The wrong equations: a reply to Gildenhuys

Bruce Glymour

How to cite this manuscript

Published Version Information

The Wrong Equations: A Reply to Gildenhuys

Bruce Glymour[†]

Abstract:

Glymour (2006) claims that classical population genetic models can reliably predict short and medium run population dynamics only given information about future fitnesses those models cannot themselves predict, and that in consequence the causal, ecological models which can predict future fitnesses afford a more foundational description of natural selection than do population genetic models. This paper defends the first claim from objections offered by Gildenhuys (2011).

1. Introduction.

In 'Wayward Modeling' (Glymour, 2006; hereafter WM), I argued against what I called the core commitment, the idea that population genetics[1] provides the core formal machinery for describing and understanding natural selection and the evolutionary events it produces. That commitment is challenged on the grounds that population genetic models cannot in general be used to reliably predict short and medium run evolutionary dynamics, are in certain important respects explanatorily incompetent, and induce avoidable conceptual confusions. The first charge is sustained by the following considerations: (1) population genetic models can reliably predict short and medium range evolutionary dynamics only given information about the time evolution of fitnesses, (2) classical population genetic models do not (generally) explicitly track the

---

[†] Dept. of Philosophy, Kansas State University, Manhattan KS 66503; glymour@ksu.edu
[1] There, and here, 'population genetics' is meant to include finite Markov chain models, diffusion approximations to them, and the algebraic models derived from them, though the problems at issue are faced broadly by any model in which fitnesses are assumed to be governed by a stationary stochastic process.

causes of fitness, but instead assume those fitnesses follow a predictable pattern (e.g. are constant, cyclical or drawn from a constant distribution), (3) the available data on survival and reproductive success are generally inadequate to reliably model fitnesses in this way, and so (4) in consequence population genetic models cannot in general reliably predict the future fitnesses or short and medium term evolutionary dynamics they generate. Because future fitnesses are essential for predicting evolutionary dynamics but can themselves be predicted only by explicitly causal ecological models of reproductive success, I conclude in WM that such causal models provide more foundational descriptions of natural selection.

The defense in WM of claim 3 above rests on a crucial inferential problem besetting population genetics. In most actual populations, the causes of per-capita rates of reproductive success for types in a population vary over time, changing the generation-specific probability densities over per-capita rates of success in ways that imply temporal variation in type-fitnesses, values for which are required if population genetic models are to yield reliable predictions. There are a variety of methods, i.e. ways of modeling fitness, employed by population genetic models to accommodate such variation—one may estimate a constant fitness by averaging over contexts, or one may assume that fitnesses follow some cycle, and so on. Unfortunately, the resulting models of fitness can generally neither be specified nor identified from the data on survival and reproductive success—there is simply not enough of it. For example, one standard method for accommodating this kind of variation is to model type-fitnesses as random variables governed by a constant probability density. But because sample size is typically small, one has to make apriori assumptions about the form of the density (e.g. is the distribution

lognormal or binomial or whatever), and even so the data are typically insufficient to reliably identify, i.e. estimate parameter values for, the density (see WM, pp. 373-376 for a sustained discussion).

In a recent paper (Gildenhuys, 2011), Peter Gildenhuys offers a number of objections to the thesis in WM. While he agrees with claim 1 above, that population genetic models can reliably predict evolutionary dynamics only given information about the time evolution of fitnesses, he holds that fairly standard population genetic models already in the literature are often adequate to 'handle' (Gildenhuys 2011, pp 821-827) spatial and temporal variation in fitness. He further argues that it is relatively straightforward to extend population genetic models so that they explicitly represent the causes of fitness, thereby enabling the extended models to predict future fitnesses and, using them, evolutionary dynamics (Gildenhuys pg 825 and pp 832-834). Finally, he claims that the predictive failures of population genetics are irrelevant to assessing the adequacy of these models as representations of natural selection (c.f. Gildenhuys, pg 827).

The third of the above complaints may appeal to some, but should give pause to those who, with me, regard explanations of evolutionary dynamics as central to evolutionary biology. The second complaint is cogent, though somewhat tendentiously posed. In WM I recommend the use of causal ecological models to predict fitnesses, and take those models to provide a better representation of selection than the bits of population genetics appended to them; the extended models recommended by Guildenhuys are thus of the kind endorsed in WM, and I welcome as fellow travelers

those who employ them to represent selection. The first complaint, however, is seriously and importantly in error, and in the body of this paper I respond to it.

2. Extant Machinery.

Gildenhuys claims that models of frequency dependent selection (c.f. Gildenhuys 2011, pp. 824-825), niche models (c.f. Gildenhuys 2001, pp. 821-824), and models in which fitnesses are drawn at random from a constant probability density (c.f. Gildenhuys 2011, pp. 825-827) commonly suffice to model spatial and temporal variation in fitness. Gildenhuys most fully elaborates an example of the third kind, and as this example is in particular illustrative of the difficulties at issue, I will focus most closely on it. I note in passing that niche models face, to a first approximation, exactly the problems to be discussed below, only multiplied by the number of niches in the model. Readers interested in the empirical difficulties besetting attempts to fit niche models are referred to Schmidt and Rand (2001) for an illustrative example. On the other hand, frequency dependent models can sometimes avoid these difficulties, exactly because they employ causal models of fitness. As I note in WM (pp 380-381), when type frequencies are the dominant cause of fitness, as e.g. in sexual selection, frequency dependent models will track the relevant causes of fitness, and so will allow predictions of future fitnesses. But quite commonly, type frequencies are not the only important causes of fitness.

In these much more common cases in which type frequencies do not exhaust the relevant causal factors, something else is needed. Gildenhuys' leading example is a model developed in Turelli, Shemske and Bierzychudek, 2001 (hereafter TSB), which treats fitnesses as random variables drawn from a constant probability distribution. And,

unsurprisingly, the model advanced in TSB demands more of the data than can actually be supplied; the result is a model that a) cannot actually be tested against, or even correctly fit to, the data, b) fails to predict observed evolutionary dynamics, and c) cannot be used, on the data, to answer the question about natural selection for which it was developed. After briefly recounting the essential idea in TSB, I explain why the model it develops fails so radically.

TSB is an effort to determine, for a particular species of plants (*Linanthus parryae*), whether a stable polymorphism (white versus blue flowers) is sustained by temporal variation in selection alone, or by both temporal and spatial variation in selection (c.f. TSB, pg. 1238). To make the determination, a population genetic model, assuming temporal variation in fitness only, is specified. The idea is that the model will predict the maintenance of the polymorphism if and only if particular parameters take values in a certain range. If estimates of the parameters from the data are in the range, the model will, on the data, predict and explain the presence of the polymorphism, and we can then infer that fitnesses vary temporally but not spatially. Conversely, if the estimates place the parameter values outside the range in which, according to the model, the polymorphism would be sustained, we can reject the model in that its predictions conflict with the observed stability of the polymorphism. Since the implicit alternative is a model (not actually developed in TSB) with both temporal and spatial variation in fitnesses, rejecting the specified model amounts to accepting the hypothesis that fitnesses vary both spatially and temporally. We can thus use the estimated parameter values, in conjunction with the TSB model, to determine whether selection varies in time only or in both space and time—this, note, is the understanding of selection sought by Turelli and

colleagues, and the point of developing the population genetic model in question is to provide exactly that understanding.

The specified model assumes that whatever these temporally varying selection processes may be, they generate a probability density over fitnesses such that the ratio of generation specific type fitnesses, $W_{white,t}/W_{blue,t}$, is drawn at random from a constant lognormal distribution. From the unfitted model, TSB derive a triplet of conditions for the maintenance of the polymorphism, viz:

1) $\quad E\left(z_t \dfrac{W_{white,t}}{W_{blue,t}} \right) > 1$

2) $\quad E\left(z_t \dfrac{W_{blue,t}}{W_{white,t}} \right) > 1$

3) $\quad E\ln\left[1 + a_t \left(\dfrac{W_{blue,t}}{W_{white,t}} - 1 \right) \right] > 0$

where E denotes the expectation function, $W_{white,t}$ is the fitness of the white morph at time t, $W_{blue,t}$ is the fitness of the blue morph at time t, a_t is contribution of the current generation to the seed-bank, measured as a proportion of the current size of the seed bank, and z_t the ratio of a_t to its expected value, i.e. $z_t = a_t/E(a_t)$. Condition 1 then requires that the expected value of the product of the white/blue fitness ratio to z be above 1 and condition 2 requires that the expected value of the product of the blue/white fitness ratio and z also be above 1. Both can be satisfied if fitnesses and z are appropriately correlated, as can condition 3, which requires, loosely, that a is relatively small when the blue morph is less fit than the white morph, and relatively big when the converse is true.

If it can be shown that either of the first two conditions *is not* satisfied, the model predicts that the polymorphism is not stable, contrary to observation, and hence the

model, and with it the hypothesis that selection only varies temporally, can be rejected in favor of the hypothesis that selection varies in both space and time. Alternatively, if it can be shown that all three conditions condition *are* satisfied, the inference that selection varies only temporally is supported—at least, the hypothesis is consistent with the data, and so cannot be rejected.

To test whether the conditions are satisfied, it is conjectured (their word) that $E(a_t)$ is between .05 and .2, which in conjunction with data yields an estimate z_t for each of seven years of observation (the data set is larger, but usable data were obtained in only seven years). Finally, fitnesses for the white and blue morphs are estimated for each of the seven years.

TSB concludes that all three conditions are probably satisfied, and consequently that the hypothesis of temporal variation alone suffices to explain the stable polymorphism. Our question is whether this conclusion is empirically well grounded. More precisely, does the model permit reliable estimates of future fitnesses from the available data; can the model, so fitted, be used to predict short and medium run evolutionary dynamics; does it in fact allow a reliable determination of whether selection varies only temporally or both spatially and temporally; and finally, does it give us an adequate understanding of selection in the study population? The answer to all of the above is 'no'.

First, note that the assumption that the fitness ratios are drawn from a constant lognormal distribution is crucial. If the assumption is false, the model does not reliably predict future fitnesses, misrepresents the time evolution of fitnesses, and fails to provide diagnostic test conditions. But TSB do not test this assumption against alternatives, and

in fact they cannot: they do not have sufficient data for any such test. As it turns out the frequency distribution of $W_{blue,t}/W_{white,t}$ in the seven generation sample used by TSB is consistent with a lognormal distribution (p values of .363 and .260 respectively for Cramer-von Mises and Anderson-Darling tests). But the data are also consistent with, for example, Weibull and gamma distributions (p values respectively of .222 and .250 for a Cramer-von Mises test, and .193 and .250 respectively for an Anderson-Darling test). This is unsurprising; to reliably determine the parent distribution from a sample that might be drawn from a lognormal distribution, but might also be drawn from a Weibull or gamma distribution, one generally needs a sample of 20 or more.[2] So the test of the 'temporal variation only' hypothesis relies on an auxiliary for which we have no empirical warrant; we have no such warrant because, exactly as charged in WM, to perform such a test we need much more data than we have.

Granting, for the sake of argument, that the fitness ratios have a lognormal distribution, is there enough data to identify the relevant parameters required to test the claim that selection varies only temporally? Again, no. Using just the observed type fitnesses, condition 1 is satisfied, but conditions 2 and 3 are not: the average weighted fitness ratio of blue to white (the left hand side of 2, above, called "weighted average RSN" by TSB), is .92. Hence 2 is not satisfied. Because 2 is a necessary condition for 3, 3 is not satisfied either. Strictly speaking, on the data the model predicts that polymorphism is not stable, contrary to observation. But point estimates are problematic, and it is better to compute confidence intervals around the point estimate of the average weighted fitness ratio, and check whether those intervals exclude values above 1 for the average weighted fitness ratio of the blue morph to the white.

---

[2] See Marshall et al. (2001) for discussions and simulation results.

Unfortunately, from a sample of size 7, reasonable confidence intervals cannot be directly computed. TSB therefore report results from a bootstrap estimation procedure, i.e. new data are simulated from the old data, and from the new data 95% confidence intervals are estimated. The resulting 95% confidence intervals *do* include values for which the polymorphism is stable (i.e. values satisfying both 2 and 3). Hence, the implicit null cannot be rejected, and, as TSB say, the data "are consistent with the hypothesis that fluctuating selection as modeled above [i.e. temporally varying selection alone] maintains the flower color polymorphism." (pg. 1293). But the data are also consistent with the claim to the contrary (i.e. the confidence intervals also include values that do *not* satisfy 2 and 3). Notwithstanding the conclusions drawn in TSB, what should be said is that the model, on the data, simply does not ground a reliable inference even about the qualitative question of whether the selection pressures maintaining the polymorphism are merely temporally varying or also varying in space.

Neither does the model predict short and medium run dynamics. There are two large changes in type frequencies during the observational period; the model predicts only one of them. TSB are at pains to explain why the fitted model does not predict particularly well. They consider several possibilities, but concentrate most closely on unmodeled features of the seed bank, which features they think the most likely source of error. Specifically, genotype may influence seed longevity, viability or germination directly or interactively with some unmeasured environmental condition, or age, or both (c.f. TSB, pg 1295). Any of these causal connections between environment, genotype, phenotype and fitness would lead to violations of the model assumptions, requiring complexification in one way or another. As TSB write: "Despite the extreme fluctuations

in the observed relative fitnesses, the morph frequencies have changed relatively little. This reflects the strong damping effect of the seed bank…[which damping] implies that even a study that spans more than a decade may be insufficient for studying the range of environmental conditions responsible for the polymorphisms …thorough studies of adult fitness spanning more than a decade would have to be supplemented by even more elaborate and sustained studies of seed bank demography and selection to understand all of the biology relevant to the transient behavior of this 'simple' polymorphism."(TSB, pg 1296). Put more tendentiously, they do not yet understand enough about the causes of fitness, i.e. selection, in this population to responsibly deploy a population genetic model representing that selection.

3. Conclusion.

Thus, the model cited by Gildenhuys as illustrative of the population genetical machinery for representing varying selection simply fails to provide a well grounded understanding of the very selection process it is supposed to represent. It fails to (reliably) answer the question it is employed to answer, because the time evolution of the fitnesses is complex and the causal processes driving that evolution are not well understood. The data on fitness, despite being drawn from a decade of observation on tens of thousands of plants, is inadequate to estimate the density over fitnesses: its form is simply assumed, and its parameters estimated from simulations. Further, the model does not in fact predict the observed evolutionary dynamics. Of course, this, or some other, population genetic model could answer the question of interest, and even predict the observed evolutionary dynamics, but only by deploying fitnesses derived from a prior

understanding of selection given by explicitly causal ecological models, here of selection

in the seed bank. It is worth nothing, however, that such detailed causal, ecological

models of fitness would answer the substantive question about whether selection varies in

space and time, or time only, and they would do so directly, without appeal to population

genetics. That is, with respect to the question of interest, the TSB model uses the wrong

variables related by the wrong equations employing the wrong kinds of parameters. The

difficulties besetting the effort in TSB to use a population genetic model to represent

actual an actual case of selection are not unique, but instead ubiquitous: routinely, data on

survival and reproductive success are inadequate to estimate future fitnesses, either

directly or by means of simple statistical models. Gildenhuys is wrong to imply

otherwise.

References.

Gildenhuys P (2011) Righteous modeling: the competence of classical population genetics, *Biology and Philosophy*.

Glymour B (2006) Wayward Modeling: Population Genetics and Natural Selection *Philosophy of Science* 73:369-389.

Marshall, AW, Meza, JC and Olkin, I (2001) Can Data Recognize Its Parent Distribution? *Journal of Computational and Graphical Statistics* 10:555-580.

Schmidt P, Rand D (2001) Adaptive Maintenance of Genetic Polymorphism in an Intertidal Barnacle: Habitat- and Life-Stage-Specific Survivorship of *MPI* Genotypes, *Evolution* 55:1336-1344.

Turelli, M., D. Schemske, and P. Bierzychudek (2001) Stable Two-Allele Polymorphisms Maintained by Fluctuating Fitnesses and Seed Banks: Protecting the Blues in *Linanthus Parryae Evolution* 55:1283-1298.