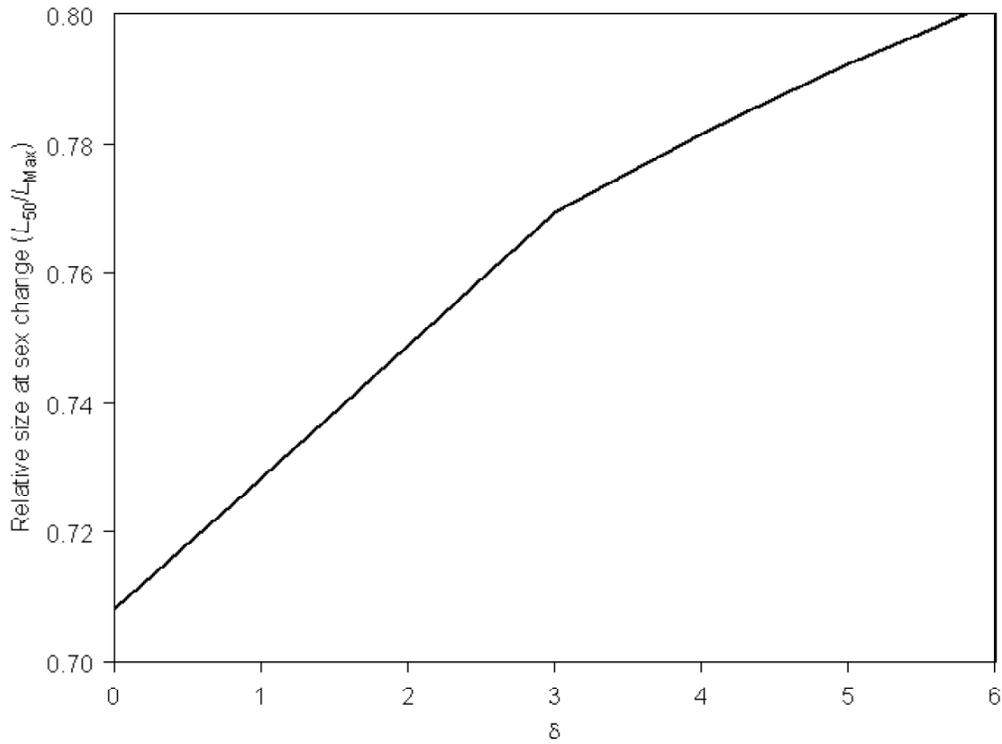
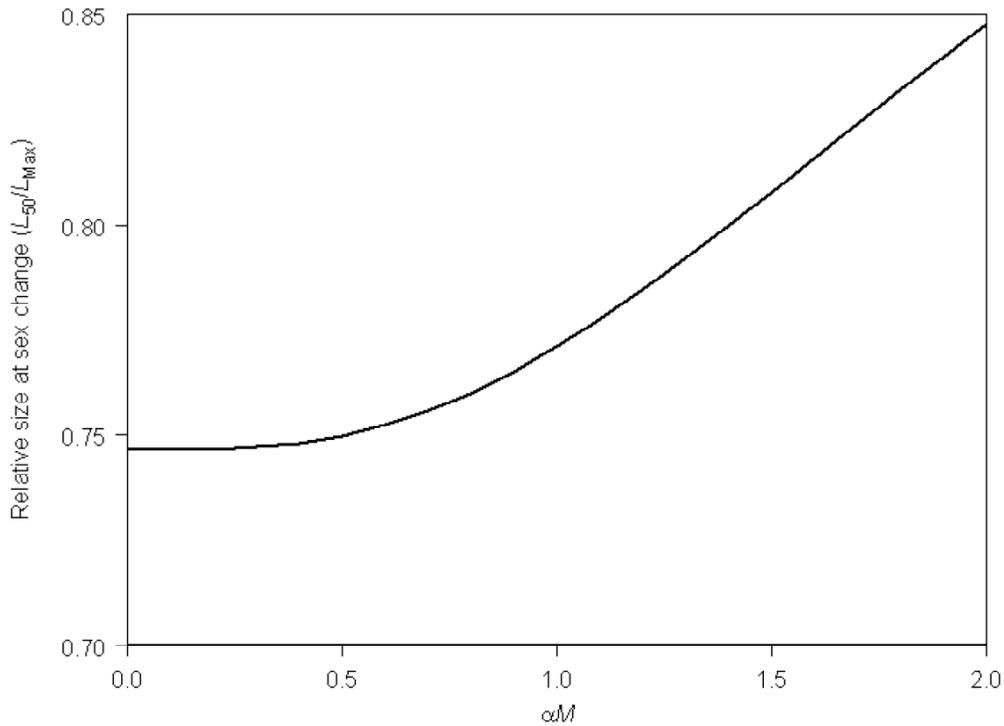


Figure 6. Having estimated the three dimensionless quantities ($\alpha M \approx 0.96$, $k/M \approx 0.64$, $\beta \approx 3$), we explore how variation in each of these translates into variation in the ESS relative size of sex change (L_{50}/L_{Max}). A (above) αM and k/M held fixed, β varied. B (below) k/M and β held fixed, αM varied. (cont. over . . .)



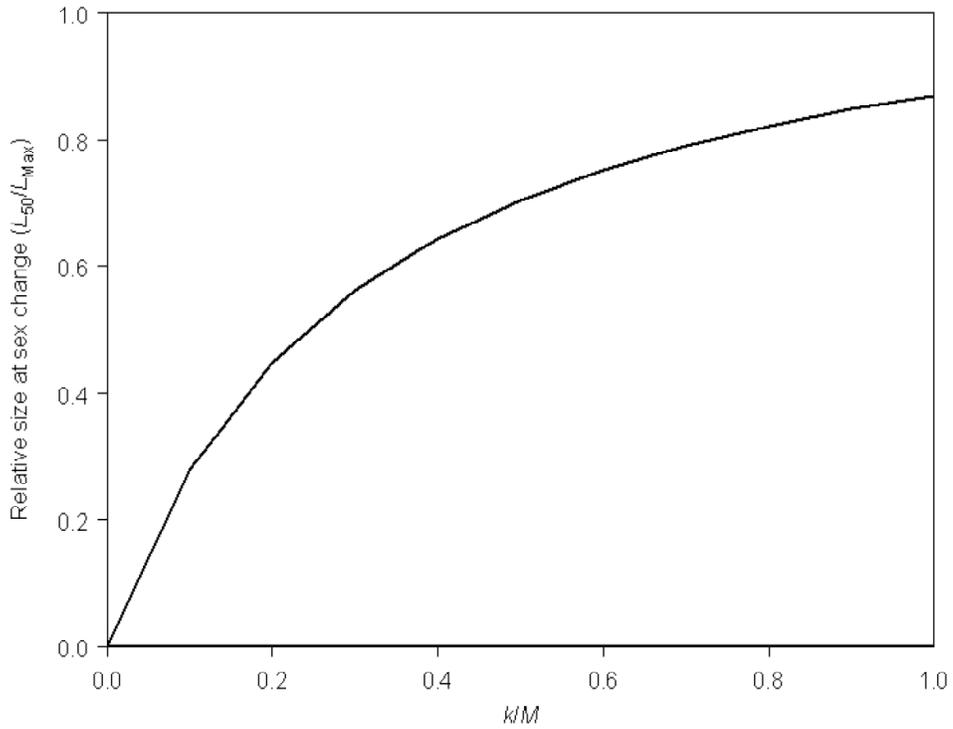
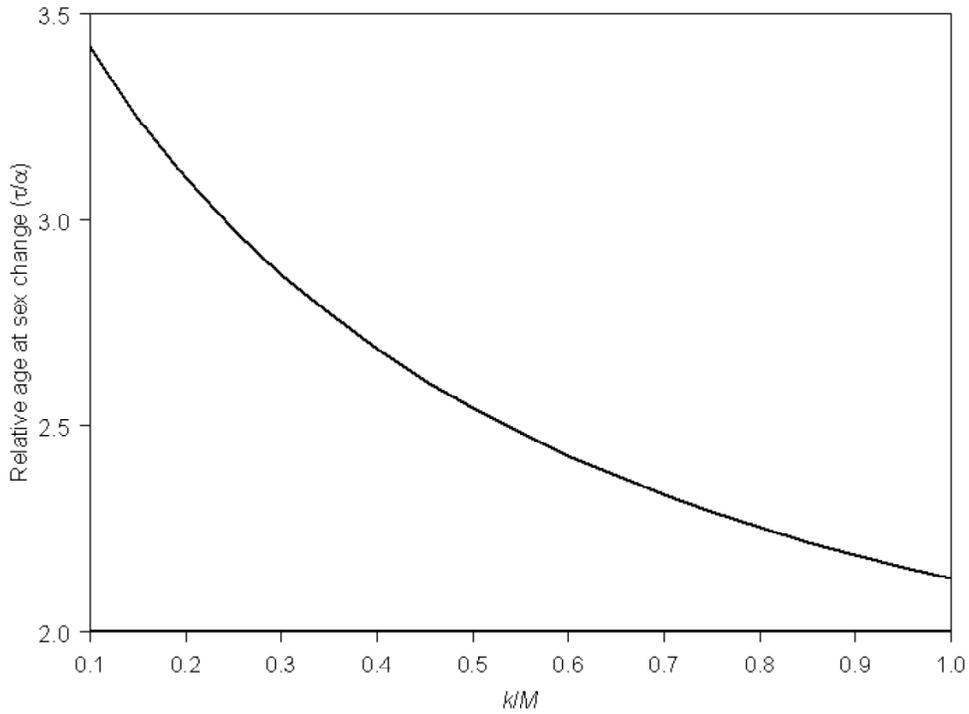


Figure 6 (cont.). Having estimated the three dimensionless quantities ($\beta M \approx 0.96$, $k/M \approx 0.64$, $\beta \approx 3$), we explore how variation in each of these translates into variation in the ESS relative size of sex change (L_{50}/L_{Max}). C (above) βM and β held fixed, k/M varied. D (below) the relationship between k/M and the ESS relative age at sex change (τ^*/α) is investigated for βM and β held constant.



allow k to vary. As the Bertalanffy growth coefficient is increased, the size at all ages is increased, and so if we assume no impact on the ESS relative age at sex change (\bar{a}/\bar{L}) we would expect an increased relative size at sex change (L_{50}/L_{Max}). In fact, the ESS relative age at sex change is a decreasing function of k (figure 3.6D). This is because the increase in size due to increased k is more pronounced at earlier ages, hence the reproduction of the first sex is improved the most by this increase, so that less time need be spent reproducing as that sex. This means that in increasing k the ESS age at sex change is reduced, but the size at that age is increased. The net effect is a positive correlation between k/M and the relative size at sex change.

A more quantitative approach is to use the model to simulate sex change data for a range of variation in the underlying dimensionless quantities, to see how much variation corresponds to that observed in the real dataset. We simulate 52 species of sex changing fish, each assigned values for \bar{M} , k/M and \bar{L} . Within each dataset, two of these dimensionless quantities are held fixed at their estimates from the previous section, while the other takes a pseudorandom value independently drawn for each species from the normal distribution with mean given by the estimated value, and standard deviation σ . Because it is biologically implausible for \bar{M} and k/M to take negative values, we draw these from a normal distribution truncated at the origin which, for the parameters we will explore, involves removing a trivial proportion of the probability distribution. Introducing the actual data on asymptotic size (L_{50}), the Charnov-Skúladóttir model of sex change is then used to generate the ESS relative size (L_{50}/L_{Max}) at sex change for each of the simulated fish species. For each dataset, an r^2 statistic can be generated to describe how well the simulated data conforms to the prediction of a slope of unity in a plot of $\text{Log}[L_{50}]$ against $\text{Log}[L_{Max}]$. This procedure is used to explore a range of variation (standard deviation σ) in each of the invariant quantities \bar{M} , k/M and \bar{L} .

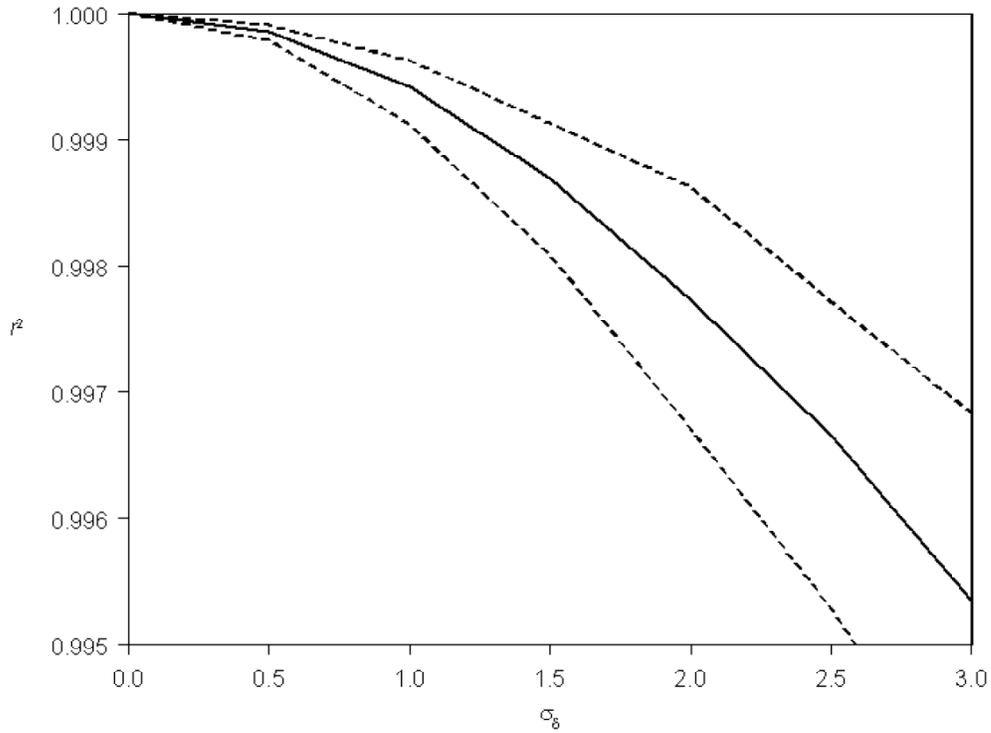
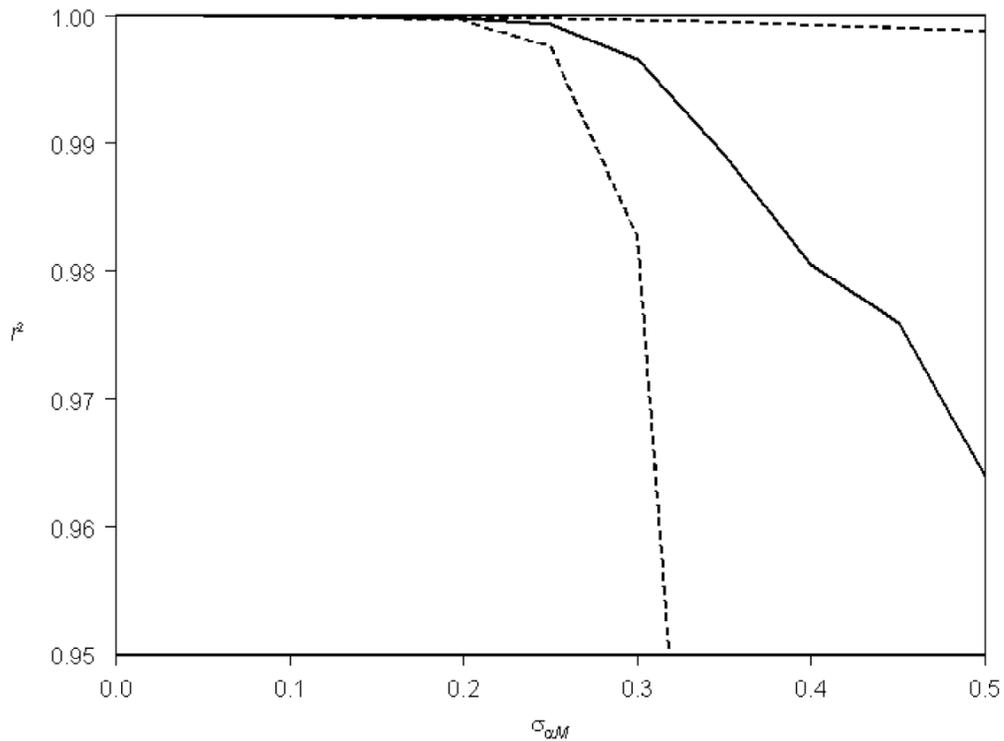


Figure 3.7. Sensitivity analysis. Two of the three dimensionless parameters ($\alpha M \approx 0.96$, $k/M \approx 0.64$ & $\beta \approx 3$) are held fixed at their estimates, while the third is drawn from a normal distribution with mean given by above estimate and standard deviation σ , to generate size at sex change data for 52 spp of fish. The r^2 statistic (cont. over . . .)



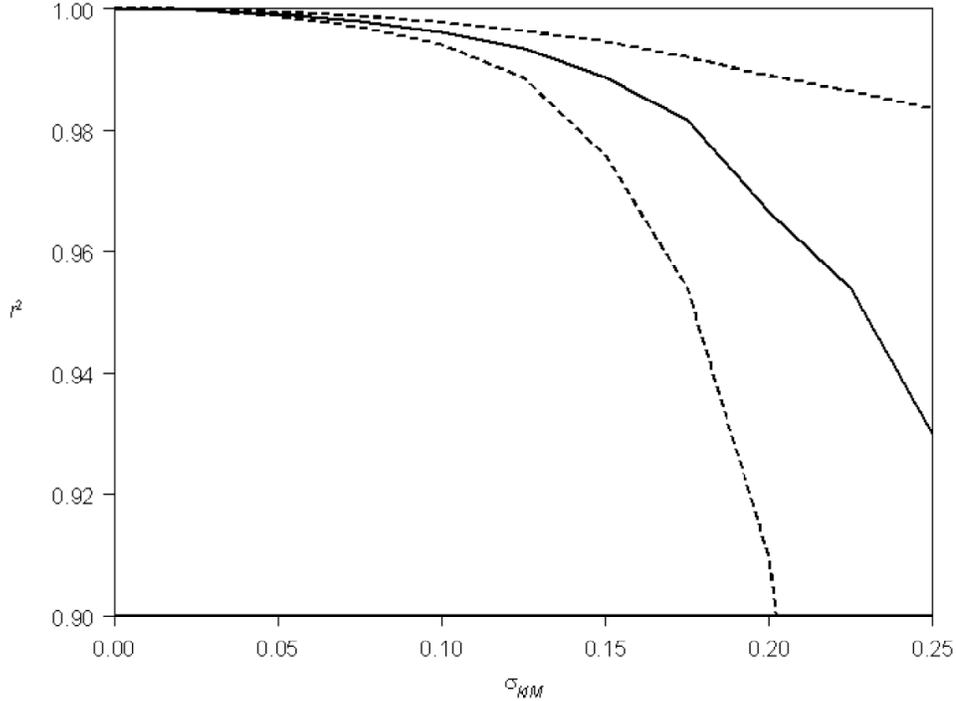


Figure 3.7 (cont) . . . for best-fit invariant (i.e. slope 1 in plot of $\text{Log}[L_{\text{Mat}}]$ against $\text{Log}[L_{\text{Max}}]$) is determined for each σ and each of the dimensionless parameters in turn (each estimate based upon 200 replicates). Solid line is mean r^2 , and dashed line delineates region in which 95% of replicates fell. A (previous page, top) variation in σ . B (previous page, bottom) variation in σM . C (this page, above) variation in k/M .

From figure 3.7 we can read off the estimate of the standard deviation for each of the life-history parameters by seeing what value of σ corresponds to the observed $r^2 = 0.973$. Figure 3.7A confirms that the invariant relative size at sex change is expected to hold even with extensive variance in the male fitness exponent (σ) when the other life-history parameters do not vary. Figure 3.7B reveals that variation in σM corresponding to a standard deviation of around 0.45 (47% of the estimated mean, $E[\sigma M]=0.96$) can account for the observed variation in the relative size at sex change – which is rather a lot of variation. Figure 3.7C reveals that a standard deviation in k/M of around 0.18 (28% of the estimated mean, $E[k/M]=0.64$) can account for the observed variation in L_{50}/L_{Max} . It should be noted that these results are upper limits on the amount of variation, as they assume only one parameter is variable, whereas in reality, there will be some variation in each.

Discussion

We have formally derived the life history invariants predicted by Charnov and Skúladóttir (2000), who modeled sex change conditional on an individual's size (and hence age). These are invariant: (1) relative age at sex change (τ/τ_0); (2) relative size at sex change (L_{50}/L_{Max}); and (3) breeding sex ratio (N_1/N_2). Previously, Buckingham's (1914) π theorem had been invoked in order to show that, in principle, the appropriate fitness function could be expressed in terms of dimensionless quantities. We have noted that the units with which we measure time do not influence the dimensionless ESS relative timing of sex change, and thus employed a simple 'switching variables' technique to explicitly state the appropriate dimensionless fitness function. Additionally, we have shown that these invariants can be predicted with a different approach, when sex change is assumed to occur in response to social cues.

Allsop & West (2003a,b) showed invariance in the relative size at sex change across all sex changing organisms for which there is data, and the relative age at sex change in fish. These results were criticised by Buston et al. (2004), who argued that randomization tests should be used instead of standard methodology (Harvey and Pagel 1991; Charnov 1993; Brown et al. 2000). We have argued that: (a) their randomization test was not truly null; (b) the data do not fit their model, and (c) more appropriate tests support the invariant conclusions of Allsop and West (2003a,b). Furthermore, we suggest that a more powerful approach and approach is avoid randomization tests based upon possibly arbitrary assumptions, and instead examine how much variance in the different parameters would explain the observed data.

We then carried out a numerical sensitivity analysis in order to determine the relative consequences of variation in the dimensionless parameters that can influence the relative size at sex change. These results showed that the invariant prediction depends primarily upon invariance in τ/M and k/M , and that variation in τ has little consequence for the ESS size at sex change. This result illustrates clearly one of the major problems with Buston et

al.'s 'null' model – it was not null because it effectively assumed an invariant $\square M$ and k/M , and only allowed \square to vary, so we would expect it to predict the observed data.

How much variation in $\square M$ and k/M are consistent with the observed data on relative size at sex change? We estimated the variation in each of these parameters that is consistent with the observed variation in the timing of sex change in the fish data set. We found that: $\square M \approx 0.96$ (with standard deviation ± 0.45) and $k/M \approx 0.64$ (± 0.18). This suggests that there can be a relatively large amount of variation in $\square M$, but less in k/M . These results are upper limits on the amount of variation, as they assume only one parameter is variable, whereas in reality, there will be some variation in each. More generally, Allsop and West (2003a) argued that their invariant result suggested a fundamental similarity across all animals in the underlying forces that select for sex change. Our results suggest that the fundamental similarities are: (a) the basic assumptions of the Charnov & Skulladotir (2000) model, and (b) the value of k/M , and to a lesser extent the other dimensionless variables.

Our results also lead to the prediction that the value of $\square M$ differs in sex changing fish from other fish species. Specifically, the published values for fish in general give $\square M \approx 2$ and $k/M \approx 0.6$. In contrast, we predict that $\square M \approx 1$ for sex changing species. We have verified this prediction by estimating the product of these two putative invariants from the relative size at maturity data in sex changing fish, confirming that $\square M \square k/M = \square k \approx 0.6$, around half of the value expected for fish in general. More investigation, both theoretical and empirical, is needed to explain this difference between sex changers and other fish.

We conclude with two general points. First, the debate over the usefulness of applying the life history invariant approach to sex change cuts to the heart of the philosophy of statistics in the biological sciences. As we are dealing with biology it is clear that there are really no true invariants in the physical sense. However, there are a number of statistically invariant relationships that hold across taxa for reasons that are not immediately apparent, and these require explanation. Second, we have demonstrated that

invariant theory can be used to estimate the values of, and variation in, important biological parameters. This is especially useful when it allows us to get at parameters which would be difficult or laborious to measure directly. Another much studied example of this from evolutionary theory more generally, is using sex ratios and sex ratio theory to estimate inbreeding rates in protozoan parasites (Read et al. 1992; West et al. 2001; Nee et al. 2002).

4. Spite and the scale of competition[†]

Abstract

In recent years there has been a large body of theoretical work examining how local competition can reduce and even remove selection for altruism between relatives. However, it is less well appreciated that local competition favours selection for spite, the relatively neglected ugly sister of altruism. Here, we use extensions of social evolution theory that were formulated to deal with the consequences for altruism of competition between social partners, to illustrate several points on the evolution of spite. Specifically, we show that: (1) the conditions for the evolution of spite are less restrictive than previously assumed; (2) previous models which have demonstrated selection for spite often implicitly assumed local competition; (3) the scale of competition must be allowed for when distinguishing different forms of spite (Hamiltonian versus Wilsonian); (4) local competition can enhance the spread of spiteful greenbeards; (5) the theory makes testable predictions for how the extent of spite should vary dependent upon population structure and average relatedness.

Altruism and spite

Social behaviours can be categorized according to the direct fitness consequences they entail for the actor and recipient (figure 1.1; Hamilton 1964, 1970, 1971). A behaviour increasing the direct fitness of the actor is mutualistic if the recipient also benefits, and selfish if the recipient suffers a loss. It is easy to see how such behaviours can be favoured by natural selection. Behaviours which reduce the direct fitness of the actor – altruism if the recipient enjoys a benefit, spite if the recipient suffers a loss – are less easy to explain. Hamilton (1963, 1964) introduced the concept of inclusive fitness and showed

[†] Published as: Gardner A. & West S.A. (in press). Spite and the scale of competition. *Journal of Evolutionary Biology* (see Appendix).

that while certain behaviours are detrimental to the individual, they may result in a net increase in the actor's genes in the population. Altruism can be favoured by natural selection despite a direct fitness cost (C) to the actor if the benefit (B) accruing to the recipient is sufficiently large and if the genetic relatedness (R) of the recipient to the actor is sufficiently positive. Specifically, when Hamilton's (1963, 1964) rule, $R B > C$, is satisfied. A spiteful behaviour, entailing a negative benefit ($B < 0$) to the recipient and a positive cost ($C > 0$) to the actor, is similarly favoured if $R B > C$, which would require a negative relatedness ($R < 0$) between actor and recipient.

Relatedness and spite

Hamilton (1963) argued that under the assumption of weak selection the appropriate measure of relatedness (R) coincides with Wright's (1922) coefficient of relationship. Wright's coefficient is a function of the correlation between individuals and the correlation within individuals with respect to their genes at a given locus. Since these correlations have popularly been interpreted in terms of Malécot's (1948) probability of identity by descent, and negative probabilities are not permitted, negative relatedness seems to be mathematically impossible (Hamilton 1970, 1996; although see Wright 1969, p178). Yet Hamilton (1963) understood that relatedness (R) was in principle a regression coefficient – a fact which is now generally appreciated (reviewed by Seger 1981, Michod 1982, Grafen 1985a, Queller 1985, 1992, Frank 1998) – and this was first made explicit in his elegant reformulation of Hamilton's rule (Hamilton 1970) using Price's (1970) equation. Specifically, relatedness is the regression (slope) of the recipient's genetical breeding value on that of the actor (Hamilton 1970, 1972; Taylor & Frank 1996, Frank 1997a, 1998). Since regressions can be negative as well as positive (and zero), relatedness can feasibly take any real value (from negative infinity to positive infinity). Discussions with Price led Hamilton to acknowledge that negative relatedness can plausibly arise between social partners, and hence spite can be favoured by natural selection (Hamilton 1970, 1996, Frank 1995).

How does negative relatedness arise? Grafen's (1985a) geometric view of relatedness reveals that relatedness between an actor and a potential recipient depends crucially upon the genetical composition of the whole population. This can be illustrated by assuming that a recipient carries the actor's genes with average frequency p , and the population frequency of the actor's genes is \bar{p} . If the recipient carries the actor's genes at a frequency greater than the population frequency of those genes ($p > \bar{p}$) then an increase in its reproductive success translates into increased frequency of the actor's genes in the population, and hence a positive inclusive fitness benefit to the actor ($R B > 0$; figure 4.2A). Conversely, if the recipient carries the actor's genes at a frequency lower than the population frequency of those genes ($p < \bar{p}$) then an increase in its reproductive success translates into decreased frequency of the actor's genes in the population, and hence a negative inclusive fitness benefit for the actor ($R B < 0$; figure 4.2B). The point here is that the difference between these two situations can arise purely due to variation in the frequency of the actor's genes in the population (variable \bar{p}), even with a fixed proportion of genes shared between the actor and recipient (fixed p): relatedness is relative, with the population as a whole providing the reference.

This also illustrates how negative relatedness can arise. Since both situations described above involve a positive benefit ($B > 0$) to the recipient, the coefficient of relatedness which transforms recipient success into inclusive fitness of the actor must be positive in the former instance ($R > 0$; figure 4.2A) and negative in the latter ($R < 0$; figure 4.2B). The other possibility is that relatedness is zero when the recipient carries the same frequency of the actor's gene as does the population as a whole ($p = \bar{p}$), so that relatedness to the average population member (and hence to the population itself) is zero (figure 4.2C).

But, how large a negative relatedness is likely to arise? Consider an individual who lives in a population of size N , and who is then related to a fraction $1/N$ of the population (i.e. itself) by an amount 1 and is related to the other fraction $(N - 1)/N$ by an amount R . The relatedness to the population as a whole must be zero (Grafen 1985a), and hence must

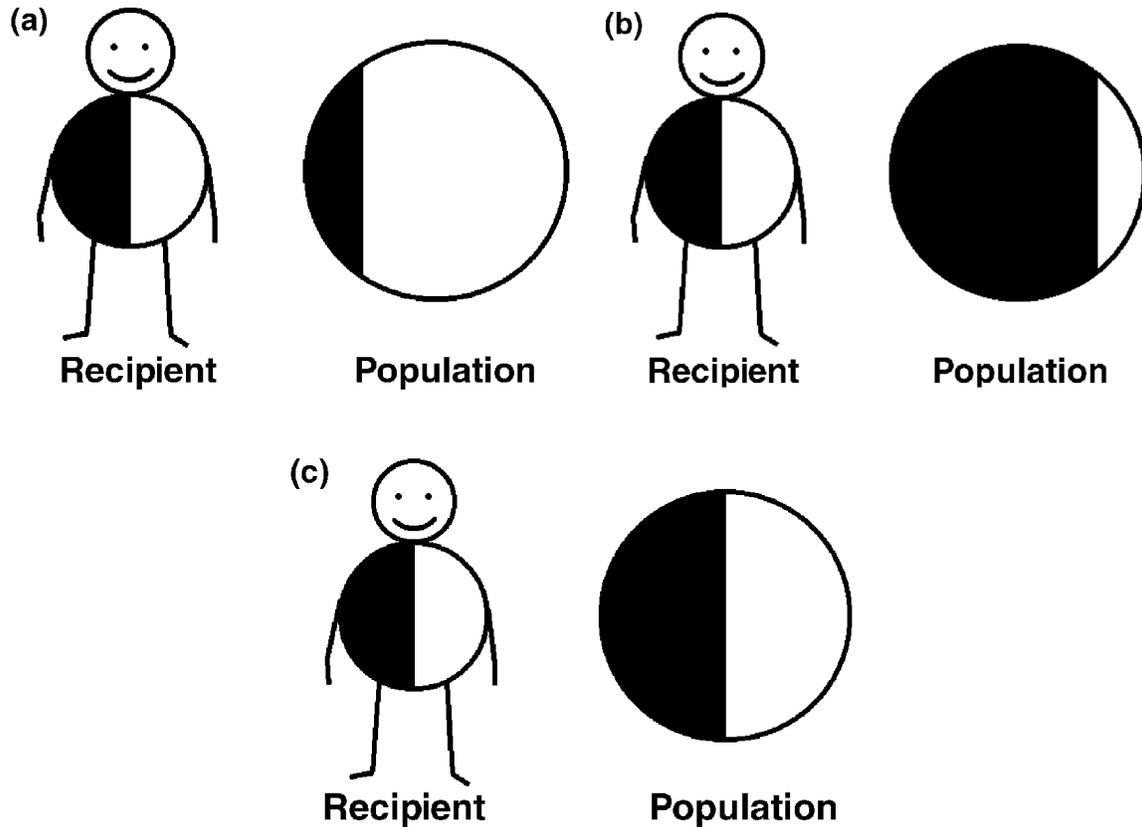


Figure 4.2. The geometric view of relatedness. The actor's genes (shaded) are present in the recipient at frequency p and in the population as a whole at frequency \bar{p} . Enhancing the direct fitness of the recipient ($B > 0$) pushes the population gene frequency towards p , and so if $p > \bar{p}$ (as in 4.2A) the frequency of the actor's genes increase, giving a positive inclusive fitness benefit ($R B > 0$) which implies positive relatedness ($R > 0$) between actor and recipient. If $p < \bar{p}$ (4.2B) then the population frequency of the actor's genes decreases, giving a negative inclusive fitness benefit ($R B < 0$) and hence negative relatedness ($R < 0$). When $p = \bar{p}$ (4.2C) the population frequency does not change, giving no inclusive fitness benefit ($R B = 0$) and hence zero relatedness ($R = 0$).

satisfy $(1/N) + ((N - 1)/N)R = 0$. Rearrangement gives $R = -1/(N - 1)$, i.e. the average relatedness between the actor and its social partners is negative (Hamilton 1975, Grafen 1985a, Pepper 2000). If the focal individual can identify, and refrain from being spiteful to, a number of positively related genealogically-close social partners (kin discrimination), then the relatedness to recipients becomes even more negative (Hamilton 1975). For very small populations (small N ; figure 4.3), negative relatedness can be nontrivial, and hence individuals might be expected to pay reasonable costs in order to inflict damage to their social partners. Negative relatedness (and hence spite) is therefore

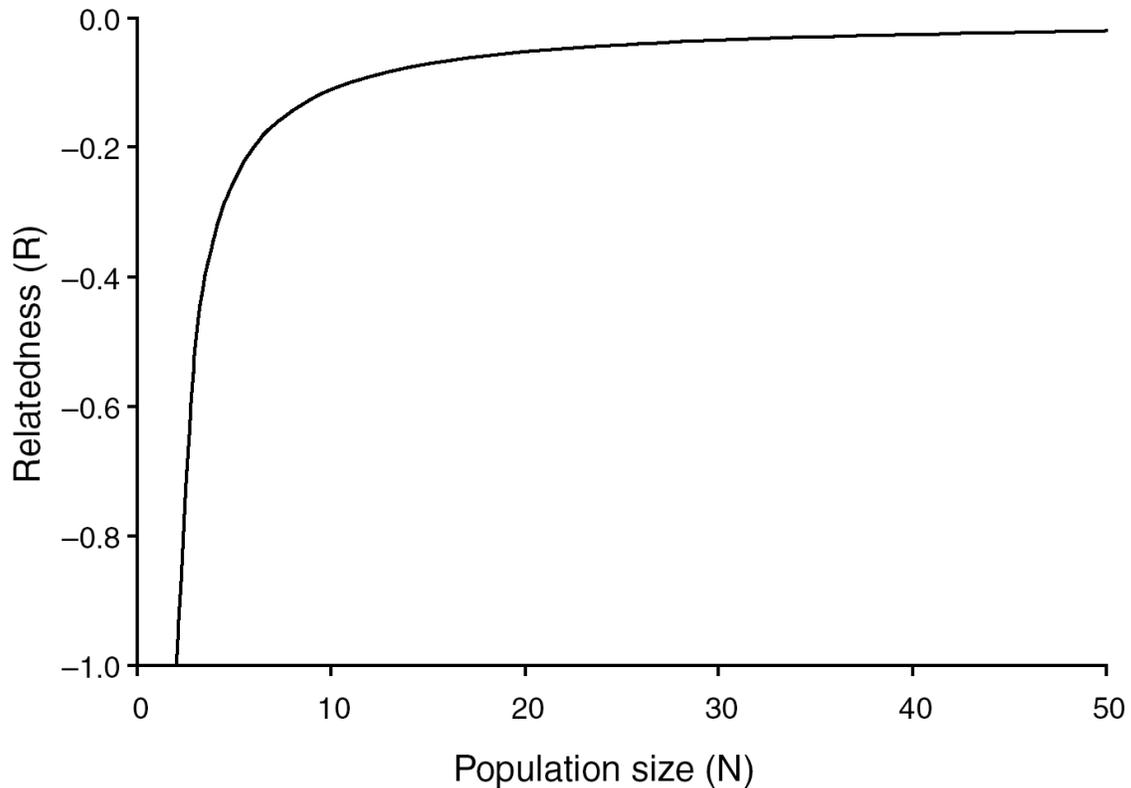


Figure 4.3. The average relatedness (R) between population members as a function of population size (N), when there is no kin discrimination. Since relatedness by any member to the population as a whole is zero, and this includes positive relatedness to itself, relatedness to the other individuals is necessarily negative, specifically $R = -1/(N-1)$. This is minimised at $R = -1$ when $N = 2$, but quickly tends to zero as N increases towards more plausible values.

possible, but this tiny-population condition caused Hamilton (1971) to regard spite as merely the “final infection that kills failing twigs of the evolutionary tree”, and not a general phenomenon contributing to adaptive evolution (Hamilton 1996).

Scale of competition

However, the situation may not be so bleak for spite. There has recently been much interest in how local competition between relatives can reduce and even remove selection for altruism between relatives (reviewed by Queller 1992, West et al. 2002a). This work was spurred by the possibility that with limited dispersal in a viscous population,

individuals would tend to associate with kin, so that kin selection theory might suggest positive relatedness between social partners, and hence conditions favourable for the evolution of altruism (Hamilton 1964, 1971, 1972, 1975, 1996). However, this relies on the implicit assumption of density-dependent regulation being global (hard selection; Wallace 1968), with no increased competition, due to increased productivity, within more

altruistic groups (Boyd 1982, Wade 1985). In contrast, if density-dependent regulation occurs at the level of the social group (soft selection, Wallace 1968; see also Haldane 1924), then the increased success of the recipient must be paid for by the group. Without kin discrimination, the relatedness of the actor to the other members of the group will have been equally raised by population viscosity. Hence, population viscosity will not necessarily favour indiscriminate altruism (Hamilton 1971, 1975, Taylor 1992a,b).

This effect of local competition between relatives can be incorporated into Hamilton's rule in a number of ways (Grafen 1984, Queller 1994, Frank 1998, West et al., 2002a). Queller (1994) reformulated the coefficient of relatedness in order to take this into account, giving a new measure which he described as "not just a statement about the genetic similarity of two individuals, it is also a statement about who their competitors are". Here, relatedness between actor and recipient is a regression as before, however it is now defined relative to a reference population of competitors, a proportion which are locals, and the remainder being average members of the global population. Obviously if all competition is global, the reference population is the global population, allowing for positive relatedness between social partners. At the other extreme, if all competition is at the level of the social group, relatedness to the average member of the social group will be zero. Frank (1998) chose not to redefine relatedness, but instead introduced a separate scale of competition parameter to be incorporated into the benefit component of Hamilton's rule in order to predict when social behaviours will be favoured by selection. This parameter (a) is simply the proportion of competitors which are local as opposed to global. Soft selection (local competition) had been relatively neglected in social evolution theory prior to these developments, and this contrasts with population genetics, where it has received much attention (Roughgarden 1979).

Although the importance of the scale of competition in the application of kin selection to altruism issues is now acknowledged (see West et al. 2002a for a recent review, and Griffin et al. 2004 for an empirical example), its implications for spite are underappreciated. Increasingly local competition, as well as disfavouring altruism, can enhance selection for spite. Hamilton was correct when he stated that spite should be restricted to tiny populations, however the ‘population’ of interest is that of the competitive arena. If competition is global, so that there is hard selection at the level of the social arena, then relatedness is measured with respect to the population as a whole. But as competition becomes increasingly local, the reference population shrinks towards the size of the social arena, which may contain only a few individuals (small N) and/or a significant proportion of identifiable positively related kin, such that the negative relatedness towards the other potential recipients is non-trivial, enhancing the selective value of spite.

Another way of seeing this is by considering a crucial difference between altruism and spite. Within a social group, individuals with greater altruism than the group average have reproductive success lower than the group average, but if more altruistic groups are more productive, altruists may have higher absolute success than nonaltruists when averaging over the whole population. When competition is global, fitness is proportional to absolute success, so that altruism can be a winning strategy. Increasingly local competition means that fitness is increasingly dominated by success relative to the social group average, and so altruism is less favoured. Conversely, spiteful behaviour incurs a direct cost and reduces the success of social partners, so that more spiteful individuals can have higher success relative to the group average, but suffer a reduction in absolute success. When competition is global and fitness is proportional to absolute success, spite cannot be favoured, but as competition becomes increasingly local fitness is increasingly determined by success relative to social partners, so that spite can be a winning strategy.

Illustrative overview

So far we have employed the standard approach of taking Hamilton's rule to be a given (for example, see Orlove 1975) and using this as an entry point into the analysis of social evolution. However, it is often more appropriate and rigorous to derive the rule using a direct fitness approach, particularly when the aim is to resolve problematic conceptual issues. We use the direct (neighbour-modulated) fitness maximization techniques of Taylor & Frank (1996) and Frank (1998) to derive Hamilton's rule, in order to (1) distinguish two different forms of spite, and (2) address the suggestion of Boyd (1982) that spite is often actually selfishness because it indirectly increments fitness through reducing the intensity of competition. The key to this is to distinguish possible direct benefits of spite that might accrue to positively related third parties, and indirect effects due to relaxed competition.

Let social groups comprise n equally abundant 'families', with kin recognition allowing discrimination of the proportion $1/n = k$ of the social group which are 'kin' from the remaining $1-k$ which are 'non-kin'. Spite directed at non-kin carries a cost (some function c), inflicts a negative benefit upon the victim (b), and also potentially directly benefits (d) individuals within one's family, so that personal success might be written as

$$S_{focal} = 1 + b[(1-k)z] - c[x] + d[k(1-k)y], \quad [4.1]$$

where x is the focal individual's spite strategy, y is the average strategy of its kin (including itself), and z is the average strategy played by the non-kin members of its social group. The local average and the average for the whole population are given by

$$S_{local} = 1 + k(b[(1-k)z] - c[y] + d[k(1-k)y]) + (1-k)(b[(1-2k)z + ky] - c[z] + d[k(1-k)z]) \quad [4.2]$$

$$S_{global} = 1 + b[(1-k)\bar{z}] - c[\bar{z}] + d[k(1-k)\bar{z}],$$

where \bar{z} is the average spite strategy played in the whole population. Following Frank's (1998) approach to including competition in models of social evolution, fitness can be expressed as success relative to that of the average competitor, i.e.

$$w = \frac{S_{focal}}{aS_{local} + (1-a)S_{global}}, \quad [4.3]$$

where the scale of competition parameter (a) is defined as the proportion of competition which occurs locally, i.e. at the level of the social group. Selection favours more spite whenever marginal fitness is positive ($dw/dx > 0$). As outlined by Taylor and Frank (1996, and Frank 1998), marginal fitness is given by the chain rule:

$$\frac{dw}{dx} = \frac{\partial w}{\partial x} + r_y \frac{\partial w}{\partial y} + r_z \frac{\partial w}{\partial z}, \quad [4.4]$$

where ∂ denotes a partial derivative, and $r_y = dy/dx$ and $r_z = dz/dx$ are the slopes of social partner phenotype on own phenotype (for kin and non-kin respectively). Assuming only minor variants ($x \approx y \approx z \approx \bar{z}$), and denoting $b[\] = db[\]/d\circ$, $c[\] = dc[\]/d\circ$ and $d[\] = dd[\]/d\circ$, we find that marginal fitness is positive ($dw/dx > 0$) when

$$\left(r_z \left[a(kr_y + (1-k)r_z) \right] \right) (1-k)b[\] + \left(r_y \left[a(kr_y + (1-k)r_z) \right] \right) k(1-k)d[\] > \left(1 \left[a(kr_y + (1-k)r_z) \right] \right) c[\]. \quad [4.5]$$

Note that the relatedness to the average competitor relative to the whole population is $\hat{r} = a(kr_y + (1-k)r_z)$, and the marginal costs and benefits of spite are $B = (1-k)b[\]$, $C = c[\]$, and $D = k(1-k)d[\]$. After making these substitutions, rearrangement of [4.5] obtains the condition

$$\frac{r_z \left[\hat{r} \right]}{1 \left[\hat{r} \right]} B + \frac{r_y \left[\hat{r} \right]}{1 \left[\hat{r} \right]} D > C. \quad [4.6]$$

The r terms denote relatedness of individuals with respect to their spite phenotypes, relative to the population average, \bar{r} . If R is used to denote relatedness *sensu* Queller (1994), i.e. measured relative to the average competitor, then [4.6] is simply

$$R_1B + R_2D > C. \quad [4.7]$$

This is the three-party extension to Hamilton's rule for spiteful interactions given by Foster et al. (2000), although here it is the consequence of an analysis rather than the starting point. R_1 is the relatedness to the victims of spite, and R_2 is the relatedness to the third party which receives any direct benefits. A major source of confusion over Hamilton's rule involves the meaning of the terms B and C (and in the above expression, D), and so it is worth pointing out that these are not fixed parameters – they are marginal values.

This form of the rule can be used to discriminate Hamiltonian and Wilsonian forms of spite (Hamilton 1970, 1971, Wilson 1975, Foster et al. 2000, 2001). Feeling that negative relatedness was implausible, Wilson (1975) proposed that spite directed against non-negatively related individuals could be favoured if it also delivered a benefit to a sufficiently positively related third party. In terms of the above notation, such Wilsonian spite occurs when $D > 0$ and $R_2 > 0$, and does not require a negatively related victim ($R_1 < 0$). Hamiltonian spite occurs when the victim is negatively related ($R_1 < 0$, and hence $R_1B > 0$; Hamilton 1970, 1971), and hence a direct benefit to positive relations ($D > 0$) is not always required in order for the spite to be favoured. From expression [4.6] we can see that: (1) negative relatedness depends on the ability to discriminate individuals who are less related than the average competitor (so that $r < \hat{r}$); and (2) the magnitude of this negative relatedness increases as competition becomes more localized (increasing a , and hence increasing \hat{r}). Clearly, there is potential for spiteful behaviours to involve both negative relatedness to victims and positive benefits to positive relations, and hence a mixture of Hamiltonian and Wilsonian spite (Foster et al. 2000, 2001).

Related to this distinction, we can address the suggestion of Boyd (1982) that spite is actually less likely to occur under local competition, since the resulting relaxed competition gives an indirect benefit to spiteful individuals, so that many cases of spite would in fact be selfishness. The expression [4.7] reveals that the relaxation of competition due to spite is absorbed into the negative relatedness term, when relatedness is measured relative to the average competitor. Boyd's indirect benefit to the spiteful individual does not make the action selfish, in the same way that this indirect benefit accrued to other positive relatives does not mean that the spiteful behaviour is Wilsonian.

It is important to note that the above is not a general model for spite, but is rather an example included for the purpose of illustration. For instance, we have assumed additivity of fitness components, and equally abundant families. For this reason, it is always more rigorous to do a direct fitness analysis for particular models of interest in order to obtain the appropriate Hamilton's rule, rather than using the rule as a starting point.

Biological applications

Applying the theory to biological examples, we show: (1) that previous models which have successfully demonstrated selection for spite have tended to implicitly assume local competition; (2) behaviours previously interpreted as indirect altruism or Wilsonian spite might turn out to involve negative relatedness and hence Hamiltonian spite; (3) spiteful greenbeards are more likely to reach their threshold frequency, above which they are favoured by selection, when competition is localized; (4) there are several general predictions which will help us identify situations where spite is likely to be found, and (5) these predictions are amenable to empirical testing.

Spiteful models assume local competition

Theoretical models that show that spiteful behaviour can be favoured often assume that some or all of competition is local. However, this has rarely been acknowledged as an important factor contributing to the success of spite. For example:

1. Reinhold (2003) used an inclusive fitness analysis to investigate fatal fighting in fig wasps. This model shows selection for spite when competition is completely local. Some fig wasps have a lifecycle, such that wingless males hatch, mate and die within the confines of the fruit, and the mated females disperse to be the foundresses of new figs (Hamilton 1979, Cook et al. 1997). This leads to an asymmetric scale of competition, such that males compete locally (for mates) and females compete globally (for figs in which to lay eggs), the consequences of which for sex allocation theory have been much studied (Herre et al. 2001). In some species, this local competition for mates is accompanied by lethal combat between heavily armoured males, which have mandibles capable of decapitating each other (Hamilton 1979, Murray 1987, West et al. 2001b). Reinhold (2003) predicted that if males could discriminate between relatives and non-relatives (kin recognition) then they would be selected to fight with males who are non-relatives. This cannot be explained simply as selfishness because there is generally a net direct fitness cost of fighting (the difference in the direct fitness component of Reinhold's equations 2.1 & 2.2 for the terms T_1 & T_2). However, it can be explained as Hamiltonian spite, because the local competition means there is a negative relatedness towards opponents. Following Reinhold's notation, n males compete locally for matings, including a focal actor who is related to a proportion y of the other males (his brothers) by r and to the remaining $(n - 1)(1 - y)$ males (nonkin) by zero. Rescaling such that the focal individual is related to competitors on average by zero, we find that the relatedness to his brothers is $[n(1 - y)r - (1 - r)y]/[(n - 1)(1 - r)y]$ and to the unrelated males is $-[1 + (n - 1)ry]/[(n - 1)(1 - r)y]$, i.e. a negative quantity. The importance of spite in this system depends upon the possibility of kin discrimination between male fig wasps, which has yet to be tested for.

2. Gardner et al. (2004) presented a model of chemical (bacteriocin) warfare between microbes. Bacteriocins are the most abundant of a range of antimicrobial compounds produced by bacteria, and are found in all major bacterial lineages (Riley & Wertz 2002). They are a diverse family of proteins with a range of antimicrobial killing activity including enzyme inhibition, nuclease activity and pore formation in cell membranes (Reeves 1972, Riley & Wertz 2002). They are distinct from other antimicrobials in that their lethal activity is often limited to the same species of the producer, suggesting a major role in competition with conspecifics (Riley et al. 2003). Since bacteriocin synthesis is energetically expensive and release can entail death of the producer cell (for instance, colicin production by *Escherichia coli*) production of bacteriocins is costly ($C > 0$). Bacteriocins kill susceptible bacteria, and hence these recipients suffer a negative benefit ($B < 0$). Hence bacteriocin production can be regarded as a spiteful trait. Since kin of the producer cell are immune to its bacteriocins, there is effective kin discrimination, and the potential for recipients to be negatively related to the producer. Specifically, this relatedness is $R = - (a k)/(1 - a k)$ where k is the proportion of the social group which are clonal kin of the producer, and a is the proportion of competition which occurs locally. This reveals the importance of local competition in the evolution of spiteful behaviour. Specifically, (a) spiteful bacteriocin production is only selected for when there is some local competition ($a > 0$; since $R = 0$ when $a = 0$), and (b) as the degree of local competition (a) increases the evolutionary stable strategy (Maynard Smith & Price 1973) is to increasingly allocate resources to spiteful bacteriocin production (Gardner et al. 2004).

3. Cytoplasmic Incompatibility (CI), the phenomenon whereby maternally-transmitted *Wolbachia* (and other) bacteria occurring in male hosts sterilize uninfected female hosts upon mating (O'Neil et al 1997), has been interpreted as a form of spite (Hurst 1991, Foster et al. 2001). Infected females are compatible with infected males, and so there is effective discrimination of carriers and non-carriers of the parasite. The question of whether it can be favoured by selection has received much attention (Prout 1994, Turelli 1994, Frank 1997b). Frank (1997b) demonstrated that selection can favour CI in structured host populations. In his model, the sterilization of uninfected females relaxes

competition for the infected progeny produced by the group. In particular, Frank highlighted the importance of kin associations, so that related bacteria are carried by several hosts within the group. Less emphasis was given to the assumption of density dependent regulation at the group level, so that all competition is local ($a = 1$). Similar reasoning can be applied to the evolution of such selfish elements as maternal-effect lethal distorter genes (Beeman et al. 1992, Hurst 1993, Hurst et al. 1996, Foster et al. 2001), in which the killing of non-carriers relaxes competition among the carriers of the killer allele.

Hamiltonian and Wilsonian spite

Inequality [4.7] can be used to discriminate between Hamiltonian and Wilsonian forms of spite, and assess their relative importance when both occur (i.e. when spite is directed at negatively related individuals but also accrues a net inclusive fitness benefit by directly enhancing the success of positive relations). In particular, using measures of relatedness that take into account the effects of competition, we can reinterpret many putative examples of Wilsonian spite as Hamiltonian spite or a mixture of the two. For instance, Foster et al. (2000, 2001) present two spiteful behaviours presented by the eusocial insects which they describe as Wilsonian: worker policing and sex allocation manipulation.

Often in eusocial hymenopteran societies, worker individuals do not have the opportunity to mate, but nevertheless have functioning ovaries, and can therefore produce unfertilized eggs which may develop as haploid males (Wilson 1971, Bourke 1988). Worker policing, the phenomenon whereby workers eat the eggs of other workers in their colony (Ratnieks 1988), is well documented (Ratnieks & Visscher 1989, Foster & Ratnieks 2000, 2001, Foster et al. 2002, Barron et al. 2001). Foster et al. (2000, 2001) argue that this costly policing behaviour enhances the inclusive fitness of the actor as it frees up resources for the queen's sons (their brothers), to which they are more related than the sons of other workers (their nephews), and hence the spite is of the Wilsonian form. However, given that competition between the progeny for resources is within the colony, it is appropriate

to measure relatedness with respect to this local competitive arena when assessing the inclusive fitness consequences for this particular behaviour. This means that the victim of the policing (a nephew) is less related than average (all brothers and nephews) and hence negatively related to the actor (i.e. $R_1 < 0$). Consequently, if relatedness is measured at the scale of competition, worker policing can be interpreted as involving Hamiltonian spite.

The haplodiploid genetics of the hymenoptera means that in eusocial species the workers can be more related to their diploid sisters than their haploid brothers. This means that, while the queen prefers equal sex allocation among reproductives, the workers would rather there was a female bias (Trivers & Hare 1976). In some species the workers create this bias by killing of male progeny (Sundström et al. 1996, Passera & Aron 1996, Chapuisat & Keller 1999, Hammond et al. 2003). Foster et al. (2000, 2001) suggest this killing of male progeny is Wilsonian spite that benefits the colony's female progeny. However, the local competition for resources within the colony, plus the fact that males are devalued relative to females in terms of relatedness, means that the recipient of the spite is negatively related to the actor ($R_1 < 0$). Again, this behaviour may be reinterpreted involving Hamiltonian spite.

Application of the theory should also allow reinterpretation of behaviours which have not been considered spiteful (Hamiltonian or otherwise) in the past. Precocious larval development in polyembryonic parasitoid wasps (Godfray 1992, Grbic et al. 1992, Hardy et al. 1993, Ode & Strand 1995) seems to constitute a previously overlooked example of spite. Typically, two eggs, one male and one female, are laid on or in the body of the host insect, which then divide asexually to produce a brood of genetically identical brothers and genetically identical sisters. Local competition for resources limits the number of adult wasps emerging from the host, suggesting that there is scope for negative relatedness between the sexes within the brood. Upon inspection, some of the larvae that have not emerged as adults are found to have developed precociously, giving up their own future reproduction in order to murder opposite-sex siblings developing in the same host. Asymmetric dispersal (which generates a sex difference in the scale of competition), and asymmetric relatedness (brothers are more related to sisters than vice versa) seem to

be responsible for evolutionary resolution of this conflict in favour of the sisters, such that most precocious larvae are female.

Local competition can enhance the success of spiteful greenbeards

Greenbeards are phenotypic markers for genetic composition that allow individuals to identify positive relations more effectively than through discrimination of genealogical kin from nonkin (Hamilton 1964, 1971, Dawkins 1976). A greenbeard gene causes three things: (1) a phenotypic trait, (2) recognition of this trait in others, and (3) preferential treatment of those recognized – see Queller et al. (2003) for an example of a single gene which satisfies these three conditions. From the perspective of the greenbearded actor, social partners displaying the phenotype carry his gene and hence are positively related to him, and those who do not display the phenotype do not carry his gene, and are therefore negatively related to him, with respect to that locus. Greenbeards can therefore increase in frequency either by directing altruism towards the positive relations or else by directing spite towards the negative relations. However, nontrivial negative relatedness is only possible when the greenbeard allele is at a substantial frequency in the population, as Hamilton (1971) understood, making it difficult to imagine how a spiteful greenbeard could initially be selected. This problem is not felt by altruistic greenbeards, who have maximal relatedness between bearers of the gene even when the greenbeard is at low frequency in the population. The understanding that it is the arena of competition that provides the appropriate reference, rather than the population as a whole, means that the spread of spiteful greenbeards can be more easily understood, and the attainment of the threshold frequency does not have to rely upon assumptions such as extreme stochastic fluctuations.

Foster et al. (2000, 2001) discuss the example of the red fire ant (*Solenopsis invicta*; see also Keller & Ross 1998 and Hurst & McVean 1998), in which workers with genotype Bb, under the influence of their greenbeard (b) gene, murder negatively related BB queens and hence increase the frequency of the b gene in the population (homozygotes for the greenbeard gene are absent as the bb genotype is lethal). It is easy to see how the

frequency of the b allele in among the small number of locally competing queens could, through sampling error, exceed the threshold even as the global frequency approaches zero.

Where should we expect spite?

The extensions to spite theory, and biological examples of spite discussed above, suggest several clues as to where we should expect such behaviours to occur. Hamilton (1970, 1971) noted that spite should be more prevalent when actors are in a position to inflict damage to others at little cost to themselves, and so it is unsurprising that many examples turn up among non-reproductives in the eusocial insects, those individuals who have little or nothing to lose with respect to their direct fitness (Foster et al. 2000, 2001). A major factor which has received much attention is the ability to identify one's negative relations. This can be achieved through recognition of genealogically close individuals (kin discrimination) or by means of phenotypic markers for genetic composition (greenbeards). We also emphasize that spite should be looked for in situations where competition is mostly local (among social partners), and in viscous populations.

Empirical testing of spite theory

Previous debate over spite has focused primarily on whether it occurs. However, some of the more recent examples, such as worker policing in the eusocial insects and bacteriocin production by bacteria, provide possibilities for testing whether the relative occurrence of spite varies as predicted by social evolution theory. Indeed, much of the data from the eusocial insects fits well with the predictions of theory (Chapuisat & Keller 1999, Ratnieks et al. 2001). Here we emphasise two general points.

1. We have used Hamilton's rule to give an overall conceptual view. However, if particular cases are to be analysed then it is often much easier and more rigorous to start with an equation for direct (neighbour-modulated) fitness based upon the relevant biology, and then derive predictions (Taylor & Frank 1996, Frank 1998). Hamilton's rule

in some form usually appears as a consequence of such an approach, and provides a conceptual tool that can be used for interpretation of the results (Frank 1998, Pen & Weissing 2000, West & Buckling 2003).

2. A relatively general prediction that arises from different models is that the incidence of spite should be dome-shaped in relation to the degree of kinship within a social group. If the proportion of kin (including oneself) in the group is vanishingly small then no spite is favoured, since the non-kin recipients of spite will have the same relatedness, on average, as the average group member (i.e. zero). Similarly, when the actor associates solely with clonal kin, spite is also not favoured, as there are no negatively related individuals present. However, when the degree of kinship takes intermediate value, some degree of spite might be favoured because some individuals will necessarily be less related to the actor than others, such that some will have below-average (and hence negative) relatedness. This result was found in both the bacteriocin (Gardner et al. 2004) and fig wasp mortal combat (Reinhold 2003) examples discussed above. The relatedness differential also selects for spiteful sex allocation manipulation (brothers are less related than sisters) and worker policing (nephews are less related than sons and brothers) discussed above. As well as suggesting where we might find spite occurring in nature, these models give predictions that could be tested with observational or experimental studies.

Summary

Spite has been neglected by social evolution theory because a common, implicit assumption (global competition) in evolutionary models tends to diminish its selective advantage. We have demonstrated that many previously analysed behaviours can be readily interpreted as involving spite. Furthermore, theory has been developed to such a degree that we can make testable predictions about where spite is likely to be found and how it relates to the degree of competition and kinship between social partners.

5. Bacteriocins, spite and virulence[†]

Abstract

There has been much interest in using social evolution theory to predict the damage to a host from parasite infection, termed parasite virulence. Most of this work has focused on how high kinship between the parasites infecting a host can select for more prudent exploitation of the host, leading to negative relationship between virulence and parasite kinship. However, it has also been shown that if parasites can cooperate to overcome the host, then high parasite kinship within hosts can select for greater cooperation and higher growth rates, hence leading to a positive relationship between virulence and parasite kinship. We examine the impact of a spiteful behaviour, chemical (bacteriocin) warfare between microbes, on the evolution of virulence, and find a new relationship: virulence is maximised when the frequency of kin among parasites' social partners is low or high, and is minimized at intermediate values. This emphasises how biological details can fundamentally alter the qualitative nature of theoretical predictions made by models of parasite virulence.

Introduction

There is a large theoretical literature applying evolutionary theory to explain the damage that parasites cause to their hosts (van Baalen & Sabelis 1995, Frank 1996a, Gandon *et al.* 2001, Day & Burns 2003). Parasite virulence presents a fundamental trade-off in that parasites must deplete host resources in order to grow and transmit to new hosts, yet over-exploitation can result in host mortality and associated reduction in resource availability (Frank 1996a). This is the 'tragedy of the commons' (Hardin 1968), in which

[†] Published as: Gardner A., West S.A. & Buckling, A. 2004. Bacteriocins, spite and virulence. *Proceedings of The Royal Society of London Series B – Biological Sciences* **271**, 1529-1535 (see Appendix).

individuals are expected to display altruistic self-restraint only if they are sufficiently related to their group (Frank 1998). A classic result of virulence theory is that intensity of exploitation and hence damage to hosts correlates negatively with kinship among the parasites infecting a host (Hamilton 1972, Bremerman & Pickering 1983, Frank 1992, 1996a). This occurs because a lower relatedness leads to greater competition for resources, which selects for faster growth rates to obtain a greater proportion of the host resources, and these higher parasite growth rates lead to higher virulence.

However, empirical support for this prediction is severely lacking (Herre 1993, 1995, Chao *et al.* 2000, Read & Taylor 2001, Read *et al.* 2002, Griffin & West 2002, Davies *et al.* 2002). One possible explanation for this is that variation in the underlying biological details can lead to alternative relationships (Frank 1996, Schjorring & Koella 2003, Ganusov & Antia 2003). In particular, it has been shown that if parasites can cooperate to overcome their hosts defenses then the opposite prediction is favoured – a positive relationship between parasite kinship and virulence (Chao *et al.* 2000, Brown *et al.* 2002, West & Buckling 2003). For example, West & Buckling (2003) modeled the evolution of the production of costly public goods (siderophores) which promote bacterial growth during iron-starvation in an infection. Not surprisingly, the altruistic production of siderophores is expected to be maximised when kinship is highest, yet this leads to enhanced growth and therefore host damage precisely where previous theory predicted self-restraint and hence low virulence.

Just as altruistic behaviour can promote parasite growth and hence enhance virulence, it is reasonable to assume that spiteful interactions (interference competition) between parasites could reduce the vigour of an infection and associated host damage. Here we consider such a spiteful trait: bacteriocin production. Bacteriocins are the most abundant of a range of antimicrobial compounds facultatively produced by bacteria, and are found in all major bacterial lineages (Riley & Wertz 2002). They are a diverse family of proteins with a range of antimicrobial killing activity, many of which can be produced by a single bacterium, including enzyme inhibition, nuclease activity and pore formation in cell membranes (Reeves 1972, Riley & Wertz 2002). Unlike other antimicrobials, the

lethal activity of bacteriocins is often (but not always) limited to members of the same species as the producer, suggesting a major role in competition with conspecifics (Riley *et al.* 2003). Intraspecific competition may also help to explain the observed variation in the types of bacteriocins produced by different strains of the same species. For example, at least twenty-five bacteriocins (colicins) have been identified in populations of *Escherichia coli*, with different populations producing unique combinations (Riley & Gordon 1999). Clone mates are protected from the toxic effects of bacteriocins as a result of genetic linkage between the bacteriocin gene and an immunity gene that encodes a factor that deactivates the bacteriocin (Riley & Wertz 2002).

In addition to the benefits of bacteriocin production (killing competitors), there are also costs (Chao & Levin 1981, Kerr *et al.* 2002, Reeves 1972). This cost may simply be diversion of resources from other cellular functions, but in many gram-negative bacteria, such as *Escherichia coli*, cell death is required for the release of bacteriocins (Reeves 1972, Riley & Wertz 2002). Such costs (and costs associated with bacteriocin immunity) are critical for coexistence, between bacteriocin-producing, sensitive and resistant strains (Czárán *et al.* 2002, Kerr *et al.* 2002, Czárán & Hoekstra 2003). Here, we investigate how key parameters affect the relative costs and benefits of bacteriocin production, hence the level favoured by natural selection, and the impact this has on disease virulence. Specifically, we consider how bacteriocin production evolves in response to the average kinship (r) of competing bacteria and the scale of competition relative to the effective range of bacteriocins (a)

Models, methods & analyses

Equal abundance model

We first consider a social arena, defined as the spatial range of bacteriocin warfare, comprising n equally abundant lineages drawn independently from the asexually reproducing bacterial population. A proportion $r = 1/n$ of the bacteria within a focal bacterium's social arena are its clonemates, or 'kin'. The remaining $1 - r$ are derived

from the other $n - 1$ lineages, and are ‘non-kin’. Using a game theoretic approach, we consider the fitness of a vanishingly rare mutant which allocates an amount of resources y into bacteriocin production within a population with average allocation z , in order to determine the ‘unbeatable’ (Hamilton 1967) or ‘evolutionarily stable’ (Maynard Smith & Price 1973) allocation strategy y^* . An amount of bacteriocin ry within the social arena is attributable to the focal lineage, and rz to each of the other lineages. The focal lineage is then subjected to an amount $(1 - r)z$ of unrelated bacteriocin to which it is susceptible, and for each of the $n - 1$ other lineages, $(1 - r)z + r(y - z)$. A lineage picked at random from the population as a whole experiences, on average, $(1 - r)z$ unrelated bacteriocin. Lineages are immune to their own bacteriocins, and although resistance (nonsusceptibility of a lineage to a bacteriocin which it does not itself produce) is not explicitly discussed in this model, the resulting reduction in susceptibility can be regarded as included in the general growth functions. The growth rate of a lineage, G , is given by the sum of two components, H and I . H reflects the cost of bacteriocin production, being a positive, decreasing function of the focal lineage’s allocation to bacteriocin production, y . Our predictions rely on no specific form for H ; when a specific relationship is required for illustrative purposes (figures 5.1 - 5.3), we use $H = 1 - y$. I models the reduction in growth due to mortality by unrelated bacteriocins, being a positive, decreasing, linear or decelerating function of the amount (Y) of unrelated bacteriocin it is subjected to. Our predictions rely on no specific form for I ; when a specific relationship is required for illustrative purposes (figures 5.1 - 5.3), we use $I = 1 - Y^{1/2}$. We combine the terms H and I additively to give overall growth ($G = H + I$) for mathematical convenience, as it permits greater tractability than using a multiplicative scheme ($G = H \square I$), and does not qualitatively change the results (see *General model*). Using the construction of Frank (1998), fitness is determined by the growth of the lineage relative to the average competitor of that lineage:

$$w = \frac{G_{focal}}{a G_{local} + (1 - a) G_{global}}. \quad [5.1]$$

The parameter a defines the (spatial) scale at which competition for resources takes place. This model therefore allows competition for resources and bacteriocin interaction to take place at different scales. Specifically, a proportion a of competition for resources occurs locally, within the scale of bacteriocin interaction, and the $(1 - a)$ remainder occurs globally. At the extremes: if $a=1$ then competition for resources and bacteriocin interaction occur at the same scale (soft selection at the level of the social group); if $a=0$ then competition is at the global level (hard selection at the level of the social group). G_{focal} , G_{local} and G_{global} are respectively the growth rate of the focal lineage, the local average, and the global average. These are, in full

$$G_{focal} = H[y] + I[(1 - r)z]$$

$$G_{local} = r(H[y] + I[(1 - r)z]) + (1 - r)(H[z] + I[(1 - r)z] + r(y - z)) \quad [5.2]$$

$$G_{global} = H[z] + I[(1 - r)z]$$

Equations [5.1] and [5.2] illustrate the fundamental trade-off in our model. Bacteriocin production by the focal lineage is: (a) costly, because it lowers the growth rate of the focal lineage (G_{focal}), and (b) beneficial, because it lowers the growth rate of competitors G_{local} .

Substituting [5.2] into [5.1] we obtain fitness function $w[y, z]$. If we assume only minor variants ($y \approx z$; Taylor & Frank 1996, Frank 1998) the marginal fitness is found to be

$$\left. \frac{dw}{dy} \right|_{y=z} = \frac{(1 - ar)H'[z] - ar(1 - r)I'[(1 - r)z]}{H[z] + I[(1 - r)z]} \quad [5.3]$$

where $H' < 0$ is the derivative of H with respect to its parameter (e.g. y in the instance of the mutant), and may be interpreted as the marginal cost ($-C$) of producing bacteriocins. $I' < 0$ is the derivative of I with respect to its parameter (e.g. $(1 - r)z$ for the amount of bacteriocin attacking the focal mutant), and is the negative 'benefit' accrued by the

recipient of spiteful behaviour – summing over all the recipients, the benefit is $B = (1 - r)I'[(1 - r)z]$. Increased bacteriocin production (y) is favoured whenever $dw/dy > 0$ is satisfied, yielding Hamilton's (1963, 1964, 1970) rule:

$$\square \frac{ar}{1 - ar} B > C \quad [5.4]$$

As is often the case (Taylor & Frank 1996, Frank 1998), inspection of the direct marginal fitness (equation [5.3]) yields a form of Hamilton's (1963) rule $RB > C$ (equation [5.4]). In this: (a) relatedness is negative and given by $R = -(ar)/(1 - ar)$; (b) the negative 'benefit', summed over all recipients, is $B = (1 - r)I'[(1 - r)z]$ where $I'[Y]$ is the derivative $dI[Y]/dY$ and represents the marginal reduction in growth of a lineage which is poisoned by an amount Y of foreign bacteriocins. To understand how a negative relatedness can arise, we will use the result of Queller (1994) that average relatedness to one's competitors is zero. Recalling that the scale of competition (a) is defined as the proportion of competition which is local, consider an arena of competition in which a proportion of competitors a are social partners, and of these a proportion r belong to the focal lineage. Then a proportion ar of competitors are clonally related to the spiteful actor by 1, and a proportion $1 - ar$ are related by some unknown coefficient R . Applying Queller's insight, we know that $ar \square 1 + (1 - ar) \square R = 0$, and rearranging we obtain $R = -(ar)/(1 - ar)$. Hence:

RESULT 1. The evolution of bacteriocin production involves a negative relatedness between actor and recipient, and hence fits Hamilton's (1970) original definition of a spiteful behaviour. Relatedness between non-kin social partners is given by $R = -(ar)/(1 - ar)$, where a is the proportion of competition which is local, and r is the proportion of social partners which are clonal kin. This equation gives negative values for relatedness, except when either of (or both) a and r are zero, in which case relatedness equals zero.

We now ask how the ESS bacteriocin production (y^*) changes over a range of bacterial kinship (r) and intensity of local resource competition (a). Substituting $r = 0$ into [5.3]

obtains $H'[z]/(H[z]+I[z])$, which is negative and hence $y^* = 0$. When $r = 1$, [5.3] becomes $(1 - a) H'[z]/(H[z]+I[z])$ which is negative and so $y^* = 0$. When $a = 0$, [A1.1] gives H' , which is negative so that $y^* = 0$. Therefore the presence of more than one lineage ($0 < r < 1$) and some degree of local competition ($a > 0$) are essential for nonzero allocation to bacteriocin production. If we denote the RHS of [5.3] by J , then the ESS $z = y^*$ satisfies $J = 0$. Using implicit differentiation, we can write

$$\frac{dy^*}{dr} = - \frac{\partial J / \partial r}{\partial J / \partial y^*} \quad [5.5]$$

where ∂ denotes partial derivatives. For y^* to be convergence stable (i.e. in a population close to y^* , mutants closer to y^* are favoured by selection), the denominator on the RHS of [5.5] must be negative (Taylor 1996). Hence, assuming convergence stability, dy^*/dr has the same sign as $\partial J / \partial r$ (Pen 2000). Evaluating the partial derivative at $r = 0$ (and hence $y^* = 0$) yields $-a(H[0]+I[0])(H'[0]+I'[0])/(H[0]+I[0])^2$ which is positive when $a > 0$. If we assume no discontinuities in y^* , then this indicates that when there is some degree of local competition, and intermediate relatedness, ESS bacteriocin production (y^*) will be nonzero.

RESULT 2. Enhanced bacteriocin production is favoured at intermediate kinship (r). The evolutionarily stable strategy (ESS) is $y^* = 0$ at $r = 0$ & 1, and is maximized somewhere in the range $0 < r < 1$ (figure 5.1 for numerical examples). When the focal lineage occupies only a tiny proportion ($r \approx 0$) of the social arena, its impact on competitor growth is negligible, and hence the benefit through competitor-killing does not outweigh the cost of bacteriocin production. When the focal lineage dominates the social group ($r \approx 1$), the density of cells susceptible to its bacteriocin is too low for the benefit of competitor killing to outweigh the production costs.

Using the same procedure, we may find the partial derivative of J with respect to the scale of competition, a .

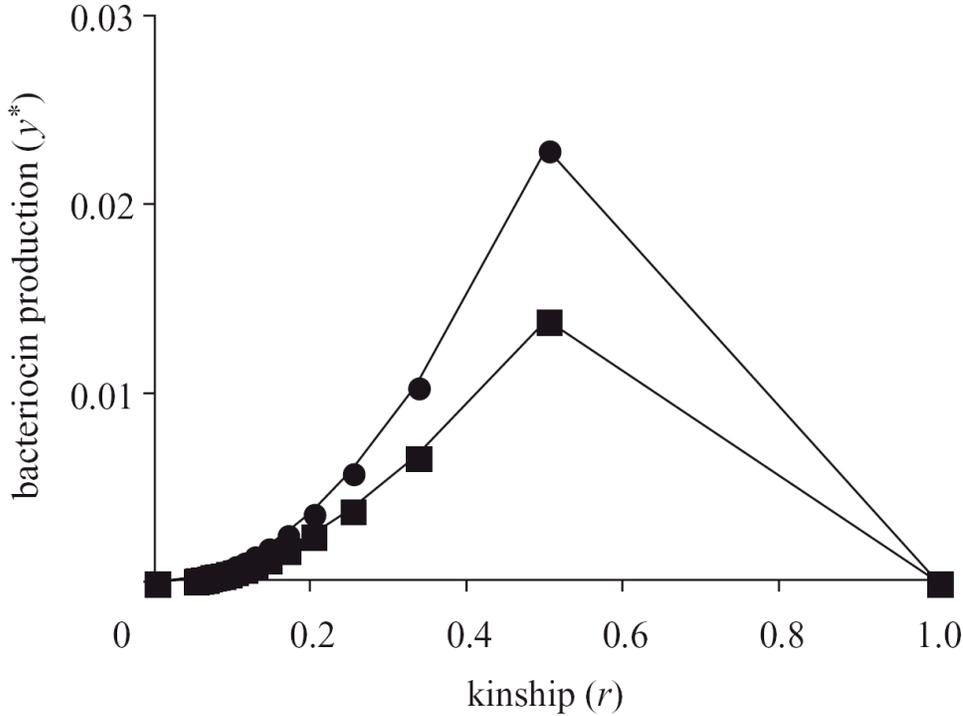


Figure 5.1. The ESS production of bacteriocins (y^*) as a function of the average kinship (r) between bacteria. Values are obtained numerically using the *equal abundance* model, assuming that bacterial growth is the sum of growth components $H = 1 - y$ and $I = 1 - Y^{1/2}$ (where the focal bacterium produces an amount y of its own bacteriocins, and receives an amount Y from its social partners) and the proportion of competition which is local is $a = 0.5$ (filled squares) and 0.6 (filled circles). Intermediate kinship (r) and increasingly local competition (high a) favour enhanced bacteriocin production.

$$\frac{\partial J}{\partial a} = \frac{rH[y^*] + r(1-r)I[(1-r)y^*]}{H[y^*] + I[y^*]} \quad [5.6]$$

which is positive for all $0 < r < 1$, and hence bacteriocin production is an increasing function of the scale of competition (a) when kinship is intermediate.

RESULT 3. Enhanced bacteriocin production is favoured as the scale of competition a is increased (and hence competition for resources becomes more local) for all $0 < r < 1$.

This occurs because fitness can be enhanced in two ways: (1) maximizing own growth (G_{focal}) and (2) reducing the growth of local competitors (G_{local}). When competition is entirely global ($a = 0$), there is no benefit in reducing growth of local competitors, so that the ESS is the strategy that maximizes focal growth (by reducing bacteriocin production).

As competition becomes more local ($a > 0$), production of bacteriocin is increasingly favoured in order to reduce the growth of the local competitors.

Variable abundance model

We now relax the assumption of equally abundant lineages, looking at the situation where only two lineages occupy the social arena, so that the focal lineage comprises a proportion r or $1 - r$ of the bacterial cells with equal probability. The appropriate fitness function is then

$$w = r \frac{G_{focal1}}{aG_{local1} + (1 - a)G_{global}} + (1 - r) \frac{G_{focal2}}{aG_{local2} + (1 - a)G_{global}} \quad [5.7]$$

where

$$G_{focal1} = H[y] + I[(1 - r)z]$$

$$G_{focal2} = H[y] + I[rz]$$

$$G_{local1} = r(H[y] + I[(1 - r)z]) + (1 - r)(H[z] + I[ry]) \quad [5.8]$$

$$G_{local2} = (1 - r)(H[y] + I[rz]) + r(H[z] + I[(1 - r)y])$$

$$G_{global} = H[z] + rI[(1 - r)z] + (1 - r)I[rz]$$

Following the same procedure as before, we obtain

$$\left. \frac{dw}{dy} \right|_{y=z} = \frac{ar(1-r) \left[r(H[z] + I[(1-r)z])I[rz] + (1-r)(H[z] + I[rz])I[(1-r)z] \right]}{(H[z] + rI[(1-r)z] + (1-r)I[rz])^2} + \frac{(1-a(1-2r(1-r)))H[z] + r(1-ar)I[(1-r)z] + (1-r)(1-a(1-r))I[rz]}{(H[z] + rI[(1-r)z] + (1-r)I[rz])^2} \quad [5.9]$$

Setting $r = 0$ yields $(1-a)H'[z]/(H[z]+I[0])$ which is always negative and hence $y^* = 0$ at $r = 0$. Setting $r = 1$ yields $(1-a)H'[z]/(H[z]+I[0])$ which is always negative, so $y^* = 0$ at $r = 1$. And when $a = 0$, we obtain $H'[z]/(H[z]+rI[(1-r)z]+(1-r)I[rz])$ which is always negative, so that $y^* = 0$ when $a = 0$.

As before, if we define J as the RHS of [A1.8] when $z = y^*$, then it is possible to show that for $a > 0$, $\partial J/\partial r = dy^*/dr = 0$ is only satisfied for $r = 1/2$. Since $y^* = 0$ at $r = 0$ & $r = 1$, and assuming no discontinuities over the range of r , we can conclude that y^* monotonically increases over the range $0 < r < 1/2$ and monotonically decreases over the range $1/2 < r < 1$. Thus, we recover the prediction that ESS bacteriocin production will be maximized at intermediate bacterial kinship ($0 < r < 1$) – see figure 5.2 for numerical examples.

The partial derivative of J with respect to the scale of competition is $\partial J/\partial a = -(r(1-r)(r(H[y^*]+I[(1-r)y^*])I'[ry^*]+(1-r)(H[y^*]+I[ry^*])I'[(1-r)y^*])+(1-2r(1-r)H[y^*]+r^2I[(1-r)y^*]+(1-r)^2I[ry^*])H'[y^*])/(H[y^*]+rI[(1-r)y^*]+(1-r)I[ry^*])^2$, which is positive for all $0 < r < 1$, and hence bacteriocin production is an increasing function of the scale of competition (a) at intermediate kinship.

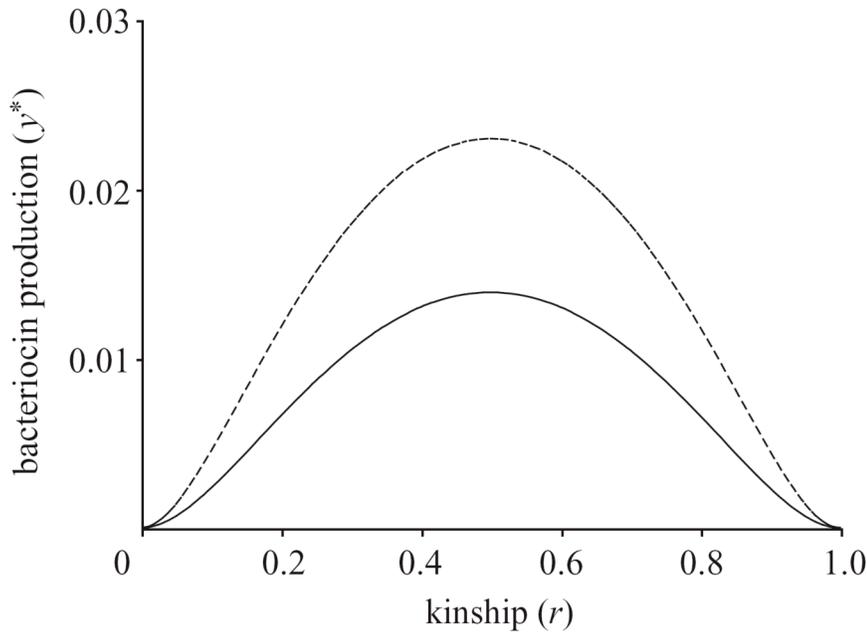


Figure 5.2. The ESS production of bacteriocins (y^*) as a function of the average kinship (r) between bacteria. Values are obtained numerically using the *variable abundance* model, assuming that bacterial growth is the sum of growth components $H = 1-y$ and $I = 1-Y^{1/2}$ (where the focal bacterium produces an amount y of its own bacteriocins, and receives an amount Y from its social partners) and the proportion of competition which is local is $a = 0.5$ (solid line) and 0.6 (dotted line). Intermediate kinship (r) and increasingly local competition (high a) favour enhanced bacteriocin production.

Host mortality

The above model is appropriate for free living bacteria, bacteria grown on agar plates, or parasitic bacteria in which host mortality doesn't influence the ESS production of bacteriocin. For parasitic bacteria, this would be appropriate when the extra host mortality due to the infection impinges very little upon bacterial success, or when there are a large number of social groups within the host, such that any lineage's growth rate has negligible impact on the mortality of the host. A simple model, relaxing these assumptions, considers that direct fitness of the focal lineage is given by the product $S \square T$ where S represents host survival (i.e. the time over which transmission is possible) and is a linearly decreasing function of the average growth rate of lineages in the host, and T is the transmission rate achieved by the focal lineage, i.e. its growth rate relative to competitors, the fitness measure given by equation 5.1. A parameter b is introduced to

denote the proportion of the bacterial population within the host which are in the focal arena of social (bacteriocin) interaction. $b = 0$ corresponds to when the social arena comprises a vanishingly small proportion of the total infection, and $b = 1$ corresponds to the arena of bacteriocin interaction being the entire infection. As in our first model, we assume n equally abundant lineages. The appropriate fitness function is

$$w = S[G_{host}] \frac{G_{focal}}{aG_{local} + (1-a)G_{global}} \quad [5.10]$$

Where the growth rate of a random lineage within the host is on average

$$G_{host} = bG_{local} + (1-b)G_{global} \quad [5.11]$$

Virulence (v) can be defined as the reduction in S relative to a host with zero bacterial growth ($G_{host} = 0$), i.e. $v = S[0] - S[G_{host}]$.

If $b = 0$, so that the social arena comprises a vanishing proportion of the bacterial population within the host, then $G_{host} = G_{global}$ and S is a constant with respect to y , so that marginal fitness is given by [A1.1]. For $b > 0$, and assuming only minor variants ($y \approx z$, $G_{focal} \approx G_{local} \approx G_{global} \approx G_{host} \approx G$), marginal fitness is

$$\frac{dw}{dy} = S[G]rb(H[z] + (1-r)I[(1-r)z]) + S[G] \frac{(1-ar)H[z] - ar(1-r)I[(1-r)z]}{G} \quad [5.12]$$

The second component on the RHS is proportional to the marginal fitness [5.3], and represents the trade-off between the cost and competitor-killing capabilities of bacteriocins. When $a = 0$, this component reduces to $(S[G] H'[z])/G$, which is always negative, reflecting the disadvantage of spite when competition is global. The first component, positive and proportional to rb , is the selection pressure for enhanced killing and costly production when growth of the focal lineage and its neighbours impact nontrivially upon host mortality. As r tends to zero, marginal fitness is negative ($S[G]$

$H'[z]/G$) as the behaviour of the focal lineage has no impact on host mortality and there is no advantage to be had from directing spite at local competitors (relatedness to non-kin in the social arena is zero). At $r = 1$, the second component is negative ($S[G](1 - a)H'[z]/G$) reflecting the fitness cost of bacteriocin production, and the first component is positive ($S'[G]H'[z]$) reflecting the enhanced fitness due to the reduction in host mortality. Note that this positive pressure is due entirely to the costs of bacteriocin production, and not through its bacteriocidal activity; this is due to an artificiality in the model such that the bacteria have no means of reducing own growth other than producing costly bacteriocin. Since no gain in terms of competitor killing is to be had from producing bacteriocins at $r = 1$, we expect $y^* = 0$.

If $y^* = 0$ at $r = 0$ & 1 , then since H and I are decreasing functions of y^* , it is here that $G_{host} = H + I$ is maximised. Since S decreases with increasing G_{host} , S is minimised at $r = 0, 1$. If we define virulence as the reduction in host survival relative to that for a host in which bacterial growth is zero ($v = S_{max} - S$), then virulence is maximised when S is minimised ($v_{max} = S_{max} - S_{min}$), i.e. at the extremes of relatedness, $r = 0$ and $r = 1$.

When a and b are both zero, so that there is no selection for spite nor for reduced virulence, [5.12] reduces to $(S[G] H'[z])/G$ which is negative and hence $y^* = 0$.

RESULT 4. Virulence (v) is maximised at the extremes of kinship ($r = 0$ & $r = 1$), and is minimised at intermediate values ($0 < r < 1$) – see figure 5.3 for numerical examples. This is due to the maximization of bacteriocin production at intermediate kinship, such that absolute growth of bacteria is reduced here but not at more extreme values, so that virulence is more pronounced whenever bacteria tend to socialize mostly, or not at all, with their kin.

General model

Relaxing the assumption of additive growth components, and making no further assumptions about the components of growth beyond that bacteriocin production reduces

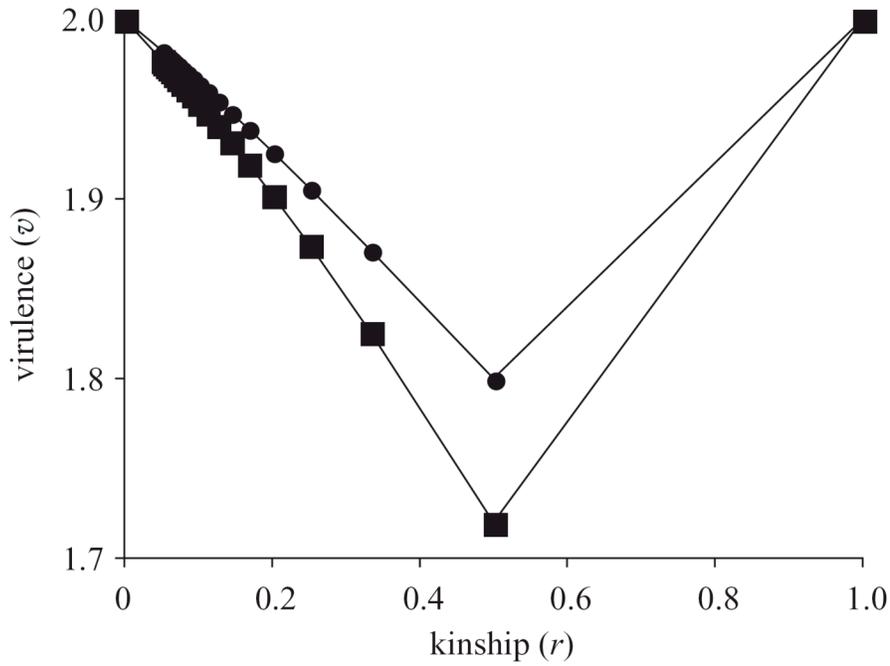


Figure 5.3. The virulence (v) as a function of the average kinship (r) between bacteria. Values are obtained numerically using the *host mortality* model, assuming that bacterial growth is the sum of growth components $H = 1 - y$ and $I = 1 - Y^{1/2}$ (where the focal bacterium produces an amount y of its own bacteriocins, and receives an amount Y from its social partners), host survival is $S = 3 - G_{\text{host}}$ (G_{host} is the overall bacterial growth in the host), the intensity of local competition is $a = 0.5$, and the range of bacteriocin warfare with respect to the whole infection is $b = 0.1$ (filled circles) and 0.2 (filled squares). Virulence is minimized at intermediate kinship (r) and when the range of bacteriocin warfare (b) is large.

the growth of the focal lineage (G_{focal}) and its nonkin social partners (G_{social}), we can recover the major predictions made in this study. Consider the fitness function [5.1].

Marginal fitness can be written

$$\frac{dw}{dy} = \frac{\left(aG_{\text{local}} + (1 - a)G_{\text{global}} \right) \frac{dG_{\text{focal}}}{dy} - G_{\text{focal}} \frac{d\left(aG_{\text{local}} + (1 - a)G_{\text{global}} \right)}{dy}}{\left(aG_{\text{local}} + (1 - a)G_{\text{global}} \right)^2} \quad [5.13]$$

Assuming only minor variants, so that $y \approx z$, and $G_{\text{focal}} \approx G_{\text{social}} \approx G_{\text{local}} \approx G_{\text{global}} \approx G$, we have

$$\frac{dw}{dy} = \frac{dG_{focal}}{dy} - a(1-r) \frac{dG_{social}}{dy} / G \quad [5.14]$$

Fitness increases with enhanced bacteriocin production when $dw/dy > 0$. dG_{focal}/dy is negative due to the production costs of bacteriocin, and dG_{social}/dy is negative because non-kin social partners experience higher mortality as bacteriocin production by the focal lineage is increased. [5.14] therefore demonstrates the tradeoff between the direct cost of bacteriocin production and the benefit of competitor killing. The benefit is zero when $a = 0$ and/or when $r = 1$, so that marginal fitness is $\{(1 - a r) dG_{focal}/dy\}/G < 0$ for all y , meaning that the ESS bacteriocin production is at $y^* = 0$. Also, the impact of the focal lineage's bacteriocin on competitor growth approaches zero as the focal lineage accounts for a vanishing proportion of the social group, i.e. at $r = 0$, $dG_{social}/dy = 0$, and so here the marginal fitness is negative, and $y^* = 0$. Regardless then of the precise details describing how the growth of the focal lineage and its nonkin social partners decline with enhanced bacteriocin production, provided that they do decline, we can state that the ESS is $y^* = 0$ when kinship is zero or complete ($r = 0, 1$) and when competition is entirely global ($a = 0$).

Discussion

We have shown that the production of bacteriocin is expected to be enhanced when kinship (r) is of intermediate value (result 2, figures 5.1 & 5.2). Since bacteriocin production is expected to correlate with low bacterial growth rates, virulence will tend to be minimized at intermediate r and maximised when bacteria compete only with non-kin ($r = 0$) or only with kin ($r = 1$). We therefore predict a U-shaped relationship between virulence and kinship (result 4, figure 5.3), contrary to previous models that variously predict monotonically increasing or decreasing virulence as kinship is increased. This emphasizes that the qualitative outcome of virulence evolution crucially depends on the biological details, such as whether parasites are able to improve their success through prudent growth (Frank 1996a), or cooperative contributions to public goods (Brown *et al.* 2002, West & Buckling 2003), or through anti-competitor toxin production.

Our result is intuitive if we consider that when kinship (r) is low the influence of the focal lineage on the growth of its social partners will be negligible, and so reduced allocation of resources into bacteriocin production is favoured. Conversely, when kinship is high, the proportion of cells in the social arena which are susceptible to bacteriocin-killing is small, and so the benefit of producing bacteriocin is less than the cost that this entails. At intermediate kinship bacteriocin production is favoured because competition with non relatives is important, and bacteriocin production by the focal lineage can significantly decrease the growth of the non-competitors. The result (3) that the ESS bacteriocin production is an increasing function of the degree to which competition is local (a ; figures 5.1 & 5.2) is also intuitive in that when competition is increasingly local the benefits accrued by reducing the growth of local competitors are enhanced.

The costly allocation of resources into bacteriocin production qualifies as an example of Hamiltonian spite (Hamilton 1970, 1996, Hurst 1991, Foster *et al.* 2001, Gardner & West (in press)). It is well accepted that altruism can be adaptive despite a direct fitness cost so long as the beneficiary of altruism is sufficiently positively related to the actor (i.e. a positive R and a positive B , and $RB > C$). Hamiltonian spite is when a costly behaviour is favoured because it has a cost to the recipient (negative B), and the recipient is negatively related to the actor (negative R , and $RB > C$). How can negative relatedness arise? Negative relatedness to some individuals is inevitable when positively-related individuals exist in the same competitive arena. The reason for this is that since the relatedness of an actor to a randomly chosen individual from its competitive arena is on average zero (Queller 1994), the existence of positive relations within that arena implies the existence of negatively-related competitors (result 1). In this situation, spiteful behaviour will be favoured if it can be preferentially directed at these negatively-related competitors, and $RB > C$ is satisfied. The specificity of bacteriocin action allows it to potentially fill this criterion, because it will preferentially harm non-relatives who are not resistant to that particular bacteriocin – i.e. bacteriocins harm individuals who are negatively related to the producer. Although the anti-competitor function of the bacteriocins suggests that this

is selfishness at the level of the clonal lineage, it is certainly spiteful at the level of the self-destructing bacterium producing the toxins.

To conclude, we have shown theoretically how kinship and the scale of competition determine levels of bacteriocin production favoured by natural selection. Contrary to previous work, we find a U-shaped relationship between kinship and virulence. The results are qualitatively the same whether bacteria have fixed strategies for bacteriocin production or if bacteriocin production is facultatively adjusted in response to kin recognition. These predictions could be tested by: (i) correlating bacteriocin production with average kinship in natural populations; or (ii) experimentally evolving bacteria under different degrees of kinship and scales of competition. Furthermore, our predictions are not limited to bacteriocin production by bacteria. A variety of microbes, including yeasts (Schmitt & Breinig 2002) and halophilic archaea (Cheung *et al.* 1997) are known to produce toxins that tend to target conspecifics.

6. Cooperation and punishment, especially in humans[†]

Abstract

Explaining altruistic cooperation is one of the greatest challenges faced by sociologists, economists, and evolutionary biologists. The problem is determining why an individual would carry out a costly behavior that benefits another. Possible solutions to this problem include kinship, repeated interactions, and policing. Another solution that has recently received much attention is the threat of punishment. However, punishing behavior is often costly for the punisher, and so it is not immediately clear how costly punishment could evolve. We use a direct (neighbor-modulated) fitness approach to analyze when punishment is favored. This methodology reveals that, contrary to previous suggestions, relatedness between interacting individuals is not crucial to explaining cooperation through punishment. In fact, increasing relatedness directly disfavors punishing behavior. Instead, the crucial factor is a positive correlation between the punishment strategy of an individual and the cooperation it receives. This could arise in several ways, such as when facultative adjustment of behavior leads individuals to cooperate more when interacting with individuals who are more likely to punish. More generally, our results provide a clear example of how the fundamental factor driving the evolution of social traits is a correlation between social partners and how this can arise for reasons other than genealogical kinship.

Introduction

Explaining cooperation at all levels of biological complexity remains one of the greatest problems for evolutionary biology (Hamilton 1964; Buss 1987; Maynard Smith and Szathmary 1995). The question is: why would an individual perform a costly altruistic

[†] Published as: Gardner, A. & West, S.A. (in press). Cooperation and punishment, especially in humans. *American Naturalist* (see Appendix).

behavior that benefits another individual? The solutions to this problem that have attracted the most attention are when social partners are related (kin selection, in a general sense; Hamilton 1963, 1964, 1970) or when there is some mechanism for repressing competition between groups, such as through repeated interactions/reputation (reciprocity; Trivers 1971; Alexander 1979, 1987; Frank 2003a), policing (Ratnieks 1988; Frank 1995b, 2003a), and systems of rewards or punishments (Oliver 1980; Sigmund et al. 2001). The fundamental similarity between all these mechanisms is that they involve positive correlations between the behaviors played by social partners, which are crucial for the evolution of social behaviors (Hamilton 1975; Grafen 1985a; Nee 1989; Frank 1998; Woodcock and Heath 2002).

Here, we are concerned with whether and how punishment can favor cooperation and how this translates into a selective benefit for punishers. The possible role of punishment has recently attracted much theoretical attention, especially with respect to its possible role in favoring cooperation among humans (Hirshleifer and Rasmusen 1989; Boyd and Richerson 1992; Sober and Wilson 1998; Sell and Wilson 1999; Fehr and Gächter 2000). However, the mechanism underlying these previous models is often not clear, and the models have been developed with little reference to related theory such as in the animal punishment literature (Clutton-Brock and Parker 1995; Clutton-Brock 1998) and Frank's (1998, 2003a) recent synthesis of social evolution theory. The basic idea is that if punishment is sufficiently frequent and harsh, it can successfully maintain cooperative behavior. However, this solution forces us to consider the motivation of the punisher. Since a behavior that promotes a public good such as cooperation is in itself a second-order public good and is not expected to be without cost to the actor, punishment is open for exploitation by second-order free-riding individuals who cooperate but who fail to punish defectors (Oliver 1980). Punishment of second-order free riders can be invoked, but this opens up the possibility of third- and higher-order free riding (Ostrom 1990). Failure to maintain participation in a high-level public-goods game unravels participation in the lower levels. At first glance, punishment seems not to be a helpful addition to the problem of cooperation because all that is achieved is the replacement of one public-goods dilemma for another. However, it is generally true that punishment is cheap relative to the

cost of cooperation. Consequently, it has been argued that any mechanism invoked to explain participation in public-goods games will more easily favor punishing (and hence also cooperation) than it would cooperation alone (Sober and Wilson 1998).

A Darwinian account of the evolution of cooperation through punishment requires that the punisher directly or indirectly receives a net benefit through punishing. Although costly punishment can ultimately enhance the direct fitness of the punisher if interactions tend to be extended or repeated with the same social partner (Frank 2003a; e.g., sanctioning in plant-rhizobium mutualisms: Denison 2000; West et al. 2002*c*, 2002*d*; Kiers et al. 2003), animals including humans punish even when there is no mechanism ensuring repeat encounters (Fehr and Gächter 2002). Genealogical relationship between social partners is often considered low or absent, and so kin selection is given little attention in the existing literature. The favored Darwinian mechanisms that have received the most attention are group selection (Gintis 2000) and cultural group selection (Heinrich and Boyd 2001). A recent simulation study (Boyd et al. 2003) has suggested that since the incidence of defection declines as punishment becomes more frequent, the costs of punishment decline as it becomes common, so that even modest group selection may plausibly maintain punishment in humans.

In this chapter, we show that the evolution of punishment and cooperation may be investigated using the powerful direct fitness maximization techniques of Taylor and Frank (1996b) and Frank (1998). This allows us to clarify the mechanisms at work and link previous theory to Frank's (1998, 2003a) general framework. In particular, we link kin selection, group selection, and cultural group selection in terms of a generalized view of relatedness. We then reveal that it is not the relatedness between social partners per se that facilitates the evolution of punishing behavior. What is crucial is that there is a positive correlation between the punishment strategy played and cooperation received by an individual. Although such an association could arise from viscous population structure and interactions between kin, it may arise for other reasons. In particular, we demonstrate that even in the absence of relatedness it is possible for such an association, due to facultative adjustment of cooperative behavior, to maintain punishment through selection acting at the level of the individual, rendering group selection and elaborate cultural

practices unnecessary. More generally, the fact that a positive correlation between the behaviors of social partners is the fundamental factor favoring cooperation has been obscured by a focus on how this correlation can be produced by kinship, through the interactions of close relatives (Hamilton 1975; Frank 1998). Our results provide a clear example of how such positive correlations can arise without kin association.

Models and Analyses

Basic Model

We now present a simple model describing the co-evolution of cooperation and punishment. This is intended to elucidate the general selection pressures involved—it is the simplest model that captures the essentials of the problem. We discuss our model in terms of humans because this is where much of the recent theoretical literature has been focused. However, the implications are general and could be applied to a variety of organisms. A role for punishment in the evolution of cooperation has been suggested in a variety of animals, including insects, birds, primates, and other mammals (Clutton-Brock and Parker 1995). We give some specific examples in the discussion when considering how our model may be tested empirically.

For simplicity, we suppose that individuals interact in pairs, with one (random) member of the pair being denoted player 1 and the other player 2. Player 1 may choose to cooperate (e.g., sharing food), in which case she loses fitness c and player 2 gains fitness b , or to defect (e.g., refusing to share food), such that neither player loses nor gains fitness from the interaction. Player 2 may respond to defection in two ways: either she punishes (e.g., by physically injuring player 1) at a cost a to herself in order to reduce player 1's fitness by d , or else she forgives (e.g., does nothing) in which case neither player gains nor loses fitness. The expected direct fitness of a focal individual might then be written as:

$$w = \bar{w} + cx + bX - (1 - X)ya - (1 - x)Yd \quad [6.1]$$

where the constant \bar{w} is baseline fitness, x is the frequency with which that individual cooperates, X is the mean frequency of cooperation among her social partners, y is the frequency with which the individual punishes, given that her partner defects, and Y is the mean punishment strategy played by her social partners, that is, the probability that the focal individual is punished given that she defects. We assume that all competition is global. An important point is that punishment acts to directly reduce both the fitness of the actor and the fitness of her social group. Punishment is therefore fundamentally different from the policing models of Frank (1995b, 1996b, 2003a) because policing directly reduces actor fitness but increases group fitness.

Coevolution of Cooperation and Punishment

We will consider the simultaneous evolutionary optimization of cooperation and punishment analogous to the evolution of policing analysis of Frank (1995b), using the direct (neighbor-modulated) fitness maximization method of Taylor and Frank (1996) and Frank (1998). A small increase in a behavior is favored by selection if the derivative of fitness with respect to that behavior (termed "marginal fitness") is >0 and disfavored when this derivative is <0 . Differentiating the focal individual's fitness function [6.1] with respect to her cooperating (x) and punishing (y) strategies obtains:

$$\frac{dw}{dx} = \bar{w} + Yd + \frac{dX}{dx}(b + ya) - \frac{dy}{dx}(1 - X)a - \frac{dY}{dx}(1 - x)d \quad [6.2A]$$

$$\frac{dw}{dy} = \bar{w} + (1 - X)a - \frac{dY}{dy}(1 - x)d + \frac{dx}{dy}(Yd - c) + \frac{dX}{dy}(b + ya) \quad [6.2B]$$

The terms dX/dx and dY/dy are the coefficients of relatedness, with respect to cooperation and punishment, respectively, between the focal individual and her social partners (Taylor

and Frank 1996; Frank 1998). Technically, the derivative is of the conditional expectation of the social partner's strategy, given the strategy played by the focal individual, with respect to the latter. The other derivative terms are dy/dx and dx/dy , which are the regression of an individual's punishing strategy on its own cooperation strategy, and vice versa, and dY/dx and dX/dy , which are the regressions of a partner's punishing strategy on its own cooperation strategy and a partner's cooperation strategy on its own punishment strategy, respectively.

Let us consider first the origin of cooperation and punishment in a population that is otherwise fixed for defection ($\bar{x} = 0$) and forgiveness ($\bar{y} = 0$). In such circumstances the trait-on-trait regressions are always nonnegative, which is important for interpretation of the analytical results that follows. To see why, consider the regression of cooperation received on cooperation strategy played: $dX/dx = (X - \bar{x})/(x - \bar{x}) \approx X/x$. Since cooperation strategies are nonnegative, the numerator (X) is nonnegative, and since the variant by definition plays a different cooperation strategy from the wild type (which plays zero cooperation), the denominator (x) is positive. Hence, $dX/dx \geq 0$. The same argument can be used to show that this is true for the other trait-on-trait regressions. Assuming only minor variants ($x \approx X \approx \bar{x}$, $y \approx Y \approx \bar{y}$; Taylor and Frank 1996; Frank 1998) and making the substitutions $\bar{x} = 0$ and $\bar{y} = 0$, the marginal fitness with respect to cooperation (equation [6.2A]) reduces to:

$$\frac{dw}{dx} = -c + \frac{dX}{dx}b - \frac{dy}{dx}a - \frac{dY}{dx}d. \quad [6.3]$$

This shows there is a direct cost (c) and a kin-selected benefit ($dX/dx \times b$) of cooperation, plus costs relating to the associated increase in costly punishing ($dy/dx \times a$) and also in being punished ($dY/dx \times d$); see figure 6.1A. Cooperation is maintained even in the absence of punishment when Hamilton's (1964) rule $dX/dx \times b > c$ holds, so we will consider the more interesting situation where it does not, such that $\frac{dw}{dx}$ is always negative.

Similarly, the marginal fitness with respect to punishment (equation [6.2B]) is:

$$\frac{dw}{dy} = -a - \frac{dY}{dy}d - \frac{dx}{dy}c + \frac{dX}{dy}b. \quad [6.4]$$

Again, this is easily understood. Punishing incurs a direct cost (a) and indirect costs ($dY/dy \times d$ from being punished by related individuals and $dx/dy \times c$ from the correlated commitment to cooperation). The benefit $dX/dy \times b$ is gained through the association between the punishment strategy played and the cooperation received (see figure 6.1B). Only when this is sufficiently large may a rare variant with some small frequency of punishing behavior be able to invade. In other words, a positive association between the punishment strategy played and the cooperation received by a focal individual is a necessary but not sufficient condition for the evolutionary origin of punishment.

Result 1. A positive association between punishment strategy played and cooperation received is crucial for the evolutionary origin of punishing behavior.

We will now investigate the evolutionary maintenance of cooperation and punishment by considering $\bar{x} \leq 1$ and $\bar{y} \leq 1$. Again, the trait-on-trait regressions will all be nonnegative: for example, $dX/dx = (X - \bar{x})/(x - \bar{x}) \approx (X - 1)/(x - 1)$. Cooperation received (X) cannot be >1 , so the numerator ($X - 1$) is ≤ 0 . Since the cooperation variant does not play the wild-type strategy (always cooperate) and cannot play a more cooperative strategy than that, the denominator ($x - 1$) is always negative. Hence, $dX/dx \geq 0$. Making the substitutions $\bar{x} \leq 1$ and $\bar{y} \leq 1$, the marginal fitness with respect to cooperation (equation [6.2A]) is now given by:

$$\frac{dw}{dx} = -c + d + \frac{dX}{dx}(b + a). \quad [6.5]$$

Here cooperation carries a direct cost (c) and a benefit (d , due to avoiding punishment) when punishment of defectors is assured. It also gives kin-selected benefits ($dX/dx \times b$ and

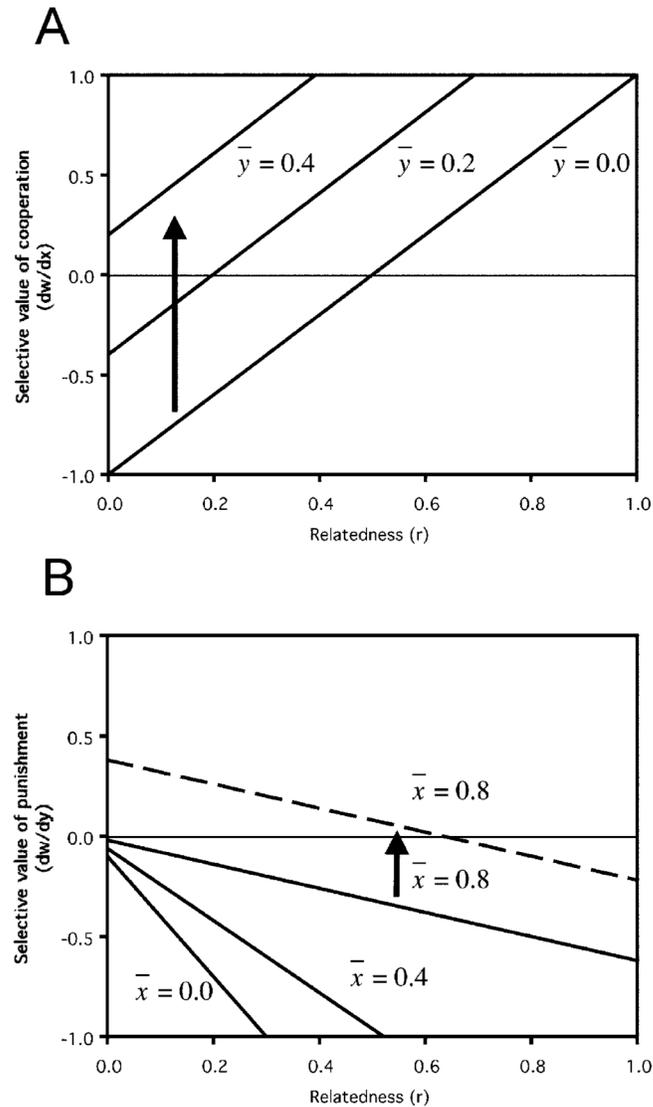


Figure 6.1. (A) Selective value of cooperation (dw/dx) as a function of relatedness and the resident punishing strategy (\bar{y}) when there is no association between traits ($dw/dy = dY/dx = 0$); $dw/dx > 0$ indicates that enhanced cooperation is favoured, and $dw/dx < 0$ indicates that it is disfavoured. Increasing relatedness (r) enhances selection for cooperation; in the absence of punishment, cooperation is favoured when $rb > c$. Increasing punishment also favours cooperation, so cooperation may be favoured even when relatedness is 0, if $\bar{y} > c/d$. (B) Selective value of punishment (dw/dy) as a function of relatedness and the resident cooperation strategy (\bar{x}); $dw/dy > 0$ indicates that enhanced punishment is favoured, and $dw/dy < 0$ indicates that it is disfavoured. Assuming no association between traits ($dx/dy = dX/dy = 0$), we see that punishment is always disfavoured; increased relatedness enhances the selective disadvantage of punishment; and increased cooperation reduces the selective disadvantage of punishment. Punishment may be favourable if there is a positive association between the punishment strategy played and the cooperation received by an individual ($dX/dy > 0$); the broken line indicates $dX/dy = 0.2$. For (A) and (B) we assume $a = 0.1$, $b = 2$, $c = 1$, and $d = 3$.

$dX/dx \times a$) due to the correlated cooperation received from social partners and the fitness saved from not having to punish defectors. Punishment cannot be an effective deterrent when the fitness of a punished defector is greater than that of a cooperator, so that we will restrict attention to the situation $d > c$. Here, the marginal fitness will always be positive, and so selection will act to maintain cooperation. The marginal fitness with respect to punishment (equation [6.2B]) is:

$$\frac{dw}{dy} = [(1 - \bar{x})a - \frac{dY}{dy}(1 - \bar{x})d + \frac{dx}{dy}(d - c) + \frac{dX}{dy}(b + a)]. \quad [6.6]$$

The costs of punishment include the direct cost ($[1 - \bar{x}] \times a$) and the kin-selected cost ($[1 - \bar{x}] \times dY/dy \times d$) plus the cost incurred by the associated cooperation ($dx/dy \times c$). The benefits of punishment are due to the correlated decrease in one's own defection and hence the frequency with which the focal individual is punished ($dx/dy \times d$) and also the correlated increase in cooperation received from social partners ($dX/dy \times b$) and, conversely, the fitness saved by not having to punish partners ($dX/dy \times a$). If $dx/dy = dX/dy = 0$ so that there is no correlation between the punishment and cooperation played by an individual, nor between the punishment played and cooperation received, then the marginal fitness with respect to punishment is small but negative, and hence full punishment is not stable. It is interesting to note that relatedness dY/dy works to undermine the stability of punishment; as an individual's punishment strategy is increased, so too is the punishment received from social partners. If the between-trait associations are positive and of sufficient magnitude, then full punishment can be evolutionarily stable. Otherwise, selection will act to reduce punishment in the population.

Result 2. A positive association between punishment strategy played and cooperation received is crucial for the evolutionary maintenance of punishing behavior.

We now check to see whether punishment is easier to maintain than it is to initially invade an otherwise forgiving population, by evaluating $dw/dy|_{\bar{x}=\bar{y}=1} - dw/dy|_{\bar{x}=\bar{y}=0}$, that is, subtracting the right-hand side (RHS) of equation [6.4] from the RHS of equation [6.6] to

obtain:

$$a \frac{dY}{dy} + \frac{dx}{dy} + a + \frac{dX}{dy}, \quad [6.7]$$

which is positive, so that the RHS of equation [6.4] is less than the RHS of equation [6.6], and hence the condition for increased punishment to be favored ($dw/dy > 0$) is more easily satisfied in a population of cooperators and punishers than in a population of defectors and forgivers. Similarly, the RHS of equation [6.3] is always negative under the relevant circumstances (i.e., when $dX/dx \times b < c$), and the RHS of equation [6.5] is always positive, so that the condition for enhanced cooperation to be favored ($dw/dx > 0$) is also more easily satisfied in punishing populations than in populations rife with defection and forgiveness.

Result 3. Punishing behavior is more easily maintained than it is originally evolved. Note that this assumes that relatedness and the between-trait regressions are constants. A fully dynamic analysis relaxing this assumption would require that we specify a more detailed (and hence less general) model and so is not pursued here because we aim only to abstract and elucidate the selection pressures involved in the evolution of punishment and cooperation.

Example: Cooperation as a Facultative Response to Punishment

The Model. We have found that relatedness between social partners is not crucial for costly punishment to be favored (indeed, increasing relatedness disfavors punishment) and that it is another association, the regression of the cooperation received on the punishment strategy played, that provides the benefit of punishment. To illustrate these findings, we examine the evolution of punishment when there is no relatedness between individuals ($dY/dy = 0$) and when cooperation is facultatively adjusted to one's punishment environment (which we will see can give $dX/dy > 0$).

We assume that individuals are randomly organized into social groups of size N , such that relatedness between group members is 0. In each social encounter, individuals pair with a random member from their group, with one of the partners playing the role of player 1 and the other being player 2. In contrast with the previous model, we consider the cooperation strategy of player 1 to be facultative and hence a function of her punishment environment. Assuming no partner recognition and therefore no adjustment of cooperation to her current partner's punishment strategy, the cooperation strategy played by the focal individual (in half of her social interactions) is expressed as a function of the average punishment strategy played by all of her social partners: $x = f[\bar{y}]$. Since each of her social partners experiences a punishing environment that includes the focal individual (and hence average punishment strategy among their social partners is $\bar{y} + [y - \bar{y}]/[N - 1]$), they will play cooperation strategy $X = f[\bar{y} + (y - \bar{y})/(N - 1)]$.

If individuals cooperate optimally, we expect the function $f[Y]$ to be such that it maximizes the fitness of player 1 when player 2 plays punishment strategy Y . It is easy to show that this optimum is given by:

$$f^*[Y] = \begin{cases} 0 & c > Yd \\ 1 & c < Yd \end{cases} \quad \text{if } \quad , \quad [6.8]$$

such that defection is favored when the cost of cooperation outweighs the threat of punishment ($c > Yd$), and cooperation is favored when the cost of cooperation is outweighed by the threat of punishment ($c < Yd$). This step function is both mathematically inconvenient and biologically unreasonable, so we will use the model of McNamara et al. (1997; see also Kokko 2003) to describe nearly optimized cooperation as:

$$f[Y] = \frac{1}{1 + \exp[\beta(c - Yd)]} = \frac{1}{1 + \exp[\beta(Yd - c)]} , \quad [6.9]$$

where ϵ is the degree of behavioral error and $\epsilon = dw/dx = Yd - c$ ensures that the frequency of non-optimal behavior declines as its impact on fitness becomes more important. The facultative cooperation function (equation [9]) approaches the step function (equation [6.8]) for vanishing behavioral error ($\epsilon \rightarrow 0$), and for larger error ($\epsilon > 0$), it takes a continuous sigmoidal form which flattens out to a constant 1/2 as the error tends to infinity (figure 6.2). For mathematical convenience, we will assume vanishing (but nonzero) behavioral error ($\epsilon \rightarrow 0$).

Altering fitness function (equation [6.1]) for this example model, we have the fitness of an individual who plays punishment strategy y , in a population with mean punishment strategy \bar{y} , given by:

$$w = \epsilon cf[\bar{y}] + bf\left[\frac{y}{\bar{y}}\right] + \frac{y - \bar{y}}{N - 1} \left[a(1 - f\left[\frac{y}{\bar{y}}\right]) + \frac{y - \bar{y}}{N - 1} y \epsilon d(1 - f[\bar{y}]) \bar{y} \right]. \quad [6.10]$$

The mean fitness of the population is:

$$\bar{w} = \epsilon cf[\bar{y}] + bf[\bar{y}] + a(1 - f[\bar{y}]) \bar{y} + d(1 - f[\bar{y}]) \bar{y}, \quad [6.11]$$

so we expect a rare variant playing punishment strategy y to increase in frequency in a population with mean punishment strategy \bar{y} when the fitness differential $\Delta w = w - \bar{w}$ is positive, that is, when:

$$\Delta w = b \left[f\left[\frac{y}{\bar{y}}\right] - f[\bar{y}] \right] + \frac{y - \bar{y}}{N - 1} \left[\left(a(1 - f\left[\frac{y}{\bar{y}}\right]) - a(1 - f[\bar{y}]) \right) \bar{y} + \frac{y - \bar{y}}{N - 1} y \epsilon d(1 - f[\bar{y}]) \bar{y} \right] > 0. \quad [6.12]$$

Origin of Punishment. We first consider the evolutionary stability (Maynard Smith & Price 1973) of forgiveness, by determining under what circumstances no variant with punishment strategy $y > 0$ can invade a population with mean punishment strategy $\bar{y} \rightarrow 0$.

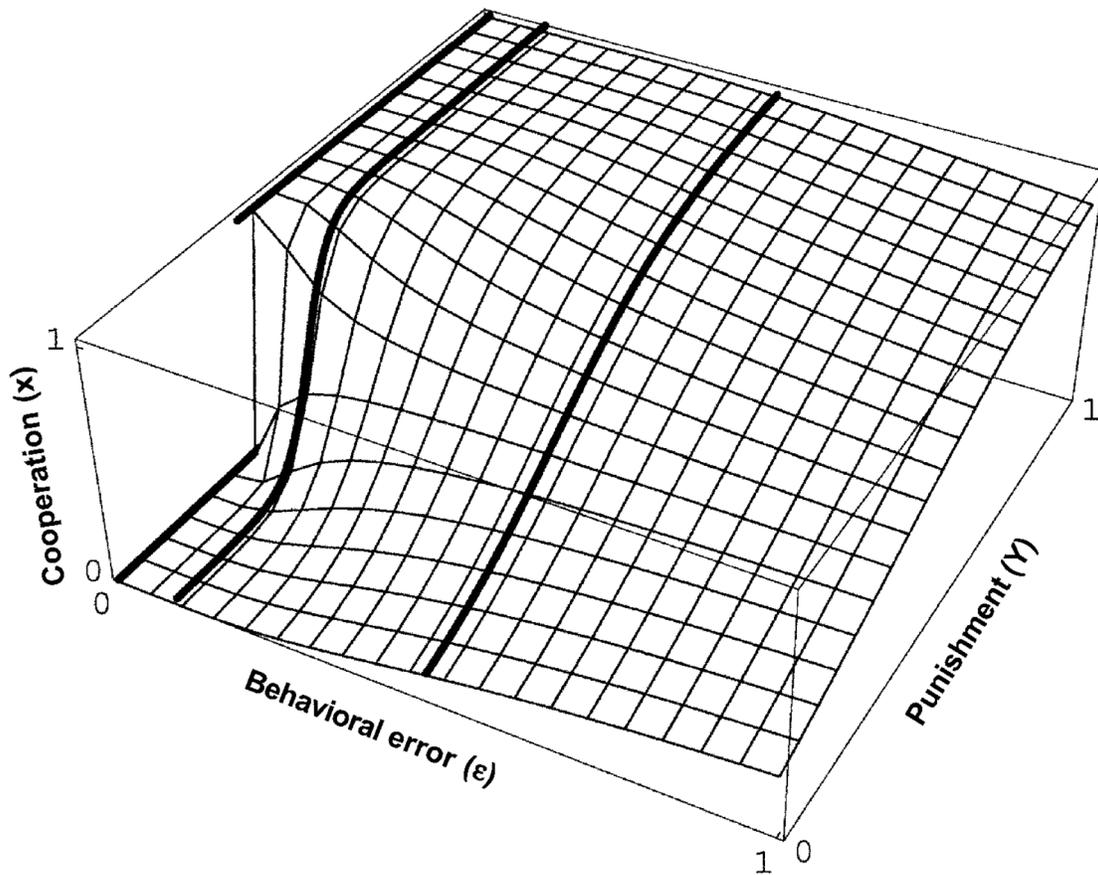


Figure 6.2. Frequency with which an individual cooperates (x) as a function of the punishment strategy of its social partners (Y) and the degree of behavioural error (e), according to the example facultative model. Values are obtained numerically, assuming $c = 1$ and $d = 3$. The bold lines indicates $\beta = 0.01, 0.10, \text{ and } 0.50$.

Substituting the cooperation function (equation [6.9]) into the fitness differential (equation [6.12]) obtains:

$$\Delta w = b \frac{1}{1 + \exp\left[\frac{c - (y/(N - 1))d}{\beta}\right]} - \frac{1}{1 + \exp\left[\frac{c}{\beta}\right]} - a \frac{1}{1 + \exp\left[\frac{c - (y/(N - 1))d}{\beta}\right]} \quad [6.13]$$

Recalling that the behavioral error is vanishingly small ($e \rightarrow 0$), we find that when the threat of punishment posed to social partners of the punishing variant is less than the cost

of cooperation ($(yd)/(N - 1) < c$), then equation [6.13] reduces to $-ya$, which is negative, and hence the rare variant cannot invade. This is because defection is the rule in the social groups of both the wild type and the variant, giving population mean fitness $\bar{w} \approx \square$ and rare variant fitness $w \approx \square - ya$. When the threat of punishment is greater than the cost of cooperation ($(yd)/(N - 1) > c$), then equation [6.13] reduces to b , which is positive, and hence the rare variant can invade. Here, the rare punisher has managed to push her social group over the punishment threshold such that cooperation is now the optimal strategy. The average social group is fully defecting, so $\bar{w} \approx \square$, but the rare variant is now a recipient of cooperative behavior and only rarely encounters a defector requiring punishment, so that her fitness is $w \approx \square + b$. Note that although the variant receives cooperation, she maximizes her fitness by always defecting (since her unrelated social partners are all forgivers) and hence pays no cost of cooperation. If no y satisfies the above invasion condition, then forgiveness is an evolutionarily stable strategy (ESS; Maynard Smith & Price 1973). This is assured when $(N - 1)c > d$, so that not even a fully punishing variant ($y = 1$) can invade. Evolutionary stability of forgiveness is therefore assured unless

$$d > (N - 1)c. \tag{6.14}$$

Result 4. In the above model, punishment is unlikely to invade forgiveness unless the population is structured into very small groups.

Maintenance of Punishment. To determine whether punishment is an ESS, we let the wild type adopt the strategy of full punishment ($\bar{y} = 1$) and consider the success of rare variants playing $y < 1$. Substituting the facultative cooperation function (equation [6.9]) into the fitness differential (equation [6.12]) obtains:

$$\begin{aligned} \Delta w = & b \frac{1}{1 + \exp[(c - (1 - y)/(N - 1)d)/\square]} - \frac{1}{1 + \exp[(c - d)/\square]} \\ & + a \frac{1}{1 + \exp[(c - d)/\square]} - \frac{1}{1 + \exp[(c - (1 - y)/(N - 1)d)/\square]} y. \end{aligned} \tag{6.15}$$

First consider "ineffective punishment" ($c > d$). When behavioral error is vanishing ($\epsilon \rightarrow 0$), the fitness differential (equation [6.15]) reduces to $a(1 - y)$, which is positive, and hence the more forgiving variant will always invade. This is because even when defection is always met with punishment, the defector has greater fitness than the cooperator, so that in all social groups defection is rife. The resident strategy incurs the cost of full punishment, and so the mean fitness of the population is $\bar{w} \approx \epsilon - a$, whereas the more forgiving variant avoids this at least part of the time, giving fitness $w \approx \epsilon - ya$. Now consider "effective punishment" ($d > c$), such that punished defectors receive lower fitness than cooperators. The resident now enjoys the benefits of cooperation and only infrequently encounters erroneous defection requiring punishment. If the rare variant forgives to such a degree that her social partners optimize by defection; that is, when $c - (1 - (1 - y)/(N - 1))d > 0$, the fitness differential (equation [6.15]) reduces to $-(b + ya)$ since she loses the benefits of cooperation and punishes a proportion y of her social partners. This is negative, and so the rare variant cannot invade. If the variant's forgiveness is not sufficient to warrant a switch to defection among her social partners, equation [6.15] becomes $-(b + ya) \exp\{c - [1 - (1 - y)/(N - 1)]d\}$, which is vanishingly small but nevertheless negative, and hence the rare variant cannot invade. This is true because with vanishing behavioral error ($\epsilon \rightarrow 0$) the frequency of defection in the fully punishing group is a vanishing fraction of the frequency of defection in the more forgiving group, so that the fitness saved from not punishing so frequently does not outweigh the fitness lost through the reduction of received cooperation. Relaxation of the infinitesimal error assumption (figure 6.3) shows that this result is robust, even for large social groups. The variant can therefore only invade an otherwise fully punishing population when punishment is ineffective, so that punishment is an ESS when:

$$d > c . \quad [6.16]$$

Result 5. In the above model, punishment is maintained by selection once it has become common if the cost of cooperation (c) is less than the cost of being punished (d).

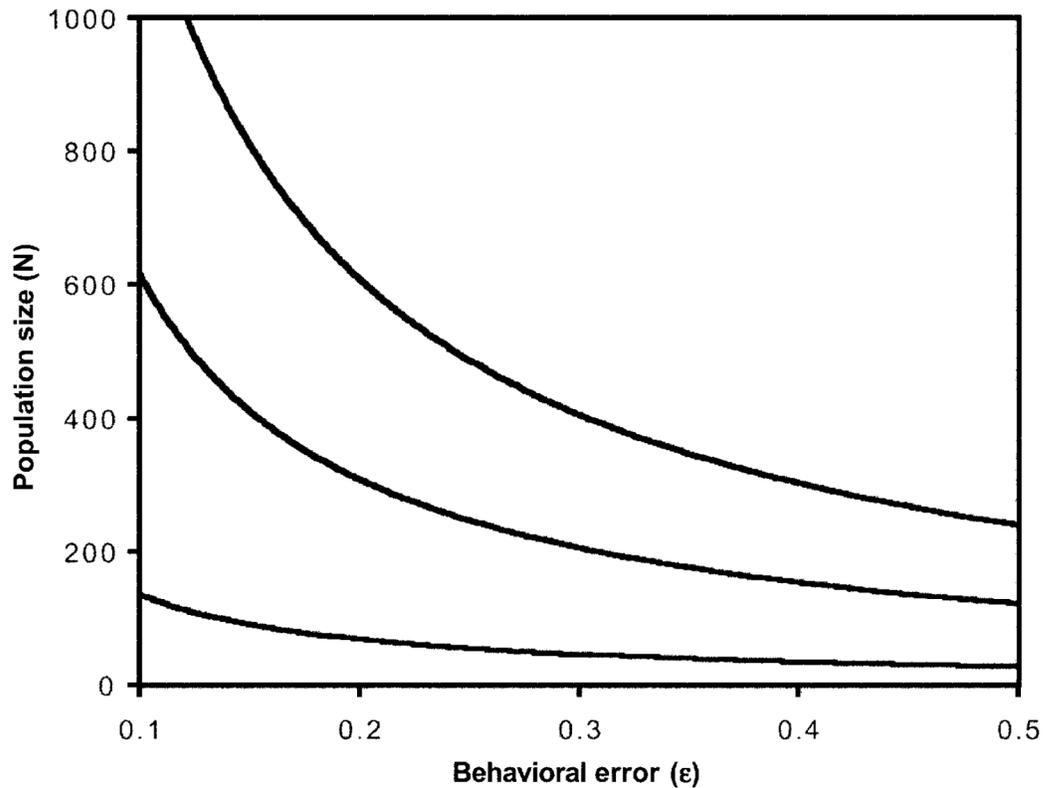


Figure 6.3. Maximum group size (N) permitting the evolutionary stability of punishment ($\bar{y} = 1$) as a function of behavioural error (ϵ) and the cost of punishing (a), according to the example facultative model, assuming $b = 2$, $c = 1$, and $d = 3$. Upper line, $a = 0.01$; middle line $a = 0.10$; bottom line, $a = 0.50$.

Discussion

Punishment and Cooperation

We have shown that full punishment can be an evolutionarily stable strategy only if there is a positive association between the punishment played and the cooperation received by an individual. This could arise if populations are viscous so that social partners tend to be genealogical relatives, but other mechanisms are possible, for example, when individuals facultatively adjust their level of cooperation in response to the local threat of punishment. We have also provided analytical support for the suggestion of Boyd et al. (2003) that the

cost of punishment declines as it becomes common in the population and hence punishing behavior might be maintained more easily than it is initially evolved.

These results suggest three general implications. First, it can be easier for some cooperation to evolve by another mechanism (e.g., altruism between relatives) and then punishment evolve to favor and maintain higher levels of cooperation. An analogous conclusion has been made for some other mechanisms that do not rely on interactions between relatives, such as group augmentation (Kokko et al. 2001; Griffin and West 2002). Second, within the specific context of explaining human cooperation, punishment could have evolved at a time when social structure was more conducive to punishment (small groups of interacting individuals). Once common, punishment could be retained even when interactions began to occur within much larger groups of humans. Third, the opposite frequency dependence is true for systems based on rewarding cooperation rather than punishing defection—the cost of rewarding escalates as more individuals cooperate, whereas we have shown the cost of punishing decreases as more individuals cooperate. This might go some way to explaining why punishment as opposed to rewarding is prevalent in nature (e.g., Clutton-Brock and Parker 1995).

How can our model be tested? Our major result is that costly punishment can be favored if there is a positive association between the punishment played and the cooperation received by an individual (results 1 and 2). This could be hard to test directly, especially experimentally, because of limitations on how an individual's level of punishment could be manipulated. However, some of the fundamental assumptions and predictions of our model that underly this result could be tested more easily. In particular, are lower levels of cooperation more likely to lead to punishment, as appears to occur in superb fairy wrens (Mulder and Langmore 1993), naked mole rats (Reeve 1992), and *Polistes* wasps (Reeve and Gamboa 1987)? Second, are individuals more likely to cooperate when they are punished, as may occur in *Polistes* wasps (Reeve and Gamboa 1987)? Third, do individuals try to signal that they cooperate more than they actually do, as occurs in white-winged choughs (Boland et al. 1997)? Fourth, do systems in which social partners are more related tend to display less punishment, analogous with Frank's (1995, 2003) result that investment into policing correlates negatively with relatedness?

Relatedness and Kin Selection

This analysis has made use of the understanding that the coefficient of relatedness appropriate to the direct fitness formulation of Hamilton's rule is a regression measure describing the association between actor and social partner phenotypes (reviewed by Seger 1981; Michod 1982; Grafen 1985a; Queller 1985; 1992; Frank 1998). Such associations are generally due to genealogical closeness and hence genetic similarity, so that the maximization of neighbor-modulated or inclusive fitness is popularly referred to as "kin selection" (Maynard Smith 1964). Group selection can be responsible for the evolution of an altruistic trait only insofar as the benefit to the group is large enough, the cost to the individual is low enough, and there is substantial between-group as opposed to within-group variation in trait values. Since the proportion of the total variance that is attributable to between-group differences is the coefficient of relatedness appropriate for whole-group traits, Hamilton's rule can be used to predict when group selection will favor the trait (i.e., when $\text{relatedness} \times \text{benefit} > \text{cost}$). Thus, kin selection and group selection are mathematically equivalent ways of conceptualizing the same evolutionary process, a point that previously has been analyzed in much detail (Price 1972a; Hamilton 1975; Wade 1985; Frank 1986, 1998; Queller 1992; Reeve and Keller 1999). Consequently, it is puzzling that kin selection has been largely ignored in the human altruistic punishment literature on the grounds that relatedness is too low, while group selection has often been regarded as important (e.g., Gintis 2000). Furthermore, because relatedness is a regression of recipient phenotype on actor phenotype, it transcends genetics and applies even when the cause of phenotypic similarity is simply imitation, for example, as in the cultural group selection proposed by Heinrich and Boyd (2001). In this sense, "kin selection" is something of a misnomer because it draws attention to only one cause of the statistical association that is relatedness, as Hamilton (1975) realized.

As this analysis has shown, positive relatedness is not really the key ingredient for the evolutionary success of punishment. Punishing behavior is costly to the individual and protects the social group from the breakdown of cooperation, and hence it has been described as a form of altruism (Sober and Wilson 1998). It might then be expected that where it is successful, altruistic punishment is being maintained by kin selection.

However, punishment is quite a different form of public good from cooperation—it is directly disadvantageous at the group level because it reduces the fitness of the focal individual and her social partners. The benefit it brings is indirect because it merely creates a coercive social environment in which cooperation is favored. It therefore differs from Frank's (1995b, 1996b, 2003a) recent models of competition-repression in which investment into policing behavior translates directly into enhanced group fitness. In our model, punishment is only of selective value when there is a sufficiently strong correlation between punishment strategy played and cooperation received (dX/dy ; figure 6.1B). This highlights a fundamental nonequivalence of first- and higher-order public goods.

A positive correlation between punishment played and cooperation received might arise in a viscous population where genealogical kin tend to associate with each other, so that the social partners of punishers are also punishers ($dY/dy > 0$) and therefore punishers are expected to be coerced into cooperating more than forgivers ($dx/dy > 0$). This association combines with relatedness to ensure that an increase in punishing behavior is associated with an increase in the amount of cooperation received ($dX/dy > 0$). The pressure for enhanced punishment is therefore not strictly kin selection but rather something more akin to "niche construction" (Odling-Smee et al. 1996), in the sense that the behavior modifies the social environment in such a way as to alter the selective pressures acting upon other traits. It is worth noting that localized competition in viscous populations adds extra complexity to models of kin selection (see Taylor 1992a, 1992b; Wilson et al. 1992; Queller 1994; Frank 1998; Griffin and West 2002; West et al. 2002a; Gardner and West (in press) for extensive discussion of its impact on the evolution of social behaviors). In our analysis, we have assumed that all competition occurs at the level of the whole population, and we leave local competition as an open problem for the future.

We may easily demonstrate that relatedness is not necessary for the evolution of costly punishment by considering mechanisms that generate positive associations between the punishment played and the cooperation received despite zero relatedness, for example, the facultative model of cooperation introduced above. We discovered that in the absence of relatedness, partner recognition, reputation, and any mechanism whereby an individual may bias her interactions or tailor her behavior in response to her immediate social

partner, punishment might be maintained by selection acting directly at the level of the individual. This is because when punishment is already frequent, the fitness saved by forgiving is minimal and may be overwhelmed by the concomitant decline in the amount of cooperation received because of the decrease in selection for cooperation among social partners. This example model is intended for illustration only and is designed to demonstrate how a net benefit for punishment might be achieved even when individuals do not interact with relatives. More complicated scenarios are therefore possible, and of particular interest is the effect of enhanced behavioral error (increasing ϵ). Numerical analysis of the example model reveals that increasing the frequency of maladaptive behavior reduces the likelihood that individual level selection will be able to maintain altruistic punishment in very large groups (figure 6.3), although the results presented above are qualitatively robust so long as behavioral error (ϵ) and the cost of punishing (a) are small. The degree to which individuals are expected to behave optimally is contentious, but punishment is indeed characterized by its cheapness (Sober & Wilson 1998).

Conclusion

We have given analytical support to the suggestion that the cost of punishment declines as it becomes a common strategy, so that punishment is more easily maintained than it is originally evolved. We showed that it is not relatedness per se that is important in ensuring that punishing behavior enhances fitness but rather that a positive correlation between punishment played and cooperation received by an individual is crucial. We also revealed that facultative adjustment of cooperation can give rise to such a positive association even in the absence of relatedness between social partners. Finally, we demonstrated that the direct benefits accrued when cooperation is facultative may be large enough for selection acting at the individual level alone to maintain punishment among humans, rendering elaborate population dynamics and cultural practices unnecessary. More generally, our results provide a specific example of how positive correlations between the behaviors played by social partners can arise and favor cooperation for reasons other than kinship. Major tasks for the future include clarifying

the links between punishment and reproductive skew theory (Johnstone 2000; Clutton-Brock et al. 2001; Langer et al. 2004) and developing more specific models for specific situations or organisms.

7. Social evolutionary multi-locus methodology

Abstract

In general, social evolution theory is concerned with correlations between individuals. When co-evolution of multiple social traits is studied, associations between these traits are typically ignored. This is analogous to the assumption of linkage equilibrium in population genetics theory. It is appreciated that allowing for the evolution of linkage disequilibrium can qualitatively change the behaviour of an evolving system. This has prompted the development of a general multi-locus notation and corresponding methodology. Currently, a general methodology for describing co-evolution of social traits is lacking, despite recent interest in such models. Analyses which have allowed for between-trait associations have done so at the expense of dynamic sufficiency. We develop the multi-locus methodology by allowing for genetic associations between as well as within individuals, and relate this to the theoretical foundations of social evolution. In the process, we highlight the subtlety of Price's theorem and Hamilton's rule. The methodology also provides a general framework for building dynamically sufficient models of social evolution that allow for associations between traits. We illustrate these developments by application of the methodology to the co-evolution of cooperation and punishment in humans.

Introduction

Although social evolution theory is fundamentally concerned with associations between individuals, analyses of the co-evolutionary dynamics of social traits are typically made tractable by assuming statistical independence between the traits. For example, Frank (1995b, 1996b, 2003a) describes the co-evolutionary dynamics of competitiveness and policing behaviour under the assumption that there is no association between these traits within individuals. Such independence is analogous to the assumption of linkage

equilibrium in population genetic multi-locus models. It has long been understood by population geneticists that this assumption can lead to qualitatively misleading predictions. Recent co-evolutionary social evolution analyses have highlighted the importance of statistical associations between different traits in different social partners, for example in the evolution of costly punishment (Gardner & West 2004a; chapter 6) and cooperation based on systems of arbitrary markers (Axelrod et al. 2004). In both studies, allowing for associations between these traits carried the cost of losing dynamic sufficiency. A general methodology for dealing with such social co-evolutionary problems, particularly one in which dynamic sufficiency is restored, is currently lacking.

Barton and Turelli (1991) and Kirkpatrick et al. (2002) have developed a general methodology for describing evolutionary change at multiple gene positions, for arbitrary ploidy, dominance, epistasis, transmission rules and lifecycles. Central to this methodology is the description of the genetical composition of a population in terms of associations (generalised from the traditional conception of linkage disequilibrium) between gene positions. The methodology is of such generality that it implicitly allows for associations between individuals, and so we may add ‘arbitrary social interactions’ to the above list. The purpose of this paper is to make explicit the social evolutionary aspects of this methodology, to relate this to the foundations of social evolution theory (and in doing so dispelling some misconceptions), and to show that methodology can be used as a general tool for conducting dynamically sufficient analyses of social co-evolutionary problems.

In order to guide the reader through this chapter, in this section we will provide a brief summary of the following sections. In the next section we introduce the multi-locus methodology and explain why it is needed in the study of population genetics. This involves an introduction to the notation, which is summarised in table 7.1, and a description of evolutionary change due to selection and transmission. The section closes with an explanation of how the assumption of quasi-linkage equilibrium (QLE) uses the multi-locus methodology to greatly simplify multi locus problems.

Symbol	Definition
z	A phenotypic value
w	Fitness (a special case of z)
\mathbf{i}	A generic gene position
X_i	Arbitrary allelic value at an instance of \mathbf{i}
ϕ_i	Arbitrary reference value for \mathbf{i}
\square_i	Allelic deviation ($\square_i = X_i - \phi_i$)
\mathbf{A}, \mathbf{U} or \mathbf{V}	An arbitrary set of gene positions
\mathbf{G}	Set of all gene positions contributing to phenotype
\mathbf{W}	Set of all gene positions contributing to fitness
$\square_{\mathbf{A}}$	Allelic deviation for a set of gene positions \mathbf{A} ($\square_{\mathbf{A}} = \sum_{\mathbf{i} \in \mathbf{A}} \square_i$)
$D_{\mathbf{A}}$	Association gene positions in \mathbf{A} ($D_{\mathbf{A}} = E[\square_{\mathbf{A}}]$); the general definition of linkage disequilibrium
$\square_{\mathbf{A}}$	Partial regression of phenotype (z) on the association ($D_{\mathbf{A}}$) for set \mathbf{A} ; a genotype \square phenotype map
$a_{\mathbf{A}}$	As $\square_{\mathbf{A}}$, for special case where $z = w/\bar{w}$; the multi-locus selection coefficient
$t_{\mathbf{U} \square \mathbf{A}}$	Transmission coefficient; the probability that set of positions \mathbf{A} after transmission derived from set \mathbf{U} before transmission

We then introduce the reader to the foundations of social evolution theory, namely Price's (1970) theorem and special cases – Robertson's (1966) secondary theorem of natural selection, Fisher's (1930) fundamental theorem of natural selection, and Hamilton's (1963, 1964, 1970) rule. The multi-locus statements of change due to selection and transmission are related to the corresponding framework of Price. Hamilton's rule is derived from multi-locus considerations, illustrating an analogy between relatedness and linkage disequilibrium. Extensions of Hamilton's rule incorporating non-additivity of fitness components (synergy; Queller 1985) are considered. A general restatement of Hamilton's rule which makes explicit all predictors (and associations between predictors) is given.

Following this, we introduce the problem of costly punishment, which has received much attention in connection with human behaviour, and has been explored recently by Gardner and West (2004a, chapter 6). We work through Gardner and West's simple model, employing the social evolutionary multi-locus techniques, recovering and

strengthening their results. In particular, the model is made dynamically sufficient, yet rendered tractable using the theoretical developments of this chapter.

Finally, there follows a discussion of what has been achieved in extending the multi-locus methodology to social evolution theory. We also point out some potentially interesting extensions for the future.

Multilocus methodology

Why have a multilocus methodology?

The formal basis of evolutionary theory rests in population genetics (Queller 1984; Grafen 2002). This is the study of purely mechanical population processes such as natural selection, mutation, migration and random drift (Crow & Kimura 1970). Proper predictions of the course of evolutionary change requires a full description of population composition at a given time step. The number of distinct genotypes increases exponentially with the number of loci, and so multi-locus analyses, whether analytical or simulation based, can be overwhelming and intractable. A common simplifying approach (e.g. Haldane 1964) is to assume statistical independence between loci (“linkage equilibrium”), so that the large number of genotype frequencies can be reconstructed from a smaller number of gene frequencies. Yet statistical associations between loci (“linkage disequilibria”) cannot in general be ignored, as they will often be created by population processes, and it is appreciated that indirect selection caused by direct selection on linked loci can dramatically alter the course of evolution. Such indirect “hitch-hiking” effects are essential for understanding the evolution of sex and recombination, the evolution of female mate preferences, gene flow through hybrid zones, how hitch-hiking impacts on patterns of genetic diversity, adaptive arguments for the evolution of dominance, and much more (see review by Barton 2000, and references therein). It is therefore essential that we follow the frequencies of each of the distinct genotypes.

Expressing evolutionary change in terms of changes in genotype frequencies can obscure the dynamics of quantities that are of more immediate interest, for example gene frequencies and population mean trait values. An alternative is to follow the gene frequencies and all linkage disequilibria, which is the approach adopted by the multi-locus methodology of Barton and Turelli (1991) and Kirkpatrick et al. (2002), and independently developed by Christiansen (1999). This approach involves tracking exactly the same number of evolving variables as if we were following genotype frequencies, but it lends itself to a quantitative genetic approach which neatly and naturally partitions the various causes of evolutionary change. It also leads the way for a powerful simplifying assumption, that of “quasi-linkage equilibrium” (Kimura 1965, Nagylaki 1993), which reduces the multi-locus problem to the same degree of complexity as the assumption of linkage equilibrium, yet retains a great deal more realism.

The notation

The power of the multi-locus methodology lies in its generality, but this can make discussion of the interpretation of the notation somewhat confusing. To aid the reader, we summarise the key notation in table 7.1. The following excursion into the notation will highlight only the features which are of most immediate interest to the aims of this paper – that is, making the possibilities for modelling social evolution explicit. For a comprehensive but exquisitely readable account, the interested reader is directed to Kirkpatrick et al. (2002).

The Barton-Turelli approach describes the genetical composition of the population in terms of the allelic values at the various *positions* where genes can reside and also the *associations* between these positions. For example, a model involving haploids with two biallelic loci might involve a separate position for each locus (and associated allelic values, or allele frequencies, for each) and also a term describing the linkage disequilibrium between the two loci. But positions are not synonymous with loci – for instance, in a diploid context with genomic imprinting, it may be necessary to distinguish the maternal and paternal instances of the same locus as two separate positions. Thus, we

might more correctly describe this as a multi-position, as opposed to multi-locus, methodology. Alternatively, ‘multi-locus’ is correct, so long as we understand ‘locus’ to simply mean ‘gene position’.

Of interest to us are the deviations ($\square_i = X_i - \wp_i$) of a gene’s allelic value (X_i) from some arbitrary reference value (\wp_i) for a given position (**i**). It will often be convenient to define the reference value as the average allelic value ($\wp_i = \bar{X}_i$) for that position, so that the allelic deviations are simply deviations from the average. Associations over sets of positions are described in terms of the average product of these deviations, for example the association between locus **i** and locus **j** is $D_{ij} = E[\square_{ij}] = E[\square_i \square_j] = E[(X_i - \wp_i)(X_j - \wp_j)]$ which, if we define the reference values as average values for that position, is the standard covariance definition for linkage disequilibrium ($D_{ij} = E[X_i X_j] - E[X_i]E[X_j]$; Lewontin 1974). However, since this approach allows us to generalise the concept of associations beyond linkage disequilibrium, we can talk of associations between any positions and not simply loci. Also, the flexible notation allows us to easily define the association between three or more positions – for example $D_{ijk} = E[\square_{ijk}] = E[\square_i \square_j \square_k]$ – and also for a single position. When the reference value is the average allelic value for the position, then the association at a single position is zero ($D_i = E[\square_i] = E[X_i - \wp_i] = E[X_i - \bar{X}_i] = 0$). Finally, there is no reason why associated positions should not be resident in different individuals. The major aim of this paper is to expand upon this crucial point, and to forge conceptual links between the understanding of population genetic associations and the social evolutionary concept of relatedness.

Once we have defined the genetical composition of individuals and the population in general, we can describe phenotypes. A phenotype (z) is defined as:

$$z = \bar{z} + \sum_{A \in G} \square_A (\square_A \square D_A) + \square_c, \quad [7.1]$$

where \bar{z} is the mean phenotypic value for the population, **G** is the set of all positions which contribute to the phenotype, \square_k is the partial regression of phenotype on the

deviation term (\bar{D}_A) for the set of loci \mathbf{A} (holding all other deviations fixed), and \bar{D} is the uncorrelated error. A special case of particular interest is when the phenotype of interest is fitness itself. We may express relative fitness as:

$$\frac{w}{\bar{w}} = 1 + \sum_{\mathbf{A} \subseteq \mathbf{W}} a_{\mathbf{A}} (\bar{D}_{\mathbf{A}} - \bar{D}_{\mathbf{A}}) + \bar{D}_v \quad [7.2]$$

Where the $a_{\mathbf{A}}$ terms are the multilocus methodology's generalised selection coefficients, and may be described as the partial regressions (i.e. holding all other associations constant) of relative fitness on the deviation ($\bar{D}_{\mathbf{A}}$) for a particular set of positions (\mathbf{A}), and \mathbf{W} is the set of all positions contributing to fitness. Barton and Turelli (1991) use these definitions to generate expressions for the change in associations, which we will summarise in the next subsection.

Describing changes in associations

The change in an association due to selection is:

$$\Delta_S \bar{D}_{\mathbf{A}} = \sum_{\mathbf{U} \subseteq \mathbf{W}} a_{\mathbf{U}} (\bar{D}_{\mathbf{U}\mathbf{A}} - \bar{D}_{\mathbf{U}} \bar{D}_{\mathbf{A}}) \quad [7.3]$$

(Barton & Turelli 1991, Kirkpatrick et al. 2002). We will derive this expression from Price's theorem in the next section. Note that this also allows us to describe the change in average allelic values at a single position. Since $\bar{D}_i = \bar{X}_i$, setting reference values to position averages ($\bar{D}_i = \bar{X}_i$), we have:

$$\Delta_S \bar{X}_i = \sum_{\mathbf{U} \subseteq \mathbf{W}} a_{\mathbf{U}} \bar{D}_{\mathbf{U}i} \quad [7.4]$$

If we are concerned with biallelic loci with allelic values $X_i = 1$ at frequency p_i and $X_i = 0$ at frequency $q_i = 1 - p_i$, then the right hand side of equation [7.4] is also the change in allele frequency p_i . Change due to transmission is defined as:

$$\Delta_T D_A = \sum_U t_{U \rightarrow A} D_U - D_A \quad [7.5]$$

(Kirkpatrick et al. 2002), where the $t_{U \rightarrow A}$ coefficients represent the probability that the set of positions A were drawn from source set of position U during the transmission event. Again, we will derive this in the next section, from Price's theorem. Note that, analogous to the derivation of expression [7.4] from [7.3], we can use expression [7.5] to describe the change in the average allelic value / allele frequency at a position, due to transmission.

It is important to note that reference values (φ_i) are not automatically updated during the selection or transmission event, so that if we used the average allelic values (\bar{X}_i) as reference values before the event, the associations after the event (D_A) are still expressed in terms of deviations from the average allelic values before the event. In order to re-express these in terms of deviations about the current average value, we need to update reference values, as outlined by Kirkpatrick et al. (2002). Kirkpatrick et al. also describe how deterministic population processes such as mutation and migration can be incorporated into the above scheme, so that the selection and transmission expressions are sufficiently general to describe such changes. However, this is not of immediate interest to this paper, and so will not be considered further.

Quasi-linkage equilibrium (QLE)

Kimura (1965) revealed that multi-locus systems often rapidly settle into a state such that linkage disequilibria terms (measured in a particular way, which allows them to be independent of allele frequencies) become virtually constant. He referred to this state as quasi-linkage equilibrium (QLE; see also Nagylaki 1993). As mentioned previously, the

complexity of the multi-locus analysis can be made much simpler if we make the assumption that QLE has been reached. Indeed, the possibility of using this gambit provided the motivation for developing the general multi-locus framework, which is specifically geared towards facilitating the QLE assumption. Essentially, we consider that the linkage disequilibria (D_A) evolve over a much faster time scale than the allele frequencies (p_i). By separating the time scales we may set all the linkage disequilibrium terms to their equilibrium values for a given set of allele frequencies, and from there determine how the set of allele frequencies changes from one time step to the next. Thus, the linkage disequilibria are implicit, but are not ignored. The QLE assumption is therefore more valid than simply assuming that all the loci are statistically independent, but manages to reduce the problem to the same level of simplicity. The approximation achieved by the QLE is expected to be accurate when selection is weak relative to recombination. As noted by Barton and Turelli (1991) and Kirkpatrick et al (2002), the QLE gives surprisingly accurate predictions well beyond the situations in which its assumptions are likely to fail.

Social evolution theory

What is social evolution?

Social evolution theory is concerned with the evolution of traits which impact on the fitness of individuals other than their bearer, especially when there are correlations between the traits of interacting individuals (social partners). Classically, these are categorised according to the sign of the marginal fitness effects for the bearer of the trait and this individual's social partners – figure 1.1. Mutually beneficial (+/+) interactions are *mutualistic*, those which benefit the bearer at the expense of the recipient (+/-) are *selfish*, those which benefit the recipient at the expense of the bearer (-/+) are *altruistic*, and those which are mutually harmful (-/-) are *spiteful* (Hamilton 1964, Trivers 1985). The social evolution literature abounds in intentional language, and so it may be somewhat surprising to find that its theoretical foundations lie in the study of passive,

mechanical processes which constitutes population genetics. The link between the two fields is provided by Price's theorem (Price 1970, Frank 1998, Grafen 1999).

Price equation

The Price (1970) equation, one of the three major contributions G.R. Price made to evolutionary biology during his short, tragic career (Frank 1995a), is an exact and general statement of evolutionary change, and applies to any mapping of subsets and their phenotypes between sets. A set of subsets (say, a population of individuals) is denoted the parental set, and another set is the offspring set. Subsets are indexed $i \in I$ in each set, with each subset having a frequency p_i , phenotypic value z_i and fitness w_i in the parent set and a frequency p'_i and phenotype z'_i in the offspring set. Offspring subsets are mapped to parental subsets by matching indices. Thus the phenotypic change due to transmission between a parent indexed i and its offspring (also indexed i) is $\Delta z_i = z'_i - z_i$. We are interested in the change in the mean phenotype of the population. This is:

$$\Delta \bar{z} = \bar{z}' - \bar{z} = \sum_i p'_i z'_i - \sum_i p_i z_i = \sum_i p_i \frac{w_i}{\bar{w}} (z_i + \Delta z_i) - \sum_i p_i z_i = \sum_i p_i \frac{w_i}{\bar{w}} z_i - \bar{z} + \sum_i p_i \frac{w_i}{\bar{w}} \Delta z_i \quad [7.6]$$

yielding the Price equation:

$$\Delta \bar{z} = Cov_i[w_i / \bar{w}, z_i] + E_i[(w_i / \bar{w}) \Delta z_i]. \quad [7.7]$$

This is a complete, exact, general statement of evolutionary change. It holds for arbitrary ploidy (including mixed ploidies), any mating system, mode of inheritance, social systems, etc. The conventional interpretation of Price's equation is that the covariance term represents change in the mean trait value of the population due to the differential reproductive success of subsets (say, selection between individuals) and the expectation term represents the change due to transmission between subsets and their offspring (say, details of inheritance). The key to understanding the Price equation is to understand that it says very little explicitly, that a great deal is implicit, and that it is most useful as a

conceptual aid. For this reason, unless dealing with the fundamentals of evolutionary theory, it is not advisable to begin an analysis with Price's equation. Rather, the equation can be used for interpretation of results, as it helps to partition the various causes of change.

As it stands, Price's equation lacks dynamic sufficiency, meaning that it cannot be used to predict the course of evolution beyond a single generation. For example, given information about a population's composition in terms of genotype fitnesses and frequencies, Price's equation can be used to predict the change in gene frequencies, but not the associations between loci. However, since the equation can be used to follow the evolution of any trait, we can follow the change in linkage disequilibrium terms. The problem is that this requires knowledge of higher order associations. In general, dynamic sufficiency requires that higher order moments of population composition can be decomposed into lower order moments (Barton & Turelli 1987, Frank 1998). Such a decomposition requires a complete model of population composition (in terms of a finite number of allelic states) and processes. Given a complete model of the population, Price's approach can be made dynamically sufficient. The multi-locus methodology, in particular the recursions for changes in the association terms (D_A) due to selection and transmission, is the result of applying Price's theorem to a complete model of population genetics, as noted by Barton and Turelli (1991). The complete model of the population is specified by the multi-locus selection coefficients (a_A), genotype-phenotype mappings (\square_A), transmission rules ($t_{U \square A}$), and generalized linkage disequilibria (D_A).

To demonstrate this, we now derive the multi-locus expressions for change in associations due to selection (equation [7.3]) and transmission (equation [7.5]) from the above Price equation [7.7]. Consider a population with focal association D_A undergoing first a selection event (to give $D_A[\square]$) and then a transmission event (to give $D_A[\square\square]$). The change in the population mean deviation due to selection ($\square_s E[\square_A] = \square_s D_A = D_A[\square] - D_A$) is:

$$\begin{aligned}
\Delta_S D_A &= Cov(w/\bar{w}, \Delta_A) \\
&= \sum_{U \subseteq W} \sum_{\{V: V \neq U, V \subseteq W\}} \Delta_{(w/\bar{w})|U} \cdot Cov(\Delta_U, \Delta_A) \\
&= \sum_{U \subseteq W} a_U (D_{UA} - D_U D_A)
\end{aligned} \tag{7.8}$$

where the complicated term appearing in the second line is a partial regression of relative fitness (w/\bar{w}) on the product of the deviations (Δ_U) for a set of positions (U), holding all other deviation terms fixed, and so is simply the multi-locus selection coefficient, a_U . The change due to transmission ($\Delta_T D_A = D_A - D_A$) according to Price is:

$$\begin{aligned}
\Delta_T D_A &= E[(w/\bar{w}) \Delta_A] \\
&= E \left[\sum_{U \subseteq W} \Delta_U + \sum_{V \subseteq W} a_V (\Delta_V - D_A) + \sum_{U:U=A} \sum_{V:V \subseteq W} t_{U|A} D_U \Delta_V \right] \\
&= \sum_{U:U=A} t_{U|A} D_U \Delta_A + \sum_{U:U=A} \sum_{V \subseteq W} t_{U|A} a_V (D_U D_V - D_U D_V) + \sum_{V \subseteq W} a_V (D_{AV} - D_A D_V) \tag{7.9} \\
&= \sum_{U:U=A} t_{U|A} D_U \Delta_A + \sum_{V \subseteq W} a_V (D_{AV} - D_A D_V) \\
&= \sum_{U:U=A} t_{U|A} D_U \Delta_A
\end{aligned}$$

Using these recursions the change in each of the association terms (D_A) can be determined, giving a complete description of the population in the next time step. The recursions can therefore be applied again, to give a full compositional description of populations placed further in the future. At any stage, expression [7.1] (the genotype-to-phenotype map) can be applied to give a complete phenotypic description of the population. Thus, a Price equation analysis can be dynamically sufficient, given a complete (closed) model for it to work upon. The combination of Price's equation and a completely general notation in which to fully describe the composition of the population is the multi-locus methodology.

Levels of selection versus neighbour-modulated fitness

One of the immediate applications of Price's equation is to the theory of group selection (Price 1972a, Wade 1985). Consider that the subsets (indexed i) in the above derivation of Price's equation are now made up of smaller sub-subsets (indexed $j \in J$). The subsets can be described in terms of the properties of their component sub-subsets, i.e.

$$w_i = \sum_j (q_{ij}/q_i)w_{ij} \quad [7.10]$$

and

$$z_i = \sum_j (q_{ij}/q_i)z_{ij}, \quad [7.11]$$

where q_{ij} is the frequency of the j^{th} sub-subset in the whole set, and hence q_{ij}/q_i is the frequency of the j^{th} sub-subset in the i^{th} subset. The transmission term in the Price equation [7.7] can therefore be regarded as being made up of the reproduction and redistribution of the sub-subsets during the reproduction of the subsets (i.e. a lower level selection event) plus a component describing changes in the properties of these sub-subsets themselves (a lower level transmission event):

$$\begin{aligned} \Delta \bar{z} &= Cov_I[w_i/\bar{w}, z_i] + E_I[(w_i/\bar{w})\Delta z_i] \\ &= Cov_I[w_i/\bar{w}, z_i] + E_I[Cov_J[w_{ij}/\bar{w}, z_{ij} | i] + E_J[\dots]] \end{aligned} \quad [7.12]$$

The lower level transmission can be further expanded to involve selection between even lower levels, and associated transmission, and so on for an arbitrary number of levels of selection. For simplicity, and to make the potentially confusing general description above more concrete, we will focus on only two levels – individuals and groups of individuals. We will also assume that individuals have perfect heredity ($\Delta z_{ij} = 0$). The Price equation [7.7] therefore takes the form:

$$\begin{aligned}
\bar{z} &= Cov_I[w_i/\bar{w}, z_i] + E_I[Cov_J[w_{ij}/\bar{w}, z_{ij} | i]] \\
&= \sum_i q_i (w_i/\bar{w}) z_i \bar{z} + \sum_i q_i \sum_j (q_{ij}/q_i) (w_{ij}/\bar{w}) z_{ij} (w_i/\bar{w}) z_i \\
&= \sum_i \sum_j q_{ij} (w_{ij}/\bar{w}) z_{ij} \bar{z} \\
&= Cov[w_{ij}/\bar{w}, z_{ij}]
\end{aligned}
\tag{7.13}$$

Thus the separate group and individual level selection terms can be summarised in a single individual selection covariance form. The key to understanding this selection covariance is that w_{ij} is the individual's total fitness, which contains information about that individual's relative success within its group, and the group's relative success within the whole population. In the context of the evolution of altruism, altruists suffer a within-group disadvantage ($Cov_J[w_{ij}/\bar{w}, z_{ij} | i] < 0$) due to exploitation by more selfish social partners, and a group level advantage ($Cov_I[w_i/\bar{w}, z_i] > 0$) due to their altruism, which might under some conditions give a total advantage for altruism ($Cov[w_{ij}/\bar{w}, z_{ij}] > 0$). The 'neighbour-modulated' fitness (w_{ij} ; Hamilton 1964) will reflect any tendency for altruistic individuals to associate with other altruists such that the benefits of socialising of altruistic neighbours might outweigh the immediate costs of altruism, to derive a net fitness benefit. The condition under which this is met is Hamilton's (1963) rule, $RB > C$, which we will derive in a later section from such neighbour-modulated fitness considerations. This illustrates a fundamental equivalence between group selection and kin selection – mathematically they are the same process. The kin selection versus group selection debate is therefore empirically empty, yet it still rages in many of social evolution's sister disciplines (Bergstrom 2002).

Using neighbour modulated fitness to model social evolution is equivalent to using a 'levels of selection' approach. Wenseleers et al. (2004) provides a recent example of how social evolutionary problems – in their example, worker policing – can be tackled from these different angles, illustrating their equivalence. In developing the social evolutionary aspects of the multi-locus methodology, we can equivalently take two approaches: (1) a levels of selection view, where we assign fitnesses to groups of individuals according to

the genetics of these individuals, and model the within group selection process within the generalised multi-locus framework for transmission; or (2) a neighbour-modulated fitness view, where we assign fitnesses to individuals according to their genetical composition and that of their social partners. In practice, levels of selection approaches are rather more cumbersome and technically problematic than neighbour-modulated fitness approaches, and so, following the trend of the social evolution literature, we will focus on the latter for the remainder of this paper.

Secondary theorem of natural selection & the phenotypic gambit

Social evolution theory, as with evolutionary ecology in general, mostly concerns itself only with the operation of selection on phenotypes, which is only one part of total evolutionary change. This is neatly summarised by Price’s covariance term, or what is often referred to as Robertson’s (1966) secondary theorem of natural selection:

$$\Delta_s \bar{z} = Cov[w/\bar{w}, z]. \tag{7.14}$$

This phenotypic gambit (Grafen 1984), whereby changes due to transmission are ignored, allows a tractable analysis of evolutionary problems where we have no information about the genetic architecture of a trait. The gambit pays off, since many predictions of social evolution theory are astoundingly well supported by empirical observation, in a quantitative rather than simply qualitative sense. For example, sex allocation theory provides among the best evidence for adaptation in the real world (West & Herre 1998, Frank 2002). In addition to pragmatics, the focus on only a partial change has a more fundamental basis, and follows the precedent of R.A. Fisher (1930, 1941).

Fundamental theorem of natural selection & individual as maximising agent analogy

In Fisher’s view, the “most important application of this analysis (which can be applied to any measurable character) is to give a rational account of the action of natural selection” (Fisher 1941). This rational account formed the basis of chapter 2 of his (1930) book, *The genetical theory of natural selection*. The result, which he described as the

fundamental theorem of natural selection, perplexed biologists for decades until it was explained by Price (1972b) and Ewens (1989) – see Edwards (1994) for a complete history of this theorem. Although Price did not use this particular approach, the true meaning of the fundamental theorem is most easily illustrated using his selection covariance mathematics (Edwards 1994, Frank 1998). As Fisher (1941) hinted, the fundamental theorem is simply a special case of the secondary theorem of natural selection, where the focal trait (z) is fitness (w) itself. Hence:

$$\square_s \bar{w} = Cov[w/\bar{w}, w] = Var[w]/\bar{w}. \quad [7.15]$$

As Fisher (1930) put it, “the rate of increase of fitness of any species is equal to the genetic variance in fitness”, and since variances are non-negative then the mean fitness of the population increases when there is variation in fitness. By framing the derivation in this way, we see that the fundamental theorem is a statement of only a partial change (Price 1972b, Ewens 1989, Frank & Slatkin 1992, Edwards 1994). Fisher (1930) deliberately excludes changes in mean fitness due to the “deterioration of the environment”, which are neatly summarised by Price’s (1970) transmission term (in equation [7.7]). Of course, the fundamental theorem could never claim to be a complete description of evolutionary change in mean fitness. Much attention has been devoted to demonstrating how the intricacies of genetical systems can lead to decreases in the mean fitness of population (Moran 1964), but simply acknowledging the existence of natural disasters should convince that mean fitness does not always increase. Price (1972b) was disappointed with this conclusion, as was Ewens (1989), who popularised this interpretation. If the fundamental theorem is only a partial statement, then what is its significance?

The significance is that this partial change represents the engine of adaptation, a mathematical description of Darwin’s improbability generator, natural selection (Grafen 2003). Fisher has isolated, from the complicated and perhaps intrinsically unpredictable total evolutionary change in fitness, the purely mechanical process which gives rise to adaptedness and hence the appearance of design. Thus, the fundamental theorem provides

the beginnings of a formal logical basis for Darwin’s analogy that natural selection should cause individuals to behave as if they were designed to maximise their fitness (Grafen 1999, 2002). It provides a licence for biologists to make use of this analogy, and hence forms the fundamental basis of all evolutionary optimisation theory (Grafen 2003), which side-steps the details of population genetics to ask: which strategy should an organism employ in order to maximise its fitness?

With the adoption of Darwin’s analogy comes the language of agency, which has been used extensively and profitably within evolutionary ecology. The use of such intentional language highlights a major problem for evolutionary biologists – the pervasiveness of apparently ‘altruistic’ behaviours in the natural world. W.D. Hamilton (1963, 1964, 1970) solved the problem by introducing the concept of neighbour-modulated fitness and an associated condition – Hamilton’s rule – to describe when social behaviours are selectively favoured.

Hamilton’s rule

As with the fundamental theorem of natural selection, much attention has been devoted to demonstrating the non-validity of Hamilton’s rule (reviewed by Grafen 1985a). We shall see that the rule is a mathematically true statement, and that it is only the action of natural selection which is of interest.

Implicit in the secondary theorem is the impact on fitness (w) of all the determinants of fitness which are correlated with the focal trait (z) – earlier we introduced the concept of neighbour-modulated fitness, whereby the phenotype of a social partner is included as a determinant of fitness. Hamilton (1964, 1970) makes such social determinants explicit in the derivation of his rule. From the secondary theorem, we have

$$\Delta_S \bar{z} = Cov[w/\bar{w}, z] = \Delta_{w,z} Var[z] / \bar{w} \quad [7.16]$$

The regression of relative fitness on one's own phenotype value ($\beta_{w,z}$) can be further partitioned to give:

$$\beta_{S,z} = (\beta_{w,z \cdot Z} + \beta_{w,Z \cdot z} \beta_{Z,z}) \text{Var}[z] / \bar{w} \quad [7.17]$$

Where the partial regression of fitness on own phenotype value (i.e. holding fixed the social partner phenotype, Z) is $\beta_{w,z \cdot Z} = -C$, i.e. the cost of the social phenotype; the partial regression of fitness on social partner phenotype (i.e. holding own phenotype fixed) is $\beta_{w,Z \cdot z} = +B$, i.e. the benefit of having social partners with the trait; and the regression of social partner phenotype on own phenotype is $\beta_{Z,z}$ which is the coefficient of relatedness (R ; Hamilton 1970, Grafen 1985a) between the focal individual and its social partners. Assuming that there is some variance in phenotype, selection acts to increase the average phenotypic value of the population when

$$R B > C. \quad [7.18]$$

This derivation of the rule has been phrased in terms of a focal individual's direct, neighbour-modulated fitness (Hamilton 1964) – the costs and benefits accrue directly to the focal individual, and the relatedness term describes how the phenotypes of social partners relates to one's own phenotype. A potential problem with neighbour-modulated fitness is that it cannot properly be regarded as a measure that is maximised by an individual agent. There are two reasons for this. Firstly, the association (relatedness, R) which ensures that, for example, an individual who displays more altruism than average enjoys the company of social partners who are more altruistic than average can in general only be thought of in terms of correlation rather than causation. The focal individual does not directly manipulate the social behaviours of her partners. Thus, there may be a correlation, but not necessarily a causal relationship, between an individual's strategy and fitness. Secondly, a worker in a social insect colony may altruistically forego her own reproduction in order to help the queen raise progeny. Clearly, such a strategy does not maximise the neighbour-modulated fitness of the altruist. Properly understood, neighbour-modulated fitness is a measure associated with the strategy (gene, breeding

value, etc) itself, which is averaged over instances of the strategy across the various classes of individual, and is not associated with any particular individual. This may be described as the "gene's eye view" (Dawkins 1976). In order to salvage the individual as maximising agent analogy for social behaviours, Hamilton (1964) introduced the concept of "inclusive fitness". Rather than measuring an individual's direct success as a function of its social strategy and the correlated strategies of its neighbours, inclusive fitness measures all the effects of a focal actor's behaviour on the reproductive success of recipients, each increment being weighted according to the relatedness between the actor and the recipient. Here, relatedness is regarded as a measure of fidelity of transmission of one's own genes through the reproduction of social partners as opposed to the direct alternative (Frank 1997). Inclusive fitness is then associated with particular actors, and is a direct outcome of their behaviours, so it represents a true individual maximand. Neighbour-modulated fitness and inclusive fitness are simply alternative methods of book-keeping, and are equally valid approaches (Frank 1997a, 1998). It is interesting to note that, either way, the correct definition of Hamilton's (1970) coefficient of relatedness is a regression, and is not in general a probability measure such as the probability that genes are identical by descent (Malecot 1948).

As discussed above in relation to Price's theorem, Hamilton's rule is a rather subtle statement of evolutionary change, with many details implicitly tidied away into its three components, so naïve applications of the rule are likely to lead to difficulties – as we shall see. In undertaking a social evolutionary analysis it is advisable to begin with a concrete model of, for example, neighbour-modulated fitness, rather than beginning with Hamilton's rule. Hamilton's rule should appear as a result of the analysis, and provides a useful conceptual aid (Taylor & Frank 1996, Frank 1998).

We now derive an example Hamilton's rule for a simple model, using the multi-locus machinery. Let haploid individuals who socialise in pairs have a biallelic locus controlling their social behaviour. An allele with value $X_i = 1$ has population frequency p_i and causes: a direct fitness cost to self (who is denoted 1); a direct fitness benefit to partner (denoted 2); where these two components of fitness are additive. The alternative

is a null allele, with value $X_i = 0$, and frequency $q_i = 1 - p_i$. The change in the frequency of the altruistic allele due to selection is:

$$\Delta_S p_i = \Delta_S D_i = a_{i_1} D_{i_1 i_1} + a_{i_2} D_{i_1 i_2} = (a_{i_1} + a_{i_2} R) p_i q_i \quad [7.19]$$

i.e. selection causes an increase in the frequency of the altruistic allele when $R a_{i_2} > -a_{i_1}$. The multi-locus selection coefficients a_{i_1} and a_{i_2} provide the (relativised) cost and benefit terms of Hamilton's rule. This simple derivation illustrates a connection between the concept of relatedness in social evolution theory and linkage disequilibrium in population genetics – both are associations between gene positions. As we discovered in the derivation of Hamilton's rule [7.18], relatedness is a regression, and moreover in this example model it is the regression associated with the covariance term describing the 'linkage disequilibrium' between individuals within a locus. This might seem rather obvious, but it is important to explicitly point out, as there is extensive misunderstanding as to what the coefficient of relatedness is.

Relatedness is often interpreted as a probability measure, though this is not in general correct – as we have seen, it is a regression coefficient. For example, negative probabilities are not permissible, yet negative relatedness is, and this allows for the evolution of spiteful behaviours (Hamilton 1970, Grafen 1985a, Foster et al 2001, Gardner & West (in press), chapter 4). Why is this misconception so prevalent? Hamilton (1963) understood that the coefficient of relatedness was in principle a regression coefficient, but argued that under weak selection it could be approximated by Wright's (1922) correlation coefficient of relationship. The coefficient of relationship is expressed in terms of path coefficients describing the genetic associations between and within individuals, and these have popularly been interpreted as probabilities of IBD (Malecot 1948). It is interesting to note that Wright used the example of negative path coefficients between uniting gametes when there is outbreeding to illustrate his disagreement with Malecot's probability of IBD approach (Wright 1969, Nee et al. 2002). Malecot's interpretation, which may be valid in certain circumstances though not in general, is presented as mathematical fact in such classic texts as Crow & Kimura's (1970)

Introduction to population genetics, and so the misconception has taken root in the heart of population genetics. Thus we are left in the bizarre situation where population geneticists are happy to talk about negative linkage disequilibrium but not negative relatedness – although the above derivations (hopefully) demonstrate their conceptual equivalence. While on the subject of what Hamilton's relatedness measure is not, is worth pointing out that it is not a measure of genealogical closeness. Hamilton (1964) illustrated this point with the famous 'green beard' (Dawkins 1976) thought experiment. See Queller et al. (2003) and Keller & Ross (1998) for empirical examples of green beards.

Given the conceptual link between relatedness and linkage disequilibrium, it may be fruitful to imagine that the same forces shaping linkage disequilibrium will be acting analogously upon relatedness. Linkage disequilibrium can arise due to drift and incomplete recombination, and similarly positive relatedness between social partners will often arise due to drift and incomplete dispersal generating within-group associations. It is also well appreciated that epistatic interaction between loci can be a cause of linkage disequilibrium, so we could expect positive relatedness to arise when, say, cooperative groups have synergistic success. For example, Frank (1994) has shown that synergistic selection for more mutualistic groups can generate positive relatedness between social partners, even where genealogical closeness is ruled out (for example, because the social partners belong to separate species). A slight technical difficulty here is that epistasis is defined as a departure from multiplicity, whereas synergy terms in social evolution models (encountered in the next subsection) are described as deviations from additivity. How far the analogy between linkage disequilibrium and relatedness will stretch is unclear, and merits attention.

Extending Hamilton's rule

In the derivation of Hamilton's rule [7.18] for the example model above, we assumed that components of fitness (costs, benefits) combined additively. Allowing for such an additional interaction term extends the rule, so that we have:

$$\Delta_S p_i = a_{i_1} D_{i_1 i_1} + a_{i_2} D_{i_1 i_2} + a_{i_{12}} D_{i_1 i_1 i_2} > 0. \quad [7.20]$$

It is worthwhile to pause here and consider what the above extension represents, as previous discussion surrounding this type of approach have been misleading and are borne out of misunderstandings about Hamilton's rule. Queller (1984, 1985) stipulated an equivalent extension – his approach is effectively a special case whereby all the reference values are zero ($\phi_i = 0$), whereas with the multi-locus approach they are entirely arbitrary, though it is natural to set them equal to the mean trait value ($\phi_i = \bar{X}_i$) – and this was specially motivated by the possibility of non-additivity of fitness components within social interactions. He described new 'synergy' coefficients, related to the $D_{i_1 i_2}$ term (which in our scheme may be interpreted as the population average association between an individual's allelic value at locus i and the association at i between that individual and her social partner), which is multiplied by a term relating to the interaction payoff. Grafen (1985a, b) argued that such complicating terms are actually implicit in the existing cost (C) and benefit (B) of Hamilton's rule, so that the rule already sufficiently handles such scenarios.

To illustrate, consider a two player game with payoff matrix as illustrated in figure 7.1. Queller (1984) essentially argues that the 'Hamilton's rule' $Rb > c$ is insufficient to predict whether or not selection favours cooperation, and so corrects the rule by adding a synergy coefficient to multiply the interaction payoff (d). But $Rb > c$ is not the Hamilton's rule we derived in the previous section, it is a straw man, and is easily shown to be deficient. For example, if we allow only the pure strategies *Cooperate* (C ; $z = 1$) and *Defect* (D ; $z = 0$), with respective frequencies p and q , then we have fitnesses:

$$\begin{aligned} w_C &= \bar{w} + (R + (1 - R)p)(b - c + d) - (1 - R)(1 - p)c \\ w_D &= \bar{w} + (1 - R)pb \end{aligned} \quad [7.21]$$

Applying the secondary theorem, we find that selection favours an increase in the *Cooperate* strategy when:

		Player 2 strategy	
Player 1 strategy	Cooperate	Defect	
Cooperate	$b - c + d$	$-c$	
Defect	$+b$	0	

Figure 7.1. Payoff to player 1 from a social interaction between players 1 and 2, as a function of their social strategies. Payoffs are in addition to baseline fitness.

$$\begin{aligned}
&Cov(w,z) > 0 \\
&Rb \square c + (R + (1 \square R)p)d > 0 \\
&R(b + qd) > c \square pd \\
&RB > C
\end{aligned}
\tag{7.22}$$

where the three components of Hamilton’s rule fulfil their proper definitions: $R = \square_{z,z}$, $B = \square_{w,z \star z}$ and $C = \square_{w,z \star Z}$. Thus Hamilton’s rule, if correctly understood, is equipped to deal with such game theoretic scenarios, even when fitness effects are large such that interaction terms (d) are nontrivial. With this in mind, expression [7.20] is not presented as the addition of a correction term to complete a deficient Hamilton’s rule, but rather it is an extension of the existing components into their implicit subcomponents, making explicit what is already there. Hamilton’s rule provides, in unmodified form, the general framework in which to understand social evolutionary problems as sought by, for example, Charlesworth (2000) and Wenseleers & Ratnieks (2001). The cost and benefit terms are somewhat complicated, so that Hamilton’s rule cannot quite claim “to be applied painlessly to solve particular problems” (Charlesworth 2000), but this misses the point. As remarked upon previously, Hamilton’s rule should not be used as a starting point for an analysis, but rather it should appear as a result of applying more standard and

concrete methodologies (Taylor & Frank 1996, Frank 1998), such as population genetics or game theory. By framing the results of an analysis in terms of Hamilton’s rule, we have translated the problems into the common language of social evolution theory, allowing for simple comparisons and contrasts regardless of the diversity of biological scenarios and analytical approaches.

Simply by applying the multi-locus notation, we can extend Hamilton’s rule for arbitrary numbers of gene positions (for example, multiple loci and multiple social partners) and for arbitrary statistical associations and fitness interactions between these positions. From the secondary theorem, the change in the population average for any trait (z) which is attributed to the action of selection is given by the covariance of relative fitness and trait value. In general, this is:

$$\Delta_S \bar{z} = Cov[w/\bar{w}, z] = \sum_{A \subseteq G} \sum_{U \subseteq G} \Delta_A a_U (D_{AU} - D_A D_U). \quad [7.23]$$

As we have seen, Hamilton’s rule is simply a restatement of the secondary theorem which gives particular attention to the association between social partners. Making a separate generalised multi-locus Hamilton’s rule makes little sense. Expression [7.23] suffices, provided that we keep in mind that sets of gene positions A and U may span several individuals, so that certain associations may be interpreted as linkage disequilibria, others as relatedness, and others as between-locus between-individual associations. To illustrate, we will consider a simple social evolutionary multi-locus problem which makes reference to each of these different associations in the next section.

Example: cooperation and punishment

The problem

A major concern of evolutionary biologists is to explain the prevalence of cooperation from primordial replicators to human and animal society, given that at every level of

biological organization there is the possibility of selfish behaviour disrupting group harmony (Maynard Smith & Szathmary 1995). Human cooperation in particular poses a major problem, as in general it is felt that relatedness (and the probability of repeat interactions) is not high enough to support altruism (Fehr & Fischbacher 2003 provide a recent review). One solution which has recently received much attention is the threat of punishment (Boyd & Richerson 1992, Sober & Wilson 1998, Fehr & Gächter 2000) – although this poses its own problems. Punishing incurs costs for the punisher, and so it too has been regarded as altruistic (Fehr & Gächter 2002). Such altruistic punishment has been observed repeatedly in empirical studies of human behaviour (for example, Fehr & Gächter 2002). As yet it defies explanation. One argument (e.g. Sober & Wilson 1998) suggests that punishment will often be cheaper than cooperation, so that kin selection can maintain altruistic punishment even when relatedness is too low for cooperation to be directly favoured. Hence, kin selection maintains punishment which maintains cooperation.

Gardner and West (2004a; chapter 6) rejected this explanation, showing that since punishment directly harms both punisher and punished, increased relatedness between social partners directly disfavours the evolution of punishment. For this reason, such a lose-lose interaction might be better described as spite (for example, see Johnstone & Bshary 2004). Gardner and West suggested that a different association between individuals – the association between the cooperation strategy of one's social partners and one's own punishment strategy – can allow for the evolution of punishment. A verbal argument suggested that when relatives tend to interact, linkage disequilibrium will arise since punishers are more likely to be associated with punishers and hence are under stronger selection to cooperate, and this linkage disequilibrium coupled with relatedness would give rise to the crucial association between the traits between social partners. Gardner and West's social evolutionary analysis was incapable of following the evolutionary dynamics of such associations, and so the idea remains unexamined.

Illustrative Model

Following Gardner and West (2004a; chapter 6), we examine a simple model which captures all the necessary details. Haploid individuals interact in pairs, with one individual from each pair randomly assigned the role of Player 1, and the other Player 2. Player 1 either cooperates (incurring personal cost c and giving Player 2 a benefit b) or defects (no pay-off for either player), and Player 2 may respond to defection either with punishment (incurring personal cost a and inflicting a cost d for Player 1) or else forgiveness (no pay-off for either player). Cooperation and punishment strategies are encoded by biallelic loci i and j respectively, with allele $X_i = 1$ giving cooperation / $X_i = 0$ giving defection, and $X_j = 1$ giving punishment / $X_j = 0$ giving forgiveness, and these alleles have population frequency p_i / q_i and p_j / q_j respectively. Mating is at random (no associations between uniting gametes) and recombination between the two loci occurs at rate r . For simplicity, we will treat the relatedness (R) between social partners as a parameter, rather than an evolving variable.

Multilocus analysis

The fitness function is:

$$w = 1 - \frac{c}{2} X_{i_1} + \frac{b}{2} X_{i_2} - \frac{a}{2} (1 - X_{i_2}) X_{j_1} - \frac{d}{2} (1 - X_{i_1}) X_{j_2}. \quad [7.24]$$

Making the substitution $\Delta_i = X_i - p_i$, (i.e. $\phi_i = p_i$) and dividing both sides by \bar{w} , gives the form of expression [7.2]. Here, the multi-locus selection coefficients are those coefficients multiplying the corresponding allelic deviations Δ_A . These are:

$$\begin{aligned} a_{i_1} &= (\Delta c + dp_j) / 2\bar{w} & a_{i_2} &= (b + ap_j) / 2\bar{w} \\ a_{j_1} &= \Delta a q_i / 2\bar{w} & a_{j_2} &= \Delta d q_i / 2\bar{w} \\ a_{i_1 j_2} &= d / 2\bar{w} & a_{i_2 j_1} &= a / 2\bar{w} \end{aligned} \quad [7.25]$$

Where:

$$\begin{aligned}\bar{w} &= 1 + \frac{b \square c}{2} p_i \square \frac{a+d}{2} q_i p_j + \frac{a}{2} D_{i,j_2} + \frac{d}{2} D_{i_2,j} \\ &= 1 + \frac{b \square c}{2} p_i \square \frac{a+d}{2} q_i p_j + \frac{a+d}{2} R D_{ij}\end{aligned}\tag{7.26}$$

Note that, because of the symmetry of the model, the association between individuals between loci is the same in both directions: $D_{i,j_2} = D_{i_2,j}$. Further, this association is equal to $R D_{ij}$, i.e. the between locus between individual association is given by the product of the association within individuals between loci (linkage disequilibrium, D_{ij}) and the regression between individuals within loci (relatedness, R). This is shown by considering that:

$$\begin{aligned}D_{i,j_2} &= Cov[\square_{j_1}, \square_{j_2}] \\ &= \square_{\square_{j_1} \square_{j_2}} Var[\square_{j_2}] \\ &= (\square_{\square_{j_1} \square_{j_2} \cdot \square_{j_2}} + \square_{\square_{j_1} \square_{j_2} \cdot \square_{j_2} \square_{j_2} \square_{j_2}}) Var[\square_{j_2}] \\ &= (0 + R \square_{\square_{j_2} \square_{j_2}}) Var[\square_{j_2}] \\ &= R D_{ij}\end{aligned}\tag{7.27}$$

And by symmetry, the same is true for $D_{i_2,j}$. By substituting into expressions [7.3] and [7.5], correcting for the update in reference values after selection, we obtain the change in the frequency of the cooperation allele, the frequency of the punishment allele, and the linkage disequilibrium between the two loci.

Before proceeding to examine the invasion and maintenance conditions for punishing behaviour, we will first examine the direction of change in linkage disequilibrium ($\square D_{ij}$) when it is initially absent ($D_{ij} = 0$). Gardner and West gave a verbal argument suggesting that positive linkage disequilibrium should result when social partners are related, such that punishers associate with punishers and hence are more heavily selected to be

cooperators. Substituting $D_{ij} = 0$ into the recursion for linkage disequilibrium evolution gives:

$$D_{ij} = (1 - r)R \frac{a + d}{2\bar{w}} + \frac{(Rb - c + (Ra + d)p_j)(a + Rd)q_i}{4\bar{w}^2} p_i q_i p_j q_j \quad [7.28]$$

When there is zero relatedness ($R = 0$) between social partners and no linkage disequilibrium ($D_{ij} = 0$), after a single generation the linkage disequilibrium will be $(1 - r)a(p_j d - c)p_i q_i p_j q_j / 4\bar{w}^2$, i.e. it will increase if $p_j d > c$, and decrease if $p_j d < c$. This reveals that there is another pressure generating associations between the loci: the additivity of fitness components translates into a deviation from multiplicity (epistasis) and hence linkage disequilibrium. For weak selection ($a, b, c, d \ll 1$; hence negligible epistasis), and positive relatedness ($R > 0$) between social partners $R > 0$, then (to leading order) the linkage disequilibrium (D_{ij}) increases from zero to $\sim (1 - r)R(a + d)p_i q_i p_j q_j$ after a single generation. This is true for stronger selection given a sufficient frequency of cooperators (it is exactly true for $p_i \ll 1$, regardless of the strength of selection). This is the effect predicted by Gardner and West.

Let us first consider the evolutionary origin of punishment. When punishing is rare, cooperation is also disfavoured, so we may set $p_i = \epsilon p_i$ and $p_j = \epsilon p_j$, where the ϵ denotes an infinitesimal quantity. This being the case, we may also set the linkage disequilibrium $D_{ij} = \epsilon D_{ij}$. This allows us to ignore higher order terms, linearising the system. The resulting recursions can be summarised in matrix form:

$$\begin{bmatrix} \epsilon p_i \\ \epsilon p_j \\ \epsilon D_{ij} \end{bmatrix} = \begin{bmatrix} (Rb - c)/2 & 0 & (1 - R)a/2 \\ 0 & 1 - (a + Rd)/2 & (R(a + b + d) - c)/2 \\ 0 & 0 & (1 - r)(1 + (Rb - c + (1 - R)a)/2) \end{bmatrix} \begin{bmatrix} \epsilon p_i \\ \epsilon p_j \\ \epsilon D_{ij} \end{bmatrix} \quad [7.29]$$

Or, more compactly, as $\mathbf{x}' = \mathbf{M}\mathbf{x}$. The three eigenvalues are solutions (λ) of the characteristic equation $\text{Det}[\mathbf{M} - \lambda \mathbf{I}] = 0$ (where \mathbf{I} is the 3×3 identity matrix), and are: 1

$-(Rb - c)/2$, $(1 - r)(1 + (Rb - c + (1 - R)a)/2)$, and $1 - (a + Rd)/2$. The punishment allele invades when the leading eigenvalue (the solution with the largest magnitude) exceeds 1. Noting that we are only interested in situations where $Rb < c$ (because otherwise punishment is not required for the evolution of cooperation), the three eigenvalues are all less than 1. Hence, in this simple model, the punishment gene cannot invade the population.

We now consider the evolutionary maintenance of punishment, by setting $p_i = 1 - p_j$, $p_j = 1 - p_i$ and $D_{ij} = D_{ji}$. The recursions, summarised in matrix form, are:

$$\begin{bmatrix} p_i \\ p_j \\ D_{ij} \end{bmatrix} = \begin{bmatrix} \frac{Ra/2 + (1-R)b/2 + d/2}{1 + b/2 - c/2} & 0 & \frac{R(a+d)/2}{1 + b/2 - c/2} \\ 0 & 1 & \frac{Rb/2 + (1-R)(a+d)/2 - c/2}{1 + b/2 - c/2} \\ 0 & (1-r) \frac{1 - (1-R)(a+b+d)/2}{1 + b/2 - c/2} & 0 \end{bmatrix} \begin{bmatrix} p_i \\ p_j \\ D_{ij} \end{bmatrix} \quad [7.30]$$

The eigenvectors for the above matrix are:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} \frac{R(a+d)/2}{R(a+d)/2 + (1-R)a/2 - r(1-(1-R)(a+b+d)/2)} \\ \frac{(Rb - c + (1-R)(a+d))/2}{(Rb - c + (1-R)(a+d))/2 + r(1-(1-R)(a+b+d)/2)} \\ 1 \end{bmatrix} \quad [7.31]$$

Two of the three eigenvectors for the matrix are trivial, involving no variation at one of the two loci. The third is non-trivial only for $R > 0$. It is biologically meaningful only if it corresponds to non-negative allele frequencies, i.e. the first and second elements have the same sign. The corresponding eigenvalue is:

$$\lambda = (1-r) \frac{1 - (1-R)(a+b+d)/2}{1 + (b-c)/2} \quad [7.32]$$

when this is less than 1, the perturbation in allele frequencies and linkage disequilibrium is neutralised, so that the population of cooperator-punishers is resistant to invasion.

Lets first consider tight linkage relative to intensity of selection, i.e. $r \ll a, b, c, d$. Then the eigenvector is, approximate to leading order: $\{R(a+d)/(Rd-(1-2R)a), 1, 1\}$. This is biologically plausible for sufficiently small a . Making the appropriate substitutions into the eigenvalue, we find that this is less than 1 when $Rb + (1-R)(a+d) > c$, i.e. there are regions of parameter space allowing for the maintenance of punishment by selection. Looking to the other extreme, where selection is weak relative to recombination ($r \gg a, b, c, d$), the eigenvector is, to leading order, approximately: $\{-R(a+d)/2r, (Rb + (1-R)(a+d)-c)/2r, 1\}$. The first element is always negative, so to give a biologically plausible allele frequency would require a rescaling of the eigenvector such that we would have a negative linkage disequilibrium. From [7.28] we expect linkage disequilibrium to increase from zero, so we can rule out this solution as meaningless. Hence, if selection is weak then punishment is not resistant to invasion by a more forgiving allele. This is because the crucial association (between one's punishment and one's social partner's cooperation) is proportional to the linkage disequilibrium between the cooperation and punishment loci within individuals. Strong linkage disequilibrium cannot arise when selection is weak relative to recombination.

Discussion

Evolutionary problems involving multiple traits are problematic in that associations between traits can cause direct selection on one trait to result in indirect selection on the associated traits. Social evolution theory is concerned with the consequences of associations between individuals, but typically ignores associations between traits. For simplicity, co-evolving traits are assumed to be statistically independent, for example in the policing models of Frank (1995b, 1996b, 2003a). However, recent theory has emphasized the importance of between-trait associations in the evolution of costly punishment (Gardner & West 2004a; chapter 6) and altruism based on arbitrary tags (Axelrod et al. 2004). These studies extended Hamilton's rule to multiple co-evolving

social traits, but unfortunately lost dynamic sufficiency in the process. In extending the multi-locus methodology to social evolutionary problems we have provided a general framework within which we can construct extended Hamilton's rules describing the action of selection on co-evolving social traits. Additionally, the methodology also provides the means for making such an analysis dynamically sufficient. This has been illustrated by re-examining Gardner and West's model of cooperation and punishment using the new theoretical tool. Although we have focused on simple one-gene-for-one-trait models, the notation is sufficiently general to allow for social evolutionary problems involving arbitrary numbers of traits with arbitrary genetic architectures.

We have adopted a neighbour-modulated fitness view, derived from Price's theorem. Application of the Price equation to social evolution theory also suggests an alternative approach for describing the evolution of social behaviours. The levels of selection approach decomposes evolutionary change according to selection events at different scales of biological organisation. Selection at lower levels is described in terms of transmission at higher levels. Both approaches are equally valid – group selection is mathematically equivalent to kin selection. In extending the multi-locus methodology to social evolution we may validly employ either approach. In general, social evolution theory has found the latter approach to be the most useful for modelling actual biological problems, and so we have adopted a neighbour-modulated view for much of this paper. We have concerned ourselves with asking how relatedness between social partners influences the evolution of social traits, and have not in general enquired as to the evolution of relatedness itself. To do so would require that we specify a model of the segregation of individuals within and between groups, which is difficult in the current neighbour-modulated fitness scheme but is ably handled by the general selection / transmission framework of the multi-locus methodology when applied to groups. Hence, a multi-locus levels of selection approach will be more appropriate for certain problems, and deserves attention.

In showing that the multi-locus methodology is implicit in Price's (1970) scheme, we have illustrated the subtlety of Price's approach. Hamilton's (1963, 1964, 1970) rule,

which derives directly from Price's equation, is a similarly subtle statement of evolutionary change. Failure to recognize the subtlety and complexity of Hamilton's rule has led to statements to the effect that it is deficient, provides only approximate predictions, and that in complicated models it needs to be modified by adding novel components. Such attempts at fixing Hamilton's rule can be understood in terms of the multi-locus methodology, and we have shown that all that has been achieved is to make existing implicit components explicit. Of course, illuminating the hidden is of value, provided that we understand this is all that is being done. Hamilton's rule is a true, general statement describing the action of natural selection on social traits, and thus it provides a unifying principle and common framework within which the whole of social evolution theory may be understood. We emphasize that, because of the hidden subtleties, Hamilton's rule is not usually appropriate for use as a starting point in analyses of social evolution. A more concrete approach, such as starting with a population genetics model, or writing down a direct neighbour-modulated fitness function, is less beset with pitfalls, and if done correctly should result in Hamilton's rule in some form dropping out of the analysis. It can then be used as a conceptual aid (Taylor & Frank 1996, Frank 1998). The same applies to Price's theorem, which is generally unwieldy when employed to analyse particular problems, and is often more appropriate for understanding the results of an evolutionary analysis.

Gardner and West (2004a, chapter 6) highlighted the importance of the association between a focal individual's punishing strategy and the cooperation displayed by its social partners in favouring the evolution of costly punishment. A verbal model suggested this could be a manifestation of having both an association between the traits within an individual, and an association within traits between individuals. A multi-locus analysis confirms that both positive relatedness and linkage disequilibrium (due to incomplete recombination) are crucial for the association to arise. We have found that in some cases the tendency of punishers to associate with cooperators will overcome the direct disadvantages of punishment, namely the personal cost plus the disadvantage of being punished by more punishing relatives. This is more likely when selection is strong relative to recombination, such that significant linkage disequilibrium can evolve.

Punishment, as described in this model, is quite distinct from the policing models of Frank (1995b, 1996b, 2003a). Policing provides a direct benefit to one's group through prevention of competitive behaviours, whereas punishing modifies the social environment such that cooperation is favoured. Thus, it is perhaps better to consider it as an example of niche construction (Odling-Smee et al. 1996), as opposed to a kin selected trait.

Gardner and West's (2004a; chapter 6) dynamically insufficient analysis necessarily treats this key association as a population parameter. Given a certain fixed degree of association, they demonstrated that punishment is more easily favoured when common than when rare, so that the maintenance of punishment is relatively easy whereas the conditions under which it can invade are less readily satisfied. The dynamically-sufficient multi-locus analysis, follows the between-trait between-individual association as an evolving variable, reveals that in this model punishment can never invade but can be maintained, lending more weight to this result.

The generality of the existing multi-locus methodology notation implicitly allows for arbitrary social interactions, as we have seen. What other extensions might be fruitfully explored? So far, we have only considered interacting social partners which belong to the same population, although there is no reason why they could not belong to separate species. The generalised understanding of relatedness allows for Hamilton's rule to be applied to mutualisms (Frank 1994), and so this suggests an extension which could readily be integrated into the existing social evolutionary multi-locus framework. Additionally, the general notation available for describing the transmission of inherited factors and their contributions to fitness allows for the possibility of following cultural evolution, and perhaps more interestingly, gene-culture co-evolution. We have discussed the social multi-locus dynamics of co-evolving cooperation and punishment traits in terms of niche construction (Odling-Smee et al. 1996). The multi-locus notation provides a sufficiently general framework in which to examine and unify such processes.

8. Recombination and the evolution of mutational robustness: a two-locus model

Abstract

Mutational robustness is the degree to which a phenotype (such as fitness) is resistant to mutational perturbations. Essentially, robustness reduces the selection coefficient associated with deleterious mutations, providing an immediate benefit for the mutated individual. However, robust systems decay due to the accumulation of deleterious mutations that would otherwise have been cleared by selection. This decay has received very little attention in the evolution of robustness literature. At equilibrium, a population or asexual lineage will have a mutation load which is invariant with respect to the selection coefficient of deleterious alleles, so the benefit of robustness (at the level of the population or asexual lineage) is temporary. Previous work has shown that robustness can be favoured when robustness loci segregate independently of the mutating loci they act upon. I examine a simple multi-locus model that allows for intermediate rates of recombination and inbreeding to show that increasing the effective recombination rate can allow for the evolution of greater genetic robustness.

Introduction

The first ideas concerning phenotypic robustness were articulated by Waddington (1940) and Schmalhausen (1949) and were borne out of observations of the remarkable constancy of developmental traits in the face of both environmental and genetic perturbations, a phenomenon described by Waddington as ‘canalisation’. The explanation proposed by Waddington was adaptive. He reasoned that traits under stabilising selection towards some intermediate optimum should benefit from any mechanism that prevents deviation from that optimum due to either heritable (genetic) or non-heritable

(environmental) perturbations. The perturbations that are of interest to us here are heritable; specifically, deleterious mutations.

The evolution of genetic robustness is conceptually similar to the adaptive evolution of dominance first proposed by Fisher (1928). In both cases it is the heritable deviation from the wild type that is being buffered, and the selective advantage of the modifier is of the order of the mutation rate. Fisher believed that, although the selective advantage was weak, in a large population with a number of recessive mutations the accumulated selective pressure would win out, a belief not shared by Wright (Hartl 1989). Another related phenomenon which has received attention in the literature is the evolutionary transition from haploidy to diploidy. The benefit afforded by an extended diploid phase may be through the ‘masking’ of recessive or partially recessive deleterious mutations, although this would be a short-term benefit as the mutation load at equilibrium could be up to twice that for haploids depending on the degree of dominance (Crow & Kimura 1965; but see Kondrashov & Crow 1991, Perrot et al 1991). Together with the evolution of genetic robustness these scenarios involve evolutionary modification of the genetic system itself driven by the immediate benefit of alleviating the affects of deleterious mutations. Interestingly, models of diploidy evolution are incompatible with Fisher’s model for the evolution of dominance because they assume that deleterious mutations are at least partially recessive (Perrot et al 1991).

A classic result which motivates this study is that the equilibrium mutation load (L^* ; Haldane 1937) of a population undergoing irreversible deleterious mutation (at rate μ) is invariant with respect to the selection coefficient (s) of the deleterious allele. According to Price’s (1970) theorem, the change in the frequency (p) of the deleterious allele (which we will denote by allelic value $X=1$, to distinguish it from the wildtype, $X=0$, which has frequency $q = 1 - p$) is

$$\Delta p = Cov[w/\bar{w}, X] + E[(w/\bar{w})\Delta X] = \mu \frac{spq}{1 - sp} + \frac{\mu q}{1 - sp} \quad [8.1]$$

indicating that there is a stable equilibrium ($\frac{dp}{dt} = 0$, $\frac{d^2p}{dt^2} < 0$) at $p = \mu/s$ (provided $s > \mu$), and an unstable equilibrium ($\frac{dp}{dt} = 0$, $\frac{d^2p}{dt^2} > 0$) at $p = 1$ (which becomes stable if $\mu \geq s$). Denoting the stable equilibrium $p^* = \mu/s$, the mean fitness of the population at this stable point is $\bar{w}^* = p^* \times (1-s) + (1-p^*) \times 1 = 1 - \mu$, and hence the equilibrium mutation load ($L^* = 1 - \bar{w}^* = \mu$) is not a function of the selection coefficient. While it may be temporarily advantageous to reduce the selection coefficient, this leads to the accumulation of deleterious mutations that would otherwise have been cleared by selection, and so the population or asexual lineage with enhanced robustness does not improve its equilibrium mutation load. Thus there is no long term benefit for being robust. This mutational decay of robust systems has received only limited attention (Frank 2003b). If robustness has a cost, then it will in the long term cause a net disadvantage for the population or asexual lineage. In an asexual population, we predict eventual loss of robust lineages. However, in a sexual population a robust lineage might have a relative advantage despite robustness bringing a net cost to the population as a whole. Recombination between the robustness loci and those loci which are under deleterious mutation decouples the immediate benefit of robustness from the longer-term cost of generating a higher mutation load. There are two advantages of recombination: (1) the robust genome can discard the excess deleterious mutations it has accumulated, and (2) these are inflicted upon the non-robust lineages where they will cause enhanced damage to fitness, improving the relative fitness of the robust lineages in the population. This has received some attention, and the above reasoning is confirmed by contrasting the predictions of models of complete linkage (Hermisson et al 2002) with those which assume free recombination (Dawson 1999). However, results for robustness evolution with intermediate recombination rates (r) are lacking (de Visser et al. 2003).

We examine a simple model which captures the essence of this problem. The dynamics of the system are described using a multi-locus methodology (developed by Barton & Turelli 1991 and Kirkpatrick et al. 2002) which highlights the linkage disequilibrium between loci, and provides a general notation which will be helpful for extending the analysis to more complicated models. We will focus on the gradual evolution of robustness by examining when small increases or decreases in robustness strategy are

favoured. In particular, we will generate a general description for intermediate evolutionary stable strategy (ESS; Maynard Smith & Price 1973), given that one exists, and relate this to recombination and inbreeding rates.

Models and Analyses

Multi-locus model

We consider a simple model which captures all the important features of this problem – a large population of sexual haploids, with a lifecycle which involves (i) selection, followed by (ii) mutation, and finally (iii) mating to form diploid zygotes which undergo meiosis to form the next generation of haploid individuals. A biallelic locus i suffers recurrent mutation from the null allele ($X_i = 0$) to the mutant ($X_i = 1$) at rate μ , which incurs a fitness decrement s . The frequencies of the mutant and null are, respectively, p_i and $q_i = 1 - p_i$. A second locus j confers robustness, and takes either of two forms. The first ($X_j = 0$) allele confers a degree of robustness k_x which reduces the selection coefficient of the deleterious mutation from s to $(1 - k_x)s$. It also suffers a direct cost, c_x . The alternative ($X_j = 1$) allele confers robustness k_y and incurs cost c_y . The allele frequencies are q_j and p_j respectively, where $p_j + q_j = 1$. Consider that locus j determines the expression of a chaperonin type molecule which to some extent restores function to the mutated gene product of locus i . Parameters c and k are functions of expression strategy: increased expression enhances robustness k but carries a production cost c . Two strategies are considered, x and y , encoded by the respective alleles at locus j . We will assume that the direct effects of the loci multiply to give genotype fitness, and that μ is suitably small for us to ignore the possibility of fixation of the deleterious mutation.

Selection

Fitness can be written in the form:

$$w = (1 - X_i)(1 - X_j)(1 - c_x) + X_i(1 - X_j)(1 - (1 - k_x)s)(1 - c_x) + (1 - X_i)X_j(1 - c_y) + X_iX_j(1 - (1 - k_y)s)(1 - c_y) \quad [8.2]$$

Defining $\Delta = X - p$ as the deviation of an allelic value from the population expectation, we may expand the fitness function as outlined by Kirkpatrick et al. (2002) to obtain multi-locus selection coefficients and mean fitness of the population:

$$\begin{aligned} a_i &= \Delta s \left((1 - (c_x + k_x - c_x k_x)q_i - (c_y + k_y - c_y k_y)p_j) \right) \bar{w} \\ a_j &= (c_x - c_y \Delta s \left((c_x + k_x - c_x k_x) - (c_y + k_y - c_y k_y) \right)) \bar{w} \\ a_{ij} &= s \left((c_x + k_x - c_x k_x) - (c_y + k_y - c_y k_y) \right) \bar{w} \end{aligned} \quad [8.3]$$

$$\begin{aligned} \bar{w} &= 1 - sp_i - c_x q_j - c_y p_j - s(c_x + k_x - c_x k_x)(D_{ij} - p_i q_j) \\ &\quad + s(c_y + k_y - c_y k_y)(D_{ij} - p_i q_j) \end{aligned}$$

Where $D_{ij} = E[\Delta_i \times \Delta_j]$, and is the linkage disequilibrium between loci i and j (Barton & Turelli 1991, Kirkpatrick et al. 2002). We now determine the change due to selection for each of the three variables of this system: p_i , p_j and D_{ij} .

The change in the frequency of the deleterious mutation which is due to selection is:

$$\begin{aligned} \Delta_S p_i &= a_i D_{ii} + a_j D_{ij} + a_{ij} D_{ijj} \\ &= a_i p_i q_i + a_j D_{ij} + a_{ij} (1 - 2p_i) D_{ij} \end{aligned} \quad [8.4]$$

(Barton & Turelli 1991, Kirkpatrick et al. 2002).

The change in allele frequency, due to selection, at the robustness locus is:

$$\begin{aligned} \Delta_S p_j &= a_i D_{ij} + a_j D_{jj} + a_{ij} D_{ijj} \\ &= a_i D_{ij} + a_j p_j q_j + a_{ij} (1 - 2p_j) D_{ij} \end{aligned} \quad [8.5]$$

Following the multi-locus methodology of Barton & Turelli (1991) and Kirkpatrick et al. (2002), and taking care to update reference values for the allelic deviations to current population averages, the change in the linkage disequilibrium due to selection is:

$$\begin{aligned} \Delta_S D_{ij} &= a_i D_{ij} + a_j D_{ij} + a_{ij} (D_{ij} - D_{ij}^2) \Delta_S p_i \Delta_S p_j \\ &= a_i (1 - 2p_i) D_{ij} + a_j (1 - 2p_j) D_{ij} \\ &\quad + a_{ij} (p_i q_i p_j q_j + (1 - 2p_i)(1 - 2p_j) D_{ij} - D_{ij}^2) \Delta_S p_i \Delta_S p_j \end{aligned} \quad [8.6]$$

Mutation

Denoting the frequency of the deleterious variant after selection p_i^* , the change in frequency due to mutation is:

$$\Delta_M p_i = p_i^* - (1 - \mu) p_i = \mu (1 - p_i). \quad [8.7]$$

Since the j locus does not undergo mutation, $\Delta_M p_j = 0$. The change in the linkage disequilibrium due to mutation is:

$$\Delta_M D_{ij} = -\mu D_{ij} \quad [8.8]$$

Transmission

Transmission – the union of gametes, crossing over, and fair meiosis – does not alter the allele frequencies in this model, but it does impact on the linkage disequilibrium.

Following Kirkpatrick et al. (2002), the linkage disequilibrium between two positions after a transmission event is the expectation of the linkage disequilibria between the positions that were the source of the genes before transmission, weighting by the probability that the genes came from each source. In our model, this is:

$$D_{ij} = rD_{ij} + (1 - r)D_{ij}, \quad [8.9]$$

i.e. with probability r there has been a recombination event, such that one mating partner donated the i gene and the other the j gene, and with probability $1 - r$ there has not been a recombination event, such that the two genes derive from the same parent. In the former instance, the linkage disequilibrium between the two genes was D_{ij} , i.e. the association between the i and j genes between mating partners after mutation, and in the latter instance it is simply D_{ij} – the association between i and j within the same individual after mutation. The association between loci between mating partners emerges because there is an association between loci within individuals (linkage disequilibrium) and an association between individuals within loci (relatedness). The over-all association can be quantified as follows:

$$D_{ij} = Cov[X_{i1}, X_{j2}] = \beta_{i1j2} Var[X_j], \quad [8.10]$$

where β_{i1j2} is the regression of X_{i1} on X_{j2} , and $Var[X_j]$ is the variance in allelic values at locus j , i.e. $p_j q_j$. The regression coefficient can be expanded in terms of partial regressions:

$$\beta_{i1j2} = \beta_{i1j2 \cdot i_2} + \beta_{i1i_2 \cdot j_2} \beta_{i_2j_2}, \quad [8.11]$$

where: the partial regression of X_{i1} on X_{j2} holding X_{i_2} constant is zero ($\beta_{i1j2 \cdot i_2} = 0$, because any association between X_{i1} and X_{j2} is mediated by the between-individual within-locus association); the partial regression of X_{i1} on X_{i_2} is the regression coefficient of relatedness ($\beta_{i1i_2 \cdot j_2} = R$; Hamilton 1963, 1970) – in this context of relatedness between mating partners it is also the coefficient of inbreeding (f ; Wright 1922, Nee et al. 2002); and $\beta_{i_2j_2} = D_{ij} / Var[X_j]$ is simply the regression between the loci within an individual. Substituting into expression [8.10] obtains:

$$D_{ij} = R \sum_{i_2 j_2} \text{Var}[X_{j_2}] = R D_{ij}^0, \quad [8.12]$$

i.e. the association between-loci between-individuals is simply the product of linkage disequilibrium and relatedness. Substituting into expression [8.9] obtains the linkage disequilibrium after transmission:

$$D_{ij} = (1 - r(1 - R)) D_{ij}^0 = (1 - r_e) D_{ij}^0, \quad [8.13]$$

where $r_e = r(1 - R)$ is the ‘effective rate of recombination’.

Evolution of robustness

We have obtained recursions describing the change in the frequencies of the deleterious mutation (p_i) and robustness modifier (p_j) and the linkage disequilibrium (D_{ij}) over a single generation incorporating selection, mutation and transmission. Ultimately we are not interested in the dynamics of two alleles conferring different degrees of robustness. Rather, we wish to understand how robustness, as a phenotype, evolves. To achieve this we will make some additional assumptions. Firstly, we will consider that mutations at the robustness locus generate vanishingly small changes in robustness strategy. We will assume a continuum of strategies, $0 \leq z \leq 1$, and allow each strategy to be represented by a separate allele. The allele with strategy z has robustness effect $k[z]$ and incurs cost $c[z]$. We will generate a description for z^* , defined as the strategy whereby small variants about z^* will not invade a population playing strategy z^* , i.e. local ESS. To do this, we shall modify our multi-locus analysis in the following ways. The $X_j = 0$ allele plays strategy x , and hence generates an amount of robustness $k[x]$ and incurs a cost $c[x]$, whereas the $X_j = 1$ allele plays strategy $y = x + \Delta x$, where $\Delta x \ll 0$, giving $k[y] = k[x] + \Delta x k'[x]$ and $c[y] = c[x] + \Delta x c'[x]$. We will consider that the former allele is the population ‘resident’, and the latter allele is a vanishingly rare prospective ‘invader’.

Earlier, we assumed that μ was sufficiently small for us not to have to worry about fixation of the deleterious mutation. Since we now consider vanishing variation about the population mean robustness strategy, this condition can be expressed as $\mu < (1-k[x])s$. Given only minor robustness variants, the deleterious mutation should remain close to its equilibrium frequency, i.e. $p_i = \mu/(1-k[x]s) + \Delta p_i$, where Δp_i is a vanishingly small quantity. Making this substitution, and summarising the changes in the allele frequency at the robustness locus and linkage disequilibrium due to selection, mutation and recombination, obtains:

$$p_j^{\Delta} = \frac{(1-c[x])(1-k[x])\Delta(1-k[y])}{(1-c[x])(1-k[x])\Delta(1-c[x]-k[x]+c[x]k[x])} p_j + \frac{s(1-c[y])(1-k[y])(1-k[x])}{(1-c[x])(1-k[x])\Delta(1-c[x]-k[x]+c[x]k[x])} D_{ij} \quad [8.14]$$

$$D_{ij}^{\Delta} = \frac{(1-c[y])(k[x]-k[y])(1-r_e)\Delta((1-k[x])s-\Delta)}{(1-c[x])(1-k[x])^2 s(1-\Delta)} p_j + \frac{(1-c[y])(1-r_e)((1-k[x])(1-(1-k[y])s)-\Delta(k[x]-k[y]))}{(1-c[x])(1-k[x])(1-\Delta)} D_{ij}$$

Note that neither of these recursions are functions of Δp_i ; thus, we may set the frequency of the deleterious allele close to its equilibrium, and disregard its exact frequency. The above recursions may be summarised in matrix form, as $\mathbf{M}\cdot\mathbf{v} = \mathbf{v}$. The leading eigenvalue (λ) associated with the matrix gives the rate of increase of the rare robustness variant, and hence is variant's reproductive value. This is found by solving the characteristic equation $\text{Det}[\mathbf{M}-\lambda\mathbf{I}]=0$ where $\text{Det}[\mathbf{N}]$ is the determinant of matrix \mathbf{N} and \mathbf{I} is the 2×2 identity matrix. For the moment, we are interested in the intermediate ESS z^* , such that the minor variant is neutral ($\lambda=1$). The characteristic equation is then $\text{Det}[\mathbf{M}-\mathbf{I}] = 0$, which can be written down and is rather complicated. Substituting in $x = z^*$ and $y = z^* + \Delta x$ gives an expression of the form

$$F[c[z^*], c[\Delta z^*], k[z^*], k[\Delta z^*], \Delta, s, r_e] \Delta k + O[\Delta k^2] = 0 \quad [8.15]$$

Dropping the higher order terms of Δx , and noting that $\Delta x \neq 0$, we can solve for r , giving:

$$r_e = \frac{(1 - k[z^*])c[z^*](1 - k[z^*])s \Delta}{(1 - c[z^*])k[z^*]\Delta(1 - k[z^*])(1 - (1 - k[z^*])s)c[z^*]} = G. \quad [8.16]$$

The putative internal ESS is implicit in the equation $r_e = G$, and turns out to be difficult to explore directly. We note that there are no intermediate solutions ($0 < z^* < 1$) for $r_e = 0$. We can look at the relation between z^* and r_e through more indirect means. Translating the cost function by increasing $c[z]$ but holding $c'[z]$ fixed increases the cost of robustness. Partial differentiation of G with respect to c examines how r_e must change in order for z^* to remain fixed given the change in c . We have:

$$\frac{\partial G}{\partial c} = \frac{(1 - k[z^*])c[z^*]k[z^*]\Delta((1 - k[z^*])s \Delta)}{((1 - k[z^*])(1 - (1 - k[z^*])s)c[z^*]\Delta(1 - c[z^*])k[z^*]\Delta)^2}. \quad [8.17]$$

This is always positive for $0 < c[z^*], k[z^*] < 1$ and $c'[z^*], k'[z^*] > 0$. Hence, increasing the cost of robustness must be met with an increase in effective recombination rate. If we can accept *a priori* that increasing the cost of robustness will result in a decrease in the ESS z^* , then we can infer that increasing effective recombination rate facilitates the evolution of costly robustness, giving an increase in the ESS z^* . This is supported by numerical exploration of parameter space.

Expression [8.16] can be solved numerically to give the putative internal ESS z^* for any parameter set and cost and robustness functions – some examples are given in figures 8.1A and 8.2A. The assumption of vanishing variation is somewhat artificial, and so we have used simulations to test the predictions using a similar two-locus model which allows for a continuum alleles which are simultaneously extant (simulation results are presented in figures 8.1A and 8.2A). We find that the numerical solutions to the analytical prediction given by [8.16] and the results of the simulations are generally very good (figure 8.1A), although occasionally the simulations do exhibit qualitatively different behaviours. Often, the two agree for lower values of effective recombination

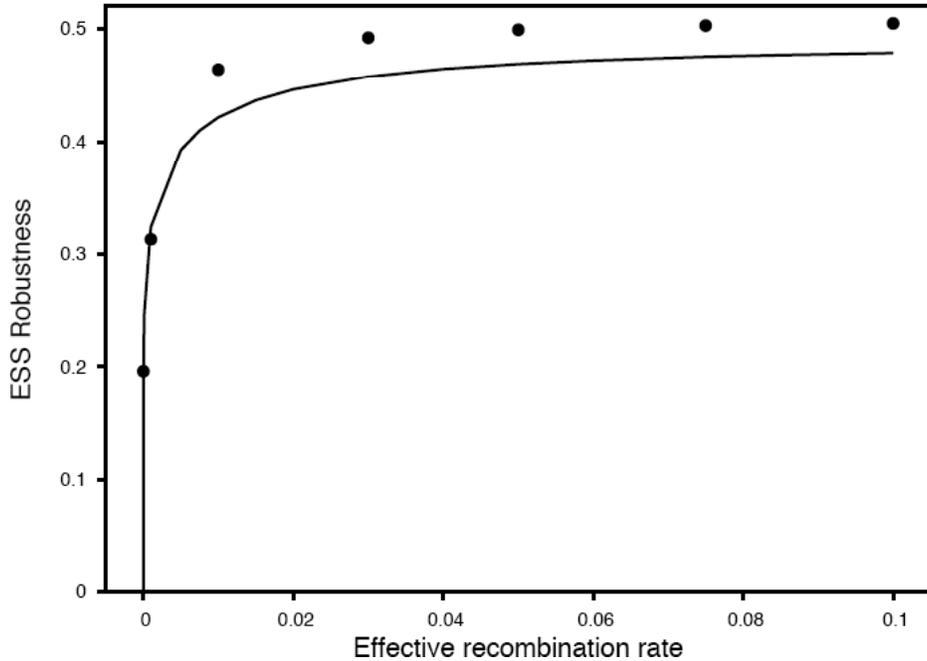
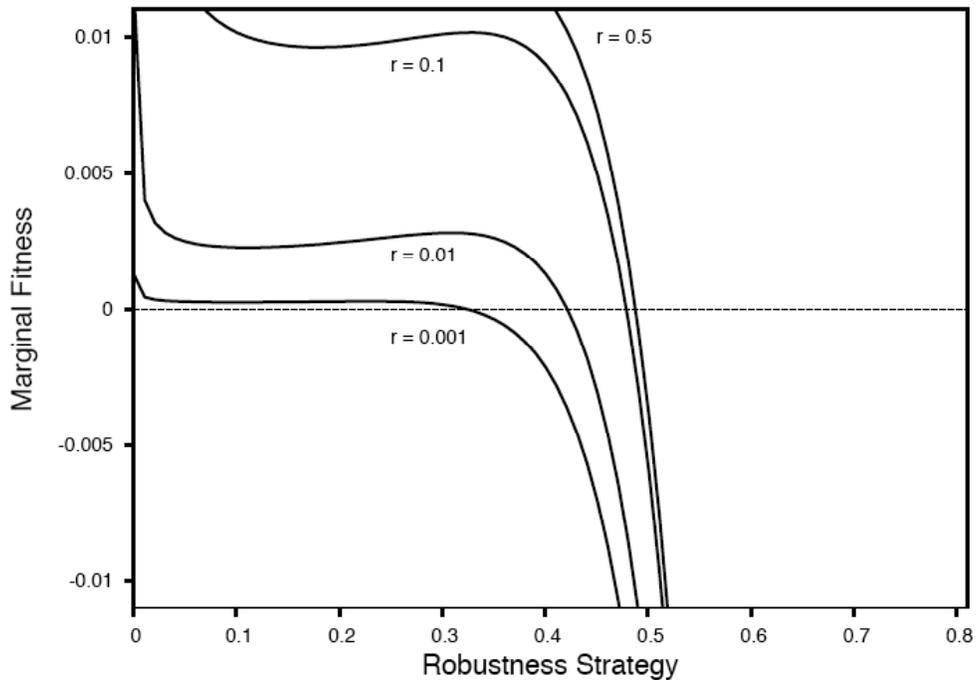


Figure 8.1. A (above) numerical solutions for equation [8.16] for cost and robustness functions $c[x]=x^{10}$ & $k[x]=x^{1/2}$, deleterious selection coefficient $s=0.1$, mutation rate $\mu=0.01$ and a range of effective recombination rate, r_e . Dots are results from simulations. B (below) numerical solutions to marginal fitness function, $d\bar{y}/dy$, evaluated at $y=x$. $d\bar{y}/dy>0$ indicates increased robustness is favoured, $d\bar{y}/dy<0$ indicates less robustness is favoured. Note that the internal stable equilibrium (ESS) is an increasing function of r_e .



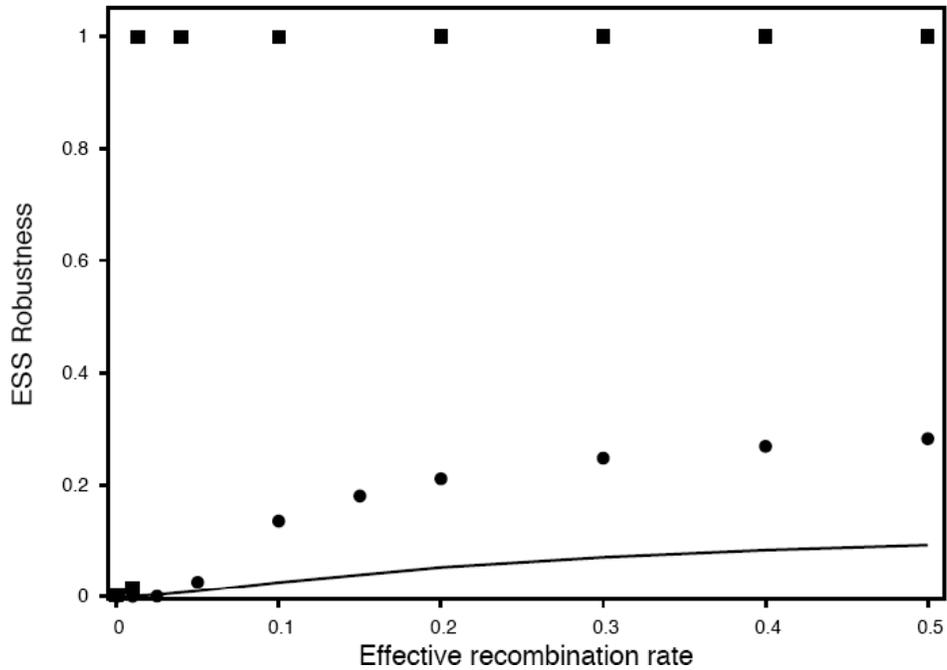
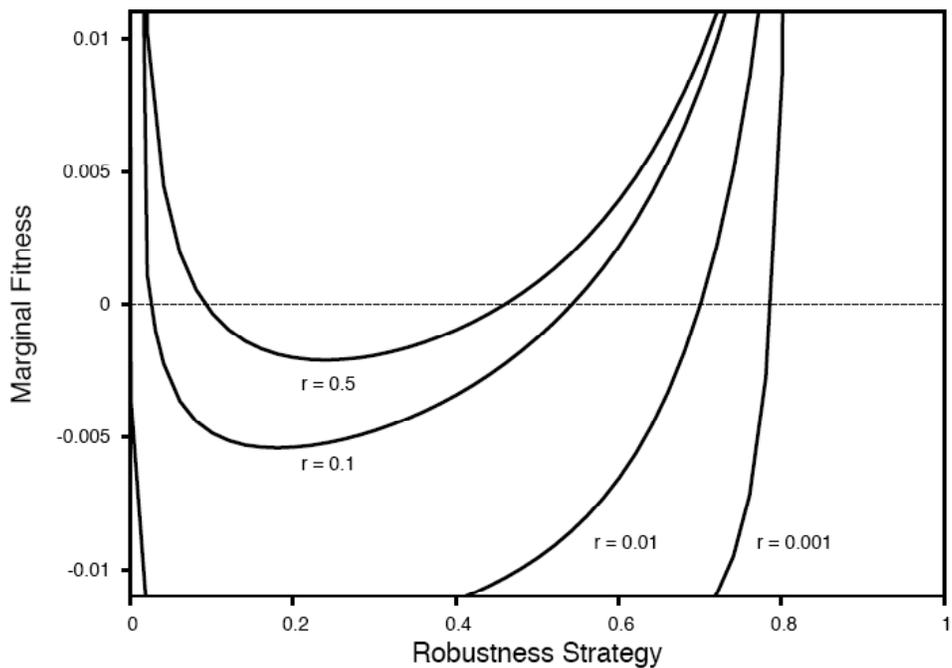


Figure 8.2. A (above) numerical solutions for equation [8.16] for cost and robustness functions $c[x]=0.015x$ & $k[x]=x^{1/2}$, deleterious selection coefficient $s=0.1$, mutation rate $\mu=0.01$ and a range of effective recombination rate, r_e . Data points are results from simulations initialised with vanishing robustness (dots), or mean robustness $x=0.5$ (squares). B (below) numerical solutions to marginal fitness function, $d\bar{x}/dy$, evaluated at $y=x$. $d\bar{x}/dy>0$ indicates increased robustness is favoured, $d\bar{x}/dy<0$ indicates less robustness is favoured. Note that the internal stable equilibrium (ESS) is an increasing function of r_e , and the unstable equilibrium is a decreasing function of r_e .



rate, beyond which the simulations exhibit a threshold behaviour (see, for example, figure 8.2A) with almost-complete robustness apparently being favoured.

As mentioned previously, the asymptotic rate of increase of the rare variant (λ) is the solution with the largest magnitude for the characteristic equation $\text{Det}[\mathbf{M}-\lambda\mathbf{I}] = 0$. Numerical inspection of the two solutions reveals that one consistently gives biologically plausible rates of increase for when small variants are assumed (i.e. $\lambda \approx 1$). Marginal fitness ($m = d\lambda/dy$, evaluated at $y = x$) can be determined analytically, and is a complicated function of the parameters and distributions. It reveals whether a population playing resident strategy x will favour the invasion of a slightly more robust variant (i.e., when $m > 0$) or whether lower robustness is favoured ($m < 0$). Setting r_e to zero, marginal fitness reduces to $c'[x]/(1-c[x])$, which is always negative and hence a reduction in robustness is favoured for all x . Thus, the ESS is zero robustness when the effective rate of recombination is zero.

Numerical investigation of marginal fitness reveals a positive-feedback effect, whereby increased robustness is often intrinsically favoured when the population exhibits a great deal of robustness (for example, figure 8.2B). This is due to the association between the more robust variant and the deleterious mutation being proportional to the effective selection coefficient of the deleterious mutation ($s_e = (1-k[x])s$). This association is one of the costs associated with robustness. When the resident strategy is robust (large k) then the effective selection coefficient is small, hence only a small association arises. Unless the cost of robustness is prohibitively large (as in figure 8.2B, which has the cost of robustness accelerating to $c \approx 1$ as $x \approx 1$), the marginal fitness may be increasing for high x , suggesting an ESS $z^* = 1$. This is in addition to the internal stable equilibrium at lower x . Which end point the population ultimately reaches is likely to be a function of its initial state. In the simulations we have initialised the population such that the population mean robustness strategy is 0.5. Examining figure 8.2A, we found that the ESS z^* was a step function of the effective recombination rate, with a threshold at $r_e \approx 0.013$.

Examining figure 8.2B, we find that for $r_e < 0.013$ the marginal fitness is negative at $x =$

0.5, whereas for $r_e > 0.013$ marginal fitness becomes positive at $x = 0.5$. Given that it is more plausible biologically for initial robustness to be low, we have re-done the simulations with initialisation such that 99% of individuals have zero robustness, and the remaining robustness alleles make up the remaining 1%. Here we find that the population gets stuck at the stable equilibrium, and the simulation results are in good agreement with our analytical prediction [8.16].

Discussion

We have modelled the evolution of costly mutational robustness in a simple two locus model for when recombination (r) between the two loci is intermediate. Previously, only the extremes of zero recombination (Hermisson et al 2002) and free recombination (Dawson 1999) have been considered. We have also incorporated relatedness (R) between mating partners (inbreeding), giving an ‘effective rate of recombination’ parameter ($r_e = r(1-R)$). An analytical statement relating the internal ESS robustness strategy (z^*) to the effective rate of recombination has been obtained. Consistent with previous theory, we find that costly robustness cannot be favoured when there is no recombination between the robustness locus and the loci that are the targets of the robustness. In addition, we show that, where one exists, the internal unbeatable robustness strategy is an increasing function of effective recombination rate. We have modelled the evolution of this two-locus system by introducing vanishing variants, one at a time, around the resident strategy. This artificial game theoretical approach appears to be justified, as the predictions find good support in simulated data which relaxes this assumption.

Why do we predict enhanced robustness with effective rate of recombination? The key to understanding this is to see robustness as a ‘selfish’ trait, having immediate benefits but ultimately overwhelming costs. It is similar to the classic tragedy of the commons (Hardin 1968) of the social evolution literature, whereby exploitation of a public good (driven by selfishness of individuals) leads to the destruction of that public good (which is a bad outcome for every member of the group). Social evolution theory reveals that self

restraint, which averts the tragedy, is increasingly favoured as individuals are more related, because the cost of selfish behaviour is increasingly paid by one's relatives, reducing inclusive fitness. Robustness provides an immediate benefit in reducing the harmfulness of a deleterious allele, but it leads to an accumulation of deleterious mutations. When relatedness between mating partners is high and recombination rates are low, this accumulation of deleterious alleles is focussed on one's own genome or in the genomes of relatives, reducing the selfish advantage of robustness. When relatedness is lower and recombination rates higher, the costs of excess deleterious alleles are suffered by the population in general and not by the selfish perpetrators in particular, leading to a relative fitness advantage for robust lineages.

Of central importance to this study is Haldane's (1937) mutation load invariant. Because the equilibrium mutation load ($L^* = \frac{U}{s}$) is invariant with respect to the selection coefficient (s) of the deleterious mutation, it cannot be alleviated (in the long term) by modifiers of robustness which reduce the magnitude of s . It is easy to see why the invariant exists – decreasing the deleterious effects of mutations reduces the efficacy with which natural selection removes them from the population, hence they become more frequent. Similar 'no pain, no gain' invariants are predicted for the cost of selection (given by the negative natural log of the initial frequency of the favoured allele, regardless of the strength of selection; Kimura 1961) and also in some simple models of parasite virulence (where parasites become more aggressive in their exploitation of the host as their impact on host mortality is reduced; e.g. Frank 1996a). As we have seen, evolving robustness does not in the long term improve the mean fitness of the population, as the equilibrium mutation load is invariant with respect to the selection coefficient of deleterious alleles. In fact, the mean fitness of the population is predicted to decline, as the costs of robustness remain after the short term benefits disappear. This being the case, the model predicts increased maladaptation in sexual / outbred genomes, whereas asexual / inbred genomes should be more efficient and less afflicted with the mutationally-decayed remains of robust networks.

Currently, no convincing empirical evidence has been published that demonstrates that genetic robustness exists as an adaptation. One reason for this is that, while it is possible to demonstrate that heritable variation is buffered in particular organisms, it is not easy to determine whether genetic robustness is the primary function or merely a side-effect of evolution for environmental robustness (Rutherford and Lindquist 1998, Queitsch *et al* 2002). In particular, experimental evolution of RNA molecules has shown that genetic robustness can result from direct selection for environmental robustness (Ancel and Fontana 2000, Burch and Chao 2004). However, the evolution of genetic robustness as a primary function may be plausible if there is migration between sub-populations in a heterogeneous environment (Stearns 2002). Migration rates can be much higher than mutation rates and therefore provide a stronger selective pressure for the buffering of maladapted alleles. It is with a view to extending the analysis to more complicated multi-locus models that we have employed the methodology of Barton and Turelli (1991) and Kirkpatrick *et al.* (2002), which permits arbitrary complexity within a single notational framework.

9. Discussion

Each of the chapters in this thesis contained their own extensive discussion. The aim of this chapter is to briefly review what has been achieved in each of the preceding chapters, and to highlight some emerging general points.

Chapter 2. Even more extreme fertility insurance and the sex ratios of protozoan blood parasites

In chapter 2, I examined a sex allocation problem – the trade-off between the production of male and female gametocytes in malaria and other protozoan blood parasites – where previous theory has achieved a rather poor fit with the empirical data. Specifically, much less female bias is observed than is predicted by standard local mate competition (LMC) theory, which assumes (1) limitless male fecundity and (2) large mating groups. The theory of fertility insurance, whereby female bias is curbed in order to ensure fertilization opportunities for these females when either male fecundity is limited or mating groups are small, has gone some way to explaining the disparity. In the context of protozoan blood parasites, both of these standard assumptions of LMC theory are often invalid, and so I have examined the implications for sex allocation when neither are met. I found that the interaction of these two pressures for fertility insurance causes a much smaller female bias than had previously been supposed. Empirical workers are now examining the importance of fertility insurance – for example Merino et al. (2004) show that antimalarial drugs lead to lower *Haemoproteus* density in blue tits, and that this is

associated with reduced female bias. Thus, the addition of some extra biological details have, in this instance, greatly increased the predictive power of LMC theory.

Chapter 3. A dimensionless invariant for relative size at sex change: explanations and implications

In chapter 3, I examined the opposite situation. In the context of the timing of sex change in sequential hermaphrodite animals, the predictions of sex allocation theory has had exceptional success, accounting for >90 % of the variation in relative of timing of sex change across several phyla, despite massive variation in supposedly relevant biological details, and several orders of magnitude in body size. I formalized the dimensionless theory underlying these predictions, generating a fitness function which is expressed in terms of the key dimensionless quantities ($\frac{M}{L}$, k/M and α , where α is age at maturity, M is instantaneous mortality rate, k is the Bertalanffy growth coefficient, and α is an exponent relating male size to fecundity) which are thought to underly the sex allocation strategy. I addressed recent criticism of the dimensionless approach in this biological context (Buston et al. 2004), and related to this I have highlighted the problems associated with generating null hypotheses for such theory. I also suggest that much of the criticism stems from a simple semantic disagreement as to what degree of invariance is expected from an invariant. As we are dealing with biology and not physics, clearly invariance is never absolute. Yet these striking near-invariant relationships remain, and are highly statistically significant. With this in mind, I suggest that the proper way forward is to employ the dimensionless approach to quantify the average values and

variation in the key underlying parameters, some of which will be very difficult or intrinsically impossible to measure directly. Using a sensitivity analysis I found that two of the three dimensionless parameters of the model appear to be relatively invariant ($\square M \approx 0.64 \pm 0.18$ s.d., $k/M \approx 0.96 \pm 0.45$ s.d.), while the third (\square) may vary considerably without affecting the invariance in relative timing of sex change.

Chapter 4. Spite and the scale of competition

In chapter 4, I related theory regarding the impact of localized competition on the evolution of altruism to a largely neglected theory of negative relatedness and the evolution of spiteful behaviours (Hamilton 1970, Grafen 1985a, Foster et al. 2001). The effects of local competition have been introduced into social evolutionary models in several ways, sometimes incorporating indirect competitive effects separately into Hamilton's (1963, 1964, 1970) rule (Grafen 1984, Frank 1998), and sometimes rescaling relatedness itself (Queller 1994), to recover a simple Hamilton's rule ($RB > C$). Several studies, employing the latter approach, have showed theoretically (Queller 1994) and experimentally (West et al. 2001b, Griffin et al. 2004) that local competition reduces the relatedness between social partners, and hence inhibits the evolution of altruism. I have extended this theory to show that local competition can facilitate the emergence of negative relatedness ($R < 0$) between social partners, such that spite ($B < 0$, $C > 0$) can be favoured by selection (i.e. $RB > C$). This development in spite theory allows re-interpretation of several social behaviours in terms of spite, allows us to make quantitative predictions of spite evolution, and suggests where spite may be favoured.

Spite should be looked for particularly where there is (1) strong competition between social partners and (2) the capacity for kin recognition, so as to avoid directing spite towards one's positive relations. Some examples include bacteriocin production in bacteria (Gardner et al. 2004, chapter 5) and the evolution of the sterile soldier caste in polyembryonic parasitoid wasps (Gardner & West 2004b, Appendix).

Chapter 5. Bacteriocins, spite and virulence

In chapter 5, I applied the theory of spite to chemical (bacteriocin) warfare in bacteria. Bacteriocin production entails production costs ($C > 0$) for the producer cell, which often has to commit suicide in the process. Bacteriocins have a toxic effect on neighbouring bacteria ($B < 0$). Their specificity, such that only conspecifics or closely related species are affected, suggests a role in interference competition. Additionally, due to linkage with an immunity locus, clones of the producer which also carry the gene will be unscathed. Hence, the system appears to satisfy the two requirements for spite: strong local competition and the capacity to kin discriminate. A simple model demonstrates that the evolutionary stable strategy (ESS; Maynard Smith & Price 1973) in terms of bacteriocin production is an increasing function of the intensity of local competition. Also, bacteriocin production is predicted to be a dome-shaped function with respect to the frequency of clonal kin within the range of bacteriocin attack. This is due to the need for both susceptible non-kin social partners (the victims) and immune kin (to benefit from the relaxed resource competition). This result is related to the evolution of virulence caused by parasitic bacteria. In the standard 'tragedy of the commons' (Hardin 1968)

models for the evolution virulence (Frank 1996a), the classic prediction is that increased kinship between parasites will favour low virulence. However, given a positive association between the vigour of bacterial growth and virulence, the bacteriocin model suggests a U-shaped relation between bacterial kinship and virulence. In this context, adding some extra biological detail generates a qualitatively different prediction. This may help explain why empirical support for the standard virulence theory is often lacking (Herre 1993, Frank 1996a, Griffin & West 2002).

Chapter 6. Cooperation and punishment, especially in humans

In chapter 6, I examined the problem of costly punishment. Punishment is invoked as an explanation for cooperation within large groups of unrelated individuals, for example in human society. However, the pressures underlying the evolution of punishment itself are obscure. A verbal argument (due to Sober & Wilson 1998) runs as follows: costly punishment is an altruistic public good, and just as other public goods are favoured by kin/group selection, punishment is favoured when there is relatedness between social partners. Since punishing is cheap relative to cooperation, it easily will evolve even when there is a small amount of relatedness between social partners, thus it will promote cooperation in contexts where cooperation will not itself be favoured. I found a flaw in this argument. The direct effect of punishing is to hurt oneself and one's social partner. 'Altruistic' punishment is therefore rather misleading, and perhaps the behaviour is better explained in terms of spite (for example, see Johnstone & Bshary 2004). This means that this model of punishment is fundamentally different from Frank's (1995b, 1996b, 2003a)

models of policing as a mechanism of repression competition. Policing prevents antisocial behaviour, and so directly benefits the group. An extended Hamilton's rule which incorporates the co-evolving punishment and cooperation traits revealed that any benefit for punishment is generated by a positive association between the punishment strategy of the focal individual and the cooperation strategy of the social partner. This sort of between-trait between-individual 'linkage disequilibrium' is typically ignored in social evolutionary models (for example, Frank 1995b) in order to achieve dynamic sufficiency. The extended Hamilton's rule suggested several possibilities: (1) increasing relatedness directly disfavours the evolution of punishment; (2) punishment may be favoured by niche construction (Odling-Smee 1996), if the cooperative environment that it favours is inherited by the descendants of punishers; (3) punishment appears to be more easily maintained than initially evolved, because punishers do very little punishing once cooperation is established (this was also noted by Boyd et al. 2003); and (4) the opposite is true for rewarding behaviour, suggesting a reason why punishing rather than rewarding is the norm in the natural world (for example, Clutton-Brock & Partker 1995). However, the dynamic insufficiency of the model means that these possibilities cannot properly be explored, leaving them as open questions for the future.

Chapter 7. Social evolution multi-locus methodology

In chapter 7, I borrowed a multi-locus methodology (developed by Barton & Turelli 1991, Kirkpatrick et al. 2002) from the population genetics literature, in order to analyse the co-evolution of social traits (as, for example, in the model of chapter 6) without

losing information about the associations between these traits. I showed that the multi-locus methodology intimately related to the foundations of social evolution theory: they share the same quantitative genetic approach to describing fitness, phenotypes and the causes of evolutionary change, and the multi-locus methodology emerges as a result of combining Price's (1970) theorem with a completely general notation of population composition and processes. Of key importance is the generalized conception of linkage disequilibrium, which the multi-locus methodology defines for any arbitrary set of gene positions, and hence also incorporates associations between individuals. I used the multi-locus notation to give a general account of how an extended Hamilton's rule (such as employed in chapter 6, and independently by Axelrod et al. 2004) can be constructed to describe social evolution for an arbitrary number of associated gene positions which contribute to an arbitrary number of traits. I emphasize that such details are implicitly allowed for in the original Hamilton's rule, which retains its mathematical truth. The methodology is then employed to analyse a dynamically-sufficient model for the co-evolution of cooperation and punishment, which lends weight to the conclusions of chapter 6.

Chapter 8. Recombination and the evolution of mutational robustness: a two-locus model

In chapter 8, I applied the multi-locus methodology to the evolution of mutational robustness. A two-locus model captured the tension between the benefits of robustness and the emerging cost due to carrying an increased load of mutations. Since the

equilibrium mutation load of the population is invariant with respect to robustness strategy, costly robustness is detrimental to the population in the long term. Previous work had focused on complete linkage (Hermisson et al. 2002) or freely segregating loci (Dawson 1999). I examined a condition which is necessarily satisfied by an intermediate ESS robustness strategy, and used this to show that increasing recombination rate and decreasing inbreeding rate promote the evolution of costly robustness. I related the tension between immediate benefit and delayed, group-level cost to the ‘tragedy of the commons’ (Hardin 1968, Frank 1998) literature in social evolution.

General remarks

Some general themes have emerged from these studies:

Social evolution is based on associations.

Fundamental to social evolution are behaviours which impact on the fitness of individuals other than the behavior. Of particular interest is when social partners are phenotypically and/or genetically correlated. Much attention has been given to genealogical closeness (kinship) as a source of such correlations. This has resulted in what might have been called ‘social selection’ being instead referred to as ‘kin selection’ (Maynard Smith 1964). This is unfortunate, as it obscures the generality of the social evolutionary framework, and the extensive applications of Hamilton’s (1963, 1964, 1970) rule. This point is made in chapter 6, and it is demonstrated that facultative behaviour can lead to

positive relatedness ($R > 0$) in Hamilton's rule, even in the absence of kinship. Chapter 6 also extends the concept of the social association beyond relatedness as it is currently defined, to allow for associations between different traits in different social partners. The result is a novel two-trait Hamilton's rule, which has made an independent appearance recently in the context of cooperation based on systems of co-evolving tags (Axelrod et al. 2004). Chapter 7 illustrated that relatedness, linkage disequilibrium and such between-trait between-individual associations can be understood within the same general framework. Hamilton's rule is shown to be an extremely subtle statement, which has implicitly allowed for such extensions since its proper derivation (Hamilton 1970).

Biological details sometimes matter.

Social evolution theory has enjoyed astonishing success, in terms of explaining the observed variation in social behaviours. Given this success, it is probable that poor empirical support in certain areas of social evolution theory are not due to a misunderstanding in how selection operates, but rather it is more likely that some crucial details of the system's biology have been overlooked. Noting that some of the biological assumptions of classical theory will not be valid in every circumstance, and developing the theory accordingly, can dramatically improve the explanatory power of the theory. For example, simultaneously relaxing the assumptions of limitless male fecundity and large mating groups allows for much improvement in the predictive power of local mate competition theory in explaining sex allocation in protozoan blood parasites (chapter 2). Such developments have huge practical importance as, for example, a quantitatively

accurate theory which relates malaria inbreeding rates to malaria sex ratios can conversely relate malaria sex ratios (easy to measure) to malaria inbreeding rates (difficult to measure), which can be important in understanding the epidemiology of this disease (Nee et al. 2002). Using sex allocation as inspiration, and noting that biological details do appear to make a huge difference in models of virulence evolution (for example, bacteriocin production in chapter 5, siderophore production as examined by West & Buckling 2003) there is hope that developing the theory of virulence evolution will eventually achieve more predictive success than it currently enjoys.

Biological details sometimes do not matter.

Studies such as Allsop & West's (2003a) description of a sex change invariant which holds across phyla imply that sometimes a great deal of variation in biological details does prevent a simple model from having terrific predictive power. Such invariant relationships should be exploited wherever they are found, as they potentially shed a great deal of light on the underlying biology. For example, in chapter 3 I examined how variation in three dimensionless parameters of the sex change model translates into variation in the relative size at sex change. Given the observed invariance in size at sex change, I was able to obtain estimates of these underlying parameters. Often such parameters will be difficult or impossible to measure directly, and so the existence of invariants provides an opportunity to measure these through indirect means. I would also argue that the invariant suggests the basic sex change model is correct, i.e. that we have correctly understood the trade-off between male and female reproductive function.

Don't worry too much about semantics.

Arguments over semantics can impede theoretical progress, and should be avoided. Classically, this is exemplified by the kin selection versus group selection debate, which still rumbles on (e.g. Keller 1999, Bergstrom 2002) despite simple mathematics showing that these are two sides of the same process (Price 1972a, Grafen 1984, Wade 1985, Frank 1986, Queller 1992, Hamilton 1975). As mentioned above, in chapter 3 I avoided being drawn in by arguments as to what constitutes an invariant in biology, and used a striking nearly-invariant relationship to examine the biology underlying sex change. In chapter 4 I re-examined the theory of spite evolution. There is room for much disagreement as to whether some or all or none of the behaviours mentioned in that chapter are really spiteful, as it is possible to reinterpret these in terms of altruism or spitefulness. I have chosen to use the framework of spite for several reasons: (1) the direct effects are losses in reproductive success to both actor and recipient, and so according to the standard classification (figure 1.1) we should call this spite (Trivers 1985); (2) it seems inappropriate to describe such behaviours as suicidally unleashing toxins on one's neighbours (bacteriocin production, chapters 4 and 5), and sacrificing any possibility of future reproduction in order to murder an embryonic host-mate (soldier caste in polyembryonic parasitic wasps, chapter 4 and Appendix 'Spite among siblings'), as altruism; (3) the key ingredient for Hamiltonian spite, negative relatedness (Hamilton 1970, Foster et al. 2001) naturally falls out, along with Hamilton's rule, from a direct fitness analysis of such behaviours (chapters 4 and 5), and leads to simple interpretation; and (4) this spiteful interpretation of the local competition results allow for greater consistency and clarity – for example, we can use as a rule-of-thumb that “local competition will tend to inhibit altruism and will tend to promote spite”, which is conceptually simpler than “local competition will inhibit some forms of altruism and promote some other forms of altruism”.

To conclude, social evolutionary biology owes much of its success to its firm, conceptually simple, theoretical underpinnings. The theory boasts a unifying framework -

centred around Price's theorem and Hamilton's rule - which is sufficiently general to address problems of arbitrary complexity. However, the work of social evolution theorists is by no means done. An appreciation for the subtlety of the paradigm is essential for: (1) development of simple, explicit models for social systems of interest; (2) ensuring rigorous theory-driven empirical research; and (3) making full use of our observations to contribute to a better understanding of nature and ourselves.

Bibliography

- Alexander, R. D. 1979. *Darwinism and human affairs*. University of Washington Press, Seattle.
- Alexander, R.D. 1987. *The biology of moral systems*. Aldine de Gruyter, New York.
- Allsop, D. J., & West, S.A. 2003a. Changing Sex at the Same Relative Body Size. *Nature* 425, 783 - 784.
- Allsop, D. J., & West, S.A. 2003b. Constant relative age and size at sex change in sequentially hermaphroditic fish. *J. Evol. Biol.* 16, 921-929.
- Allsop, D. J., & West, S.A. 2004a. Sex Ratio Evolution in Sex Changing Animals. *Evolution* 58, 1019 - 1027.
- Allsop, D. J., & West, S.A. 2004b. Sex allocation in the sex changing marine goby *Coryphopterus personatus* on atoll fringing reefs. *Evolutionary Ecology Research*. 6, 843-855.
- Allsop, D. J., & West, S.A. 2004c. Evolutionary biology - Sex change and relative body size in animals - Reply. *Nature* 428, 2.
- Ancel, L.W. & Fontana, W. 2000. Plasticity, evolvability and modularity in RNA. *J. Exp. Zool.* 288, 242-283.
- 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.

- Barron, A., Oldroyd, B.P. & Ratnieks, F.L.W. 2001. Worker policing and anarchy in *Apis*. *Behav Ecol Sociobiol* 50, 199-208.
- Barton, N. 2000. Genetic hitch-hiking. *Phil Trans Roy Soc* 355, 1553-1562.
- Barton, N.H. & Turelli, M. 1987. Adaptive landscapes, genetic distance, and the evolution of quantitative characters. *Genetical Research* 49, 157-174.
- Barton, N.H. & Turelli, M. 1991. Natural and sexual selection on many loci. *Genetics* 127, 229-255.
- Beeman, R.W., Friesen, K.S. & Denell, R.E. 1992. Maternal-effect selfish genes in flour beetles. *Science* 256, 89-92.
- Bergstrom, T.C. 2002. Evolution of social behavior: individual and group selection. *J Econ Perspect* 16, 67-88.
- Bertalanffy, L. v. 1938. A quantitative theory of organic growth (Inquiries on growth laws. II). *Human Biol.* 10, 181-213.
- Beverton, R. J. H. 1963. Maturation, growth and mortality of clepeid and engraulid stocks in relation to fishing. *Rapp. P. -V. Reun. Cons. Int. Explor. Mer.* 154, 44-67.
- Beverton, R. J. H. 1992. Patterns of reproductive strategy parameters in some marine teleost fishes. *J. Fish Biol.* 41:137-160.
- Boland, C. R. J., Heinsohn, R., & Cockburn, A. 1997. Deception by helpers in cooperatively breeding white-winged choughs and its experimental manipulation. *Behavioral Ecology and Sociobiology* 41, 251-256.

Bourke, A.F.G. 1988. Worker reproduction in the higher eusocial Hymenoptera. *Q Rev Biol* 63, 291-3112.

Boyd, R. 1982. Density-dependent mortality and the evolution of social interactions. *Animal Behaviour* 30, 972-982.

Boyd, R., & Richerson, P.J. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology* 13, 171-195.

Boyd, R., Gintis, H., Bowles, S., and Richerson, P.J. 2003. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the USA* 100, 3531-3535.

Bremerman, H.J. & Pickering, J. 1983. A game-theoretical model of parasite virulence. *Journal of Theoretical Biology* 100, 411-426.

Brown, J. H., West, G.B., & Enquist, B.J. 2000. Scaling in Biology: Patterns and Processes, Causes and Consequences. pp. 1 - 24 in J. H. Brown and G. B. West, eds. *Scaling in Biology*. Oxford University Press.

Brown, S.P., Hochberg, M.E. & Grenfell, B.T. 2002. Does multiple infection select for raised virulence? *Trends in Microbiology* 10, 401-405.

Brule, T., Deniel, C., Colas-Marrufo, T., & Sanchez-Crespo, M. 1999. Red grouper reproduction in the southern Gulf of Mexico. *Trans. Am. Fish. Soc.* 128, 385-402.

Buckingham, E. 1914. On physically similar systems: illustrations of the use of dimensional equations. *Phys. Rev.* 4, 345-376.

Burch, C.L. & Chao, L. 2004. Epistasis and its relationship to canalization in the RNA virus phi-6. *Genetics* 167, 559-567.

- Buss, L. W. 1987. *The evolution of individuality*. Princeton University Press, Princeton, N.J.
- Buston, P. M., Munday, P. L., & Warner, R. R. 2004. Evolutionary biology - Sex change and relative body size in animals. *Nature* 428, 1.
- Chao, L. & Levin, B.R. 1981. Structured habitats and the evolution of anticompetitor toxins in bacteria. *PNAS* 78, 6324-6328.
- Chao, L., Hanley, K.A., Burch, C.L., Dahlberg, C., Turner, P.E. 2000. Kin selection and parasite evolution: higher and lower virulence with hard and soft selection. *Quarterly Review of Biology* 75, 261-275.
- Chapuisat, M. & Keller, L. 1999. Testing kin selection with sex allocation data in eusocial Hymenoptera. *Heredity* 82, 473-478.
- Charlesworth, B. 2000. Book review: Levels of selection in evolution. *Heredity* 84, 493.
- Charnov, E. L. 1982a. *The Theory of Sex Allocation*. Princeton University Press, Princeton.
- Charnov, E. L. 1982b. Alternative Life-Histories in Protogynous Fishes: A General Evolutionary Theory. *Mar. Ecol.-Prog. Ser.* 9, 305-307.
- Charnov, E. L. 1991. Dimensionless numbers and the assembly rules for life histories. *0962-8436* 332, 41-48.
- Charnov, E. L. 1993. *Life History Invariants*. Oxford University Press, Oxford.

Charnov, E. L., & Berrigan, D. 1990. Dimensionless numbers and life history evolution: age of maturity versus the adult lifespan. *Evol. Ecol.* 4, 273-275.

Charnov, E. L., Gotshall, D. & Robinson, J. 1978. Sex ratio: Adaptive response to Population Fluctuations in Pandalid Shrimp. *Science* 200, 204-205.

Charnov, E.L. & Skúladóttir, U. 2000. Dimensionless invariants for the optimal size (age) of sex change. *Evolutionary Ecology Research* 2, 1067-1071.

Cheung, J., Danna, K., O'Connor, E., Price, L. & Shand, R. 1997. Isolation, sequence, and expression of the gene encoding halocin H4, a bacteriocin from the halophilic archaeon *Haloferax mediterranei* R4. *J. Bact.* 179, 548-551.

Christiansen, F.B. 1999. *Population genetics of multiple loci*. Wiley & Sons Ltd, Chichester.

Clutton-Brock, T. H. 1998. Reproductive skew, concessions and limited control. *Trends in Ecology & Evolution* 13, 288-292.

Clutton-Brock, T. H., & Parker, G.A. 1995. Punishment in animal societies. *Nature* 373, 209-216.

Clutton-Brock, T. H., Brotherton, P. N. M., Russell, A. F., O'Riain, M. J., Gaynor, D., Kinsky, R., Griffin, A., Manser, M., Sharpe, L., McIlrath, G.M., Small, T., Moss, A. & Monfort, S. 2001. Cooperation, control, and concession in meerkat groups. *Science* 291, 478-481.

Cook, J.M., Compton, S.G., Herre, E.A. & West, S.A. 1997. Alternative mating tactics and extreme male dimorphism in fig wasps. *Proc R Soc Lond B* 264, 747-754.

Crabtree, R. E., & Bullock, L. H. 1998. Age, growth, and reproduction of black grouper, *Mycteroperca bonaci*, in Florida waters. *Fish. Bull.* 96, 735-753.

Crow, J.F. & Kimura, M. 1965. Evolution in sexual and asexual populations. *American Naturalist* 99, 439-450.

Crow, J.F. & Kimura, M. 1970. *An introduction to population genetics theory*. New York, Harper & Row.

Czárán, T.L., Hoekstra, R.F. & Pagie, L. 2002. Chemical warfare between microbes promotes biodiversity. *PNAS* 99, 786-790.

Czárán, T.L. & Hoekstra, R.F. 2003. Killer-sensitive coexistence in metapopulations of micro-organism. *Proceedings of the Royal Society of London Series B – Biological Sciences* 270, 1373-1578.

Davies, C.M., Fairbrother, E. & Webster, J.P. 2002. Mixed strain schistosome infections of snails and the evolution of parasite virulence. *Parasitology* 124, 31-38.

Dawkins, R. 1976. *The selfish gene*. Oxford University Press.

Dawson, K.J. 1999. The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. *Theoretical Population Biology* 55, 1-22.

Day, T. & Burns, J.G. 2003. A consideration of patterns of virulence arising from host-parasite coevolution. *Evolution* 57, 671-676.

Day, T. & Taylor, P.D. 1997. von Bertalanffy's growth equation should not be used to model age and size at maturity. *American Naturalist* 149, 381-393.

- Denison, R. F. 2000. Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *American Naturalist* 156, 567-576.
- De Visser, J.A.G.M., Hermisson, J., Wagner, G.P., Meyers, L.A., Bagheri-Chaichian, H., Blanchard, J., Chao, Cheverud, J.M., L., Elena, S.F., Fontana, W., Gibson, G., Hansen, T.F., Krakauer, D., Lewontin, R.C., Ofria, C., Rice, S.H., von Dassow, G., Wagner, A., & Whitlock, M.C. 2003. Evolution and detection of genetic robustness. *Evolution* 57, 1959-1972.
- Ewens, W.J. 1989. An interpretation and proof of the fundamental theorem of natural selection. *Theoretical Population Biology* 36, 167-180.
- Edwards, A.W.F. 1994. The fundamental theorem of natural selection. *Biological Reviews* 69, 443-474.
- Fehr, E., & Gächter, S. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90, 980-994.
- Fehr, E., & Gächter, S. 2002. Altruistic punishment in humans. *Nature* 415, 137-140.
- Ferreira, B. P., & Russ, G. R. 1995. Population-Structure of the Leopard Coralgroup, *Plectropomus leopardus*, on Fished and Unfished Reefs Off Townsville, Central Great-Barrier-Reef, Australia. *Fish. Bull.* 93:629-642.
- Fisher, R.A. 1928. The possible modification of the response of the wild type to recurrent mutations. *American Naturalist* 62, 115-126.
- Fisher, R.A. 1930. *The genetical theory of natural selection*. Oxford University Press.
- Fisher, R.A. 1941. Average excess and average effect of a gene substitution. *Annals of Eugenics* 11, 53-63.

Foster, K.R., Gulliver, J. & Ratnieks, F.L.W. 2002. Why workers do not reproduce: worker policing in the European hornet *Vespa crabro*. *Insectes Sociaux* 49, 41-44.

Foster, K.R. & Ratnieks, F.L.W. 2000. Facultative worker policing in a wasp. *Nature* 407, 692-693.

Foster, K.R. & Ratnieks, F.L.W. 2001. Convergent evolution of worker policing by egg eating in the honey bee and common wasp. *Proc R Soc Lond B* 268, 169-174.

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

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

Frank, S. A. 1986. Hierarchical selection theory and sex ratios I. General solutions for structured populations. *Theoretical Population Biology* 29, 312-342.

Frank, S.A. 1992. A kin selection model for the evolution of virulence. *Proceedings of the Royal Society of London Series B – Biological Sciences* 250, 195-197.

Frank, S.A. 1994. Genetics of mutualism: the evolution of altruism between species. *Journal of Theoretical Biology* 170, 393-400.

Frank, S.A. 1995a. George Price's contributions to evolutionary genetics. *Journal of Theoretical Biology* 175, 373-388.

Frank, S.A. 1995b. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* 377, 520-522.

Frank, S.A. 1996a. Models of parasite virulence. *Quarterly Review of Biology* 71, 37-78.

Frank, S.A. 1996b. Policing and group cohesion when resources vary. *Animal Behaviour* 52, 1163-1169.

Frank, S.A. 1997a. The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* 51, 1712-1729.

Frank, S.A. 1997b. Cytoplasmic incompatibility and population structure. *Journal of Theoretical Biology* 184, 327-330.

Frank, S.A. 1998. *Foundations of social evolution*. Princeton University Press: Princeton.

Frank, S.A. 2002. A touchstone in the study of adaptation. *Evolution* 56, 2561-2564.

Frank, S.A. 2003a. Repression of competition and the evolution of cooperation. *Evolution* 57, 693-705.

Frank, S.A. 2003b. Genetic variation of polygenic characters and the evolution of genetic degeneracy. *Journal of Evolutionary Biology* 16, 138-142.

Frank, S.A. & Slatkin, M. Fisher's fundamental theorem of natural selection. *Trends in Ecology and Evolution* 7, 92-95.

Gandon, S., Mackinnon, M.J., Nee, S. & Read, A.F. 2001. Imperfect vaccines and the evolution of pathogen virulence. *Nature* 414, 751-756.

Ganusov, V.V. & Antia, R. 2003. Trade-offs and the evolution of virulence of microparasites: do details matter? *Theoretical Population Biology* 64, 211-220.

- Gardner, A. & West, S.A. 2004a. Cooperation and punishment, especially in humans. *American Naturalist* (available online)
- Gardner, A. & West, S.A. 2004b. Spite among siblings. *Science* 305. 1413-1414.
- Gardner, A. & West, S.A. in press. Spite and the scale of competition. *Journal of Evolutionary Biology* (available online).
- Gardner, A., West, S.A. & Buckling, A. 2004. Bacteriocins, spite and virulence. *Proc R Soc Lond B.*, 271, 1529-1535.
- Gemmill, A. W., Skorping, A., & Read, A. F. 1999. Optimal timing of first reproduction in parasitic nematodes. *J. Evol. Biol.* 12, 1148-1156.
- Ghiselin, M. T. 1969. The evolution of hermaphroditism amongst animals. *Quart. Rev. Biol.* 44, 189-208.
- Gillanders, B. M. 1995. Reproductive biology of the protogynous hermaphrodite *Achoerodus viridis* (Labridae) from south-eastern Australia. *Mar. Freshw. Res.* 46, 999-1008.
- Gintis, H. 2000. Strong reciprocity and human sociality. *Journal of Theoretical Biology* 206,169-179.
- Godfray, H.C.J. 1992. Strife among siblings. *Nature* 360, 213-214.
- Godfray, H. C. J. 1994. *Parasitoids. Behavioural and Evolutionary Ecology*. Princeton University Press, Princeton.
- Grafen, A. 1984. Natural selection, kin selection and group selection. In: Krebs, J.R. and Davies, N.B. (eds.), *Behavioural ecology*, 2nd edition. Blackwell: Oxford, pp 62-84.

Grafen, A. 1985a. A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology* 2, 28-89.

Grafen, A. 1985b. Hamilton's rule OK. *Nature* 318, 310-311.

Grafen, A. 1999. Formal Darwinism, the individual-as-maximising-agent analogy, and bet-hedging. *Proc R Soc Lond B* 266, 799-803.

Grafen, A. 2002. A first formal link between the Price equation and an optimization program. *Journal of Theoretical Biology* 217, 75-91.

Grafen, A. 2003. Fisher the evolutionary biologist. *The Statistician* 52, 319-329.

Grandcourt, E. M. 2002. Demographic characteristics of a selection of exploited reef fish from the Seychelles: preliminary study. *Mar. Freshw. Res.* 53, 123-130.

Grbic, M. Ode, P.J. & Strand, M.R. 1992. Sibling rivalry and brood sex ratios in polyembryonic wasps. *Nature* 360, 254-256.

Griffin, A.S. & West, S.A. 2002. Kin selection: fact and fiction. *TREE* 17, 15-21.

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

Haldane, J.B.S. 1924. A mathematical theory of natural and artificial selection part I. *Transactions of the Cambridge Philosophical Society* 23, 19-41.

Haldane, J.B.S. 1937. The effect of variation on fitness. *American Naturalist* 71, 337-349.

Haldane, J.B.S. 1964. A defense of beanbag genetics. *Perspect Biol Med* 19, 343-359.

Hamilton, W.D. 1963. The evolution of altruistic behaviour. *American Naturalist* 97, 354-356.

Hamilton, W.D. 1964. The genetical evolution of social behaviour I. *Journal of Theoretical Biology* 7, 1-16.

Hamilton, W.D. 1967. Extraordinary sex ratios. *Science* 156, 477-488.

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

Hamilton, W.D. 1971. Selection of selfish and altruistic behaviour in some extreme models. In: Eisenberg, J.F. and Dillon, W.S. (eds), *Man and beast: comparative social behaviour*. Smithsonian Press: Washington, DC. pp 57-91.

Hamilton, W.D. 1972. Altruism and related phenomena, mainly in social insects. *Annual Review of Ecology and Systematics* 3, 193-232.

Hamilton, W.D. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox, R. (ed), *Biosocial anthropology*. Malaby Press: London. pp 133-153.

Hamilton, W.D. 1979. Wingless and fighting males in fig wasps and other insects. In: Blum, M.S. & Blum, N.A. (eds), *Reproductive competition, mate choice, and sexual selection in insects*. Academic Press, New York. pp 167-220.

Hamilton, W.D. 1996. *Narrow roads of geneland volume 1: evolution of social behaviour*. Freeman: Oxford.

Hammond, R.L., Bruford, M.W. & Bourke, A.F.G. 2003. Ant workers selfishly bias sex ratios by manipulating female development. *Proc R Soc Lond B* 269, 173-178.

Hardin, G. 1968. The tragedy of the commons. *Science* 162, 1243-1248.

Hardy, I.C.W., Ode, P.J. & Strand, M.R. 1993. Factors influencing brood sex-ratios in polyembryonic hymenoptera. *Oecologia* 93, 343-348

Hartl, D.L. The physiology of weak selection. *Genome* 31, 183-189.

Harvey, P. H., & Pagel, M. D. 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press

Heinrich, J., & Boyd, R. 2001. Why people punish defectors. *Journal of Theoretical Biology* 208, 7989.

Hermisson, J., Redner, O., Wagner, H. & Baake, E. 2002. Mutation-selection balance: ancestry, load, and maximum principle. *Theoretical Population Biology* 62, 9-46.

Herre, E.A. 1987. Optimality, plasticity and selective regime in fig wasp sex-ratios. *Nature* 329, 627-629.

Herre, E.A. 1993. Population-structure and the evolution of virulence in nematode parasites of fig wasps. *Science* 259, 1442-1445.

Herre, E.A. 1995. Factors influencing the evolution of virulence: nematode parasites of fig wasps as a case study. *Parasitology* 111, S179-S191.

Herre, E.A., Machado, C. & West, S.A. 2001. Selective regime and fig wasp sex ratios: towards sorting rigor from pseudo-rigor in tests of adaptation. In: Orzack, S. & Sober, E. (eds), *Adaptation and optimality*. Cambridge University Press. pp 191-218.

Hirshleifer, D., & Rasmusen, E. 1989. Cooperation in a repeated prisoner's dilemma with ostracism. *Journal of Economic Behavior and Organization* 12, 87-106.

Hurst, G.D.D. & McVean, G.A.T. 1998. Selfish genes in a social insect. *Trends Ecol Evol* 13, 434-435.

Hurst L.D. 1991. The evolution of cytoplasmic incompatibility or when spite can be successful. *Journal of Theoretical Biology* 148, 269-277.

Hurst, L.D. 1993. *Scat+* is a selfish gene analogous to *Medea* of *Tribolium castaneum*. *Cell* 75, 407-408.

Hurst, L.D., Atlan, A. & Bengtsson, B.O. 1996. Genetic conflicts. *Q Rev Biol* 71, 317-364.

Johnstone, R. A. 2000. Models of reproductive skew: a review and synthesis. *Ethology* 106, 5-26.

Johnstone, R.A. & Bshary, R. 2004. Evolution of spite through indirect reciprocity. *Proc R Soc Lond B* 271, 1917-1922.

Keller, L. 1999. *Levels of selection in evolution*. Princeton University Press.

Keller, L. & Ross, K.G. 1998. Selfish genes: a green beard in the red fire ant. *Nature* 394, 573-575.

Kerr, B., Riley, M.A., Feldman, M.W. & Bohannan, B.J.M. 2002. Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* 418, 171-174.

Kiers, E. T., Rouseau, R. A., West, S. A., & Denison, R. F. 2003. Host sanctions and the legume-rhizobium mutualism. *Nature* 425, 7881.

Kimura, M. 1961. Natural selection as the process of accumulating genetic information in adaptive evolution. *Genetical Research* 2, 127-140.

Kimura, M. 1965. Attainment of quasi linkage equilibrium when gene frequencies are changing by natural selection. *Genetics* 52, 875-890.

Kirkpatrick, M., Johnson, T. & Barton, N.H. 2002. General models of multilocus evolution. *Genetics* 161, 1727-1750.

Kokko, H. 2003. Are reproductive skew models evolutionarily stable? *Proceedings of the Royal Society of London B* 270, 265-270. First citation in article

Kokko, H., Johnstone, R. A. & Clutton-Brock, T. H. 2001. The evolution of cooperative breeding through group augmentation. *Proceedings of the Royal Society of London B* 268, 187-196.

Kondrashov, A.S. & Crow, J.F. 1991. Haploidy or diploidy – which is better? *Nature* 351, 314-315.

Langer, P., Hogendoorn, K., & Keller, L. 2004. Tug-of-war over reproduction in a social bee. *Nature* 428, 844-847.

Leigh, E. G., Charnov, E. L., & Warner, R. R. 1976. Sex ratio, sex change and natural selection. *Proc. Natl. Acad. Sci.* 73, 3655-3660.

Lewontin, R.C. 1974. *The genetic basis of evolutionary change*. Columbia University Press, New York.

Lorenzo, J. M., Pajuelo, J. G., Mendez-Villamil, M., Coca, J. & Ramos, A.G. 2002. Age, growth, reproduction and mortality of the striped seabream, *Lithognathus mormyrus* (Pisces, Sparidae), off the Canary Islands (Central-east Atlantic). *J. Appl. Ichthyol.* 18, 204-209.

Mackie, M. 2000. Reproductive biology of the halfmoon grouper, *Epinephelus rivulatus*, at Ningaloo Reef, Western Australia. *Environ. Biol. Fishes* 57, 363-376.

Malécot, G. 1948. *Les mathematiques de l'heredite*. Masson, Paris.

Marino, G., Azzurro, E., Massari, A., Finoia, M. G. & Mandich, A. 2001. Reproduction in the dusky grouper from the southern Mediterranean. *J. Fish Biol.* 58, 909-927.

Maynard Smith, J. 1964. Group selection and kin selection. *Nature* 201, 1145-1147.

Maynard Smith, J. & Price, G.R. 1973. The logic of animal conflict. *Nature* 246, 15-18.

Maynard Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge.

Maynard Smith, J. & Szathmáry, E. 1995. *The major transitions in evolution*. Oxford University Press, Oxford.

McNamara, J. M., Webb, J. N., Collins, E. J., Székely, T. & Houston, A.I. 1997. A general technique for computing evolutionary stable strategies based on errors in decision-making. *Journal of Theoretical Biology* 189, 211-225.

Merino, S., Tomas, G., Moreno, J., Sanz, J.J. Arriero, E. & Folgueira, C. 2004. Changes in Haemoproteu sex ratios: fertility insurance or differential sex lifespan? *Proc R Soc Lond B* 271, 1605-1609.

- Michod, R.E. 1982. The theory of kin selection. *Ann Rev Ecol Syst* 13, 23-55.
- Millius, S. 2004. When to change sex. *Science News* 165, 40-41.
- Moller, A. & Jennions, M. 2002. How much variance can be explained by ecologists and evolutionary biologists? *Oecologia* 132, 492-500.
- Moran, P.A.P. On the nonexistence of adaptive topographies. *Ann Hum Genet* 27, 383-393.
- Mulder, R. A. & Langmore, N. E. 1993. Dominant males punish helpers for temporary defection in superb fairy wrens. *Animal Behaviour* 45, 830-833.
- Murray, M.G. 1987. The closed environment of the fig receptacle and its influence on male conflict in the old-world fig wasp, *Philotrypesis pilosa*. *Anim Behav* 35, 488-506.
- Nagelkerke, C. J. & Hardy, I. C. W. 1994. The influence of developmental mortality on optimal sex allocation under local mate competition. *Behav. Ecol.* 5, 401-411.
- Nagylaki, T. 1993. The evolution of multilocus systems under weak selection. *Genetics* 134, 627-647.
- Nakashima, Y., Kuwamura, T. & Yogo, Y. 1995. Why be a both ways sex changer? *Ethology* 101, 301-307.
- Nee, S. 1989. Does Hamilton's rule describe the evolution of reciprocal altruism? *Journal of Theoretical Biology* 141, 81-91.
- Nee, S., West, S. A. & Read, A. F. 2002. Inbreeding and parasite sex ratios. *Proc. Roy. Soc. Lond. B* 269, 755-760.

- Ode, P.J. & Strand, M.R. 1995. Progeny and sex allocation decisions of the polyembryonic wasp *Copidosoma floridanum*. *J. Anim. Ecol.* 64, 213-224.
- Odling-Smee, F. J., Laland, K. N. & Feldman, M. W. 1996. Niche construction. *American Naturalist* 147, 641-648.
- Oliver, P. 1980. Rewards and punishments as selective incentives for collective action: theoretical investigations. *American Journal of Sociology* 85:1356-1375.
- O'Neill, S.L., Hoffmann, A.A. & Werren, J.H. 1997. *Influential passengers: inherited microorganisms and arthropod reproduction*. Oxford University Press.
- Orlove, M.J. 1975. A model of kin selection not invoking coefficients of relationship. *Journal of Theoretical Biology* 49, 289-310.
- Osgood, S. M., Eisen, R. J. & Schall, J. J. 2002. Gametocyte sex ratio of a malaria parasite: experimental test of heritability. *J. Parasitol.* 88, 494-498.
- Ostrom, E. 1990. *Governing the commons*. Cambridge University Press, New York.
- Paperna, I. & Landau, I. 1991. Haemoprotozoans (Haemosporidia) of lizards. *Bull. Mus. Natl. Hist. Nat.* 13, 309-349.
- Passera, L. & Aron, S. 1996. Early sex discrimination and male brood elimination by workers of the Argentine ant. *Proc R Soc Lond B* 263, 1041-1046.
- Paul, R. E. L., Brey, P. T. & Robert, V. 2002. Plasmodium sex determination and transmission to mosquitoes. *Trends Parasitol.* 18, 32-38.
- Paul, R. E. L., Coulson, T. N., Raibaud, A. & Brey, P. T. 2000. Sex determination in malaria parasites. *Science* 287, 128-131.

- Paul, R. E. L., Raibaud, A. & Brey, P. T. 1999. Sex ratio adjustment in *Plasmodium gallinaceum*. *Parassitologia* 41, 153-158.
- Pauly, D. 1980. On the inter-relationships between natural mortality, growth parameters, and mean environmental temperature in 175 fish stocks. *Journal du Conseil* 39, 175 - 192.
- Pen, I. 2000. Reproductive effort in viscous populations. *Evolution* 54, 293-297.
- Pen, I. & Weissing, F.J. 2000. Towards a unified theory of cooperative breeding: the role of ecology and life history re-examined. *Proc R Soc Lond B* 267, 2411-2418.
- Pickering, J., Read, A. F., Guerrero, S. & West, S. A. 2000. Sex ratio and virulence in two species of lizard malaria parasites. *Evol. Ecol. Res.* 2, 171-184.
- Policansky, D. 1982. Sex Change in Plants and Animals. *Annu. Rev. Ecol. Syst.* 13, 471-495.
- Price, G.R. 1970. Selection and covariance. *Nature* 227, 520-521.
- Price, G. R. 1972a. Extension of covariance selection mathematics. *Annals of Human Genetics* 35, 485-490.
- Price, G.R. 1972b. Fisher's fundamental theorem made clear. *Ann Hum Genet* 36, 129-140.
- Prout, T. 1994. Some evolutionary possibilities for a microbe that causes incompatibility in its host. *Evolution* 48, 909-911.

Queitsch, C., Sangster, T.A. & Lindquist, S. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature* 417, 618—624.

Queller, D.C. 1984. Kin selection and frequency dependence: a game-theoretic approach. *Biol J Linn Soc* 23, 133-143.

Queller, D.C. 1985. Kinship, reciprocity, and synergism in the evolution of social behaviour. *Nature* 318, 366-367.

Queller, D.C. 1992. Does population viscosity promote kin selection? *Trends Ecol Evol* 7, 322-324.

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

Queller, D.C., Ponte, E., Bozzaro, S. & Strassmann, J.E. 2003. Single-gene greenbeard effects in the social amoeba, *Dictyostelium discoideum*. *Science* 299, 105-106.

Ratnieks, F.L.W. 1988. Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. *American Naturalist* 132, 217-236.

Ratnieks, F.L.W. & Visscher, P.K. 1989. Worker policing in the honeybee. *Nature* 342, 796-797.

Ratnieks, F.L.W., Monnin, T. & Foster, K.R. 2001. Inclusive fitness theory: novel predictions and tests in eusocial Hymenoptera. *Ann Zool Fennici* 38, 201-214.

Read, A. F., Narara, A., Nee, S., Keymer, A. E. & Day, K. P. 1992. Gametocyte sex ratios as indirect measures of outcrossing rates in malaria. *Parasitology* 104, 387-395.

Read, A. F., Smith, T. G., Nee, S. & West, S. A. 2002a. Sex ratios of malaria parasites and related protozoa. In *Sex Ratio Handbook* (ed. I. C. W. Hardy), pp. 314-332. Cambridge University Press, Cambridge.

Read, A.F., Mackinnon, M.J., Anwar, M.A. & Taylor, L.H. 2002b. Kin selection models as explanations of malaria. In *Virulence management: the adaptive dynamics of pathogen-host interactions* (U Dieckmann, JAJ Metz, MW Sabelis & K Sigmund eds), pp 165-178. Cambridge University Press.

Read, A.F. & Taylor, L.H. 2001. The ecology of genetically diverse infections. *Science* 292, 1099-1102.

Reece, S. E. & Read, A. 2000. Malaria sex ratios. *Trends Ecol. Evol.* 15, 259-260.

Reeve, H. K. 1992. Queen activation of lazy workers in colonies of the eusocial naked mole-rat. *Nature* 358, 147-149.

Reeve, H. K., & Gamboa, J. 1987. Queen regulation of worker foraging in paper wasps: a social feedback-control system (*Polistes fuscatus*, Hymenoptera, Vespidae). *Behaviour* 102,147-167.

Reeve, H. K., & Keller, L. 1999. Levels of selection: burying the units-of-selection debate and unearthing the crucial new issues. Pages 314 in L. Keller, ed. *Levels of selection in evolution*. Princeton University Press, Princeton, N.J.

Reeves, P. 1972. *The bacteriocins*. Springer-Verlag: NY.

Reinhold, K. 2003. Influence of male relatedness on lethal combat in fig wasps: a theoretical analysis. *Proc R Soc Lond B* 270, 1171-1175.

Riley, M.A., Goldstone, C.M., Wertz, J.E. & Gordon, D. 2003. A phylogenetic approach to assessing the targets of microbial warfare. *Journal of Evolutionary Biology* 16, 690-697.

Riley, M.A. & Gordon, D.M. 1999. The ecological role of bacteriocins in bacterial cooperation. *Trends in Microbiology* 7, 129-133.

Riley, M.A. & Wertz, J.E. 2002. Bacteriocins: evolution, ecology, and application. *Annual Review of Microbiology* 56, 117-137.

Robert, V., Read, A. F., Essong, J., Tchuinkam, T., Mulder, B., Verhave, J.-P. & Carnevale, P. 1996. Effect of gametocyte sex ratio on infectivity of *Plasmodium falciparum* to *Anopheles gambiae*. *Trans. Roy. Soc. Trop. Med. Hyg.* 90, 621-624.

Robertson, A. 1966. A mathematical model of the culling process in dairy *cattle*. *Anim Prod* 8, 95-108.

Robertson, D. R. 1972. Social control of sex reversal in a coral reef fish. *Science*. 177, 1007-1009.

Robertson, D. R., & Choat, J. H. 1974. Protogynous hermaphroditism and social systems in Labrid fish. *Proc. Second. Intl. Coral Reef Symposium* 1. 1, 217-225.

Robertson, D. R., & Warner, R. R. 1978. Sexual patterns in the labroid fishes of the western Caribbean, II: The Parrotfishes (Scaridae). *Smithsonian Contrib. Zool.* 255, 1-26.

Roughgarden, J. 1979. *Theory of population genetics and evolutionary ecology: an introduction*. Macmillan: New York.

Rutherford, S.L. & Lindquist, S. 1998. Hsp90 as a capacitor for morphological evolution. *Nature* 396, 336-342.

Schall, J. J. 1989. The sex ratio of Plasmodium gametocytes. *Parasitology* 98, 343-350.

Schall, J. J. 2000. Transmission success of the malaria parasite Plasmodium mexicanum into its vector: role of gametocyte density and sex ratio. *Parasitology* 121, 575-580.

Scharer, L. & Vizoso, D.B. 2003. Earlier sex change in infected individuals of the protogynous reef fish *Thalassoma bifasciatum*. *Behav. Ecol. Sociobiol.* 55, 137-143.

Schjorring, S. & Koella, J.C. 2003. Sub-lethal effects of pathogens can lead to the evolution of lower virulence in multiple infections. *Proceedings of the Royal Society of London Series B – Biological Sciences* 270, 189-193.

Schmalhausen, J.J. 1949. *Factors of evolution*. Blakiston, Philadelphia, PA.

Schmitt, M.J. & Breinig, F. 2002. The viral killer system in yeast: from molecular biology to application. *FEMS Microbiol. Rev.* 26, 257-276.

Seger, J. 1981. Kinship and covariance. *Journal of Theoretical Biology* 91, 191-213.

Seger, J. & Stubblefield, J.W. 1996. Optimization and adaptation. In 'Adaptation' (G.V. Lauder & M.R. Rose eds) pp 93-123. Academic Press.

Sell, J. & Wilson, R. K. 1999. The maintenance of cooperation: expectations of future interaction and the trigger of group punishment. *Social Forces* 77, 1551-1570.

Shapiro, D. Y. 1980. Role of Females in the Initiation of Sex Change in a Coral-Reef Fish. *Am. Zool.* 2, 826-826.

- Shapiro, D.Y. 1981. Size, maturation and the social control of sex reversal in the coral reef fish *Anthias squamipinnis* (Peters). *Journal of Zoology* 193, 105-128.
- Shapiro, D. Y., & Lubbock, R. 1980. Group sex ratio and sex reversal. *J. Theor. Biol.* 82, 411-426.
- Shutler, D., Bennett, G. F. & Mullie, A. 1995. Sex proportions of *Haemoproteus* blood parasites and local mate competition. *Proc. Natl. Acad. Sci. USA* 92, 6748-6752.
- Shutler, D. & Read, A. F. 1998. Local mate competition, and extraordinary and ordinary blood parasite sex ratios. *Oikos* 82, 417-424.
- Sigmund, K., Hauert, C. & Nowak, M. A. 2001. Reward and punishment. *Proceedings of the National Academy of Sciences of the USA* 98, 10757-10762.
- Skúladóttir, U., & Petursson, G. 1999. Defining populations of northern shrimp, *Pandalus borealis* (Kroyer 1938), in Icelandic waters using maximum length and maturity ogive of females. *Rit Fiskideildar* 16, 247-262.
- Sober, E. & Wilson, D. S. 1998. *Unto others: the evolution and psychology of unselfish behavior*. Harvard University Press, Cambridge, Mass.
- Stearns, S.C. 2002. Progress on canalization. *Proc Nat Acad Sci USA* 99, 10229-10230.
- Stephens, D.W. & Dunbar, S.R. 1993 Dimensional analysis in behavioural ecology. *Behavioural Ecology* 4, 172-183.
- Sundström, L., Chapuisat, M. & Keller, L. 1996. Conditional manipulation of sex ratios by ant workers: a test of kin selection theory. *Science* 274, 993-995.

Taylor, L. H. 1997. *Epidemiological and Evolutionary Consequences of Mixed-Genotype Infections of Malaria Parasites*. PhD thesis, University of Edinburgh.

Taylor, P. D. 1981. Intra-sex and inter-sex sibling interactions as sex-ratio determinants. *Nature* 291, 64-66.

Taylor, P.D. 1992a. Altruism in viscous populations – an inclusive fitness approach. *Evol Ecol* 6, 352-356.

Taylor, P.D. 1992b. Inclusive fitness in a heterogeneous environment. *Proc R Soc Lond B* 249, 299-302.

Taylor, P.D. 1996. Inclusive fitness arguments in genetic models of behaviour. *Journal of Mathematical Biology* 34, 654-674.

Taylor, P. D. & Bulmer, M. G. 1980. Local mate competition and the sex ratio. *J. theor. Biol.* 86, 409-419.

Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *Journal of Theoretical Biology* 180, 27-37.

Tobin, A. J., Sheaves, M. J. & Molony, B. W. 1997. Evidence of protandrous hermaphroditism in the tropical sparid *Acanthopagrus berda*. *J. Fish Biol.* 50, 22-33.

Trivers, R. L. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46, 35-57.

Trivers, R.L. & Hare, H. 1976. Haplodiploidy and the evolution of the social insects. *Science* 191: 249-263.

Trivers, R.L. 1985. *Social evolution*. Benjamin/Cummings, Menlo Park, CA.

- Turelli, M. 1994. Evolution of incompatibility-inducing microbes and their hosts. *Evolution* 48, 1500-1513.
- van Baalen, M. & Sabelis, M.W. 1995. The scope for virulence management – a comment on Ewald's view on the evolution of virulence. *Trends in Microbiology* 3, 414-416.
- Villamil, M. M., Lorenzo, J. M., Pajuelo, J. G., Ramos, A. G. & Coca, J. 2002. Aspects of the life history of the salema, *Sarpa salpa* (Pices, Sparidae), off the Canarian Archipelago (central-east Atlantic). *Environ. Biol. Fishes* 63, 183-192.
- Waddington, C.H. 1942. Canalization of development and the inheritance of acquired characters. *Nature* 150, 563-565.
- Wade, M.J. 1985. Soft selection, hard selection, kin selection and group selection. *American Naturalist* 125, 61-73.
- Wallace, B. 1968. Polymorphism, population size, and genetic load. In: Lewontin, R.C. (ed), *Population biology and evolution*. Syracuse University Press: Syracuse, NY. pp 87-108.
- Warner, R. R. 1984. Mating behavior and Hermaphroditism in Coral Reef Fishes. *American Scientist* 72, 128-136.
- Warner, R. R. 1988a. Sex Change in Fishes - Hypotheses, Evidence, and Objections. *Environ. Biol. Fishes* 22, 81-90.
- Warner, R. R. 1988b. Sex Change and the Size-Advantage Model. *Trends in Ecology & Evolution* 3, 133-136.

Warner, R. R., & Swearer, S.E. 1991. Social-Control of Sex-Change in the Bluehead Wrasse, *Thalassoma-Bifasciatum* (Pisces, Labridae). *Biol. Bull.* 181, 199-204.

Warner, R. R., Robertson, D. R. & Leigh, E. G. 1975. Sex change and Sexual Selection: The reproductive biology of a labrid fish is used to illuminate a theory of sex change. *Science.* 190, 633-638.

Warner, R. R., & Robertson, D. R. 1978. Sexual patterns in the labroid fishes of the western Carribean, I: The wrasses (Labridae). *Smithsonian Contrib. Zool.* 254, 1-27.

Wenseleers, T. & Ratnieks, F.L.W. 2001. Towards a general theory of conflict: the sociobiology of mendelian segregation. In “*Conflict from cell to colony*”, T. Wenseleers PhD thesis, University of Leuven, Belgium.

Wenseleers, T., Helantera, H., Hart, A.G. & Ratnieks, F.L.W. (2004) Worker reproduction and policing in insect societies. An ESS analysis. *Journal of Evolutionary Biology* 17, 1035-1047.

Werren, J.H. 1980. Sex-ratio adaptations to local mate competition in a parasitic wasp. *Science* 208, 1157-1159.

West, S.A. & Buckling, A. 2003. Cooperation, virulence and siderophore production in bacterial parasites. *Proc R Soc Lond B* 270, 37-44.

West, S. A. & Herre, E. A. 1998. Stabilizing selection and variance in fig wasp sex ratios. *Evolution* 52, 475-485.

West, S. A., Smith, T. G. & Read, A. F. 2000a. Sex allocation and population structure in apicomplexan (protozoa) parasites. *Proc. Roy. Soc. Lond. B* 267, 257-263.

West, S. A., Herre, E. A. & Sheldon, B. C. 2000b. The benefits of allocating sex. *Science* 290, 288-290.

West, S. A., Reece, S. E. & Read, A. F. 2001a. The evolution of gametocyte sex ratios in malaria and related apicomplexan (protozoan) parasites. *Trends Parasitol.* 17, 525-531.

West, S.A., Murray, M.G., Machado, C., Griffin, A.S. & Herre, E.A. 2001b. Testing Hamilton's rule with competition between relatives. *Nature* 409, 510-513.

West, S.A., Pen, I. and Griffin, A.S. 2002a. Cooperation and competition between relatives. *Science* 296, 72-75.

West, S. A., Smith, T. G., Nee, S. & Read, A. F. 2002b. Fertility insurance and the sex ratios of malaria and related hemosporein blood parasites. *J. Parasitol.* 88, 258-263.

West, S. A., Kiers, E. T., Simms, E. L. & Denison, R. F. 2002c. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proceedings of the Royal Society of London B* 269, 685-694.

West, S. A., Kiers, E. T., Pen, I. & Denison, R.F. 2002d. Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *Journal of Evolutionary Biology* 15, 830-837.

Wilson, D. S., Pollock, G. B. & Dugatkin, L. A. 1992. Can altruism evolve in purely viscous populations? *Evolutionary Ecology* 6, 331-341.

Wilson E.O. 1971. *The insect societies*. Harvard Press: Camb, Mass.

Wilson, E.O. 1975. *Sociobiology: the new synthesis*. Harvard Press: Cambridge, Mass.

Woodcock, S., & Heath, J. 2002. The robustness of altruism as an evolutionary strategy. *Biology and Philosophy* 17, 567-590.

Wright, S. 1922. Coefficients of inbreeding and relationship. *American Naturalist* 56, 330-338.

Wright, S. 1969. *Evolution and the genetics of populations II: the theory of gene frequencies*. University of Chicago Press, Chicago.

IS EVOLVABILITY INVOLVED IN THE ORIGIN OF MODULAR VARIATION?

ANDY GARDNER^{1,2} AND WILLEM ZUIDEMA^{1,3}

¹*Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, United Kingdom*

²*E-mail: Andy.Gardner@ed.ac.uk*

³*Language Evolution and Computation Research Unit, Theoretical and Applied Linguistics, University of Edinburgh, 40 George Square, Edinburgh EH8 9LL, United Kingdom*

Abstract.—Lipson et al. (2002) presented an elegant linear algebraic formalism to define and study the evolution of modularity in an artificial evolving system. They employed simulation data to support their suggestion that modularity arises spontaneously in temporally fluctuating systems in response to selection for enhanced evolvability. We show analytically and by simulation that their correlate of modularity is itself under selection and so is not a reliable indicator of selection for modularity per se. In addition, we question the relation between modularity and evolvability in their simulations, suggesting that this modularity cannot confer enhanced evolvability.

Key words.—Adaptability, canalization, fluctuating selection, pleiotropy, robustness.

Received January 22, 2003. Accepted January 27, 2003.

Modularity is a major principle of design and abounds in nature. Functional separation of modules—from eukaryote organelles to *Drosophila* limbs to human cognitive faculties—may give robustness to changing inputs and facilitate future improvement. The question of the evolutionary origins of such modularity is important and the recent simulation study of Lipson et al. (2002) is therefore a welcome contribution. They introduce a potentially extremely useful formalism that allows one to quantify modularity and study its evolutionary origins. Environmental variables are described by a vector **E**, and phenotypic traits by a vector **P**. A matrix **A**, which premultiplies **E** to give **P**, then describes the organismal process of transforming environmental input into phenotypic output.

Lipson et al. argue that the “blockiness” of **A** and its correlate, the number of zero elements, are measures of modularity. By assigning fitnesses to realized phenotypes depending on their distance from an arbitrarily chosen optimum, Lipson et al. (2002) study the evolution of modularity. Their simulations show that the frequency of zero elements in the matrices deviates from the expected value (1/3, the frequency of zero elements at initialization and among random mutations) when the environment changes rapidly. Lipson et al. attribute these results to a “second order (delayed) pressure for decomposition for adaptability,” (p.1554) that is, the uncoupling of traits to allow independent optimization of each and hence increased ability to adapt to new environments. Enhanced evolvability is concluded to be a cause, as well as a fortunate outcome, of the preponderance of zero-element-rich matrices. We disagree with this conclusion and believe that an alternative explanation exists. In addition, we feel that modularity cannot influence evolvability in their study.

In the simulations of Lipson et al., the element values of **E** are restricted to -1 and $+1$ and the element values of **A** are restricted to -1 , 0 , and $+1$. The elements of the phenotype vector **P** are therefore restricted to the range $-n \rightarrow n$, where n is the number of dimensions of the vectors (eight in the simulations of Lipson et al.). They restrict the elements of **F**, the arbitrary optimal phenotype, to -1 and $+1$. The optimal phenotypes are therefore restricted to a small subset of

all possible phenotypes, centered on the origin. We find that matrices with many zero elements tend to produce phenotypes that are closer to the zero vector, and therefore on average closer to the optimal phenotypes (mathematical details are given in the Appendix).

Rather than appealing to enhanced evolvability, the preponderance of zero-rich matrices can be explained by the advantage delivered to any **A** that can maintain a phenotype close to the origin, despite environmental perturbation (i.e., canalization; Waddington 1942). In Figure 1 we give the probability distribution of the value of an element of **P** as a function of ζ , the number of zero elements in the corresponding row of **A**. As ζ increases, the value of the focal element of **P** is more tightly distributed about the origin. Figure 2 reveals the relation between ζ and the mean scalar residual (negatively correlated with Lipson et al.’s measure of fitness) in a focal dimension: increasing ζ reduces the residual and thus increases fitness. Conducting simulations of our own, we have been able to demonstrate frequencies of zero elements significantly greater than 1/3, even when mutation is suppressed. Hence, individual lineages may thrive or decline, but cannot evolve and therefore cannot be under selection for enhanced evolvability (see Fig. 3 and Table 1).

Moreover, in the set-up of Lipson et al., it is unclear why enhanced evolvability is expected to play any role. Each element of the vector **P** is the result of (dot-) multiplying a separate row vector from **A** with **E**. Contrary to the suggestions of Lipson et al., manipulating the elements of such a row vector has no effect on the value of other elements of **P**. This means that when evolving **A** in the context of a certain environment **E** and a certain target phenotype **F**, every element of the actual phenotype **P** can be optimized independently. Interestingly, a different use of the same formalism was suggested by Lipson et al. and avoids this problem. Under this alternative scheme, vector **E** describes the genotype and matrix **A** describes the genetic architecture of the phenotype (e.g., pleiotropy), a framework similar to the multiple quantitative trait model proposed by Taylor and Higgs (2000). By allowing both **E** and **A** to evolve, one can study the evolution of modularity and evolvability under, for example, fluctuations in **F**.

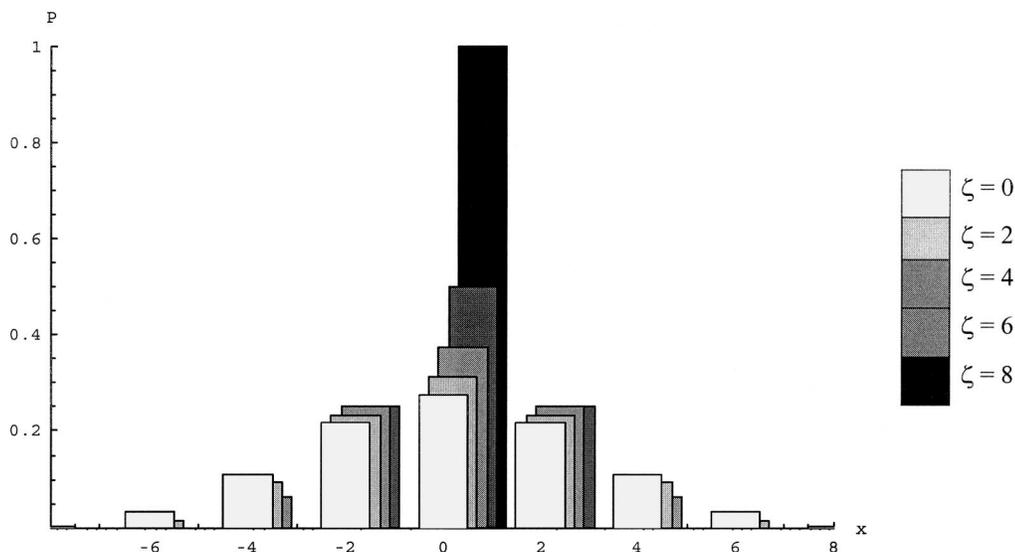


FIG. 1. The probability distribution of the value of \mathbf{P}_κ as a function of the number of zero elements in the κ^{th} row of the 8×8 ternary matrix \mathbf{A} , ζ . Here $n (= 8)$ and every value of $\zeta (= 0, 2, 4, 6, 8)$ are even, so the values of \mathbf{P}_κ are restricted to the set of even integers.

This is not to say that modularity is not under selection. It is possible that modularity confers robustness of fitness in response to the form of environmental change investigated by Lipson et al. When matrices are highly modular, such that there is a one-to-one correspondence between environmental characteristic and phenotypic trait, alteration of only one aspect of the environment will perturb the phenotype in one dimension only. Matrices that are less modular have environmental components each affecting more than one trait, and more than one trait being affected by several environmental components. They are therefore perturbed in multiple dimensions whenever a single aspect of the environment is altered. Because Lipson et al. change the sign of only one element of \mathbf{E} at each environmental alteration, it is conceivable that selection for fitness robustness has given rise to an increase in modularity in their simulations. However, this is

quite a different pressure than the supposed selection for enhanced evolvability.

In summary, Lipson et al. have presented an exciting and novel formalism that may yield quantitative, as well as qualitative insights into the evolution of evolvability and other problems. However, in their application of the model they have: (1) failed to demonstrate selection for modularity per se; and (2) not clearly established a link between modularity and evolvability. We suggest that enhanced evolvability can be neither a cause nor an outcome of the increase in their correlate of modularity.

ACKNOWLEDGMENTS

We thank N. Barton, A. Kalinka, and S. West for helpful discussion and comments on the manuscript, and H. Lipson for assistance in our re-creation of the evolutionary algorithm of the Lipson et al. (2002) paper. This work was supported

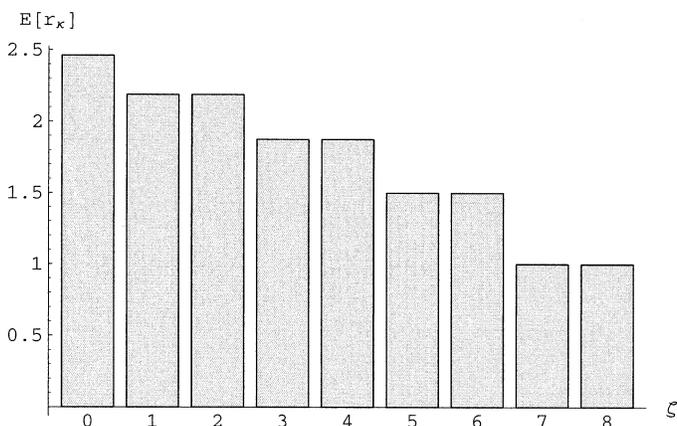


FIG. 2. The expectation of the residual r_κ as a function of ζ for an 8×8 ternary matrix. By ensuring that phenotypic vectors are more tightly distributed around the origin, and hence closer to the optimum, matrix rows with more zero elements achieve reduced residual, on average.

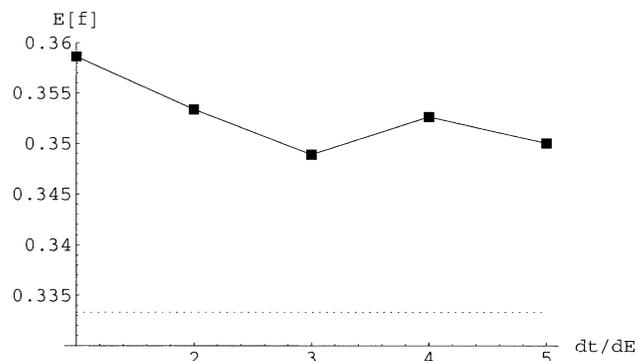


FIG. 3. The frequency of zero elements, averaged over 400 replicates, after 20 generations of evolution for a population of $50 \times 8 \times 8$ matrices over a range of rates of environmental change dt/dE . The broken line indicates the null prediction $1/3$. Simulations were devoid of mutation, but otherwise the evolutionary algorithm remained the same as that of Lipson et al.

TABLE 1. Simulation data and the one-tailed sign test for significant departure from null prediction “frequency of zero elements = 1/3”.

dt/dE	Mean frequency of zero elements (from 400 replicates)	No. of replicates (out of 400) with frequency of zero elements >1/3	P
1	0.359	268	4.700×10^{-12}
2	0.353	243	9.979×10^{-6}
3	0.349	233	5.639×10^{-4}
4	0.353	250	3.266×10^{-7}
5	0.350	228	2.946×10^{-3}

by the Biotechnology and Biological Sciences Research Council (AG) and a Marie Curie fellowship from the European Commission (WZ).

LITERATURE CITED

- Lipson, H., J. B. Pollack, and N. P. Suh. 2002. On the origin of modular variation. *Evolution* 56:1549–1556.
 Taylor, C. F., and P. G. Higgs. 2000. A population genetics model for multiple quantitative traits exhibiting pleiotropy and epistasis. *J. Theor. Biol.* 203:419–437.
 Waddington, C. H. 1942. Canalization of development and the inheritance of acquired characters. *Nature* 150:563–565.

Corresponding Editor: R. Harrison

APPENDIX

The Distribution of \mathbf{P}_κ

\mathbf{A} is a $n \times n$ ternary matrix (element values are -1 , 0 , and $+1$) and \mathbf{E} is a n -element column vector with element values $+1$ and -1 . The product of the premultiplication of \mathbf{E} by \mathbf{A} gives the phenotype vector \mathbf{P} . The κ^{th} element of \mathbf{P} is given by $\mathbf{P}_\kappa = \mathbf{A}_\kappa \cdot \mathbf{E} = \sum_i \mathbf{A}_{\kappa i} \cdot \mathbf{E}_i = \zeta \cdot 0 + m \cdot (+1) + (n - \zeta - m) \cdot (-1)$ where ζ is the number

of zero elements in \mathbf{A}_κ and $m \sim \text{Bin}(n - \zeta, 1/2)$ is the number of same-sign pairs of $\mathbf{A}_{\kappa i}$ and \mathbf{E}_i (i.e., those pairs of elements multiplying to give $+1$). Rearranging, the probability distribution of \mathbf{P}_κ is found to be

$$P[\mathbf{P}_\kappa = x] = \binom{n - \zeta}{\frac{n - \zeta - x}{2}} 2^{\zeta - n}, \quad (\text{A1})$$

for $n = 8$, the distribution of \mathbf{P}_κ as a function of ζ is shown in Figure 1.

$E[r_\kappa]$ as a function of ζ

Lipson et al. define fitness as a decreasing function of the (scalar) distance between realized phenotype \mathbf{P} and an arbitrary optimum \mathbf{F} . The residual in the κ^{th} dimension is $r_\kappa = |\mathbf{F}_\kappa - \mathbf{P}_\kappa|$ where \mathbf{F}_κ takes value $+1$ or -1 with equal probability. The probability density function of r_κ is then

$$\begin{aligned} P[r_\kappa = y] &= \frac{1}{2} P[|\mathbf{P}_\kappa| - 1 = y] + \frac{1}{2} P[|\mathbf{P}_\kappa| + 1 = y] \\ &= \frac{1}{2} (P[|\mathbf{P}_\kappa| = y + 1] + P[|\mathbf{P}_\kappa| = y - 1]). \end{aligned} \quad (\text{A2})$$

Because \mathbf{P}_κ is symmetrical about the origin, $P[\mathbf{P}_\kappa = z] = P[\mathbf{P}_\kappa = -z]$ and so for $z > 0$, $P[|\mathbf{P}_\kappa| = z] = 2 P[\mathbf{P}_\kappa = z]$, that is, for $y > 1$,

$$P[r_\kappa = y] = P[\mathbf{P}_\kappa = y + 1] + P[\mathbf{P}_\kappa = y - 1]. \quad (\text{A3})$$

For $y \leq 1$;

$$\begin{aligned} P[r_\kappa = 1] &= P[\mathbf{P}_\kappa = -2]P[F_\kappa = -1] + P[\mathbf{P}_\kappa = +2]P[F_\kappa = +1] \\ &\quad + P[\mathbf{P}_\kappa = 0] = P[\mathbf{P}_\kappa = +2] + P[\mathbf{P}_\kappa = 0] \\ P[r_\kappa = 0] &= P[\mathbf{P}_\kappa = -1]P[F_\kappa = -1] + P[\mathbf{P}_\kappa = +1]P[F_\kappa = +1] \\ &= P[\mathbf{P}_\kappa = +1]. \end{aligned} \quad (\text{A4})$$

Because $r_\kappa = \mathbf{P}_\kappa \pm 1$, and \mathbf{P}_κ is restricted to values of the same parity as $n - \zeta$, r_κ is only evaluated for those integers with parity opposite to $n - \zeta$. For $n = 8$, the mean of r_κ is revealed as a function of ζ in Figure 2.



Even more extreme fertility insurance and the sex ratios of protozoan blood parasites

A. Gardner*, S.E. Reece, S.A. West

Institute of Cell, Animal and Population Biology, Ashworth Laboratories, University of Edinburgh, Edinburgh EH9 3JT, UK

Received 27 August 2002; accepted 19 March 2003

Abstract

Theory developed for malaria and other protozoan parasites predicts that the evolutionarily stable gametocyte sex ratio (z^* ; proportion of gametocytes that are male) should be related to the inbreeding rate (f) by the equation $z^* = (1 - f)/2$. Although this equation has been applied with some success, it has been suggested that in some cases a less female biased sex ratio can be favoured to ensure female gametes are fertilized. Such fertility insurance can arise in response to two factors: (i) low numbers of gametes produced per gametocyte and (ii) the gametes of only a limited number of gametocytes being able to interact. However, previous theoretical studies have considered the influence of these two forms of fertility insurance separately. We use a stochastic analytical model to address this problem, and examine the consequences of when these two types of fertility insurance are allowed to occur simultaneously. Our results show that interactions between the two types of fertility insurance reduce the extent of female bias predicted in the sex ratio, suggesting that fertility insurance may be more important than has previously been assumed.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Fertility insurance; Local mate competition; Malaria; Sex allocation; Stochastic model

1. Introduction

One of the many successful applications of sex allocation theory has been the study of how competition for mates between related males can favour the evolution of female biased sex ratios (Charnov, 1982; Godfray, 1994; Hamilton, 1967; West et al., 2000a, b). Recent years have seen an increasing interest in applying this theory (local mate competition; LMC) to malaria and related protozoan parasites (Read et al., 2002; West et al., 2001). Here, the appropriate prediction is that the evolutionarily stable (ES; Maynard Smith, 1982) gametocyte sex ratio (z^* ; proportion of gametocytes that are male) should be related to the inbreeding rate (f) by the equation $z^* = (1 - f)/2$ (Hamilton, 1967; Nee et al., 2002; Read et al., 1992). When there is complete inbreeding ($f = 1$; i.e. a single lineage or clone is selfing), the ES strategy is to produce the minimum number of males required to fertilize the available female gametes and thus, maximize the number of zygotes. Conversely,

when gametes in the mating pool are of a mixture of lineages, f decreases and the sex ratio increases in order for each lineage to maximize its genetic representation in the zygote population. The relationship between the inbreeding rate and sex ratio has been able to explain a number of sex ratio patterns in Apicomplexan parasite populations (reviewed by West et al., 2001; Read et al., 2002). However, there are a number of observations that cannot be explained by this equation. In particular: (1) across *Haemoproteus* populations in birds the sex ratio does not correlate with an expected correlate of the inbreeding rate (prevalance; Shutler et al., 1995; Shutler and Read, 1998); (2) in malaria parasites, sex ratios within and between infections can be extremely variable (Osgood et al., 2002; Paul et al., 1999, 2000, 2002; Pickering et al. 2000; Schall, 1989; Taylor, 1997), and less female biased sex ratios can lead to greater transmission success (Robert et al., 1996).

A potential explanation for these contradictory observations is “fertility insurance”—the production of a less female biased sex ratio to ensure that all female gametes are fertilized (West et al., 2002). Before describing how fertility insurance can influence the ES sex ratio it is necessary to describe the background

*Corresponding author. Tel.: +44-01316505508; fax: +44-01316506465.

E-mail address: andy.gardner@ed.ac.uk (A. Gardner).

biology. In malaria and related Haemosporin parasites, haploid sexual stages (gametocytes) are taken up from the host in the blood meal of a vector. Once inside the midgut, the haploid gametocytes differentiate into haploid gametes and fuse to form zygotes. These resulting diploid zygotes undergo meiosis and asexual proliferation before migrating to the vector's salivary glands where they wait to enter a new vertebrate host. Each female gametocyte (macro-gametocyte) will differentiate into 1 female gamete, whereas each male gametocyte (micro-gametocyte) will produce several motile male gametes. The number of viable gametes produced per male gametocyte varies enormously across species—4–8 in mammalian malaria parasites (Read et al., 1992); ~ 2 in some lizard malarial (Schall, 2000); $5 \rightarrow 1000$ in *Eimeriorin* intestinal parasites (West et al., 2000a, b).

Fertility insurance can occur for two broad reasons—which are summarized here but discussed more fully in West et al. (2002). First, the number of male gametes produced per gametocyte (c) may be a limiting factor (Read et al., 1992). If the mean number of viable gametes produced per male gametocyte is c , then the ES sex ratio must be $z^* \geq 1/(c+1)$, otherwise there will not be enough male gametes to fertilize the female gametes (Fig. 1A; Read et al., 1992). Second, the ability of gametes to interact may be a limiting factor. West et al. (2002) investigated this possibility by assuming that the number of gametocytes whose gametes can interact (q) is restricted. In this case a less female biased sex ratio is favoured to avoid the stochastic absence of males in a mating group of q gametocytes (Fig. 1B; West et al., 2002). A low q could occur for a number of reasons including low male gamete motility, high gametocyte or gamete mortality, low gametocyte density, or small blood meals (Shutler and Read, 1998; Paul et al., 1999, 2000, 2002; Reece and Read, 2000; West et al., 2001, 2002). Recent attention has focused on how the host immune response may influence and vary the importance of these factors (Paul et al., 1999, 2000, 2002; Reece and Read, 2000).

In order to make their analyses mathematically tractable, previous studies have considered the influence of these two forms of fertility insurance separately. When examining the influence of male gametocyte fecundity (c), Read et al. (1992) assumed that the gametes from an infinite pool of gametocytes can interact ($q = \infty$), and when examining the influence of the number of gametocytes whose gametes can interact (q), West et al. (2002) assumed that male gamete fecundity was not a limiting factor ($c = \infty$; i.e. one male gametocyte is able to provide enough gametes to fertilize all of the female gametes in its mating group arising from q gametocytes). It has subsequently been assumed that the overall effect of these two factors can be examined by seeing which is more constraining, and

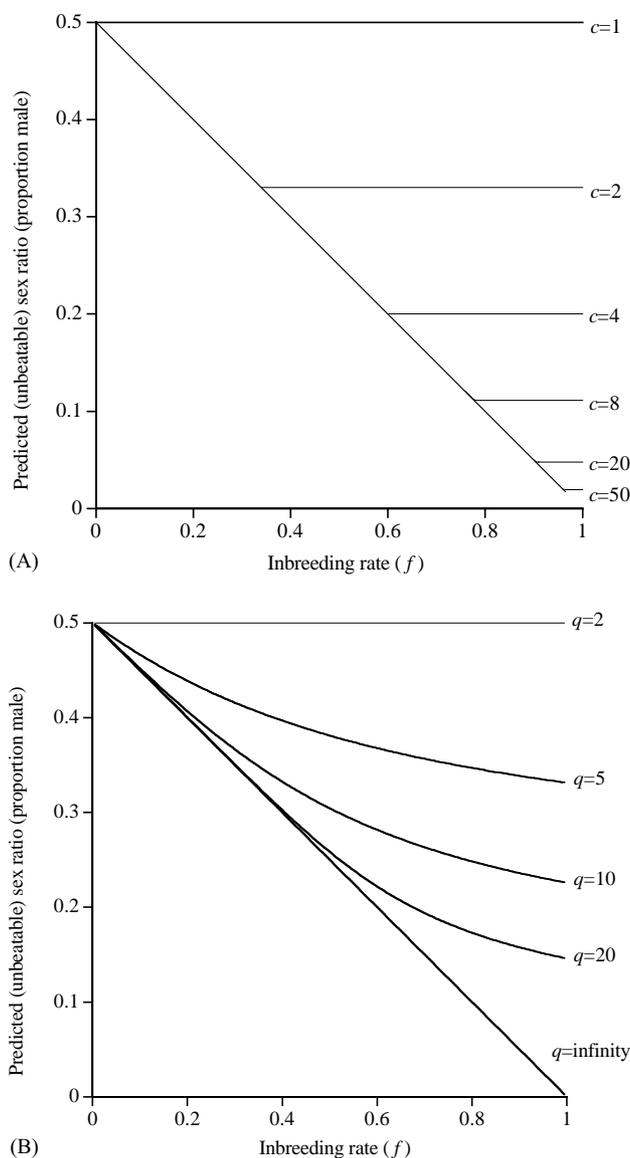


Fig. 1. The relationship between the predicted unbeatable sex ratio (proportion of gametocytes that are male; z^*) and the inbreeding rate (f). (A) Unbeatable sex ratio when the number of gametes produced by each male gametocyte (c) varies and gametes from all gametocytes in a very large group can interact ($q \rightarrow \infty$; Read et al. 1992). (B) Unbeatable sex ratio when the number of gametocytes whose gametes can interact (q) is limited and the number of gametes produced by each male gametocyte (c) is not limiting (West et al., 2002).

favours the least female biased sex ratio (West et al., 2002). However, there is the possibility that these factors may interact—when both c and q are low, even if there are males in a mating group, these males may not be able to provide enough gametes to fertilize all the female gametes. Although this scenario could logically occur, it is not clear whether this interaction will significantly influence the ES sex ratio. We use a stochastic analytical model to address this problem and consider how the unbeatable sex ratio is influenced by the interaction of finite values for both c and q . We use life history

terminology associated with malaria parasites, but our results are applicable to any Apicomplexan parasite with dimorphic sexual stages.

2. Methods

We consider an infinite population of vertebrates harbouring malaria parasites and supporting an infinite number of blood-feeding dipteran vectors (effects due to finite numbers of vertebrate hosts is negligible unless the number of hosts are extremely small; Taylor and Bulmer, 1980). Every host contains an infinite pool of haploid gametocytes circulating in the peripheral blood, comprising n independent lineages (all notation is given in Table 1). Within a lineage, all gametocytes are clonally derived from a single sporozoite founder individual. Each lineage produces a proportion z of male gametocytes and $1 - z$ of female gametocytes, where z is determined by a single biallelic nuclear gene. A common ‘Null’ allele exists at frequency $1 - m$ and

has $z = z_N$, and an infinitely rare ‘Mutant’ allele exists at frequency m and has $z = z_M$. We may assign each infected host individual to one of $n + 1$ classes on the basis of the number of Mutant lineages carried. Each host is fed upon by a large number of vectors, transmitting q gametocytes to each vector in the process. Once in the midgut of the vector, each male gametocyte gives rise to c male gametes and female gametocytes each give rise to a single female gamete. Random syngamy ensues, and the resulting next generation of zygotes are, following Read et al. (1992), assumed to reflect the genetic composition of the next generation of infections. It is worth noting that although each vector contains a single mating group of size q the predictions of this analysis will hold for any number of such groups, provided that there is no exchange of gametes between mating groups.

The fitness of the Null is the mean success of a Null lineage from each host-class weighted by the number of Null lineages in the host-class and the frequency of that host-class. As the mutant is infinitely rare, so that $m \rightarrow 0$, the fitness of the Null is dominated by its success in vectors feeding upon hosts containing no Mutant lineages

$$w_N \approx \frac{1}{n} S_{N,0} = f S_{N,0}, \quad (1)$$

where $S_{N,0}$ is the mean number of zygotic Null alleles produced per vector feeding on a host harbouring zero Mutant lineages, and f is the degree of inbreeding. The Mutant never occurs in such hosts, and almost never occurs in hosts with other Mutant lineages, so its fitness is dominated by its success in vectors feeding upon hosts with 1 Mutant lineage and $n - 1$ Null lineages

$$w_M \approx S_{M,1}, \quad (2)$$

where $S_{M,1}$ is the mean number of zygotic Mutant alleles derived from a vector feeding on a host containing one Mutant infection only. The Mutant invades if $w_M > w_N$ and so the ES sex ratio z^* is the value of z_N , such that $\omega = w_N/w_M$ is not less than unity for all $0 \leq z_M \leq 1$. Exact solutions for $S_{N,0}$ and $S_{M,1}$ will be determined, so that for known q , c and f pairs of sex ratio strategies may be compared.

A vector feeding on a Null-only host is assured of obtaining q Null gametocytes in its bloodmeal. $\mu_N \sim Bi(q, z_N, \cdot)$ are male, and the remaining $\phi_N = q - \mu_N$ are female, so that there are $c\mu_N$ male gametes and ϕ_N female gametes able to interact in the midgut. The number of zygotes, ζ , is the smaller of these two values, and since zygotes are diploid the number of Null alleles formed in that vector is 2ζ :

$$S_{N,0} = \sum_{\mu_N=0}^q \binom{q}{\mu_N} z_N^{\mu_N} (1 - z_N)^{q-\mu_N} 2 \min\{c\mu_N, q - \mu_N\}. \quad (3)$$

Table 1
Definition of each parameter/variable referred to in the methods and appendix

Symbol	Definition
$Bi(k, \pi)$	Binomial distribution: k trials and probability of success π
C	Number of viable male gametes per male gametocyte
F	Inbreeding coefficient; $f = n^{-1}$
g_X	Number of X-allele male gametes remaining viable
$HypGeo(\alpha, \beta, \gamma)$	Hypergeometric distribution: α trials, and β potential successes out of γ
M	The Mutant allele
M	Population frequency of the mutant
N	The Null allele
N	Number of independent lineages per vertebrate host
P	Probability of male gamete survival
Q	Number of gametocytes whose gametes can interact in the vector
$S_{X,y}$	Success of the X-allele in a host containing y Mutant infections
w_X	Absolute fitness of the X-allele
Z	Sex ratio (proportion male gametocytes per lineage)
z^*	Evolutionarily stable (ES) sex ratio
z_X	Sex ratio employed by the X-allele
λ	Species-specific number of gametes released per male gametocyte
ϕ_X	Number of X-allele females in a mating group
μ_X	Number of X-allele males in a mating group
τ_X	Total number of X-allele gametocytes in a mating group
$\varpi_{X,y}$	Frequency of X-alleles in successful male ($y = 1$) or female ($y = 0$) gametes
ω	Relative fitness of the Null, w_N/w_M ; Mutant invades if $\omega < 1$
ζ	Number of zygotes produced by the mating group

A vector feeding on a host containing 1 Mutant and $n - 1$ Null lineages will obtain q gametocytes of which $\tau_M \sim Bi(q, f)$ are Mutant and $\tau_N = q - \tau_M$ are Null. These will comprise $\mu_M \sim Bi(\tau_M, z_M)$ Mutant males and $\phi_M = \tau_M - \mu_M$ Mutant females, and $\mu_N \sim Bi(\tau_N, z_N)$ Null males and $\phi_N = \tau_N - \mu_N$ Null females. The number of zygotes, ζ , is then the lower of the two values $c(\mu_M + \mu_N)$ and $\phi_M + \phi_N$, meaning that there are ζ successful male gametes and ζ successful female gametes. Of the former, a proportion $\varpi_{M,1} \sim HypGeo(\zeta, c\mu_M, c(\mu_M + \mu_N))/\zeta$ will be Mutant, and of the latter a proportion $\varpi_{M,0} \sim HypGeo(\zeta, \phi_M, \phi_M + \phi_N)/\zeta$ will be Mutant. The success of the Mutant is simply $\zeta(\varpi_{M,1} + \varpi_{M,0})$ (Taylor, 1981; Charnov, 1982), i.e.:

$$S_{M,1} = \sum_{\tau_M=0}^q \sum_{\mu_M=0}^{\tau_M} \sum_{\mu_N=0}^{q-\tau_M} \binom{q}{\tau_M} f^{\tau_M} (1-f)^{q-\tau_M} \binom{\tau_M}{\mu_M} z_M^{\mu_M} (1-z_M)^{\tau_M-\mu_M} \binom{q-\tau_M}{\mu_N} z_N^{\mu_N} (1-z_N)^{q-\tau_M-\mu_N} \min\{c(\mu_M + \mu_N), q - \mu_M - \mu_N\} (E[\varpi_{M,1}] + E[\varpi_{M,0}]), \tag{4a}$$

where

$$E[\varpi_{M,1}] = \begin{cases} \frac{\mu_M}{\mu_M + \mu_N} & \text{if } \mu_M + \mu_N > 0, \\ 0 & \mu_M + \mu_N = 0, \end{cases} \tag{4b}$$

$$E[\varpi_{M,0}] = \begin{cases} \frac{\tau_M - \mu_M}{q - \mu_M - \mu_N} & \text{if } q - \mu_M - \mu_N > 0, \\ 0 & q - \mu_M - \mu_N = 0. \end{cases} \tag{4c}$$

These expressions reveal whether the Mutant allele can invade a population fixed for the Null. We determined the ES sex ratio iteratively, such that the value of z_N in each round is the sex ratio of the successfully invading Mutant or successfully defending Null of the previous round, and z_M is a randomly assigned value. After an indefinite number of rounds the Null will assume and subsequently retain the value of z^* , so that at any time the currently unbeaten z can be tested for evolutionary stability by plotting ω for z_N equal to the putative z^* against all $0 \leq z_M \leq 1$ and rejecting if $\omega < 1$ for any z_M .

To check our expressions, we derived Eqs. (3) and (4) for the special cases where q or c are infinite, i.e. corresponding to the analyses of Read et al. (1992) and West et al. (2002). These equations are presented in the appendix, and in all cases gave the same results as the previous analyses.

3. Results and discussion

We have discriminated between two types of fertility insurance, in response to: (i) low male gamete fertility (low c), and (ii) the ability of gametes to interact (low q). Previous theoretical work has examined the effect of these two types of fertility insurance separately. Specifically, West et al. (2002) assumed that when both of these factors are operating, the effect for sex ratio evolution can be determined by seeing which leads to a greater reduction in the predicted female bias (i.e. which of Figs. 1A and B predicts the least female biased sex ratio). In contrast, our model explicitly allows for both types of fertility insurance to act simultaneously, and hence allows for any interactions. In Figs. 2–4 we give

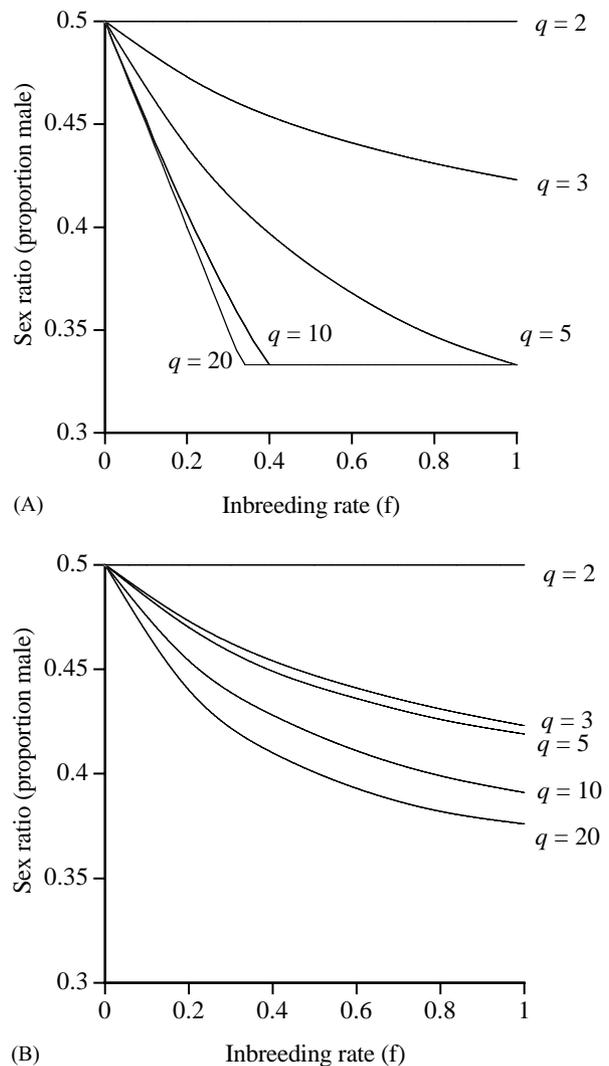
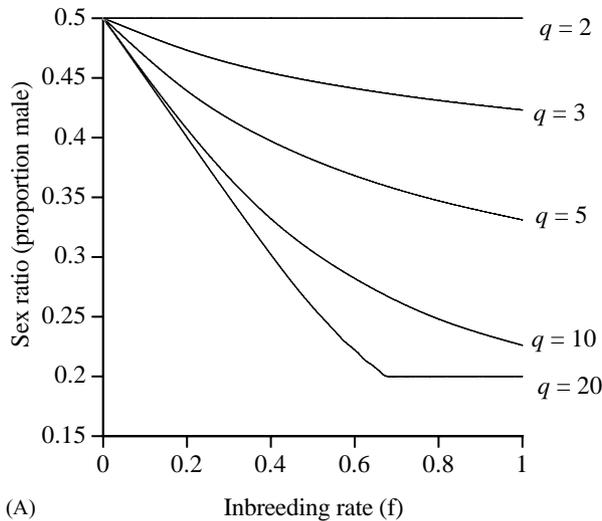
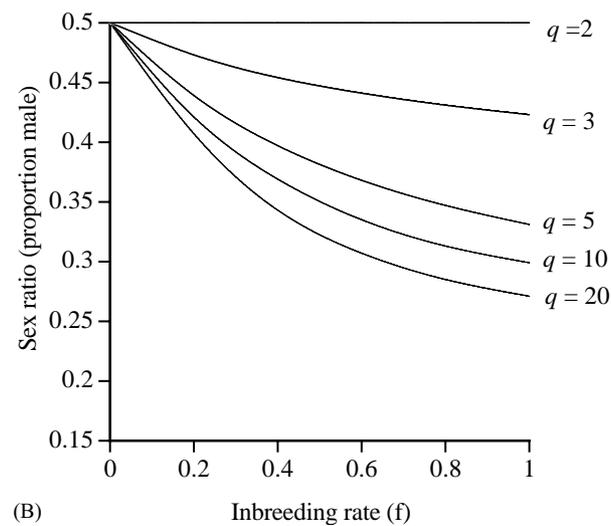


Fig. 2. (A) Relationship between predicted sex ratio and inbreeding rate, for given values of q when $c = 2$ assuming no interaction between the two types of fertility insurance and (B) relationship between ES sex ratio and inbreeding rate arising from Eqs. (1)–(4), for given values of q when $c = 2$.



(A)

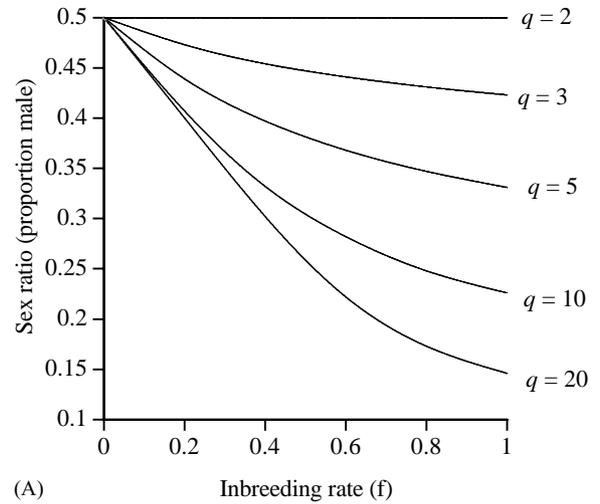


(B)

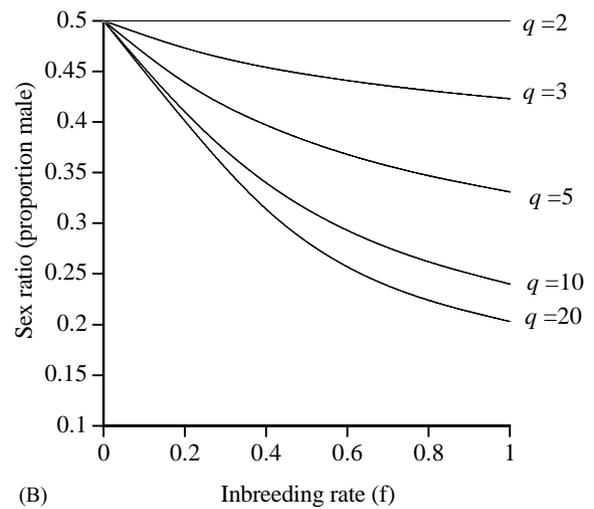
Fig. 3. (A) Relationship between predicted sex ratio and inbreeding rate, for given values of q when $c = 4$ assuming no interaction between the two types of fertility insurance and (B) Relationship between ES sex ratio and inbreeding rate arising from Eqs. (1)–(4), for given values of q when $c = 4$.

example predictions when the two types of fertility insurance are allowed to act separately as previously assumed by West et al. (2002) (part A of the figures) or simultaneously in our model (part B of the figures). Our results show that when both c and q are low, the ES sex ratio may be higher than predicted when considering these two effects separately.

Why does our model predict a less female biased sex ratio? It has been assumed that one male gametocyte will be able to provide enough gametes to fertilize all the female gametes in the mating group that arises from q gametocytes. This is not the case if $(q - 1) > c$. More generally, the male gametocytes will not be able to fertilize all the female gametes when $(q - \mu) > c\mu$, where μ is the number of male gametocytes in a mating group. This risk of not having enough males to fertilize the



(A)



(B)

Fig. 4. (A) Relationship between predicted sex ratio and inbreeding rate, for given values of q when $c = 8$ assuming no interaction between the two types of fertility insurance and (B) relationship between ES sex ratio and inbreeding rate arising from Eqs. (1)–(4), for given values of q when $c = 8$.

females in the group leads to less female biased sex ratios being favoured. Another way of conceptualising this is that a finite q increases the potential for low c to be a problem—when gametes can not interact as successfully (finite q), a mating group may contain only a single or small number of male gametocytes, and so the gamete fecundity (c) of these males is more likely to be a limiting factor.

Our model shows that the interaction between the two types of fertility insurance can have a surprisingly large influence on the ES sex ratio. In the examples that we give, the predicted sex ratio can be up to 0.1 higher (Fig. 2, when $c = 2$, $q = 10$ and $f = 0.3$). In this instance the sex ratio deviates from equality (0.5) by approximately half the amount inferred by West et al. (2002). Although increasing c proportionally reduces the degree of female bias, the complex interplay between male fecundity and

size of mating groups makes it difficult to relate the magnitude of this effect to q . In the limit, as q increases towards infinity, the effect dissipates as the predictions converge with those of Read et al. (1992). However, as q rises it increases the propensity for c to become limiting. The effect is therefore a dome-shaped function of q , although the exact relationship crucially depends upon the particular parameter values.

We also extended our model to allow stochastic variability in the number of viable gametes per gametocyte (c); see appendix, Eqs. (A.5) and (A.6). This could occur through variation in the number of gametes produced per gametocyte, or through mortality. Adding in this stochasticity (for invariant $E[c]$) gives further reduction in the female bias predicted, although this effect is negligible in all but the smallest of mating groups. However, a novel prediction arises from this form of stochasticity, as it allows the investigation of the mean value of $c < 1$, so that male fecundity is lower than that of females. In this case, a male biased sex ratio is favoured. For the case of $q \rightarrow \infty$ Eqs. (A.3) and (A.4) remain valid even for $c < 1$, and male biased ES sex ratios are easily demonstrated. Switching the roles of males and females in the classic LMC relation, the result of Read et al. (1992) can be extended so that, as before, for $c \geq 1$ $z^* = \max\{(1-f)/2, 1/(c+1)\}$, yet now for $c \leq 1$ $z^* = \min\{(1+f)/2, 1/(c+1)\}$. This prediction contrasts with standard LMC models constructed for insects (e.g. Nagelkerke and Hardy, 1994; West and Herre, 1998), where male biased sex ratios are never predicted, due to the assumption that one male can mate any number of females (analogous to assuming $c = \infty$). Male biased sex ratios have been observed in some samples of lizard malaria (Paperna and Landau, 1991), although the necessarily small sample sizes mean that these observations should be treated with caution.

To conclude, our analysis has revealed that fertility insurance can be a more potent evolutionary buffer to female biased sex ratios in malaria and related parasites than previously suggested. Clearly, the outstanding problem is to obtain empirical estimates of c and q , and how their values are influenced by factors such as host immune responses. We have recently reviewed the existing literature on this (West et al., 2002), and sadly very little is known.

Acknowledgements

We would like to thank Nick Barton, Brian Charlesworth, Toby Johnson, Rie Paul and Andrew Read for comments and useful discussion, and the Biotechnology & Biotechnological Sciences Research Council, the Natural Environment Research Council and The Royal Society for financial Support.

Appendix

In West et al. (2002) the implications of finite mating group size for fertility insurance were made amenable for mathematical treatment by assuming infinite male fecundity. This represents a special case of our model, such that $c = \infty$ and Eqs. (3) and (4) reduce to

$$S_{N,0} = \sum_{\mu_N=0}^q \binom{q}{\mu_N} z_N^{\mu_N} (1 - z_N)^{q-\mu_N} 2\zeta, \tag{A.1a}$$

where

$$\zeta = \begin{cases} q - \mu_N & \text{if } \mu_N > 0, \\ 0 & \mu_N = 0 \end{cases} \tag{A.1b}$$

and

$$S_{M,0} = \sum_{\tau_M=0}^q \sum_{\mu_M=0}^{\tau_M} \sum_{\mu_N=0}^{q-\tau_M} \binom{q}{\tau_M} f^{\tau_M} (1-f)^{q-\tau_M} \binom{\tau_M}{\mu_M} z_M^{\mu_M} (1 - z_M)^{\tau_M-\mu_M} \binom{q-\tau_M}{\mu_N} z_N^{\mu_N} (1 - z_N)^{q-\tau_M-\mu_N} \zeta (E[\varpi_{M,1}] + E[\varpi_{M,0}]), \tag{A.2a}$$

where

$$\zeta = \begin{cases} q - \mu_M - \mu_N & \text{if } \mu_M + \mu_N > 0, \\ 0 & \mu_M + \mu_N = 0, \end{cases} \tag{A.2b}$$

$$E[\varpi_{M,1}] = \begin{cases} \frac{\mu_M}{\mu_M + \mu_N} & \text{if } \mu_M + \mu_N > 0, \\ 0 & \mu_M + \mu_N = 0, \end{cases} \tag{A.2c}$$

$$E[\varpi_{M,0}] = \begin{cases} \frac{\tau_M - \mu_M}{q - \mu_M - \mu_N} & \text{if } q - \mu_M - \mu_N > 0, \\ 0 & q - \mu_M - \mu_N = 0. \end{cases} \tag{A.2d}$$

Conversely, in the deterministic analysis of Read et al. (1992), the fertility insurance consequences of limited male fecundity were investigated under the assumption of infinite mating group size. This special case, $q = \infty$, reduces Eqs. (3) and (4) to give

$$S_{N,0} = 2q \min\{cz_N, (1 - z_N)\} \tag{A.3}$$

and

$$S_{M,1} = q \min\{c(z_M f + z_N(1 - f)), (1 - z_M)f + (1 - z_N)(1 - f)\} \times \left(\frac{z_M f}{z_M f + z_N(1 - f)} + \frac{(1 - z_M)f}{(1 - z_M)f + (1 - z_N)(1 - f)} \right). \tag{A.4}$$

Although both $S_{N,0}$ and $S_{M,1}$ are linear functions of q , and therefore have infinite solutions, the relative fitness of the Null allele may still be evaluated as ω is the ratio of the two and hence is finite. The predictions converge with those of Read et al. (1992) for $c \geq 1$, but being more

general, are able to predict the male biased ES sex ratio when males fecundity is more limiting than that of females, so that $c < 1$.

We considered the possibility of stochastic male fecundity, specifically, how accurately do expressions (3) and (4) predict the ES sex ratio when the value of c represents the expectation of a random variable? Assuming that males all produce the same species-specific number (χ) of gametes, each with independent probability p of being viable for fertilization, Eqs. (3) and (4) become

$$S_{N,0} = \sum_{\mu_N=0}^q \sum_{g_N=0}^{\chi \mu_N} \binom{q}{\mu_N} z_N^{\mu_N} (1 - z_N)^{q - \mu_N} \binom{\chi \mu_N}{g_N} p^{g_N} (1 - p)^{\chi \mu_N - g_N} 2 \min\{g_N, q - \mu_N\} \quad (\text{A.5})$$

and

$$S_{M,1} = \sum_{\tau_M=0}^q \sum_{\mu_M=0}^{\tau_M} \sum_{\mu_N=0}^{q - \mu_M} \sum_{g_M=0}^{\chi \mu_M} \sum_{g_N=0}^{\chi \mu_N} \binom{q}{\tau_M} f^{\tau_M} (1 - f)^{q - \tau_M} \binom{\tau_M}{\mu_M} z_M^{\mu_M} (1 - z_M)^{\tau_M - \mu_M} \binom{q - \tau_M}{\mu_N} z_N^{\mu_N} (1 - z_N)^{q - \tau_M - \mu_N} \binom{\chi \mu_M}{g_M} \binom{\chi \mu_N}{g_N} p^{g_M + g_N} (1 - p)^{\chi(\mu_M + \mu_N) - g_M - g_N} \min\{g_M + g_N, q - \mu_M - \mu_N\} (E[\varpi_{M,1}] + E[\varpi_{M,0}]), \quad (\text{A.6a})$$

where

$$E[\varpi_{M,1}] = \begin{cases} \frac{g_M}{g_M + g_N} & \text{if } g_M + g_N > 0, \\ 0 & g_M + g_N = 0, \end{cases} \quad (\text{A.6b})$$

$$E[\varpi_{M,0}] = \begin{cases} \frac{\tau_M - \mu_M}{q - \mu_M - \mu_N} & \text{if } q - \mu_M - \mu_N > 0, \\ 0 & q - \mu_M - \mu_N = 0. \end{cases} \quad (\text{A.6c})$$

References

Charnov, E.L., 1982. The Theory of Sex Allocation. Princeton University Press, Princeton, NJ.
 Godfray, H.C.J., 1994. Parasitoids. Behavioural and Evolutionary Ecology. Princeton University Press, Princeton, NJ.
 Hamilton, W.D., 1967. Extraordinary sex ratios. Science 156, 477–488.
 Maynard Smith, J., 1982. Evolution and the Theory of Games. Cambridge University Press, Cambridge.
 Nagelkerke, C.J., Hardy, I.C.W., 1994. The influence of developmental mortality on optimal sex allocation under local mate competition. Behav. Ecol. 5, 401–411.

Nee, S., West, S.A., Read, A.F., 2002. Inbreeding and parasite sex ratios. Proc. R. Soc. London B 269, 755–760.
 Osgood, S.M., Eisen, R.J., Schall, J.J., 2002. Gametocyte sex ratio of a malaria parasite: experimental test of heritability. J. Parasitol. 88, 494–498.
 Paperna, I., Landau, I., 1991. *Haemoproteus* (haemosporidia) of lizards. Bull. Mus. Natl. Hist. Nat. 13, 309–349.
 Paul, R.E.L., Raibaud, A., Brey, P.T., 1999. Sex ratio adjustment in *Plasmodium gallinaceum*. Parasitologia 41, 153–158.
 Paul, R.E.L., Coulson, T.N., Raibaud, A., Brey, P.T., 2000. Sex determination in malaria parasites. Science 287, 128–131.
 Paul, R.E.L., Brey, P.T., Robert, V., 2002. *Plasmodium* sex determination and transmission to mosquitoes. Trends Parasitol. 18, 32–38.
 Pickering, J., Read, A.F., Guerrero, S., West, S.A., 2000. Sex ratio and virulence in two species of lizard malaria parasites. Evol. Ecol. Res. 2, 171–184.
 Read, A.F., Narara, A., Nee, S., Keymer, A.E., Day, K.P., 1992. Gametocyte sex ratios as indirect measures of outcrossing rates in malaria. Parasitology 104, 387–395.
 Read, A.F., Smith, T.G., Nee, S., West, S.A., 2002. Sex ratios of malaria parasites and related protozoa. In: Hardy, I.C.W. (Ed.), Sex Ratio Handbook. Cambridge University Press, Cambridge, pp. 314–332.
 Reece, S.E., Read, A., 2000. Malaria sex ratios. Trends Ecol. Evol. 15, 259–260.
 Robert, V., Read, A.F., Essong, J., Tchuinkam, T., Mulder, B., Verhave, J.-P., Carnevale, P., 1996. Effect of gametocyte sex ratio on infectivity of *Plasmodium falciparum* to *Anopheles gambiae*. Trans. R. Soc. Trop. Med. Hyg. 90, 621–624.
 Schall, J.J., 1989. The sex ratio of *Plasmodium* gametocytes. Parasitology 98, 343–350.
 Schall, J.J., 2000. Transmission success of the malaria parasite *Plasmodium mexicanum* into its vector: role of gametocyte density and sex ratio. Parasitology 121, 575–580.
 Shutler, D., Read, A.F., 1998. Local mate competition, and extraordinary and ordinary blood parasite sex ratios. Oikos 82, 417–424.
 Shutler, D., Bennett, G.F., Mullie, A., 1995. Sex proportions of *Haemoproteus* blood parasites and local mate competition. Proc. Natl Acad. Sci. USA 92, 6748–6752.
 Taylor, L.H., 1997. Epidemiological and evolutionary consequences of mixed-genotype infections of malaria parasites. Ph.D. Thesis, University of Edinburgh.
 Taylor, P.D., 1981. Intra-sex and inter-sex sibling interactions as sex-ratio determinants. Nature 291, 64–66.
 Taylor, P.D., Bulmer, M.G., 1980. Local mate competition and the sex ratio. J. theor. Biol. 86, 409–419.
 West, S.A., Herre, E.A., 1998. Stabilizing selection and variance in fig wasp sex ratios. Evolution 52, 475–485.
 West, S.A., Herre, E.A., Sheldon, B.C., 2000a. The benefits of allocating sex. Science 290, 288–290.
 West, S.A., Smith, T.G., Read, A.F., 2000b. Sex allocation and population structure in apicomplexan (protozoa) parasites. Proc. R. Soc. London B 267, 257–263.
 West, S.A., Reece, S.E., Read, A.F., 2001. The evolution of gametocyte sex ratios in malaria and related apicomplexan (protozoan) parasites. Trends Parasitol. 17, 525–531.
 West, S.A., Smith, T.G., Nee, S., Read, A.F., 2002. Fertility insurance and the sex ratios of malaria and related hemosporein blood parasites. J. Parasitol. 88, 258–263.

Bacteriocins, spite and virulence

Andy Gardner^{1*}, Stuart A. West¹ and Angus Buckling²

¹*Institute of Cell, Animal and Population Biology, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK*

²*Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK*

There has been much interest in using social evolution theory to predict the damage to a host from parasite infection, termed parasite virulence. Most of this work has focused on how high kinship between the parasites infecting a host can select for more prudent exploitation of the host, leading to a negative relationship between virulence and parasite kinship. However, it has also been shown that if parasites can cooperate to overcome the host, then high parasite kinship within hosts can select for greater cooperation and higher growth rates, hence leading to a positive relationship between virulence and parasite kinship. We examine the impact of a spiteful behaviour, chemical (bacteriocin) warfare between microbes, on the evolution of virulence, and find a new relationship: virulence is maximized when the frequency of kin among parasites' social partners is low or high, and is minimized at intermediate values. This emphasizes how biological details can fundamentally alter the qualitative nature of theoretical predictions made by models of parasite virulence.

Keywords: social evolution; neighbour-modulated fitness; negative relatedness; Hamiltonian spite; scale of competition; interference competition

1. INTRODUCTION

There is a large theoretical literature applying evolutionary theory to explain the damage that parasites cause to their hosts (van Baalen & Sabelis 1995; Frank 1996; Gandon *et al.* 2001; Day & Burns 2003). Parasite virulence presents a fundamental trade-off in that parasites must deplete host resources to grow and transmit to new hosts, yet over-exploitation can result in host mortality and an associated reduction in resource availability (Frank 1996). This is the 'tragedy of the commons' (Hardin 1968), in which individuals are expected to display altruistic self-restraint only if they are sufficiently related to their group (Frank 1998). A classic result of virulence theory is that intensity of exploitation and hence damage to hosts correlates negatively with kinship among the parasites infecting a host (Hamilton 1972; Bremerman & Pickering 1983; Frank 1992, 1996). This occurs because a lower relatedness leads to greater competition for resources, which selects for faster growth rates to obtain a greater proportion of the host resources, and these higher parasite growth rates lead to higher virulence.

However, empirical support for this prediction is severely lacking (Herre 1993, 1995; Chao *et al.* 2000; Read & Taylor 2001; Davies *et al.* 2002; Griffin & West 2002; Read *et al.* 2002). One possible explanation for this is that variation in the underlying biological details can lead to alternative relationships (Frank 1996; Ganusov & Antia 2003; Schjørring & Koella 2003). In particular, it has been shown that if parasites can cooperate to overcome their host's defences then the opposite prediction is favoured—a positive relationship between parasite kinship and virulence (Chao *et al.* 2000; Brown *et al.* 2002; West & Buckling 2003). For example, West & Buckling (2003)

modelled the evolution of the production of costly public goods (siderophores) that promote bacterial growth during iron starvation in an infection. Not surprisingly, the altruistic production of siderophores is expected to be maximized when kinship is highest, yet this leads to enhanced growth and therefore host damage precisely where previous theory predicted self-restraint and hence low virulence.

Just as altruistic behaviour can promote parasite growth and hence enhance virulence, it is reasonable to assume that spiteful interactions (interference competition) between parasites could reduce the vigour of an infection and associated host damage. We consider such a spiteful trait: bacteriocin production. Bacteriocins are the most abundant of a range of antimicrobial compounds facultatively produced by bacteria, and are found in all major bacterial lineages (Riley & Wertz 2002). They are a diverse family of proteins with a range of antimicrobial killing activity, many of which can be produced by a single bacterium, including enzyme inhibition, nuclease activity and pore formation in cell membranes (Reeves 1972; Riley & Wertz 2002). Unlike other antimicrobials, the lethal activity of bacteriocins is often (but not always) limited to members of the same species as the producer, suggesting a major role in competition with conspecifics (Riley *et al.* 2003). Intraspecific competition may also help to explain the observed variation in the types of bacteriocin produced by different strains of the same species. For example, at least 25 bacteriocins (colicins) have been identified in populations of *Escherichia coli*, with different populations producing unique combinations (Riley & Gordon 1999). Clone mates are protected from the toxic effects of bacteriocins by genetic linkage between the bacteriocin gene and an immunity gene that encodes a factor that deactivates the bacteriocin (Riley & Wertz 2002).

In addition to the benefits of bacteriocin production (killing competitors), there are also costs (Reeves 1972;

* Author for correspondence (andy.gardner@ed.ac.uk).

Chao & Levin 1981; Kerr *et al.* 2002). This cost may simply be a diversion of resources from other cellular functions, but in many Gram-negative bacteria, such as *E. coli*, cell death is required for the release of bacteriocins (Reeves 1972; Riley & Wertz 2002). Such costs (and costs associated with bacteriocin immunity) are critical for coexistence, between bacteriocin-producing, sensitive and resistant strains (Czárán *et al.* 2002; Kerr *et al.* 2002; Czárán & Hoekstra 2003). We investigate how key parameters affect the relative costs and benefits of bacteriocin production, hence the level favoured by natural selection, and the impact this has on disease virulence. Specifically, we consider how bacteriocin production evolves in response to the average kinship (r) of competing bacteria and the scale of competition relative to the effective range of bacteriocins (a).

2. MODELS, METHODS AND ANALYSES

(a) *Simplest scenario*

We first consider a social arena, defined as the spatial range of bacteriocin warfare, comprising n equally abundant lineages drawn independently from the asexually reproducing bacterial population. A proportion $r = 1/n$ of the bacteria within a focal bacterium's social arena are its clone-mates, or 'kin'. The remaining $1 - r$ are derived from the other $n - 1$ lineages, and are 'non-kin'. Using a game theoretic approach, we consider the fitness of a vanishingly rare mutant that allocates an amount of resources y into bacteriocin production within a population with average allocation z , so as to determine the 'unbeatable' (Hamilton 1967) or 'evolutionarily stable' (Maynard Smith & Price 1973) allocation strategy y^* . An amount of bacteriocin ry within the social arena is attributable to the focal lineage, and rz to each of the other lineages. The focal lineage is then subjected to an amount $(1 - r)z$ of unrelated bacteriocin to which it is susceptible, and for each of the $n - 1$ other lineages, $(1 - r)z + r(y - z)$. A lineage picked at random from the population as a whole experiences, on average, $(1 - r)z$ unrelated bacteriocin. Lineages are immune to their own bacteriocins, and although resistance (non-susceptibility of a lineage to a bacteriocin which it does not itself produce) is not explicitly discussed in this model, the resulting reduction in susceptibility can be regarded as included in the general growth functions. The growth rate of a lineage, G , is given by the sum of two components, H and I . H reflects the cost of bacteriocin production, being a positive, decreasing function of the focal lineage's allocation to bacteriocin production, y . Our predictions rely on no specific form for H ; when a specific relationship is required for illustrative purposes (figures 1–3), we use $H = 1 - y$. I models the reduction in growth owing to mortality by unrelated bacteriocins, being a positive, decreasing, linear or decelerating function of the amount (Y) of unrelated bacteriocin it is subjected to. Our predictions rely on no specific form for I ; when a specific relationship is required for illustrative purposes (figures 1–3), we use $I = 1 - Y^{1/2}$. We combine the terms H and I additively to give overall growth ($G = H + I$) for mathematical convenience, as it allows greater tractability than using a multiplicative scheme ($G = H \times I$), and does not qualitatively change the results (see Appendix B). Using the construction of Frank (1998), fitness is determined by the growth of the lineage relative to the average competitor of that lineage:

$$w = \frac{G_{\text{focal}}}{a G_{\text{local}} + (1 - a) G_{\text{global}}} \quad (2.1)$$

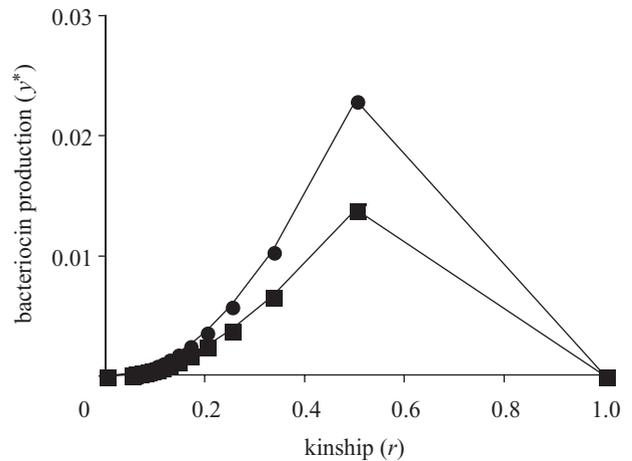


Figure 1. The ESS production of bacteriocins (y^*) as a function of the average kinship (r) between bacteria. Values are obtained numerically using the model described in § 2a, assuming that bacterial growth is the sum of growth components $H = 1 - y$ and $I = 1 - Y^{1/2}$ (where the focal bacterium produces an amount y of its own bacteriocins, and receives an amount Y from its social partners) and the intensity of local competition which is local is $a = 0.5$ (filled squares) and 0.6 (filled circles). Intermediate kinship (r) and increasingly local competition (high a) favour enhanced bacteriocin production.

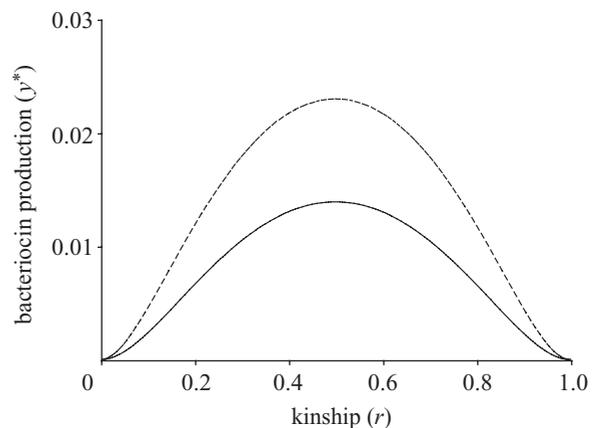


Figure 2. The ESS production of bacteriocin (y^*) as a function of the average kinship (r) between bacteria. Values are obtained numerically using the two-lineage model described in Appendix A, assuming that bacterial growth is the sum of growth components $H = 1 - y$ and $I = 1 - Y^{1/2}$ (where the focal bacterium produces an amount y of its own bacteriocins, and receives an amount Y from its social partners) and the intensity of local competition which is local is $a = 0.5$ (solid line) or 0.6 (dotted line). Intermediate kinship (r) and increasingly local competition (high a) favour enhanced bacteriocin production.

The parameter a defines the (spatial) scale at which competition for resources takes place. This model therefore allows competition for resources and bacteriocin interaction to take place at different scales. Specifically, a proportion a of competition for resources occurs locally, within the scale of bacteriocin interaction, and the $(1 - a)$ remainder occurs globally. At the extremes: if $a = 1$ then competition for resources and bacteriocin interaction occur at the same scale (soft selection at the level of the social group); if $a = 0$ then competition is at the global level

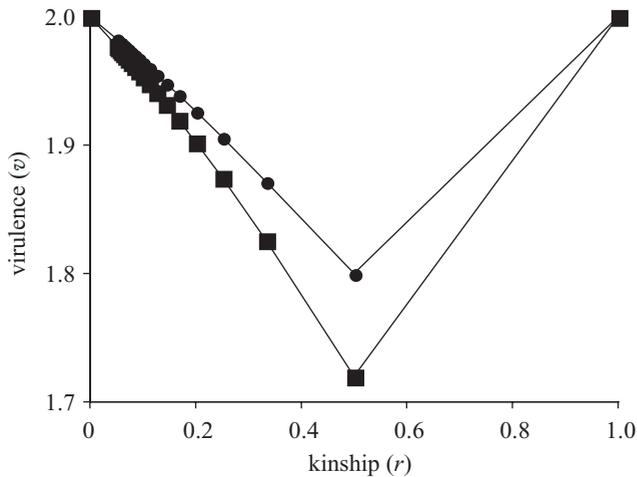


Figure 3. The virulence (z) as a function of the average kinship (r) between bacteria. Values are obtained numerically using the host mortality model described in § 2b, assuming that bacterial growth is the sum of growth components $H = 1 - y$ and $I = 1 - Y^{1/2}$ (where the focal bacterium produces an amount y of its own bacteriocins, and receives an amount Y from its social partners), host survival is $S = 3 - G_{\text{host}}$ (where G_{host} is the overall bacterial growth in the host), the intensity of local competition is $a = 0.5$, and the range of bacteriocin warfare with respect to the whole infection is $b = 0.1$ (filled circles) and 0.2 (filled squares). Virulence is minimized at intermediate kinship (r) and when the range of bacteriocin warfare (b) is large.

(hard selection at the level of the social group). G_{focal} , G_{local} and G_{global} are, respectively, the growth rate of the focal lineage, the local average and the global average. These are, in full

$$\begin{aligned} G_{\text{focal}} &= H[y] + I[(1-r)z], \\ G_{\text{local}} &= r(H[y] + I[(1-r)z]) + (1-r)(H[z] + I[(1-r)z \\ &\quad + r(y-z)]), \\ G_{\text{global}} &= H[z] + I[(1-r)z]. \end{aligned} \quad (2.2)$$

Equations (2.1) and (2.2) illustrate the fundamental trade-off in our model. Bacteriocin production by the focal lineage is: (i) costly, because it lowers the growth rate of the focal lineage (G_{focal}); and (ii) beneficial, because it lowers the growth rate of competitors G_{local} .

Employing the direct fitness maximization technique of Taylor & Frank (1996; Frank 1998), we obtain the following results (details in Appendix A; numerical examples are given in figure 1).

Result 1: enhanced bacteriocin production is favoured at intermediate kinship (r). The evolutionarily stable strategy (ESS) is $y^* = 0$ at $r = 0$ and 1 , and is maximized somewhere in the range $0 < r < 1$. When the focal lineage occupies only a tiny proportion ($r \rightarrow 0$) of the social arena, its impact on competitor growth is negligible, and hence the benefit through competitor killing does not outweigh the cost of bacteriocin production. When the focal lineage dominates the social group ($r \rightarrow 1$), the density of cells susceptible to its bacteriocin is too low for the benefit of competitor killing to outweigh the production costs.

Result 2: enhanced bacteriocin production is favoured as the scale of competition a is increased (and hence competition for resources becomes more local) for all $0 < r < 1$. This occurs because fitness can be enhanced in two ways: (i) maximizing own growth (G_{focal}); and (ii) reducing the growth of local

competitors (G_{local}). When competition is entirely global ($a = 0$), there is no benefit in reducing the growth of local competitors, so that the ESS is the strategy that maximizes focal growth (by reducing bacteriocin production). As competition becomes more local ($a > 0$), production of bacteriocin is increasingly favoured so as to reduce the growth of the local competitors.

We also consider a model in which the abundance of the focal lineage can vary continuously over the range $0 \leq r \leq 1$, and the other cells all belong to one other lineage (see Appendix A and figure 2). We recover the same results, finding that ESS bacteriocin production is maximized at intermediate kinship (at $r = 1/2$, because of the symmetry of this model) and increases as competition becomes more localized (i.e. as a increases).

As is often the case (Taylor & Frank 1996; Frank 1998), inspection of the direct marginal fitness (equation (A 1)) yields a form of Hamilton's (1963) rule $RB > C$ (equation (A 2)). In this: (i) relatedness is negative and given by $R = -(ar)/(1-ar)$; (ii) the negative 'benefit', summed over all recipients, is $B = (1-r)I'[(1-r)z]$ where $I'[Y]$ is the derivative $dI[Y]/dY$ and represents the marginal reduction in growth of a lineage which is poisoned by an amount Y of foreign bacteriocins. To understand how a negative relatedness can arise, we will use the result of Queller (1994) that average relatedness to one's competitors is zero. Recalling that the scale of competition (a) is defined as the proportion of competition which is local, consider an arena of competition in which a proportion of competitors a are social partners, and of these a proportion r belong to the focal lineage. Then a proportion ar of competitors are clonally related to the spiteful actor by 1 , and a proportion $1-ar$ are related by some unknown coefficient R . Applying Queller's insight, we know that $ar \times 1 + (1-ar) \times R = 0$, and rearranging we obtain $R = -(ar)/(1-ar)$. Hence:

Result 3: the evolution of bacteriocin production involves a negative relatedness between actor and recipient, and hence fits Hamilton's (1970) original definition of a spiteful behaviour. Relatedness between non-kin social partners is given by $R = -(ar)/(1-ar)$, where a is the proportion of competition that is local, and r is the proportion of social partners that are clonal kin. This equation gives negative values for relatedness, except when either (or both) a and r are zero, in which case relatedness equals zero.

(b) Host mortality

The above model is appropriate for free-living bacteria, bacteria grown on agar plates, or parasitic bacteria in which host mortality does not influence the ESS production of bacteriocin. For parasitic bacteria, this would be appropriate when the extra host mortality due to the infection impinges very little upon bacterial success, or when there are many social groups within the host, such that any lineage's growth rate has a negligible impact on the mortality of the host. A simple model, relaxing these assumptions, considers that direct fitness of the focal lineage is given by the product $S \times T$, where S represents host survival (i.e. the time over which transmission is possible) and is a linearly decreasing function of the average growth rate of lineages in the host. T is the transmission rate achieved by the focal lineage, i.e. its growth rate relative to competitors, the fitness measure given by equation (2.1). A parameter, b , is introduced to denote the proportion of the bacterial population within the host that is in the focal arena of social (bacteriocin) interaction; $b = 0$ corresponds to when the social arena comprises a vanishingly small

proportion of the total infection, and $b = 1$ corresponds to the arena of bacteriocin interaction being the entire infection. As in our first model, we assume n equally abundant lineages. The appropriate fitness function is

$$w = S[G_{\text{host}}] \frac{G_{\text{focal}}}{aG_{\text{local}} + (1-a)G_{\text{global}}}, \quad (2.3)$$

where the growth rate of a random lineage within the host is on average

$$G_{\text{host}} = bG_{\text{local}} + (1-b)G_{\text{global}}. \quad (2.4)$$

Virulence (v) can be defined as the reduction in S relative to a host with zero bacterial growth ($G_{\text{host}} = 0$), i.e. $v = S[0] - S[G_{\text{host}}]$. The following result is obtained (see Appendix A for details, and figure 3 for numerical examples).

Result 4: virulence (v) is maximized at the extremes of relatedness ($r = 0$ and $r = 1$), and is minimized at intermediate values $0 < r < 1$. This is because of the maximization of bacteriocin production at intermediate values of r , such that absolute growth of bacteria is reduced here but not at more extreme values, so that virulence is more pronounced whenever bacteria tend to socialize mostly, or not at all, with their kin.

3. DISCUSSION

We have shown that the production of bacteriocin is expected to be enhanced when kinship (r) is of intermediate value (result 1; figures 1 and 2). Because bacteriocin production is expected to correlate with low bacterial growth rates, virulence will tend to be minimized at intermediate r and maximized when bacteria compete only with non-kin ($r = 0$) or only with kin ($r = 1$). We therefore predict a U-shaped relationship between virulence and kinship (result 4; figure 3), contrary to previous models that variously predict monotonically increasing or decreasing virulence as kinship is increased. This emphasizes that the qualitative outcome of virulence evolution crucially depends on the biological details, such as whether parasites are able to improve their success through prudent growth (Frank 1996), or cooperative contributions to public goods (Brown *et al.* 2002; West & Buckling 2003), or through anti-competitor toxin production.

Our result is intuitive if we consider that when kinship (r) is low the influence of the focal lineage on the growth of its social partners will be negligible, and so reduced allocation of resources into bacteriocin production is favoured. By contrast, when kinship is high, the proportion of cells in the social arena that are susceptible to bacteriocin killing is small, and thus the benefit of producing bacteriocin is less than the cost that this entails. At intermediate kinship, bacteriocin production is favoured because competition with non-relatives is important, and bacteriocin production by the focal lineage can significantly decrease the growth of the non-competitors. Result 2, that the ESS bacteriocin production is an increasing function of the degree to which competition is local (a ; figures 1 and 2), is also intuitive in that when competition is increasingly local the benefits accrued by reducing the growth of local competitors are enhanced.

The costly allocation of resources into bacteriocin production qualifies as an example of Hamiltonian spite (Hamilton 1970, 1996; Hurst 1991; Foster *et al.* 2001; Gardner & West 2004). It is well accepted that altruism

can be adaptive despite a direct fitness cost provided the beneficiary of altruism is sufficiently positively related to the actor (i.e. a positive R and a positive B , and $RB > C$). Hamiltonian spite is when a costly behaviour is favoured because it has a cost to the recipient (negative B), and the recipient is negatively related to the actor (negative R , and $RB > C$). How can negative relatedness arise? Negative relatedness to some individuals is inevitable when positively related individuals exist in the same competitive arena. The reason for this is that because the relatedness of an actor to a randomly chosen individual from its competitive arena is, on average, zero (Queller 1994), the existence of positive relations within that arena implies the existence of negatively related competitors (Result 3). In this situation, spiteful behaviour will be favoured if it can be preferentially directed at these negatively related competitors, and $RB > C$ is satisfied. The specificity of bacteriocin action allows it to potentially fill this criterion, because it will preferentially harm non-relatives who are not resistant to that particular bacteriocin; i.e. bacteriocins harm individuals who are negatively related to the producer. Although the anti-competitor function of the bacteriocins suggests that this is selfishness at the level of the clonal lineage, it is certainly spiteful at the level of the self-destructing bacterium producing the toxins.

To conclude, we have shown theoretically how kinship and the scale of competition determine levels of bacteriocin production favoured by natural selection. Contrary to previous work, we find a U-shaped relationship between kinship and virulence. The results are qualitatively the same whether bacteria have fixed strategies for bacteriocin production or if bacteriocin production is facultatively adjusted in response to kin recognition. These predictions could be tested by: (i) correlating bacteriocin production with average kinship in natural populations; or (ii) experimentally evolving bacteria under different degrees of kinship and scales of competition. Furthermore, our predictions are not limited to bacteriocin production by bacteria. A variety of microbes, including yeasts (see Schmitt & Breinig 2002) and halophilic archaea (see Cheung *et al.* 1997) are known to produce toxins that tend to target conspecifics.

We thank N. Barton and three anonymous reviewers for comments. Funding was provided by BBSRC, NERC and The Royal Society.

APPENDIX A

(a) *Simplest scenario*

Substituting equation (2.2) into equation (2.1) we obtain fitness function $w[y, z]$. If we assume only minor variants ($y \approx z$; Taylor & Frank 1996) the marginal fitness is found to be

$$\left. \frac{dw}{dy} \right|_{y=z} = \frac{(1-ar)H'[z] - ar(1-r)I'[(1-r)z]}{H[z] + I[(1-r)z]}. \quad (A 1)$$

Where $H' < 0$ is the derivative of H with respect to its parameter (e.g. y in the instance of the mutant), and may be interpreted as the marginal cost ($-C$) of producing bacteriocins. $I' < 0$ is the derivative of I with respect to its parameter (e.g. $(1-r)z$ for the amount of bacteriocin attacking the focal mutant), and is the negative 'benefit'

accrued by the recipient of spiteful behaviour—summing over all the recipients, the benefit is $B = (1 - r)I'[(1 - r)z]$. Increased bacteriocin production (y) is favoured whenever $dw/dy > 0$ is satisfied, yielding Hamilton's rule:

$$-\frac{ar}{1 - ar}B > C. \tag{A 2}$$

Substituting $r = 0$ into equation (A 1) obtains $H'[z]/(H[z] + I[z])$, which is negative and hence $y^* = 0$. When $r = 1$, equation (A 1) becomes $(1 - a)H'[z]/(H[z] + I[z])$ which is negative and so $y^* = 0$. When $a = 0$, equation (A 1) gives $H'[z]/(H[z] + I[(1 - r)z])$, which is negative so that $y^* = 0$. Therefore, the presence of more than one lineage ($0 < r < 1$) and some degree of local competition ($a > 0$) are essential for non-zero allocation to bacteriocin production. If we denote the right-hand side (RHS) of equation (A 1) by \mathcal{F} , then the ESS $z = y^*$ satisfies $\mathcal{F} = 0$. Using implicit differentiation, we can write

$$\frac{dy^*}{dr} = -\frac{\delta\mathcal{F}/\delta r}{\delta\mathcal{F}/\delta y^*} \tag{A 3}$$

where δ denotes partial derivatives. For y^* to be convergence stable (i.e. in a population close to y^* , mutants closer to y^* are favoured by selection), the denominator on the RHS of equation (A 3) must be negative (Taylor 1996). Hence, assuming convergence stability, dy^*/dr has the same sign as $\delta\mathcal{F}/\delta r$ (Pen 2000). Evaluating the partial derivative at $r = 0$ (and hence $y^* = 0$) yields $-a(H[0] + I[0])(H'[0] + I'[0])/(H[0] + I[0])^2$, which is positive when $a > 0$. This indicates that when there is some degree of local competition, and intermediate relatedness, bacteriocin production will be nonzero. Using the same procedure, we may find the partial derivative of \mathcal{F} with respect to the scale of competition, a :

$$\frac{\delta\mathcal{F}}{\delta a} = -\frac{rH'[y^*] + r(1 - r)I'[(1 - r)y^*]}{H[y^*] + I[y^*]}, \tag{A 4}$$

which is positive for all $0 < r < 1$, and hence bacteriocin production is an increasing function of the scale of competition (a) when kinship is intermediate.

We now relax the assumption of equally abundant lineages, looking now at the situation where only two lineages occupy the social arena, so that the focal lineages comprise a proportion r or $1 - r$ of the bacterial cells with equal probability. The appropriate fitness function is then

$$zw = r \frac{G_{\text{focal1}}}{aG_{\text{local1}} + (1 - a)G_{\text{global}}} + \frac{(1 - r)G_{\text{focal2}}}{aG_{\text{local2}} + (1 - a)G_{\text{global}}} \tag{A 5}$$

where

$$\begin{aligned} G_{\text{focal1}} &= H[y] + I[(1 - r)z], \\ G_{\text{focal2}} &= H[y] + I[rz], \\ G_{\text{local1}} &= r(H[y] + I[(1 - r)z]) + (1 - r)(H[z] + I[ry]), \\ G_{\text{local2}} &= (1 - r)(H[y] + I[rz]) + r(H[z] + I[(1 - r)y]), \\ G_{\text{global}} &= H[z] + rI[(1 - r)z] + (1 - r)I[rz]. \end{aligned} \tag{A 6}$$

Following the same procedure as before, we obtain

$$\left. \frac{dw}{dy} \Big|_{y=z} = \left\{ \begin{aligned} &ar(1 - r) \left(\begin{aligned} &r(H[z] + I[(1 - r)z])I'[rz] \\ &+(1 - r)(H[z] + I[rz])I'[(1 - r)z] \end{aligned} \right) \\ &+ \left(\begin{aligned} &(1 - a(1 - 2r(1 - r)))H[z] \\ &+r(1 - ar)I[(1 - r)z] \\ &+(1 - r)(1 - a(1 - r))I[rz] \end{aligned} \right) H'[z] \end{aligned} \right\} / \{H[z] + rI[(1 - r)z] + (1 - r)I[rz]\}^2. \tag{A 7}$$

Setting $r \rightarrow 0$ yields $(1 - a)H'[z]/(H[z] + I[0])$ which is always negative and hence $y^* = 0$ at $r = 0$. Setting $r \rightarrow 1$ yields $(1 - a)H'[z]/(H[z] + I[0])$ which is always negative, so $y^* = 0$ at $r = 1$. And when $a \rightarrow 0$, we obtain $H'[z]/(H[z] + rI[(1 - r)z] + (1 - r)I[rz])$ which is always negative, so that $y^* = 0$ when $a = 0$.

As before, if we define \mathcal{F} as the RHS of equation (A 7) when $z = y^*$, then it is easy to show that for $a > 0$, $\delta\mathcal{F}/\delta r = dy^*/dr = 0$ is satisfied for only $r = 1/2$. Since $y^* = 0$ at $r = 0$ and $r = 1$, and assuming no discontinuities over the range of r , we can conclude that y^* monotonically increases over the range $0 < r < 1/2$ and monotonically decreases over the range $1/2 < r < 1$.

The partial derivative of \mathcal{F} with respect to the scale of competition is $\delta\mathcal{F}/\delta a = -(r(1 - r)(r(H[y^*] + I[(1 - r)y^*]) \times I'[ry^*] + (1 - r)(H[y^*] + I[ry^*])I'[(1 - r)y^*]) + (1 - 2r(1 - r))H[y^*] + r^2I[(1 - r)y^*] + (1 - r)^2I[ry^*])H'[y^*]) / (H[y^*] + rI[(1 - r)y^*] + (1 - r)I[ry^*])^2$, which is positive for all $0 < r < 1$, and hence bacteriocin production is an increasing function of the scale of competition (a) at intermediate kinship.

(b) Host mortality

Previously we constructed a fitness function (equation (2.3)) appropriate to the situation where bacterial growth impacts upon host mortality (virulence) and hence introduces a novel selection pressure. We also introduced a parameter b scaling the social arena with respect to the host. If $b = 0$, so that the social arena comprises a vanishing proportion of the bacterial population within the host, then $G_{\text{host}} = G_{\text{global}}$ and S is a constant with respect to y , so that marginal fitness is given by equation (A 1). For $b > 0$, and assuming only minor variants ($y \approx z$, $G_{\text{focal}} \approx G_{\text{local}} \approx G_{\text{global}} \approx G_{\text{host}} \approx G$), marginal fitness is

$$\begin{aligned} \frac{dw}{dy} &= S'[G]rb(H'[z] + (1 - r)I'[(1 - r)z]) \\ &+ S[G] \frac{(1 - ar)H'[z] - ar(1 - r)I'[(1 - r)z]}{G}. \end{aligned} \tag{A 8}$$

The second component on the RHS is proportional to the marginal fitness (equation (A 1)), and represents the trade-off between the cost and competitor-killing capabilities of bacteriocins. When $a = 0$, this component reduces to $(S[G]H'[z])/G$, which is always negative, reflecting the disadvantage of spite when competition is global. The first component, positive and proportional to rb , is the selection pressure for enhanced killing and costly production when growth of the focal lineage and its neighbours impact non-trivially upon host mortality. As r tends to zero, marginal fitness is negative $(S[G]H'[z])/G$ as the behaviour of the focal lineage has no impact on host mortality and there is no advantage to be had from directing spite at local competitors (relatedness to non-kin in the social arena is zero). At $r = 1$, the second component is

negative ($S[G](1-a)H'[z]/G$) reflecting the fitness cost of bacteriocin production, and the first component is positive ($S'[G]H'[z]$) reflecting the enhanced fitness due to the reduction in host mortality. Note that this positive pressure is due entirely to the costs of bacteriocin production, and not through its bacteriocidal activity; this is due to an artificiality in the model such that the bacteria have no means of reducing own growth other than producing costly bacteriocin. Because no gain in terms of competitor killing is to be had from producing bacteriocins at $r=1$, we expect $y^*=0$.

If $y^*=0$ at $r=0$ and 1, then since H and I are decreasing functions of y^* , it is here that $G_{\text{host}}=H+I$ is maximized. Because S decreases with increasing G_{host} , S is minimized at $r=0, 1$. If we define virulence as the reduction in host survival relative to that for a host in which bacterial growth is zero ($v=S_{\text{max}}-S$), then virulence is maximized when S is minimized ($v_{\text{max}}=S_{\text{max}}-S_{\text{min}}$), i.e. at the extremes of relatedness, $r=0$ and $r=1$.

When a and b are both zero, so that there is no selection for spite nor for reduced virulence, equation (A 8) reduces to $(S[G]H'[z])/G$ which is negative and hence $y^*=0$.

APPENDIX B

Relaxing the assumption of additive growth components, and making no further assumptions about the components of growth beyond bacteriocin production reducing the growth of the focal lineage (G_{focal}) and its non-kin social partners (G_{social}), we can recover the major predictions made in this study. Consider the fitness function (equation (2.1)). Marginal fitness can be written

$$\frac{dz}{dy} = \frac{(aG_{\text{focal}} + (1-a)G_{\text{global}}) \frac{dG_{\text{focal}}}{dy} - G_{\text{focal}} \frac{d(aG_{\text{local}} + (1-a)G_{\text{global}})}{dy}}{(aG_{\text{focal}} + (1-a)G_{\text{global}})^2}. \quad (\text{B } 1)$$

Assuming only minor variants, so that $y \approx z$, and $G_{\text{focal}} \approx G_{\text{social}} \approx G_{\text{local}} \approx G_{\text{global}} \approx G$, we have

$$\frac{dz}{dy} = \left((1-ar) \frac{dG_{\text{focal}}}{dy} - a(1-r) \frac{dG_{\text{social}}}{dy} \right) / G. \quad (\text{B } 2)$$

Fitness increases with enhanced bacteriocin production when $dz/dy > 0$. dG_{focal}/dy is negative owing to the production costs of bacteriocin, and dG_{social}/dy is negative because non-kin social partners experience higher mortality as bacteriocin production by the focal lineage is increased. Equation (B 2) therefore demonstrates the trade-off between the direct cost of bacteriocin production and the benefit of competitor killing. The benefit is zero when $a=0$ and/or when $r=1$, so that marginal fitness is $\{(1-ar)dG_{\text{focal}}/dy\}/G < 0$ for all y , meaning that the ESS bacteriocin production is at $y^*=0$. Also, the impact of the focal lineage's bacteriocin on competitor growth approaches zero as the focal lineage accounts for a vanishing proportion of the social group, i.e. at $r=0$, $dG_{\text{social}}/dy=0$, and so here the marginal fitness is negative, and $y^*=0$. Therefore, regardless of the precise details describing how the growth of the focal lineage and

its non-kin social partners decline with enhanced bacteriocin production, provided they do decline, we can state that the ESS is $y^*=0$ when kinship is zero or complete ($r=0, 1$) and when competition is entirely global ($a=0$).

REFERENCES

- Bremerman, H. J. & Pickering, J. 1983 A game-theoretical model of parasite virulence. *J. Theor. Biol.* **100**, 411–426.
- Brown, S. P., Hochberg, M. E. & Grenfell, B. T. 2002 Does multiple infection select for raised virulence? *Trends Microbiol.* **10**, 401–405.
- Chao, L. & Levin, B. R. 1981 Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proc. Natl Acad. Sci. USA* **78**, 6324–6328.
- Chao, L., Hanley, K. A., Burch, C. L., Dahlberg, C. & Turner, P. E. 2000 Kin selection and parasite evolution: higher and lower virulence with hard and soft selection. *Q. Rev. Biol.* **75**, 261–275.
- Cheung, J., Danna, K., O'Connor, E., Price, L. & Shand, R. 1997 Isolation, sequence, and expression of the gene encoding halocin H4, a bacteriocin from the halophilic archaeon *Haloferax mediterranei* R4. *J. Bacteriol.* **179**, 548–551.
- Czárán, T. L. & Hoekstra, R. F. 2003 Killer-sensitive coexistence in metapopulations of micro-organisms. *Proc. R. Soc. Lond. B* **270**, 1373–1378. (DOI 10.1098/rspb.2003.2328.)
- Czárán, T. L., Hoekstra, R. F. & Pagie, L. 2002 Chemical warfare between microbes promotes biodiversity. *Proc. Natl Acad. Sci. USA* **99**, 786–790.
- Davies, C. M., Fairbrother, E. & Webster, J. P. 2002 Mixed strain schistosome infections of snails and the evolution of parasite virulence. *Parasitology* **124**, 31–38.
- Day, T. & Burns, J. G. 2003 A consideration of patterns of virulence arising from host–parasite coevolution. *Evolution* **57**, 671–676.
- Foster, K. R., Wenseleers, T. & Ratnieks, F. L. W. 2001 Spite: Hamilton's unproven theory. *Annls Zool. Fennici* **38**, 229–238.
- Frank, S. A. 1992 A kin selection model for the evolution of virulence. *Proc. R. Soc. Lond. B* **250**, 195–197.
- Frank, S. A. 1996 Models of parasite virulence. *Q. Rev. Biol.* **71**, 37–78.
- Frank, S. A. 1998 *Foundations of social evolution*. Princeton University Press.
- Gandon, S., Mackinnon, M. J., Nee, S. & Read, A. F. 2001 Imperfect vaccines and the evolution of pathogen virulence. *Nature* **414**, 751–756.
- Ganusov, V. V. & Antia, R. 2003 Trade-offs and the evolution of virulence of microparasites: do details matter? *Theor. Popul. Biol.* **64**, 211–220.
- Gardner, A. & West, S. A. 2004 Spite and the scale of competition. *J. Evol. Biol.* **17**. (In the press.)
- Griffin, A. S. & West, S. A. 2002 Kin selection: fact and fiction. *Trends Ecol. Evol.* **17**, 15–21.
- Hamilton, W. D. 1963 The evolution of altruistic behaviour. *Am. Nat.* **97**, 354–356.
- Hamilton, W. D. 1967 Extraordinary sex ratios. *Science* **156**, 477–488.
- Hamilton, W. D. 1970 Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**, 1218–1220.
- Hamilton, W. D. 1972 Altruism and related phenomena, mainly in social insects. *A. Rev. Ecol. Syst.* **3**, 193–232.
- Hamilton, W. D. 1996 *Narrow roads of geneland I: evolution of social behaviour*. Oxford: Freeman.
- Hardin, G. 1968 The tragedy of the commons. *Science* **162**, 1243–1248.
- Herre, E. A. 1993 Population structure and the evolution of virulence in nematode parasites of fig wasps. *Science* **259**, 1442–1445.

- Herre, E. A. 1995 Factors influencing the evolution of virulence: nematode parasites of fig wasps as a case study. *Parasitology* **111**, S179–S191.
- Hurst, L. D. 1991 The evolution of cytoplasmic incompatibility or when spite can be successful. *J. Theor. Biol.* **148**, 269–277.
- Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. M. 2002 Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature* **418**, 171–174.
- Maynard Smith, J. & Price, G. R. 1973 The logic of animal conflict. *Nature* **246**, 15–18.
- Pen, I. 2000 Reproductive effort in viscous populations. *Evolution* **54**, 293–297.
- Queller, D. C. 1994 Genetic relatedness in viscous populations. *Evol. Ecol.* **8**, 70–73.
- Read, A. F. & Taylor, L. H. 2001 The ecology of genetically diverse infections. *Science* **292**, 1099–1102.
- Read, A. F., Mackinnon, M. J., Anwar, M. A. & Taylor, L. H. 2002 Kin selection models as explanations of malaria. In *Virulence management: the adaptive dynamics of pathogen–host interactions* (ed. U. Dieckmann, J. A. J. Metz, M. W. Sabelis & K. Sigmund), pp. 165–178. Cambridge University Press.
- Reeves, P. 1972 *The bacteriocins*. New York: Springer.
- Riley, M. A. & Gordon, D. M. 1999 The ecological role of bacteriocins in bacterial cooperation. *Trends Microbiol.* **7**, 129–133.
- Riley, M. A. & Wertz, J. E. 2002 Bacteriocins: evolution, ecology, and application. *A. Rev. Microbiol.* **56**, 117–137.
- Riley, M. A., Goldstone, C. M., Wertz, J. E. & Gordon, D. 2003 A phylogenetic approach to assessing the targets of microbial warfare. *J. Evol. Biol.* **16**, 690–697.
- Schjørring, S. & Koella, J. C. 2003 Sub-lethal effects of pathogens can lead to the evolution of lower virulence in multiple infections. *Proc. R. Soc. Lond. B* **270**, 189–193. (DOI 10.1098/rspb.2002.2233.)
- Schmitt, M. J. & Breinig, F. 2002 The viral killer system in yeast: from molecular biology to application. *FEMS Microbiol. Rev.* **26**, 257–276.
- Taylor, P. D. 1996 Inclusive fitness arguments in genetic models of behaviour. *J. Math. Biol.* **34**, 654–674.
- Taylor, P. D. & Frank, S. A. 1996 How to make a kin selection model. *J. Theor. Biol.* **180**, 27–37.
- van Baalen, M. & Sabelis, M. W. 1995 The scope for virulence management—a comment on Ewald’s view on the evolution of virulence. *Trends Microbiol.* **3**, 414–416.
- West, S. A. & Buckling, A. 2003 Cooperation, virulence and siderophore production in bacterial parasites. *Proc. R. Soc. Lond. B* **270**, 37–44. (DOI 10.1098/rspb.2002.2209.)

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.

tidomimetic library was screened to find molecules that could compete with the binding of the SMAC peptide to the Bir domain of different forms of IAP. After further chemical modification of a candidate molecule, Li *et al.* generated compound 3 that, like SMAC, has a high avidity for different forms of IAP including X-chromosome encoded IAP (XIAP), cellular IAP-1, and cellular IAP-2. Compound 3 blocked the interaction of XIAP with active caspase 9. In previous work, SMAC was shown to act synergistically with a death receptor called TRAIL to induce tumor-selective apoptosis (10). Impressively, treatment of glioblastoma cells with a combination of the ligand for the TRAIL receptor and compound 3 resulted in apoptosis of the tumor cells, whereas normal cells were not

harmed. Li *et al.* (4) also demonstrated that compound 3 could potentiate apoptosis in cells treated with TNF- α (tumor necrosis factor- α) without activation of the nuclear transcription factor NF- κ B. Because TNF- α mediates host responses in acute and chronic inflammatory conditions, these results suggest that compound 3 may have potential for treating inflammatory diseases (11). Although the efficacy of compound 3 was not evaluated in vivo, the authors are using compound 3 as a lead structure for the refinement of future therapeutic compounds with better pharmacological properties.

Peptidomimetics are only now emerging as a powerful solution for overcoming the limitations imposed by the physical properties of native peptides. Walensky *et al.* (3)

and Li *et al.* (4) demonstrate provocative proof-of-concept approaches to the design of peptidomimetics that may have a decided impact on future therapeutics that target disease by modulating specific protein-protein interactions.

References

1. C. A. Schmitt, *Nature Rev. Cancer* **3**, 286 (2003).
2. J. C. Reed, *Cancer Cell* **3**, 17 (2003).
3. L. D. Walensky *et al.*, *Science* **305**, 1466 (2004).
4. L. Li *et al.*, *Science* **305**, 1471 (2004).
5. J. A. Patch, A. E. Barron, *Curr. Opin. Chem. Biol.* **6**, 872 (2002).
6. N. N. Danial, S. J. Korsmeyer, *Cell* **116**, 205 (2004).
7. V. A. Levin, *J. Med. Chem.* **23**, 682 (1980).
8. J. S. Wadia, R. V. Stan, S. F. Dowdy, *Nature Med.* **10**, 310 (2004).
9. G. S. Salvesen, C. S. Duckett, *Nature Rev. Mol. Cell Biol.* **3**, 401 (2002).
10. S. Fulda *et al.*, *Nature Med.* **8**, 808 (2002).
11. R. M. Pope, *Nature Rev. Immunol.* **2**, 527 (2002).

ECOLOGY

Spite Among Siblings

Andy Gardner and Stuart A. West

"Sometimes I work my brother over...I make him squirm, I've made him cry. He doesn't know how I do it. I'm smarter than he is. I don't want to do it. It makes me sick."

—John Steinbeck, *East of Eden*

Although sibling conflict abounds in the literary world—from the Bible to Steinbeck—it also features prominently in the real world. Recent research from the laboratories of Strand and Hardy (1–3) on sibling conflict among parasitic wasps sheds light on that most puzzling of social behaviors—spite.

Social behaviors are those that affect the fitness of multiple individuals (4). The social behavior that has provoked the most interest is altruism, in which an action incurs a direct fitness cost for the actor and provides a benefit for the actor's social partners. Hamilton showed that altruism is favored when individuals are helping their close relatives, and hence still passing on their genes to the next generation, albeit indirectly. A pleasingly simple and elegant method for quantifying this idea of kin selection is Hamilton's rule, which states that an altruistic behavior will be favored if the cost to the actor (C) is outweighed by the product of the benefit (B) and the genetic relatedness (R) to the social partners, resulting in $RB > C$ (5). Hamilton, however, also

pointed out that his rule has a more sinister interpretation (6). His rule can be twisted to predict that spiteful behavior—which hurts both the actor and the recipient—may be favored when there is sufficient negative relatedness between the social partners.

Negative relatedness may seem like a bizarre concept, but it simply means that the recipient of a particular behavior is less related than other competitors to the actor (6–8). It has generally been assumed that spite is unlikely to be an important evolutionary force because the conditions required to obtain significant negative relatedness are too restrictive. Nonetheless, theoretical interest in spiteful behavior rumbles on. It is clear that spite can evolve given the right conditions: (i) when there is strong competition for local resources among social partners and (ii) when individuals have the capacity to recognize (and refrain from being spiteful to) their close kin (6, 7). In recent work, Strand, Hardy, and their colleagues (1–3) investigated a biological system that appears to satisfy both conditions—the sterile soldier caste of polyembryonic parasitic wasps.

These small wasps deposit their eggs into the eggs of moths, and the wasp larvae develop within the moth caterpillars (see the figure). A single wasp egg proliferates asexually (clonally) to produce multiple larvae such that, when the host contains larvae from several eggs, the limited food resources within the caterpillar will permit only a fraction of those larvae to complete development and emerge as adults. Thus, there is intense competition for resources

among the larvae within the host, satisfying the first condition for spite.

The majority of the wasp larvae develop normally, whereas others develop precociously to form a soldier caste that differs morphologically and behaviorally from normal wasp siblings (see the figure). Donnell *et al.* (1) demonstrate that the mechanism underlying caste formation in the clonally developing wasp population involves asymmetric inheritance of germ cells. Embryos that develop into normal larvae inherit the germ line, whereas embryos that develop into soldiers do not, making them obligately sterile—the cost of developing as a soldier. Upon hatching, soldiers distribute themselves throughout the host and launch aggressive attacks on other larvae, murdering their unfortunate victims. This has the potential to be spite and not altruism because the benefits of reduced competition accrue to all larvae in the host and not preferentially to closer relatives (7).

This would be an adaptive spiteful behavior if soldiers preferentially attacked the larvae they are least related to in the caterpillar, which requires kin recognition, the second condition for spite (7). In a new study, Giron and colleagues (2, 3) demonstrate that soldiers are indeed capable of recognizing their kin, and the investigators then elucidate the mechanism. First, they varied kinship by introducing either full (but not genetically identical) sisters and brothers or unrelated larvae into a host caterpillar containing a developing female brood of wasp larvae (2). The introduced larvae were labeled with a fluorescent tracer, and attack rates were assessed by measuring how many of the resident soldiers ingested labeled larval tissue (see the figure). As predicted, the researchers found a strong negative correlation between attack rates and kinship.

The authors are in the School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, UK. E-mail: andy.gardner@ed.ac.uk

In a companion study (3), the investigators shed light on the mechanism of this kin recognition faculty. They reveal that the key element is the extraembryonic membrane surrounding each larva during its development in the caterpillar host. They show that attack rates correlated negatively with kinship when the membrane was present, but not when the membrane was removed. In addition, by transplanting membranes between larvae they were able to fool the soldiers, whose attack rates correlated negatively with the kinship of the membrane donor but not with the larva encased inside. Mechanisms of kin recognition are unstable because deceptive variants arise that signal strong kinship to everyone; such variants can become common. However, the importance of the membrane in protecting larvae from host immune attack means that rare variants are intrinsically favored and that common variants are disadvantageous, providing a robust, honest signal of kinship. This may be true for many endoparasites, rendering such species masters of kin recognition.

One potentially puzzling result is that manipulation of resource availability by starving the host caterpillars did not influ-

ence the level of aggression exhibited by the wasp soldier caste (2). Possibly because competition is always local, resource availability does not influence how soldiers vary their relatedness-dependent behavior. Alternatively, soldier larvae may not be able to assess the intensity of competition for resources, either because doing so is difficult or because natural variation in competition is negligible and there has been no need for this faculty to evolve. Future work on how local competition for resources relates to soldier aggression could benefit from explicit theoretical modeling, as well as alternative methods for varying the scale of competition such as selection experiments (9) or comparative studies across species and populations. Nonetheless, the existence of an aggressive

A Life cycle of a polyembryonic wasp



A female wasp deposits her eggs into a caterpillar host egg.

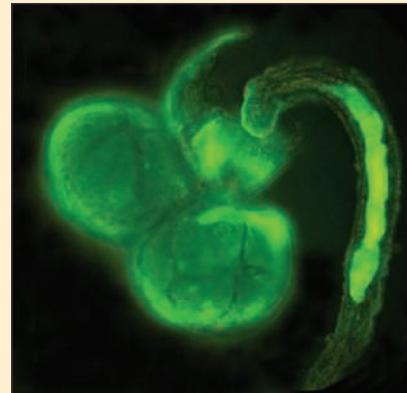


The wasp larvae undergo clonal proliferation within the caterpillar host.



Larvae develop either precociously into soldiers (left) or normally (right).

B Caught in the spiteful act



A soldier wasp larva ingests its fluorescently labeled victim.

soldier caste among parasitic wasps provides evidence that spite does exist in the real world, as Hamilton predicted it would.

References

1. D. Donnell, L. S. Corley, G. Chen, M. R. Strand. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10095 (2004).
2. D. Giron, D. W. Dunn, I. C. W. Hardy, M. R. Strand. *Nature* **430**, 676 (2004).
3. D. Giron, M. R. Strand. *Proc. R. Soc. B (Suppl.) Biol. Lett.*, 17 June 2004 (10.1098/rsb1.2004.0205).
4. S. A. Frank. *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, NJ, 1998).
5. W. D. Hamilton. *Am. Nat.* **97**, 354 (1963).
6. W. D. Hamilton. *Nature* **228**, 1218 (1970).
7. A. Gardner, S. A. West. *J. Evol. Biol.*, 22 July 2004 (10.1111/j.1420-9101.2004.00775).
8. A. Grafen. *Oxford Surv. Evol. Biol.* **2**, 28 (1985).
9. A. S. Griffin, S. A. West, A. Buckling. *Nature* **430**, 1024 (2004).

PLANETARY SCIENCE

Looking into the Giant Planets

Jonathan J. Fortney

Images of Jupiter and Saturn from telescopes and space probes only show the outermost layers of these giant planets. Learning about their interiors, which consist mostly of hydrogen (H) and helium (He) and make up over 90% of the planetary mass in the solar system, is more challenging. Recent model studies (1–3) show how new measurements from the Cassini spacecraft—now in orbit around Saturn—could lead to a better understanding of the interior of Saturn and, by extension, all giant planets.

The most important input into giant planet models is the equation of state—that

is, the relation between pressure and density—of hydrogen. Uncertainties in the equation of state translate directly into uncertainties in the estimated size of the “heavy element” (elements more massive than He) cores of the giant planets and the abundances of elements in their hydrogen-rich envelopes (1). Two groups have measured the shock-induced compressibility of deuterium, a heavy isotope of H, but there is a 50% discrepancy between their data sets (4, 5). As Saumon and Guillot (1) show in a recent paper in *The Astrophysical Journal*, this uncertainty profoundly affects inferences about the composition of the planets and the sizes of their cores. These quantities must be known before we can understand the process of giant planet formation and properties of the early solar system.

The authors created static models of Jupiter and Saturn that match all available

constraints, including mass, radius, oblateness, rotation period, atmospheric temperature, and gravitational moments for each planet. They also used a wide range of possible equations of state for H to allow for the disparate experimental data sets. According to their model, Jupiter’s core is 0 to 11 Earth masses. Saturn’s core is likely larger, between 9 and 22 Earth masses. (For comparison, Jupiter is 317.8 Earth masses and Saturn 95.2 Earth masses.) Overall, Jupiter is enriched in heavy elements by a factor of 1.5 to 6 relative to the Sun, and Saturn by a factor of 6 to 14. The most striking of these results is that we cannot be sure whether Jupiter has a core.

The greatest uncertainty in the structure of Jupiter comes from unsatisfactory understanding of liquid metallic H at Mbar pressures. In contrast, for Saturn, poor knowledge of its gravitational moments, which describe how the planet’s mass responds to its rotation, is the main obstacle. Gravitational moments are determined by measuring small accelerations of a spacecraft as it passes near a planet. During Cassini’s 4-year mission, error bars on the low-degree gravitational mo-

The author is with the Planetary Systems Branch, NASA Ames Research Center, Moffett Field, CA 94035, USA. E-mail: jfortney@arc.nasa.gov

MINI REVIEW

Spite and the scale of competition

A. GARDNER & S. A. WEST

*Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh, UK**Keywords:*

Hamiltonian spite;
 hard selection;
 kin competition;
 negative relatedness;
 Price equation;
 soft selection;
 Wilsonian spite.

Abstract

In recent years there has been a large body of theoretical work examining how local competition can reduce and even remove selection for altruism between relatives. However, it is less well appreciated that local competition favours selection for spite, the relatively neglected ugly sister of altruism. Here, we use extensions of social evolution theory that were formulated to deal with the consequences for altruism of competition between social partners, to illustrate several points on the evolution of spite. Specifically, we show that: (i) the conditions for the evolution of spite are less restrictive than previously assumed; (ii) previous models which have demonstrated selection for spite often implicitly assumed local competition; (iii) the scale of competition must be allowed for when distinguishing different forms of spite (Hamiltonian vs. Wilsonian); (iv) local competition can enhance the spread of spiteful greenbeards; and (v) the theory makes testable predictions for how the extent of spite should vary dependent upon population structure and average relatedness.

Altruism and spite

Social behaviours can be categorized according to the direct fitness consequences they entail for the actor and recipient (Fig. 1; Hamilton, 1964, 1970, 1971). A behaviour increasing the direct fitness of the actor is mutualistic if the recipient also benefits, and selfish if the recipient suffers a loss. It is easy to see how such behaviours can be favoured by natural selection. Behaviours which reduce the direct fitness of the actor – altruism if the recipient enjoys a benefit, spite if the recipient suffers a loss – are less easy to explain. Hamilton (1963, 1964) introduced the concept of inclusive fitness and showed that while certain behaviours are detrimental to the individual, they may result in a net increase in the actor's genes in the population. Altruism can be favoured by natural selection despite a direct fitness cost (C) to the actor if the benefit (B) accruing to the recipient is sufficiently large and if the genetic relatedness (R) of the recipient to the actor is sufficiently positive.

Correspondence: Andy Gardner, Institute of Cell, Animal & Population Biology, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK.
 Tel.: +44 131 650 5508; fax: +44 131 650 6564;
 e-mail: andy.gardner@ed.ac.uk

Specifically, when Hamilton's (1963, 1964) rule, $RB > C$, is satisfied. A spiteful behaviour, entailing a negative benefit ($B < 0$) to the recipient and a positive cost ($C > 0$) to the actor, is similarly favoured if $RB > C$, which would require a negative relatedness ($R < 0$) between actor and recipient.

Relatedness and spite

Hamilton (1963) argued that under the assumption of weak selection the appropriate measure of relatedness (R) coincides with Wright's (1922) correlation coefficient of relationship. Wright's coefficient is a function of the association between individuals and the association within individuals with respect to their genes at a given locus. Since these associations have popularly been interpreted in terms of Malécot's (1948) probability of identity by descent, and negative probabilities are not permitted, negative relatedness seems to be mathematically impossible (Hamilton, 1970, 1996; although see Wright, 1969, p. 178). Yet Hamilton (1963) understood that relatedness (R) was in principle a regression coefficient – a fact which is now generally appreciated (reviewed by Seger, 1981; Michod, 1982; Grafen, 1985; Queller, 1985, 1992; Frank, 1998) – and this was first made explicit in his elegant

		Effect on recipient	
		+	-
Effect on actor	+	mutualism	selfishness
	-	altruism	spite

Fig. 1 A classification of social behaviours.

reformulation of Hamilton's (1970) rule using Price's (1970) equation. Specifically, relatedness is the regression (slope) of the recipient's genetical breeding value on that of the actor (Hamilton, 1970, 1972; Taylor & Frank, 1996; Frank, 1997a, 1998). As regressions can be negative as well as positive (and zero), relatedness can feasibly take any real value (from negative infinity to positive infinity). Discussions with Price led Hamilton to acknowledge that negative relatedness can plausibly arise between social partners, and hence spite can be favoured by natural selection (Hamilton, 1970, 1996; Frank, 1995).

How does negative relatedness arise? Grafen's (1985) geometric view of relatedness reveals that relatedness between an actor and a potential recipient depends crucially upon the genetical composition of the whole population. This can be illustrated by assuming that a recipient carries the actor's genes with average frequency p , and the population frequency of the actor's genes is \bar{p} . If the recipient carries the actor's genes at a frequency greater than the population frequency of those genes ($p > \bar{p}$) then an increase in its reproductive success translates into increased frequency of the actor's genes in the population, and hence a positive inclusive fitness benefit to the actor ($RB > 0$; Fig. 2a). Conversely, if the recipient carries the actor's genes at a frequency lower than the population frequency of those genes ($p < \bar{p}$) then an increase in its reproductive success translates into decreased frequency of the actor's genes in the population, and hence a negative inclusive fitness benefit for the actor ($RB < 0$; Fig. 2b). The point here is that the difference between these two situations can arise purely due to variation in the frequency of the actor's genes in the population (variable \bar{p}), even with a fixed proportion of genes shared between the actor and recipient (fixed p): relatedness is relative, with the population as a whole providing the reference.

This also illustrates how negative relatedness can arise. As both situations described above involve a positive benefit ($B > 0$) to the recipient, the coefficient of relatedness which transforms recipient success into inclusive fitness of the actor must be positive in the former instance ($R > 0$; Fig. 2a) and negative in the latter ($R < 0$;

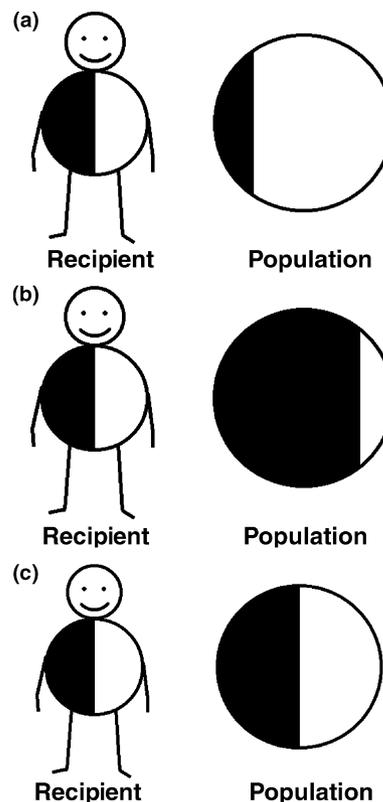


Fig. 2 The geometric view of relatedness. The actor's genes (shaded) are present in the recipient at frequency p and in the population as a whole at frequency \bar{p} . Enhancing the direct fitness of the recipient ($B > 0$) pushes the population gene frequency towards p , and so if $p > \bar{p}$ (a) the frequency of the actor's genes increase, giving a positive inclusive fitness benefit ($RB > 0$) which implies positive relatedness ($R > 0$) between actor and recipient. If $p < \bar{p}$ (b), then the population frequency of the actor's genes decreases, giving a negative inclusive fitness benefit ($RB < 0$) and hence negative relatedness ($R < 0$). When $p = \bar{p}$ (c) the population frequency does not change, giving no inclusive fitness benefit ($RB = 0$) and hence zero relatedness ($R = 0$).

Fig. 2b). The other possibility is that relatedness is zero when the recipient carries the same frequency of the actor's genes as does the population as a whole ($p = \bar{p}$), so that relatedness to the average population member (and hence to the population itself) is zero (Fig. 2c).

But, how large a negative relatedness is likely to arise? Consider an individual who lives in a population of size N , and who is then related to a fraction $1/N$ of the population (i.e. itself) by an amount 1 and is related to the other fraction $(N - 1)/N$ by an amount R . The relatedness to the population as a whole must be zero (Grafen, 1985), and hence must satisfy $(1/N) + [(N - 1)/N]R = 0$. Rearrangement gives $R = -1/(N - 1)$, i.e. the average relatedness between the actor and its social partners is negative (Hamilton, 1975; Grafen, 1985; Pepper, 2000). If the focal individual can identify, and refrain from being spiteful to, a number of positively related genealogically close social

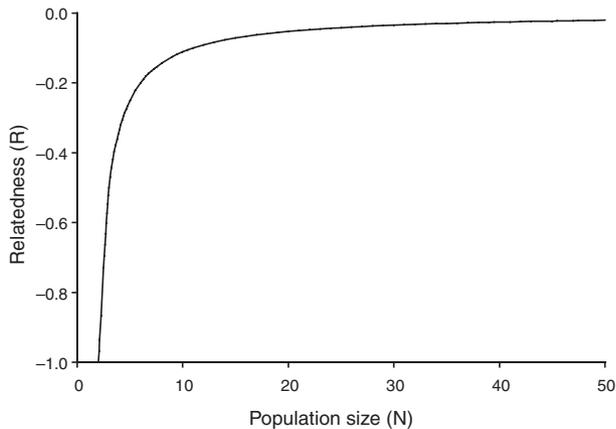


Fig. 3 The average relatedness (R) between population members as a function of population size (N), when there is no kin discrimination. Since relatedness by any member to the population as a whole is zero, and this includes positive relatedness to itself, relatedness to the other individuals is necessarily negative, specifically $R = -1/(N - 1)$. This is minimized at $R = -1$ when $N = 2$, but quickly tends to zero as N increases towards more plausible values.

partners (kin discrimination), then the relatedness to recipients becomes even more negative (Hamilton, 1975). For very small populations (small N ; Fig. 3), negative relatedness can be nontrivial, and hence individuals might be expected to pay reasonable costs in order to inflict damage to their social partners. Negative relatedness (and hence spite) is therefore possible, but this tiny population condition caused Hamilton (1971) to regard spite as merely the ‘final infection that kills failing twigs of the evolutionary tree’, and not a general phenomenon contributing to adaptive evolution (Hamilton, 1996).

Scale of competition

However, the situation may not be so bleak for spite. There has recently been much interest in how local competition between relatives can reduce and even remove selection for altruism between relatives (reviewed by Queller, 1992; West *et al.*, 2002). This work was spurred by the possibility that with limited dispersal in a viscous population, individuals would tend to associate with kin, so that kin selection theory might suggest positive relatedness between social partners, and hence conditions favourable for the evolution of altruism (Hamilton, 1964, 1971, 1972, 1975, 1996). However, this relies on the implicit assumption of density-dependent regulation being global (hard selection; Wallace, 1968), with no increased competition, due to increased productivity, within more altruistic groups (Boyd, 1982; Wade, 1985). In contrast, if density-dependent regulation occurs at the level of the social group (soft selection, Wallace, 1968; see also Haldane, 1924), then the increased success of the recipient must be paid for by the group. Without kin discrimination, the relatedness of the actor to the

other members of the group will have been equally raised by population viscosity. Hence, population viscosity will not necessarily favour indiscriminate altruism (Hamilton, 1971, 1975; Taylor, 1992a,b).

This effect of local competition between relatives can be incorporated into Hamilton’s rule in a number of ways (Grafen, 1984; Queller, 1994; Frank, 1998; West *et al.*, 2002). Queller (1994) reformulated the coefficient of relatedness in order to take this into account, giving a new measure which he described as ‘not just a statement about the genetic similarity of two individuals, it is also a statement about who their competitors are’. Here, relatedness between actor and recipient is a regression as before, however it is now defined relative to a reference population of competitors, a proportion of which are locals, and the remainder being average members of the global population. Obviously if all competition is global, the reference population is the global population, allowing for positive relatedness between social partners. At the other extreme, if all competition is at the level of the social group, relatedness to the average member of the social group will be zero. Frank (1998) chose not to redefine relatedness, but instead introduced a separate scale of competition parameter to be incorporated into the benefit component of Hamilton’s rule in order to predict when social behaviours will be favoured by selection. This parameter (a) is simply the proportion of competitors which are local as opposed to global. Soft selection (local competition) had been relatively neglected in social evolution theory prior to these developments, and this contrasts with population genetics, where it has received much attention (Roughgarden, 1979).

Although the importance of the scale of competition in the application of kin selection to altruism is now acknowledged (see West *et al.*, 2002 for a recent review, and Griffin *et al.*, 2004 for an empirical example), its implications for spite are underappreciated. Increasingly local competition, as well as disfavouring altruism, can enhance selection for spite. Hamilton was correct when he stated that spite should be restricted to tiny populations; however, the ‘population’ of interest is that of the competitive arena. If competition is global, so that there is hard selection at the level of the social arena, then relatedness is measured with respect to the population as a whole. But as competition becomes increasingly local, the reference population shrinks towards the size of the social arena, which may contain only a few individuals (small N) and/or a significant proportion of identifiable positively related kin, such that the negative relatedness towards the other potential recipients is nontrivial, enhancing the selective value of spite.

Another way of seeing this is by considering a crucial difference between altruism and spite. Within a social group, individuals with greater altruism than the group average have reproductive success lower than the group average, but if more altruistic groups are more productive, altruists may have higher absolute success than

nonaltruists when averaging over the whole population. When competition is global, fitness is proportional to absolute success, so that altruism can be a winning strategy. Increasingly local competition means that fitness is increasingly dominated by success relative to the social group average, and so altruism is less favoured. Conversely, spiteful behaviour incurs a direct cost and reduces the success of social partners, so that more spiteful individuals can have higher success relative to the group average, but suffer a reduction in absolute success. When competition is global and fitness is proportional to absolute success, spite cannot be favoured, but as competition becomes increasingly local fitness is increasingly determined by success relative to social partners, so that spite can be a winning strategy.

Illustrative overview

So far we have employed the standard approach of taking Hamilton's rule to be a given (for example, see Orlove, 1975) and using this as an entry point into the analysis of social evolution. However, it is often more appropriate and rigorous to derive the rule using a direct fitness approach, particularly when the aim is to resolve problematic conceptual issues. We use the direct (neighbour-modulated) fitness maximization techniques of Taylor & Frank (1996) and Frank (1998) to derive Hamilton's rule, in order to (i) distinguish two different forms of spite, and (ii) address the suggestion of Boyd (1982) that spite is often actually selfishness because it indirectly increments fitness through reducing the intensity of competition. The key to this is to distinguish possible direct benefits of spite that might accrue to positively related third parties, and indirect effects due to relaxed competition.

Let social groups comprise n equally abundant 'families', with kin recognition allowing discrimination of the proportion $1/n = k$ of the social group which are 'kin' from the remaining $1 - k$ which are 'nonkin'. Spite directed at nonkin carries a cost (some function c), inflicts a negative benefit upon the victim (b), and also potentially directly benefits (d) individuals within one's family, so that personal success might be written as:

$$S_{\text{focal}} = 1 + b[(1 - k)z] - c[x] + d[k(1 - k)y], \quad (1)$$

where x is the focal individual's spite strategy, y is the average strategy of its kin (including itself), and z is the average strategy played by the nonkin members of its social group. The local average and the average for the whole population are given by:

$$\begin{aligned} S_{\text{local}} &= 1 + k\{b[(1 - k)z] - c[y] + d[k(1 - k)y]\} \\ &\quad + (1 - k)\{b[(1 - 2k)z + ky] - c[z] + d[k(1 - k)z]\}, \\ S_{\text{global}} &= 1 + b[(1 - k)\bar{z}] - c[\bar{z}] + d[k(1 - k)\bar{z}], \end{aligned} \quad (2)$$

where \bar{z} is the average spite strategy played in the whole population. Following Frank's (1998) approach to including competition in models of social evolution,

fitness can be expressed as success relative to that of the average competitor, i.e.

$$w = \frac{S_{\text{focal}}}{aS_{\text{local}} + (1 - a)S_{\text{global}}}, \quad (3)$$

where the scale of competition parameter (a) is defined as the proportion of competition which occurs locally, i.e. at the level of the social group. Selection favours more spite whenever marginal fitness is positive ($dw/dx > 0$). As outlined by Taylor & Frank (1996), and Frank (1998), marginal fitness is given by the chain rule:

$$\frac{dw}{dx} = \frac{\partial w}{\partial x} + r_y \frac{\partial w}{\partial y} + r_z \frac{\partial w}{\partial z}, \quad (4)$$

where ∂ denotes a partial derivative, and $r_y = dy/dx$ and $r_z = dz/dx$ are the slopes of social partner phenotype on own phenotype (for kin and nonkin respectively). Assuming only minor variants ($x \approx y \approx z \approx \bar{z}$), and denoting $b' = db[\circ]/d\circ$, $c' = dc[\circ]/d\circ$ and $d' = dd[\circ]/d\circ$, we find that marginal fitness is positive ($dw/dx > 0$) when

$$\begin{aligned} &\{r_z - a[kr_y + (1 - k)r_z]\}(1 - k)b' \\ &+ \{r_y - a[kr_y + (1 - k)r_z]\}k(1 - k)d' \\ &> \{1 - a[kr_y + (1 - k)r_z]\}c'. \end{aligned} \quad (5)$$

Note that the relatedness to the average competitor relative to the whole population is $\hat{r} = a[kr_y + (1 - k)r_z]$, and the marginal costs and benefits of spite are $B = (1 - k)b'$, $C = c'$, and $D = k(1 - k)d'$. After making these substitutions, rearrangement of eqn 5 obtains the condition

$$\frac{r_z - \hat{r}}{1 - \hat{r}}B + \frac{r_y - \hat{r}}{1 - \hat{r}}D > C. \quad (6)$$

The r terms denote relatedness of individuals with respect to their spite phenotypes, relative to the population average, \bar{z} . If R is used to denote relatedness *sensu* Queller (1994), i.e. measured relative to the average competitor, then eqn 6 is simply

$$R_1B + R_2D > C. \quad (7)$$

This is the three-party extension to Hamilton's rule for spiteful interactions given by Foster *et al.* (2000), although here it is the consequence of an analysis rather than the starting point. R_1 is the relatedness to the victims of spite, and R_2 is the relatedness to the third party which receives any direct benefits. A major source of confusion over Hamilton's rule involves the meaning of the terms B and C (and in the above expression, D), and so it is worth pointing out that these are not fixed parameters – they are marginal values.

This form of the rule can be used to discriminate Hamiltonian and Wilsonian forms of spite (Hamilton, 1970, 1971; Wilson, 1975; Foster *et al.*, 2000, 2001). Feeling that negative relatedness was implausible, Wilson (1975) proposed that spite directed against non-negatively related individuals could be favoured if it also delivered a benefit to a sufficiently positively related

third party. In terms of the above notation, such Wilsonian spite occurs when $D > 0$ and $R_2 > 0$, and does not require a negatively related victim ($R_1 < 0$). Hamiltonian spite occurs when the victim is negatively related ($R_1 < 0$, and hence $R_1 B > 0$; Hamilton, 1970, 1971), and hence a direct benefit to positive relations ($D > 0$) is not always required in order for the spite to be favoured. From eqn 6 we can see that: (i) negative relatedness depends on the ability to discriminate individuals who are less related than the average competitor (so that $r < \hat{r}$); and (ii) the magnitude of this negative relatedness increases as competition becomes more localized (increasing a , and hence increasing \hat{r}). Clearly, there is potential for spiteful behaviours to involve both negative relatedness to victims and positive benefits to positive relations, and hence a mixture of Hamiltonian and Wilsonian spite (Foster *et al.*, 2000, 2001).

Related to this distinction, we can address the suggestion of Boyd (1982) that spite is actually less likely to occur under local competition, as the resulting relaxed competition gives an indirect benefit to spiteful individuals, so that many cases of spite would in fact be selfishness. Equation 7 reveals that the relaxation of competition due to spite is absorbed into the negative relatedness term, when relatedness is measured relative to the average competitor. Boyd's indirect benefit to the spiteful individual does not make the action selfish, in the same way that this indirect benefit accrued to other positive relatives does not mean that the spiteful behaviour is Wilsonian.

It is important to note that the above is not a general model for spite, but is rather an example included for the purpose of illustration. For instance, we have assumed additivity of fitness components, and equally abundant families. For this reason, it is always more rigorous to do a direct fitness analysis for particular models of interest in order to obtain the appropriate Hamilton's rule, rather than using the rule as a starting point.

Biological applications

Applying the theory to biological examples, we show: (i) that previous models which have successfully demonstrated selection for spite have tended to implicitly assume local competition; (ii) behaviours previously interpreted as indirect altruism or Wilsonian spite might turn out to involve negative relatedness and hence Hamiltonian spite; (iii) spiteful greenbeards are more likely to reach their threshold frequency, above which they are favoured by selection, when competition is localized; (iv) there are several general predictions which will help us identify situations where spite is likely to be found, and (v) these predictions are amenable to empirical testing.

Spiteful models assume local competition

Theoretical models that show that spiteful behaviour can be favoured often assume that some or all of competition

is local. However, this has rarely been acknowledged as an important factor contributing to the success of spite. For example:

1. Reinhold (2003) used an inclusive fitness analysis to investigate fatal fighting in fig wasps. This model shows selection for spite when competition is completely local. Some fig wasps have a lifecycle, such that wingless males hatch, mate and die within the confines of the fruit, and the mated females disperse to be the foundresses of new figs (Hamilton, 1979; Cook *et al.*, 1997). This leads to an asymmetric scale of competition, such that males compete locally (for mates) and females compete globally (for figs in which to lay eggs), the consequences of which for sex allocation theory have been much studied (Hamilton, 1967; Herre *et al.*, 2001). In some species, this local competition for mates is accompanied by lethal combat between heavily armoured males, which have mandibles capable of decapitating each other (Hamilton, 1979; Murray, 1987; West *et al.*, 2001). Reinhold (2003) predicted that if males could discriminate between relatives and nonrelatives (kin recognition) then they would be selected to fight with males who are nonrelatives. This cannot be explained simply as selfishness because there is generally a net direct fitness cost of fighting (the difference in the direct fitness component of Reinhold's equations 2.1 and 2.2 for the terms T_1 & T_2). However, it can be explained as Hamiltonian spite, because the local competition means there is a negative relatedness towards opponents. Following Reinhold's notation, n males compete locally for matings, including a focal actor who is related to a proportion y of the other males (his brothers) by r and to the remaining $(n-1)(1-y)$ males (nonkin) by zero. Rescaling such that the focal individual is related to competitors on average by zero, we find that the relatedness to his brothers is $[n(1-y)r - (1-ry)] / [(n-1)(1-ry)]$ and to the unrelated males is $-[1 + (n-1)ry] / [(n-1)(1-ry)]$, i.e. a negative quantity. The importance of spite in this system depends upon the possibility of kin discrimination between male fig wasps, which has yet to be tested for.

2. Gardner *et al.* (2004) presented a model of chemical (bacteriocin) warfare between microbes. Bacteriocins are the most abundant of a range of antimicrobial compounds produced by bacteria, and are found in all major bacterial lineages (Riley & Wertz, 2002). They are a diverse family of proteins with a range of antimicrobial killing activity including enzyme inhibition, nuclease activity and pore formation in cell membranes (Reeves, 1972; Riley & Wertz, 2002). They are distinct from other antimicrobials in that their lethal activity is often limited to the same species of the producer, suggesting a major role in competition with conspecifics (Riley *et al.*, 2003). As bacteriocin synthesis is energetically expensive and release can entail death of the producer cell (for instance, colicin production by *Escherichia coli*) production of bacteriocins is costly ($C > 0$). Bacteriocins kill susceptible bacteria, and hence these recipients suffer a negative

benefit ($B < 0$). Hence bacteriocin production can be regarded as a spiteful trait. As kin of the producer cell are immune to its bacteriocins, there is effective kin discrimination, and the potential for recipients to be negatively related to the producer. Specifically, this relatedness is $R = -(ak)/(1 - ak)$ where k is the proportion of the social group which are clonal kin of the producer, and a is the proportion of competition which occurs locally. This reveals the importance of local competition in the evolution of spiteful behaviour. Specifically, (i) spiteful bacteriocin production is only selected for when there is some local competition ($a > 0$; as $R = 0$ when $a = 0$), and (ii) as the degree of local competition (a) increases the evolutionary stable strategy (Maynard Smith & Price, 1973) is to increasingly allocate resources to spiteful bacteriocin production (Gardner *et al.*, 2004).

3. Cytoplasmic Incompatibility (CI), the phenomenon whereby maternally transmitted *Wolbachia* (and other) bacteria occurring in male hosts sterilize uninfected female hosts upon mating (O'Neill *et al.*, 1997), has been interpreted as a form of spite (Hurst, 1991; Foster *et al.*, 2001). Infected females are compatible with infected males, and so there is effective discrimination of carriers and noncarriers of the parasite. The question of whether it can be favoured by selection has received much attention (Prout, 1994; Turelli, 1994; Frank, 1997b). Frank (1997b) demonstrated that selection can favour CI in structured host populations. In his model, the sterilization of uninfected females relaxes competition for the infected progeny produced by the group. In particular, Frank highlighted the importance of kin associations, so that related bacteria are carried by several hosts within the group. Less emphasis was given to the assumption of density-dependent regulation at the group level, so that all competition is local ($a = 1$). Similar reasoning can be applied to the evolution of such selfish elements as maternal-effect lethal distorter genes (Beeman *et al.*, 1992; Hurst, 1993; Hurst *et al.*, 1996; Foster *et al.*, 2001), in which the killing of noncarriers relaxes competition among the carriers of the killer allele.

Hamiltonian and Wilsonian spite

Equation 7 can be used to discriminate between Hamiltonian and Wilsonian forms of spite, and assess their relative importance when both occur (i.e. when spite is directed at negatively related individuals but also accrues a net inclusive fitness benefit by directly enhancing the success of positive relations). In particular, using measures of relatedness that take into account the effects of competition, we can reinterpret many putative examples of Wilsonian spite as Hamiltonian spite or a mixture of the two. For instance, Foster *et al.* (2000, 2001) present two spiteful behaviours presented by the eusocial insects which they describe as Wilsonian: worker policing and sex allocation manipulation.

Often in eusocial hymenopteran societies, worker individuals do not have the opportunity to mate, but nevertheless have functioning ovaries, and can therefore produce unfertilized eggs which may develop as haploid males (Wilson, 1971; Bourke, 1988). Worker policing, the phenomenon whereby workers eat the eggs of other workers in their colony (Ratnieks, 1988), is well documented (Ratnieks & Visscher, 1989; Foster & Ratnieks, 2000, 2001; Barron *et al.*, 2001; Foster *et al.*, 2002). Foster *et al.* (2000, 2001) argue that this costly policing behaviour enhances the inclusive fitness of the actor as it frees up resources for the queen's sons (their brothers), to which they are more related than the sons of other workers (their nephews), and hence the spite is of the Wilsonian form. However, given that competition between the progeny for resources is within the colony, it is appropriate to measure relatedness with respect to this local competitive arena when assessing the inclusive fitness consequences for this particular behaviour. This means that the victim of the policing (a nephew) is less related than average (all brothers and nephews) and hence negatively related to the actor (i.e. $R_1 < 0$). Consequently, if relatedness is measured at the scale of competition, worker policing can be interpreted as involving Hamiltonian spite.

The haplodiploid genetics of the hymenoptera means that in eusocial species the workers can be more related to their diploid sisters than their haploid brothers. This means that, while the queen prefers equal sex allocation among reproductives, the workers would rather there was a female bias (Trivers & Hare, 1976). In some species the workers create this bias by killing male progeny (Passera & Aron, 1996; Sundström *et al.*, 1996; Chapuisat & Keller, 1999; Hammond *et al.*, 2003). Foster *et al.* (2000, 2001) suggest this killing of male progeny is Wilsonian spite that benefits the colony's female progeny. However, the local competition for resources within the colony, plus the fact that males are devalued relative to females in terms of relatedness, means that the recipient of the spite is negatively related to the actor ($R_1 < 0$). Again, this behaviour may be reinterpreted involving Hamiltonian spite.

Application of the theory should also allow reinterpretation of behaviours which have not been considered spiteful (Hamiltonian or otherwise) in the past. Precocious larval development in polyembryonic parasitoid wasps (Godfray, 1992; Grbic *et al.*, 1992; Hardy *et al.*, 1993; Ode & Strand, 1995) seems to constitute a previously overlooked example of spite. Typically, two eggs, one male and one female, are laid on or in the body of the host insect, which then divide asexually to produce a brood of genetically identical brothers and genetically identical sisters. Local competition for resources limits the number of adult wasps emerging from the host, suggesting that there is scope for negative relatedness between the sexes within the brood. Upon inspection, some of the larvae that have not emerged as

adults are found to have developed precociously, giving up their own future reproduction in order to murder opposite-sex siblings developing in the same host. Asymmetric dispersal (which generates a sex difference in the scale of competition), and asymmetric relatedness (brothers are more related to sisters than vice versa) seem to be responsible for evolutionary resolution of this conflict in favour of the sisters, such that most precocious larvae are female.

Local competition can enhance the success of spiteful greenbeards

Greenbeards are phenotypic markers for genetic composition that allow individuals to identify positive relations more effectively than through discrimination of genealogical kin from nonkin (Hamilton, 1964, 1971; Dawkins, 1976). A greenbeard gene causes three things: (i) a phenotypic trait, (ii) recognition of this trait in others, and (iii) preferential treatment of those recognized – see Queller *et al.* (2003) for an example of a single gene which satisfies these three conditions. From the perspective of the greenbearded actor, social partners displaying the phenotype carry his gene and hence are positively related to him, and those who do not display the phenotype do not carry his gene, and are therefore negatively related to him, with respect to that locus. Greenbeards can therefore increase in frequency either by directing altruism towards the positive relations or else by directing spite towards the negative relations. However, nontrivial negative relatedness is only possible when the greenbeard allele is at a substantial frequency in the population, as Hamilton (1971) understood, making it difficult to imagine how a spiteful greenbeard could initially be selected. This problem is not felt by altruistic greenbeards, which have maximal relatedness between bearers of the gene even when the greenbeard is at low frequency in the population. The understanding that is the arena of competition that provides the appropriate reference, rather than the population as a whole, means that the spread of spiteful greenbeards can be more easily understood, and the attainment of the threshold frequency does not have to rely upon assumptions such as extreme stochastic fluctuations.

Foster *et al.* (2000, 2001) discuss the example of the red fire ant (*Solenopsis invicta*; see also Keller & Ross, 1998 and Hurst & McVean, 1998), in which workers with genotype Bb, under the influence of their greenbeard (b) gene, murder negatively related BB queens and hence increase the frequency of the b gene in the population (homozygotes for the greenbeard gene are absent as the bb genotype is lethal). It is easy to see how the frequency of the b allele among the small number of locally competing queens could, through sampling error, exceed the threshold even as the global frequency approaches zero.

Where should we expect spite?

The extensions to spite theory, and biological examples of spite discussed above, suggest several clues as to where we should expect such behaviours to occur. Hamilton (1970, 1971) noted that spite should be more prevalent when actors are in a position to inflict damage to others at little cost to themselves, and so it is unsurprising that many examples turn up among nonreproductives in the eusocial insects, those individuals who have little or nothing to lose with respect to their direct fitness (Foster *et al.*, 2000, 2001). A major factor which has received much attention is the ability to identify one's negative relations. This can be achieved through recognition of genealogically close individuals (kin discrimination) or by means of phenotypic markers for genetic composition (greenbeards). We also emphasize that spite should be looked for in situations where competition is mostly local (among social partners), and in viscous populations.

Empirical testing of spite theory

Previous debate over spite has focused primarily on whether spite occurs. However, some of the more recent examples, such as worker policing in the eusocial insects and bacteriocin production by bacteria, provide possibilities for testing whether the relative occurrence of spite varies as predicted by social evolution theory. Indeed, much of the data from the eusocial insects fits well with the predictions of the theory (Chapuisat & Keller, 1999; Ratnieks *et al.*, 2001). Here we emphasize two general points.

1. We used Hamilton's rule to give an overall conceptual view. However, if particular cases are to be analysed, then it is often much easier and more rigorous to start with an equation for direct (neighbour-modulated) fitness based upon the relevant biology, and then derive predictions (Taylor & Frank, 1996; Frank, 1998). Hamilton's rule in some form usually appears as a consequence of such an approach, and provides a conceptual tool that can be used for interpretation of the results (Frank, 1998; Pen & Weissing, 2000; West & Buckling, 2003; Gardner *et al.*, 2004).

2. A relatively general prediction that arises from different models is that the incidence of spite should be dome shaped in relation to the degree of kinship within a social group. If the proportion of kin (including oneself) in the group is vanishingly small then no spite is favoured, as the nonkin recipients of spite will have the same relatedness, on average, as the average group member (i.e. zero). Similarly, when the actor associates solely with clonal kin, spite is also not favoured, as there are no negatively related individuals present. However, when the degree of kinship takes intermediate value, some degree of spite might be favoured because some individuals will necessarily be less related to the actor than others, such that some will have below-average

(and hence negative) relatedness. This result was found in both the bacteriocin (Gardner *et al.*, 2004) and fig wasp mortal combat (Reinhold, 2003) examples discussed above. The relatedness differential also selects for spiteful sex allocation manipulation (brothers are less related than sisters) and worker policing (nephews are less related than sons and brothers) discussed above. As well as suggesting where we might find spite occurring in nature, these models give predictions that could be tested with observational or experimental studies.

Conclusion

Spite has been neglected by social evolution theory because a common, implicit assumption (global competition) in evolutionary models tends to diminish its selective advantage. We have demonstrated that many previously analysed behaviours can be readily interpreted as involving spite. Furthermore, theory has been developed to such a degree that we can make testable predictions about where spite is likely to be found and how it relates to the degree of competition and kinship between social partners.

Acknowledgments

We thank N. Barton, K. Foster, S. Frank, M. Kirkpatrick, N. Mehdiabadi, S. Nee, T. Sands, D. Shuker, W. Zuidema and an anonymous reviewer for comments and discussion, and NERC, BBSRC and The Royal Society for funding.

References

- Barron, A., Oldroyd, B.P. & Ratnieks, F.L.W. 2001. Worker policing and anarchy in *Apis*. *Behav. Ecol. Sociobiol.* **50**: 199–208.
- Beeman, R.W., Friesen, K.S. & Denell, R.E. 1992. Maternal-effect selfish genes in flour beetles. *Science* **256**: 89–92.
- Bourke, A.F.G. 1988. Worker reproduction in the higher eusocial Hymenoptera. *Q. Rev. Biol.* **63**: 291–311.
- Boyd, R. 1982. Density-dependent mortality and the evolution of social interactions. *Anim. Behav.* **30**: 972–982.
- Chapuisat, M. & Keller, L. 1999. Testing kin selection with sex allocation data in eusocial Hymenoptera. *Heredity* **82**: 473–478.
- Cook, J.M., Compton, S.G., Herre, E.A. & West, S.A. 1997. Alternative mating tactics and extreme male dimorphism in fig wasps. *Proc. R. Soc. Lond. B* **264**: 747–754.
- Dawkins, R. 1976. *The Selfish Gene*. Oxford University Press, Oxford.
- Foster, K.R. & Ratnieks, F.L.W. 2000. Facultative worker policing in a wasp. *Nature* **407**: 692–693.
- Foster, K.R. & Ratnieks, F.L.W. 2001. Convergent evolution of worker policing by egg eating in the honey bee and common wasp. *Proc. R. Soc. Lond. B* **268**: 169–174.
- Foster, K.R., Ratnieks, F.L.W. & Wenseleers, T. 2000. Spite in social insects. *Trends Ecol. Evol.* **15**: 469–470.
- Foster, K.R., Wenseleers, T. & Ratnieks, F.L.W. 2001. Spite: Hamilton's unproven theory. *Ann. Zool. Fennici* **38**: 229–238.
- Foster, K.R., Gulliver, J. & Ratnieks, F.L.W. 2002. Why workers do not reproduce: worker policing in the European hornet *Vespa crabro*. *Insectes Sociaux* **49**: 41–44.
- Frank, S.A. 1995. George Price's contributions to evolutionary genetics. *J. Theor. Biol.* **175**: 373–388.
- Frank, S.A. 1997a. The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* **51**: 1712–1729.
- Frank, S.A. 1997b. Cytoplasmic incompatibility and population structure. *J. Theor. Biol.* **184**: 327–330.
- Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press, Princeton, NJ.
- Gardner, A., West, S.A. & Buckling, A. 2004. Bacteriocins, spite and virulence. *Proc. R. Soc. Lond. B* **271**: 1529–1535.
- Godfray, H.C.J. 1992. Strife among siblings. *Nature* **360**: 213–214.
- Grafen, A. 1984. Natural selection, kin selection and group selection. In: *Behavioural Ecology*, 2nd edn (J. R. Krebs & N. B. Davies, eds), pp. 62–84. Blackwell, Oxford.
- Grafen, A. 1985. A geometric view of relatedness. *Oxf. Surv. Evol. Biol.* **2**: 28–89.
- Grbic, M., Ode, P.J. & Strand, M.R. 1992. Sibling rivalry and brood sex ratios in polyembryonic wasps. *Nature* **360**: 254–256.
- Griffin, A.S., West, S.A. & Buckling, A. 2004. Cooperation and competition in pathogenic bacteria. *Nature* (in press).
- Haldane, J.B.S. 1924. A mathematical theory of natural and artificial selection part I. *Trans. Camb. Philos. Soc.* **23**: 19–41.
- Hamilton, W.D. 1963. The evolution of altruistic behaviour. *Am. Nat.* **97**: 354–356.
- Hamilton, W.D. 1964. The genetical evolution of social behaviour I. *J. Theor. Biol.* **7**: 1–16.
- Hamilton, W.D. 1967. Extraordinary sex ratios. *Science* **156**: 477–488.
- Hamilton, W.D. 1970. Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**: 1218–1220.
- Hamilton, W.D. 1971. Selection of selfish and altruistic behaviour in some extreme models. In: *Man and Beast: Comparative Social Behaviour* (J. F. Eisenberg & W. S. Dillon, eds), pp. 57–91. Smithsonian Press, Washington, DC.
- Hamilton, W.D. 1972. Altruism and related phenomena, mainly in social insects. *Annu. Rev. Ecol. Syst.* **3**: 193–232.
- Hamilton, W.D. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. In: *Biosocial Anthropology* (R. Fox, ed.), pp. 133–153. Malaby Press, London.
- Hamilton, W.D. 1979. Wingless and fighting males in fig wasps and other insects. In: *Reproductive Competition, Mate Choice, and Sexual Selection in Insects* (M. S. Blum & N. A. Blum, eds), pp. 167–220. Academic Press, New York.
- Hamilton, W.D. 1996. *Narrow Roads of Geneland Volume 1: Evolution of Social Behaviour*. Freeman, Oxford.
- Hammond, R.L., Bruford, M.W. & Bourke, A.F.G. 2003. Ant workers selfishly bias sex ratios by manipulating female development. *Proc. R. Soc. Lond. B* **269**: 173–178.
- Hardy, I.C.W., Ode, P.J. & Strand, M.R. 1993. Factors influencing brood sex-ratios in polyembryonic Hymenoptera. *Oecologia* **93**: 343–348.
- Herre, E.A., Machado, C. & West, S.A. 2001. Selective regime and fig wasp sex ratios: towards sorting rigor from pseudo-rigor in tests of adaptation. In: *Adaptation and Optimality* (S. Orzack & E. Sober, eds), pp. 191–218. Cambridge University Press, Cambridge.

- Hurst, L.D. 1991. The evolution of cytoplasmic incompatibility or when spite can be successful. *J. Theor. Biol.* **148**: 269–277.
- Hurst, L.D. 1993. *Scat+* is a selfish gene analogous to *Medea* of *Tribolium castaneum*. *Cell* **75**: 407–408.
- Hurst, G.D.D. & McVean, G.A.T. 1998. Selfish genes in a social insect. *Trends Ecol. Evol.* **13**: 434–435.
- Hurst, L.D., Atlan, A. & Bengtsson, B.O. 1996. Genetic conflicts. *Q. Rev. Biol.* **71**: 317–364.
- Keller, L. & Ross, K.G. 1998. Selfish genes: a green beard in the red fire ant. *Nature* **394**: 573–575.
- Malécot, G. 1948. *Les mathématiques de l'hérédité*. Masson, Paris.
- Maynard Smith, J. & Price, G.R. 1973. The logic of animal conflict. *Nature* **246**: 15–18.
- Michod, R.E. 1982. The theory of kin selection. *Ann. Rev. Ecol. Syst.* **13**: 23–55.
- Murray, M.G. 1987. The closed environment of the fig receptacle and its influence on male conflict in the Old-World fig wasp, *Philotrypsis pilosa*. *Anim. Behav.* **35**: 488–506.
- O'Neill, S.L., Hoffmann, A.A. & Werren, J.H. 1997. *Influential Passengers: Inherited Microorganisms and Arthropod Reproduction*. Oxford University Press, Oxford.
- Ode, P.J. & Strand, M.R. 1995. Progeny and sex allocation decisions of the polyembryonic wasp *Copidosoma floridanum*. *J. Anim. Ecol.* **64**: 213–224.
- Orlove, M.J. 1975. A model of kin selection not invoking coefficients of relationship. *J. Theor. Biol.* **49**: 289–310.
- Passera, L. & Aron, S. 1996. Early sex discrimination and male brood elimination by workers of the Argentine ant. *Proc. R. Soc. Lond. B* **263**: 1041–1046.
- Pen, I. & Weissing, F.J. 2000. Towards a unified theory of cooperative breeding: the role of ecology and life history re-examined. *Proc. R. Soc. Lond. B* **267**: 2411–2418.
- Pepper, J.W. 2000. Relatedness in trait group models of social evolution. *J. Theor. Biol.* **206**: 355–368.
- Price, G.R. 1970. Selection and covariance. *Nature* **227**: 520–521.
- Prout, T. 1994. Some evolutionary possibilities for a microbe that causes incompatibility in its host. *Evolution* **48**: 909–911.
- Queller, D.C. 1985. Kinship, reciprocity, and synergism in the evolution of social behaviour. *Nature* **318**: 366–367.
- Queller, D.C. 1992. Does population viscosity promote kin selection? *Trends Ecol. Evol.* **7**: 322–324.
- Queller, D.C. 1994. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**: 70–73.
- Queller, D.C., Ponte, E., Bozzaro, S. & Strassmann, J.E. 2003. Single-gene greenbeard effects in the social amoeba *Dictyostelium discoideum*. *Science* **299**: 105–106.
- Ratnieks, F.L.W. 1988. Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. *Am. Nat.* **132**: 217–236.
- Ratnieks, F.L.W. & Visscher, P.K. 1989. Worker policing in the honeybee. *Nature* **342**: 796–797.
- Ratnieks, F.L.W., Monnin, T. & Foster, K.R. 2001. Inclusive fitness theory: novel predictions and tests in eusocial Hymenoptera. *Ann. Zool. Fennici* **38**: 201–214.
- Reeves, P. 1972. *The Bacteriocins*. Springer-Verlag, New York.
- Reinhold, K. 2003. Influence of male relatedness on lethal combat in fig wasps: a theoretical analysis. *Proc. R. Soc. Lond. B* **270**: 1171–1175.
- Riley, M.A. & Wertz, J.E. 2002. Bacteriocins: evolution, ecology, and application. *Annu. Rev. Microbiol.* **56**: 117–137.
- Riley, M.A., Goldstone, C.M., Wertz, J.E. & Gordon, D. 2003. A phylogenetic approach to assessing the targets of microbial warfare. *J. Evol. Biol.* **16**: 690–697.
- Roughgarden, J. 1979. *Theory of Population Genetics and Evolutionary Ecology: an Introduction*. Macmillan, New York.
- Seeger, J. 1981. Kinship and covariance. *J. Theor. Biol.* **91**: 191–213.
- Sundström, L., Chapuisat, M. & Keller, L. 1996. Conditional manipulation of sex ratios by ant workers: a test of kin selection theory. *Science* **274**: 993–995.
- Taylor, P.D. 1992a. Altruism in viscous populations – an inclusive fitness approach. *Evol. Ecol.* **6**: 352–356.
- Taylor, P.D. 1992b. Inclusive fitness in a heterogeneous environment. *Proc. R. Soc. Lond. B* **249**: 299–302.
- Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *J. Theor. Biol.* **180**: 27–37.
- Trivers, R.L. & Hare, H. 1976. Haplodiploidy and the evolution of the social insects. *Science* **191**: 249–263.
- Turelli, M. 1994. Evolution of incompatibility-inducing microbes and their hosts. *Evolution* **48**: 1500–1513.
- Wade, M.J. 1985. Soft selection, hard selection, kin selection and group selection. *Am. Nat.* **125**: 61–73.
- Wallace, B. 1968. Polymorphism, population size, and genetic load. In: *Population Biology and Evolution* (R. C. Lewontin, ed.), pp. 87–108. Syracuse University Press, Syracuse, NY.
- West, S.A. & Buckling, A. 2003. Cooperation, virulence and siderophore production in bacterial parasites. *Proc. R. Soc. Lond. B* **270**: 37–44.
- West, S.A., Murray, M.G., Machado, C., Griffin, A.S. & Herre, E.A. 2001. Testing Hamilton's rule with competition between relatives. *Nature* **409**: 510–513.
- West, S.A., Pen, I. & Griffin, A.S. 2002. Cooperation and competition between relatives. *Science* **296**: 72–75.
- Wilson, E.O. 1971. *The Insect Societies*. Harvard Press, Cambridge, MA.
- Wilson, E.O. 1975. *Sociobiology: the New Synthesis*. Harvard Press, Cambridge, MA.
- Wright, S. 1922. Coefficients of inbreeding and relationship. *Am. Nat.* **56**: 330–338.
- Wright, S. 1969. *Evolution and the Genetics of Populations II: the Theory of Gene Frequencies*. University of Chicago Press, Chicago, IL.

Received 3 December 2003; revised 13 April 2004; accepted 12 May 2004

Cooperation and Punishment, Especially in Humans

Andy Gardner* and Stuart A. West†

Institute of Evolutionary Biology, School of Biological Sciences,
University of Edinburgh, West Mains Road, Edinburgh EH9 3JT,
United Kingdom

*Submitted January 8, 2004; Accepted August 4, 2004;
Electronically published October 18, 2004*

ABSTRACT: Explaining altruistic cooperation is one of the greatest challenges faced by sociologists, economists, and evolutionary biologists. The problem is determining why an individual would carry out a costly behavior that benefits another. Possible solutions to this problem include kinship, repeated interactions, and policing. Another solution that has recently received much attention is the threat of punishment. However, punishing behavior is often costly for the punisher, and so it is not immediately clear how costly punishment could evolve. We use a direct (neighbor-modulated) fitness approach to analyze when punishment is favored. This methodology reveals that, contrary to previous suggestions, relatedness between interacting individuals is not crucial to explaining cooperation through punishment. In fact, increasing relatedness directly disfavors punishing behavior. Instead, the crucial factor is a positive correlation between the punishment strategy of an individual and the cooperation it receives. This could arise in several ways, such as when facultative adjustment of behavior leads individuals to cooperate more when interacting with individuals who are more likely to punish. More generally, our results provide a clear example of how the fundamental factor driving the evolution of social traits is a correlation between social partners and how this can arise for reasons other than genealogical kinship.

Keywords: kin selection, neighbor-modulated fitness, repression of competition, public-goods game, human evolution, policing.

Explaining cooperation at all levels of biological complexity remains one of the greatest problems for evolutionary biology (Hamilton 1964; Buss 1987; Maynard Smith and Szathmáry 1995). The question is, Why would an individual perform a costly altruistic behavior that benefits another individual? The solutions to this problem that

have attracted the most attention are when social partners are related (kin selection, in a general sense; Hamilton 1963, 1964, 1970) or when there is some mechanism for repressing competition between groups (see table 1), such as through repeated interactions/reputation (reciprocity; Trivers 1971; Alexander 1979, 1987; Frank 2003), policing (Ratnieks 1988; Frank 1995, 2003), and systems of rewards or punishments (Oliver 1980; Sigmund et al. 2001). The fundamental similarity between all these mechanisms is that they involve positive correlations between the behaviors played by social partners, which are crucial for the evolution of social behaviors (Hamilton 1975; Grafen 1985; Nee 1989; Frank 1998; Woodcock and Heath 2002).

Here, we are concerned with whether and how punishment can favor cooperation and how this translates into a selective benefit for punishers. The possible role of punishment has recently attracted much theoretical attention, especially with respect to its possible role in favoring cooperation among humans (Hirshleifer and Rasmusen 1989; Boyd and Richerson 1992; Sober and Wilson 1998; Sell and Wilson 1999; Fehr and Gächter 2000). However, the mechanism underlying these previous models is often not clear, and the models have been developed with little reference to related theory such as in the animal punishment literature (Clutton-Brock and Parker 1995; Clutton-Brock 1998) and Frank's (1998, 2003) recent synthesis of social evolution theory. The basic idea is that if punishment is sufficiently frequent and harsh, it can successfully maintain cooperative behavior. However, this solution forces us to consider the motivation of the punisher. Since a behavior that promotes a public good such as cooperation is in itself a second-order public good and is not expected to be without cost to the actor, punishment is open for exploitation by second-order free-riding individuals who cooperate but who fail to punish defectors (Oliver 1980). Punishment of second-order free riders can be invoked, but this opens up the possibility of third- and higher-order free riding (Ostrom 1990). Failure to maintain participation in a high-level public-goods game unravels participation in the lower levels. At first glance, punishment seems not to be a helpful addition to the problem of cooperation because all that is achieved is the replacement of one public-goods dilemma for another.

* Corresponding author; e-mail: andy.gardner@ed.ac.uk.

† E-mail: stu.west@ed.ac.uk.

Table 1: A simple classification of some mechanisms that promote the evolution of cooperative behaviors

Selection pressure	Fundamental concept	Costs	Benefits
Kin selection	Relatedness between social partners	Cost for actor	Benefit for recipient
Reciprocal altruism	Repression of competition	Cost for actor	Future benefit for actor
Policing	Repression of competition	Cost for actor	Benefits for group
Punishment	Repression of competition	Cost for actor and recipient	Indirect benefit through increased cooperation

However, it is generally true that punishment is cheap relative to the cost of cooperation. Consequently, it has been argued that any mechanism invoked to explain participation in public-goods games will more easily favor punishing (and hence also cooperation) than it would cooperation alone (Sober and Wilson 1998).

A Darwinian account of the evolution of cooperation through punishment requires that the punisher directly or indirectly receives a net benefit through punishing. Although costly punishment can ultimately enhance the direct fitness of the punisher if interactions tend to be extended or repeated with the same social partner (Frank 2003; e.g., sanctioning in plant-rhizobium mutualisms: Denison 2000; West et al. 2002*b*, 2002*c*; Kiers et al. 2003), animals including humans punish even when there is no mechanism ensuring repeat encounters (Fehr and Gächter 2002). Genealogical relationship between social partners is often considered low or absent, and so kin selection is given little attention in the existing literature. The favored Darwinian mechanisms that have received the most attention are group selection (Gintis 2000) and cultural group selection (Heinrich and Boyd 2001). A recent simulation study (Boyd et al. 2003) has suggested that since the incidence of defection declines as punishment becomes more frequent, the costs of punishment decline as it becomes common, so that even modest group selection may plausibly maintain punishment in humans.

In this article, we show that the evolution of punishment and cooperation may be investigated using the powerful direct fitness maximization techniques of Taylor and Frank (1996) and Frank (1998). This allows us to clarify the mechanisms at work and link previous theory to Frank's (1998, 2003) general framework. In particular, we link kin selection, group selection, and cultural group selection in terms of a generalized view of relatedness. We then reveal that it is not the relatedness between social partners per se that facilitates the evolution of punishing behavior. What is crucial is that there is a positive correlation between the punishment strategy played and cooperation received by an individual. Although such an association could arise from viscous population structure and interactions between kin, it may arise for other reasons. In particular, we demonstrate that even in the absence of relatedness it is possible for such an association, due to

facultative adjustment of cooperative behavior, to maintain punishment through selection acting at the level of the individual, rendering group selection and elaborate cultural practices unnecessary. More generally, the fact that a positive correlation between the behaviors of social partners is the fundamental factor favoring cooperation has been obscured by a focus on how this correlation can be produced by kinship, through the interactions of close relatives (Hamilton 1975; Frank 1998). Our results provide a clear example of how such positive correlations can arise without kin association.

Models and Analyses

Basic Model

We now present a simple model describing the coevolution of cooperation and punishment. This is intended to elucidate the general selection pressures involved—it is the simplest model that captures the essentials of the problem. We discuss our model in terms of humans because this is where much of the recent theoretical literature has been focused. However, the implications are general and could be applied to a variety of organisms. A role for punishment in the evolution of cooperation has been suggested in a variety of animals, including insects, birds, primates, and other mammals (Clutton-Brock and Parker 1995). We give some specific examples in the discussion when considering how our model may be tested empirically.

For simplicity, we suppose that individuals interact in pairs, with one (random) member of the pair being denoted player 1 and the other player 2. Player 1 may choose to cooperate (e.g., sharing food), in which case she loses fitness c and player 2 gains fitness b , or to defect (e.g., refusing to share food), such that neither player loses nor gains fitness from the interaction. Player 2 may respond to defection in two ways: either she punishes (e.g., by physically injuring player 1) at a cost a to herself in order to reduce player 1's fitness by d , or else she forgives (e.g., does nothing) in which case neither player gains nor loses fitness. The expected direct fitness of a focal individual might then be written as

$$w = \alpha - cx + bX - (1 - X)ya - (1 - x)Yd, \quad (1)$$

where the constant α is baseline fitness, x is the frequency with which that individual cooperates, X is the mean frequency of cooperation among her social partners, y is the frequency with which the individual punishes, given that her partner defects, and Y is the mean punishment strategy played by her social partners, that is, the probability that the focal individual is punished given that she defects. We assume that all competition is global. An important point is that punishment acts to directly reduce both the fitness of the actor and the fitness of her social group. Punishment is therefore fundamentally different from the policing models of Frank (1995, 1996, 2003) because policing directly reduces actor fitness but increases group fitness.

Coevolution of Cooperation and Punishment

We will consider the simultaneous evolutionary optimization of cooperation and punishment analogous to the evolution of policing analysis of Frank (1995), using the direct (neighbor-modulated) fitness maximization method of Taylor and Frank (1996) and Frank (1998). A small increase in a behavior is favored by selection if the derivative of fitness with respect to that behavior (termed “marginal fitness”) is >0 and disfavored when this derivative is <0 . Differentiating the focal individual’s fitness function (eq. [1]) with respect to her cooperating (x) and punishing (y) strategies obtains

$$\frac{dw}{dx} = -c + Yd + \frac{dX}{dx}(b + ya) - \frac{dy}{dx}(1 - X)a - \frac{dY}{dx}(1 - x)d, \quad (2a)$$

$$\frac{dw}{dy} = -(1 - X)a - \frac{dY}{dy}(1 - x)d + \frac{dx}{dy}(Yd - c) + \frac{dX}{dy}(b + ya). \quad (2b)$$

The terms dX/dx and dY/dy are the coefficients of relatedness, with respect to cooperation and punishment, respectively, between the focal individual and her social partners (Taylor and Frank 1996; Frank 1998). Technically, the derivative is of the conditional expectation of the social partner’s strategy, given the strategy played by the focal individual, with respect to the latter. The other derivative terms are dy/dx and dx/dy , which are the regression of an individual’s punishing strategy on its own cooperation strategy, and vice versa, and dY/dx and dX/dy , which are the regressions of a partner’s punishing strategy on its own cooperation strategy and a partner’s cooperation strategy on its own punishment strategy, respectively.

Let us consider first the origin of cooperation and pun-

ishment in a population that is otherwise fixed for defection ($\bar{x} \rightarrow 0$) and forgiveness ($\bar{y} \rightarrow 0$). In such circumstances the trait-on-trait regressions are always nonnegative, which is important for interpretation of the analytical results that follows. To see why, consider the regression of cooperation received on cooperation strategy played: $dX/dx = (X - \bar{x})/(x - \bar{x}) \approx X/x$. Since cooperation strategies are nonnegative, the numerator (X) is nonnegative, and since the variant by definition plays a different cooperation strategy from the wild type (which plays zero cooperation), the denominator (x) is positive. Hence, $dX/dx \geq 0$. The same argument can be used to show that this is true for the other trait-on-trait regressions. Assuming only minor variants ($x \approx X \approx \bar{x}$, $y \approx Y \approx \bar{y}$; Taylor and Frank 1996; Frank 1998) and making the substitutions $\bar{x} \rightarrow 0$ and $\bar{y} \rightarrow 0$, the marginal fitness with respect to cooperation (eq. [2a]) reduces to

$$\frac{dw}{dx} = -c + \frac{dX}{dx}b - \frac{dy}{dx}a - \frac{dY}{dx}d. \quad (3)$$

This shows there is a direct cost (c) and a kin-selected benefit ($dX/dx \times b$) of cooperation, plus costs relating to the associated increase in costly punishing ($dy/dx \times a$) and also in being punished ($dY/dx \times d$); see figure 1A. Cooperation is maintained even in the absence of punishment when Hamilton’s (1964) rule $dX/dx \times b > c$ holds, so we will consider the more interesting situation where it does not, such that equation (3) is always negative.

Similarly, the marginal fitness with respect to punishment (eq. [2b]) is

$$\frac{dw}{dy} = -a - \frac{dY}{dy}d - \frac{dx}{dy}c + \frac{dX}{dy}b. \quad (4)$$

Again, this is easily understood. Punishing incurs a direct cost (a) and indirect costs ($dY/dy \times d$ from being punished by related individuals and $dx/dy \times c$ from the correlated commitment to cooperation). The benefit $dX/dy \times b$ is gained through the association between the punishment strategy played and the cooperation received (see fig. 1B). Only when this is sufficiently large may a rare variant with some small frequency of punishing behavior be able to invade. In other words, a positive association between the punishment strategy played and the cooperation received by a focal individual is a necessary but not sufficient condition for the evolutionary origin of punishment.

Result 1. A positive association between punishment strategy played and cooperation received is crucial for the evolutionary origin of punishing behavior.

We will now investigate the evolutionary maintenance of cooperation and punishment by considering $\bar{x} \rightarrow 1$ and

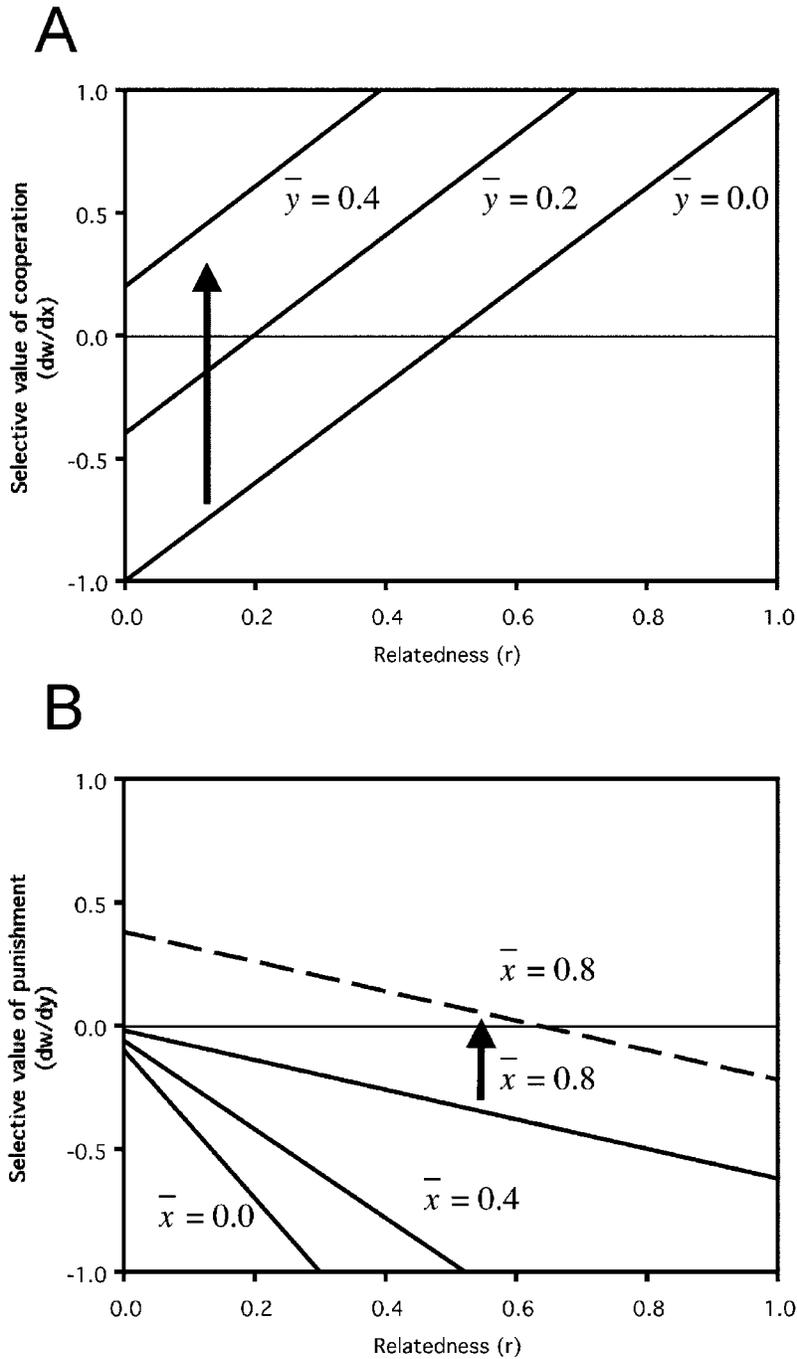


Figure 1: A, Selective value of cooperation (dw/dx) as a function of relatedness and the resident punishing strategy (\bar{y}) when there is no association between traits ($dy/dx = dY/dx = 0$); $dw/dx > 0$ indicates that enhanced cooperation is favored, and $dw/dx < 0$ indicates that it is disfavored. Increasing relatedness (r) enhances selection for cooperation; in the absence of punishment, cooperation is favored when $rb > c$. Increasing punishment also favors cooperation, so cooperation may be favored even when relatedness is 0, if $\bar{y} > c/d$. B, Selective value of punishment (dw/dy) as a function of relatedness and the resident cooperation strategy (\bar{x}); $dw/dy > 0$ indicates enhanced punishment is favored, and $dw/dy < 0$ indicates that it is disfavored. Assuming no association between traits ($dx/dy = dX/dy = 0$), we see that punishment is always disfavored, that increased relatedness enhances the selective disadvantage of punishment, and that increased cooperation reduces the selective disadvantage of punishment. Punishment may be favorable if there is a positive association between the punishment strategy played and the cooperation received by an individual ($dX/dy > 0$); the broken line indicates $dX/dy = 0.2$. For A and B, we assume $a = 0.1$, $b = 2$, $c = 1$, and $d = 3$.

$\bar{y} \rightarrow 1$. Again, the trait-on-trait regressions will all be non-negative: for example, $dX/dx = (X - \bar{x})/(x - \bar{x}) \approx (X - 1)/(x - 1)$. Cooperation received (X) cannot be >1 , so the numerator ($X - 1$) is ≤ 0 . Since the cooperation variant does not play the wild-type strategy (always cooperate) and cannot play a more cooperative strategy than that, the denominator ($x - 1$) is always negative. Hence, $dX/dx \geq 0$. Making the substitutions $\bar{x} \rightarrow 1$ and $\bar{y} \rightarrow 1$, the marginal fitness with respect to cooperation (eq. [2a]) is now given by

$$\frac{dw}{dx} = -c + d + \frac{dX}{dx}(b + a). \quad (5)$$

Here cooperation carries a direct cost (c) and a benefit (d , due to avoiding punishment) when punishment of defectors is assured. It also gives kin-selected benefits ($dX/dx \times b$ and $dX/dx \times a$) due to the correlated cooperation received from social partners and the fitness saved from not having to punish defectors. Punishment cannot be an effective deterrent when the fitness of a punished defector is greater than that of a cooperator, so that we will restrict attention to the situation $d > c$. Here, the marginal fitness will always be positive, and so selection will act to maintain cooperation. The marginal fitness with respect to punishment (eq. [2b]) is

$$\begin{aligned} \frac{dw}{dy} = & -(1 - \bar{x})a - \frac{dY}{dy}(1 - \bar{x})d \\ & + \frac{dx}{dy}(d - c) + \frac{dX}{dy}(b + a). \end{aligned} \quad (6)$$

The costs of punishment include the direct cost ($[1 - \bar{x}] \times a$) and the kin-selected cost ($[1 - \bar{x}] \times dY/dy \times d$) plus the cost incurred by the associated cooperation ($dx/dy \times c$). The benefits of punishment are due to the correlated decrease in one's own defection and hence the frequency with which the focal individual is punished ($dx/dy \times d$) and also the correlated increase in cooperation received from social partners ($dX/dy \times b$) and, conversely, the fitness saved by not having to punish partners ($dX/dy \times a$). If $dx/dy = dX/dy = 0$ so that there is no correlation between the punishment and cooperation played by an individual, nor between the punishment played and cooperation received, then the marginal fitness with respect to punishment is small but negative, and hence full punishment is not stable. It is interesting to note that relatedness dY/dy works to undermine the stability of punishment; as an individual's punishment strategy is increased, so too is the punishment received from social partners. If the between-trait associations are positive and of sufficient magnitude, then full punishment

can be evolutionarily stable. Otherwise, selection will act to reduce punishment in the population.

Result 2. A positive association between punishment strategy played and cooperation received is crucial for the evolutionary maintenance of punishing behavior.

We now check to see whether punishment is easier to maintain than it is to initially invade an otherwise forgiving population, by evaluating $dw/dy|_{\bar{x}, \bar{y}=1} - dw/dy|_{\bar{x}, \bar{y}=0}$, that is, subtracting the right-hand side (RHS) of equation (4) from the RHS of equation (6) to obtain

$$d \left(\frac{dY}{dy} + \frac{dx}{dy} \right) + a \left(1 + \frac{dX}{dy} \right), \quad (7)$$

which is positive, so that RHS equation (4) is less than RHS equation (6), and hence the condition for increased punishment to be favored ($dw/dy > 0$) is more easily satisfied in a population of cooperators and punishers than in a population of defectors and forgivers. Similarly, the RHS of equation (3) is always negative under the relevant circumstances (i.e., when $dX/dx \times b < c$), and the RHS of equation (5) is always positive, so that the condition for enhanced cooperation to be favored ($dw/dx > 0$) is also more easily satisfied in punishing populations than in populations rife with defection and forgiveness.

Result 3. Punishing behavior is more easily maintained than it is originally evolved. Note that this assumes that relatedness and the between-trait regressions are constants. A fully dynamic analysis relaxing this assumption would require that we specify a more detailed (and hence less general) model and so is not pursued here because we aim only to abstract and elucidate the selection pressures involved in the evolution of punishment and cooperation.

Example: Cooperation as a Facultative Response to Punishment

The Model. We have found that relatedness between social partners is not crucial for costly punishment to be favored (indeed, increasing relatedness disfavors punishment) and that it is another association, the regression of the cooperation received on the punishment strategy played, that provides the benefit of punishment. To illustrate these findings, we examine the evolution of punishment when there is no relatedness between individuals ($dY/dy = 0$) and when cooperation is facultatively adjusted to one's punishment environment (which we will see can give $dX/dy > 0$).

We assume that individuals are randomly organized into social groups of size N , such that relatedness between group members is 0. In each social encounter, individuals pair with a random member from their group, with one

of the partners playing the role of player 1 and the other being player 2. In contrast with the previous model, we consider the cooperation strategy of player 1 to be facultative and hence a function of her punishment environment. Assuming no partner recognition and therefore no adjustment of cooperation to her current partner's punishment strategy, the cooperation strategy played by the focal individual (in half of her social interactions) is expressed as a function of the average punishment strategy played by all of her social partners: $x = f(\bar{y})$. Since each of her social partners experiences a punishing environment that includes the focal individual (and hence average punishment strategy among their social partners is $\bar{y} + [y - \bar{y}]/[N - 1]$), they will play cooperation strategy $X = f(\bar{y} + (y - \bar{y})/(N - 1))$.

If individuals cooperate optimally, we expect the function $f(Y)$ to be such that it maximizes the fitness of player 1 when player 2 plays punishment strategy Y . It is easy to show that this optimum is given by

$$f^*(Y) = \begin{cases} 0 & c > Yd \\ \text{if} & \\ 1 & c < Yd \end{cases}, \quad (8)$$

such that defection is favored when the cost of cooperation outweighs the threat of punishment ($c > Yd$), and cooperation is favored when the cost of cooperation is outweighed by the threat of punishment ($c < Yd$). This step function is both mathematically inconvenient and biologically unreasonable, so we will use the model of McNamara et al. (1997; see also Kokko 2003) to describe nearly optimized cooperation as

$$f(Y) = \frac{1}{1 + \exp(-\Delta/\varepsilon)} = \frac{1}{1 + \exp[-(Yd - c)/\varepsilon]}, \quad (9)$$

where ε is the degree of behavioral error and $\Delta = dw/dx = Yd - c$ ensures that the frequency of nonoptimal behavior declines as its impact on fitness becomes more important. The facultative cooperation function (eq. [9]) approaches the step function (eq. [8]) for vanishing behavioral error ($\varepsilon \rightarrow 0$), and for larger error ($\varepsilon > 0$), it takes a continuous sigmoidal form which flattens out to a constant $1/2$ as the error tends to infinity (fig. 2). For mathematical convenience, we will assume vanishing (but non-zero) behavioral error ($\varepsilon \rightarrow 0$).

Altering fitness function (eq. [1]) for this example model, we have the fitness of an individual who plays punishment strategy y , in a population with mean punishment strategy \bar{y} , given by

$$\begin{aligned} w &= \alpha - cf(\bar{y}) + bf\left(\bar{y} + \frac{(y - \bar{y})}{(N - 1)}\right) \\ &\quad - a\left[1 - f\left(\bar{y} + \frac{(y - \bar{y})}{(N - 1)}\right)\right]y \\ &\quad - d(1 - f(\bar{y}))\bar{y}. \end{aligned} \quad (10)$$

The mean fitness of the population is

$$\begin{aligned} \bar{w} &= \alpha - cf(\bar{y}) + bf(\bar{y}) \\ &\quad - a(1 - f(\bar{y}))\bar{y} - d(1 - f(\bar{y}))\bar{y}, \end{aligned} \quad (11)$$

so we expect a rare variant playing punishment strategy y to increase in frequency in a population with mean punishment strategy \bar{y} when the fitness differential $\Delta w = w - \bar{w}$ is positive, that is, when

$$\begin{aligned} \Delta w &= b\left[f\left(\bar{y} + \frac{(y - \bar{y})}{(N - 1)}\right) - f(\bar{y})\right] \\ &\quad - a\left[1 - f\left(\bar{y} + \frac{(y - \bar{y})}{(N - 1)}\right)\right]y - (1 - f(\bar{y}))\bar{y} > 0. \end{aligned} \quad (12)$$

Origin of Punishment. We first consider the evolutionary stability (Maynard Smith and Price 1973) of forgiveness, by determining under what circumstances no variant with punishment strategy $y > 0$ can invade a population with mean punishment strategy $\bar{y} \rightarrow 0$. Substituting the cooperation function (eq. [9]) into the fitness differential (eq. [12]) obtains

$$\begin{aligned} \Delta w &= b\left[\frac{1}{1 + \exp\{(c - [y/(N - 1)]d)/\varepsilon\}} - \frac{1}{1 + \exp(c/\varepsilon)}\right] \\ &\quad - a\left[1 - \frac{1}{1 + \exp\{(c - [y/(N - 1)]d)/\varepsilon\}}\right]y. \end{aligned} \quad (13)$$

Recalling that the behavioral error is vanishingly small ($\varepsilon \rightarrow 0$), we find that when the threat of punishment posed to social partners of the punishing variant is less than the cost of cooperation ($[yd]/[N - 1] < c$), then equation (13) reduces to $-ya$, which is negative, and hence the rare variant cannot invade. This is because defection is the rule in the social groups of both the wild type and the variant, giving population mean fitness $\bar{w} \approx \alpha$ and rare variant fitness $w \approx \alpha - ya$. When the threat of punishment is greater than the cost of cooperation ($[yd]/[N - 1] > c$), then equation (13) reduces to b , which is positive, and hence the rare variant can invade. Here, the rare punisher has managed to push her social group over the punishment threshold such that cooperation is now the optimal strat-

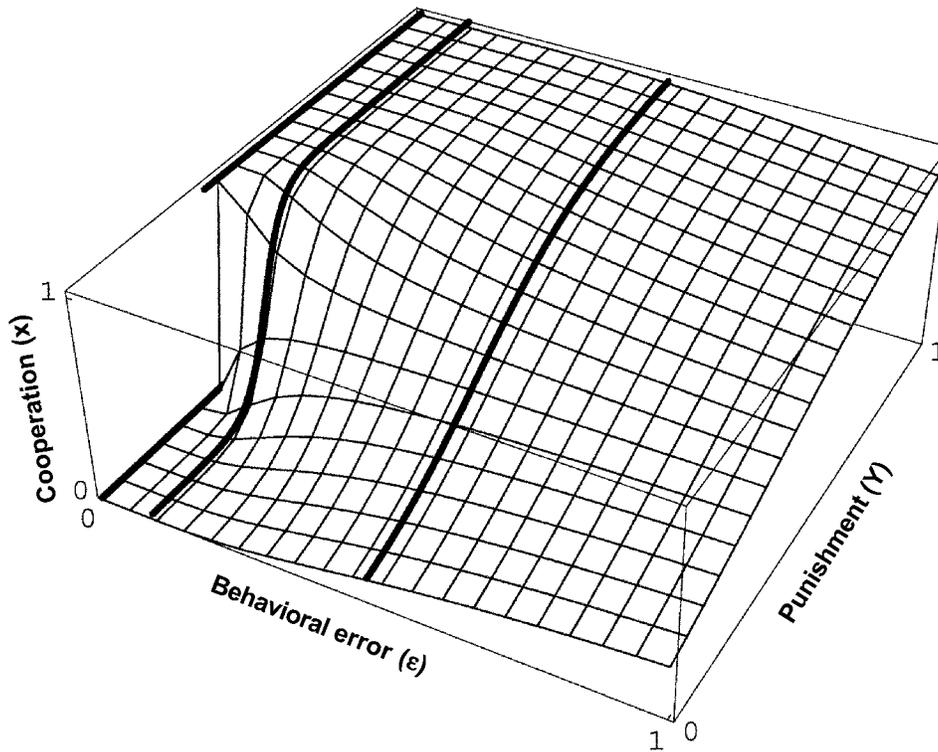


Figure 2: Frequency with which an individual cooperates (x) as a function of the punishment strategy of its social partners (Y) and the degree of behavioral error (ϵ), according to the example facultative model. Values are obtained numerically, assuming $c = 1$ and $d = 3$. The bold lines indicate $\epsilon = 0, 0.1, \text{ and } 0.5$.

egy. The average social group is fully defecting, so $\bar{w} \approx \alpha$, but the rare variant is now a recipient of cooperative behavior and only rarely encounters a defector requiring punishment, so that her fitness is $w \approx \alpha + b$. Note that although the variant receives cooperation, she maximizes her fitness by always defecting (since her unrelated social partners are all forgivers) and hence pays no cost of cooperation. If no y satisfies the above invasion condition, then forgiveness is an evolutionarily stable strategy (ESS; Maynard Smith and Price 1973). This is assured when $(N - 1)c > d$, so that not even a fully punishing variant ($y = 1$) can invade. Evolutionary stability of forgiveness is therefore assured unless

$$d > (N - 1)c. \quad (14)$$

Result 4. In the above model, punishment is unlikely to invade forgiveness unless the population is structured into very small groups.

Maintenance of Punishment. To determine whether punishment is an ESS, we let the wild type adopt the strategy of full punishment ($\bar{y} \rightarrow 1$) and consider the success of

rare variants playing $y < 1$. Substituting the facultative cooperation function (eq. [9]) into the fitness differential (eq. [12]) obtains

$$\Delta w = b \left\{ \frac{1}{1 + \exp \left(\frac{c - [1 - (1 - y)/(N - 1)]d}{\epsilon} \right)} - \frac{1}{1 + \exp \left[\frac{(c - d)}{\epsilon} \right]} \right\} + a \left\{ 1 - \frac{1}{1 + \exp \left[\frac{(c - d)}{\epsilon} \right]} - \left[1 - \frac{1}{1 + \exp \left(\frac{c - [1 - (1 - y)/(N - 1)]d}{\epsilon} \right)} \right] y \right\}. \quad (15)$$

First consider “ineffective punishment” ($c > d$). When behavioral error is vanishing ($\epsilon \rightarrow 0$), the fitness differential (eq. [15]) reduces to $a(1 - y)$, which is positive, and hence the more forgiving variant will always invade. This is because even when defection is always met with punishment, the defector has greater fitness than the cooperator, so that in all social groups defection is rife. The resident strategy incurs the cost of full punishment, and so the mean fitness of the population is $\bar{w} \approx \alpha - a$, whereas the more forgiving variant avoids this at least part of the time, giving fitness $w \approx \alpha - ya$. Now consider “effective punishment” ($d > c$),

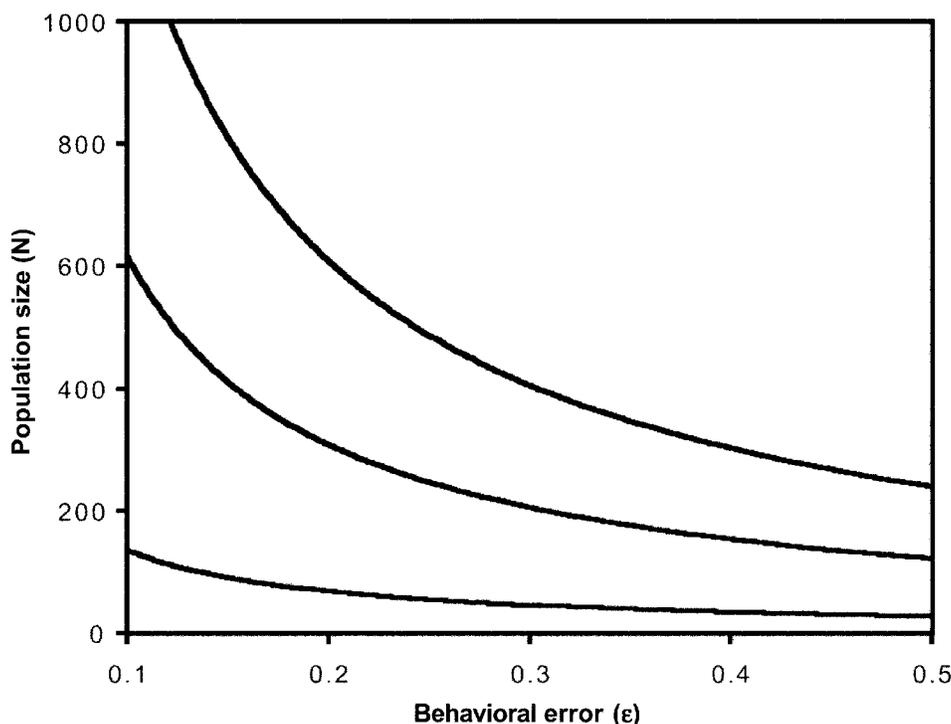


Figure 3: Maximum group size (N) permitting the evolutionary stability of punishment ($\bar{y} = 1$) as a function of behavioral error (ε) and the cost of punishing (a), according to the example facultative model, assuming $b = 2$, $c = 1$, and $d = 3$. Upper line, $a = 0.01$; middle line, $a = 0.10$; bottom line, $a = 0.50$.

such that punished defectors receive lower fitness than cooperators. The resident now enjoys the benefits of cooperation and only infrequently encounters erroneous defection requiring punishment. If the rare variant forgives to such a degree that her social partners optimize by defection; that is, when $c - [1 - (1 - y)/(N - 1)]d > 0$, the fitness differential (eq. [15]) reduces to $-(b + ya)$ since she loses the benefits of cooperation and punishes a proportion y of her social partners. This is negative, and so the rare variant cannot invade. If the variant's forgiveness is not sufficient to warrant a switch to defection among her social partners, equation (15) becomes $-(b + ya) \exp\{c - [1 - (1 - y)/(N - 1)]d\}$, which is vanishingly small but nevertheless negative, and hence the rare variant cannot invade. This is true because with vanishing behavioral error ($\varepsilon \rightarrow 0$) the frequency of defection in the fully punishing group is a vanishing fraction of the frequency of defection in the more forgiving group, so that the fitness saved from not punishing so frequently does not outweigh the fitness lost through the reduction of received cooperation. Relaxation of the infinitesimal error assumption (fig. 3) shows that this result is robust, even for large social groups. The variant can therefore only

invade an otherwise fully punishing population when punishment is ineffective, so that punishment is an ESS when

$$d > c. \quad (16)$$

Result 5. In the above model, punishment is maintained by selection once it has become common if the cost of cooperation (c) is less than the cost of being punished (d).

Discussion

Punishment and Cooperation

We have shown that full punishment can be an evolutionarily stable strategy only if there is a positive association between the punishment played and the cooperation received by an individual. This could arise if populations are viscous so that social partners tend to be genealogical relatives, but other mechanisms are possible, for example, when individuals facultatively adjust their level of cooperation in response to the local threat of punishment. We have also provided analytical support for the suggestion of Boyd et al. (2003) that the cost of punishment declines

as it becomes common in the population and hence punishing behavior might be maintained more easily than it is initially evolved.

These results suggest three general implications. First, it can be easier for some cooperation to evolve by another mechanism (e.g., altruism between relatives) and then punishment evolve to favor and maintain higher levels of cooperation. An analogous conclusion has been made for some other mechanisms that do not rely on interactions between relatives, such as group augmentation (Kokko et al. 2001; Griffin and West 2002). Second, within the specific context of explaining human cooperation, punishment could have evolved at a time when social structure was more conducive to punishment (small groups of interacting individuals). Once common, punishment could be retained even when interactions began to occur within much larger groups of humans. Third, the opposite frequency dependence is true for systems based on rewarding cooperation rather than punishing defection—the cost of rewarding escalates as more individuals cooperate, whereas we have shown the cost of punishing decreases as more individuals cooperate. This might go some way to explaining why punishment as opposed to rewarding is prevalent in nature (e.g., Clutton-Brock and Parker 1995).

How can our model be tested? Our major result is that costly punishment can be favored if there is a positive association between the punishment played and the cooperation received by an individual (results 1 and 2). This could be hard to test directly, especially experimentally, because of limitations on how an individual's level of punishment could be manipulated. However, some of the fundamental assumptions and predictions of our model that underly this result could be tested more easily. In particular, are lower levels of cooperation more likely to lead to punishment, as appears to occur in superb fairy wrens (Mulder and Langmore 1993), naked mole rats (Reeve 1992), and *Polistes* wasps (Reeve and Gamboa 1987)? Second, are individuals more likely to cooperate when they are punished, as may occur in *Polistes* wasps (Reeve and Gamboa 1987)? Third, do individuals try to signal that they cooperate more than they actually do, as occurs in white-winged choughs (Boland et al. 1997)? Fourth, do systems in which social partners are more related tend to display less punishment, analogous with Frank's (1995, 2003) result that investment into policing correlates negatively with relatedness?

Relatedness and Kin Selection

This analysis has made use of the understanding that the coefficient of relatedness appropriate to the direct fitness formulation of Hamilton's rule is a regression measure describing the association between actor and social partner

phenotypes (reviewed by Seger 1981; Michod 1982; Grafen 1985; Queller 1985; 1992; Frank 1998). Such associations are generally due to genealogical closeness and hence genetic similarity, so that the maximization of neighborhood-modulated or inclusive fitness is popularly referred to as "kin selection" (Maynard Smith 1964). Group selection can be responsible for the evolution of an altruistic trait only insofar as the benefit to the group is large enough, the cost to the individual is low enough, and there is substantial between-group as opposed to within-group variation in trait values. Since the proportion of the total variance that is attributable to between-group differences is the coefficient of relatedness appropriate for whole-group traits, Hamilton's rule can be used to predict when group selection will favor the trait (i.e., when relatedness \times benefit $>$ cost). Thus, kin selection and group selection are mathematically equivalent ways of conceptualizing the same evolutionary process, a point that previously has been analyzed in much detail (Price 1972; Hamilton 1975; Wade 1985; Frank 1986, 1998; Queller 1992; Reeve and Keller 1999). Consequently, it is puzzling that kin selection has been largely ignored in the human altruistic punishment literature on the grounds that relatedness is too low, while group selection has often been regarded as important (e.g., Gintis 2000). Furthermore, because relatedness is a regression of recipient phenotype on actor phenotype, it transcends genetics and applies even when the cause of phenotypic similarity is simply imitation, for example, as in the cultural group selection proposed by Heinrich and Boyd (2001). In this sense, "kin selection" is something of a misnomer because it draws attention to only one cause of the statistical association that is relatedness, as Hamilton (1975) realized.

As this analysis has shown, positive relatedness is not really the key ingredient for the evolutionary success of punishment. Punishing behavior is costly to the individual and protects the social group from the breakdown of cooperation, and hence it has been described as a form of altruism (Sober and Wilson 1998). It might then be expected that where it is successful, altruistic punishment is being maintained by kin selection. However, punishment is quite a different form of public good from cooperation—it is directly disadvantageous at the group level because it reduces the fitness of the focal individual and her social partners. The benefit it brings is indirect because it merely creates a coercive social environment in which cooperation is favored. It therefore differs from Frank's (1995, 1996, 2003) recent models of competition-repression in which investment into policing behavior translates directly into enhanced group fitness. In our model, punishment is only of selective value when there is a sufficiently strong correlation between punishment strategy played and cooperation received (dX/dy ; fig. 1B).

This highlights a fundamental nonequivalence of first- and higher-order public goods.

A positive correlation between punishment played and cooperation received might arise in a viscous population where genealogical kin tend to associate with each other, so that the social partners of punishers are also punishers ($dY/dy > 0$) and therefore punishers are expected to be coerced into cooperating more than forgivers ($dx/dy > 0$). This association combines with relatedness to ensure that an increase in punishing behavior is associated with an increase in the amount of cooperation received ($dX/dy > 0$). The pressure for enhanced punishment is therefore not strictly kin selection but rather something more akin to "niche construction" (Odling-Smee et al. 1996), in the sense that the behavior modifies the social environment in such a way as to alter the selective pressures acting upon other traits. It is worth noting that localized competition in viscous populations adds extra complexity to models of kin selection (see Taylor 1992*a*, 1992*b*; Wilson et al. 1992; Queller 1994; Frank 1998; Griffin and West 2002; West et al. 2002*a*; Gardner and West 2004 for extensive discussion of its impact on the evolution of social behaviors). In our analysis, we have assumed that all competition occurs at the level of the whole population, and we leave local competition as an open problem for the future.

We may easily demonstrate that relatedness is not necessary for the evolution of costly punishment by considering mechanisms that generate positive associations between the punishment played and the cooperation received despite zero relatedness, for example, the facultative model of cooperation introduced above. We discovered that in the absence of relatedness, partner recognition, reputation, and any mechanism whereby an individual may bias her interactions or tailor her behavior in response to her immediate social partner, punishment might be maintained by selection acting directly at the level of the individual. This is because when punishment is already frequent, the fitness saved by forgiving is minimal and may be overwhelmed by the concomitant decline in the amount of cooperation received because of the decrease in selection for cooperation among social partners. This example model is intended for illustration only and is designed to demonstrate how a net benefit for punishment might be achieved even when individuals do not interact with relatives. More complicated scenarios are therefore possible, and of particular interest is the effect of enhanced behavioral error (increasing ϵ). Numerical analysis of the example model reveals that increasing the frequency of maladaptive behavior reduces the likelihood that individual level selection will be able to maintain altruistic punishment in very large groups (fig. 3), although the results presented above are qualitatively robust so long as behav-

ioral error (ϵ) and the cost of punishing (a) are small. The degree to which individuals are expected to behave optimally is contentious, but punishment is indeed characterized by its cheapness (Sober and Wilson 1998).

Conclusion

We have given analytical support to the suggestion that the cost of punishment declines as it becomes a common strategy, so that punishment is more easily maintained than it is originally evolved. We showed that it is not relatedness per se that is important in ensuring that punishing behavior enhances fitness but rather that a positive correlation between punishment played and cooperation received by an individual is crucial. We also revealed that facultative adjustment of cooperation can give rise to such a positive association even in the absence of relatedness between social partners. Finally, we demonstrated that the direct benefits accrued when cooperation is facultative may be large enough for selection acting at the individual level alone to maintain punishment among humans, rendering elaborate population dynamics and cultural practices unnecessary. More generally, our results provide a specific example of how positive correlations between the behaviors played by social partners can arise and favor cooperation for reasons other than kinship. Major tasks for the future include clarifying the links between punishment and reproductive skew theory (Johnstone 2000; Clutton-Brock et al. 2001; Langer et al. 2004) and developing more specific models for specific situations or organisms.

Acknowledgments

We thank N. Barton, T. Clutton-Brock, F. Dionisio, S. Frank, A. Kalinka, H. Kokko, S. Nee, J. Pepper, D. Wilson, and W. Zuidema for their comments on the manuscript. Funding was provided by the Biotechnology and Biological Sciences Research Council, the Natural Environment Research Council, and the Royal Society.

Literature Cited

- Alexander, R. D. 1979. *Darwinism and human affairs*. University of Washington Press, Seattle.
- . 1987. *The biology of moral systems*. Aldine de Gruyter, New York.
- Boland, C. R. J., R. Heinsohn, and A. Cockburn. 1997. Deception by helpers in cooperatively breeding white-winged choughs and its experimental manipulation. *Behavioral Ecology and Sociobiology* 41:251–256.
- Boyd, R., and P. J. Richerson. 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology* 13:171–195.
- Boyd, R., H. Gintis, S. Bowles, and P. J. Richerson. 2003.

- The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the USA* 100:3531–3535.
- Buss, L. W. 1987. *The evolution of individuality*. Princeton University Press, Princeton, N.J.
- Clutton-Brock, T. H. 1998. Reproductive skew, concessions and limited control. *Trends in Ecology & Evolution* 13: 288–292.
- Clutton-Brock, T. H., and G. A. Parker. 1995. Punishment in animal societies. *Nature* 373:209–216.
- Clutton-Brock, T. H., P. N. M. Brotherton, A. F. Russell, M. J. O’Riain, D. Gaynor, R. Kansky, A. Griffin, et al. 2001. Cooperation, control, and concession in meerkat groups. *Science* 291:478–481.
- Denison, R. F. 2000. Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *American Naturalist* 156:567–576.
- Fehr, E., and S. Gächter. 2000. Cooperation and punishment in public goods experiments. *American Economic Review* 90:980–994.
- . 2002. Altruistic punishment in humans. *Nature* 415:137–140.
- Frank, S. A. 1986. Hierarchical selection theory and sex ratios I. General solutions for structured populations. *Theoretical Population Biology* 29:312–342.
- . 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* 377:520–522.
- . 1996. Policing and group cohesion when resources vary. *Animal Behaviour* 52:1163–1169.
- . 1998. *Foundations of social evolution*. Princeton University Press, Princeton, N.J.
- . 2003. Repression of competition and the evolution of cooperation. *Evolution* 57:693–705.
- Gardner, A., and S. A. West. 2004. Spite and the scale of competition. *Journal of Evolutionary Biology* (in press).
- Gintis, H. 2000. Strong reciprocity and human sociality. *Journal of Theoretical Biology* 206:169–179.
- Grafen, A. 1985. A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology* 2:28–89.
- Griffin, A. S., and S. A. West. 2002. Kin selection: fact and fiction. *Trends in Ecology & Evolution* 17:15–21.
- Hamilton, W. D. 1963. The evolution of altruistic behavior. *American Naturalist* 97:354–356.
- . 1964. The genetical evolution of social behavior. I, II. *Journal of Theoretical Biology* 7:1–52.
- . 1970. Selfish and spiteful behavior in an evolutionary model. *Nature* 228:1218–1220.
- . 1975. Innate social aptitudes of man: an approach from evolutionary genetics. Pages 133–153 in R. Fox, ed. *Biosocial anthropology*. Malaby, London.
- Heinrich, J., and R. Boyd. 2001. Why people punish defectors. *Journal of Theoretical Biology* 208:79–89.
- Hirshleifer, D., and E. Rasmusen. 1989. Cooperation in a repeated prisoner’s dilemma with ostracism. *Journal of Economic Behavior and Organization* 12:87–106.
- Johnstone, R. A. 2000. Models of reproductive skew: a review and synthesis. *Ethology* 106:5–26.
- Kiers, E. T., R. A. Rouseau, S. A. West, and R. F. Denison. 2003. Host sanctions and the legume-rhizobium mutualism. *Nature* 425:78–81.
- Kokko, H. 2003. Are reproductive skew models evolutionarily stable? *Proceedings of the Royal Society of London B* 270:265–270.
- Kokko, H., R. A. Johnstone, and T. H. Clutton-Brock. 2001. The evolution of cooperative breeding through group augmentation. *Proceedings of the Royal Society of London B* 268:187–196.
- Langer, P., K. Hogendoorn, and L. Keller. 2004. Tug-of-war over reproduction in a social bee. *Nature* 428:844–847.
- Maynard Smith, J. 1964. Group selection and kin selection. *Nature* 201:1145–1147.
- Maynard Smith, J., and G. R. Price. 1973. The logic of animal conflict. *Nature* 246:15–18.
- Maynard Smith, J., and E. Szathmáry. 1995. *The major transitions in evolution*. Oxford University Press, Oxford.
- McNamara, J. M., J. N. Webb, E. J. Collins, T. Székely, and A. I. Houston. 1997. A general technique for computing evolutionary stable strategies based on errors in decision-making. *Journal of Theoretical Biology* 189: 211–225.
- Michod, R. E. 1982. The theory of kin selection. *Annual Review of Ecology and Systematics* 13:23–55.
- Mulder, R. A., and N. E. Langmore. 1993. Dominant males punish helpers for temporary defection in superb fairy wrens. *Animal Behaviour* 45:830–833.
- Nee, S. 1989. Does Hamilton’s rule describe the evolution of reciprocal altruism? *Journal of Theoretical Biology* 141:81–91.
- Odling-Smee, F. J., K. N. Laland, and M. W. Feldman. 1996. Niche construction. *American Naturalist* 147:641–648.
- Oliver, P. 1980. Rewards and punishments as selective incentives for collective action: theoretical investigations. *American Journal of Sociology* 85:1356–1375.
- Ostrom, E. 1990. *Governing the commons*. Cambridge University Press, New York.
- Price, G. R. 1972. Extension of covariance selection mathematics. *Annals of Human Genetics* 35:485–490.
- Queller, D. C. 1985. Kinship, reciprocity, and synergism in the evolution of social behavior. *Nature* 318:366–367.
- . 1992. Quantitative genetics, inclusive fitness, and group selection. *American Naturalist* 139:540–558.

- . 1994. Relatedness in viscous populations. *Evolutionary Ecology* 8:70–73.
- Ratnieks, F. L. W. 1988. Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. *American Naturalist* 132:217–236.
- Reeve, H. K. 1992. Queen activation of lazy workers in colonies of the eusocial naked mole-rat. *Nature* 358:147–149.
- Reeve, H. K., and J. Gamboa. 1987. Queen regulation of worker foraging in paper wasps: a social feedback-control system (*Polistes fuscatus*, *Hymenoptera*, *Vespiidae*). *Behaviour* 102:147–167.
- Reeve, H. K., and L. Keller. 1999. Levels of selection: burying the units-of-selection debate and unearthing the crucial new issues. Pages 3–14 in L. Keller, ed. *Levels of selection in evolution*. Princeton University Press, Princeton, N.J.
- Seger, J. 1981. Kinship and covariance. *Journal of Theoretical Biology* 91:191–213.
- Sell, J., and R. K. Wilson. 1999. The maintenance of cooperation: expectations of future interaction and the trigger of group punishment. *Social Forces* 77:1551–1570.
- Sigmund, K., C. Hauert, and M. A. Nowak. 2001. Reward and punishment. *Proceedings of the National Academy of Sciences of the USA* 98:10757–10762.
- Sober, E., and D. S. Wilson. 1998. *Unto others: the evolution and psychology of unselfish behavior*. Harvard University Press, Cambridge, Mass.
- Taylor, P. D. 1992a. Altruism in viscous populations: an inclusive fitness approach. *Evolutionary Ecology* 6:352–356.
- . 1992b. Inclusive fitness in a heterogeneous environment. *Proceedings of the Royal Society of London B* 249:299–302.
- Taylor, P. D., and S. A. Frank. 1996. How to make a kin selection model. *Journal of Theoretical Biology* 180:27–37.
- Trivers, R. L. 1971. The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57.
- Wade, M. J. 1985. Soft selection, hard selection, kin selection, and group selection. *American Naturalist* 125:61–73.
- West, S. A., I. Pen, and A. S. Griffin. 2002a. Cooperation and competition between relatives. *Science* 296:72–75.
- West, S. A., E. T. Kiers, I. Pen, and R. F. Denison. 2002b. Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *Journal of Evolutionary Biology* 15:830–837.
- West, S. A., E. T. Kiers, E. L. Simms, and R. F. Denison. 2002c. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proceedings of the Royal Society of London B* 269:685–694.
- Wilson, D. S., G. B. Pollock, and L. A. Dugatkin. 1992. Can altruism evolve in purely viscous populations? *Evolutionary Ecology* 6:331–341.
- Woodcock, S., and J. Heath. 2002. The robustness of altruism as an evolutionary strategy. *Biology and Philosophy* 17:567–590.

Associate Editor: Bernard J. Crespi