PHILOSOPHICAL
TRANSACTIONS
OF
THE ROYAL
SOCIETY B

Phil. Trans. R. Soc. B (2009) **364**, 3135–3141
doi:10.1098/rstb.2009.0056

Review

Formalizing Darwinism and inclusive fitness theory

Alan Grafen*

St John's College, Oxford OX1 3JP, UK

Inclusive fitness maximization is a basic building block for biological contributions to any theory of the evolution of society. There is a view in mathematical population genetics that nothing is caused to be maximized in the process of natural selection, but this is explained as arising from a misunderstanding about the meaning of fitness maximization. Current theoretical work on inclusive fitness is discussed, with emphasis on the author's 'formal Darwinism project'. Generally, favourable conclusions are drawn about the validity of assuming fitness maximization, but the need for continuing work is emphasized, along with the possibility that substantive exceptions may be uncovered. The formal Darwinism project aims more ambitiously to represent in a formal mathematical framework the central point of Darwin's *Origin of Species*, that the mechanical processes of inheritance and reproduction can give rise to the appearance of design, and it is a fitting ambition in Darwin's bicentenary year to capture his most profound discovery in the lingua franca of science.

Keywords: formal Darwinism; fitness optimization; Price equation; uncertainty; dynamic insufficiency

1. INTRODUCTION

Darwin's *Origins of species* was published 150 years and 10 months ago and brought fully within the ambit of science subjects such as anatomy and physiology. A conference on optimal wing design, or comparative renal morphology, would certainly have fitted that 150-year anniversary. However, the Evolution of Society relies on selection of mental attributes, emotions and cooperation. It was not until 1871, in the *Descent of man*, that Darwin claimed the moral universe for biology, and in some ways 2021 would be a more appropriate date for this meeting.

However, this is also the 200th anniversary of Darwin's birth, and so all of his work can justly be celebrated on that score. My paper is appropriately traced to the *Origin*, as I have been asked to speak on inclusive fitness and on formalizing Darwinism. Inclusive fitness is a basic element of the modern understanding of natural selection and goes back directly to the ideas in the *Origin*, without the need of the further developments of the *Descent*.

Inclusive fitness is now a building block of our current understanding of natural selection, and in a meeting on the Evolution of Society, it is likely to be taken for granted, simply assumed, and then used to erect more advanced and complex ideas. It is the building block that tells us, when we focus on just the behaviour of one individual, how selection will bear on her actions. The higher reaches of the subject will ask how individuals get into their situations, how the collective behaviours will interact and whether emergent properties do in fact emerge.

My first purpose is to issue a mainly reassuring message—yes, it is sensible to use inclusive fitness as a building block—but with some reservations. We are not quite sure what inclusive fitness is in any but very simple circumstances, and relatedness might be more complicated than we think. Theoretical work on inclusive fitness can help us in extending the circumstances in which inclusive fitness is known to work, in telling us how to calculate relatedness and in warning us in what kinds of cases inclusive fitness may be liable to break down.

A second purpose is to say not all qualified biologists agree that inclusive fitness maximization is a sound biological principle, contrary to my first reassuring point. The orthodox position among mathematical population geneticists is that natural selection does not lead to any maximization principle at all. There has been a history of misunderstanding over what fitness maximization means, and theoretical work on inclusive fitness theory can help to sort out that misunderstanding.

Third, I will discuss current research on the theory of inclusive fitness, emphasizing my own formal Darwinism project. This theoretical programme aims to help us understand inclusive fitness and seeks to explain what biologists mean by inclusive fitness maximization. In addition, it also has a grander goal, which may be thought relevant at an anniversary meeting of this kind, of formalizing Darwin's core argument in the *Origin* in a fully mathematical, fully rigorous framework. After all, if Newton, Maxwell and Einstein

*alan.grafen@sjc.ox.ac.uk

One contribution of 16 to a Discussion Meeting Issue 'The evolution of society'.

have their ideas encapsulated in mathematical terms, why should not Darwin too? The obvious answer to this question is that the physicists first enunciated their theories mathematically. A second answer is that Darwin's ideas are so rich and varied, a dry mathematical account cannot capture everything. Let me accept this point and say that the mathematical framework is designed to represent one central argument of the *Origin*, namely, that the mechanical processes of inheritance and reproduction can give rise to the appearance of design.

2. WHAT IS INCLUSIVE FITNESS?

To read elementary accounts of inclusive fitness in undergraduate textbooks, one would not think there was any problem in the definition of inclusive fitness or in the security of its place in biological theory. Nor would one doubt that natural selection leads to the maximization of inclusive fitness by organisms, or at least, the 'as-if maximization'. I want at the beginning of this part of the discussion to point out how much of the advance in behavioural ecology since 1960 depends on applying the idea of fitness maximization and, where social behaviour is involved, inclusive fitness maximization. The concept of adaptation as analysed by Williams (1966), clutch-size optimization as studied by Lack (1968) and all subsequent organismal optimization theory, all the innovative theories of Trivers and co-workers (including Trivers 1971, 1972, 1974; Trivers & Willard 1973; Trivers & Hare 1976) and the use of optimization ideas such as evolutionarily stable strategies (Maynard Smith & Price 1973) and inclusive fitness itself (Hamilton 1964, 1970) only make sense if there is a maximand of natural selection.

Having established that a lot is at stake, I need to persuade you that there is reason to think inclusive fitness as a concept is not fully and logically established. The obvious place to start is with mathematical population genetics, and the clear message of Ewens (2004) that there is no quantity maximized by the operation of natural selection. In fact, one of the recurrent sports of mathematical population geneticists since 1960 has been showing that natural selection does not lead to maximization of anything (pioneered by Moran (1964) and reviewed by Ewens (2004)).

Fortunately, for behavioural ecologists and most students of the behaviour of whole organisms, this very negative conclusion is based on a misunderstanding of what fitness optimization means. It is natural for those with mathematical training, when starting with a dynamical system such as gene frequency change equations, and faced with claims of maximization, to think of established mathematical tools widely used by physicists such as Lyapunov functions and gradient functions. However, these are very far from what biologists mean by fitness maximization, and it is worth spending a moment to see why. Consider a Lyapunov function, which in this context is a function that attaches a real number to each point in genotype frequency space. Its crucial property is that as the system evolves through time, the associated real number never increases. A gradient function also attaches a real number to each point of genotype frequency space and has the stronger property that the dynamic path through genotype frequency space always takes the direction of fastest increase of the associated real number. Now the reason that these do not reflect a biologist's concept of fitness maximization is not hard to see. Both these functions are about a choice of direction in genotype frequency space and a direction that is taken by the whole population. The biological concept of fitness maximization is quite different. It is about a choice of some phenotypic trait, perhaps size or sex ratio, clarity of cornea or strength of bone, and where the trait is a property of an individual. Further, the choice is subject to constraints from physiology, physics and information. Thus, these are quite different kinds of optimization ideas. It is not surprising that this confusion should have arisen, and it could be argued that Fisher (1930) did not help with his choice of verbal expression of his fundamental theorem of natural selection. But there is no longer any excuse for perpetuating this misunderstanding.

What, then, is the correct understanding of the biologist's concept of fitness maximization? I argue (Grafen 2002) that we need to set out an optimization programme, which is a mathematical tool familiar in operations research, game theory and economics. This specifies an instrument—the variable whose value is to be chosen; a constraint set—the set of values from which the instrument is to be chosen and a maximand—a function of the instrument that says how successful that value of the instrument is. The choice of instrument and the constraint set are determined by the biological system being studied, but where is the maximand to come from? In other words, how are we to define fitness?

The known processes of natural selection are gene frequency change. We therefore need to begin with the dynamic equations of gene frequency change and try to prove links to the optimization programme. If we can prove strong enough links, including defining the maximand, then that will show how natural selection relates to fitness maximization. That, in essence, is the logic of my formal Darwinism project, which currently consists of five core papers (Grafen 1999, 2000, 2002, 2006*a*,*b*), a bunch of applications (Grafen 2007*a*,*c*; Grafen & Archetti 2008; Gardner & Grafen 2009) and two introductory and expository papers, one non-mathematical (Grafen 2007*b*) and one mathematical (Grafen 2008).

3. CURRENT WORK ON INCLUSIVE FITNESS

I now want to say a few words about current theoretical work on inclusive fitness. As well as my own project, there are three main branches of which I am aware. Two are on the pure side of population genetics. Peter Taylor and co-workers (recent papers include Taylor 1996; Day & Taylor 1998, 2000; Irwin & Taylor 2000, 2001; Taylor & Irwin 2000; Taylor *et al.* 2000, 2007*a*,*b*; Wild & Taylor 2004) have consistently extended the range of mathematical models in which inclusive fitness is defined and predicts gene frequency change. A second school of François Rousset, Laurent Lehmann and their co-workers (recent works include Rousset & Billiard 2000; Rousset 2004; Lehmann *et al.* 2006, 2007*a*,*b*; Lehmann & Balloux 2007) has developed very powerful methodologies for analysing gene frequency change using inclusive fitness and applied them to tough theoretical problems. See Gardner *et al.* (2007) for a similar approach. For theoreticians, the details of this work are very important, but for more practical biologists, there are two messages. The positive message is the range of circumstances in which inclusive fitness is known to apply is always being extended. The negative message is that range is still quite small, and there is a long way to go to cover the situations that most empirical biologists would consider usual. A further useful aspect of this theoretical work is that both groups redefine exactly what inclusive fitness is as they extend it, providing a deeper and more refined conceptual understanding.

The third branch is associated with the name of Steve Frank (including Taylor & Frank 1996; Frank 1997*a*,*b*, 1998). Frank and co-workers provide very powerful tools for biologists who wish to apply the idea of inclusive fitness. If you have a sex ratio problem, or a dispersal problem, and you want to know how to understand it in inclusive fitness terms, this is the body of work to consult.

My own formal Darwinism project, as discussed earlier, is based on linking gene frequency change to optimization programmes. A basic model presents these links for non-social behaviour, in discrete non-overlapping generations, but with arbitrary uncertainty and arbitrary ploidies (Grafen 2002). Further papers deal with extensions such as the existence of classes such as sexes or sizes (Grafen 2006*b*). and social behaviour (Grafen 2006*a*). The tasks for the future include allowing continuous time and overlapping generations and uniting all the extensions into a single over-arching model.

These bodies of work vary along a number of dimensions. A key difference is that the two more theoretical branches (Taylor and Rousset) retain the 'gold standard' property of population genetics models known as dynamic sufficiency. This restricts their range to models with very precise assumptions. The two more applied branches (Frank and Grafen) have abandoned the gold standard, for what might be called a 'plastic standard', to indicate that it aims for applied usefulness rather than decoration. They operate with fewer assumptions about gene frequencies, with the consequence that their conclusions apply more widely when they can find them, but there are many kinds of conclusions they cannot attain, because of missing information. The tradeoff is that the unattainable conclusions are mainly about highly technical dynamic things such as interior equilibria and linkage disequilibrium, which are not the focus of empirical work at the organismal level, and not even demonstrably useful to it. The attainable conclusions are about quantities more likely to be significant at a meeting like this, such as the maximization of inclusive fitness and optimized trait values. The theory is certainly ripe for an overview, in which one key question would be: what lessons should be drawn from current theory about the widespread assumption among behavioural ecologists and others, that organisms act so as to maximize their inclusive fitness?

4. INCLUSIVE FITNESS CONTRASTED WITH KIN SELECTION

The terms inclusive fitness (introduced by Hamilton 1964) and kin selection (introduced by Maynard Smith 1964) now have long histories and are often discussed as though their meanings are clear and fixed. However, authors differ as to those meanings, and at the level of work discussed in this article, their meanings have to be considered as a subject of discussion. Let us begin with what can be regarded as uncontroversial about the terms. Inclusive fitness is certainly the name of a mathematical quantity that attaches to individuals (or possibly to genotypes), whose function is to capture how natural selection acts on social behaviour by taking the place of Darwinian fitness in the simpler case of non-social behaviour. Kin selection is the name of a process, parallel to Darwinian selection or natural selection, that causes individuals to behave differently (and generally more favourably) towards more closely than to less closely related conspecifics.

These points are easy to agree upon, but many aspects are left in the air. I would add to inclusive fitness the requirement that it is a quantity that natural selection tends to cause individuals to act as if maximizing, just as Darwinian fitness tends to be maximized in the non-social case. This is a controversial point for reasons elaborated earlier, namely, that the sense of maximization has not always been understood. Furthermore, the tendency towards maximization, its strength and power and its exact nature will depend on further assumptions whose delineation is an important part of theoretical work. A crucial point for contrast with kin selection is that inclusive fitness maximization can be shown for cases in which the interactants have no special kin links, and this is discussed further below. The definition of inclusive fitness is a precise issue, and one can expect theoretical work to apply with mathematical exactness. Kin selection, on the other hand, is a loose term. Its strongest useful attachment may now be towards facts—there is overwhelming evidence, too much to cite here, that individuals of many species do behave more favourably towards relatives than towards non-relatives. These differences between the terms reflect their origins.

The contrast that will be suggested here is that a mathematical quantity called inclusive fitness can be defined, such that gene frequency dynamics tend to cause individuals to act as if maximizing inclusive fitness. Sometimes this results in a tendency to act favourably towards individuals in a way that is fully explained by their links of common ancestry—in this case, we would say that inclusive fitness underlies kin selection, and the two are in harmony. Sometimes, however, the as-if maximization of inclusive fitness will lead individuals to act more favourably to other individuals based on some other feature. The simplest case would be green beard genes, when individuals with green beards act favourably towards other individuals bearing the same trait. Here inclusive fitness is still, in a relevant sense, being maximized, but we would not want to call this process kin selection—and so in this case inclusive fitness and kin selection are not in harmony.

These prefatory remarks will be expanded in the rest of this section, which can be read as an elaboration of arguments made by and positions adopted by Hamilton (1975), who characteristically anticipated many sophisticated points not understood by others for decades, updated in the light of more recent theoretical work. I pursue the theme through the development of the formal Darwinism project to incorporate social interactions (Grafen 2006*a*). The aim of this section is to give a flavour of the complications that might arise in justifying inclusive fitness maximization, and to point to where more research would be useful.

The extension of the project to incorporate social behaviour follows Hamilton's two main derivations (Hamilton 1964, 1970). It shows that, under the assumptions of finite population, finite uncertainty and additivity of fitness interactions, selection of social behaviour can be interpreted as maximizing the arithmetic average of relative inclusive fitness. Here, relative means relative to the population mean number of offspring. This very general conclusion supports the idea that inclusive fitness is the natural way to understand selection on social behaviour and is a sufficient tool.

The theory shows how to calculate relatednesses, but it does not straight away give the kind of relatednesses with which most biologists are familiar, that is, the kind in which sibs are related by a half, parents and offspring are related by a half and cousins are related by an eighth. The relatednesses provided by the theory I will call 'inclusive fitness relatednesses' because it is those relatednesses that guarantee the link between inclusive fitness and selection. The familiar relatednesses I will call 'ancestral relatedness' because they are calculated from patterns of common ancestry.

What kinds of objects are the inclusive fitness relatednesses? How *are* they calculated if not from common ancestry? We need to know the genotypes of all individuals in the population, and relatedness itself is a weighted regression coefficient of the recipient's genotypic value on the actor's genotypic value. The weights depend on the fitness increments that represent the effects on offspring production of the social action whose selection is being studied. A relatedness is calculated for a given allele, for a given information state of the actor, and for a given social action. A selection of early work on these ideas is Crozier (1970), Orlove (1975), Orlove & Wood (1978), Michod & Hamilton (1981) and Seger (1981).

But ancestral relatednesses do not depend on a particular allele, on the information state of the actor or on the given social action. To justify the familiar relatednesses from the abstract theory therefore requires further assumptions, and it is an open area of theory just what assumptions are required. Some cases are simple. With a panmictic population and random mating, and individuals selected as interactants solely with reference to their kin links, the inclusive fitness relatednesses and the ancestral relatednesses are equal. But it would be useful to know more. What kinds of properties of dispersal group formation, or of choice of interactants, guarantee that the two kinds of relatedness are equal? Those are assumptions that are needed to justify the current emphasis on ancestral relatednesses.

Let us now turn to ask what follows when the ancestral and inclusive fitness relatednesses are not equal. In most cases, we would find that the inclusive fitness relatednesses were different for different loci and even for different alleles at the same locus. Selection at different loci would therefore be pulling in different directions. The simplest example of this phenomenon is green beard genes, but we press on to more general considerations. The power of selection to create adaptations and design requires that selection operates at many loci. If loci vary a lot in relatedness, then social behaviour is likely not to be very well designed. Selection wastes itself by opposing itself at different loci. A building with different sets of builders working to different blueprints, with one group taking down what another is in the process of erecting, is unlikely to develop a complex and functioning design.

In fact, there is one major feature of genomes that does produce these 'different groups of workers'. The analytical arguments so far have all assumed that the loci in question share their pattern of inheritance, but different patterns of inheritance do produce conflicts in the phenomenon of intra-genomic conflict (Haig & Westoby 1989; Burt & Trivers 2005). In these cases, it is common for one group of workers (in vertebrates the autosomal genes) to tie up the other groups (genes on sex chromosomes and mitochondrial chromosomes) as part of their work to complete the building. So long as one group of loci is numerically much larger than the others, this is the probable outcome, and in that case, complex design again becomes possible.

The significance of relatednesses being equal across a large solid majority of the genome is therefore very great. If it holds, then we can expect selection to be pushing in the same direction across the whole genome and to find an organism whose parts all function together, to maximize inclusive fitness. Of course, the methods of estimating relatedness from sequenced genes initiated by Queller & Goodknight (1989) could be adapted to allow an empirical investigation of whether relatedness is indeed the same across loci.

I conclude by stating briefly the contrast I am proposing between inclusive fitness and kin selection. Theory increasingly shows that inclusive fitness applies very broadly under wide assumptions, but the relatednesses required could, in principle, be affected by many factors. It makes sense to say that kin selection is operating simply, when the only or dominant force determining the relatednesses is common ancestry. In that case, relatednesses will be equal across alleles and loci, and selection will act in concert across the genome and across the organs of the body. Inclusive fitness will be a property of the individual. Where other factors influence relatedness to a significant extent, inclusive fitness is still being maximized, but with different selective effects at different alleles and different loci, and so an individual will not have a single inclusive fitness. An important task of theory is to establish under what conditions the simpler, coherent, situation obtains, in which inclusive fitness theory implies simple kin selection.

5. NEW WAYS TO EVOLVE ALTRUISM

Hamilton (1964) was excited to discover a theory of altruism and has generated a large literature as a consequence. Many later authors have claimed to discover a new and separate theory of altruism, and all of them have been wrong. They have mistaken the strength of Hamilton's achievement: he did not produce just one way of evolving altruism, rather he produced an analysis of how selection acts on social behaviour. All social behaviour must therefore conform to his conclusions.

Let us review two recent examples of 'new ways to evolve altruism'. Killingback *et al.* (2006) produced a model of a grouped population with variably sized groups and claimed that some kinds of social games showed altruism at work because of that variability. But the example fell within the assumptions of Grafen (2006*a*), and so the results must conform to inclusive fitness theory. Grafen (2007*a*) produced the relevant analysis and showed that common ancestry produced relatednesses that fully explained the results of the model.

The second example is interesting new work using graph theory, where Ohtsuki *et al.* (2006) claim that their conclusions bear some resemblance to, but are distinct from, inclusive fitness. A number of papers (Grafen 2007*c*; Lehmann *et al.* 2007*a*; Grafen & Archetti 2008) show that the graph theory work can be illuminatingly understood as fully in line with inclusive fitness theory.

The powerful analysis of Hamilton (1964, 1970), supported by the mathematically more explicit derivation of Grafen (2006*a*), allows an analysis of the natural selection of social behaviour within reasonable assumptions. This is a single theory of social behaviour for biology and is widely known and understood. Authors who produce new biological models of social behaviour would greatly assist readers by setting each new model in that canonical context.

6. FURTHER QUESTIONS

The titles of the papers in this meeting suggest further questions to pursue in a theory of formal Darwinism. The theory of inclusive fitness in Grafen (2006*a*) applies regardless of whether the population has sexual or asexual reproduction and regardless of ploidy. It does assume that the population is of uniform ploidy, but even this assumption is made only to allow a simpler notation. Thus, it should apply to asexual populations including bacteria, and the formulae for relatedness are applicable in that case.

However, this abstract connection leaves many questions unanswered. In diploids, relatedness is usually calculated using kinship links, whereas in asexual haploids, these calculations give either zero (for unrelated) or one (for clone-members). One mollifying factor, which allows graded outcomes, incorporates the chance that a specified individual is a clone-mate or not. But, an unfamiliar factor is that mutation may need to be tracked to calculate relatednesses. The concept of clone-mate is probably not well defined. Practical and applicable ways of calculating relatedness in bacteria will be very useful in analysing their social behaviour and will be discovered only through understanding the nature of their social behaviour more closely.

A second further question is whether relatedness can be defined across species. Certainly, current models of which I am aware do not allow this. But then those models contain only one species and therefore have no inter-species interactions. It is tempting to hope that a green-beard-like mechanism could operate across species boundaries.

Third, the foregoing discussion of the importance of uniformity of relatedness across loci raises a question about memes. In the genetic theory, the number of offspring is the same for all loci (assuming that they all have the same pattern of inheritance), and it is in relatedness, once social behaviour is considered, that we saw the possibility of discordance. But with cultural inheritance, the number of offspring will be different for each culturgen, and we should therefore not expect to find well-designed cultural phenotypes.

7. DISCUSSION

I begin the concluding discussion with the statement that the existing theory suggests that it is reasonable to proceed with caution in assuming fitness maximization, while recognizing there may be theoretical discoveries that limit the range of circumstances in which we can expect fitness to be approximately maximized. This applies to fitness in general and to inclusive fitness when social behaviour is considered. The current methods of calculating relatednesses are probably usually fine, but this too is subject to revision. To be a little more concrete, inbreeding tends to make relatednesses different for different alleles at the same locus, a population that is not mating at random can be viewed as inbreeding, and so all real populations suffer from one of the potential causes of difficulties with fitness maximization.

Next, it is worth considering what the purposes of formalizing Darwinism are. Most immediately, from the point of view of this meeting, it would be useful to know more about when the assumption of fitness maximization is reasonable, and how fitness should be defined. In the presence of social interactions, that includes the calculation of relatednesses so that inclusive fitness can be calculated. It is also important that the theoretical results should apply to circumstances such as conditional behaviour, appropriate use of information and realistic population structures and include environmental uncertainty and possibly overlapping generations. Some of these factors have been dealt with, while some have not, as discussed earlier, but my present point is that this theoretical justification for ongoing empirical work is one reason for formalizing Darwin.

A second reason is to set down exactly what Darwin's point was, in formal terms, to avoid doubt and make it easier for other scientists to learn about. It was noted in Darwin's obituary published in the Society's Proceedings that the *Origin* is a difficult book to read and that many more people think they understand evolution by natural selection than they in fact do. Darwin could not have formalized his point mathematically. As well as not being a good mathematician, the necessary mathematics had not yet been invented, and anyway, it is necessary to know about Mendelian genetics, which had not been discovered in 1859. My approach uses measure theory, which also had not been invented. But mathematics is the lingua franca of science, and if we want physicists and mathematicians to understand what biologists are doing, and why they are doing it, it is in my view essential to express the key points mathematically. Then they have no excuse for not understanding. This is almost a defensive reason, and it applies to many biologists as well. There are two tendencies I have encountered among biologists about Darwin that lead to a less than full-throated acceptance of his work. First, there is what we might call 'radical empirical provisionalism', which says, in effect, 'Darwin may or may not have been right—we need to do more experiments to find out'. Second, there is a 'theoretical reserve', which says 'Models of natural selection don't bear out what Darwin said, so he may been approximately right, but only further theoretical work can tell how approximately'. A fully rigorous treatment of Darwin's central point would set out Darwin's achievement in a clear and unambiguous way.

Finally, my favourite reason for formalizing Darwin is not any of those things. The grand theories of physics are all equations, with a few words to interpret the meanings of the symbols into the reader's language of choice. Darwin's idea, afforced by Mendelian genetics, is an extremely important scientific discovery. We will understand it better, more precisely, more generally, if we have a formal mathematical framework in which the idea can be expressed. Not a model that is an example, but a model that captures the idea at its full level of generality. One advance of Darwinism is represented by this meeting, by the ambition to explain more of the natural world in Darwinian terms. Another advance is to understand his central ideas more fully and more generally. These two directions of work support and inform each other and allow us to benefit, with ever increasing effectiveness, from the intellectual legacy of that remarkable thinker, Charles Darwin.

I am grateful to Stu West, Andy Gardner, Joao Alpedrinha and Claire El Mouden for very helpful comments on the manuscript.

REFERENCES

Burt, A. & Trivers, R. L. 2005 *Genes in conflict: the biology of selfish genetic elements*. Cambridge, MA: Harvard University Press.

Crozier, R. H. 1970 Coefficients of relationship and the identity of genes by descent in the Hymenoptera. *Am. Nat.* **104**, 216–217. (doi:10.1086/282654)

Day, T. & Taylor, P. D. 1998 Unifying genetic and game theoretic models of kin selection for continuous traits. *J. Theor. Biol.* **194**, 391–407. (doi:10.1006/jtbi.1998.0762)

Day, T. & Taylor, P. D. 2000 A generalization of Pontryagin's maximum principle for dynamic evolutionary games among relatives. *Theor. Popul. Biol.* **57**, 339–356. (doi:10.1006/tpbi.2000.1459)

Ewens, W. J. 2004 *Mathematical population genetics I. Theoretical introduction*. Berlin, Germany: Springer.

Fisher, R. A. 1930 *The genetical theory of natural selection*. Oxford, UK: Oxford University Press. OUP published in 1999 a variorum edition of the 1930 and 1958 editions.

Frank, S. A. 1997*a* Multivariate analysis of correlated selection and kin selection, with an ESS maximization method. *J. Theor. Biol.* **189**, 307–316. (doi:10.1006/jtbi.1997.0516)

Frank, S. A. 1997*b* The Price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *Evolution* **51**, 1712–1729. (doi:10.2307/2410995)

Frank, S. A. 1998 *The foundations of social evolution*. Princeton, NJ: Princeton University Press.

Gardner, A. & Grafen, A. 2009 Capturing the superorganism: a formal theory of group adaptation. *J. Evol. Biol.* **22**, 659–671. (doi:10.1111/j.1420-9101.2008.01681.x)

Gardner, A., West, S. A. & Barton, N. H. 2007 The relation between multilocus population genetics and social evolution theory. *Am. Nat.* **169**, 207–226. (doi:10.1086/510602)

Grafen, A. 1999 Formal Darwinism, the individual-as-maximising-agent analogy, and bet-hedging. *Proc. R. Soc. Lond. B* **266**, 799–803. (doi:10.1098/rspb.1999.0708)

Grafen, A. 2000 Developments of Price's equation and natural selection under uncertainty. *Proc. R. Soc. Lond. B* **267**, 1223–1227. (doi:10.1098/rspb.2000.1131)

Grafen, A. 2002 A first formal link between the Price equation and an optimisation program. *J. Theor. Biol.* **217**, 75–91. (doi:10.1006/jtbi.2002.3015)

Grafen, A. 2006*a* Optimisation of inclusive fitness. *J. Theor. Biol.* **238**, 541–563. (doi:10.1016/j.jtbi.2005.06.009)

Grafen, A. 2006*b* A theory of Fisher's reproductive value. *J. Math. Biol.* **53**, 15–60. (doi:10.1007/s00285-006-0376-4)

Grafen, A. 2007*a* Detecting kin selection at work using inclusive fitness. *Proc. R. Soc. B* **274**, 713–719. (doi:10.1098/rspb.2006.0140)

Grafen, A. 2007*b* The formal Darwinism project: a mid-term report. *J. Evol. Biol.* **20**, 1243–1254. (doi:10.1111/j.1420-9101.2007.01321.x)

Grafen, A. 2007*c* Inclusive fitness on a cyclical network. *J. Evol. Biol.* **20**, 2278–2283. (doi:10.1111/j.1420-9101.2007.01413.x)

Grafen, A. 2008 The simplest formal argument for fitness optimization. *J. Gene.* **87**, 421–433. (doi:10.1007/S12041-008-0064-9)

Grafen, A. & Archetti, M. 2008 Natural selection of altruism in inelastic homogeneous populations. *J. Theor. Biol.* **252**, 694–710. (doi:10.1016/j.jtbi.2008.01.021)

Haig, D. & Westoby, M. 1989 Parent-specific gene expression and the triploid endosperm. *Am. Nat.* **134**, 147–155. (doi:10.1086/284971)

Hamilton, W. D. 1964 The genetical evolution of social behaviour. *J. Theor. Biol.* **7**, 1–52. (doi:10.1016/0022-5193(64)90038-4)

Hamilton, W. D. 1970 Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**, 1218–1220. (doi:10.1038/2281218a0)

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

Irwin, A. & Taylor, P. D. 2000 Evolution of dispersal in stepping-stone populations with overlapping generations. *Theor. Popul. Biol.* **58**, 321–328. (doi:10.1006/tpbi.2001.1490)

Irwin, A. & Taylor, P. D. 2001 Evolution of altruism in stepping-stone populations with overlapping generations. *Theor. Popul. Biol.* **60**, 315–325. (doi:10.1006/tpbi.2001.1533)

Killingback, T., Bieri, J. & Flatt, T. 2006 Evolution in group-structured populations can resolve the tragedy of the commons. *Proc. R. Soc. B* **273**, 1477–1481. (doi:10.1098/rspb.2006.3476)

Lack, D. 1968 *Ecological adaptations for breeding in birds*. London, UK: Methuen.

Lehmann, L. & Balloux, F. 2007 Natural selection on fecundity variance in subdivided populations: kin selection meets bet hedging. *Genetics* **176**, 1–17. (doi:10.1534/genetics.106.066910)

Lehmann, L., Perrin, N. & Rousset, F. 2006 Population demography and the evolution of helping behaviors. *Evolution* **60**, 1137–1151.

Lehmann, L., Keller, L. & Sumpter, D. 2007*a* The evolution of helping and harming on graphs: the return of the inclusive fitness efiect. *J. Evol. Biol.* **20**, 2284–2295. (doi:10.1111/j.1420-9101.2007.01414x)

Lehmann, L., Roze, D., Rousset, F. & Keller, L. 2007*b* Strong reciprocity or strong ferocity? A population genetic view of the evolution of altruistic punishment. *Am. Nat.* **170**, 21–36. (doi:10.1086/518568)

Maynard Smith, J. 1964 Group selection and kin selection. *Nature* **201**, 1145–1147. (doi:10.1038/2011145a0)

Maynard Smith, J. & Price, G. R. 1973 The logic of animal conflict. *Nature* **246**, 15–18. (doi:10.1038/2460145a0)

Michod, R. E. & Hamilton, W. D. 1981 Coefficients of relatedness in sociobiology. *Nature* **288**, 694–697. (doi:10.1038/288694a0)

Moran, P. A. P. 1964 On the nonexistence of adaptive topographies. *Ann. Hum. Genet.* **27**, 383–393.

Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M. 2006 A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502–505. (doi:10.1038/nature04605)

Orlove, M. J. 1975 A model of kin selection not involving coefficients of relationship. *J. Theor. Biol.* **49**, 289–310.

Orlove, M. J. & Wood, C. 1978 Coefficients of relationship and coefficients of relatedness in kin selection: a covariance form for the RHO formula. *J. Theor. Biol.* **73**, 679–686. (doi:10.1016/0022-5193(78)90129-7)

Queller, D. C. & Goodknight, K. 1989 Estimation of relatedness from allozyme data. *Evolution* **43**, 258–275. (doi:10.2307/2409206)

Rousset, F. 2004 *Genetic structure and selection in subdivided populations*. Princeton, NJ: Princeton University Press.

Rousset, F. & Billiard, S. 2000 A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J. Evol. Biol.* **13**, 814–825. (doi:10.1046/j.1420-9120-9101.2000.00219.x)

Seger, J. 1981 Kinship and covariance. *J. Theor. Biol.* **91**, 191–213. (doi:10.1016/0022-5193(81)90380-5)

Taylor, P. D. 1996 Inclusive fitness arguments in genetic models of behaviour. *J. Math. Biol.* **34**, 654–674. (doi:10.1007/BF02409753)

Taylor, P. D. & Frank, S. A. 1996 How to make a kin selection model. *J. Theor. Biol.* **180**, 27–37. (doi:10.1006/jtbi.1996.0075)

Taylor, P. D. & Irwin, A. 2000 Overlapping generations can promote altruistic behavior. *Evolution* **54**, 1135–1141.

Taylor, P. D., Irwin, A. & Day, T. 2000 Inclusive fitness in finite deme-structured and stepping-stone populations. *Selection* **1**, 83–93.

Taylor, P. D., Day, T. & Wild, G. 2007*a* Evolution of cooperation in a finite homogeneous graph. *Nature* **447**, 469–472. (doi:10.1038/nature05784)

Taylor, P. D., Day, T. & Wild, G. 2007*b* From inclusive fitness to fixation probability in homogeneous structured populations. *J. Theor. Biol.* **249**, 101–110. (doi:10.1016/j.jtbi.2007.07.006)

Trivers, R. L. 1971 The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57. (doi:10.1086/406755)

Trivers, R. L. 1972 Parental investment and sexual selection. In *Sexual selection and the descent of man, 1871–1971* (ed. B. Campbell), chap. 7, pp. 136–179. Chicago, IL: Aldine.

Trivers, R. L. 1974 Parent–offspring conflict. *Am. Zool.* **14**, 249–264. (doi:10.1093/icb/14.1.249)

Trivers, R. L. & Hare, H. 1976 Haplodiploidy and the evolution of the social insects. *Science* **191**, 250–263. (doi:10.1126/science.1108197)

Trivers, R. L. & Willard, D. E. 1973 Natural selection of parental ability to vary the sex of offspring. *Science* **179**, 90–92. (doi:10.1126/science.179.4068.90)

Wild, G. & Taylor, P. D. 2004 Fitness and evolutionary stability in game theoretic models of finite populations. *Proc. R. Soc. Lond. B* **271**, 2345–2349. (doi:10.1098/rspb.2004.2862)

Williams, G. C. 1966 *Adaptation and natural selection*. Princeton, NJ: Princeton University Press.