A First Formal Link between the Price Equation and an Optimisation Program

Alan Grafen, alan.grafen@sjc.ox.ac.uk

St John's College

Oxford OX1 3JP

United Kingdom

February 26, 2002

Accepted by *Journal of Theoretical Biology*.

Running Headline: Price equation and optimisation

Abstract

The Darwin unification project is pursued. A meta-model encompassing an important class of population genetic models is formed by adding an abstract model of the number of successful gametes to the Price equation under uncertainty. A class of optimisation programs are defined to represent the 'individual-as-maximising-agent analogy' in a general way. It is then shown that for each population genetic model there is a corresponding optimisation program with which formal links can be established. These links provide a secure logical foundation for the commonplace biological principle that natural selection leads organisms act as if maximising their 'fitness', provides a definition of 'fitness', and clarifies the limitations of that principle. The situations covered do not include frequency dependence or social behaviour, but the approach is capable of extension.

1 Introduction

Population genetics is the part of biology in which the operations of natural selection should be studied, and their consequences worked out. Most biologists who study adaptiveness in the field hold the view, following Darwin (1859), that natural selection leads organisms to act as if they are maximising their fitness. This view is not supported by the current state of population genetics, despite recent exegesis and developments, discussed in Section 2.2, of the first candidate supporting theory, the Fundamental Theorem of Natural Selection of Fisher (1930).

The discrepancy has many consequences for the current status of biology. Lacking a formalised version of Darwin's theory, biologists have a collection of theories based on optimisation which have no common root in basic theory. Thus widely used and taught ideas in many biology courses, optimal foraging, ESSs and inclusive fitness, have no single underlying theory to match the single underlying intuition. There are arguments about what kind of average of fitness

is maximised by natural selection, arithmetic or geometric. Challenges to orthodoxy have to be met with ad hoc arguments, rather than a simple application of a general Darwinian result. Finally, the central concept of Darwinian fitness remains ill-defined, and a source of conceptual difficulty.

There is valuable work establishing the foundations of ESS theory (Hammerstein, 1996) and inclusive fitness (Taylor, 1996), which reveal the nature of the ideas and their logical bases. However, these are not unifying foundations, and do not attempt to justify the optimisation view as a whole.

Darwins original argument has been successfully expounded and advanced in words e.g. Williams (1966), Dawkins (1976), but there is no accepted mathematical representation of these advances. Grafen (1999) proposed a project to construct one, effectively finding a replacement for the fundamental theorem that would justify the general biological view that natural selection leads organisms to act as if maximising their fitness. Grafen (2000) made progress in developing the covariance selection mathematics of Price (1970, 1972) to include arbitrary uncertainty. The present paper now fulfils a further step in the project by establishing the first formal link between the Price equation and optimisation programs.

Section 2 first discusses why this link is important, and in particular why establishing the nature of the 'target of selection' in an appropriately elaborated Price equation, as Grafen (2000) did, is not enough. Sections 3 to 7 introduce and use a highly technical mathematical apparatus, and will not be found easy to read. Accordingly, Section 2 also contains an overview of the technical arguments, and an explanation for why such an apparatus is required. Later, Section 8 reviews the argument in non-technical terms. Thus it should be possible for a reader to jump straight from Section 2 to Section 8 and retain a sense of the argument of the paper.

A curious parallelism in later sections is worthy of note. The population genetic development of Sections 3 and 4 will treat organisms as machines whose morphology and behaviour can be studied, but are taken as given. The ap-

proach here uses Price's covariance selection mathematics (Price, 1970, 1972a) as developed by Grafen (2000), but is general enough to include a wide range of simple population genetic models. See Grafen (2000) for a discussion of the dynamic insufficiency of Price's method in this context: it does not compromise the exactness of the results obtained here.

The optimisation development in Section 5 will treat organisms as designed solutions to problems. The approach will use the formal structures of optimisation programs, and will be motivated by examples from behavioural ecology, for the sake of concreteness. However, the ideas are very general to the adaptationist approach in biology.

The parallelism is that these sections develop two very different kinds of formal structure, which have both been used extensively by biologists, and are often used to discuss the same phenomena. Indeed, biologists have frequently employed ideas from both types of structure simultaneously, in a productive promiscuity. Here, similar ideas have to be treated with different words and separate notations, because a formal justification for adaptationism requires us to distinguish sharply between dynamics and optimisation. The formal links established in Section 5, and discussed in Section 8, will seem at first like weak truisms to readers who take the validity of the optimisation analogy for granted. Their whole point, however, is precisely to justify, to as large an extent as can be logically defended, the validity of that analogy.

This paper develops a formal link between population genetics and optimisation programs in a special case. The apparatus required is so elaborate that it seemed right to pause there, which more or less attains the level of representing the adaptationist arguments of Darwin (1859). The last remaining technical restriction at that level is to discrete generations. However, a major purpose is to prepare for extensions that will encompass the twentieth century additions to adaptationist theory, and so bring to fulfilment the 'formal Darwinism unification project' of Grafen (1999).

2 Motivation

This section's first aim is to clarify the importance of employing optimisation programs, and constructing a formal link between them and population genetic models. Second a comparison is made between the present exercise and Fisher's fundamental theorem of natural selection. The final subsections provide an overview of the technical arguments and make the case that the abstraction gained by the sophisticated mathematics employed is important.

2.1 The inadequacy of 'targets of selection'

The Price equation in its simplest form states that the change in a p-score equals the covariance across individuals between the p-score and relative fitness, formally, using notation of Grafen (2000) but ignoring uncertainty for the purposes of illustration,

$$\Delta p = \mathbb{C}_i \left[p_i, v_i \right] \tag{1}$$

Following Grafen (2000), we assume perfect transmission, that is, no mutation, no gametic selection, fair meiosis, and that all contributing loci have the same mode of inheritance.

It is tempting to conclude that selection leads individuals to act as maximising v_i, their relative fitness, and if this were so there would be no need to introduce optimisation programs. The term with which p_i appears in the covariance was called the 'target of selection' by Grafen (2000). However, by considering social behaviour as an example, we can see that natural selection need not lead to maximisation by individuals of the target of selection. The technical arguments of this paper will not include social behaviour, but the purpose is to develop an approach that will be ready to include it.

The arguments that lead to Equation (1) still apply when there are social interactions of the kind envisaged by Hamilton (1964), and for which he showed natural selection caused individuals to act as if maximising their inclusive fitness.

It is instructive to examine just where the difficulty arises. An individual's v_i may depend on her own actions, and also on those of other individuals. The Price equation acts as an after-the-fact accounting scheme, which is exactly true in a diverse set of circumstances, but does not recognise issues of control. In the case of interacting with sibs, suppose that an individual increased her own v_i and provided a positive element to the sum that constitutes the covariance in Equation (1); if she also reduced the v_i of her relatives, she would contribute negative elements to that same sum. Thus whether a trait spreads would not be revealed solely by the effect of an action on an individual's own v_i. Put in another way, the individual could increase her v_i by altering the genotypes of her relatives, so that they were altruistic to her. These genotypes are not under her direct control, and yet she should receive some benefit for having altruistic genes because these are likely to be shared with relatives and so indirectly she does increase her v_i. The target of selection in the Price equation therefore does not reveal how selection will act on individual behaviour when interactions with relatives is concerned. In view of the absence of the concepts of control or of the set of possible phenotypes, this is, at least on reflection, only to be expected.

In order, then, to uncover an optimisation principle in the workings of natural selection, we will link the Price equation to optimisation programs that leave no room for doubt what it is to establish that an individual is acting to maximise a given quantity. In the present paper, this doubt will be removed for uncertainty and for varying ploidy levels.

The Price equation combines great generality with great subtlety of interpretation, because it leaves completely unspecified the links between genotype and phenotype, and between phenotype and fitness. Our aim is leave the genotype-phenotype link completely arbitrary, as this allows the equation to hold over more or less arbitrary genetic architectures. But we will elaborate a model of phenotypes and their link to fitness, to be incorporated into the Price equation, in order to specify whose genotype causes which effects on fitness. This model in combination with the Price equation will allow links to be made with

optimisation models.

The Price equation, then, will play a central part in the thesis of this paper, but needs to be combined with a model of phenotypes and how they link to fitness. In order to maintain the same generality as is found on the genotypic side, this phenotypic model will be very general and abstract.

2.2 The Fundamental Theorem of Natural Selection

Fisher (1930) believed that his fundamental theorem demonstrated a maximisation principle in the workings of natural selection, and the current exercise is necessary because that principle is not sufficiently elaborated and is not established sufficiently generally. Before discussing those deficits, the exegetical literature on the fundamental theorem is briefly reviewed. Price (1972b) explained the meaning and derivation of the fundamental theorem, but doubted its biological significance. (He did so in Fisher's own terms, and did not employ the covariance selection mathematics he was developing at the same time (Price, 1970, 1972a) and which forms a basic element of the theory of the current paper.) Ewens (1989) expanded Price's argument, and the subject is reviewed by Edwards (1994). Lessard (1997) claims a more authentic interpretation of Fisher's argument, while Frank (1997, 1998) presents his own exposition and development of the fundamental theorem.

Deriving the fundamental theorem is very easy using covariance selection mathematics, as shown by Frank (1997, 1998). Relative number of successful gametes is a trait like any other, and so must have an additive genetic component (Falconer, 1981), which will equal some weighted sum of allelic values, and therefore specifies some particular p-score. With that p-score, $\mathbb{C}_i\left[p_i, v_i\right] = \mathbb{C}_i\left[p_i, p_i\right]$, leading through an application of the Price equation in Equation (1), ignoring uncertainty, and assuming perfect transmission, to $\Delta p = \mathbb{C}_i\left[p_i, p_i\right] = \mathbb{V}_i[p_i]$. Hence, the change in the mean additive genetic value of fitness equals the additive genetic variance in fitness.

The important conclusion for our purposes follows from the next step. As

variances are by definition non-negative, and will usually be positive, it follows that the additive genetic value of mean fitness never decreases, and will usually increase. By ignoring the distinction between the additive genetic value of fitness and fitness itself, we could conclude that mean population fitness must increase. This erroneous conclusion, hotly contested by Fisher, formed the basis of Wright's adaptive landscape (see Ewens, 1989).

Price distinguishes, following what can be seen in retrospect as clear indications from Fisher, the partial change in fitness due to changes in allele frequencies, which the fundamental theorem shows must be positive, from the remainder of the change in fitness, which is attributed to environmental change. This extended sense of 'environmental change', including changes in genotype frequencies, as the effect of an allele on phenotype can be altered by its genic environment. Lessard (1997) proposes a revision of the exact nature of this distinction.

The primary complicating effect of these environmental factors is to alter the weightings of alleles in the additive genetic value of fitness from one generation to the next. The additive genetic values that increase from generation t to generation $t+1$, and from $t+1$ to $t+2$, may then be different characters, and so neither is guaranteed to increase between generations t and $t+2$. So the fundamental theorem doesn't guarantee that there is a character that increases over more than one generation.

Price found the fundamental theorem disappointing because it discussed only a partial change in fitness, and ignored environmental change. Ewens (1992), Frank (1997, 1998) and Lessard (1997) have extended Fisher's result to incorporate all of gene frequency change. The approach in the current paper focusses solely on the partial fitness with which Fisher's original version dealt. One defence is that the natural selection part of evolution tends to improve design, while the environmental change part is essentially non-constructive. The results of this paper, like Fisher's, hold true even when those environmental changes are present. The limitations arise in the interpretations of the results,

and will be made explicit in Section 8.

Thus the current paper is a formal exploration of Fisher's sense that the partial result is the important one for the link between natural selection and design. The completeness sought by Price (1972b), and pursued by Ewens (1992), Frank (1997, 1998) and Lessard (1997) will be important for providing a complete account of gene and genotype frequency changes.

Let us return to the deficits in the fundamental theorem. The optimisation side is insufficiently elaborated for modern purposes, which operate with sophisticated use of information, and averaging over uncertainty. The current paper makes that good by establishing formal links with optimisation programs. Frequency dependence and social behaviour are further complications which the current paper does not consider, but the elaboration of optimisation programs is a preparation for them.

The second deficit is that the fundamental theorem is insufficiently general. It does not explicitly consider uncertainty in fitness, differing ploidies, or multiple loci. These restrictions are mainly important because a general version should not need to concern itself with these details any more than Darwin did. Another specificity, not generalised in this paper, but for which the apparatus developed is a preparation, is that Fisher's derivation assumes there is only one type of offspring.

One aspect of difference is that Fisher's derivation is in continuous time. Ewens (1989) showed that the original derivation in continuous time is easily extended to discrete time. The current paper's approach may be capable of generalisation to include continuous and discrete time in one formulation.

Fisher's approach looked to physics and statistical mechanics, and Fisher emphasised the formal similarity of his maximisation principle to that of entropy. The approach of the present paper looks to economics and the mathematics of optimisation.

The two approaches are, however, similar in their ambition to produce a 'constitutive model' of the order-creating capacity of natural selection, that

is, a model that is not merely an example of natural selection at work, but rather aims to contain the whole argument that natural selection leads to the appearance of design. This aspiration led to Fisher's naming of the fundamental theorem, and the goal of the 'formal Darwinism unification project' is to make good that aspiration under very general conditions, employing more modern mathematical techniques.

2.3 Overview of the technical arguments

The aim of the following sections is to link two formalisms, population genetics and optimisation programs, in order to establish that a wide class of population genetic models admit an interpretation of their equilibria in terms of individual optimisation. Here the argument is sketched, its technical complexity is justified, and attention is drawn to notable parts.

The first stage is the development of the population genetic formalism in Sections 3 and 4. The essence is adding a model of phenotypes and their link to fitness in the presence of uncertainty. The number of successful gametes is modelled as a function with three arguments:

1. the realised-phenotype. This in turn depends on the whole phenotype (viewed as a norm of reaction) and the local environment of individual i.
2. chance factors specific to individual i
3. population-wide chance factors

The local environment of individual i and the chance factors specific to individual i are i-specific functions of the state of nature. The population-wide chance factors are represented by an i-independent function of the state of nature. It is vital that the function of the three arguments is the same function for all i.

The omissions from the list of arguments are also important. The phenotype of individuals other than i does not appear and so cannot influence the number of successful gametes. Nor can they appear 'through the back door' in second or

third arguments, because the state of nature is assumed not to contain information about the genotypes of individuals, and so is unable to contain information about phenotypes. (This point is discussed in more depth in the more technical sections.)

The absence of the phenotypes of others renders the model incapable of representing various important biological topics, including the evolution of social actions and parental care. Extending the list of arguments is one way to elaborate the model to incorporate these topics.

A vital assumption is then introduced about the population genetic model, to be called *pairwise exchangeability*. This effectively decouples the individual genotype from the local environments and individual-specific chance effects, imposing an assumption that while individuals may differ in phenotypes, "time and chance happeneth to them all". One relevance of this is to ensure that we do not have on average an association between, for example, one allele and particularly favourable environments. The major reason is to ensure that all individuals face the same environmental challenges, and so are having to solve the same problems.

The second stage is the development of the optimality framework in Section 5. Here, ideas that are already commonplace in biology are given an unusually formal presentation, borrowing optimisation programs from economics and game theory. A series of optimisation programs sees the biological model becoming more sophisticated. The first is suitable for a simple optimal foraging example with deterministic gain functions and a simple optimum. By the end, the organism is allowed to observe a random variable, and she maximises an expected payoff function conditional on the observation, updating her information in an optimally Bayesian way in light of the cues available to her. It is characteristic of the strategic approach that concern focusses on information, and on constraints on action.

The third stage shows how to construct, for any population genetic model belonging to the framework established in Section 4, a corresponding optimisa-

tion model belonging the framework established in Section 5. Some concepts are simply identified, such as the two regimes of uncertainty. The key point in the construction is the reduction from a population of individuals in the population genetic framework to a single decision-taker in the optimisation framework. First, each individual is provided with a maximand in Section 5. Then it is shown what needs to be assumed to ensure that all the maximands are the same, and it turns out that 'pairwise exchangeability' will suffice provided one or other of a pair of additional conditions holds. Equation 5 derives the maximand from the population genetic framework. Its technical interest, however, is not so much its form, but rather that we can prove conditions under which the maximand is the same for each individual. It is therefore the subscripts, rather than the variables, that hold the key to the technical significance of the formula, and in some crucial places (such as the function f) it is the absence of a subscript.

Finally, links are established between the population genetic model and its corresponding optimisation model. These links justify the identification of the maximand with Darwinian fitness, and provide a logical foundation for the use of optimality ideas in discussing the operation of natural selection, while placing strong limits on them.

The next subsection discusses the technical complexity of the argument, and briefly introduces the states of nature approach and measure theory. Readers happy to proceed non-technically are invited to take in as much of the next subsection as possible, and then skip to Section 8.

2.4 Technical complexity and generality

The argument unfortunately involves some measure theory, and potentially infinite-dimensional spaces (Schechter, 1997, contains all the relevant mathematics, though it cannot be recommended as introductory). The justification is that the argument is very general, which itself has three main purposes. After discussing these purposes, a brief introduction is made to the representation of

uncertainty by the 'states of nature' approach, and to measure theory.

The first reason to be general is to show that the optimisation link with natural selection is not just a coincidence in a special case, but a fundamental fact about a class of selection processes. Furthermore, the formal Darwinism unification project aims to provide a technical representation of the commonsense, informal, arguments first proposed by Darwin (1859), and accepted by generations of biologists since. The formal argument should work in the same way for finite and infinite populations; for haploid populations, diploid populations and mixtures; for one-locus, two-locus and multi-locus traits; and for cases with and without environmental stochasticity, with finite or infinite sets of possible environments. Darwin did not take these cases separately, and neither should we. It is worth noting that, although the apparatus is complex, the argument is simple, reflecting the persuasive nature of the original verbal argument.

Another advantage of generality is that the theoretical developments here can be viewed as 'meta-models', that is, as models of models. The aim is to show that a wide class of existing population genetic models admit of an optimisation interpretation, and to show how to construct the corresponding optimisation model. This purpose is fulfilled in proportion to the generality of the model.

Finally, the model is not yet general enough. A general argument provides a better source for further development than a special case. For example, inclusive fitness and ESS theory could be incorporated with careful extensions of the model, and ideally both would be incorporated simultaneously.

The main technical complexity is the use of measure theory. In its application to the population, this allows finite and infinite populations to be handled at the same time. The conceptual foundation of all probability is measure theory, which is therefore implicit in many biological models already.

Measure theory is also applied to the states of nature, which are themselves perhaps unfamiliar to many biologists. The state of nature approach to handling uncertainty is to define a variable, in this case ω, to represent all the uncertainty. If the only uncertainty is the average temperature, then ω can be a

real number denoting the temperature. If each individual in the population has its own micro-climate, then ω could contain a specification of the temperature of each individual. In an infinite population, that would be an infinite number of temperatures. Having defined ω, we can then allow functions to have ω as an argument, to represent how uncertainty affects certain features. We also need to specify the probability distribution of ω, and this is done here using measure theory.

The basics of measure theory will be superficially reviewed, using the state of nature ω as an example. First we require the set in which ω takes its values, and we will call it Ω. If this were a finite set, we could specify the probability of each possible value separately. But in an infinite set (consider the example of a Normal Distribution on the real line) it is common that the probability of any particular value is zero. To circumvent this difficulty, we decide to attach probabilities to subsets of Ω rather than to individual points. For example, the probability that a Normal variable lies between two values is given by the area under the curve between them. Under certain conditions this allows the definition of a probability density, but we remain general and stay with the idea of assigning probabilities to subsets of Ω. For technical reasons, we cannot assign probabilities to all subsets of Ω (unless Ω is a very small set), as this generates paradoxes in which probabilities of events do not add up as they should. In measure theory, therefore, we specify a set of subsets of Ω, called a σ-algebra, over which we will define probabilities. In the present paper, that set will usually be indicated with a bar over the original set name, so the set of subsets of Ω over which we will define probabilities will be written $\bar{\Omega}$. Finally, the specification of the probabilities must give a probability for each element of $\bar{\Omega}$, and this is a function from $\bar{\Omega}$ into the real line. For states of nature, we call this function m. The triple $(\Omega, \bar{\Omega}, m)$ is known as a measure space.

The essence of defining this measure space is that we have assumed no more than that a variable taking values in the set Ω is capable of having probabilities defined on it. If Ω does not have this property, then it is unlikely to make

biological sense to say that the state of nature takes values in Ω. Thus we have retained complete generality in working with the measure space.

The level at which this model engages with uncertainy in fitness is more abstract than that of the illuminating exegesis by Frank and Slatkin (1990) of the nature of uncertainty in fitness, but is entirely compatible with it. The treatment here simply does not grapple with the same issues because those important levels of biological detail are not germane to the methodological task in hand.

The complex mathematical apparatus has its purpose, then, in allowing the argument to be simultaneously general and simple. This simplicity resides not in its ease of comprehension, which will admittedly be hindered by the complexity of the apparatus, but in the logical structure of the argument once the apparatus is set up.

3 Price's Equation and Measure Spaces

The starting point for the population genetic modelling will be Grafen (2000)'s development of Price's equation (Price, 1970, 1972a), but more formally expressed. Needed mathematical results will be cited in the textbook of Schechter (1997) as 'Sch' followed by a chapter and section number. The population I of individuals is indexed by i. The expectation over this population $\mathbb{E}_i$ used by Grafen (2000) implicitly assumes the existence of a 'measure space' $(I, \bar{I}, \mu)$, which consists of the set of individuals I, a set $\bar{I}$ of subsets of I with special properties (called a 'σ-algebra'), and a measure function μ that assigns a probability to each member of $\bar{I}$ (Sch 11.39). Price's equation applies to mixed ploidies (Grafen, 2000), and we assume that the ploidy of individuals is represented by a function $d\colon I \to \{1, 2, 3 \dots\}$. There is a probability measure $\tilde{\mu}$ over the measurable space $(I, \bar{I})$ that is a ploidy-weighting of μ itself (Sch 21.38(i)), and it is with respect to $\tilde{\mu}$ that our expectations will be taken. Assume a function $p\colon I \to \mathbb{R}$ denoting the p-score of individual i. A p-score can represent the frequency of

a particular allele, or an arbitrary weighted sum of such frequencies, so that the additive genetic value of each character can be represented as a p-score (see Grafen, 1985, for further details). Another measure space $(\Omega, \bar{\Omega}, m)$ is defined conditional on $(I, \bar{I}, \mu)$ to contain the states of nature indexed by ω. A function $w\colon I \times \Omega \to \mathbb{R}^+$ denotes the number of successful gametes per haploid set of individual i in state of nature ω. Arguments of functions that are members of I will be written as subscripts, and those that are states of nature will be written as superscripts, and the same convention will be used for expectations. Thus we write d_i, p_i, and ${w_i}^\omega$. The average number of successful gametes per haploid set in state of nature ω will be written as $w^\omega = \mathbb{E}_i\,[{w_i}^\omega]$, and the relative number for individual i as ${v_i}^\omega = {w_i}^\omega / w^\omega$.

All sets are assumed to have a σ-algebra associated with them, forming a measurable space. All functions are assumed to be measurable on the spaces over which they are defined. Functions over product spaces are assumed measurable with respect to the product σ-algebra (Sch 21.6). The measure on $I \times \Omega$ is assumed to be the natural product measure of $\tilde{\mu}$ and m (Sch 21.40).

The form of the Price equation to be employed shows $\mathbb{E}^\omega\,[\Delta p^\omega]$, the expected change in mean p-score from this generation to the next, in terms of a covariance over individuals, as follows:

$$\mathbb{E}^\omega\,[\Delta p^\omega] = \mathbb{C}_i\,[p_i, \mathbb{E}^\omega\,[{v_i}^\omega]] \tag{2}$$

Grafen (2000) shows that this equation holds under the assumption of perfect transmission, that is, no mutation, no gametic selection, fair meiosis and that all the loci contributing to the p-score have the same mode of inheritance.

The measure space approach allows our notation to apply equally to finite and infinite sets I and Ω, making the assumption only that those sets are capable of having probabilities defined over them. It is often easier to write and read statements about probabilities of sets than integrals or sums.

The state-of-nature structure allows individuals in the population not only to be subject to the occurrence of chance events, but also to be different from each

other. For example, a population can be geographically dispersed in one, two or three dimensions, contiguously or otherwise, location may affect reproductive success, and the effect of different phenotypes may be different in different locations. A population may be a 'metapopulation', with different subpopulations having different properties of various kinds. This broad generality will in the present work be strongly constrained by assumptions made later, but the approach does make the current models more general, and prepares the way for future development.

We make a strong assumption about ω, which is that it does not contain information about the genotypes of individuals. Specifically, it refers to individuals only by their element in I. In describing the outcome of meiosis (for more detailed discussion of the significance of this point, see Grafen, 2000) it refers to the alleles of individual i by a labelling that is independent of the actual alleles present. One implication is that the completed state of nature relies both on ω and on the array of genotypes, and to that extent calling ω the state of nature is an abuse of notation. This separation of genotypes from other factors (or conceivably some other precise statement of their interrelation) is essential for operating with the Price equation at this level of generality.

One key area of this separation of information is that when a function has one argument related to the phenotype and another relating to the state of nature, we know that phenotypic information is not 'leaking' through the state of nature. An assumption of 'pairwise exchangeability' will be introduced that effectively randomises against each other the phenotypes and the local environments. If the local environment contained information about genotype or phenotype, this would result in inconsistencies. The assumption of separation of information is thus of central importance to the whole argument.

Finally, we make the additional assumption that for $i, j \in I$, we have $\mu(\{i\}) = \mu(\{j\})$, and notate it as μ_1. This implies that $\tilde{\mu}(\{i\})/d_i = \tilde{\mu}(\{j\})/d_j$, which we will denote $\tilde{\mu}_1$. Hence $\tilde{\mu}(\{i\}) = d_i\tilde{\mu}_1$. If I is infinite, then $\mu_1 = \tilde{\mu}_1 = 0$.

4 Population genetic models

The Price equation in Equation (2) can be understood as an accounting statement, without any view about causation. We now take such a view by adopting a model of ${w_i}^\omega$. The genotype of an individual is assumed to determine the phenotype, but in a completely unspecified way. The phenotype is a set of potentialities only, and is assumed to affect ${w_i}^\omega$ only through the realised-phenotype, which depends on both the phenotype and on the local environment. Formally, the set of possible local environments is R, and a function $r\colon I \times \Omega \to R$ represents the local environment of individual i in state of nature ω. A phenotype $a_i\colon R \to A$ then allows the realised-phenotype, which is assumed to lie in some set A, to be shown as $a_i({r_i}^\omega)$. We formally consider there to be a function $a\colon I \times R \to A$. The term local indicates only that the relevant features of the environment may be different for each individual, and in particular need have no geographical connotations. Wallace (1990) discusses the history of the concept of 'norm of reaction'.

As well as the realised-phenotype, chance influences will also affect ${w_i}^\omega$. Those that apply to all individuals take a value in some set H and are represented by $h\colon \Omega \to H$. Those that apply separately to individuals lie in some set U, and are represented by $u\colon I \times \Omega \to U$.

Let there be a function $f\colon A \times U \times H \to \{0, 1, 2 \dots\}$ that models the number of successful gametes and so leads to

$$\begin{aligned} {w_i}^\omega &= f(a_i({r_i}^\omega), {u_i}^\omega, h^\omega)/d_i \\ {v_i}^\omega &= \frac{f(a_i({r_i}^\omega), {u_i}^\omega, h^\omega)/d_i}{\mathbb{E}_k\left[f(a_k({r_k}^\omega), {u_k}^\omega, h^\omega)/d_k\right]} \end{aligned} \tag{3}$$

The omissions from f are crucial. The genotypes and phenotypes of individuals in I other than i are not permitted to influence ${w_i}^\omega$, thus ruling out the selection of social behaviour from the reach of the model. (The possibility that we may wish to define f only over a subset of $A \times U \times H$ is ignored here for simplicity.)

The final form of the Price equation for present purposes is then:

$$\mathbb{E}^{\omega}\,\Delta p^{\omega} = \mathbb{C}_i\left[p_i, \mathbb{E}^{\omega}\frac{f(a_i({r_i}^{\omega}), {u_i}^{\omega}, h^{\omega})/d_i}{\mathbb{E}_k\left[f(a_k({r_k}^{\omega}), {u_k}^{\omega}, h^{\omega})/d_k\right]}\right] \tag{4}$$

Summarising, the target of selection is the arithmetic average over all uncertainties of the number of successful gametes of individual i per haploid set, relative to the average over individuals in the realised state-of-nature. The realised-phenotype is allowed to depend on local environment ${r_i}^{\omega}$, and the number of successful gametes is allowed to depend on uncertainties, both general, h^{ω}, and individual, ${u_i}^{\omega}$.

A new assumption of *pairwise exchangeability* is now introduced. Denote by $\mathcal{A}$ the product σ-algebra defined over $H \times (R \times U)^I$ (Sch 9.15 to 9.18), and define $\phi\colon \Omega \to H \times (R \times U)^I$ by

$$\phi(\omega) = (h^{\omega}, ({r_i}^{\omega}, {u_i}^{\omega})_{i\in I})$$

There is a measure α defined on the measurable space $(H \times (R \times U)^I, \mathcal{A})$ imputed by ϕ from the measure space $(\Omega, \bar{\Omega}, m)$ (Sch 9.16). Now consider the set of mappings $T^{j,k}\colon H \times (R \times U)^I \to H \times (R \times U)^I$ for $j, k \in I$ defined by $T^{j,k}(h, (r_i, u_i)_{i\in I}) = (h', (r'_i, u'_i)_{i\in I})$ and

$$h' = h \qquad (r'_i, u'_i) = \begin{cases} (r_j, u_j) & i = k \\ (r_k, u_k) & i = j \\ (r_i, u_i) & \text{otherwise} \end{cases}$$

Thus $T^{j,k}$ simply swaps round the chance events for individuals j and k. The assumption of pairwise exchangeability is that all the mappings $T^{j,k}$ preserve the measure α on $(H \times (R \times U)^I, \mathcal{A})$. Loosely, each pair of individuals has a symmetric distribution of chance events, conditional on the outcomes for all others in the population.

The biological interpretation of exchangeability is that offspring at the stage of censussing for the Price equation are indistinguishable except by their genotype and their phenotype and their consequences. In particular, there are no

historical factors that make individuals different from each other, thus excluding cultural inheritance and differential parental care. A more sophisticated mathematical structure will be required to capture the essential features of these more sophisticated biological ideas. Restricting the set of mappings that preserve α, to make weaker exchangeability assumptions, will be an important technical element in further developments.

Pairwise exchangeability is weaker than independence. For example, it allows individuals to be allocated to physical locations, and for no two individuals to be allocated to the same location. It would also allow individuals to be paired, and then exactly one from each pair to die at random.

This section has defined the 'mechanical processes' of Grafen (1999) in an important class of situations. The range of population genetic models that are included in the Price equation framework is very wide (Grafen, 2000), but the model introduced here has restricted that range considerably. It has already been pointed out that one individual's phenotype cannot affect the number of successful gametes of another. Further, although genomic imprinting is still formally included, the model effectively precludes any interesting selection taking place at an imprinted locus. There is no allowance for mutation. Nevertheless, many simple population genetic models do fall within the scope outlined in this section, and the model contains clear indications of where generalisations need to, and can, be made.

5 Optimisation programs

It is a long tradition in biology to match form and function. The mathematical formulation of purpose is most highly developed in economics and game theory, and optimisation programs will be borrowed and used here to represent the logical barebones of the form-and-function approach. Formal expressions of optimisation programs are used in some parts of economics (see for example the textbook of Mas-Colell *et al.*, 1995), and occasionally in biology (Grafen, 1998).

A generic program is given below and named 'ProgG'. Words consistent with examples in the literature on optimal foraging (Stephens and Krebs, 1986) will be used for concreteness. A strategy x is to be chosen to maximise a currency g, subject to the constraint that x belongs to the set X. The function $g\colon X \to \mathbb{R}$ is called more generally the maximand. ProgG is written as

$$\begin{array}{ll} x & \max\ g(x), \\ & x \in X \end{array} \qquad \text{ProgG}$$

ProgG is adequate to represent simple situations in which it is reasonable to assume that each phenotype always has exactly the same success. In an early example (Charnov, 1976), a forager moves from patch to patch with a deterministic gain function that depends on time in the patch, and a fixed inter-patch travel time, and is assumed to choose the stay time to maximise the rate of gain of energy.

With this simple example, the strategy x belongs to X, the set of non-negative real numbers, and the function $g(x)$ is the gain rate produced by staying in each patch for time x. Optimisation ideas in biology are almost always used with natural selection in the near-background, and there are two relationships that we might hope will hold between this optimisation program and the processes of natural selection. First, if there are various stay-times present in the population, then they will change in relative frequency according to their respective values of $g(x)$. Second, over a longer timescale, if there is a set of possible stay times that occur occasionally by mutation, and if we set X equal to that set, we may hope that the outcome of selection at equilibrium will be equal to (or at least close to) a *solution* to ProgG. A solution is defined as a value x^* such that $g(x^*) \geq g(x)$ for $x \in X$. 'On-going selection' and 'eventual equilibrium' cover the important areas of evolutionary interpretation of the optimisation program.

We develop ProgG to represent a wider range of biological arguments about natural selection, in three steps. Williams (1966) made a distinction between the

effects of design and the effects of chance in determining 'fitness', and used the example of two identical twins, one of whom is struck by lightning and dies. The effects of chance are generally important in biological arguments, such as those about how natural selection acts in the presence of uncertainty. To incorporate uncertainty, we allow the maximand to depend on a random variable, say γ, defined over a measure space $(\Gamma, \bar{\Gamma}, n)$, allowing expectations $\mathbb{E}^{\gamma}$ to be defined. The distribution of γ will have an important role in determining the relative successes of different phenotypes. It is usual to assume in ESS theory and optimisation theory that an individual maximises the arithmetic average over uncertainty, and so we now embody this assumption by making the maximand the arithmetic average of a function of x and γ. Now, $g\colon X \times \Gamma \to \mathbb{R}$ is assumed measurable over γ for each x, and the generalisation to include uncertainty is therefore

$$\begin{array}{ll} x & \max\ \mathbb{E}^{\gamma}\left[g(x,\gamma)\right], \\ & x \in X \end{array} \qquad \text{ProgU}$$

The interpretation of ProgU is that $g(x,\gamma)$ is the number of offspring of an individual playing strategy x when the state-of-nature turns out to be γ, and that an individual maximises the arithmetic average of g through choice of x.

The next step is to allow 'conditional strategies'. The idea is that some aspects of γ may be observable as a cue. Formally, suppose s, the value of some function $\sigma\colon \Gamma \to S$, is observed, then the action chosen may depend on s. The strategy x becomes a function so that $x(s)$ denotes the action taken. The currency of an individual will depend only on the action taken, and not on the parts of the strategy that were not played. We can allow the $x(s)$ to be chosen separately for each s, but we will also allow for constraints, so that an organism cannot react completely differently to different cues. It may be that an organism can make minor behavioural differences according to s, but cannot alter morphology determined before s was observed. A stag may not be able to alter its skeletal size in response to a rainy summer, but may be able to alter

its diurnal feeding patterns.

We will write the complete strategy, with an action chosen for each s, as $x = (x(s))_{s\in S}$. Separate constraints would be represented by $x(s) \in X_s$ for each s. Possibly linked constraints can be represented in a completely general way by

$$(x(s))_{s\in S} = x \in X \subset (X_s)_{s\in S}$$

This requires a formal revision of our definition of the maximand. Formally, letting $\Gamma_s = \{\gamma \in \Gamma \colon \sigma^\gamma = s\}$, now $g \colon \bigcup_s (X_s \times \Gamma_s) \to \mathbb{R}$ depends only on the action taken, and not on the whole strategy.

Cues can also provide information to an individual, and so it is important to allow the distribution over uncertainty as it affects an individual to be different for different cues. The expectation of $g(x(\sigma^\gamma), \gamma)$ with respect to γ conditional on $\sigma^\gamma = s$ is defined (Sch 29.14), and we write it as $\mathbb{E}^{\gamma|s}$.

The program for one particular cue s is therefore

$$\begin{aligned} x(s) \quad & \max \ \mathbb{E}^{\gamma|s} \left[g(x(s), \gamma) \right], \\ & x(s) \in X_s \end{aligned} \qquad \text{ProgUC}$$

If we consider two different values of s, then there are three salient differences between the two cases, which are more easily presented if we temporarily agree to write $\gamma = (s, c)$, a decomposition into observed and unobserved parts of γ. First, the maximand may be different, in that $g(x, (s_1, c))$ and $g(x, (s_2, c))$ may be different functions of x, with quite different maxima. Second, the set of possible actions may be different, as X_{s_1} may not equal X_{s_2}. Finally, the cue may provide only information, so that $g(x, (s, c))$ may depend on c but not on s. Even though the values of g do not depend on s, the expectations $\mathbb{E}^{\gamma|s_1}$ and $\mathbb{E}^{\gamma|s_2}$ may give different probability weightings to the different values of c, and so make $\mathbb{E}^{\gamma|s_1}\left[g(x, (s, c))\right]$ and $\mathbb{E}^{\gamma|s_2}\left[g(x, (s, c))\right]$ quite different functions of x.

If organisms could choose their actions in response to each s quite separately, or if we considered just one s in isolation, then we might hope for the same links

to natural selection for ProgUC as for ProgG or ProgU. In general, however, it is necessary to balance performance in response to different cues against each other, as in the final program:

$$(x(s))_{s\in S} \quad \max \ \mathbb{E}^{\gamma}\left[g(x(\sigma^{\gamma}),\gamma)\right], \qquad \text{ProgUCJ}$$
$$(x(s))_{s\in S} \in X$$

This program represents a moderately sophisticated biological situation in some generality. Cues can determine available strategies but in an arbitrarily linked way, and can also provide information about uncertainty. The value of σ^{γ} is a random variable over which expectations can be defined, and the maximand of ProgUCJ can be written $\mathbb{E}^{s}\mathbb{E}^{\gamma|s}\left[g(x(\sigma^{\gamma}),\gamma)\right]$, showing that the maximands of ProgUC for given s are weighted, when combined into ProgUCJ, according to the probabilities of the different values of s.

ProgUCJ is a formal version of the working hypothesis of many empirical and theoretical biologists when they assume that organisms 'maximise their fitness', which Grafen (1999) called the 'individual-as-maximising-agent analogy'. The main purpose of the next section is to show how to construct an optimisation program in the class outlined in this section from a population genetic model belonging to the class described in the previous section.

6 Construction of a corresponding optimisation program

The population genetic assumptions and quantities of Section 4 are taken as fundamental, and we construct an optimisation program in those terms. At this stage we construct ProgUCJ for one particular individual, namely k. The significance of the construction will become clear in the following section, when links are proved between the two models.

Begin by identifying the measure space $(\Gamma,\bar{\Gamma},n)$ with $(\Omega,\bar{\Omega},m)$. Also identify the set of cues S with the set of local environments R, and σ^{γ} with ${r_k}^{\omega}$ for $\omega=\gamma$.

We assume that $X_s \subset A$ for all s. Table 1 shows some of the correspondences of notation.

The next goal is to define g as a function of $x\colon R \to A$ and γ. To emphasise that we are focussing on individual k, the maximand will be written g_k. Now the kernel of the maximand can be defined in terms of the kernel of the target of selection, setting $\omega = \gamma$, as

$$g_k(x,\gamma) = \frac{f(x({r_k}^\omega), {u_k}^\omega, h^\omega)}{\tilde{\mu}_1(f(x({r_k}^\omega), {u_k}^\omega, h^\omega) - f(a_k({r_k}^\omega), {u_k}^\omega, h^\omega)) + \mathbb{E}_i\left[f(a_i({r_i}^\omega), {u_i}^\omega, h^\omega)/d_i\right]} \tag{5}$$

This is indeed a measurable function of ω as required, because f is a measurable function. We have expressed g_k in a form that makes all the dependence on k explicit, so that had we focussed on individual j instead, we need only substitute j for k in the right hand side.

The distributions of $g_k(x,\omega)$ and $g_j(x,\omega)$ will be the same if the distributions in $\mathbb{R}^3$ of the vectors

$$\begin{pmatrix} f(x({r_k}^\omega), {u_k}^\omega, h^\omega) \\ \tilde{\mu}_1(f(x({r_k}^\omega), {u_k}^\omega, h^\omega) - f(a_k({r_k}^\omega), {u_k}^\omega, h^\omega)) \\ \mathbb{E}_i\left[f(a_i({r_i}^\omega), {u_i}^\omega, h^\omega)/d_i\right] \end{pmatrix}$$

and

$$\begin{pmatrix} f(x({r_j}^\omega), {u_j}^\omega, h^\omega) \\ \tilde{\mu}_1(f(x({r_j}^\omega), {u_j}^\omega, h^\omega) - f(a_j({r_j}^\omega), {u_j}^\omega, h^\omega)) \\ \mathbb{E}_i\left[f(a_i({r_i}^\omega), {u_i}^\omega, h^\omega)/d_i\right] \end{pmatrix}$$

are the same. The mapping $T^{j,k}$ is measure-preserving by the assumption of pairwise exchangeability, and it will therefore suffice to show that $T^{j,k}$ carries one vector into the other. This will be so under two alternative sets of conditions. First, that I is infinite. For then $\tilde{\mu}_1 = 0$, so the middle term is always zero, and both j and k have zero weight in the expectation. Then the first term has indices swapped, and the other terms are unaltered. The alternative condition is that the a_i are all equal. For the first two terms it is then simply a matter of swapping indices. The ploidy-weighting in the expectation cancels with the

division by ploidy to show that the expectation is unchanged, for it equals the number (or density) of successful gametes divided by the number (or density) of haploid sets in I. Swapping the number of successful gametes of individuals j and k changes neither of these quantities.

This is the point at which the exchangeability assumption reduces the population in one framework to the single implicit decision-taker in the other.

Turning to the expectations, the equality of distributions implies immediately that

$$\mathbb{E}^{\gamma}\left[g_k(x(\sigma^{\gamma}),\gamma)\right] = \mathbb{E}^{\omega}\left[\frac{f(x({r_k}^{\omega}),{u_k}^{\omega},h^{\omega})}{\tilde{\mu}_1(f(x({r_k}^{\omega}),{u_k}^{\omega},h^{\omega})-f(a_k({r_k}^{\omega}),{u_k}^{\omega},h^{\omega}))+\mathbb{E}_i\left[f(a_i({r_i}^{\omega}),{u_i}^{\omega},h^{\omega})/d_i\right]}\right] \tag{6}$$

Agree to denote this expectation as $\tilde{g}_k(x)$.

Thus we may write $g_k(a_k({r_k}^{\omega}),\omega)$ for the value of the maximand realised by individual k in state of nature ω, and the expected value as $\tilde{g}_k(a_k)$. The 'what-if' question of what the realised maximand and expected maximand would have been had x been played instead are answered by $g_k(x({r_k}^{\omega}),\omega)$ and $\tilde{g}_k(x)$, respectively. These expressions are values of the maximand in the optimisation framework, but this section has defined them in population genetic terms. They will allow formal links to be established between the two frameworks.

7 Formal links

For any population genetics model included in the framework of Section 4, we have now constructed an optimisation program. In this section, we prove links that justify the application of concepts of optimisation to the population genetics model.

The concepts in terms of which the formal links are constructed must first be introduced. On the population genetic side, 'there is no scope for selection' will mean that $\mathbb{E}^{\omega}\left[\Delta p^{\omega}\right]=0$ for all possible p-scores. This implies in turn that

$\mathbb{E}^{\omega}\left[{v_i}^{\omega}\right]$ is equal for all i (technically, for almost all i). 'There is no potential for positive selection in relation to a set X' will mean that there is no phenotype $x \in X$ which would have been favoured by selection had it been present. More formally, suppose that individuals in the set $J \in \bar{I}$ with $\mu(J) > 0$ had their phenotypes a_i substituted with x, and also define a function q^J to equal 1 for $i \in J$ and 0 otherwise. Then define $\bar{v}_i^{\omega}$ as the values of ${v_i}^{\omega}$ that would result from Equation 3 with that substitution. The absence of potential for positive selection is indicated by the following condition holding for all $x \in X$ and J:

$$\mathbb{C}_i\left[q_i^J, \mathbb{E}^{\omega}\left[\bar{v}_i^{\omega}\right]\right] \leq 0$$

These two concepts are framed entirely in terms of the dynamics of gene frequency change. Now we turn to a concept on the optimisation side. 'a_i solves ProgUCJ for individual i in relation to a set X' means $\tilde{g}_i(a_i) \geq \tilde{g}_i(x)$ for all $x \in X$.

Some of the results to be obtained will depend on the 'same-program' assumption that either I is infinite or the a_i are all equal, in line with the discussion in Section 6. This assumption has the crucial implication that the solution set and value of ProgUCJ will be same for all individuals. If x^* is a solution, then we may write formally that $\tilde{g}_i(x^*) \geq \tilde{g}_i(a_i)$ with equality if a_i is also a solution. If a_i and a_j are both solutions to their respective programs, then the equivalence of the programs provides the central equality to imply that

$$\tilde{g}_i(a_i) = \tilde{g}_i(x^*) = \tilde{g}_j(x^*) = \tilde{g}_j(a_j)$$

Note that these equalities hold in cases in which a_i and a_j are not the unique solutions to their respective programs.

With these concepts, and using the fact that by the definitions of the optimisation concepts, $\mathbb{E}^{\omega}\left[{v_i}^{\omega}\right] = \tilde{g}_i(a_i)$, we straightforwardly obtain the following results. The first and fourth result, and the second part of the second result, depend on the 'same-program' assumption.

1. If each a_i solves ProgUCJ in relation to a set X, then there is no scope

for selection, and no potential for positive selection in relation to the set X.

2. If all a_i attain the same value of the maximand in ProgUCJ, but do not solve it in relation to a set X, then there is no scope for selection, but there is potential for positive selection in relation to the set X.

3. If the a_i attain different values of the maximand in ProgUCJ, then there is scope for selection, and the expected change in every allele frequency and in the additive genetic value of every character equals its covariance across individuals with the attained value of the maximand $\tilde{g}_i(a_i)$.

4. Suppose there is no scope for selection, and no potential for positive selection in relation to a set X. Then each individual acts rationally in the sense that each a_i solves ProgUCJ in relation to the set X.

These properties justify identifying the realised attained maximand and expected attained maximand, $g_i(a_i, \gamma)$ and $\tilde{g}_i(a)$, with realised and expected Darwinian fitness. The first three results derive population genetic conclusions from statements about maximising Darwinian fitness. The fourth shows that absence of selective forces in the entirely mechanical processes of population genetics correspond to behaviour that is consistent with a standard formulation of rationality.

Some readers may also find that the results seem 'obviously true'. However, they are likely to reach this conclusion by taking for granted optimisation ideas about the effect of natural selection. The whole purpose of the argument of this paper is to justify optimisation ideas about natural selection, without assuming them in the first place. The results are in some ways weak, and it will be shown in the next section that some obvious stronger results are not available simply because they are not true. It is only by justifying optimisation ideas without assuming them that their limitations can be established.

It is important to remember that the links established in this section are true for arbitrary genetic architecture, including arbitrary linkage, under the

assumption of perfect transmission. The links aim to represent the strongest results that do hold with that degree of generality.

8 Interpretation of the results

The four linking results found in the previous section establish that certain kinds of optimality ideas apply to the population genetics model. The aim of this section is to consider in a non-technical way what these results mean.

The main result is that each population genetic model satisfying the conditions of Section 4 has a corresponding optimality model of the kind described in Section 5. The particular optimality model is exactly defined from the population genetic model according to the conversions of Section 6. The importance of the existence of this corresponding optimality model lies in the links that it has with the original model. Essentially, it provides an intepretation of some aspects of the population genetic model in optimality terms.

The precise links that permit these interpretations are given directly. To readers who take for granted that population genetic models have optimality interpretations, these links will seem like weak truisms, but it is precisely the intuition of these optimality interpretations that we are in the process of justifying. Interest lies in the fact that the links can be made formally, and also in the limitations of the links that the formal version makes clear.

The first result in informal terms is that 'If each individual acts optimally, then no selection occurs and no potential selection would occur'. Technical points are that if there is more than one global optimum, then the individuals may be playing different optima, and that the lack of selection applies to every locus (more strictly, to every locus that shares the inheritance pattern of the loci in the p-score).

The sense in which no selection occurs is somewhat weak, however. Although the expected changes in gene frequency are all zero, it is quite possible for genotype frequencies to change. One simple example of the exception in-

volves a population which is a mixture of AA and BB homozygotes, where both genotypes produce an optimal phenotype, but where the absent heterozygote does not. Next generation, the result tells us that the gene frequencies of A and B have not changed. However, mating between the two homozygotes would produce suboptimal heterozygotes, and the population would no longer consist of optimal phenotypes. Thus although the result establishes a positive linking result in very general terms, it does not show that once all phenotypes are optimal, the population remains that way.

The second result in informal terms is that 'If each individual acts suboptimally, but equally so, then no selection occurs but potential selection would occur'. The significance is that the link between the optimisation program and the population genetics model do not hold only at solutions to the optimisation. There are also out-of-equilibrium links.

The third result in informal terms states that 'If individuals vary in the value of the maximand they attain, then the expected change in every gene frequency and in the additive genetic value of every character equals its covariance across individuals with the value of the maximand'. Hence all gene frequencies change in line with the value of the maximand attained. This shows that the optimisation program captures important aspects of the course of evolution in completely non-equilibrium situations. By now, the justification for agreeing to call the maximand 'Darwinian fitness' is very strong.

The first three results have argued from an assumption about the optimisation program to a conclusion about gene frequencies. The final result does the opposite, and in informal terms is 'If there is no selection and no potential selection, then each individual in the population acts optimally'. This final result gives a result that follows from the mechanics of inheritance and reproduction, and draws a conclusion that each individual acts rationally in the quite sophisticated sense of the final optimisation program of Section 5. Specifically, the individuals must follow the rules for 'Decision problems with observations' set out in section 8.8 of de Groot (1970)'s text on 'Optimal Statistical Decisions'.

The organisms behave as if following advice for statisticians taking decisions in the presence of uncertainty. Further, the quantity to be maximised is the arithmetic average over all uncertainties of the relative number of successful gametes transmitted to the next generation.

The optimisation program is tightly specified, with the instrument, maximand and constraint set strictly determined by the population genetics model it is based on. Only some of the optimisation analogies made for natural selection, therefore, can be founded on the results of this paper.

This fourth conclusion, too, is subject to the example of the population of AA and BB optimally behaving homozygotes. The condition is satisfied and the conclusion holds, but the population nevertheless evolves away from optimality when heterozygotes appear in the next generation.

It is interesting to consider another contrary case, that of sickle-cell anaemia (Allison, 1954). At the equilibrium with over-dominance, the hypotheses of the first two results do not hold. The third hypothesis holds, but the conclusion is that the covariance of each gene frequency with the maximand equals zero, and so there is no change in gene frequencies. The hypothesis of the final conclusion does not hold, because there is a potential for selection. An allele specifying the same phenotype as the heterozygote would indeed spread in the population. Hence, we rightly fail to conclude that each individual in the population has an optimal phenotype.

The two exceptional cases confirm that the linking results are true, for they hold even then, but they also illustrate that the results do not imply that the outcome of natural selection is inevitably that each individual in the population has an optimal phenotype. These cases incorporate some of the aspects of 'environmental change' as discussed in Section 2.2 in relation to Fisher's fundamental theorem. The results do provide in a very precise way a general background of optimality to the operations of natural selection, and justify the use of intentional terms in relation to the outcome of natural selection. The optimisation program exists and the links hold in the presence of arbitrary ploi-

dies, epistasis, multiple loci and alleles, arbitrary linkage maps, finite or infinite population size, and arbitrary forms of uncertainty. Because the same argument holds across all these cases, it demonstrates that the optimising tendencies are inherent in selection in general, and are not accidental occurrences in special cases.

The sense of the general background is that changes in gene frequencies (though not necessarily in genotype frequencies) change in line with the optimisation framework. It is a plausible view that the designing capacity of natural selection operates through gene frequency changes. The extra factors that contribute to genotype frequencies, including segregation and mating systems, do not contribute to design, and need not therefore enter into an expression of the basic Darwinian argument, however important they may be to a full account of the genetics of a population.

The technical argument as a whole has extended the Price equation approach, and added concepts of phenotype and of how phenotypes determine fitness, but retains the Price equation's complete agnosticism about the link between genotype and phenotype. This has allowed very general but somewhat weak results to be derived about a wide class of population genetic models.

9 Discussion

The most direct use of the theory of this paper is that a population genetic model satisfying the requirements of Section 4 can be used to generate an optimisation program using the equivalences of Section 6. Also, the reasonableness of an optimisation model can be tested by asking whether a corresponding population genetic model can be constructed that would give rise to it. Two obvious possible difficulties are the relative nature of the maximand, and the arithmetic averaging required. Ascending the scale of abstractness, it provides a general population genetic justification for modelling approaches that assume optimisation with a specific maximand claimed as 'natural', such as optimal foraging

theory (Stephens and Krebs, 1986), and can provide guidance for selection of a maximand in difficult cases.

More abstractly, the meta-models show what is meant by the claim that natural selection leads to individuals maximising their fitness, even if the extent of the validity of that claim is still in question. They will also help identify and understand the distinctive features of biological evolution, compared to cultural evolution and other analogues.

This paper has defined an optimisation program that has tight formal links with a reasonably general population genetic model, and allows a definition of Darwinian fitness in certain cases. This represents a validaton of the 'individual-as-maximising-agent' analogy. The key point is that aspects of the population genetic model have been interpreted, formally and rigorously, in terms of optimisation. There is still, however, a considerable way to go to fulfil the 'formal Darwinism unification project' (Grafen, 1999).

With the exception of the restriction to discrete generations, the theory in this paper encompasses Darwin's arguments about the design brought about by natural selection, apart from his hints about family selection in the social insects. Extending the analysis should allow the twentieth century additions to the theory to be incorporated formally, essentially providing a mathematical version of the conceptual and verbal synthesis of Dawkins (1976).

10 Acknowledgments

I am grateful to Professors Charles Batty and Paul Tod for extremely helpful conversations about measure theory, and to Dr Steven Siller for comments on an earlier draft. Professor Steven Frank and an anonymous referee made sympathetic and invaluable comments on the first submitted version.

References

Allison, A. C. (1954). Notes on sickle-cell polymorphism. *Annals of Human Genetics*, **19**, 39–57.

Charnov, E. L. (1976). Optimal foraging, the marginal value theorem. *Theoretical Population Biology*, **9**, 129–136.

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

Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.

de Groot, M. H. (1970). *Optimal Statistical Decisions*. McGraw Hill, New York.

Edwards, A. W. F. (1994). The fundamental theorem of natural selection. *Biological Reviews*, **69**, 443–474.

Ewens, W. J. (1979). *Mathematical Population Genetics*. Springer, Berlin, Heidelberg, New York.

Ewens, W. J. (1989). An interpretation and proof of the fundamental theorem of natural selection. *Theoretical Population Biology*, **36**, 167–180.

Ewens, W. J. (1992). An optimizing principle of natural selection in evolutionary population genetics. *Theoretical Population Biology*, **42**, 333–346.

Falconer, D. S. (1981). *Introduction to Quantitative Genetics (2nd edn)*. Longman, London.

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.

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.

Frank, S. A. and Slatkin, M. (1990). Evolution in a variable environment. *American Naturalist*, **136**, 244–260.

Grafen, A. (1985). A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology*, **2**, 28–89.

Grafen, A. (1998). Fertility and labour supply in *Femina economica*. *Journal of Theoretical Biology*, **194**, 429–455.

Grafen, A. (1999). Formal darwinism, the individual-as-maximising-agent analogy, and bet-hedging. *Proceedings of the Royal Society, Series B*, **266**, 799–803.

Grafen, A. (2000). Developments of Price's Equation and natural selection under uncertainty. *Proceedings of the Royal Society, Series B*, **267**, 1223–1227.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, **7**, 1–52.

Hammerstein, P. (1996). Darwinian adaptation, population-genetics and the streetcar theory of evolution. *Journal of Mathematical Biology*, **34**, 511–532.

Lessard, S. (1997). Fisher's fundamental theorem of natural selection revisited. *Theoretical Population Biology*, **52**, 119–136.

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

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

Price, G. R. (1972a). Extension of covariance selection mathematics. *Annals of Human Genetics*, **35**, 485–490.

Price, G. R. (1972b). Fisher's 'fundamental theorem' made clear. *Annals of Human Genetics*, **36**, 129–140.

Schechter, E. (1997). *Handbook of Analysis and its Foundations*. Academic Press.

Stephens, D. W. and Krebs, J. R. (1986). *Foraging Theory*. Princeton University Press, Princeton.

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

Wallace, B. (1990). Norms of reaction: do they include molecular events? *Perspectives in Biology and Medicine*, **33**, 323–334.

Williams, G. C. (1966). *Adaptation and Natural Selection*. Princeton University Press.

Population Genetics			Optimisation Theory
Meaning	Notation	Notation	Meaning
Population measure space	$(I, \bar{I}, \tilde{\mu})$	(none)	Implicit single decision-taker
General Uncertainty	ω	γ	General Uncertainty
and its Measure Space	$(\Omega, \bar{\Omega}, m)$	$(\Gamma, \bar{\Gamma}, n)$	and its Measure Space
Local Environment	${r_k}^{\omega}$	$s = \sigma^{\gamma}$	Cue
belongs to the set	R	S	belongs to the set
Phenotype	a_k	x	Strategy
Realised Phenotype	$a_k({r_k}^{\omega})$	$x(\sigma^{\gamma})$	Action
belongs to the set	A	$\bigcup_s X_s$	belongs to the set
Target of Selection	${v_k}^{\omega}$	g	Maximand

Table 1: Identities between the notations of Sections 4 and 5. The whole population I is reduced to a single implicit individual, the decision-taker, in the optimisation program.