

# Should we ask for more than *consistency* of Darwinism with Mendelism? \*

Alan Grafen  
Zoology Department and St John's College  
Oxford University

October 10, 2019

---

\*This paper is dedicated to the memory of Prof. Jean Gayon, who died during the preparation of this paper, and with whom I had the honour to serve on a doctoral panel. He believed that participation in science assisted the philosopher of science: the benefits of this approach are increasingly evident.

## Abstract

A nonmathematical exposition of the current status of the formal darwinism project is presented, linking it to the fundamental theorem of natural selection, which is regarded as Fisher’s own ‘formal darwinism project’. The purpose is to found organism-level thinking about design and adaptation, in short *Darwinism*, on what is known about the mechanics of genetic inheritance, in short *Mendelism*, and the project is to do so in as general a biological setting as possible. This view also makes sense of the name ‘fundamental theorem of natural selection’.

The views of a fruitfly about chromosomal rearrangements are of little interest, and it is with some diffidence that a biologist must contribute to a discussion of the philosophy of biology. My purpose will be simply to explain my formal darwinism project, which is referred to by some of the other discussants, in non-technical terms, and at a general level. The primary question is: what relationship should we expect, or indeed hope for, between Mendelism and Darwinism? The formal darwinism project is about that relationship, especially the foundational property (Darwin, 1859) that *natural selection is an improving process*. For a more detailed account of many of the points here that is suitable for a biological audience, with full citations, the reader is referred elsewhere (Grafen, 2007; Batty et al., 2014; Grafen, 2015a; Crewe et al., 2018). The formal darwinism project has already been discussed in a mixed biological and philosophical context (see Grafen, 2014, and associated papers): the main changes since 2014 are the publication of much more rigorous and general mathematical papers (Batty et al., 2014; Crewe et al., 2018), and the greater appreciation of the connection to Fisher (1930)’s fundamental theorem (Grafen, 2015a,b). It is also fair to point the reader to papers that criticise formal darwinism (Lehmann and Rousset, 2014a,b; Ewens and Lessard, 2015), and to which full responses have not yet been published (but see Grafen, 2018).

The principal points in my geometry need some explanation, as follows. In the crudest terms, Darwinism can be regarded as the expectation that natural selection will lead to good organismal design. I will here mean with a slight increase in sophistication the Darwinian side of the Modern Synthesis, the idea of looking at individual organisms, and understanding their design in terms of enhancing an individual’s survival and reproduction, as combined in a single quantity that is a property of an individual and is usually called

‘fitness’. This term is, unsurprisingly, much discussed and its precise meaning is often contested.

The second point is Mendelism, by which I mean the mechanics of inheritance of genes by offspring from parents, together with the population patterns of genotypes. For diploids relevant concepts include segregation, linkage, and recombination, in the mechanics; and gene frequencies and linkage disequilibrium in the patterns. In fact, for the most part, formal darwinism’s results also apply to asexual reproduction, and in general to systems in which the genes an individual passes on are representative of the genes it possesses.

How are Darwinism and Mendelism connected? Historically, we can see four major connections in the minds of biologists. First, those now almost incomprehensible wars between Mendelians and biometricians in the first two decades of the twentieth century, in which both sides agreed that Darwinism and Mendelism were incompatible. Second, Fisher (1918)’s paper ‘on the correlation between relatives on the supposition of Mendelian inheritance’, in which he showed that many small Mendelian effects would, added together, give the patterns of trait correlations that the biometricians (the Darwinians of the day) had observed. Fisher had shown that the particular disagreement was about nothing, and established a consistency between Darwinism and Mendelism. The third connection is the fundamental theorem of natural selection presented by Fisher (1930), only 12 years later. I will return to the role of the theorem but remark here that no one, apart from possibly Fisher himself, understood it at the time anyway. Finally, there is the Modern Synthesis (Huxley, 1942), whose basic tenets included that all the phenomena of evolution could be understood by the known properties of genetics. In particular, adaptation of organisms could be understood by the positive selection of relevant alleles.

Through the Modern Synthesis, the large majority of biologists became Mendelians and Darwinians. No longer contradictory, population geneticists could study Mendelian inheritance and also selection, including the spread of simple traits, mutation-selection balance, and stabilising selection, as well as complications such as epistasis and linkage disequilibrium. Other biologists studied the traits of organisms and employed Darwinian ideas of design, confident that although they didn’t know the genetics, their ideas would be supported by genetic models. But, this confidence was not justified for long, if indeed it ever was.

Certainly, as soon as biologists began to look seriously at social behaviour, or life-history theory, they strayed beyond what population genetics could do for them. They needed inclusive fitness, developed by Hamilton (1964), and reproductive value, introduced by Fisher (1927), but not incorporated into mathematical population genetics for some time (it does not appear in the index of the textbook of Ewens (2004), nor in its brief discussion of age-structured populations), to get to the design ideas they wanted, and this broke the bounds of what population geneticists at the time could provide. Thus, Darwinians needed more than the Modern Synthesis provided.

What connections would we like to see today between Mendelism and Darwinism? In principle, there is no doubt that Mendelism is basic, and Darwinism needs to be derived from it, in the same way as chemistry from physics. So, what kinds of things would we like to derive from Mendelism?

Two elementary points are:

1. Selection on a given trait will in general act in the same direction at different loci.
2. Selection on different traits will act together to produce concerted organismal design.

Most biologists who study organisms in the field take these so much for granted that they would not recognise them as problematic. They allow an organism-level thinking that is second nature to many biologists, and they lay the ground for two further desirable points, which are

3. To characterise in Mendelian terms the nature of Darwinian design. This is most simply done by defining a quantity we can call ‘fitness’, that attaches to an individual and not to a genotype, or a specific locus. To make such a definition, we would need to prove that selection is suitably connected to the fitness of the phenotypes of individuals. ‘Suitably’ means that gene frequency changes and phenotypic changes are linked to the fitnesses of individuals.

4. In particular, we would like to justify Darwin’s claim that natural selection is an improving process, and define the quantity called ‘fitness’ in such a way as to capture the sense in which natural selection improves.

Fulfilling all four points is what I propose should count as founding Darwinism on Mendelism, but are these four points actually true – should we be trying to fulfil them at all?

1 and 2 are not universally true. One simple counter-example arises when there is genetic conflict between different coreplicons (Cosmides and Tooby,

1981). A coreplicon is a set of loci that have the same inheritance pattern: four coreplicons in mammals are loci on autosomes, X-chromosomes, Y-chromosomes and mitochondria. The loci in a coreplicon also need to have the same relatednesses to other individuals, so that, for example, genomically imprinted loci are in a different coreplicon from non-imprinted, and paternally and maternally imprinted loci form separate coreplicons. A more advanced exception arises in social behaviour in an inbreeding population. The relatedness of sibs is different for different alleles, depending on their frequency, and so the selection of alleles at the same locus can be acting in conflicting directions on social traits. Another cause of problems is diploidy, and dominance. Selection can act with different strengths depending on dominance and depending on gene frequencies. Over-dominance can make it unclear what is meant by ‘direction of selection’. So, genetic complications can make the project look unpromising. However, if there is a dominant coreplicon, in the sense that it contains many more loci than the other coreplicons, judged as a ratio, then it is probably reasonable to ignore genetic conflict for most traits. This is because we expect (though this is an expectation not a theorem) that if the non-imprinted autosomes have 100 times more loci than the X-chromosome, then more relevant mutations will arise on the autosomes, there will be more genetic variation on the autosomes, and so selection on them is likely to overwhelm selection on the much smaller coreplicon. If selection acts on one coreplicon with coefficients of relatedness that are only very weakly frequency-dependent, then both points 1 and 2 will be true. Inbreeding is not usually very high, and so those complications may not have much impact. Note, though, that in continuing from points 1 and 2 to points 3 and 4, we are working with an idealisation.

I have recently come to regard the first attempt to carry out this project as Fisher (1930)’s fundamental theorem of natural selection, and now explain why. The statement of the theorem was long completely misunderstood by everyone except Fisher. Shockingly, Huxley (1942) does not mention the fundamental theorem of natural selection in the founding work of the Modern Synthesis. Later, it was essentially derided as false in its misunderstood form; then, following Price (1972), Ewens (1989) and others, the theorem after all became true, but of little discernible biological value. This is the standard position of Price (1972), Ewens (1989), Lessard (1997) and Ewens (2004), as well as Edwards (1994, 2014).

What about the fundamental theorem suggests that is an attempt to

found Darwinism on Mendelism? There are interesting circumstantial details: it would seem natural for Fisher, having proved *consistency* in 1918, to wish to prove something stronger 12 years later. An interesting note is that the fundamental theorem uses the statistical concept of ‘variance’, a term Fisher had himself invented in that same 1918 paper. A useful textual study of *The Genetical Theory* would consider what the location and uses of the theorem in the book make clear about Fisher’s own beliefs about the theorem: the name he gives it, and the comparison with the Second Law of Thermodynamics, are strongly suggestive. The main argument here, however, is that Fisher proposes (though implicitly) two very precise definitions of common terms, namely ‘due to natural selection’ and ‘fitness’, and proves a very general and pregnant result with those definitions. The theorem states that the rate of change in mean fitness due to natural selection equals the additive genetic variance in fitness: the second of these quantities cannot be negative, and the theorem therefore implies that the first cannot be negative. This looks very much like a demonstration that natural selection is an improving process, provided the definitions are reasonably consistent with general biological usage of the terms. This demonstration would be such a significant result that in my view it would be worth biology adopting those definitions as the precise quantitative versions of the informal ideas, if at all possible.

Let us look at those two definitions, starting with ‘fitness’. In an age-structured population, each age  $x$  has a reproductive value, which we may call  $v_x$ . For simplicity, we switch to considering discrete time. G.C. Williams is a key figure here, and it is important that this line of argument links significant figures in non-mathematical thinking about evolution with the concept of fitness. Williams (1966), under the impression he is following Fisher, defines the reproductive value of an individual as the sum of the age-specific reproductive values  $v_x$  for all descendants next period. Those descendants include offspring, who will all contribute  $v_0$ , and also his surviving self, which will contribute  $v_{x+1}$ , weighted by the chance of survival. This ‘Williams’ reproductive value’ is a property of an individual, and depends on the outcome of chance events and on the actions of the individual that may be influenced by its genotype, and we will denote it  $W_i$  for individual  $i$ . It is not same as the age-based reproductive values we have already denoted as  $v_x$ , which depend just on age and are calculated from aggregate population data, although the two are obviously intimately related. Then Williams suggests

that an individual at each age chooses actions (is led by natural selection to act as if choosing actions) to maximise  $W_i$ , from within its set of feasible actions. It is easy to regard  $W_i$  as fitness, especially as it uses Fisher's  $v_x$ . However, Fisher actually defined fitness for individual  $i$  in age-class  $x$  as

$$\frac{W_i - v_x}{v_x} \tag{1}$$

So long as  $v_x > 0$ , maximising one is the same as maximising the other, so for some purposes this discrepancy is inessential. Fisher chooses this definition because he needs it to get the link between Darwinism and Mendelism to work. It has very convenient technical properties that we pass over now (Grafen, 2015a). My point here is that he chooses this definition because founding Darwinism on Mendelism is his project: unless he defines fitness this way, natural selection will *not* increase mean fitness. (In fact the division by  $v_x$  is necessary because for the whole theory to work all averages are weighted by  $v_x$ . Mean fitness is therefore calculated by multiplying fitness by  $v_x$ , adding up, and dividing by the sum of  $v_x$ . The weighted mean of fitness therefore involves adding up the  $W_i$  themselves.)

The second odd definition is of natural selection or, more precisely, ‘due to natural selection’. More generally, Fisher explicitly defends the idea that natural selection is represented by changes in gene frequencies and in particular not by changes in genotype frequencies (made clear in a letter to O. Kempthorne (Bennett, 1983, p. 228), and quoted on page xiii of Bennett’s introduction to Fisher (1999)). This definition of ‘due to natural selection’ thus has repercussions for precisely what we understand by ‘natural selection’ itself. Specifically, he takes the total change in mean fitness, and divides it into two: the part due to natural selection, and the rest, which he ascribes to the environment. The definition is to do with changes in mean fitness brought about by changes in gene frequencies (not genotype frequencies), taking average effects as fixed. The details are very important when deriving the result, but need not detain us here. Price (1972) was the first to point out the ‘partial’ nature of the result, and he and later derivors of the theorem agonise about this definition and can’t make anything interesting of it. My suggestion is that this is the definition he needs to make in order to get the theorem to work. We understand Fisher if we think: ‘It’s amazing the fundamental theorem can give us such a simple connection between Darwinism and Mendelism with these definitions. Are these definitions acceptable? If

they do no violence to the concept of natural selection and the concept of fitness, then these would be very natural definitions to adopt.’ It is early days to pass final judgment on the question of doing violence. In Grafen (2015a), I derive the theorem and explore the definitions, and so far as I can see those definitions of natural selection and fitness are perfectly acceptable definitions. I am unaware of any other attempt to be mathematically precise about ‘due to natural selection’ within population genetic models. While working with examples, the question hardly arises. But if we take seriously the question of whether natural selection is an improving process *in general*, we do need such a definition. My impression is that general reasoners about natural selection have shied away from quantitative precision, and population geneticists have shied away from general questions about natural selection. Putting general reasoning together with quantitative precision seems to me at the heart of Fisher’s analytical treatment of Darwinism, and something worth exploring.

This may be the right place to recognise the treatment of the fundamental theorem by Gayon (1998), whose magisterial book on the history of Darwinism all biologists would benefit by reading: there were many significant surprises for me. There is also a philosophical literature on whether natural selection is a process or a product (see e.g. Millstein, 2017), which aims at a different kind of precision about ‘due to natural selection’ from that referred to in the previous paragraph.

The recognition that formal darwinism and the fundamental theorem are attempts at the same problem has led to some significant changes in the basic approach in formal darwinism. The earlier approach can be usefully caricatured as follows: make an inspired guess at a quantity that might represent fitness, then show it does represent fitness by proving sufficiently strong links between an equilibrium of gene frequency dynamics and the solution to an optimisation program with the proposed maximand. The new approach follows Fisher very closely, but in a more general model. Define the per-capita reproductive value of classes as the expected fraction of the distant gene pool that is descended from the average individual in that class, and then fitness following equation (1); then prove two results using that definition of fitness. First, a Price Equation, that shows that the expected frequency of every gene depends on survival and reproduction only through the expectation of fitnesses of individuals over all uncertainty. Second, a fundamental theorem showing that the expected change in the mean of fitness that is due



to natural selection equals the additive genetic variance in expected fitness. (It is important to recall that we are considering genes that belong to one coreplicon only.)

This change in approach has various advantages. It links to Fisher’s original argument very directly; it avoids the need to make an educated guess; the educated guess was in any event usually Williams’ reproductive value (notated as  $W_i$  above) and so used unweighted means of fitness, while the new approach can employ reproductive-value-weighted means, which makes for consistency within the theory. The change in approach also faces challenges: the new interpretation of the fundamental theorem is by no means universally accepted (e.g. Ewens and Lessard, 2015). and this is likely to inhibit acceptance of its extension. The earlier approach explicitly justified fitness-maximisation in a genetically special case, while now there is no need for a detailed genetics: on the other hand, the argument for fitness-maximisation is now parallel to that from the fundamental theorem case, which as just noted is far from accepted.

Formal darwinism in its most recent form (Crewe et al., 2018) deals with a finite class-structured population, with stochastic environments and continuing fluctuations in the class-distribution of the population. It fulfils points 1 to 4 by proving a Price Equation and a fundamental theorem, which are remarkably similar to the deterministic versions. Essentially we place an expectation symbol around ‘change in gene frequency’ (for the Price Equation) and ‘change in mean fitness’ (for the fundamental theorem), on the left hand sides, and then place an expectation around ‘fitness’ on the right hand side. In simple terms, we operate with averages, and the truly remarkable fact is that the theorems can be proved when we do so. The Fisherian framework passes simply into averages. By contrast, other approaches to populations with continuing fluctuations in class-distribution (e.g. Tuljapurkar, 1989) invoke leading Luyapunov exponents and provide no simple results about selection. As Darwinian biologists, both empirical and theoretical, routinely deal with averages in such cases, this is a very welcome result.

It is fair to point out that a device called the ‘Taylor switch’ is employed by Crewe et al. (2018), which generalises from work of Taylor (1990, 1996). This device assumes that genotypes affect phenotypes in only one time period, which makes the changes in gene frequency much easier to assess. It would be useful to have some work exploring the difficulties created by this artifice, specifically looking for situations in which the equilibrium

of an explicit genetic model is not what would be predicted on the basis of this ‘one-time-period-only’ device. If these situations are common, this diminishes the value of these formal darwinism results. On the other hand, it will also be possible to look for analytical or simulation results that establish that the device does not mislead across a sufficiently wide range. It is worth noting that Fisher’s fundamental theorem suffers from essentially the same problem, and that work at this level of abstraction may not be able to avoid some similar device, essentially as a consequence of the dynamic insufficiency of the genetic side of the modelling (a point further discussed by Crewe et al., 2018).

At the heart of formal darwinism lies the conviction that Darwin was right that natural selection is an improving process, and that it is worthwhile to articulate this statement as generally as possible. The resulting body of theory in formal darwinism does not displace the many other literatures about evolution and natural selection, but aims to tie together some very high level abstractions that can be usefully summarised as ‘founding Darwinism on Mendelism’. It was not understood at the inception of formal darwinism (Grafen, 1999) that it continues a project of Fisher, which is encapsulated in his fundamental theorem. This new understanding has helped further work that has adopted Fisher’s definitions (to date, specifically Grafen, 2015b; Crewe et al., 2018) ; it has also helped to make sense of the fundamental theorem itself, of the significance attached to it by Fisher (1930), and particularly of the name ‘fundamental theorem of natural selection’.

## References

- Batty, C. J. K., Crewe, P., Grafen, A., and Gratwick, R. (2014). Foundations of a mathematical theory of darwinism. *Journal of Mathematical Biology*, 69:295–334. doi: 10.1007/s00285-013-0706-2.
- Bennett, J. H., editor (1983). *Natural Selection, Heredity and Eugenics (Including Selected Correspondence of R.A. Fisher with Leonard Darwin and others)*. Oxford University Press, Oxford.
- Cosmides, L. M. and Tooby, J. (1981). Cytoplasmic inheritance and intragenomic conflict. *Journal of Theoretical Biology*, 89:83–129.

- Crewe, P., Gratwick, R., and Grafen, A. (2018). Defining fitness in an uncertain world. *Journal of Mathematical Biology*, 76:1059–1099.
- Darwin, C. R. (1859). *The Origin of Species*. John Murray, London.
- Edwards, A. W. F. (1994). The fundamental theorem of natural selection. *Biological Reviews*, 69:443–474.
- Edwards, A. W. F. (2014). R.A. Fisher’s gene-centred view of evolution and the Fundamental Theorem of Natural Selection. *Biological Reviews*, 89:135–147.
- Ewens, W. J. (1989). An interpretation and proof of the fundamental theorem of natural selection. *Theoretical Population Biology*, 36:167–180.
- Ewens, W. J. (2004). *Mathematical Population Genetics I. Theoretical Introduction*. Springer, Berlin, Heidelberg, New York.
- Ewens, W. J. and Lessard, S. (2015). On the interpretation and relevance of the fundamental theorem of natural selection. *Theoretical Population Biology*, 104:59–67.
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52:399–433.
- Fisher, R. A. (1927). The actuarial treatment of official birth records. *Eugenics Review*, 19:103–108.
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford University Press. See Fisher (1999) for a version in print.
- Fisher, R. A. (1999). *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford, UK. Variorum Edition of 1930 OUP edition and 1958 Dover edition, edited by J. Henry Bennett.
- Gayon, J. (1998). *Darwinism’s Struggle for Survival*. Cambridge University Press, Cambridge, UK.
- 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. (2007). The formal Darwinism project: a mid-term report. *Journal of Evolutionary Biology*, 20:1243–1254.
- Grafen, A. (2014). The formal darwinism project in outline. *Biology and Philosophy*, 29:155–174.
- Grafen, A. (2015a). Biological fitness and the fundamental theorem of natural selection. *American Naturalist*, 186:1–14.
- Grafen, A. (2015b). Biological fitness and the Price Equation in class-structured populations. *Journal of Theoretical Biology*, 373:62–72.
- Grafen, A. (2018). The left hand side of the Fundamental Theorem of Natural Selection. *Journal of Theoretical Biology*, 456:175–189.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, 7:1–52.
- Huxley, J. S. (1942). *Evolution: the Modern Synthesis*. Allen and Unwin, London.
- Lehmann, L. and Rousset, F. (2014a). Fitness, inclusive fitness and optimization. *Biology and Philosophy*, 29:181–195.
- Lehmann, L. and Rousset, F. (2014b). The genetical theory of social behaviour. *Philosophical Transactions of the Royal Society of London, Series B*, 369:20130357.
- Lessard, S. (1997). Fisher’s fundamental theorem of natural selection revisited. *Theoretical Population Biology*, 52:119–136.
- Millstein, R. L. (2017). Genetic Drift. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2017 – <https://plato.stanford.edu/archives/fall2017/entries/genetic-drift/> edition.
- Price, G. R. (1972). Fisher’s ‘fundamental theorem’ made clear. *Annals of Human Genetics*, 36:129–140.
- Taylor, P. D. (1990). Allele-frequency change in a class-structured population. *American Naturalist*, 135:95–106.

- Taylor, P. D. (1996). Inclusive fitness arguments in genetic models of behaviour. *Journal of Mathematical Biology*, 34:654–674.
- Tuljapurkar, S. (1989). An uncertain life: Demography in random environments. *Theoretical Population Biology*, 35:227–294.
- Williams, G. C. (1966). Natural selection, the costs of reproduction, and a refinement of Lack’s principle. *American Naturalist*, 100:687–690.