Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Theoretical Biology 238 (2006) 541–563
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Optimization of inclusive fitness

Alan Grafen*

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

Received 9 March 2005; received in revised form 7 June 2005; accepted 8 June 2005
Available online 25 July 2005

Abstract

The first fully explicit argument is given that broadly supports a widespread belief among whole-organism biologists that natural selection tends to lead to organisms acting as if maximizing their inclusive fitness. The use of optimization programs permits a clear statement of what this belief should be understood to mean, in contradistinction to the common mathematical presumption that it should be formalized as some kind of Lyapunov or even potential function. The argument reveals new details and uncovers latent assumptions. A very general genetic architecture is allowed, and there is arbitrary uncertainty. However, frequency dependence of fitnesses is not permitted. The logic of inclusive fitness immediately draws together various kinds of intra-genomic conflict, and the concept of 'p-family' is introduced. Inclusive fitness is thus incorporated into the formal Darwinism project, which aims to link the mathematics of motion (difference and differential equations) used to describe gene frequency trajectories with the mathematics of optimization used to describe purpose and design. Important questions remain to be answered in the fundamental theory of inclusive fitness.

Keywords: Natural selection; Social behaviour; Inclusive fitness; Optimality; Price equation

1. Introduction

Inclusive fitness was introduced by Hamilton (1964) and has become a foundation stone of modern biology. There is a large literature justifying and explaining inclusive fitness by Hamilton and others (e.g. Hamilton, 1963, 1964; Grafen, 1984, 1985; Queller, 1992; Taylor, 1990; Frank, 1998), which is often referred to when inclusive fitness is used. There nevertheless continues to be uncertainty about its logical standing: the original derivation has not generally been accepted as rigorous and convincing, and there have been many further versions by Hamilton and by others. The aim of this paper is to advance the justification in a number of ways, and a central technical device is to be fully explicit for the first time about the connection between population genetics and the optimization of inclusive fitness.

The reasons for the inadequacy of the original papers should be given at the outset. The uncertainty over the derivations of Hamilton (1964, 1970) is not mainly caused by flaws in the arguments contained in those papers. It is true that the 1964 paper has a difficult notation, and that the 1970 paper's specification of a relatedness between each pair of individuals is problematic. But the chief doubt arises simply because it postulates a maximization principle for inclusive fitness in a model that is more complex than the model in which Fisher (1930) postulated his Fundamental Theorem: until Fisher's work was accepted, therefore, Hamilton's result would lie in limbo. Following a long-neglected paper by Price (1972), a more general realization of what Fisher's theorem meant, and that it was true, began with Ewens (1989, 1992) and continues (Lessard, 1997). I have drawn implications for the optimization principle in Fisher's case mathematically (Grafen, 2002) and verbally (Grafen, 2003). The chief difficulty in accepting the optimization principles of Fisher and Hamilton has arisen from misunderstanding of what maximization of

*Tel.: +44 1865 277438; fax: +44 1865 277435.
E-mail address: alan.grafen@sjc.ox.ac.uk.

0022-5193/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.jtbi.2005.06.009

fitness means, hence the vital importance of being fully explicit on this point.

The position of inclusive fitness as the quantity organisms are most widely believed to be selected to maximize makes it important to provide an understandable argument in this case. The model below is therefore kept fairly simple: for example it assumes a finite population, so that summation signs are used but no integration is required. This leaves some aspects in need of a more abstract treatment, to remove the implausibility of some assumptions, but these are clearly outlined as they occur.

The major restrictions on the derivation are as follows. Frequency dependence is not permitted, and it is found necessary to make the assumption that effects of social actions on numbers of offspring add up. Clearly these are important limitations, as inclusive fitness ideas should be useful in frequency dependent and non-additive situations. There is a literature on this case (Grafen, 1979; Hines and Maynard Smith, 1979; Day and Taylor, 1998), but the question even of how inclusive fitness should be defined in general in the presence of frequency dependence has not been adequately considered. The population, though finite, will be assumed to be large when uncertainty is added to the model.

The positive advantages of the derivation are significant. The first two, very different, derivations of Hamilton (1964, 1970) are both very general in that they permit arbitrary numbers and types of social actions to be performed by each individual, and the recipients of each action can also be various. The current derivation shares this feature, while nearly all other previous derivations have limited themselves to only one kind of social action, and the recipient of one action has been either one individual with a fixed relationship to the actor, or alternatively all the other members of the actor's group. A major feature of the present paper is that every effort has been made to be fully explicit. A third advantage is that uncertainty is allowed, and conditional behaviour is explicitly articulated, which are features taken for granted by biologists but so far neglected in theoretical derivations. Finally, the conclusion that inclusive fitness is optimized, as opposed to the commoner conclusion that Hamilton's Rule is valid for a given social action, is very significant from the point of view of the design of the organism's behaviour as a whole: this point is taken up in the discussion in Section 7.

The paper pursues for inclusive fitness the same agenda as Grafen (2002) did for Darwinian fitness in expanding on the Fundamental Theorem of Natural Selection of Fisher (1930), namely to provide an explicit and mathematically rigorous link between population genetics and fitness optimization. Many groups of biologists simply accept that organisms act as if maximizing their fitness, and conduct research projects from that standpoint. This paper aims to articulate, consolidate and improve the basis for that acceptance.

2. Overview

The mathematical arguments to follow have been limited in complexity, but it is recognized that many readers will prefer to omit some of the sections. The introduction and this overview, as well as the interpretative Section 6 and the final discussion should be accessible to all. Many of the important technical points can be grasped by reading Section 3 up to 3.3. The notation becomes larger in scope, though in fact no more mathematically complex, in Section 3.4. The interpretive issues also become more subtle. Section 4 introduces optimization programs as an explicit representation of biologists' assumption of optimality, and although the section becomes quite hard, it may be of interest to some readers to see how far they can tolerate it.

The argument begins with two separate approaches, which are later linked. The connection between inclusive fitness and gene frequency change is established first, in Section 3. The concept of 'role' is introduced, which formalizes a concept neglected since the original argument of Hamilton (1964). Roles allow a fully explicit argument, based on the Price Equation, that begins with neighbour-modulated fitness as the target of selection, and ends with inclusive fitness as the target of selection. An important quantity called the 'Hamilton residual' appears in the course of the transformation and it turns out that the key property of relatednesses is that they render the Hamilton residual negligible in magnitude. Section 3.2 discusses a number of ways to choose relatednesses to bring this about, and an exceptional case where relatedness cannot be defined to annihilate the Hamilton residual is discussed in Section 3.3. The model is extended to allow arbitrary uncertainty in Section 3.4, which is important both to investigate how fitness should be averaged in the presence of environmental uncertainty, and also to permit the formal treatment of conditional behaviour.

The second approach is to construct explicit optimization programs in Section 4, beginning in Section 4.1 with a very simple generic program, which is developed to include uncertainty and conditional behaviour. Biologists take for granted that animal behaviour is conditional and that it must be average fitness that is maximized, as there is no such thing as unconditional behaviour and the life of no organism possesses certainty. It is important that to solve the final optimization program, an organism must be a sophisticated decision-taker, with a (correct) prior distribution over uncertainty, and an ability to update the prior in an optimal Bayesian way in the light of information received and to find an action that maximizes the arithmetic average payoff. The simple example in Section 4.2 illustrates how the optimization program embodies these sophistications. Finally, Section 4.3 contains a linking model that connects the fitness effects used to construct the optimization program to those in the population genetics model. This linking model requires the assumptions of additivity and actor's control. Additivity means that the effects of others on one individual's fitness combine by adding up. Actor's control means that the nature and quantitative effects of one individual's action depend only on the phenotype of that individual, and not, for example, on some capacity of the recipient to use the help provided. The payoff function in the optimization program is defined in terms of population genetic quantities, and the population of individuals in the population genetics model is related to the one single decision-taker of the optimization program through the important assumption of 'universal strategic equivalence', which essentially says all individuals face the same set of problems.

The two approaches are brought together in Section 5. Formal links are proved between the two approaches, which are quite detailed and technical. A broad interpretation is that natural selection always changes gene frequencies in the direction of increasing inclusive fitness; and that a population genetic equilibrium in which no feasible mutations can spread implies that the individuals in the population are each acting so as to maximize their inclusive fitness. One qualification is that genetic complications may mean that *genotype* frequencies do not change to increase inclusive fitness, and so inclusive fitness may not in fact increase, even though that is the direction in which *gene* frequencies have changed. It is worth noting that the implication that population mean fitness is maximized, often wrongly taken to be the meaning of fitness optimization principles, is rendered meaningless for the fitness that appears in the optimization programs in the Fisherian case by the fact that this fitness is measured relative to the population mean. Although the mean absolute fitness does happen to be maximized in the Fisherian case studied by Grafen (2002), it is not in the inclusive fitness case of the present paper. Further, the structure of the optimization programs would carry over unchanged if frequency dependence were introduced in a further development.

In proving the links in Section 5, and indeed in the whole of the argument up to that point, we discuss a set of genes that share the same transmission pattern. On the way, we pick up an additional restriction that the genes must also share the same relatednesses. Such a set of genes we will name a 'p-family'. The significance of the optimization result for the behaviour of an organism is the resultant of selection on all the organism's different p-families. The general situation is discussed in Section 6, which goes on to consider various forms of intragenomic conflict in the context of the new derivation of inclusive fitness.

The discussion in Section 7 considers this paper in the contexts of the Formal Darwinism project and other foundational work on inclusive fitness, reviews the new ideas to emerge, and ends by taking a look at the current status of inclusive fitness.

3. Inclusive fitness and gene frequency change

Mendelian-type genetics will be taken as the known process underlying evolutionary change, and accepted as fundamental. Mendelian-type means that there are haploid sets, each of which contains one copy of each locus; that the genotype of an individual consists of a number of haploid sets, but ploidy may vary between individuals; and that reproduction takes place by (optional) recombination between an individual's haploid sets that conserves all alleles but redistributes them, followed by passing on a number of the new haploid sets to an offspring, who may also receive haploid sets from one or more other parents. Thus classical Mendelian genetics is included, as is asexual reproduction, and a mixture.

Inclusive fitness is related to gene frequencies in this section, in three stages. The Price equation is developed in Section 3.1 in combination with an additive model of social interactions to provide a version of the Price equation with inclusive fitness as the 'target of selection'. This introduces 'relatednesses' as components of the 'Hamilton residual', which itself must be zero or small for the theory to apply. Section 3.2 considers how relatednesses have been and can be defined, and how these render zero or small the Hamilton residual. An exceptional case is discussed in Section 3.3, in which one of these approaches fails. In Section 3.4, the analysis and discussion of relatednesses is extended to the case of uncertainty, including enviromental uncertainty and the uncertainty of Mendelian segregation.

The assumption of additivity is made throughout this paper, but is not in general a realistic assumption. In many applications, non-additivity is an important part of the problem. The assumption is made here for two reasons. First, the general argument given here about the maximization of inclusive fitness will hold only with additivity, and this paper does not consider more detailed models in which non-additivity requires attention. Second, one basic technique for non-additive cases is to linearize fitness relationships so that an assumption of additivity is reasonable with small fitness effects. The logic of this approach relies on knowing that inclusive fitness is maximized once we have additivity, and here the work of this paper can provide support in the more realistic situation. It has already been mentioned that the question of how to define inclusive fitness in the absence of additivity has not been settled, and so fundamental theory on the non-additive case can hardly yet begin.

3.1. The Price Equation with additive social interactions

The altered basic assumption that distinguishes models of social interactions from non-social models is that one individual's number of offspring may depend on the actions of other individuals. The way this assumption is expressed formally is the key to a simple derivation. We will assume there is a set T of possible 'roles' which a recipient can occupy in relation to an actor. In simple cases based on kinship, roles might include 'sib', 'cousin' and 'unrelated'. Where the action is based on proximity, roles might include 'neighbour', or alternatively 'close neighbour' and 'distant neighbour'; and where they are based on groups, they might include 'fellow-group-member', and 'neighbouring-group-member'. In general, we will allow all these possibilities and simply work with an arbitrary set T of roles that draws on one or more of these categories. One role is special, representing the fundamental assumption that an action has an individual that is responsible for it, that is, whose genotype determines whether the action takes place and its nature. This role will be called 'actor' and represented by 'e' (for 'ego') in subscripts. We do not assume that the actor can always distinguish other individuals according to their role, though in simple cases this will be so. The logical force of the concept of role is that in modelling we have no need to discriminate more finely than between roles, and this does imply that the individual cannot discriminate more finely than them either.

Individuals will be notated as i and j, elements of the finite set I, and will be allowed to vary in ploidy. Let b_{ijt} be the extra reproduction conferred by i on j in role t. The quantity of these donated offspring will depend on the phenotypes of the individuals and, by the assumption of actor's control, b_{ijt} will depend only on the phenotype of individual i. Let γ_i represent the phenotype of individual i. When we wish to emphasize the dependence, we will write $b_{ijt}(\gamma_i)$, and when we consider what would happen if individual i played α instead, we will write $b_{ijt}(\alpha)$. However for the moment it suffices to leave the dependence on phenotype implicit, as no alternative phenotypes are being considered. We therefore represent an assumption of additivity of fitness effects by writing the number of successful gametes, per parental haploid set, of individual j as

$$w_j = 1 + \sum_i \sum_t b_{ijt}.$$

The notation of the Price equation (Price (1970)—we do *not* use the more general formulation of Price (1972)) is now briefly introduced (see Grafen, 1985, 2002, for further details). A 'p-score' is a single number for each individual that reflects the individual's genotype. The frequency of an allele is the simplest p-score, and in diploids an individual's gene frequency is either 0, $\frac{1}{2}$ or 1. In general a p-score is an average over the individual's haploid sets of a weighted sum of allele frequencies, where the weights can be arbitrarily assigned. In particular, this means that the additive genetic value (Falconer, 1981) of any given trait in any given generation is also a p-score. The p-score of individual j is denoted p_j, and the average over the population simply by p. The mean p-score in the next generation is denoted p', and the change is denoted $\Delta p = p' - p$. The notation Δp_j refers to the difference in p-score between the successful gametes of individual j and the p-score of individual j herself. The ploidy-weighted mean value of w_j is denoted w. The Price equation provides a kind of accounting identity, which acts as a very powerful formalism for studying natural selection.

We shall assume for the moment that all the alleles involved in the definition of the Price equation share a common pattern of transmission. The formalism will produce true results, though with different interpretations, no matter what that pattern is. In humans, there are at least four different patterns. Autosomal loci have one pattern, X-linked genes and Y-linked genes each have their own pattern, and mitochondrial genes have a fourth. Section 6 considers how to interpret together different Price equations describing the same population and its reproduction, but for p-scores with different transmission patterns.

The general form for the Price equation is $w\Delta p = \mathbb{C}[w_j, p_j] + \mathbb{E}[w_j \Delta p_j]$, where $\mathbb{E}[\cdot]$ and $\mathbb{C}[\cdot,\cdot]$ denote the average and covariance over the individuals in the population, in each case weighting by ploidy. The Price equation in this form applies very widely, in the face of arbitrary linkage, linkage disequilibrium, ploidy levels, non-random mating, a mixture of sexual and asexual reproduction, and population structure. In order to be able to translate out of covariances into sums we introduce the notation n_j to represent the ploidy of individual j, and $N = \sum_j n_j$ as the total ploidy of the parental population.

The Price equation can now be expanded to

$$w\Delta p = \mathbb{C}\left[p_j, \sum_i \sum_t b_{ijt}\right] + \mathbb{E}[w_j \Delta p_j]. \quad (1)$$

We now embark on a series of rearrangements whose aim is to replace summing b_{ijt} over j, and so adding up all the fitness effects that are suffered by individual j; with summing over i, so adding up all the fitness effects caused by individual i. The equation can be re-written as

$$Nw\Delta p = \sum_j n_j(p_j - p) \sum_i \sum_t b_{ijt} + N\mathbb{E}[w_j \Delta p_j].$$

The first summand can be rearranged to produce

$$\sum_j n_j(p_j - p)b_{jje} + \sum_j \sum_i \sum_{t \neq e} n_j(p_j - p)b_{ijt},$$

where we have separated effects on the actor 'ego' from effects on other roles. Reordering the summations in the second term, and changing index variable in the first we find

$$\sum_i n_i(p_i - p)b_{iie} + \sum_i \sum_{t \neq e} \sum_j n_j(p_j - p)b_{ijt}.$$

Now associate with each role $t \neq e$ a number r^t. The manipulations remain valid whatever the values of the (r^t), but the interpretation of the end result will depend crucially on how they are chosen. For the moment we add to the first term and subtract from the second term the sum $\sum_i \sum_{t \neq e} \sum_j r^t n_j(p_i - p)b_{ijt}$ to obtain, gathering terms in i in each case,

$$\sum_i n_i(p_i - p)\left(b_{iie} + \sum_{t \neq e} \sum_j r^t \frac{n_j}{n_i} b_{ijt}\right) + \sum_i \sum_{t \neq e} \sum_j n_j((p_j - p) - r^t(p_i - p))b_{ijt}.$$

The first term can be expressed using a covariance as

$$N\mathbb{C}\left[p_i, \left(b_{iie} + \sum_{t \neq e} r^t \sum_j \frac{n_j}{n_i} b_{ijt}\right)\right]$$

and Eq. (1) can now be rewritten, with a slight rearrangement of the summation signs in the third summand, as

$$w\Delta p = \mathbb{C}\left[p_i, \left(b_{iie} + \sum_{t \neq e} r^t \sum_j \frac{n_j}{n_i} b_{ijt}\right)\right] + \mathbb{E}[w_j \Delta p_j] + \frac{1}{N} \sum_{t \neq e} \left(\sum_i \sum_j n_j(p_j - p)b_{ijt} - r^t \sum_i \sum_j n_j(p_i - p)b_{ijt}\right). \quad (2)$$

This equation shows a target of selection (that is, its covariance with p-score appears in the Price equation) which is the per-ploidy value of

$$n_i b_{iie} + \sum_{t \neq e} \sum_j r^t n_j b_{ijt}.$$

If r^t can be chosen so that it is the fraction of j's reproduction that in some sense *counts* for i, then this is the total equivalent number of successful gametes of individual i, a very natural quantity to find as a maximand in an evolutionary model. Anticipating, it would not be inclusive fitness itself, as it lacks the basic unit of reproduction, but it would be the inclusive fitness *effect* in the sense of Hamilton (1964). The tightness of the link between this quantity and the change in p-score will depend on arguments that render the second and third summands on the right-hand side of Eq. (2) zero, zero on average, or small on average.

Various kinds of averaging will be relevant in dealing with the second and third summands. The second term arises in just the same way without social interactions. The term Δp_j is the difference between the p-score of individual j and the p-score of the successful gametes of individual j. If we consider the average over the Mendelian segregations that create those gametes, and assume that they are fair, then each Δp_j is on average zero independently of w_j, and so the whole term is also on average zero: any change in the mean p-score due to this term is down to randomness in meiosis, and is not systematic. Thus we will ignore the second summand for the rest of this section, and consider only the expected change in mean p-score, implicitly taking expectations over Mendelian segregations in the production of this generation's offspring. This term is dealt with more formally in Section 3.4.

The third summand is particular to models with social behaviour, and it is convenient to give this and similar terms the name 'Hamilton residual'. A key point about its structure is that there is one term for each role, apart from 'ego', and there is one relatedness for each term. The arguments that the Hamilton residual is zero or at least small will apply to each of these terms separately. The crucial logic of the concept of role is that it must permit a proof that the Hamilton residual is zero or close to it. The next subsection discusses how relatednesses have been and can be chosen.

3.2. Relatednesses

Eq. (2) holds true for all possible values of the r^t, but is most useful in linking inclusive fitness to gene frequency change if the Hamilton residual can be shown to be zero exactly, zero on average, or small. When the r^t are so chosen, they measure genetic similarity between actor and recipient, and from now on we will call them 'relatednesses'. Specifically, r^t will be the relatedness of an actor to recipients in role t. There is a considerable literature on this topic including a foundational paper by Crozier (1970) who introduced the distinction between coancestry (his 'relationship') and genetic similarity (his 'relatedness'), and many papers on how to measure relatednesses so as to make inclusive fitness work (beginning with Orlove, 1975; Orlove and Wood, 1978).

The methods of choosing the r^t, the relatednesses, can be divided into two, according to the kind of information used to calculate them. Hamilton (1964) used information about ancestral links between individuals. This, along with other methods which use a small amount of information about kinship links or the group structure of a population will be referred to as following a 'modelling approach'. This is to contrast with the 'measurement approach', which uses complete information about the genotypes of the individuals and their social interactions. We will return to a comparison of modelling and measurement once they have been given more substance through examples.

Identity-by-descent is the commonest and original (Hamilton, 1964) modelling approach. It involves in principle averaging over Mendelian segregations in previous generations, just as we have already averaged over them in the production of offspring by parents in this generation. The method by which this choice for r^t renders the Hamilton residual zero on average is discussed in some detail by Grafen (1985). Essentially, as Hamilton explained in 1964, the genotype of the recipient is viewed as a mixture: some fraction is as if drawn from the population gene pool at random, and some fraction (the relatedness) is as if identical with the actor's genotype. This implies that the deviation of the recipient's genotype from the population mean $(p_j - p)$ equals *on average* a fraction of the deviation of the actor's genotype from the population mean $(r^t(p_i - p))$, which implies as required that *on average* $(p_j - p) - r^t(p_i - p) = 0$. The uncertainty in Mendelian segregations over which these averages are taken is not incorporated into the formal structure of the model, and it would be difficult to do so, as the model conditions on the population and its genotypes. The weighting in the Hamilton residual in Eq. (2) of $(p_j - p) - r^t(p_i - p)$ with $n_j b_{ijt}$ is partly a matter of giving more importance to those individuals j in the role t that are the recipient of more benefit from i, but note that the 'weights' can be negative so that helps and hindrances to the same role can cancel each other out. As Hamilton (1964) also explained, this approach works only under an assumption of weak selection: coancestry cannot predict genetic similarity all by itself when gene frequencies have been changing.

Identity-by-descent provides useful results for close kin in an outbreeding population. However, for reasons that are well-rehearsed (Seger, 1981; Grafen, 1985), it cannot provide a sensible approach for calculating relatedness where there may be many weak ties going back many generations, as might frequently be the case between neighbours in a viscous population. Essentially, the concept of identity-by-descent depends on declaring all individuals in some recent generation as having no identity-by-descent, and then calculating on the basis only of ties since then. Seger's "paradox of inbreeding" is that if we declare the foundational generation to be long enough ago, then all pairs of organisms alive today have a relatedness of 1, as they all share complete descent from the common ancestor of all currently living creatures. This shows that the method of identity-by-descent cannot be conceptually central, and illustrates that even without selection, paths must not become too long: even drift disturbs the calculations eventually. But we do have a serviceable method of calculating r^t for some kinds of roles, and under some kinds of assumptions. Importantly, this method gives the same values of r^t for all alleles and all loci with the same transmission pattern.

Another modelling approach, used for grouped populations, is to use group sizes and immigration patterns to calculate genetic similarity, using the F-statistics of Wright (1969–1978). Note that F-statistics themselves can be calculated from identity-by-descent considerations or from measured frequencies of genotypes. They were first employed in relation to inclusive fitness by Hamilton (1971). Two members of a group will typically share genes through many lengthy linking ancestral paths. To use ordinary identity-by-descent is unsatisfactory because (i) the relevant calculations become poorer and poorer approximations as the lengths increase, owing to the increasing strength of cumulative selection and drift over lengthy paths (ii) it is less plausible that two group members know of all the ancestral links and so can act according to them, and more plausible that the information they have is simply that they belong to the same group. Recursion equations can provide measures of relatedness and F-statistics under simple assumptions. Other modelling approaches are in principle possible.

The 'measurement' approach was first advocated and employed as a complete approach by Orlove (1975) and developed as a regression coefficient by Orlove and Wood (1978). Hamilton (1972) had already mixed these methods in deriving coefficients for the case of inbreeding, estimating the inbreeding coefficient F_A, but calculating the probability of identity between random gametes of two individuals r_{AB} using identity-by-descent, in the course of presenting his regression coefficient of relatedness equal to $r_{AB}/(1 + F_A)$. In our current context, measurement takes the values of p_i and b_{ijt} as given, and calculates from them the relatedness associated with a given role, that is, it chooses the r^t to make the Hamilton residual zero. It is most natural to do this with a role such as 'neighbour' or 'fellow-group-member', but perfectly possible also with kinship roles such as 'full-sib' or 'second cousin'. Finding a value of r^t to make the t-th term in the Hamilton residual in Eq. (2) equal to zero is possible provided $\sum_i \sum_j (p_i - p) n_j b_{ijt} \neq 0$, which requires that the nb-weighted sum of actors' p-scores is not equal to the nb-weighted sum of the population mean p-score. The actors must be unrepresentative of the population in this very precise sense, and we discuss the exactly representative case further in Section 3.3. The required definition is

$$r^t = \frac{\sum_i \sum_j n_j(p_j - p)b_{ijt}}{\sum_i \sum_j n_j(p_i - p)b_{ijt}}. \tag{3}$$

This formula can be interpreted as a kind of regression, in which p_j is regressed on p_i across all pairs of individuals, with two special features. First, the regression is forced to pass through the point (p,p), and second the datapoint (p_i,p_j) is weighted by $n_j \sum_{t \neq e} b_{ijt}$. The failure when $\sum_i \sum_j n_j(p_i - p)b_{ijt} = 0$ can be explained in simple geometric terms, on the assumption that $\sum_i \sum_j n_j b_{ijt} \neq 0$. The mean values of p_i and p_j as weighted in the regression are the relevantly weighted mean p-scores for the actors and recipients, respectively. Writing them as $\bar{p}_A$ and $\bar{p}_R$, we have

$$\bar{p}_A = \frac{\sum_i \sum_j p_i n_j b_{ijt}}{\sum_i \sum_j n_j b_{ijt}}, \quad \bar{p}_R = \frac{\sum_i \sum_j p_j n_j b_{ijt}}{\sum_i \sum_j n_j b_{ijt}}.$$

Now the value of r^t is simply

$$r^t = \frac{\bar{p}_R - p}{\bar{p}_A - p}. \tag{4}$$

the slope from (p,p) to $(\bar{p}_A,\bar{p}_R)$. If it happens that the two points share the same value on the horizontal axis only, then the slope will be 'infinite'. As choosing $r^t = \infty$ does not sensibly reduce the Hamilton residual to zero in this case, this solution will not be useful to us. Note that the geometric analogy makes clear that measured relatednesses can take any positive or negative value and are not restricted to lie between zero and one, or minus one and one.

The conceptually appealing view of relatedness is as a measure of information about the recipient's genotype from the actor's point of view, and the regression slope is indeed predicting recipient's genotype from actor's genotype. This interpretation of the regression, first offered by Hamilton (1963), is developed at some length in Grafen (1985).

We now return to the comparison of the two types of approach to calculating relatedness. The modelling approaches are likely to produce relatednesses we can calculate, but only in special circumstances, where we happen to know the ancestral links between interactants. Further, they will imply only that the Hamilton residual is zero on average, and likely to be small in most cases. The measurement approaches rely on information that is most unlikely to be available, except in computer simulations of evolution. But measurement always in principle provides relatednesses that guarantee the Hamilton residual to be exactly zero. Thus measured relatednesses always make the link to gene frequency change exact, apart from the uncertainties of this generation's Mendelian segregation, but there is no guarantee that the relatednesses will be the same for all p-scores. In a moment we investigate chance variation in measured relatednesses by adding uncertainty in general to our population genetics model, and a suggestion for joint measurement is made in Section 6.3.

Our measurement and modelling approaches to dealing with the Hamilton residual in Eq. (2) should not be confused with the separate issue of which computational statistic is employed. Genetic similarity was originally presented formally as a correlation coefficient based on identity-by-descent (Hamilton, 1964), but Hamilton (1963) is quite clear that this only an approximation to the true nature of relatedness as a regression coefficient. Hamilton (1970) showed how relatedness could be defined as a regression coefficient, but continued to base it on identity-by-descent.

Forcing the regression line through the point (p,p) is technically very important (Grafen, 1985). Centring on the population mean makes the regression slope most meaningful, and most simply related to gene-frequency change. Regression slopes based on estimated intercepts require unnecessary correction terms that are complex and hard to interpret, when the opportunities for action are available to a non-representative subset of the population.

3.3. Representative actors

The t-th term in the Hamilton residual from Eq. (2) is

$$\sum_i \sum_j n_j(p_j - p)b_{ijt} - r^t \sum_i \sum_j n_j(p_i - p)b_{ijt}.$$

Choosing r^t to render this zero is always possible provided the sum multiplying r^t does not equal zero, and in this section we consider the biological significance of this exact circumstance, because it carries the implication that we cannot define r^t by 'measurement' to make inclusive fitness work. Note that under identity-by-descent arguments there is nothing special or difficult about this equality.

It is easy to understand the problem in the geometric terms of Grafen (1985), if slightly informally. The population mean, the actors' mean and the recipients' mean are three points in the one-dimensional space of possible p-scores. Actors are said to be representative when the actors' mean and the population mean are at the same point. Relatedness is the fraction of the distance from the population mean to the actors' mean at which the recipients' mean is found. This concept becomes meaningless when actors are representative. Actors may be representative for some p-scores but not for others.

We now turn to a more analytical approach. In some cases, which are unproblematic, the first sum in the Hamilton residual is zero as well as the second one. In this case the Hamilton residual equals zero whatever the value of r^t. Thus we can simply set $r^t = 0$, or, if r^t has taken a value in previous generations as gene frequencies have been changing, retain that value for the one generation when actors are representative. Both sums equalling zero will arise quite often, when the p-score is not causally involved in any social fitness effects, and is in linkage equilibrium with all the alleles that are; and will of course always happen if there are no social effects at all so $b_{ijt} = 0$ for $t \neq e$. This just means the whole social apparatus is irrelevant to that p-score, and need not concern us in principle. There is a practical issue, however. If both sums are very close to zero, as might arise through small chance effects in the b^s_{ijt}, or through small chance linkage disequilibrium, then the measured value of $r^{t,s}$ will vary wildly as two very small values that are zero on average wobble about. But by virtue of being very small, they will have little effect on net selection anyway.

The issues in principle arise, therefore, if the second sum equals zero while the first does not, for systematic reasons. In geometric terms, the actors' mean and population mean are at the same point, but the recipients' mean is elsewhere. No difficulty arises when role t satisfies the assumptions of identity-by-descent, for if the actors' mean and population mean are the same point, then so is the recipients' mean, at least on average, and the relatednesses (r^t) do not vary with s and so do deal with the average. A simple problematic kind of role would be 'A random red-headed sibling if there are red-headed siblings; otherwise a random sibling'. Here the gene for red hair benefits at the expense of all alleles, and the recipient's genotype will be systematically biassed towards red hair genes compared to the actor's genotype. Here the measurement approach allows us in general to find a relatedness that will make inclusive fitness work, though one might wonder at the utility of such an interpretation, and this is a good example of where measurement of relatedness produces accurate but unmeaningful results. It is also an example where the representativeness of actors would make the relatedness calculation fail because of dividing by zero.

It is also worth noting that if the trait is genetically staightforward in relation to the p-score, so that the higher the p-score the more the individual gives to recipients, then actors will never be representative of the population, as their p-score will be always be higher than the population average, except when the trait goes to fixation and then again both sums are zero.

Thus this failure of measurement is likely to arise only in unusual circumstances, and not when identity-by-descent assumptions are met, or when allelic effects are straightforward. So there is a very occasional failure in principle to be able to find an r^t that will render the Hamilton residual equal to zero, but it should not cause concern except in peculiar cases.

3.4. The Price equation under uncertainty

The model of the previous section holds good for fixed b_{ijt}. In order to accommodate flexibility of behaviour on the part of the organisms, and in order to deal more formally with Mendelian segregation, it is important to extend the model to allow uncertainty. We will take the population of individuals and their genotypes as fixed, but allow all other components of the model to depend on the situation that arises. Thus n_i, p_i and p do not vary with the situation s, but we will superscript other notation thus to indicate that they do depend on s: w^s_j, w^s, p'^s, Δp^s and Δp^s_j. It is important to note that variation in b^s_{iie} can allow individuals' fitnesses to vary randomly and differently across situations.

Suppose situation s arises with probability τ^s, and that the set of situations is S, so that $\sum_{s\in S}\tau^s = 1$. It is helpful to consider the situation to include information about how Mendelian segregations are resolved. Partition S into subsets $(S^u)_{u\in U}$ such that the situations within a subset differ only in Mendelian segregations. Agree to use a bracketed superscript $S^{(s)}$ to denote S^u such that $s \in S^u$. The technical representation of our assumptions of fairness and independence of meiosis is that two situations belonging to the same S^u have equal probabilities τ^s. It follows that $\sum_{s'\in S^{(s)}} \Delta p^{s'}_j = 0$ for all j and s.

As in the absence of uncertainty, the Price equation holds in the face of arbitrary linkage, linkage disequilibrium, ploidy levels, non-random mating, a mixture of sexual and asexual reproduction, and population structure.

Obtaining the formula for the change in gene frequency in situation s is straightforward. Eq. (2) is simply rewritten with the relevant symbols superscripted for situation s:

$$w^s\Delta p^s = \mathbb{C}\left[p_i, \left(b^s_{iie} + \sum_{t\neq e} r^{t,s}\sum_j \frac{n_j}{n_i} b^s_{ijt}\right)\right] + \mathbb{E}[w^s_j \Delta p^s_j]$$
$$+ \frac{1}{N}\sum_{t\neq e}\left(\sum_i\sum_j n_j(p_j - p)b^s_{ijt} - r^{t,s}\sum_i\sum_j n_j(p_i - p))b^s_{ijt}\right).$$

We focus on the expected value of Δp^s, namely $\sum_{s\in S}\tau^s\Delta p^s$. w^s must be moved to the right-hand side to make further progress, and its natural place is in the denominator of the target of selection, to make 'relative fitness'. Then multiplying by τ^s and summing over s we obtain a version of Eq. (2) that incorporates uncertainty as follows:

$$\sum \tau^s \Delta p^s = \sum_{s\in S}\tau^s\mathbb{C}\left[p_i, \left(\frac{b^s_{iie} + \sum_{t\neq E} r^{t,s}\sum_j \frac{n_j}{n_i} b^s_{ijt}}{w^s}\right)\right]$$
$$+ \sum_{s\in S}\tau^s\mathbb{E}\left[\frac{w^s_j}{w^s}\Delta p^s_j\right]$$

$$+\frac{1}{N}\sum_{s\in S}\frac{\tau^s}{w^s}\sum_{t\neq e}\left(\sum_i\sum_j n_j(p_j-p)b^s_{ijt}\right.$$
$$\left.-r^{t,s}\sum_i\sum_j n_j(p_i-p))b^s_{ijt}\right). \qquad (5)$$

The second summand can now be dealt with formally. We saw above that $\sum_{s'\in S^{(s)}}\Delta p_j^{s'}=0$, but we also know that w_j^s is equal for all situations belonging to the same S^u, as such situations differ *only* in Mendelian segregation. Thus $\sum_{s'\in S^{(s)}}(w_j^s/w^s)\Delta p_j^{s'}=0$ for all j and s, and so the covariance representing the second summand as a whole equals zero.

In the first and third summands, the sum over situations can be brought inside the other structures, as p_i does not depend on s, and so the final Price equation with uncertainty becomes

$$\sum\tau^s\Delta p^s=\mathbb{C}\left[p_i,\sum_{s\in S}\tau^s\left(\frac{b^s_{iie}+\sum_{t\neq e}r^{t,s}\sum_j\frac{n_j}{n_i}b^s_{ijt}}{w^s}\right)\right]$$
$$+\frac{1}{N}\sum_{t\neq e}\left(\sum_i\sum_j n_j(p_j-p)\sum_{s\in S}\tau^s\frac{b^s_{ijt}}{w^s}\right.$$
$$\left.-r^{t,s}\sum_i\sum_j n_j(p_i-p))\sum_{s\in S}\tau^s\frac{b^s_{ijt}}{w^s}\right). \qquad (6)$$

The target of selection here is the arithmetic average of the inclusive fitness effect relative to mean neighbour-modulated fitness, which will be referred to as 'expected relative inclusive fitness'.

What are the consequences of incorporating uncertainty into the model? It extends the model to a more realistic situation, as uncertainty will always hold. It allows the assumption of perfect transmission to be expressed and render the second summand in Eq. (5) exactly zero. Further, it extends to social behaviour the result of Grafen (2002) that expected relative fitness is the maximand in natural selection. The highest mean expected relative fitness will be selected, and the variance in relative fitness is irrelevant.

Next, how does the uncertainty affect the Hamilton residual? The only change formally between Eqs. (2) and (6) is that the definite fitness effect b_{ijt} is replaced with the expected relative fitness effect $\sum_s\tau^s b^s_{ijt}/w^s$. The case of identity-by-descent is simple, as the individuals, their genotypes and ancestral links are all fixed, and independent of the situation. Thus if a role can have its relatedness calculated by identity-by-descent, then it will not vary with the situation: formally, $r^{t,s}$ will not depend on s for such roles. Other modelling methods might allow dependence. For example, if t denoted 'neighbour', then in a very windy year organisms in some species may find themselves with a lower genetic similarity to neighbours than they would have in a very still year.

The case of measurement is more complicated. It seems likely to require dependence on s, as it relies on observing values of b^s_{ijt}. Supposing that everything can be observed except Mendelian segregation, what we know is which value of u occurs and which subset S^u of situations has occurred. Then the full measurement approach is to make the Hamilton residual equal to zero uniformly across situations by choosing

$$r^{t,s}=\frac{\sum_i\sum_j n_j(p_j-p)\sum_{s'\in S^{(s)}}\tau^{s'}b^{s'}_{ijt}/w^{s'}}{\sum_i\sum_j n_j(p_i-p)\sum_{s'\in S^{(s)}}\tau^{s'}b^{s'}_{ijt}/w^{s'}}$$
$$=\frac{\sum_i\sum_j n_j(p_j-p)b^s_{ijt}}{\sum_i\sum_j n_j(p_i-p)b^s_{ijt}}, \qquad (7)$$

where the second simpler equation is possible because for $s\in S^u$, the values of τ^s, w^s and b^s_{ijt} do not vary with s. With this definition the Hamilton residual will be zero, and so the change in Δp^s, averaged over the s belonging to a given S^u (and so over Mendelian segregation in production of offspring), is equal to the covariance between p-score and observed (or we could say realized) relative inclusive fitness. This 'measured relatedness' will therefore absorb random changes in gene frequency, and so vary randomly itself, as discussed in Section 6 where the possibility of measuring relatedness jointly for a number of p-scores is suggested.

A further consequence is that a graduation of possible actions can be appropriately represented. With just one situation, the kind of help is strongly constrained, as an action must alter the recipient's number of successful gametes by an integer. Allowing uncertainty permits the same action to lead to no gain in some situations and some gain in other situations, so the average gain can take non-integer values.

4. Optimization programs

The purpose of this section is to construct an optimization program, which can then be linked in Section 5 to gene frequency changes through Eq. (6). Optimization programs are rarely seen in the biological literature (though see Grafen, 1998, 2002, forthcoming), but are a standard tool in game theory and economics (e.g. Mas-Colell et al., 1995). To make sense of the book title *The Selfish Gene* (Dawkins, 1976) in formal terms, it is necessary to have a mathematical description of how genes behave, and also a mathematical encapsulation of selfishness. The Formal Darwinism project (Grafen, 1999, 2000, 2002, forthcoming) therefore aims to link the mathematics of motion (difference and differential equations) used to describe gene frequency trajectories with the mathematics of optimization used to describe purpose and design.

Section 4.1 develops in stages a model of the individual's decision problem along with an optimization program that can represent the sophisticated optimality that is expected of organisms in behavioural ecology and related disciplines. A simple example worked out in Section 4.2 shows how the formalities of the final optimization program represent sophistication in decision-taking. An abstract model is constructed in Section 4.3 that expresses the maximized quantity in the optimization program in terms of numbers of successful gametes of the decision-taker and others. This will allow a link to population genetics in the next section. In the course of the section, we make notationally explicit that the Price equation refers to a particular class of p-scores.

4.1. Developing an optimization program

Optimization programs focus on one implicit individual, the decision-taker. Suppose the decision-taker finds herself in one of a number of situations, and recall that S denotes the set of possible situations. In situation $s \in S$, suppose there is a finite set of decisions D_s to be taken. In a decision $d \in D_s$, there is a (finite or infinite) set of actions A_d available. A strategy is a function α that says what to do in each decision, and so is a function from the set of all decisions $D \equiv \bigcup_s D_s$ to the set of all actions $A \equiv \bigcup_d A_d$, and a feasible strategy satisfies at least $\alpha(d) \in A_d$. Then we will denote the maximand in situation $s \in S$ playing strategy α by $\pi(s, \alpha)$. After setting out optimization programs employing π, an expression will be sought for it in terms of the population genetic quantities of previous sections. For the moment, we simply explore how to represent an assumption that there is some quantity that is maximized.

Our first optimization program, for an individual in a fixed situation s, is written as follows:

$$\begin{aligned} &\alpha \max \pi(s, \alpha), \\ &\quad \alpha(d) \in A_d \quad \forall d \in D_s. \end{aligned} \tag{ProgS}$$

The name ProgS indicates that it is for individuals in one particular situation. The term before the 'max' indicates the *instrument*, in this case the decision rule α. The maximand $\pi(s, \alpha)$ appears after the 'max'. On succeeding lines appear the constraints, although in this case there is only one line, representing the feasibility constraint for each decision. Note that we have required feasibility only for the decisions in situation s.

An introduction to the mathematical structure of optimization programs can be found in economics textbooks (e.g. Mas-Colell et al., 1995). An optimization program may have a *solution*. A solution α^* to ProgS must itself obey the constraint, and is defined by the inequality $\pi(s, \alpha^*) \geqslant \pi(s, \alpha)$ for all α satisfying the constraint. There may be more than one solution, and some optimization programs have none. In the case of ProgS, because the behaviour in other situations does not affect the maximand, a solution may well have other equally good feasible strategies, which differ only on decisions that do not arise in situation s. The *value* of a program is the value of the maximand at a solution.

This first program is now developed in two important ways to make it more general. The first is to create a program for the whole decision rule by assuming that the probability-weighted arithmetic average of $\pi(s, \alpha)$ is the maximand for the whole decision rule across all situations. There has been some discussion, under the topic of bet-hedging, about whether fitnesses should be arithmetically or geometrically averaged in various different circumstances (Seger and Brockmann, 1987; Grafen, 1998, 2000). In this case, it will come out in the proof of links between the optimization program and gene frequency change that the arithmetic averaging works.

The second extension removes a restriction of the model of choice used so far. We have assumed that in decision d the actions A_d are available; and ProgS is based on assuming that all the actions in A_d are always available *irrespective of actions taken in other decisions*. It is likely, however, that decisions are inter-related, and that actions in some decisions constrain actions in others. This possibility can be formally written as follows. The set of available actions indexed by decisions is $(A_d)_{d \in D}$. We can constrain the decision rule, requiring that $(\alpha(d))_{d \in D}$ belongs to an arbitrary subset B of $(A_d)_{d \in D}$. Making both these extensions, we can write ProgASC as

$$\begin{aligned} &\alpha \max \sum_{s \in S} \tau^s \pi(s, \alpha), \\ &\quad (\alpha(d))_{d \in D} \in B \end{aligned} \tag{ProgASC}$$

and we will also use the shorthand $\pi(\alpha)$ for $\sum_{s \in S} \tau^s \pi(s, \alpha)$.

So far, the optimization programs have been built up quite independently of the gene frequency equations. Their interest depends on first, the expression we can find for $\pi(s, \alpha)$ in terms of the population genetics model, and second, the links we can establish between gene frequency changes and the optimization program. Before doing so, it is worth showing how ProgASC formalizes a widespread assumption about animal behaviour. Behavioural Ecology and other subjects essentially use this program as a working hypothesis.

4.2. Example

ProgASC has a number of important properties, and we elaborate a simple example, which even lacks social behaviour, to display some of them. Suppose an organism has to choose between eating and hiding, and that in the absence of predation its basic payoff of 1 is decreased to 0.9 if it hides, through failing to eat, but is increased to 1.05 if it eats. However, there is a predator who may be far or near. If far, then there is no chance of predation and these payoffs stand, but if the predator is near and the organism is eating, then there is a 0.2 chance that the organism itself will be eaten and have its payoff reduced to zero. There is no chance of being predated while in hiding. Thus the net payoffs are given in Table 1.

To introduce information processing, we assume that the reeds rustle either quietly or loudly. The probabilities of the predator being near or far, and the reeds rustling quietly or loudly, are given in Table 2.

To complete the specification of the program we need to decide on the value of B, the set of actions available to the organism, and two will be discussed. B_1 insists that the organism must behave as if deaf, and take the same action irrespective of the reeds. B_2 permits the organism to make entirely separate decisions in the cases of loud and quiet rustling.

We can therefore write the maximand of ProgASC when $B = B_1$ as

$$\begin{cases} 1.05 \times (\sigma_{FL} + \sigma_{FQ}) + 0.84 \times (\sigma_{NL} + \sigma_{NQ}) & \text{if Eat} \\ 0.9 & \text{if Hide} \end{cases}$$

An organism that solves this optimization program, and would also solve it for other values of the payoff matrix, acts as if it knows the value of $(\sigma_{FL} + \sigma_{FQ})$ and $(\sigma_{NL} + \sigma_{NQ})$, which are the probabilities that the predator will be far or near, respectively. Thus the organism acts as though it has a correct prior probability distribution over the location of the predator.

Table 1
Payoffs for eating and hiding when predator is far and near

	Far	Near
Eat	1.05	0.84
Hide	0.90	0.90

Net payoffs are calculated by assigning zero where the organism is predated, which has probability 0.2 when the predator is near and 0 when the predator is far, and the value of 1.05 or 0.9 when the unpredated organism has been eating or hiding, respectively.

Table 2
Probabilities of rustling and predator position

	Loud	Quiet
Far	σ_{FL}	σ_{FQ}
Near	σ_{NL}	σ_{NQ}

The probabilites are merely given symbols with subscripts denoting the situation they correspond to. The four probabilities are assumed to sum to 1.

Now consider the maximand of ProgASC when $B = B_2$, where the program allows independent decisions depending on the rustling of the reeds. It is

$$\left.\begin{matrix} 1.05 \times \sigma_{FL} + 0.84 \times \sigma_{NL} & \text{if Eat when Loud} \\ 0.9 \times (\sigma_{FL} + \sigma_{NL}) & \text{if Hide when Loud} \end{matrix}\right\} + \left.\begin{matrix} 1.05 \times \sigma_{FQ} + 0.84 \times \sigma_{NQ} & \text{if Eat when Quiet} \\ 0.9 \times (\sigma_{FQ} + \sigma_{NQ}) & \text{if Hide when Quiet} \end{matrix}\right\}$$

The first and second terms are therefore optimized separately, and the first term is proportional to the optimization program conditional on the reeds rustling loudly. The maximand of this conditional program, which represents the strategic position once it is known that the rustling is loud, is found by dividing by the probability that the rustling is loud $(\sigma_{FL} + \sigma_{NL})$, thus:

$$\begin{cases} 1.05 \times \frac{\sigma_{FL}}{\sigma_{FL}+\sigma_{NL}} + 0.84 \times \frac{\sigma_{NL}}{\sigma_{FL}+\sigma_{NL}} & \text{if Eat when Loud} \\ 0.9 & \text{if Hide when Loud} \end{cases}$$

The first point to make is that solving the whole program therefore requires solving the separate programs corresponding to the different information states (loud or quiet rustling) the organism may be in. We may safely generalize to conclude that if B permits completely separate decisions in subsets of cases, to solve the whole program requires solving each subset separately, and this represents appropriate flexibility in response to cues.

The second point is that we can view the tranformation from the unconditional program with $B = B_1$ to the two parts of the program with $B = B_2$ in terms of Bayesian updating. The probability of the predator being far is $(\sigma_{FL} + \sigma_{FQ})$ when no information is available, but is $\sigma_{FL}/(\sigma_{FL} + \sigma_{NL})$ when it is known the rustling is loud and $\sigma_{FQ}/(\sigma_{FQ} + \sigma_{NQ})$ when it is quiet. Again we may generalize: if the information received is correlated with information about relevant factors in making the decisions, solving the whole program implies optimal Bayesian updating of the prior distributions to obtain the correct posterior distributions for the relevant factors.

In the notation employed in this paper, the cues received are left implicit—they are represented in the structure of B. The non-social case of Grafen (2002) has a similar program in which the cues received by the organism are explicitly modelled, but the notation in general is more complex. The point of this example has been to show that an organism playing a solution to ProgASC is acting as though it was a sophisticated decision-taker, with correct prior distributions over uncertainty, flexible responses to cues, and appropriate Bayesian updating in response to information received.

When it is shown in Section 5 that a kind of population genetic equilibrium implies a solution to such a program, therefore, this will be a powerful link between population genetics and optimality.

Behavioural ecologists who expect organisms to maximize their inclusive fitness expect them to do so conditionally, and appropriately in the light of information received. Organisms are so sophisticated, it is hard to think of any unconditional behaviour. Thus the formal treatment of conditional behaviour in no sense goes beyond what is required for the justification of a very common working hypothesis. Conditional behaviour is the same as environmental-dependence, but with emphasis on strategic consequences rather than mechanism.

4.3. A linking model

We now embark on the second step and turn to a useful choice of the maximand in population genetic terms. The obvious possibility is a target of selection in the Price equation, and the general thrust of the paper makes quite clear that the expected relative inclusive fitness effect in Eq. (6) will be chosen. But we must first deal with three notational preliminaries.

First, we have so far been working with a set of p-scores that share the same transmission pattern (which by itself would define the coreplicon of Cosmides and Tooby, 1981) and the same relatednesses, but now we need to indicate this in our notation. Let us call such a set of p-scores a *family of p-scores*, or more briefly a p-*family*, give the p-family with which we are working a name—Q—and indicate on relevant symbols that they refer to the p-family Q. The basic variables that depend on Q are the ploidies n_j, and the number of successful gametes w_j, from the population genetic side of the argument, and the relatednesses $r^{t,s}$. A man is haploid for X- and Y-chromosomes, and diploid for autosomes. The fitnesses depend on Q because genes in different p-families can have different values of w_j: for example, a man with one daughter and one son has $w_j = 2$ for his autosomes, $w_j = 1$ for his X- and Y-chromosomes, but 0 for his mitochondria. The relatednesses vary for reasons explained further in Section 6. We will not burden n_j, w_j, b^s_{ijt} and $r^{t,s}$ with their superscript Q. It is, however, useful to notice the dependence on p-family in variables being used in bringing the population genetics and optimization side of the argument together. Thus, for example, because the ploidies n_j affect the weighting in the expectation and covariance, we need to write $\mathbb{E}^Q[\cdot]$ and $\mathbb{C}^Q[\cdot,\cdot]$.

Second, we will now use more frequently the longer notation in which we recognise that the social effects b^s_{ijt} depend on the phenotype of the actor i, so that $b^s_{ijt}(\alpha)$ represents the effect of i on j in role t when i has phenotype α. Recall that we notate the strategies actually played as γ_i, so the simpler b^s_{ijt} has stood and will continue to stand for $b^s_{ijt}(\gamma_i)$.

Third, other ingredients of the formula for inclusive fitness, w^s and, if relatednesses are measured, $r^{t,s}$, also depend on phenotypes, but we will want to ignore those dependencies when the maximand varies with an individual's changing strategy. We therefore adopt the device of using other variables in their place. Thus we will use $\rho^{t,s}$ for relatednesses, and ω^s for mean neighbour-modulated fitness, and consider these as parameters of the maximand and the optimization program. To lighten the notation, we will treat ρ as a matrix and ω as a vector, so that we can write ρ and ω instead of $(\rho^{t,s})_{t\in T,s\in S}$ and $(\omega^s)_{s\in S}$.

Returning to the substantive issue, we first note that in choosing expected relative inclusive fitness effect to be the maximand, the question arises *whose*? Let us suppose for the moment that it is individual k, and recognize this by adding a subscript k when we define the absolute inclusive fitness effect in situation s, and the relative expected inclusive fitness effect, ϕ^Q_k and Φ^Q_k, respectively, as

$$\phi^Q_k(s,\alpha,\rho) = b^s_{kke}(\alpha) + \sum_{t\neq e} \rho^{t,s} \sum_j \frac{n_j}{n_k} b^s_{kjt}(\alpha), \tag{8}$$

$$\Phi^Q_k(\alpha,\rho,\omega) = \sum_{s\in S} \tau^s \frac{\phi^Q_k(s,\alpha,\rho)}{\omega^s}. \tag{9}$$

These expressions are pivotal in the argument, and central to them is the way dependence on p-family Q is distributed. The n_j and the b^s_{jks} depend on Q at an individual level. The useful optimization program for linking to a population genetic equilibrium will of course be that in which the relatednesses and mean fitnesses in the program are chosen to be equal to their values in the equilibrium, so the appropriate choices will be $\omega = (w^s)$ and $\rho = (r^{t,s})$: for then $\Phi_k(\alpha,\rho,\omega)$ equals the expected relative inclusive fitness of individual k. The bivalent role of Φ, defined in population genetic terms but used as the maximand of the optimization program, is a major essential element of the whole argument.

The strong assumption of universal strategic equivalence is now made. The optimization program has just been linked to particular individual k, but we could have chosen any individual for this purpose. Universal strategic equivalence means that for all 'fair' ω,

$$\Phi^Q_k(\alpha,\rho,\omega) = \Phi^Q_j(\alpha,\rho,\omega) \quad \forall j,k\in I, \alpha\in B, \rho$$

and that the set of feasible strategies is the same for all individuals. See Section 5.1 for further discussion of the assumption, including the definition of 'fairness' of ω. The essential point is that we are restricting our analysis to the case in which all individuals face the same evolutionary problems, when averaged over all situations.

Thus we had in ProgASC an optimization program that represented sophisticated behaviour on the part of organisms, and we have now identified the maximand in terms of population genetic quantities. Here we formally rewrite the program with an explicit dependence on important parameters, namely the p-family used to define the maximand, the set of feasible strategies, the relatednesses and the mean fitnesses. Fortified by universal strategic equivalence, we may also drop the subscript k from Φ to obtain

$$\begin{aligned} \alpha \quad & \max\ \Phi^{Q}(\alpha, \rho, \omega), \\ & (\alpha(d))_{d\in D} \in B. \qquad\qquad (\mathrm{ProgIF}(\mathrm{Q}, \mathrm{B}, \rho, \omega)) \end{aligned}$$

This allows us to consider how selection acts on different p-families by contrasting $\mathrm{ProgIF}(Q, B, \rho^{Q}, \omega^{Q})$ and $\mathrm{ProgIF}(R, B, \rho^{R}, \omega^{R})$, where the relatednesses and mean fitnesses are shown to depend on Q as they typically will; and how it acts with different constraint sets by comparing $\mathrm{ProgIF}(Q, B_1, \rho, \omega)$ and $\mathrm{ProgIF}(Q, B_2, \rho, \omega)$, as we did in Section 4.2.

5. Links between population genetics and the optimization program

Section 3.4 constructed a population genetic model, and Section 4 proposed an optimization program and an expression for the maximand in population genetic terms. Here, by proving links between them, the connection is explicitly articulated between population genetics and the optimization of inclusive fitness. We begin by defining equilibrium concepts for the population genetics model, and then we prove four results linking the optimization program and population genetics. In order to complete the proofs, it will be necessary to use again the assumption that the actor controls the performance of the action and its quantitative consequences.

It is important to recall that our population genetic analysis holds only for one p-family, namely the one whose transmission pattern is encapsulated in the values of w_j^s and whose relatednesses are represented by the $r^{t,s}$. Recall that dependence on p-family is now indicated explicitly by a superscript, for example $\Phi^{Q}(\alpha, \rho, \omega)$ denotes the maximand of $\mathrm{ProgIF}(Q, B, \rho, \omega)$, where Q represents the p-family.

The equilibrium concepts for gene frequencies are now introduced. The first concept is 'scope for Q-selection', which means that individuals have different values of $\Phi_i^{Q}(\gamma_i, r, (w^s))$, so that p-scores belonging to Q can be found that are subject to selection on average. Note that we allow hypothetical p-scores, in which we assign a number to each individual, without asking whether there is an actual set of allelic weights that does produce that set of numbers. Thus no scope for Q-selection is defined by $\sum_{s\in S} \tau^s \Delta p^s = 0$ for all p-scores in Q. This in turn implies that $\Phi_i^{Q}(\gamma_i, r, (w^s))$ is constant for all i, as otherwise we could define a p-score to distinguish one individual i from all the others which would be under selection.

The second concept is 'potential for Q-selection', which is defined in relation to a set X of possible strategies: 'no potential for Q-selection in relation to X' means that no phenotype $\alpha \in X$ would have been favoured by selection in the p-family Q had that phenotype been present, *where we neglect the consequent effect on population mean fitness*. Formally, suppose that one individual, say h, has her strategy γ_h replaced with α. If we let Δ_i represent the difference made to Φ_i by the change in strategy, we conclude that $\Delta_h = \Phi_h^{Q}(\alpha, r, (w^s)) - \Phi_h^{Q}(\gamma_h, r, (w^s))$ while the assumption of actor's control ensures that the inclusive fitnesses of all other individuals are unchanged, so $\Delta_i = 0$ for $i \neq h$. Let q_i^h be a p-score that equals one for $i = h$ and zero otherwise, and note that its mean equals n_h/N. Then selection on this p-score with the altered phenotype proceeds according to

$$\begin{aligned} \Delta p^h &= \mathbb{C}[p_i^h, \Phi_i^{Q}(\gamma_h, r, (w^s)) + \Delta_i] \\ &= \mathbb{C}[p_i^h, \Phi_i^{Q}(\gamma_h, r, (w^s))] + \mathbb{C}[p_i^h, \Delta_i]. \end{aligned}$$

On the assumption of no scope for Q-selection, the first of the two covariances equals zero, and we proceed to obtain

$$= \mathbb{C}[p_i^h, \Delta_i] = \sum_i n_i (p_i^h - p^h)\Delta_i = \frac{n_h}{N}\left(1 - \frac{n_h}{N}\right)\Delta_h$$

The new allele would spread if

$$\frac{n_h}{N}\left(1 - \frac{n_h}{N}\right)[\Phi_h^{Q}(\alpha, r, (w^s)) - \Phi_h^{Q}(\gamma_h, r, (w^s))] > 0.$$

Some q^h would spread unless this covariance is non-positive for all h and α, so the condition for no potential for Q-selection in relation to the set X can be written as

$$\Phi_h^{Q}(\alpha, r, (w^s)) - \Phi_h^{Q}(\gamma_h, r, (w^s)) \leqslant 0 \quad \forall h \in I, \alpha \in X. \qquad (10)$$

This definition is made for all population sizes, but the neglect of the effect on population mean fitness makes most sense when the population, though finite, is large. The awkwardness, in principle, would be if the change in one individual's strategy altered the w^s considerably, so changing the balance of importance of different situations in comparing two strategies: then a strategy might in reality spread even though there was no 'potential for Q-selection'. It is like a central limit theorem assumption, that no individual should be very important, even in just one situation. Further work on small populations could be useful.

Summarizing, 'no scope for Q-selection' means that gene frequencies in the p-family Q will not change on average from this generation to the next, while 'no potential for Q-selection' means that no mutant in the p-family Q with a single phenotypic expression would have increased in frequency in expectation, and is relative to some set X of possible mutants.

Four propositions are now proved connecting population genetics and optimization programs. The major assumptions required to arrive at this point include additivity of social effects, actor's control, and universal strategic equivalence.

Proposition 1. *Suppose the population is playing strategies* $(\gamma_i)_{i\in I}$, *and that the Hamilton residual is rendered negligible by relatednesses* $(r^{t,s})$. *If each individual in the population is playing a strategy that is optimal in* $\mathrm{ProgIF}(Q, B, r, (w^s))$, *then there is no scope for Q-selection in the population genetics model, and no potential for Q-selection in relation to B.*

Proof. If all individuals solve ProgIF, which by universal strategic equivalence is the same program for all individuals, then they must achieve the same value of the maximand, hence $\Phi_k^Q(\gamma_k, r, (w^s))$ is equal for all k. This establishes no scope for selection. Further, since γ_k is a solution, $\Phi_k^Q(\alpha, r, (w^s)) \leqslant \Phi_k^Q(\gamma_k, r, (w^s))$ for all $\alpha \in B$, and through Eq. (10) it follows that there is no potential for Q-selection in relation to B. □

Proposition 2. *Suppose the population is playing strategies* $(\gamma_i)_{i\in I}$, *and that the Hamilton residual is rendered negligible by relatednesses* $(r^{t,s})$. *If all individuals in the population do not solve* $\mathrm{ProgIF}(Q, B, r, (w^s))$ *but nevertheless achieve an equal value of the maximand, then there is no scope for Q-selection in the population genetics model, but there is potential for Q-selection in relation to B.*

Proof. If each individual attains the same value of the maximand in $\mathrm{ProgIF}(Q, B, r, (w^s))$, then by universal strategic equivalence, it follows that $\Phi_k^Q(\gamma_k, r, (w^s))$ is equal for all k, establishing no scope for Q-selection. But if they do not solve the program then there is an $\alpha \in B$ such that $\Phi_h^Q(\alpha, r, (w^s)) > \Phi_h^Q(\gamma_h, r, (w^s))$ for some h, and so there is potential selection for a p-score that picks out individual h. Hence, there is potential for Q-selection in relation to B. □

Proposition 3. *Suppose the population is playing strategies* $(\gamma_i)_{i\in I}$, *and that the Hamilton residual is rendered negligible by relatednesses* $(r^{t,s})$. *If individuals in the population play strategies with different values of the maximand in* $\mathrm{ProgIF}(Q, B, r, (w^s))$, *then there is scope for Q-selection, and the expected change in each p-score in family Q equals its covariance across individuals with the attained value of the maximand.*

Proof. If individuals attain different values of the maximand, then

$$\mathbb{C}^Q[p_i, \Phi_i^Q(\gamma_i, r, (w^s))] \neq 0$$

for the p-score defined by $p_i = \Phi_i^Q(\gamma_i, r, (w^s))$, and so there is scope for Q-selection. The Price equation (6) states that the expected change in any p-score equals its covariance with the value of the maximand.

Proposition 4. *Suppose the population is playing strategies* $(\gamma_i)_{i\in I}$, *and that the Hamilton residual is rendered negligible by relatednesses* $(r^{t,s})$. *Further suppose there is no scope for Q-selection in the population genetics model, and no potential for Q-selection in relation to B. Then each individual acts rationally in the sense that each plays a strategy that solves* $\mathrm{ProgIF}(Q, B, r, (w^s))$.

Proof. If there is no scope for Q-selection and no potential for Q-selection in relation to B, it follows from Eq. (10) that $\Phi_k^Q(\gamma_k, r, (w^s)) \geqslant \Phi_k^Q(\alpha, r, (w^s))$ for all $\alpha \in B$ and k. This by definition implies that γ_k is a solution of $\mathrm{ProgIF}(Q, B, r, (w^s))$. □

A number of important points need to be made about these conclusions, which represent the first explicit connection between population genetics and the optimization of inclusive fitness. First, they link the optimization program to gene frequency change, and do not have anything to say about genotype frequencies.

Second, the emphasis on gene as opposed to genotype frequencies is of conceptual importance as well as historical interest. Fisher (1930) presented his fundamental theorem in the same way, and in a 1955 letter to O. Kempthorne explicitly defended the view that a population evolves to the extent that its gene frequencies change (Bennett, 1983, p. 228):

> if by extinction of certain insects a plant were rapidly to become generally self-fertilised and homozygous through lack of means to cross-pollination, I should, so long as the gene ratios remained unchanged, consider that the plant had not evolved but was responding passively to its changed environment

See Grafen (2003) for a further discussion. Fisher dealt with genotype frequencies when it was required for the problem at hand, including a substantive paper on linkage under polysomy (Fisher, 1947) and a book on inbreeding (Fisher, 1949). His emphasis on gene frequencies in the quotation does not stem, therefore, from an inability or dislike for the extra work involved, but rather from the conviction, fully backed up by mathematical proofs, that gene frequencies play a critical role in the formal representation of the central argument of Darwin (1859).

For present purposes, it is enough to stress that the emphasis on gene frequencies in no way compromises the exactness of the conclusions or the rigour of the arguments, and to recall that interesting evolution, say from their common ancestor to humans and to chimps, is likely to be caused by gene frequency change and not, for example, by fluctuations in linkage disequilibrium. Grafen (2002) includes the example of sickle-cell anaemia in a longer discussion of other types of genotypic change besides gene frequency, in the parallel propositions for non-social Darwinian fitness.

Third, the links do not just hold at optimality. Other foundational work aimed at linking 'phenotypic methods' with population genetic methods, such as Hammerstein (1996) and Taylor (1996), focus on the final outcome of evolution. The results here show non-equilibrium connections that reflect biologists' use of the term 'fitness'.

Next, the fourth proposition is of particular interest, as it moves from a hypotheses based purely on dynamic population genetic conditions to a conclusion about sophisticated rationality, of the kind illustrated in Section 4.2.

Finally, the results apply to *expected* gene frequency changes. Thus the observation of selection does not automatically contradict the assumptions of the model. The incorporation of uncertainty in Section 3.4 makes the results in principle stronger, but also more remote from empirical observations.

5.1. Universal strategic equivalence

There is only one decision-taker in an optimization program, but a whole population in the population genetics model. Reducing the population to a single individual involves making an assumption that all individuals in the population face the same strategic situation. This section discusses the exact nature of the assumption required, and the limitations on its biological reasonableness. A weaker assumption would require a more complex definition and justification of inclusive fitness.

The assumption was defined by requiring that

$$\Phi_k^Q(\alpha,\rho,\omega) = \Phi_j^Q(\alpha,\rho,\omega) \quad \forall j,k \in I, \;\; \alpha \in B, \rho$$

for all 'fair' ω, and we now define this fairness by requiring that there is a set of strategies $(\alpha_i)_{i\in I}$, $\alpha_i \in B$, such that the ω^s are the mean neighbour-modulated fitnesses in situation s, or formally,

$$\omega^s = \frac{\sum_j n_j\left(1 + b^s_{jje}(\alpha_j) + \sum_{t\neq e}\sum_i b^s_{ijt}\right)}{\sum_j n_j}.$$

This is a highly technical assumption, and is the price to be paid for much of the simplicity of notation. In the parallel work for non-social behaviour (Grafen, 2002), the behaviour and information are more explicitly articulated at the cost of more complex notation, and the parallel assumption of pairwise exchangeability is biologically more meaningful as a result. One key to understanding is that if ω could be chosen freely, then in many cases, the assumption could never be met and the propositions would never be true.

One important point about universal strategic equivalence is that we do not assume that individuals are equivalent within any one situation. Individuals are allowed to be lucky and unlucky, and just plain different from each other, in any one situation. It is only when averaged over all situations that their strategic position is assumed to be the same.

But it is not always reasonable to assume that individuals are identical, even while the 'veil of ignorance' obscures which situation will prevail. In many species gender is not assigned by situation, and other determining circumstances may well put individuals in different strategic positions, particularly in a model with social behaviour. Here is a simple illustrative hypothetical example showing that social interactions are likely to require non-identical individuals. In some species of social Hymenoptera, relatives cooperate to build a nest, and there is a premium on completing the nest early enough (Queller, 1989). It is likely that the number of cooperators, whom we now for simplicity assume to be sisters, will affect the chance of success. Suppose a lone female has two daughters, but that she could instead increase her sister's reproductive output by two, from two to four daughters. Counting offspring and relatives, she would give up 2 and gain 0.75×2, so by simple inclusive fitness reasoning, this would be definitely disadvantageous to the strength of 0.5 offspring. But if four daughters cooperating have a 60% chance of successfully founding a nest while two have only a 10% chance, then we do better to count successful nests. Let's assume for simplicity that nests with different numbers of foundresses are equally successful once started in terms of grandoffspring for the parent. Now the lone female would give up 0.1, but gain $0.75\times(0.6-0.1)$, so helping would make a net profit of 0.275 nests worth of grandoffspring. Thus to capture this situation, giving a 3rd and 4th offspring to a sister would have to be different from giving a 1st and 2nd, or 5th and 6th. Recognizing classes of offspring, and assigning them reproductive values (Taylor, 1990, 1996; Grafen, forthcoming), would go some way towards this. But it is even more complicated, because the success of the sister's existing offspring is affected by the altruism, as they become part of a foursome rather than of a twosome. (To reconcile the situations of the mothers and the daughters, as parents, we would need to introduce bivoltinism, with the possibility for helping arising in only one of the generations.) Most biologists faced with this case would doubtless reach the same solution. The point here is the difficulty for a single general method in applying automatically to this and similar examples.

Thus there are significant complexities awaiting the relaxation of the assumption of universal strategic equivalence. We cannot make number of sibs part of the situation, as it is an essential element of the structure of the model that the population and its constitution are fixed and independent of the situation.

It is also worth noting that the relevant models of reproductive value (Taylor, 1990; Grafen, forthcoming) not only assume an infinite population, but also have no uncertainty in them. Thus while relaxing it seems quite feasible in due course, for the moment there is no option but to make the sometimes reasonable assumption of universal strategic equivalence.

6. How relatedness varies across the genome

It is vitally important in understanding the implications for natural selection of the results of earlier sections to consider how r^t varies across the genome. Previous sections establish an optimization program with a maximand for a particular p-family. If all *p*-scores share the same maximand, then natural selection is acting on all loci and traits in concert, and we can expect sophisticated adaptations to arise in pursuit of the single organism-wide maximand. On the other hand, if the *p*-scores have different maximands, we can ask 'what maximand will the *organism* appear to be maximizing, if any?'; and we should also expect intra-organismal conflict, as some alleles and traits are selected to oppose the changes that other alleles and traits are selected to promote. The importance of the maximand varying across the genome was first recognized by Hamilton (1967) in his paper on extraordinary sex ratios, and we can view p-families as different parties in the 'parliament of genes' of Leigh (1971).

There are two basic expectations about p-families. The larger a p-family, the more alleles belong to it, and the more phenotypic effects it has, the greater subtlety of adaptation will result in line with the corresponding optimization program. A small p-family consisting of the alleles at one locus must be considered extremely weak: on its own, it has very limited power, could easily be thwarted by many other loci, and is not likely to produce a biological adaptation. It takes many loci to construct complicated organs. Second, when p-families have different maximands, the larger and stronger the p-family, the more likely it is to win out over others, and in two ways. Opposing selection in two families will tend to drive the smaller or weaker set to genetic uniformity at relevant loci, waiting for mutations to arise; thus the larger or stronger p-family will tend to prevail. There will also be selection for one p-family to take measures to prevent other p-families being expressed at all, in general or in particular contexts.

While some families may be very different from each other, it is also possible for a set of rather similar families to operate mainly together. In principle, the difference in interests between p-families Q and R could be measured by the difference in maximands $\Phi^Q(\alpha,\rho,\omega)-\Phi^R(\alpha,\rho,\omega)$, and in particular cases by some average over this function.

The succeeding subsections look at different kinds of genetic conflict, linking it to the nature of the p-families involved.

6.1. Conflict based only on transmission pattern

The first property defining p-families is pattern of transmission, and the simplest case of conflict is between *p*-scores with different transmission patterns. In humans, the relatednesses between full sibs vary for the patterns associated with autosomes, X-linked, Y-linked and mitochondrial genes, as shown in Table 3. It seems likely that the resolution of this four-way tug-of-war is that humans as organisms have essentially the same maximand as the autosomes, and the reason is simply that there are so many more of them. This distinction between these groups of genes occurs even for non-social Darwinian fitness: for example, the w_j will vary for a male human depending on whether autosomal, X-chromosome or Y-chromosome genes are being discussed. If individual j is a male with m sons and f daughters, then $w_j=m+f$ for autosomes, $w_j=f$ for X-linked genes, and $w_j=m$ for Y-linked genes, while $w_j=0$ for mitochondrial genes. In non-social selection, these differences in pattern of transmission can affect sex ratio. Including social actions, they will differentially affect altruism towards male and female sibs, for example.

Also coming under the heading of 'transmission pattern' is genomic imprinting (Haig and Westoby, 1989; Haig, 1997), though the part of the pattern that matters here is not where the genes go from here, but how they got here, and where else their clone-mates are therefore likely to be found. To apply the models of previous sections to genomic imprinting, we need to interpret 'individual' in the model as applying to either the paternally derived alleles in one physical individual, or to the maternally derived alleles in one physical individual. Essentially, each physical individual counts as two. In addition, for the genomic imprinting to have interesting consequences, there must be roles that distinguish maternal from paternal relatives. Then the different relatednesses of paternally derived and maternally derived alleles to the different roles will result in

Table 3
Relatednesses to full sibs in humans varying with transmission pattern

Pattern	Sis-sis	Sis-bro	Bro-sis	Bro-bro
Autosomal	1/2	1/2	1/2	1/2
X-linked	3/4	1/4	1/2	1/2
Y-linked	u	u	0	1
Mitochondrial	1	1	1	1

Four different transmission patterns confer four different patterns of relatednesses in humans. The symbol 'u' denotes an undefined relatedness.

Table 4
Some relatednesses under genomic imprinting

Actor	Pat. cous.	Mat. cous.	Pat. half-sib	Mat. half-sib	Full-sib	Father	Mother
Paternal	$\frac{1}{4}$	0	$\frac{1}{2}$	0	$\frac{1}{2}$	1	0
Maternal	0	$\frac{1}{4}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0	1
Does not know	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

Relatednesses to seven classes of relative, of three different types of alleles within a single physical individual, at autosomal loci. Note that the 'does not know' allele is always the average of the other two, and that roles differentiating paternal from maternal relatives are necessary to create the difference. The sharpest discordance occurs for the closest kin.

intra-genomic conflict. Table 4 shows how relatednesses differ to various classes of relative from the standpoint of parentally derived alleles, maternally derived alleles, and from random alleles (those presumed not to know their origin, or more explicitly, those assumed to be expressed in the same way regardless of their origin).

These two kinds of conflict both use identity-by-descent as the basis for calculating relatednesses. They produce 'static' differences that are constant over time and through all of the genome that shares a given transmission pattern. Thus they can be expected to give rise to constant selection resulting in serious and sustained conflicts.

We have assumed outbreeding so far in this section. Inbreeding alters relatednesses by increasing them, but more fundamentally from the point of view of this section, it does so differently for different alleles, thus splitting a single p-family into many. This is further discussed by Grafen (1985).

6.2. Varying by matching

Rather different biological issues are raised by variation of r^t between loci with the *same* transmission pattern. Apart from chance, dealt with in Section 6.3, this form of variation in r^t across the genome depends on genetic matching of some kind. The analysis of this section is based on Grafen (1990). Green beard genes were originally proposed by Hamilton (1964) and given their vivid name by Dawkins (1976), and now have a large literature devoted to them (e.g. Keller and Ross, 1998; Summers and Crespi, 2005). They will serve to illustrate the important points. Take first the simple case in which one gene at a particular locus causes its bearer to produce a distinctive marker (the green beard) and to seek out another unrelated individual with a green beard and to perform some altruistic act towards them. It is understood that only individuals with the gene have a green beard. The other individual is supposed unrelated in the sense that only the possession of the green beard distinguishes them from a random member of the population. Letting t denote the role 'fellow green beard bearer selected from otherwise random members of the population', we will study how r^t varies for the p-score that counts how many green beard genes an individual has, and we will assume that gene is rare enough that no individual has more than one copy. At the relevant locus, actor and recipient are heterozygotes with $p_j = 1/2$. Assuming that the alternative allele is null, so that $b_{ijt} = 0$ except for green beard altruism, the measurement formula in Eq. (3) then gives $r^t = 1$. At unlinked loci, relatedness equals zero, on the assumption that green beard individuals pick each other out by cues that are uncorrelated with any other genetic property. The relatedness decays from one to zero on either side of the relevant locus according to the strength of linkage disequilibrium. In general, this would produce a very small region of positive relatedness in a chromosome with a zero relatedness everywhere else. Thus the autosomes in general would have zero relatedness apart from a very small region around the relevant locus. The great bulk of the genome would all belong to a single p-family, and so even if the green beard behaviour continued for lack of a mutant arising that sported the beard but was not altruistic towards beard-wearers, the main thrust of evolution would continue as though the matching did not occur.

If we alter our assumption, and suppose that the green beard individual is picked out from a group composed partly of relatives all with the same identity-by-descent relatedness r_{IBD} at the other autosomes and partly of unrelated individuals, so that the mean relatedness to group members is r_{AVE}, then the conclusion is changed in two ways. The baseline to which green beard relatedness falls on either side of the relevant locus is also r_{AVE}, but the speed at which r^t falls from 1 now depends on linkage, on the chance that two loci have been recombined in the ancestral paths linking the two relatives. Characteristically, linkage is much stronger than linkage disequilibrium, and the falling off will take much further to happen. Thus autosomes will have a relatedness of r_{AVE}, the coancestry value, apart from a moderate sized region around the relevant locus. The majority of the genome would still likely belong to the same p-family.

On the basis of these observations, we can conclude that we would not expect sophisticated adaptations to arise in support of green beard genes that assist otherwise random members of the population. It is more likely that adaptations would arise for green beard genes that assist relatives, as the region of elevated relatedness is greater. But it is even more likely that a matching mechanism would be used to pick out a green beard from a mixed group of relatives and non-relatives; and that the evolutionary support for this would arise from the elevation of genome-wide relatedness from r_{AVE} to close to r_{IBD} that results for the whole genome from picking out a relative rather than a random member of the group.

The other causes of variation can be understood in relation to these two green beard examples. Consider first kin recognition by matching loci. This provides a high relatedness at the loci actually matched, with a falling off according to linkage: more precisely by a mixture of linkage disequilibrium and linkage proper, with the quantities depending on the distribution of coancestry with the actor among the matched individuals. A main effect is to establish a higher background relatedness, closer to r_{IBD}: it is doubtful whether the loci can be sufficiently numerous, and the matching sufficiently strong, to make the higher r^t around the matching loci really make the effective relatedness exceed r_{IBD}.

Another possible mechanism is assortment by trait. Suppose there is a p-score and that individuals interact with others possessing the same value of the p-score. Essentially the same analysis applies. Selection of that one p-score will proceed with a relatedness of one. The relatedness applying to some other p-score will depend on the correlation between the two p-scores. If the selection of interactants is from a group of unrelated individuals, then it is likely that most p-scores are uncorrelated with the matching p-score, and so there is still one dominant p-family. If interactants are selected from a group with varying ancestral relatednesses, then individuals equal on the matching p-score will tend to be relatives, and so have a higher relatedness across the genome. Simultaneous matching on a number of p-scores would more effectively discriminate kin. The evolution of group-directed altruism based on assortment, first discussed by Hamilton (1975), can be a robust phenomenon only if the assortment succeeds in establishing a substantial relatedness across the genome as a whole.

Calculations can be done for a simple case, in which individuals assort exactly on a p-score based on alleles at n different unlinked loci in linkage equilibrium, and the allelic values at each locus are drawn from a Normal distribution with zero mean and standard deviation $1/\sqrt{n}$. The p-score itself then has a mean of zero and a standard deviation of 1. The relatedness at each of the n loci is

$$\frac{z^2}{n-1+z^2}$$

in a group whose matched p-score is z. In the 'typical' case of $z = \pm 1$, this reduces to $1/n$. Thus from a relatedness of 1 at a single locus, to a relatedness of 0.1 at ten loci, in some crude sense there is the same average 'push' towards higher relatedness. Matching at a more broadly based p-score does not therefore seem to increase overall genetic similarity. Matching simultaneously for a number of p-scores would increase it. Interestingly, the equation shows an advantage to making the altruism conditional not only on matching a p-score, but simultaneously on sharing an extreme value of that p-score, as indicated in an extreme value of z.

6.3. Relatedness varying by chance

Finally, we come to chance as a cause of variation of $r^{t,s}$, which will occur with measurement methods as shown by Eq. (7). At first sight it may seem to be a strength of measurement that it gives an exact expression for the realized Δp^s (apart from the randomness of Mendelian segregation) rather than just the more approximate value for the expected Δp^s provided by identity-by-descent. However, it is really a weakness. Δp^s may vary with s for all kinds of reasons, having nothing to do with genetic similarity. For example, the same assistance of providing an item of food may have markedly different consequences depending on how much other food is available to the recipient around that time. All these changes, from whatever cause, are 'explained' by variation in $r^{t,s}$, the measured relatedness coefficients. Thus in interpreting the $r^{t,s}$ as causal, we need to be cautious about other possibilities. Nevertheless, it is possible to use measured relatednesses to illuminate simulations by providing interpretation and insight into the selective process (e.g. Frank, 1994; Axelrod et al., 2004).

Recall that the Price equation and $r^{t,s}$ are defined with respect to a particular p-score. The estimates from Eq. (7) will sometimes be very 'noisy', for example if there are only a few individuals with non-zero b^s_{ijt}, or if the p-score represents a rare allele. The estimate can then respond strongly to chance effects, and even vary wildly. In some circumstances it may be useful to form a joint estimate of the $r^{t,s}$ across a number of alleles. Suppose q^k_i is the p-score for individual i representing the frequency of allele number k, belonging to a set K of alleles not necessarily at the same locus. Let q^k be the mean of q^k_i. Then a joint estimate for all the alleles in K can be formed by

$$N^{t,s}_k = \sum_i \sum_j n_j (q^k_j - q^k) b^s_{ijt},$$

$$D^{t,s}_k = \sum_i \sum_j n_j (q^k_i - q^k) b^s_{ijt},$$

$$r^{t,s} = \frac{\sum_k D^{t,s}_k N^{t,s}_k}{\sum_k D^{t,s}_k D^{t,s}_k}.$$

This has the useful properties that (i) it agrees with identity-by-descent in very large populations when the relevant assumptions are upheld (ii) it is defined except where actors are exactly representative at all alleles in K (most unlikely for large K except when $b_{ijt} = 0$ and there are no fitness effects anyway) (iii) it gives more weight to p-scores more correlated with social actions, and so emphasizes those p-scores whose calculation contains more 'signal' over those uncorrelated with social action whose p-scores contain more 'noise'. It is equivalent to estimating r^t from a single p-score that is maximally correlated with social action. If there is no matching and no selection going on, then for p-scores uncorrelated with the the central p-score, the Hamilton residual will be small and differ unsystematically from zero.

6.4. Conclusion

These considerations of varying $r^{t,s}$ over the genome allow a number of conclusions. The logic of inclusive fitness provides an anatomy of intra-genomic conflict, allowing optimization ideas to be used of p-families, including most importantly the case where different p-families within the same body are pulling in different directions. The size of p-families is important in how these conflicts are resolved, and also in how complex an adaptation can be expected.

Perhaps the most important conclusion is that although, as seen earlier, coancestry has no logically special place in the theory of inclusive fitness, it is nevertheless the most powerful likely cause of non-zero relatednesses that are consistent over large parts of the genome, and so likely to have significant evolutionary consequences. It is not impossible in theory that combinations of linkage, linkage disequilibrium and matching could raise relatednesses throughout the genome, but theoretical discussions putting forward the idea of assortment-based altruism, even the initial treatments by Hamilton (1964, 1975), have not taken seriously the issue of the selective pressures in the bulk of the genome.

It was Hamilton (1967) who first analysed what we now call intragenomic conflict in terms of different relatednesses, and it was further developed for genomic imprinting by Haig and Westoby (1989) and Haig (1997). Thus no originality is claimed for this section, which aims only to place intragenomic conflict in its true place as a theoretically immediate corollary of the logic of inclusive fitness.

7. Discussion

Many biology students are taught that natural selection leads to organisms acting as if maximizing their inclusive fitness, which for that reason alone becomes a key concept. The discussion is directed at implications of the derivation of previous sections for the concept of inclusive fitness. In Section 7.1 the current paper is situated in relation to some previous theoretical and exegetical work on inclusive fitness, and is in places technical. Section 7.2 avoids technical language, and considers the new ideas in the paper and the current status of inclusive fitness.

7.1. Relationship to some previous theory

The current paper is set in the context of the formal Darwinism project, and then its relationship discussed to previous theoretical and exegetical works.

The Formal Darwinism project aims to provide a full, explicit and rigorous justification for ideas of fitness optimization in terms of population genetics. After two preparatory papers (Grafen, 1999, 2000), the project's first link (Grafen, 2002) treated theory that covered Darwin's ideas in the *Origin of Species* in 1859. Grafen (forthcoming) brought the theory up to the *Descent of Man* in 1871 by permitting different classes of individual. Now this paper treats the only major extension of Darwinian theory: the inclusive fitness concept of Hamilton (1964). Within the topics so far, it would be useful to extend the classes result to finite populations, and the inclusive fitness result of the current paper to infinite populations and situations. Combining the results so far is also desirable. The assumption of discrete non-overlapping generations has so far been made throughout, and removing this assumption is an important goal. The analysis of ESS theory and sequential interactions has yet to be tackled at all.

This paper can also be compared to an earlier account of mine on inclusive fitness as an evolutionary tool (Grafen, 1985). The earlier work is longer and more expository on a number of issues including relatedness and the Price equation; but was restricted in that relatednesses were always 'measured' rather than 'modelled', there was no uncertainty, the concept of p-family was present but implicit, and there were no explicit optimization programs. The concept of 'role' is a development of the 'action category' of the earlier work, and at the same time a return to the original treatment of Hamilton (1964).

A substantial and important difference is that the earlier treatment considered only one kind of social action at a time, and derived Hamilton's Rule for it. The use of Hamilton's Rule as opposed to inclusive fitness was recommended on the grounds that it was simpler to apply. The present paper aims to show that inclusive fitness is optimized, and takes the view that the extra complications are after all important. The significance is that the earlier work leaves open the possibility that each action obeys Hamilton's Rule, but different actions have different relatednesses. The present paper shows that for all actions affected by a given p-family, the alleles in the p-family all pull in the same direction. To the extent that the individual is under the control of a single p-family, we can expect the organism to be well-designed, with all its alleles and organs playing their part in maximizing inclusive fitness. Thus the link between genetics and organismal design is made much more strongly if it can be shown that inclusive fitness tends to be optimized.

It is important to place the present work in the context of other basic theory on inclusive fitness and the natural selection of social behaviour. Taylor (1990, 1996) provides a mathematical description of inclusive fitness theory in relation to gene frequency change (as part of a wider project that includes class-structured populations). A methodology employed in a series of papers by Frank (e.g. Frank, 1994, 1995a,b) was partly formalized by Taylor and Frank (1996), Frank (1997) and Frank (1998) who offer general prescriptions, and many worked examples, of how to apply inclusive fitness theory in constructing useful models of biological problems. To this end, they relax assumptions such as additivity made in the present paper, and consider population structure. They go beyond Hamilton's technical results in making optimization of fitness a usable principle, avoiding the need to build a population genetic model. In this sense they are more applicable than the present paper. On the other hand, they do not aim to provide a proof connecting population genetics and optimization of inclusive fitness, and they make tacitly assumptions uncovered here, such as universal strategic equivalence. Thus, the present paper provides general background support, and more explicit logic, of use in asking more fundamental, abstract questions, such as 'when are organisms selected to maximize their inclusive fitness?'. The earlier papers are much more useful in constructing particular inclusive fitness models, answering 'what biological conclusions follow from the optimization of inclusive fitness?'. In fact, Frank (1998) and the other papers in this section of the literature mainly employ neighbour-modulated fitness, but with the emphasis on relatednesses this approach is very much a contribution in the original Hamilton tradition.

A further important paper is by Queller (1992), who derives a very general inclusive fitness result. Essentially, he extends the measuring of relatednesses and measures the benefits and costs too. This very elegant approach shares the difficulties pointed out for measured relatednesses in Section 6.3, but even more strongly. The generality is appealing and valuable, but is achieved at a cost: the exact gene frequency change will be predicted whatever values of fitnesses and gene frequencies are supplied, and whether or not brought about by natural selection on social behaviour. The introduction of a model of how fitness is determined, as in the current paper, restricts the generality but gives in some ways a more valuable result.

We now turn to an issue that is important for population genetics, namely dynamic sufficiency. Since at least Lewontin (1974) it has been a mark of respectable population genetic models that they are dynamically sufficient, which means in our context that given the array of genotypes present in one generation, the model constructs the complete array of genotypes in the next generation. In previous papers (Grafen, 1999, 2000, 2002), I have described the Price equation as dynamically insufficient, but I now view this as misleading. The *assumptions* made at the point of applying the Price equation are insufficient to predict the whole array of genotypes in the next generation. For it is assumed how many successful gametes each individual has, but not how those gametes are combined into offspring. Thus no equation could be dynamically sufficient in that case. The important point is that the argument as a whole is fully rigorous despite the dynamic insufficiency of the assumptions. The inaccuracies that dynamic sufficiency was introduced to prevent do not arise in this application. It is a positive benefit of the Price equation that it will produce conclusions about gene (not genotype) frequencies from those dynamically insufficient assumptions. The Fundamental Theorem of Natural Selection (Fisher, 1930) shares the same property.

7.2. Current status of inclusive fitness

The main lines of the conclusions from the current paper are by no means new to biology, as its main task is to formalize ideas introduced decades ago (e.g. Hamilton, 1964, 1970; Dawkins, 1976), which have become basic principles for many biologists. However, many misunderstandings have arisen among more mathematical biologists because the original presentations were insufficiently mathematically explicit. As well as aiming to undo those misunderstandings, the formal arguments have uncovered assumptions and details, and have clarified conclusions.

What was essentially shown by Hamilton (1964, 1970) was that genes correlated with inclusive fitness would increase in frequency. The major addition in the current paper has been to articulate the next step, to claiming that inclusive fitness will be maximized as a consequence. There have been two stumbling blocks. First, heterozygote advantage is a simple example in which inclusive fitness is not actually maximized. The resolution here is to find slightly more complicated expressions of optimality conditions, which apply universally, but in some circumstances will not lead to all individuals possessing an optimal phenotype. Essentially, gene frequencies always change in the direction of increased inclusive fitness, but genotype frequencies may not. The second stumbling block has been to interpret the principle as providing a Lyapunov or potential function of the dynamic state of the population, in other words that the optimization principle is about how genotype frequencies change through time. Instead, the biologically interesting conclusion is about how an individual's fitness would change if it changed its behaviour. Both of these difficulties vanish once optimization programs are used to make wholly explicit the conclusion that natural selection tends to result in individuals that act as if maximizing their inclusive fitness.

One detail to emerge from the current paper has been that, when dealing with uncertainty, the quantity individuals will act as if maximizing is the expected relative inclusive fitness effect, where, perhaps surprisingly, 'relative' means to the average neighbour-modulated fitness. In a positive sense, we could use this detail to construct examples where dividing by the average inclusive fitness would give the wrong answer. In the more important negative sense, this detail can reassure us that most of our understanding of how inclusive fitness works has not required amendment now the argument has been fully articulated.

An important assumption uncovered in the paper has been 'universal strategic equivalence', parallel to the 'pairwise exchangeability' of Grafen (2002). Both say roughly that to justify a strategic analysis, we have needed to assume that all individuals face the same strategic situation. The need for such an assumption makes sense, as explained in Section 5.1. A natural attempt to allow a diversity of strategic situations would somehow bring together strategic situations and the classes central to reproductive value theories, in a combination of this paper and Grafen (forthcoming). This illustrates that the links proved in this paper could, and should, be improved by further work.

Other future improvements could remove the assumption of discrete non-overlapping generations, permit frequency dependence and more complex interactions, and allow non-additivity of fitness effects. Mutation could also be added to the theory.

The concept of the 'Hamilton residual', along with the use of 'roles', allows us to apply inclusive fitness theory with relatednesses defined for some roles by modelling such as identity-by-descent, and for other roles by measurement. The inclusion of uncertainty allows a more sophisticated kind of optimality to be considered, in which averages are taken and suitable conditional behaviour required.

Two points can be made about the nature of the optimization demonstrated. First, inclusive fitness is a generalization of Darwinian fitness as, if we set $b_{ijt} = 0$ for all $t \neq e$, we get Darwinian fitness optimization, which follows because ProgIF has Darwinian fitness as maximand, and the Price equation has Darwinian fitness as target of selection. Second, the nature of the connections is quite intimate, including out-of-equilibrium connections in the presence of arbitrary genetic and phenotypic variability, and it is unlikely that other maximization principles proposed in biology would be able to achieve a similar level of support. Bet-hedging (Seger and Brockmann, 1987), with its maximization of geometric mean of absolute fitnesses, would be hard pressed to integrate as neatly. MacArthur's product theorem for sex ratios (e.g. Charnov, 1982) applies to populations rather than individuals. The analogy between genes and memes could be investigated further by constructing the parallel argument for memes, and considering the plausibility of the necessary assumptions. Finally, it is doubtful that group selection would in general produce optimization programs sharing the same maximand over a range of loci, and so be shown to be a creative evolutionary force: note also that the programs developed here have the individual and not the group as the optimizing agent.

It emerges that inclusive fitness is more of a conceptual construction than neighbour-modulated fitness. There is a choice of how to argue that the Hamilton residual equals zero at least on average, and different relatednesses would result from different choices. Thus different biologists could attribute different inclusive fitnesses to the same individual organisms in the field or in a model. On the other hand, even Darwinian fitness is an abstraction: Williams (1966) argued that fitness was a property of a design and should mean the average success a design would achieve, averaged over uncertainties and over years. Measuring lifetime reproductive success of individuals is only a stage towards calculating the success of a design, and by recognizing the design differently, or choosing different years or uncertainties, biologists could assign different Darwinian fitnesses to the same design. Inclusive fitness may have a larger measure of construction compared to Darwinian fitness, but this is a price to be paid for its greater utility.

The optimization principle has been a central feature of Hamilton's argument since the 1964 paper, although it is only here made fully explicit. It is clear that the precursors of inclusive fitness to be found in papers of Fisher (1914) and Haldane (1955) were so primitive as not to have an optimization principle behind them. Hamilton intended his writings on inclusive fitness to provide a comprehensive analytical tool for social behaviour (Hamilton, 2001, p. 135). The treatment here fully confirms that intention for simple additive social interactions, and with arguments that are wholly explicit in a mathematically rigorous way. Although there is still more work to do to extend the logical reach of the derivation presented here, it is clear that relatedness and inclusive fitness are essential and ubiquitously useful tools in understanding the social behaviour of organisms and the social interactions of genes.

Acknowledgements

I am very grateful to Prof. Steven Frank for very useful conversations and comments on earlier drafts. A number of valuable comments were made at the Behaviour, Culture and Evolution seminar at UCLA, especially by Prof. Robert Boyd; and at a seminar at the University of Exeter campus in Penryn, Cornwall, England. The two referees made very helpful comments. This research was generously supported by a Research Fellowship (0468) of the Leverhulme Trust.

References

Axelrod, R., Hammond, R.A., Grafen, A., 2004. Altruism via kin-selection strategies that rely on arbitrary tags with which they co-evolve. Evolution 58, 1833–1838.

Bennett, J.H. (Ed.), 1983. Natural Selection, Heredity and Eugenics (Including Selected Correspondence of R.A. Fisher with Leonard Darwin and others). Oxford University Press, Oxford.

Charnov, E.L., 1982. The Theory of Sex Allocation. Princeton University Press, Princeton.

Cosmides, L.M., Tooby, J., 1981. Cytoplasmic inheritance and intragenomic conflict. J. Theor. Biol. 89, 83–129.

Crozier, R.H., 1970. Coefficients of relationship and the identity of genes by descent in the hymenoptera. Am. Nat. 104, 216–217.

Darwin, C.R., 1859. The Origin of Species. John Murray, London.

Dawkins, R., 1976. The Selfish Gene. Oxford University Press, Oxford.

Day, T., Taylor, P.D., 1998. Unifying genetic and game theoretic models of kin selection for continuous traits. J. Theor. Biol. 194, 391–407.

Ewens, W.J., 1989. An interpretation and proof of the fundamental theorem of natural selection. Theor. Popul. Biol. 36, 167–180.

Ewens, W.J., 1992. An optimizing principle of natural selection in evolutionary population genetics. Theor. Popul. Biol. 42, 333–346.

Falconer, D.S., 1981. Introduction to Quantitative Genetics, second ed. Longman, London.

Fisher, R.A., 1914. Some hopes of a eugenist. Eugen. Rev. 5, 309–315.

Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Oxford University Press OUP published in 1999 a variorum edition of the 1930 and 1958 editions.

Fisher, R.A., 1947. The theory of linkage in polysomic inheritance. Philos. Trans. Roy. Soc. London Ser. B 233, 55–87.

Fisher, R.A., 1949. The Theory of Inbreeding. Oliver and Boyd, Edinburgh.

Frank, S.A., 1994. Kin selection and virulence in the evolution of protocells and parasites. Proc. Roy. Soc. London Ser. B 258, 153–161.

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

Frank, S.A., 1995b. Sex allocation in solitary bees and wasps. Nature 146, 316–323.

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

Frank, S.A., 1998. The Foundations of Social Evolution. Princeton University Press, Princeton.

Grafen, A., 1979. The hawk-dove game played between relatives. Anim. Behav. 27, 905–907.

Grafen, A., 1984. Natural selection, kin selection and group selection. In: Krebs, J.R., Davies, N.B. (Eds.), Behavioural Ecology, second ed. Blackwell Scientific Publications, Oxford, UK, pp. 62–84.

Grafen, A., 1985. A geometric view of relatedness. Oxford Surveys Evol. Biol. 2, 28–89.

Grafen, A., 1990. Do animals really recognize kin? Anim. Behav. 39, 42–54.

Grafen, A., 1998. Fertility and labour supply in Femina economica. J. Theor. Biol. 194, 429–455.

Grafen, A., 1999. Formal Darwinism, the individual-as-maximising-agent analogy, and bet-hedging. Proc. Roy. Soc. Ser. B 266, 799–803.

Grafen, A., 2000. Developments of Price's Equation and natural selection under uncertainty. Proc. Roy. Soc. Ser. B 267, 1223–1227.

Grafen, A., 2002. A first formal link between the Price equation and an optimization program. J. Theor. Biol. 217, 75–91.

Grafen, A., 2003. Fisher the evolutionary biologist. J. Roy. Statist. Soc. Ser. D (The Stat.) 52, 319–329.

Grafen, A. (forthcoming). A theory of Fisher's reproductive value.

Haig, D., 1997. Parental antagonism, relatedness asymmetries, and genomic imprinting. Proc. Roy. Soc. Ser. B 264, 1657–1662.

Haig, D., Westoby, M., 1989. Parent-specific gene expression and the triploid endosperm. Am. Nat. 134, 147–155.

Haldane, J.B.S., 1955. Population genetics. New Biol. 18, 34–51.

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

Hamilton, W.D., 1964. The genetical evolution of social behaviour. J. Theor. Biol. 7, 1–52.

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., Dillon, W. (Eds.), Man and Beast: Comparative Social Behavior. Smithsonian Press, Washington, DC, pp. 57–91.

Hamilton, W.D., 1972. Altruism and related phenomena, mainly in the social insects. Ann. Rev. Ecol. Syst. 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., 2001. Narrow Roads of Gene Land. Volume 2: Evolution of Sex. Oxford University Press, Oxford.

Hammerstein, P., 1996. Darwinian adaptation, population-genetics and the streetcar theory of evolution. J. Math. Biol. 34, 511–532.

Hines, W., Maynard Smith, 1979. Games between relatives. J. Theor. Biol. 79, 19–30.

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

Leigh, E.G., 1971. Adaptation and Diversity. Freeman Cooper, San Francisco, CA.

Lessard, S., 1997. Fisher's fundamental theorem of natural selection revisited. Theor. Popul. Biol. 52, 119–136.

Lewontin, R.C., 1974. The Genetic Basis of Evolutionary Change. Columbia University Press, New York.

Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic Theory. Oxford University Press, Oxford.

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.

Price, G.R., 1970. Selection and covariance. Nature 227, 520–521.

Price, G.R., 1972. Extension of covariance selection mathematics. Ann. Hum. Genet. 35, 485–490.

Queller, D.C., 1989. The evolution of eusociality: Reproductive head starts of workers. Proc. Natl Acad. Sci. (USA) 260, 3224–3226.

Queller, D.C., 1992. A general model for kin selection. Evolution 46, 376–380.

Seger, J., 1981. Kinship and covariance. J. Theor. Biol. 91, 191–213.

Seger, J., Brockmann, H.J., 1987. What is bet-hedging? Oxford Surv. Evol. Biol. 4, 182–211.

Summers, K., Crespi, B., 2005. Cadherins in maternal-foetal interactions: red queen with a green beard? Proc. Roy. Soc. Ser. B 272, 643–649.

Taylor, P.D., 1990. Allele-frequency change in a class-structured population. Am. Nat. 135, 95–106.

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.

Williams, G.C., 1966. Adaptation and Natural Selection. Princeton University Press, Princeton, NJ.

Wright, S., 1969–1978. Evolution and the Genetics of Populations. University of Chicago Press, Chicago.