REVIEW ARTICLE

The simplest formal argument for fitness optimization

ALAN GRAFEN*

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

Abstract

The Formal Darwinism Project aims to provide a formal argument linking population genetics to fitness optimization, which of necessity includes defining fitness. This bridges the gulf between those biologists who assume that natural selection leads to something close to fitness optimization and those biologists who believe on theoretical grounds that there is no sense of fitness that can usefully be said to be optimized. The current paper's main objective is to provide a careful mathematical introduction to the project, and it also reflects on the project's scope and limitations. The central argument is the proof of close ties between the mathematics of motion, as embodied in the Price equation, and the mathematics of optimization, as represented by optimization programmes. To make these links, a general and abstract model linking genotype, phenotype and number of successful gametes is assumed. The project has begun with simple dynamic models and simple linking models, and its progress will involve more realistic versions of them. The versions given here are fully mathematically rigorous, but elementary enough to serve as an introduction.

[Grafen A. 2008 The simplest formal argument for fitness optimization *J. Genet.* **87**, 421–433]

Introduction

The concept of fitness optimization has a sometimes difficult history, not least because the concept of fitness has been found hard to define. Empirical biologists in many fields have routinely assumed since the 1970s that natural selection leads organisms to act as if (more or less) maximizing a quantity often called fitness, intended to be roughly the lifetime number of offspring, and base research projects on that foundation. Theoretical opinion has varied over time and between experts, since Fisher (1930) published his *'Fundamental theorem of natural selection'*, intended to exhibit an optimizing tendency of natural selection. But it would be fair to say that, according to most technical mathematical opinion since the second edition of Fisher's (1958) book, there is no concept of fitness under which fitness is optimized by natural selection in any useful sense; Ewens (2004) provides one authoritative viewpoint from that side.

The shadow of this technical question falls over a large and important literature. The concepts of adaptation (Williams 1966), of 'vehicle' (Dawkins 1982) and 'levels of selection' (for a nonpartisan discussion see chapter 7 of Segestrale 2000) all depend, implicitly or explicitly, on the idea that selection produces outcomes that are 'for the good of' some entity. The most natural formal structure to underpin this idea is that selection leads towards optimization. Thus the question of fitness optimization is of very wide significance in modern evolutionary biology.

I have previously argued that the disagreement has arisen because of different understandings of what is meant by fitness optimization, and shown that, once the concepts of fitness and fitness optimization are properly formalized, there is a very strong sense in which we can say that natural selection leads towards fitness optimization. There are five technical papers in this 'Formal Darwinism Project' (Grafen 1999, 2000, 2002, 2006a,b), as well as a nonmathematical exposition (Grafen 2007b). The first two applications (Grafen 2007a,c) show that fitness optimization provides a rigorous and biologically meaningful interpretation of previous models. While optimization theory and ESS theory (Maynard Smith and Price 1973; Maynard Smith 1982) have allowed biologists to model the consequences of assuming optimality, they have not satisfactorily dealt with the question of justifying optimality in the first place.

The purpose of the current paper is to fill a gap in the presentation of the project, which arises because the existing

*E-mail: alan.grafen@sjc.ox.ac.uk.

Keywords. Formal Darwinism Project; fitness optimization; Price equation; uncertainty; dynamic insufficiency.

papers are mostly highly technical, making them difficult to approach. The current paper starts formally but at the beginning, providing a 'launch ramp' for readers with ambitions on the other papers. The opportunity is also taken to pursue some general points of interpretation that arise more naturally in this more generic setting.

The next section presents a derivation of the Price equation, and a simple optimization programme, and argues that they represent and formalize the main ways of thinking about natural selection, even though they are conceptually very distinct. Then, the paper goes on to show that the simplest mathematical link between dynamic systems and optimization, namely the idea that mean fitness is a function of the dynamical state and is nondecreasing over time, is not a faithful mathematical representation of the concept of fitness optimization, even in those special cases where this property holds. Progress towards an appropriate representation requires a model linking genotype, phenotype and number of successful gametes, which is presented in the following section, but it is a very formal and general model. With this model established, links are then presented that can be established between these two ways of formalizing natural selection. They form the prototype of the central structure of the Formal Darwinism Project, but their biological significance is limited, and their meaning is fairly obvious. Therefore, the paper goes on to derive an extension to the case of uncertainy, from which some important conclusions can be drawn, demonstrating how the simplest scheme can be extended.

The Price equation and its consequences often have a rather bewilderingly high degree of generality, that can generate a sense of unease about its usefulness. The discussion begins by looking at the links proved, their meaning and their limitations. This is necessary, as this kind of argument is new in biology, and important because the results of the Formal Darwinism Project all share certain core features. The discussion goes on to review briefly the already published extensions of the basic argument, and to argue that the analytical possibilities fully justify working with gene frequency equations that are dynamically insufficient, indeed that it is necessary to do so.

The Price equation from scratch and an optimization programme

This section presents two separate formal objects. The Price equation (Price 1970; note that we do not use the more general formulation of Price 1972a) represents the dynamics of changing gene frequencies, and the optimization programme represents design. The fundamental strategy of the Formal Darwinism Project is linking two such formal objects.

The Price equation takes many different forms, and has been rederived many times with different details and assumptions (Price 1970, 1972a; Hamilton 1975; Grafen 1985; Queller 1992; Frank 1998). It has a beguiling generality. Here another derivation is given, suitable for the task at hand. Consider a population evolving over time, and two census points. For simplicity, we assume that all the individuals at the second census point are offspring of the individuals at the first census point, and that all individuals have the same ploidy. Suppose I is the finite set of individuals at the first census point, and we introduce three functions from I. p_i is the frequency within i of a given gene at its locus, and lies between 0 and 1. w_i is an integer represening the total ploidy of the gametes of individual i that contributed to offspring at the second census point (the 'successful gametes'). The gene frequency among the successful gametes of i is given by $p_i+\Delta p_i$, defining Δp_i as the discrepancy between the gene frequency in the gametes and in the parent.

Standard Pricean notation drops the subscript to denote the population average, so $p = \mathbb{E}[p_i]$ and $w = \mathbb{E}[w_i]$. We introduce $\mathbb{E}$ as the operator connoting the arithmetic average over the population, and $\mathbb{C}$ as the corresponding covariance operator. Let the gene frequency among the offspring be p' and let $\Delta p = p' - p$ be the change in gene frequency from adults to offspring.

Then, the gene frequency among the offspring is, by definition,

$$p' = \frac{\sum_i (p_i + \Delta p_i) w_i}{\sum_i w_i},$$

which we rearrange by subtracting p from both sides, obtaining

$$p' - p = \frac{\sum_i (p_i - p + \Delta p_i) w_i}{\sum_i w_i},$$

$$\frac{(\sum_i w_i)(p' - p)}{\sum_i 1} = \frac{\sum_i (p_i - p + \Delta p_i) w_i}{\sum_i 1}$$

$$w\Delta p = \frac{\sum_i (p_i - p) w_i}{\sum_i 1} + \frac{\sum_i w_i \Delta p_i}{\sum_i 1}$$

$$w\Delta p = \mathbb{C}[p_i, w_i] + \mathbb{E}[w_i \Delta p_i] \quad (1)$$

$$\Delta p = \mathbb{C}\left[p_i, \frac{w_i}{w}\right] + \mathbb{E}\left[\frac{w_i}{w}\Delta p_i\right] \quad (2)$$

The penultimate form is more usually quoted, but the final form turns out to have a particular conceptual significance (Grafen 2000). In it, the Δp can be added over generations to obtain the summed change in gene frequency and, wherever it appears, the measure of reproductive success w_i is relative to the mean value w.

Both forms of the Price equation are linear in gene frequency, and hold true for every gene frequency, as pointed out initially by Price (1970). It follows that the self-same equation applies if we replace gene frequency throughout with an arbitrary linear combination of gene frequencies. This is especially important, because it allows the Price equation to represent quantitative genetics models, and because the additive genetic component of any trait is a linear combination of gene frequencies. The term 'p-score' will be used to refer to a linear combination of gene frequencies (Grafen 1985, 2002), and the possibility that the p-score is not just the frequency of one gene will sometimes matter. Note that the

Price equation holds for all p-scores, provided all the genes involved have the same pattern of inheritance — the biologically significant issues arising when genes have different patterns of inheritance go under the name of intragenomic conflict (Burt and Trivers 2005), and are discussed in relation to the Formal Darwinism Project by Grafen (2006a).

The Price equation (1 and 2) applies as a meta-model to many population genetic models, as well as, more directly, a model of a population. Unlike most population genetic models, the index is over individuals, and the p-score and total ploidy of successful gametes are notated on an individual basis: this is a very significant feature underlying the links to fitness optimization. The second term on the right hand side will often be negligible under reasonable assumptions, as we shall see in later sections.

The Price equation gives only the change in mean p-score, and does not provide the whole array of genotypes in the next generation. This dynamic insufficiency, and the remarkable fact that so much can nevertheless be concluded from the equation, are considered in the 'Discussion'. For our purposes, the Price equation will represent the dynamic approach to p-scores, and it will be used to determine conditions under which p-scores increase, decrease, and stay the same.

Now we move on to the optimization programme, and set aside all thoughts of genes and populations. Suppose an individual's phenotype ϕ is drawn from some set $\mathbf{\Phi}$, representing physiological, physical and informational constraints. Suppose that its 'fitness' is some function $f(\phi)$ that shows how successful the phenotype is. Then we represent the idea of fitness optimization with the simple programme:

$$\begin{aligned} &\phi \max f(\phi), \\ &\phi \in \mathbf{\Phi}. \end{aligned} \tag{3}$$

These programmes have on occasion been explicitly notated in biology, but are more frequent in economics and operations research. Note that there is no sense of generations, or of genes. Indeed there is no population, and so for many reasons there is no sense of average gene frequency or average p-score.

The programme sets up a problem: what is the element ϕ of the set $\mathbf{\Phi}$ that achieves the highest value of $f(\phi)$? One advantage of elaborating the programme and actually writing it out is that we can discuss the problem without claiming that the population solves it. Teasing apart the existence of the problem from whether the maximum is attained in a population gives us an analytical distance, and allows us a language in which to explore the question of when the maximum is attained.

Optimization thinking is pervasive in biology even though optimization programmes are rare. Whenever a feature of an organism is explained as 'for' some purpose, there is at least an implicit appeal to optimization ideas. Classical morphology with its 'form and function' is therefore as much involved here as optimal foraging. Most whole-organism behavioural biology is founded on an assumption of this kind and, as mentioned in the Introduction, it underlies the concepts of adaptation and levels of selection.

In the dynamical approach, p-scores are tracked. In the optimization approach, form is linked to function. A formal justification for the optimization approach must derive from considerations of gene frequency, and that is why the Formal Darwinism Project takes as its business the construction of mathematical links between the Price equation and optimization programmes. The central task is to say what quantity derived from the dynamical approach can play the role of the maximand in the optimization approach, and so be reasonably viewed as 'fitness'.

The fundamental theorem and a false trail

Fisher (1930) presented his 'Fundamental Theorem of Natural Selection', and gave a verbal version as "the rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time". In this section, for brevity, we will use 'fitness' as Fisher did to mean number of offspring, temporarily suspending the very careful usage elsewhere in the current paper that restricts fitness to the maximand of an optimization programme. This is not wholly illogical, as the fundamental theorem represents the first attempt to link dynamics and optimization.

The change in understanding of the fundamental theorem by mathematical population geneticists is strikingly shown by the change between two editions of Ewens' textbook *'Mathematical population genetics'*. Ewens (1979) takes the theorem to state that mean fitness will never decrease, and denies its truth. Ewens (2004) distinguishes between the 'mean fitness increase theorem', abbreviated to MFIT, which is how the fundamental theorem had been understood, and the fundamental theorem itself, abbreviated to FTNS. The MFIT is still untrue except in very special circumstances, roughly additivity of all allelic effects. Meanwhile the FTNS is acknowledged to be true under very general assumptions, including multiple loci and alleles, with arbitrary mating systems, and with arbitrary linkage and linkage disequilibrium. In short, it holds under conditions remarkably similar to those of the Price equation.

If the FTNS does not assert that mean fitness never decreases, does it have an optimization interpretation? Fisher certainly thought that it did, but this idea was not pursued initially because the whole theorem fell into disrepute (for historical explanations see Edwards (1994); Grafen (2003)). The papers that later clarified and explained the newly understood theorem did not pursue the optimizing line of thought (Price 1972b; Ewens 1989, 1992; Edwards 1994). The Formal Darwinism Project aims at extensions of the FTNS in two ways: first, the population genetic assumptions are relaxed, and second and crucially, the link to optimization is rendered absolutely explicit.

This section has so far pointed out that the common interpretation held by mathematical population geneticists of fitness optimization has been that mean fitness is a function of the dynamical state that never decreases, which is true only in very restricted circumstances. We now go on to show that it was anyway never a good interpretation of the principle of fitness optimization.

Suppose the dynamical system representing p-scores has a mean fitness function m — how should we represent it in an optimization programme? We need the instrument to be the state of the dynamic system, say x, and the constraint set to be some neighbourhood of x, say $N(x)$, which could in some circumstances be the whole space. Then we introduce the mean fitness program:

$$\begin{aligned} &x \max m(x), \\ &x \in N(x). \end{aligned} \tag{4}$$

The result we could hope to prove would be of the form 'the dynamical system approaches a solution x^* of the mean fitness program (4) for some $N(x^*)$'.

However, it is already clear that this optimization bears no relationship to the biologists' concept of fitness optimization as represented in the 'simple programme' (3). The instrument for biologists is some aspect of the phenotype of the individual, such as height, sphericality of eye, or probability of mating with a male of a given tail length. The value of the instrument may vary among individuals, and there may or may not be genetic variation in it at any given time. The state of genotype frequencies is the instrument of the mean fitness programme, which does not represent possible phenotypes of an individual organism. Moreover, it is a population attribute, which does not meaningfully vary between individuals. The constraint set for biologists is a set of phenotypes that is physiologically, physically and informationally determined—what can an animal do? What can a mutation produce? What events can an animal's behaviour be conditioned on? The mean fitness programme constraint set is a set of possible genotype frequencies. Therefore, the mean fitness programme does not represent what biologists mean by fitness optimization.

There has therefore been a misunderstanding over fitness optimization, which is a key element in the difference of approach between mathematical population geneticists and organismally oriented biologists, as discussed by Schwartz (2002). It is certainly tempting to see its origin in the brief and overly-simple verbal statement of the fundamental theorem provided by Fisher (1930). I have argued elsewhere (Grafen 2007b) that the responsibility for continuing this misunderstanding lies on all sides of the argument.

The literature, discussed by Ewens (2004), in which it is shown that mean fitness does sometimes decrease is therefore not relevant to the question of whether fitness optimization as understood by biologists occurs. The formal Darwinism project offers a different and more elaborate approach that does satisfactorily formalize the biologist's sense of fitness optimization.

A model linking genotype, phenotype and number of successful gametes

The Price equation as it stands makes very widely-true statements about the operation of selection, but these statements on their own have the arid character of accounting identities. It takes the w_i, as well as the p_i and the Δp_i, simply as given. In order to make progress towards linking the Price equation to optimization programmes involving phenotypes, we now develop a very general model that applies to all p-scores, and includes phenotypes. Let the set of possible genotypes be $\mathbf{\Gamma}$, and we assume that an element $\gamma \in \mathbf{\Gamma}$ includes information about all loci and all alleles of an individual. Let the set of possible phenotypes be $\mathbf{\Phi}$, as in the optimization programme, and assume that an element ϕ specifies all aspects of an individual's phenotype. In a model without uncertainty, and without social behaviour, we can conclude that the number of successful gametes depends only on the phenotype, and the phenotype, in turn, depends only on the genotype. Let there be functions $\nu : \mathbf{\Gamma} \to \mathbf{\Phi}$ and $\omega : \mathbf{\Phi} \to \mathbb{N}$, so that $w_i = w(\phi_i) = \omega(\nu(\gamma_i))$ shows formally how the number of gametes depends on the phenotype and then genotype. We say nothing at this stage about how p_i relates either to γ_i or to w_i, although clearly p_i must be a function of γ_i. We take the natural numbers $\mathbb{N}$ to include zero.

We have taken all the simplest choices in order to work with a very spare model for illustrative purposes. Spelling out this model is important because it shows where different choices need to be made in more complex cases.

Notice that the population may be growing or shrinking in size. The biological force of density dependence depends on interactions between individuals, and these are more complicated than the model here allows. While in a dynamic model, it is useful to introduce bland density-dependence into a model to artificially create an equilibrium point, there is no purpose here in a similar manoeuvre. The results to be proved are all true in the growing or shrinking, or stable, population.

The simplest possible links

The model of the previous section links the dynamic and optimization approaches sufficiently that it is possible to prove links between them. We now use elaborate notation and multi-step arguments to prove some very obvious results. The value is that by articulating our reasoning in this way, we see which points need altering and expanding to deal with more complex cases, as shown later in the extension to uncertainty.

The first step is to construct all the elements of the optimization programme from the underlying population genetic model. The set $\mathbf{\Phi}$ is already present in the underlying model as the set of phenotypes, and we need only to construct the

function f, the candidate for fitness. It is clear that we need a function from $\mathbf{\Phi}$ to the real line, and so we choose simply that $f = \omega$, reflecting the extreme simplicity of the current assumptions. Thus we assume that 'fitness' is the number of successful gametes.

With all the elements of the optimization programme defined in dynamic terms, we now define two conditions that will be used in the links. ESS models (Maynard Smith and Price 1973; Maynard Smith 1982) invoke populations in which nearly all individuals play one candidate strategy, while the small residue plays some alternative strategy. The candidate is indeed an ESS only if every alternative strategy would decline in frequency. Formal Darwinism models also involve testing alternative strategies, but must do so differently, as their structure involves genotypes as well as phenotypes, and the proofs apply to an arbitrary given population, which is not necessarily genetically or strategically uniform.

First, we will say that there is 'no scope for selection' if every p-score has a zero change, i.e., $\Delta p = 0$ for all possible p-scores. Second, we will say that there is 'no potential for selection' in relation to a specified set of phenotypes, $\mathbf{\Phi}$, when certain conditions are met. To explain those conditions, we construct a modified population as follows. We take one individual, say with index h, and consider the consequences of altering her strategy from ϕ_h to $\alpha \in \Phi$. We let δ_i represent the difference made to w_i by this strategy alteration and under our assumptions we have $\delta_i = 0$ for $i \neq h$ and $\delta_h = \omega(\alpha) - \omega(\phi_h)$. Let p_i^h be a p-score that equals 0 for $i \neq h$ and equals one for $i = h$. Selection on this p-score would proceed as follows, in the modified population:

$$\begin{aligned}\Delta p^h &= \mathbb{C}\left[p_i^h, \omega(\phi_i) + \delta_i\right] \\ &= \mathbb{C}\left[p_i^h, \omega(\phi_i)\right] + \mathbb{C}\left[p_i^h, \delta_i\right],\end{aligned}$$

and on the assumption that all members of the population have equal $\omega(\phi_i)$ this leads to

$$\Delta p^h = \mathbb{C}\left[p_i^h, \delta_i\right] = \frac{n-1}{n^2}\delta_h.$$

If all p^h are not to increase in frequency, it follows that,

$$\delta_h = \omega(\alpha) - \omega(\phi_h) \leq 0 \text{ for all } h \text{ and all } \alpha \in \Phi. \qquad (5)$$

This is the condition for 'no potential for selection in relation to the set $\mathbf{\Phi}$' (i.e., no possible mutant would spread) under the assumption that there is already 'no scope for selection' (i.e. all the extant genotypes have equal numbers of offspring).

Let us first prove: *If every individual solves the optimization programme with constraint set $\mathbf{\Phi}$, and there is independent fair meiosis and no gametic selection, then there is no scope for selection, and no potential for selection in relation to $\mathbf{\Phi}$.* If the individuals all solve the programme then the $f(\phi_i)$ must all be equal, for otherwise every nonhighest value would not be a solution. But $f(\phi_i) = \omega(\phi_i)$ and so that w_i are all equal, giving us $\mathbb{C}[p_i, w_i] = 0$. The second term on the RHS of equation (1) will be zero under the assumptions—but only roughly, and we tidy this aspect up later. Accepting this minor sweeping under the carpet, we can conclude that $\Delta p = 0$ for all p-scores (note, we did not specify which we were dealing with), and so there is no change in the mean of any p-score. Our optimization assumption gives us that $f(\alpha) \leq f(\phi_i)$ for all i and all $\alpha \in \mathbf{\Phi}$. As we have identified the functions f and ω, this gives us $\omega(\alpha) - \omega(\phi_h) \leq 0$, as required to show by equation (5) that there is no potential selection in relation to $\mathbf{\Phi}$, thus completing the proof.

Before moving on to the next result, note that the first result has been proved for any p-score. Indeed, more than this, it applies to any function p from I to the unit interval, or, more biologically, to any possible p-score. This means that if we arbitrarily assigned a genotype at a hypothetical locus to every individual, the equations would show that there was no selection at that locus either. The significance is that the next result assumes that 'every possible p-score is not changing', and 'possible' is in the sense just explained.

Let us now prove the result in the opposite direction: *if every possible p-score is not changing, and there is no potential for selection for any phenotype in $\mathbf{\Phi}$, and if there is independent fair meiosis and no gametic selection, then all individuals solve the optimization programme.* If every possible p-score is not changing, recalling that independent fair meiosis and absence of gametic selection imply the second term in the Price equation is zero (on the same approximation as used in the first result), then the covariance term $\mathbb{C}[p_i, w_i]$ must be zero for all functions $p : I \to \mathbb{R}$. This is possible only if $w_i = w_j$ for all i, j. Hence, $f(\phi_i) = f(\phi_j)$ for all i, j. We also assume no potential for selection which gives us from (5) and the choice of $f = \omega$ that,

$$f(\alpha) \leq f(\phi_i) \text{ for all } \alpha \in \mathbf{\Phi} \text{ and all } i.$$

This shows that all the ϕ_i solve the simple programme (3), as required.

The first result takes a statement about optimization and derives a conclusion about dynamics, while the second takes a statement about dynamics, and draws a conclusion about optimization. Neither statement seems very surprising and the proofs mainly just track notation and have little substance. If this were all the approach could do, there would be little value in it. However, as we see later, by building on this basic pattern, we can both tidy up the hand-waving part about the second right-hand term in the Price equation, and derive nonobvious results.

The more formal papers (Grafen 2002, 2006a,b) have further types of link between dynamics and optimization beyond the two types proved in this section, which look at how selection operates out of equilibrium. It would add to the length and complication here, without adding much to the introductory value, to include those extra results.

Extending to uncertainty

Extension to uncertainty is undertaken in this section, both for its own sake, and as an illustration of how the very careful arguments of the previous sections can be used to substantial effect. The development here is only one part of the argument in Grafen (2002), and even the uncertainty is incorporated in a simpler way.

The first subsection considers and rejects the tempting possibility of making an informal extension to uncertainty, simply by declaring that w_i means the expected number of successful gametes of individual i.

As some of the previous notation needs to be amended, as well as new notation added, it is clearest to rub the slate clean and introduce the whole notation from scratch. Succeeding subsections cover the Price equation under uncertainty; the model linking genotype, phenotype, uncertainty and number of successful gametes; the optimization programme, and the definition of its elements in dynamic terms; and the links between the dynamic and optimization formalisms.

Difficulties of informal generalization

This section extends the model to uncertainty, but one natural reaction is to ask whether we cannot simply reinterpret the terms of the preceding section, and say 'w_i measures the expected number of successful gametes of individual i, and everything else then follows as shown'.

What can go wrong if we just interpret w_i as the expected number of successful gametes of individual i? The more formal approach requires a whole apparatus of uncertainty to make sense of this assertion, and it is work to do so. It would presume heavily upon the outcome of that work to suppose it will result in simply being able to replace w_i with its average, not to mention presuming what kind of average that should be. Further, we would not know what latent assumptions had been acquired in the process, which would lead to a fundamental unsettledness in the whole argument from that point on. Indeed, we will see that the formal approach introduces some unexpected results, but with great generality. Also, the derivation of the Price equation given earlier makes no sense in the presence of uncertainty, so we would be working with an equation without an adequate justification.

At the lowest level, this approach is the reason that w_i is asserted to take integer values. At the highest level, it is especially important in a modelling exercise with ambitions of generality and abstraction, and which produces unfamiliar results, to work with 'tight models'. The dynamic insufficiency considered in the Discussion is another reason to leave no logical lacunae.

The Price equation under uncertainty

Accepting the need for a formal treatment of uncertainty, we develop notation afresh. It is similar to the certain case, but some symbols have an extra argument because they depend on uncertainty. The basic methodology of handling uncertainty is the 'states of nature' approach. We decide which quantities should be allowed to vary, and which should not. Then we define a set of states of nature, the idea being that beforehand, there is a probability distribution over which state will turn out to be the actual state. The varying quantities will be a function of the state of nature, as well as of their existing arguments. This is a common approach in economics.

The formal definitions are as follows:

(i) A finite population I is censussed at two points in time, and we assume that all the individuals at the second census are offspring of the individuals present at the first census.

(ii) Assume a function $p : I \to \mathbb{R}$ such that p_i is the p-score of individual i (a gene frequency or arbitrary linear combination of gene frequencies), and p is the mean value of p_i.

(iii) A finite set $\mathbf{S}$ of states of nature is defined, with a probability distribution τ such that τ^s is the probability that s occurs. We assume $\sum_s \tau^s = 1$.

(iv) We assume there is a function $w : \mathbf{S} \times \mathbf{I} \to \mathbb{N}$, such that w_i^s is the number of successful gametes that contributed to the individuals at the second census point that derive from individual i at the first census point, and w^s is the mean value of w_i^s, all in state of nature s.

(v) Let $(p')_i^s$ be the p-score of the successful gametes of individual i in state of nature s, and let $\Delta p_i^s = (p')_i^s$ be the discrepancy for individual i between the p-scores of her successful gametes and of herself.

(vi) Let $(p')^s$ be the average p-score at the second census point in state of nature s, and let $\Delta p^s = (p')^s - p$ denote the change in mean p-score in state of nature s.

(vii) We introduce two expectation operators. $\mathbb{E}$ takes the expectation over the population I, weighting each individual equally, and $\mathbb{C}$ represents the corresponding covariance. $\mathcal{E}$ takes the expectation over states of nature, weighting each with its probability τ^s. As all sets are finite, these two expectation operators can happily coexist and also commute.

These formal assumptions mean that an individual's p-score is fixed, but that her reproductive success and allocation of alleles to gametes depends on chance. A state of nature therefore represents the weather and other chance events at the macroscopic level, as well as the intracellular events at meiosis and fertilization. For any given state of nature s, we can repeat the argument from the certain case to obtain the Price equation in one state of nature,

$$\Delta p^s = \mathbb{C}\left[p_i, \frac{w_i^s}{w^s}\right] + \mathbb{E}\left[\frac{w_i^s}{w^s}\Delta p_i^s\right]. \qquad (6)$$

We now present a formal assumption, whose interpretation is that the Δp_i^s differ from zero only through fair and independent Mendelian segregations. Actually, the model applies to haploids and triploids, as well as diploids, so the assumption represents a fairly obvious extension of fair and independent Mendelian segregation. Notice that stating the assumption is only possible now that uncertainty has been included in our formal framework.

The assumption is stated in a strong form corresponding to what is biologically reasonable. We will also note a weakened form that will suffice for all the requirements of the present paper.

Assumption of unbiassed transmission: The diploid version is that the discrepancy between the gene frequency of an individual and the gene frequency of its successful gametes is due only to fair Mendelian segregation, and the discrepancies are independent for all individuals. The general version is that, at a locus, each allele in an individual has an expected representation in each successful gamete proportional to its representation in the individual's genome, and allocation of alleles to gametes in one individual is independent of the allocation in all other individuals. Formally, for a p-score that is the frequency of a single allele, for each individual $i \in I$,

$$\mathcal{E}[\Delta p_i^s : (w_j^s)_{j\in I}, (\Delta p_j^s)_{j\in I\, j\neq i}] = 0.$$

In fact, if the assumption holds for each allele, then it also holds for all p-scores. Conditional expectation has its standard definition (Schechter 1997, section 29.14). In this finite case, over each subset of $\mathbf{S}$ defined by sharing values of all the conditioning variables, the average value of Δp_i^s must equal zero.

The weaker form that suffices for all requirements here drops the need for independence between individuals, and for independence of the meiotic events from the number of successful events of other individuals. Formally, the weakened form is simply that, for each i,

$$\mathcal{E}[\Delta p_i^s : w_i^s] = 0.$$

Now we multiply the Price equation (6) for given s by τ^s and sum, then reexpress using expectations and covariances.

$$\sum_s \tau^s \Delta p^s = \mathbb{C}\left[p_i, \sum_s \tau^s \frac{w^s}{w^s}\right] + \mathbb{E}\left[\sum_s \tau^s \frac{w^s}{w^s} \Delta p_i^s\right]$$
$$\mathcal{E}[\Delta p^s] = \mathbb{C}\left[p_i, \mathcal{E}\left[\frac{w_i^s}{w^s}\right]\right] \qquad (7)$$

The second term on the RHS is dropped because at that stage we make the assumption of unbiassed transmission, which shows it to be exactly zero. Equation (7) has a very simple interpretation. The expected change in population p-score equals the covariance across the population between p-score on the one hand, and on the other, the arithmetic mean of the relative number of successful gametes, where relative is to the population mean.

The model of genotype, phenotype, uncertainty and number of successful gametes

The model under certainty of genotype, phenotype and number of successful gametes presented for the certain case was extremely simple. Here, the addition of uncertainty shows one way in which that model can be developed.

New notation is first defined. We have *(i) A set* $\mathbf{\Gamma}$ *of genotypes and a set* $\mathbf{\Phi}$ *of phenotypes, and a function* $\nu = \mathbf{\Gamma} \to \mathbf{\Phi}$ *carrying the genotype* γ_i *into the phenotype* $\phi_i = \nu(\gamma_i)$ *for individual i. (ii) A set* $\mathbf{U}$ *of ways in which a state of nature can affect an individual, and a function* $u : \mathbf{I} \times \mathbf{S} \to \mathbf{U}$ *such that* $u_i(s)$ *captures all the effects on individual i of state of nature s. (iii) A function* $\omega : \mathbf{\Phi} \times \mathbf{U} \to \mathbb{N}$ *specifying the number of successful gametes in state of nature s of an individual i with phenotype* ϕ*, so that:*

$$\omega_i^s = \omega(\phi_i, u_i(s)) = \omega(\nu(\gamma_i), u_i(s)).$$

The character of the dependences on state of nature has been left very open, and so the results will apply very generally. In fact, there is a restriction, in that the phenotype depends on the genotype, but does not depend on the state of nature, thus ruling out 'norms of reaction' or 'conditional strategies'. This restriction is more apparent than real. The reader is referred to Grafen (2002) for a treatment of uncertainty that incorporates this and further sophistications. The assumption made in the present paper is in keeping with the aim of presenting a simple case.

To understand this very abstract approach, we consider what s may represent, in a special case. Some years may be good and others bad, so that w^s may be higher in some years than others. Some individuals may die (struck by lightning or caught by a predator) before reproducing, and s then represents how many and which individuals. Other individuals will find abundant food for their young, and others will find little, and s will include these too. At a cellular level, we assume that there is a way of indexing all the alleles at all the loci, and that s specifies which subset of these alleles goes into the first successful gamete of an individual, which subset into the second, the third, and so on, up to some large but finite limit. Thus s contains enough information to calculate $(p')_i^s$.

We pause to consider why the set $\mathbf{U}$ is needed. s allows the calculation of $(p')_i^s$ and of $(p')_j^s$ for distinct individuals i and j. However, they are not the same function of s, otherwise they would always be the same as each other when all other factors were equal for i and j. Chance should be able to make two otherwise identical individuals different. That is why the intermediate structure of $\mathbf{U}$ is required. $u_i(s)$ says how s makes the world look from the point of view of individual i. Then we can make $(p')_i^s$ the same function of $u_i(s)$ that $(p')_j^s$ is of $u_j(s)$. This allows the function ω to be the same function of ϕ and u for all individuals. The placing and nature of subscripts, and especially the omission of subscripts, are vital elements of these kinds of argument.

The optimization programme with uncertainty

We can repeat the optimization argument from the certain case, but allow the maximand to depend on s. Then we get the program for s:

$$\phi \max g(\phi, s), \quad \phi \in \mathbf{\Phi} \tag{8}$$

If we wished to assume that the individual observed s and then made its decision, this would be an appropriate programme. But it is more useful, because more general (see Grafen 2002), to assume that, when the decision is taken, the state of nature is unknown. The obvious route is to take a probability-weighted arithmetic average over the states of nature, and so we construct a programme for all s as follows:

$$\phi \max \sum_s \tau^s g(\phi, s), \quad \phi \in \mathbf{\Phi} \tag{9}$$

We will see that links can be proved with this weighted average as the maximand. Other possibilities exist in theory, including geometric averages, and this is the juncture at which a case for them would have to take a different path. The test is the nature of the links that can be proved to the dynamics.

The next step is to define the components of the optimization programme in dynamic terms. The strategy set and the set of states of nature with its probabilities are already assumed to be in common, so it remains to find an expression for $g(\phi, s)$. The reduction from the population of individuals in the dynamics to a single implicit decision-taker in the optimization programme is going to require an additional assumption, and it is convenient to proceed by constructing a maximand for each individual i, say $g_i(\phi, s)$. In view of the division of w_i^s by w^s throughout equation (6), the obvious way to start is to choose

$$g_i(\phi, s) = \frac{\omega(\phi, u_i(s))}{w^s}.$$

Now we aim to be able to remove the subscript from g_i so that there is a unique maximand, but to require $g_i(\phi, s) = g_j(\phi, s)$ for all ϕ and s would be much too stringent, as we do not want to insist that two individuals playing the same strategy always have the same number of successful gametes.

A more permissive possibility is to insist that $g_i(\phi, s)$ and $g_j(\phi, s)$ have the same probability distribution as s varies, for each ϕ. This is a restriction on the u_i and τ^s, which states that each individual suffers the same range of effects from uncertainty, for each given phenotype. If this is not so, then there are indeed individuals in the population facing different kinds of problems, and it would be wrong to try to represent them all in the same optimization programme. We make this assumption, and call it 'strategic equivalence'. Parallel assumptions were called 'pairwise exchangeability' by Grafen (2002) and 'universal strategic equivalence' by Grafen (2006a).

Strategic equivalence allows some individuals to be killed by lightning while others survive, but insists that each individual has the same probability of being killed by lightning. Variation in foraging success is also allowed in any one state of nature, but the distribution over all states of nature must be the same for each individual.

The maximand of the 'program for all s' (9) can now be written in terms of the dynamic quantities as,

$$\sum_s \tau^s g_i(\phi, s) = \sum_s \tau^s \frac{\omega(\phi, u_i(s))}{w^s},$$

where strategic equivalence assures us there is no actual dependence on i. Notice that the w^s in the denominator are taken as fixed values, as the lack of dependence on ϕ shows that we do not consider the effect on the average number of successful gametes of the focal individual varying her strategy. It is another advantage of a formal representation of the optimization programme that we can choose the programme for convenience, so long as we can continue to prove results about it. The optimization approach explicitly considers the possibility of an individual varying its strategy, but we can pick and choose which effects of that variation to incorporate into the programme. This differs from the approach taken by Grafen (2002) who did allow the w^s to vary as the focal individual's strategy varied, but is in keeping with Grafen (2006a).

The significance of the w^s varying with s is that mean fitness varies with state of nature. This variation will alter the weights that individuals should place on the same marginal change in their mean number of offspring within a generation. In particular, they should place more weight on a given marginal change when w^s is smaller. This introduces a frequency-dependence, as the variation of w^s with s will depend on the phenotypes present in the population.

Links under uncertainty

We proceed to prove parallel results to those of the certain case, with parallel arguments. First, we reexpress two conditions from the case of certainty. 'No scope for selection' now means that the average value of change in mean p-score, averaged over states of nature, equals zero for each possible p-score; formally, $\mathcal{E}[\Delta p^s] = 0$.

To reexpress 'no potential for selection', we consider a population in which individual h has phenotype $\alpha \in \mathbf{\Phi}$ instead of ϕ_h, and let δ_i^s be the fitness differences from the original population so that,

$$\delta_i^s = \begin{cases} 0 & i \neq h \\ \omega(\alpha, u_h(s)) - \omega(\phi_h, u_h(s)) & i = h \end{cases}.$$

As before, we let p_i^h be a p-score that equals zero for $i = h$ and one for $i = h$. The expected change in frequency of this

p-score is given by equation (7), and assuming in the second step there is no scope for selection, we obtain

$$\begin{aligned}\mathcal{E}\Delta(p^h)^s &= \mathbb{C}\left[p_i^h, \mathcal{E}\left[\frac{\omega(\phi_i, u_i(s)) + \delta_i^s}{w^s}\right]\right] \\ &= \mathbb{C}\left[p_i^h, \mathcal{E}\left[\frac{\delta_i^s}{w^s}\right]\right] \\ &= \frac{n-1}{n^2}\mathcal{E}\left[\frac{\delta_h^s}{w^s}\right] \\ &= \frac{n-1}{n^2}\left(\mathcal{E}\left[\frac{\omega(\alpha, u_h(s))}{w^s}\right] - \mathcal{E}\left[\frac{\omega(\phi_h, u_h(s))}{w^s}\right]\right).\end{aligned}$$

If no alternative phenotype $\alpha \in \mathbf{\Phi}$ can positively invade, then we must have,

$$\mathcal{E}\left[\frac{\omega(\alpha, u_h(s))}{w^s}\right] \leq \mathcal{E}\left[\frac{\omega(\phi_h, u_h(s))}{w^s}\right] \tag{10}$$

for all $\alpha \in \mathbf{\Phi}$ and all h. This is the condition of 'no potential for selection in relation to a set $\mathbf{\Phi}$'. It is now possible to prove the results.

Optimization to dynamics: *If every individual solves the 'programme for all s' (9) with constraint set $\mathbf{\Phi}$, and there is unbiassed transmission, then there is no scope for selection, and no potential for selection in relation to $\mathbf{\Phi}$.* If the individuals all solve the programme, then the $\sum_s \tau^s g_i(\phi, s)$ must all be equal, for otherwise every nonhighest value would not be a solution. But this sum equals $\sum_s \tau^s \frac{\omega(\phi_i, u_i(s))}{w^s}$ and so also equals $\mathcal{E}\left[\frac{w_i^s}{w^s}\right]$, showing that $\mathbb{C}\left[p_i, \mathcal{E}\left[\frac{w_i^s}{w^s}\right]\right] = 0$. The assumption of unbiassed transmission allows us to use equation (7), from which we conclude that $\mathcal{E}[\Delta p^s] = 0$. We have not had to specify which p-score, and so there is no expected change in the mean of an arbitrary p-score, as required. Further, the optimization also implies that equation (10) holds for all $\alpha \in \mathbf{\Phi}$, and so establishing no potential for selection in relation to $\mathbf{\Phi}$.

Dynamics to optimization: *If the expected change in every p-score equals zero, and there is no potential for selection for any phenotype in $\mathbf{\Phi}$, and if there is unbiassed transmission, then all individuals must solve the optimization programme.* If there is unbiassed transmission, we may employ equation (7), which shows that the requirement of a zero covariance for every p-score implies that the $\mathcal{E}\left(\frac{w_i^s}{w^s}\right)$ are all equal. These equal $\sum_s \tau^s \frac{\omega(\phi_i, u_i(s))}{w^s}$, which is the maximand of the 'program for all s' evaluated at the phenotype of individual i, ϕ_i. As this is therefore equal for all individuals, they achieve the same value of the maximand. We also assume no potential for selection, which gives us from equation (10), substituting with g as appropriate, that

$$\mathcal{E}[g(\alpha, s)] \leq \mathcal{E}[g(\phi_h, s)],$$

for all $\alpha \in \mathbf{\Phi}$ and all h, which immediately shows that all individuals in the original population solve the 'program for all s' (9), as required.

The main difference in the proofs compared to the certain case is that here the second term in the right hand side of the Price equation is now dealt with analytically, and so no hand-waving is required: it was necessary to have uncertainty in the model to do this.

Discussion

The discussion is focussed, in line with the paper as a whole, on a mathematical introduction to the Formal Darwinism Project. For more general discussion of the project as a whole, the reader is referred to the other papers that make up the project (Grafen 1999, 2000, 2002, 2006a,b, 2007a,b,c). I begin by interpreting the mathematical results biologically, and show what the mathematical arguments can do for us, providing an important motivation. Succeeding sections glance ahead at future work from a mathematical point of view, and offer a substantive discussion of the value of working with dynamic equations that are dynamically insufficient.

Two brief points are dealt with immediately. The whole approach does not consider mutation. The properties of the system without mutation can be considered to be unlikely to be much perturbed by the introduction of a biologically realistic small mutation rate. Second, the various papers in the Formal Darwinism Project use notation very differently, and the reader is so warned. In an exploratory series of papers, with heavy notational demands, this inconsistency seems inevitable if regrettable.

The meaning and generality of the results

The results of the certain case are roughly that 'optimality implies no actual or potential selection' and 'the absence of actual and potential selection implies optimality'. Those of the uncertain case may be restated roughly as 'optimality implies an average of zero actual or potential selection', and 'no average actual selection and no average potential selection together imply optimality'. The results omitted here but given in the more technical papers say, roughly, 'gene frequencies change so as to move towards optimality'. Taken all together, these results do not directly state conditions under which a population can be expected to exhibit optimality on the part of each of its individuals, nor even whether that is to be expected. Conclusions of these kind would require further dynamic assumptions, and that may well be a direction for future work. The results do suggest that dynamic equilibrium is likely to be at or close to a situation in which all individuals are optimal, much of the time.

From a technical point of view, a definite optimization programme has been constructed and linked to a dynamical system. That bridgehead having been established, it is possible to make dynamical arguments that explore and may confirm the suggestions of near-optimality. The current emphasis in the Formal Darwinism Project is in widening the bridgehead rather than undertaking further explorations from it.

The linking results hold for hypothetical populations satisfying the assumptions, and we now enquire into the generality of these conclusions, taking the uncertainty model as the main example. The structure of the argument is new in biology, and it is important to understand how widely the conclusions of the Formal Darwinism Project hold, as well as the nature of their limits. Two kinds of generality are important: whether 'real-life' populations are likely to meet the assumptions, and also whether models of gene frequency change are likely to meet the assumptions. The second, 'meta-model', question may allow the optimizing approach to be used as an organizing principle for dynamical models.

A very important feature of the results is that conclusions do not refer to a particular trait, but show that the given maximand is relevant to all traits; note the implication that 'the maximand is the same for all traits'. The links therefore establish an organism-wide maximand, relevant to the natural selection of all genes and all traits, that is a property of an individual. This level of the individual is important. It shows first of all that selection acts on a trait in the same way whether that trait is determined by a single locus, two loci, a few loci, or many loci. Equally, selection will act on different traits in a consistent way, because the individual's maximand is the same for all traits. We, therefore, expect to see harmony in the effects of selection. The liver, kidneys, heart, brain and skin of an individual should be acting together as if with a common goal. This extension to the whole organism is of central importance to the optimizing view of natural selection, and supplies a level of understanding that is omitted from dynamical models that consider one particular genetic architecture.

We have assumed that all individuals share the same ploidy, but this can easily be relaxed at the expense of some extra notation (Grafen 2002). The substantive assumptions are that all the loci involved have the same inheritance pattern, and that segregation is fair and Mendelian. This assumption is met if all the loci are autosomal or pseudo-autosomal; or if all the loci are Y-linked (in which case females are simply not included in the population). If we do allow ploidy to vary within the population, this assumption can be met if all the loci are X-linked. What this implies is that there is a different maximand for each mode of inheritance, formalizing the original discovery of intra-genomic conflict by Hamilton (1967), recently reviewed at book length by Burt and Trivers (2005).

The model of genotype, phenotype and number of successful gametes in both the certain and uncertain cases makes clear that there are additional assumptions at quite a high level of abstraction. We assume that the number of successful gametes of an individual depends only on its own genotype, and not on the genotypes of others: this rules out social behaviour. Further, we assume that there are no classes of individuals, such as male and female, or large and small, for then the number of successful gametes would have to depend on an individual's class, as well as on its genotype and on uncertainty. The model has a logical completeness, in which these assumptions must be made to draw any conclusions, which therefore exhibits all the assumptions required.

Finally, there is the special assumption of strategic equivalence. This is a new kind of assumption, revealed by the Formal Darwinism Project, and required whenever the population of the dynamic side of the model has to be linked to the implicit single decision-taker of the optimization programme. An assumption is required to say that, in essence, all individuals are the same. This is natural, for otherwise it would be wrong to end up with a single optimization programme. Variability in outcome between individuals is permitted provided each individual has equal starting chances in life.

The question arises of how the conclusions can fit in with the well-known difficulties with optimality that arise under heterozygote advantage and other genetic circumstances. The discussion by Grafen (2007b) is not repeated here, because the current paper is focussing on a formal introduction to the mathematical arguments.

Further extensions

The present paper gives the simplest possible links between the Price equation and optimization programmes, but previous papers give more advanced links. In this section, various published extensions are first reviewed briefly, and then a programme for future extensions is discussed.

The first set of links were proved by Grafen (2002), and correspond roughly to the *origin of species* (Darwin 1859), except for Grafen's assumption of discrete nonoverlapping generations. As well as uncertainty, included in a more general way than here, that paper contains a representation of information processing by the individual; and the population and the set of states of nature may separately be finite or infinite. Grafen (2006a) permits social behaviour by allowing an individual's number of successful gametes to depend on the genotypes of other individuals. Grafen (2006b) allows offspring to belong to different classes, such as sex or size, and derives reproductive value as a means of evaluating the contribution of different offspring to a parent's overall reproductive success.

Each extension follows the general pattern of the extension to uncertainty carried out earlier in this paper. There is a formal modification to the model linking genotype to phenotype to the number of successful gametes, adding dependencies and/or introducing further elements. The details of the steps of construction of the optimization programme, construction of appropriate links, and the proof of those links, must all be modified in the light of the altered formal framework.

Future work will involve two major tasks. The first is to add further extensions. In particular, the assumption of discrete nonoverlapping generations should be relaxed to permit overlap of generations and if possible to combine continuous and discrete time in a single formalism. The second is

to combine the different extensions into a single model, to allow social behaviour simultaneously with classes, and both simultaneously with overlapping generations. This work will be technical, but the anticipated outcome is an extremely general and unified model of natural selection.

Dynamic insufficiency

The results of previous sections are of the form 'if we know $(w_i)_{i\in I}$, and the $(p_i)_{i\in I}$, and some other things, then we can predict the change in average p-score'. But we have nowhere been required to make any further assumption about how the gametes, counted in the w_i, segregated into individual offspring. This makes our assumptions dynamically insufficient, because we are unable to produce enough information about the offspring generation to get to the position in which we can apply the Price equation again—hence, we cannot 'crank the handle' and create a simulation of evolution over the generations (I am grateful to Steve Frank for helping me see that it is the assumptions themselves, rather than the Price equation built on them, that introduce this dynamic insufficiency).

What is the significance of developing a framework based on this partial representation of the dynamics? First note that, unlike the dynamically insufficient models widespread until the 1970s (commented on by Lewontin 1974), the framework is mathematically exact in its claims, the proofs given in the uncertain case are all fully rigorous, and no approximation is involved. We simply limit our claims to those that can be rigorously established within the limited nature of the assumptions.

The advantages of developing a framework based on partial information, and which for some purposes outweigh the disadvantages of the limitations, are (i) we have access to very general results, as shown earlier in the Discussion, (ii) we can focus technical effort on issues that are important to understanding whole-organism features and behaviour, (iii) that level of detail of the dynamics seems to be a fruitful level at which to make links to optimization programmes, and (iv) that level relates very directly to biologically relevant literature, including empirical and modelling studies that employ the idea of fitness optimization but do not use information about genetics.

These four points are now expanded in turn. The extent of the generality was discussed earlier. It matters because it shows that natural selection operates in the same way across a whole range of assumptions, as one argument covers them all, and we do not have a large and varied set of arguments that just happen to give similar conclusions. This is the aspect that underlies the hope of Grafen (1999) to produce a mathematical version of the core argument of Darwin (1859, 1871), providing for that work the same kind of exhaustive mathematical representation that physical theories, such as Newton's mechanics and special relativity, have. It is the pursuit of this unifying objective that requires further development and integration of all the general assumptions so far, and adding the yet further generality of permitting overlapping generations and continuous time.

The generality also matters because the framework can act as a meta-model, and can be applied to a wide range of existing models. This has allowed the reinterpretation of two kinds of model in terms of fitness optimization by the direct application of results obtained using the partial dynamics. Grafen (2007a) reintepreted a model of Killingback *et al*. (2006), showing that the evolution of cooperation in a metapopulation model with variable group sizes was after all interpretable in terms of inclusive fitness, and Grafen (2007c) showed that an inclusive fitness analysis of cooperation evolving on a cyclical network in a model of Ohtsuki *et al*. (2006) provided a biologically more meaningful interpretation than graph theory did. These papers support the more widespread and increasing trend to insist that models of social behaviour all come within the ambit of inclusive fitness theory (see also Lehmann and Keller 2006; Lehmann *et al*. 2007). Note, however, that the principle that selection is linked to fitness optimization is more general, and covers non-social behaviour, and classes, and applications of those parts of the project are in progress.

The second advantage is that technical effort saved from the minutiae of dynamical systems can be focussed instead on biologically important issues. For example, essentially all behaviour is conditional on cues and information received, and behavioural ecologists routinely assume, in effect, that organisms use information optimally, when they assume any kind of optimization of fitness. The technical complexity saved by using only the partial information on dynamics allows an explicit representation of the use of information (Grafen 2002). To give another example, most derivations of inclusive fitness since Hamilton (1964, 1970) have focussed on just one type of social action. By using only the partial information on dynamics (and in this respect following Hamilton's original derivation), Grafen (2006a) was able to establish a principle of optimization of inclusive fitness in a model that did not restrict the number and nature of the types of social action (to be explicit, additivity of fitness effects was required, but actions could vary in benefits and costs, and in how many recipients were involved, and in their relatednesses to the actor). If some mathematical complications relating to dynamics can be done without, it can pay to avoid them and invest the effort in biologically significant sophistications.

The third advantage is that it is possible to prove links between the partial dynamics and optimization programmes. I am unaware of any such links being proved using complete dynamics. The partial dynamics are linked fairly simply, and it seems that there is a natural connection at that level of dynamical detail. As the partial links are exactly true, and therefore consistent with whatever further information would permit the dynamics to be made complete, it is most unlikely that a link with complete dynamics could produce a different optimization principle.

The final advantage is that the terms of discussion are the same as in many biological discussions. Biologists routinely measure reproduction, or proxies for reproduction, so the $(w_i)_{i\in I}$ are at least within range of potential observability, and they do relate to individual organisms. The remaining information required to complete the dynamics would require knowledge of the genomes of all the parents, and enough detail about the gametes and segregation to construct the genomes of all the offspring. Despite the advances of molecular biology, this is a completely unrealistic level of information even to aspire to, for most organisms studied behaviourally. Fortunately, it is clear from the project so far that the simpler information is adequate for many purposes (though it is true that the important question of what purposes the simpler information is not adequate for has not been much explored).

The other biological discussions that use the same terms as the Formal Darwinism Project include Darwin (1859, 1871) and Dawkins (1976). Connected to the aim of providing a rigorous mathematical version of Darwin's core arguments, mentioned above, is the goal of showing rigorously that the verbally persuasive arguments of Darwin and Dawkins have mathematically sound equivalents, thus justifying formally arguments that have inspired many biologists and sustained the intellectual level of their work. At a less verbal level, it was noted in an earlier section that the project also generalizes the *Fundamental theorem of natural selection* of Fisher (1930), elaborating and rendering fully explicit the link between dynamics and optimization, whose significance Fisher first saw and formalized.

Darwin (1859) reached his conclusion that the mechanical processes of inheritance and reproduction gave rise to the appearance of design, knowing little about how inheritance actually worked. In that light, it is perhaps less surprising that a formalized version of his conclusion does not require all the details of inheritance. It does, of course, mean that if selection of some character, and recombination rates seem a likely candidate in this connection, does depend on the further details, then the results of the Formal Darwinism Project must fail to capture how selection acts on that character. But the logical rigour of the project will mean that we will see where, in the argument, the assumption is made that excludes it. The list of exceptional cases could be a very interesting outcome of the culmination of the project.

Acknowledgements

This manuscript found life in the necessity to present the Formal Darwinism Project to a masters course in Kingston, Ontario, run by Prof. Peter Taylor and Prof. Troy Day. I realized that all the work up to then was unapproachably complicated. I am grateful to Peter and Troy for that challenge, and to the class for their thoughtful reactions to a first attempt. Dr Andy Gardner, Prof. Laurent Keller, Dr Laurent Lehmann, and Dr Francisco Úbeda de Torres made very useful comments on subsequent versions.

References

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

Darwin C. R. 1859 *The origin of species.* John Murray, London.

Darwin C. R. 1871 *The descent of man and selection in relation to sex.* John Murray, London.

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

Dawkins R. 1982 *The extended phenotype.* W. H. Freeman, Oxford.

Edwards A. W. F. 1994 *The fundamental theorem of natural selection. Biol. Rev*. **69**, 443–474.

Ewens W. J. 1979 *Mathematical population genetics.* Springer, Berlin.

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.

Ewens W. J. 2004 *Mathematical population genetics I. theoretical introduction.* Springer, Berlin.

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, Oxford.

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

Grafen A. 1985 A geometric view of relatedness. *Oxf. Surv. Evol. Biol*. **2**, 28–89.

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

Grafen A. 2000 Developments of Price's equation and natural selection under uncertainty. *Proc. R. 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. R. Stat. Soc. Ser. D (The statistican)* **52**, 319–329.

Grafen A. 2006a Optimisation of inclusive fitness. *J. Theor. Biol*. **238**, 541–563.

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

Grafen A. 2007a Detecting kin selection at work using inclusive fitness. *Proc. R. Soc. Ser. B* **274**, 713–719.

Grafen A. 2007b The formal Darwinism project: a mid-term report. *J. Evol. Biol.* **20**, 1243–1254.

Grafen A. 2007c Inclusive fitness on a cyclical network. *J. Evol. Biol*. **20**, 2278–2283.

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

Killingback T. Bieri J. and Flatt T. 2006 Evolution in group-structured populations can resolve the tragedy of the commons. *Proc. R. Soc. Ser. B* **273**, 1477–1481.

Lehmann L. and Keller L. 2006 The evolution of cooperation and altruism: a general framework and a classication of models. *J. Evol. Biol*. **19**, 1365–1376.

Lehmann L., Keller L. and Sumpter D. 2007 The evolution of helping and harming on graphs: the return of the inclusive fitness effect. *J. Evol. Biol*. **20**, 2284–2295.

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

Maynard Smith J. 1982 *Evolution and the theory of games.* Cambridge University Press, Cambridge.

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

Ohtsuki H., Hauert C., Lieberman E. and Nowak M. 2006 A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502–505.

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

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

Price G. R. 1972b Fisher's 'fundamental theorem' made clear. *Ann. Hum. Genet*. **36**, 129–140.

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

Schechter E. 1997 *Handbook of analysis and its foundations.* Academic Press, San Diego.

Schwartz J. 2002 Population genetics and sociobiology: conflicting views of evolution. *Perspect. Biol. Med.* **45**, 224–240.

Segestrale U. 2000 *Defenders of the truth.* Oxford University Press, New York.

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

Received 19 June 2008; accepted 30 June 2008

Published on the Web: 23 December 2008