University of Bristol - Explore Bristol Research
General rights

Adaptation, Fitness and the Selection-Optimality Links
Samir Okasha (University of Bristol)
Cédric Paternotte (LMU University)

1. Introduction

We are broadly sympathetic to the aims of the formal Darwinism project. Grafen's basic point – that the notion of design or adaptation must somehow be formalized if Darwin's argument that selection leads to design is to be placed on a secure foundation – is interesting and important. The use of optimization programs to capture the notion of design could perhaps be questioned (cf. J. Birch's contribution to this issue), but seems a reasonable starting point. Grafen's idea of proving links between optimization and gene frequency change also makes good sense to us. In what follows we raise a number of queries about the formal Darwinism project. Some of these are technical points, while others concern the project's implications and significance.

2. A Foundation for Behavioural Ecology?

In both the target paper and his previous papers, Grafen describes his project as providing a formal vindication of the practice of behavioural ecologists and others who assume fitness-maximization in their day-to-day biological work. However, it is not clear to us how his 'selection-optimality links' actually contribute to this goal. The behavioural ecologists' credo that evolved phenotypes will appear designed to maximize an organism's fitness (personal or inclusive) seems predicated on two assumptions: firstly, that the range of available phenotypes from which natural selection chooses is not heavily constrained; and secondly that genetic architecture will not prevent frustrate the fixation of the 'best' phenotype in the population.

Grafen's selection-optimality links do not seem to vindicate either of these assumptions. In proving these links, Grafen operates with an abstract feasible set, representing the phenotypes available to natural selection. Nothing is assumed about how big this set is, nor about how different the phenotypes are from each other. Thus Grafen's analysis is compatible with the presence of strong developmental constraints on the available phenotypic variation or with the complete absence of such constraints; in either case, his links hold true. But the fitness-maximization paradigm in behavioural ecology, and the adaptationist programme more generally, are premised on the assumption that developmental

constraints do not greatly limit the available variation; for if they did, then well-designed phenotypes could not evolve.

Turning to genetic architecture, consider the simple case where the optimal phenotype in a population is produced by a heterozygote. Since heterozygotes do not breed true, selection will not lead the optimal phenotype to be fixed in the population, i.e. sub-optimal phenotypes will be present at the population-genetic equilibrium. Adherents of fitness-maximization are therefore making the empirical bet that such situations are fairly rare, or that natural selection will eventually manage to alter the genotype-phenotype map to circumvent the problem. But as Grafen observes, his selection-optimality links hold true in the case of heterozygote superiority. Thus it is hard to see how proving those links helps to vindicate the assumption of fitness-maximization made by behavioural ecologists; since the links would be true even in a world where genetic constraints were ubiquitous and the fitness-maximization assumption untenable.

The upshot, it seems to us, is that Grafen's selection-optimality links do not quite do the job that he asks of them. The links establish a connection between natural selection and optimization; and it is true that behavioural ecologists and others assume a connection between these two things. But the connection that the latter assume is not the same as the connection that Grafen's links establish. In effect, behavioural ecologists assume that if natural selection is the only evolutionary force at work, it will lead to well-designed phenotypes; but Grafen's links do not imply that this is true, nor that it is a reasonable approximation to reality.

Grafen might reply that the stronger selection-optimization connection that behavioural ecologists posit does not always obtain, so cannot possibly be proved mathematically. This is indeed so; and it may be that Grafen's links are the most that can be proved in a fully general setting. Still, given the considerable distance between the selection-optimality connection that Grafen establishes and the one that behavioural ecologists posit, it becomes hard to see why the formal Darwinism project should be thought to provide any support for the fitness-maximization assumption on which many research projects in behavioural ecology are premised.

To be fair to Grafen, his target paper partly addresses this worry. He describes cases in which fitness-maximization doesn't occur but his links still hold true, such as the heterozygote superiority case, as 'exceptional'. But it is not clear on what basis this judgement is made. After all, epistatic interactions in general are quite common, between and within loci, so fitness epistasis cannot be assumed negligible. It might be argued that the

success of those research programs that assume fitness maximization shows that fitness epistasis must be the exception not the rule. This is quite plausible; but the research programs in question would then be justified by their empirical success, not by Grafen's links.

In his target paper, Grafen introduces a contrast between a selection-optimality link holding 'trivially' and 'substantively', where the former means that the link holds because its antecedent is false. (Recall that the links are all conditional statements of the form 'if *x* then *y*'.) Grafen observes that in the problem cases where fitness is not maximized, such as heterozygote superiority, some of the links only hold true trivially. He then suggests that the needs of the behavioural ecologists could be met if the links hold "in a significant way" (p.12).

This seems to imply that the trivial / substantive distinction coincides with the distinction between 'exceptional' cases where selection does not lead fitness to be maximized and 'ordinary' cases where it does. However this seems wrong. For note that in *every* case, two out of three of links 1, 2 and 3 must hold trivially, i.e. they cannot all hold substantively. The reason is simple: the antecedents of these three links constitute a partition of logical space, so exactly one of them is true in any case. Thus the trivial / substantive distinction, though well-defined, seems unlikely to help in characterizing the 'exceptional' cases that behavioural ecologists are assuming to be rare.

In stressing these points, we are not saying that Grafen's selection-optimality links are devoid of value. Two possible uses for these links suggest themselves, both of which Grafen hints at. One is to use the links to study the conditions under which natural selection *will* lead to fixation of the optimal phenotype from the feasible set; this would then help isolate the empirical assumption that behavioural ecologists and others are implicitly making. A second, more philosophical use is to regard the links as constituting an axiomatic theory of fitness.

3. An Axiomatic Theory of Fitness?

There is a substantial literature, in both biology and philosophy of biology, on how to 'define' fitness. Indeed this topic is rather notorious. The debate is beset by unclarity about what requirements a suitable definition should have to meet, about whether 'fitness' should be predicated of individuals, genotypes, or something else, and about how to deal with a myriad of complications including stochastic variability, clonality, overlapping generations and social behaviour. As a result, some authors appear to regard the 'correct' definition of fitness as if it were a matter of philosophical taste, to be debated verbally, rather than

something amenable to demonstrative argument. Clearly this is an undesirable situation, given the pivotal role of the fitness concept in biology.

A natural remedy is to resort to the axiomatic method. Instead of trying to define fitness from the outset, why not instead search for axioms, or conditions, that the concept of fitness must satisfy? This idea has been mooted in the past, notably by M.B. Williams (1970), but not successfully executed. (See also Wagner (2010) and Okasha, Weymark and Bossert (forthcoming)). We suggest that Grafen's five links can be used to resuscitate the axiomatic approach to fitness, and indeed the formal Darwinism project as a whole goes a long way to successfully implementing it, in particular Grafen's (2006) argument that inclusive fitness is the quantity that individuals will behave as if they are maximizing in social contexts.

For this particular aim, the fact that Grafen's links are too weak (arguably) to supply a foundation for behavioural ecology is not to the point. If anything it is an advantage. For in general, an axiomatic characterization of some concept is more interesting the weaker the axioms that it employs. In any case, even if we are unwilling to grant that Grafen's selection-optimality links constitute a complete characterization of the concept of Darwinian fitness, they are very plausibly necessary conditions.

When the selection-optimality links are viewed this way, as constituting an axiomatic theory of fitness, Grafen's admission that the uniqueness of the maximand (or objective function) is only 'tentative' (p.12) might seem like a rather large problem. How can one have a genuine axiomatization without a uniqueness proof? However, here an important distinction must be drawn. For in fact it follows from link 3 alone that the objective function is unique up to addition of a constant.[1] If we suppose that two objective functions F and G both satisfy link 3, then we can infer immediately that $b_{Gp} = b_{Fp}$, i.e. the linear regression of G on a given p-score must equal the linear regression of F on that p-score. This can only hold true for all p-scores if $b_{GF} = b_{FG} = 1$, which implies that $F = a + G$, for some constant a.

When Grafen describes the uniqueness of the objective function as a tentative conjecture, he does not mean that there is any uncertainty surrounding the class of transformations that can be applied to the objective function while preserving the truth of the links. There is no such uncertainty. What he means, rather, is that it is uncertain whether there might be some biological property, distinct from fitness itself, which also satisfies the five links. As he says, this seems improbable; but the key point is that even if there is such a

---

[1] This sort of uniqueness is known as 'translation-scale measurability' in the economics literature.

property, the links themselves imply that it must be related to fitness by addition of a constant.

4. *Individual versus Population-level Optimization*

Grafen (2007) makes the important distinction between maximization of individual and of population-level variables. In his previous papers on Formal Darwinism, Grafen makes clear that the former sort of maximization is what interests him, hence his talk of the 'individual as maximizing agent' analogy. This focus is also apparent in the target paper's closing paragraph, in which Grafen notes the philosophical implications of the project, namely its potential to shed light on the nature of individual agency and goal-directed behavior. Thus it is individuals that are doing the (as if) optimization in Grafen's models. By contrast, results such as Fisher's FTNS concern the tendency of a population-level parameter (mean fitness) to increase and thus move towards a maximum (even if that maximum itself changes constantly).

The target paper somewhat obscures this distinction by locating the Formal Darwinism project within a long tradition of works dealing with 'fitness-maximization', some of which, including Fisher's, deal with mean population fitness. This tradition notwithstanding, we think that Grafen (2007) was correct to emphasize that individual- and population-level maximization are distinct and even independent topics. First, it is well-known in evolutionary game theory that individual fitness maximization need not entail mean fitness maximization. In the evolutionary versions of 'social dilemmas' such as the Prisoner's Dilemma, the Stag Hunt or public goods games, the behaviour of optimizing agents typically leads to collectively suboptimal outcomes, in which mean fitness is not maximal. Second, mean fitness can easily be maximized even if agents fail to maximize their personal fitness. Again, evolutionary game theory provides examples such as asymmetric coordination games, in which one equilibrium guarantees the highest mean fitness although at least one type of agent would be better off in another equilibrium.

These observations are not meant to undermine any attempt to establish links between individual- and population-level maximization, but to show that such results will depend on specific situations and cannot be established in full generality. Therefore it is important to keep in mind that the literature on the natural selection-optimization connection is Janus-faced; its two strands should arguably be kept separate.

5. *Formulation of the Links*

At the risk of nit-picking, we observe that Grafen's formulation of the selection-optimality links in his target paper differs slightly from the formulation in his previous papers. In particular, link 3 in the target paper reads: "if individuals attain different values of the maximand, then the change in every gene frequency equals its covariance across individuals with those attained values" (p. 10). However in Grafen (2002 p.88) the corresponding link is stronger: it says in addition that if individuals attain different values of the maximand, then there is some gene (*p*-score) that will change in expected frequency, i.e. that there is 'scope for selection' in Grafen's phrase.[2]

This difference actually matters quite a lot. The stronger formulation of link 3 is intuitively preferable, as it forges a much closer link between selection and optimization – it says that if sub-optimal individuals are present, then the population cannot be in genetic equilibrium, i.e. gene frequencies must be changing. However at first blush this claim is simply untrue, a counterexample being provided by the case of heterozygote superiority. In that case, the population evolves to a polymorphic equilibrium at which sub-optimal individuals (homozygotes) are present, but gene frequencies are unchanging. How then can Grafen (2002) argue that there is 'scope for selection' whenever some individuals fail to solve the optimization program?

The answer is that Grafen understands the phrase 'scope for selection' in an unusually wide sense, to mean that no *possible* gene, or *p*-score, will change in frequency (Grafen, personal communication). On this understanding, at the polymorphic equilibrium resulting from heterozygote superiority there *is* scope for selection; as owing to the fitness differences between individuals, it is logically possible that, at some locus, there is some allele that will change in frequency from one generation to the next. Understood this way, there can only be 'no scope for selection' if all individuals attain equal values of the maximand.

This may seem like a rather devious way of forging a link between selection and optimality. A more natural understanding of 'scope for selection' would restrict attention to genes actually found in the population, rather than including conceivable but non-actual ones. In Okasha and Paternotte (2013), we adopted this approach, and found that it produced interestingly different results, especially in the relation to the 'group adaptation' issue discussed in Gardner and Grafen (2010).

---

[2] The expression 'scope for selection' is not used explicitly in the target article. It is define in Grafen (2002, 2006), though see Okasha and Paternotte (2013) for an important qualification.

Given the slightly peculiar construal of 'scope for selection' that is needed to derive the stronger version of link 3, it is noteworthy that Grafen has opted for the weaker version of that link in his target paper, by omitting the problematic claim that if sub-optimal individuals are present then there must be scope for selection. In effect, this means that links as stated in the target paper are compatible with the 'actual' definition of 'scope for selection' studied by Okasha and Paternotte (2013).

It is not our intent to ask which formulation of link 3 is 'correct', or which construal of the expression 'scope for selection' is better. These are not good questions to ask, particularly if we view the five links as part of an axiomatic theory of fitness, as outlined above. The interesting task, rather, is to determine what consequences follow from the different ways of formulating the links. Nonetheless it would be interesting to hear Grafen's thoughts on these matters of formulation.

6. *Final Thoughts*

Two final issues deserve mention. Firstly, it is interesting to compare the formal Darwinism project with Fisher's fundamental theorem of natural selection (FTNS), which was also an attempt to say something general about the relation between natural selection and optimization. On the modern 'partial change' interpretation of Fisher, explained by Price (1972), the FTNS expresses a general truth, but does not show that natural selection will lead to maximization of mean fitness. At most, it shows that there is a 'tendency' of natural selection to increase mean fitness, in the sense that if the environment is constant, mean fitness will increase. Notoriously though, 'environmental constancy' must be understood in a highly specific way for this assertion to be true, as constancy of the average effects (on fitness) of all alleles in the population. Unless the genetics is perfectly additive, these average effects will depend on gene frequencies, which natural selection changes; so the 'tendency' for mean fitness to increase need not be realized.

Like the FTNS, Grafen's project also aims to say something general about the relation between selection and optimization. (Though it is individual optimization, not optimization of mean fitness, that is at stake.) Like the FTNS, Grafen's project falls short of showing that selection will lead to optimization and for similar reasons, since in both cases non-additive genetics and/or frequency-dependence are the root causes of the problem. We suggest that Grafen's results achieve both more and less than Fisher's. One the one hand, the FTNS detects a maximizing 'tendency', and directly isolates the condition which must obtain if natural selection is to lead to fitness optimization (constancy of average effects); while

Grafen's project does not. On the other hand, the FTNS does not deal with the case of social behaviour, or with uncertainty, and does not explicitly incorporate the idea of individuals 'trying' to maximize an objective function. In these respects, Grafen's project constitutes an advance.

Secondly, it is interesting to ask how Grafen's ideas might be extended to a game-theoretic setting. A natural suggestion is that the replicator dynamics will supply the 'mathematics of motion', and the ESS concept, or perhaps the weaker Nash equilibrium concept, will supply the 'mathematics of optimization'. One immediate complication is that the replicator dynamics do not always take the population to an ESS, as is well-known (cf. Hofbauer and Sigmund 1998) – which suggests that selection-optimality links will be hard to come by in a game-theoretic setting. The more recent literature on adaptive dynamics reinforces this conclusion, as it teaches us that ESS states, even when they exist, need not be attractors of the underlying dynamics, even when genetic constraints are idealized away (cf. Metz et. al. 2008). *Prima facie*, it thus seems unlikely that the 'individual as maximizing agent' analogy can be applied in a game-theoretic context. It would be interesting to hear Grafen's thoughts on this issue.

References

Gardner and Grafen (2010). Caputring the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology* 22(4):659-671.

Grafen, A. (2002). A first formal link between the Price equation and an optimization program. *Journal of Theoretical Biology* 217:75-91

Grafen, A. (2006). Optimization of inclusive fitness. *Journal of Theoretical Biology* 238:541-563.

Grafen, A. (2007). The formal Darwinism project: a mid-term report. *Journal of Evolutionary Biology 20:1243-1254.*

Hofbauer, J. and Sigmund, K. (1998). *Evolutionary Games and Population Dynamics.* Cambridge University Press.

Metz, J.A.J., Mylius, S.D. and Diekmann, O. (2008). When does evolution optimize? *Evolutionary Ecology Research* 10:629-654.

Okasha, S. and Paternotte, C. (2012). Group adaptation, formal Darwinism and contextual analysis. *Journal of Evolutionary Biology* 25(6):1127-1139.

Okasha, S., Weymark, J.A. and Bossert, W. (forthcoming). Inclusive fitness maximization: An Axiomatic Approach.

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

Williams, M.B. (1970). Deducing the consequences of evolution: a mathematical model. *Journal of Theoretical Biology* 29:343-385.

Wagner, G.P. (2010). The measurement theory of fitness. *Evolution* 64-5:1358-1376.