Conditions for Evolution by Natural Selection

Peter Godfrey-Smith

Philosophy Department
Harvard University
Cambridge MA 02138, USA.

December 2006 [U]

Standard verbal formulations of necessary and sufficient conditions for evolution by natural selection require variation, heritability, and differential reproduction. These formulations are ambiguous with respect to their role and subject to counterexamples. Problem cases reveal tensions and trade-offs within standard practices of description within evolutionary biology. In some respects, the verbal summaries must be seen as *approximations* of more exact treatments. In some other ways, they embody *idealizations*. However, in yet other respects the verbal summaries can function to express elements of evolutionary theory that are hard to capture in other ways.

1. Introduction

Both biologists and philosophers often make use of simple verbal formulations of necessary and sufficient conditions for evolution by natural selection (ENS). Such summaries go back to Darwin's *Origin of Species* (1859/2001, especially the "Recapitulation"), but recent ones are more compact. Perhaps the most commonly cited

formulation is due to Lewontin (1970). These summaries tend to have three or four conditions, where the core requirement is a combination of *variation*, *heredity*, and *fitness differences*. The summaries are employed in several ways. First, they are often used in pedagogical contexts, and in showing the coherence of evolutionary theory in response to attacks from outside biology. Second, they are important in discussions of possible extensions of evolutionary principles to new domains, such as cultural change. The summaries also have intrinsic scientific and philosophical interest as attempts to capture some core principles of evolutionary theory in a highly concise way.

Despite their prominence, both the proper formulation and status of these summaries are unclear. Standard formulations are subject to counterexamples, and their relations to formal models of evolutionary change are not straightforward (Okasha 2007). So are the verbal summaries merely rough approximations that have no theoretical role of their own? Perhaps they could operate as theoretical statements in Darwin's time, but have now been superseded by more exact treatments.

I will look at three families of problem cases, and argue that each motivates different conclusions. One set of cases, involving the role of age-structure in populations, is best addressed by regarding the verbal summaries as *idealized* in a particular way. A second set of cases, involving heredity, show a role for *approximation*. Discussion of these first two sets of cases show that there are trade-offs operating, often unacknowledged, in many discussions of ENS. Simplicity and predictive clarity are being traded off against the accommodation of all genuine cases. But a third set of problems, involving random genetic drift and related phenomena, reveal a way in which a verbal summary, properly formulated, can have a more positive theoretical role. These summaries can be used to say things that cannot be said, in suitably general form, via existing formal models. At the end of the paper I offer two new formulations of the traditional three-part summary, guided by a distinction between two roles such a formulation can play.

2. Standard Formulations of the Conditions

Perhaps the most commonly cited summary of ENS is due to Richard Lewontin (1970, p. 1).

> As seen by present-day evolutionists, Darwin's scheme embodies three principles...
> 1. Different individuals in the population have different morphologies, physiologies, and behaviors (phenotypic variation).
> 2. Different phenotypes have different rates of survival and reproduction in different environments (differential fitness)
> 3. There is a correlation between parents and offspring in the contribution of each to future generations (fitness is heritable).
> These three principles embody the principle of evolution by natural selection. While they hold, a population will undergo evolutionary change.

Though often cited, later discussions by Lewontin and others do not follow this formulation closely. In particular, it is not usually seen as necessary for ENS that *fitness* be heritable. There are simple cases where fitness is not heritable, but a phenotypic trait is heritable, and that is enough for the trait to evolve. Suppose the tall individuals in generation 1 have more offspring than the short ones, and height is heritable. Then there will be change from generation 1 to 2, even if there are no fitness differences at all in generation 2. This problem with Lewontin's 1970 formulation does become an advantage with respect to one case discussed below. But when in this paper I refer to Lewontin's version of the recipe, I will usually mean the following summary, slightly modified from Lewontin (1980, p. 76).[1] Note that in the summaries quoted below in this section, I have changed the symbols used to number the authors' conditions, for ease of reference below.

> A sufficient mechanism for evolution by natural selection is contained in three propositions:
> L1. There is variation in morphological, physiological, or behavioral traits among members of a species (the principle of variation).
> L2. The variation is in part heritable, so that individuals resemble their relations more than they resemble unrelated individuals and, in particular, offspring resemble their parents (the principle of heredity).
> L3. Different variants leave different numbers of offspring either in immediate or remote generations (the principle of differential fitness).

A more elaborate formulation is offered by John Endler (1986, p. 4).

Natural selection can be defined as a *process* in which:
If a population has:
E1. variation among individuals in some attribute or trait: *variation*.
E2. a consistent relationship between that trait and mating ability, fertilizing ability, fertility, fecundity, and, or, survivorship: *fitness differences*.
E3. a consistent relationship, for that trait, between parents and their offspring, which is at least partially independent of common environmental effects: *inheritance*.

Then:
E4. the trait frequency distribution will differ among age classes or life-history stages, beyond that expected from ontogeny;
E5. if the population distribution is not at equilibrium, then the trait distribution of all offspring in the population will be predictably different from that of all parents, beyond that expected from conditions E1 and E3 alone.

Conditions E1, E2, and E3 are necessary and sufficient for natural selection to occur, and these lead to deductions E4 and E5. As a result of this process, but not necessarily, the trait distribution may change in a predictable way over many generations.

Lastly, here is an example of a textbook presentation, in Ridley (1996, pp. 71-72).

Natural selection is easiest to understand, in the abstract, as a logical argument, leading from premises to conclusion. The argument, in its most general form, requires four conditions.

R1. Reproduction. Entities must reproduce to form a new generation.
R2. Heredity. The offspring must tend to resemble their parents: roughly speaking, "like must produce like."
R3. Variation in the individual characters among the members of the population....
R4. Variation in the *fitness* of organisms according to the state they have for a heritable character. In evolutionary theory, fitness is a technical term, meaning the average number of offspring produced by an individual relative to the number of offspring left by an average member of the population. This condition means that individuals in the population with some characters must be more likely to reproduce (i.e., have higher fitness) than others....

If these conditions are met for any property of a species, natural selection automatically results. If any conditions are not met, natural selection does not occur.

First it is necessary to look at some questions about the intended *role* of these summaries. In particular, there is an ambiguity in the idea of giving "necessary and sufficient conditions for ENS." The aim may be to describe conditions that will *produce* ENS (where we know what ENS is). Or the aim may be giving conditions for some process *being a case* of ENS. So there is a distinction between *constitutive* and *causal* questions to make here. (Similarly, giving necessary and sufficient conditions for *becoming* pregnant must be distinguished from necessary and sufficient conditions for *being* pregnant.)

The usual aim of those offering conditions for ENS seems to be answering *both* kinds of question. The summaries describe a situation in which a certain kind of change in a population will occur, and the entire process is identified with ENS. The standard summaries *explain* ENS by giving a *recipe* for ENS. But it is also possible to give a summary without giving a recipe in this sense. We might describe ENS as a temporally extended process of a certain kind, without asserting a tight dependence relation between stages of the process. The initial stages of the process may not invariably (or even reliably) be followed by the latter stages, but if the initial stages *are* followed by the latter stages, then we have a case of ENS.

Summaries of ENS, ever since Darwin, have often been presented in the form of recipes for change. It is often seen as a strength of evolutionary theory that its core mechanism (or one of its core mechanisms) has a straightforward predictive character. More specifically, I take it that the usual aim is to give conditions that are *sufficient ceteris paribus* for a certain kind of change occurring. (The status of the *ceteris paribus* clause will be discussed below.)

I will finish this section with brief comments about topics that will not be discussed below. First, a few summaries require that variation be *random* (Fracchia and Lewontin 1999). I assume that no randomness condition should be required. Darwinian evolution can occur on variation that is directional, even adaptively "directed." In these cases natural selection may have less explanatory importance than it has when variation is random, but it can still exist. Second, there is no distinction made in summaries above between *natural* and *artificial* selection. This, also, will not be treated as a problem, as a dichotomy between natural and artificial selection is not of theoretical importance within

biology itself.[2] Third, some discussions of the core of Darwinism focus on *cumulative* selection (Sterelny and Griffiths 1999, Chapter 2), but I avoid any such restriction and focus on a more inclusive category. Lastly, there is an alternative foundational description of ENS that uses the idea of a "replicator" (Dawkins 1976, Hull 1980). But when intended as a fully general description that can function as an alternative to the standard summaries, this approach fails (Godfrey-Smith 2000). Though those who advocate this approach usually do not define replicators carefully, this concept involves some absolute (though vague) notion of faithful transfer of structure across generations. But what is needed for ENS is parent/offspring predictability understood in a *comparative* way. (This point will survive the problems and modifications discussed below.) It is sufficient for ENS (given other conditions) that parent and offspring be *more* similar that randomly chosen individuals of different generations. So any absolute degree of parent/offspring similarity (except 100%) will be sufficient in some contexts and insufficient in others. Replicators are also apparently meant to be asexual in their "transfer of structure," and ENS should surely be possible with sex at the focal level – without sexual reproduction being reducible to asexual replication of genes, for example. Offspring may resemble both parents more than they resemble randomly chosen members of the parental population, even if offspring do not faithfully *copy* either parent. As a consequence, replicators are not necessary for ENS.

I now turn to cases that cause problems for the standard summaries.

3. Births and Deaths

The first case concerns the measurement of differences in reproductive output.

> **Case 1, Different Generation Times:** We have a population of individuals, of types A and B. Every individual alive at the beginning of some time interval fissions to produce two offspring of the same type as the parent. Later, all individuals do the same thing again, and again. But A individuals cycle through this process faster than B individuals do, owing to their more efficient metabolism. So more A's are produced, and at the end of the time interval the frequencies of the types have changed.

This seems to be clearly a case of ENS. There is a change in the frequencies of types, due to variation in reproductive capacities and faithful inheritance of type. There are no differences in the population with respect to the *number* of offspring produced by different individuals, however. So there are no differences between the two types in the number of offspring produced per individual of that type. The only differences between the types concern the rate at which new individuals are produced per unit of *time*.[3]

The Lewontin 1980 and Ridley summaries clearly exclude this as a case of ENS, as they explicitly treat fitness as the *number* of offspring produced by an individual (or the number produced on average by individuals of a given type).[4] All the organisms in Case 1 have two offspring – eventually – and the "eventually" versus "quickly" distinction has no place in these summaries.

This is not to say that Case 1 is a problem for evolutionary theory itself. This is a simple case of an "age-structured population," and there are detailed models of such cases (Charlesworth 2004, Crow 1986 Chapter 6, Roughgarden 1979 Chapter 18). I will sketch the simplest kind of analysis that would be given.

The crucial move is that we think of reproduction as occurring in time. For simplicity I will treat time in a discrete way, measured in days. The two types (A and B) each have an $l(x)$ schedule, which specifies how likely an individual of that type is to reach age x, and an $m(x)$ schedule, which specifies how many offspring an individual of that type will have at age x.

We initially think of the A's and B's as forming two subpopulations. Each subpopulation will reach a stable age distribution and then grow multiplicatively, by a factor of λ_A and λ_B per day respectively. To determine these rates of increase for the two types, we solve the following equation for each type.

$$(1) \qquad 1 = \sum_x \lambda^{-x} l(x) m(x)$$

Let us assume the following $l(x)$ and $m(x)$ schedules, which satisfy the description above.

Type A: $l(1) = 1$; $l(2) = 0$, $l(3) = 0$...

$m(1) = 2$; $m(2) = 0$, $m(3) = 0$...

Type B: $l(1) = 1$; $l(2) = 1$, $l(3) = 0$...

$m(1) = 0$; $m(2) = 2$, $m(3) = 0$...

We find that $\lambda_A = 2$ and $\lambda_B = \sqrt{2}$. These numbers can then be used to represent the rate at which the frequency of A will grow relative to B, in the total population (Crow 1986, p. 173).

So it is possible in Case 1 to assign to each type a fitness-*like* number, that will predict what will happen in the population. These are the *rates of increase* of the two types, measured with respect to time. I return to the status of these parameters below, but the immediate point is that we see that several recipes for ENS are making a tacit *idealization*. They treat all cases of ENS as if they occurred in populations in which generations are *non-overlapping* and *synchronized* across the entire population. (This is often called a "discrete generation" model, though this should not be confused with treating time in a discrete way, as I did above.) The same idealization also operates, often unacknowledged, in much of the literature on the "propensity view of fitness" (Mills and Beatty 1979).

The term "idealization" is a controversial one in philosophy of science. I understand idealization as involving the *imagined modification* of a real system, usually in the direction of simplicity. An idealized description is one that is straightforwardly true of a fictional relative of the real system, and may also be taken (in many cases) to be approximately true of the more complex real system (Godfrey-Smith forthcoming, Thompson-Jones 2005). The present example involves a special kind of idealization, however. Some organisms *do* have non-overlapping generations synchronized across the population – annual plants such as basil do, many insects, and some others. But most organisms do not. In a case like the human population, the notion of a "generation" has no meaning as a population-level, as opposed to individual-level, phenomenon. So summaries of ENS given in the style of Lewontin (1980) and Ridley, which treat fitness as the number of offspring produced (or the expected number) can be applied *literally* to annual plants, many insects, and some other organisms. But to most organisms they must

be applied in an indirect way, via an idealization. A summary that treats fitness in terms of numbers, ignoring the role of time, can only be seen as describing an *imagined simpler relative* of the processes of ENS in those cases.

At this point, it might be thought that the right response is to modify the summaries so that they use rates of increase instead of fitnesses of the simplest kind. And it is true that the simplest "discrete generation" cases can be treated as special cases of an evolutionary process with age-structure. However, models that predict change with rates of increase (like λ_A and λ_B above) make their *own* idealizations. Above we assumed asexual reproduction, and $l(x)$ and $m(x)$ schedules that stay fixed as the population grows. Once the population is sexual, and creates new individuals by combining contributions from two parents, we cannot represent the rate at which a type increases in terms of its *own* survival and reproduction schedules. This is because any type also produces other types, and is produced by them – if, indeed, discrete "types" exist at all in the population, which may instead contain individuals who vary only quantitatively. We also assumed that the population was in a stable age distribution, though most populations will be knocked out of this distribution by natural selection itself, along with other factors. (See Ariew and Lewontin 2004 for a discussion of some of the problems that arise.)

So those who want an exact formal treatment face a choice between idealizations. The consensus among modelers seems to be that in many empirical cases, a good approximate description can be achieved by assuming either a discrete generation model or an age-structured model with rates of increase, and these two models converge a lot of the time, especially when selection is weak (Crow 1986, p. 174). Thus a modeler can pick and choose between frameworks, according to the case and the purposes at hand. But if our aim is formulating necessary and sufficient conditions for ENS itself, which we aim to see as a single kind of real process, then the role of these idealizations is problematic. In particular, it becomes impossible to treat ENS as a process that is "driven" by something like the familiar fitness differences between organisms or types invoked by Lewontin and Ridley.[5]

So far, we have seen that there are two ways to approach the formulation of a summary of ENS. One way is to make idealizations, and give a summary that will apply literally to some cases and as an idealized model to others. Then it is possible (modulo

some complications discussed below) to keep the summary simple, while also specifying a process with the kind of internal causal reliability or predictive power that was discussed in the previous section. That is, it becomes possible to give a summary in the form of a recipe.

The other approach is to avoid idealization, and try to capture every case. This is what Endler's formulation does, in effect. Endler's summary is expressed *as if* it is describing a recipe with the kind of predictive features seen in Lewontin 1980 and Ridley, but the formulation is so full of qualifications that it has little predictive power. In his "E2" clause Endler lists a number of properties that are *related* to the notion of fitness – survival, fecundity, fertility, mating ability – but he does not collapse these into a single measure that is taken to be predictive of change. He does not say that the "bottom line" for ENS is differences in expected number of offspring, or differences in rate of increase, or anything else. If there is no "bottom line," Endler is leaving it open that the "mating ability" differences might balance out the "survival" differences, for example, to yield no evolutionary change.

In fact, we see the same phenomenon in microcosm in Lewontin's formulations. In Lewontin 1970, the phrase "rates of survival and reproduction" appears, with "rate" ambiguous with respect to units. Further, one of Lewontin's examples only works on a "per unit time" interpretation. But if Lewontin 1970 is read in this way, then no "bottom line" measure of reproductive output is given, and predictive power is lost. Lewontin's 1980 formula resolves the ambiguity in favor of a "per individual" conception of fitness, which we now see to be a move sacrificing generality for predictive clarity.

So we see that the standard recipes are forced, in effect, to make strategic choices that trade generality off against a combination of simplicity and predictive power. Ridley idealizes, Endler avoids idealization at the expense of predictive "punch," and Lewontin does a little of both. What is then striking is that none of Ridley, Lewontin, or Endler *say* that this is what they are doing. Ridley and Lewontin 1980 do not confess to idealization, and Endler hangs onto the language of definite prediction even when the details of his formulation let much of the predictive air out. What these biologists are confronting is related to the trade-off between *precision* and *generality* emphasized in some recent philosophy of science (Levins 1966, Weisberg 2003). The desire that a recipe have a

clear predictive role is related to the goal of precision, and the role of generality here is obvious. But the situation is made complicated by the role of idealization with respect to structural features.

I will make a few more comments on the relation between idealized and non-idealized summaries, before moving to a new set of cases. Another role that the idealization to discrete generations achieves is the establishment of a *minimal unit* of evolutionary change, which becomes change across a single generation. This enables simple treatment of another case, that is not a problem case *per se* but a routine one whose categorization is not straightforward.

In Case 1, we saw change due solely to different rates of birth. There can also be change via differential rates of death, without any births.

> **Case 2, Culling:** We have a population of individuals, of types A and B. Across a time interval, some individuals die while some remain alive. The frequencies of types change over the interval, as B individuals die at a higher rate than A individuals. The higher death rate in B is due to their inferior ability to fight off disease.

If reproduction is required for ENS, as in Lewontin 1980 and Ridley, then this does not suffice for ENS. Endler's formulation and Lewontin 1970 allow that survivorship differences alone count as fitness differences, but if culling *alone* counts for ENS then heritability, required in those formulations, is not strictly necessary.

Of course, the most natural reply is to say that culling without reproduction is only *part* of a process of ENS, and not sufficient alone.[6] But this raises the question: if this is not enough, then what is the minimal unit that *is* enough? In a discrete generation model, as noted above, change across one generation provides a natural minimal unit. But once we are thinking of an age-structured population, there is no non-arbitrary minimal unit; there are just shorter and longer intervals of time. An interval too short will be one in which nothing of interest can happen, and an interval too long is one that might be outside the domain of micro-evolutionary theory (the theory of change within a population) altogether. But within those boundaries, there is a lot of freedom.

The right response at this point is to say that this is only a "problem" from the point of view of convenience. Modelers like to think in terms of minimal units of change,

but nothing in the theory itself is affected if there is no such unit. The natural response to Case 1 and its kin is to say that the *paradigm* cases of ENS include reproduction and extend over many typical generation lengths for the organisms in question. If someone asks what is the *minimal fragment* of such a process that counts as ENS, the question should be dismissed as empty. As we get further and further from the paradigm cases, we get further and further from having a fragment that deserves the name "a case of ENS."

The central message of this section, again, is that standard summaries of ENS often engage in idealization, and avoiding idealization trades off against simplicity and predictive power. These idealizations are present even when the language used by some authors seems intended to establish as literal and direct a mode of description as possible.

4. Heredity and Heritability

All summaries of ENS include a requirement involving the inheritance of traits. Lewontin's summaries require that variation be "heritable," where this is understood in terms of a statistical similarity between parent and offspring. Endler and Ridley are less specific (see clauses E3 and R2). In this section I use the term "heritability" to refer to a family of exact statistical measures, and "heredity" and "inheritance" to refer in a vaguer way to all phenomena involving parent/offspring similarity and the transmission of traits across generations.

The problems encountered in this section are in some ways reminiscent of those in the previous one. It is appealing to summarize ENS by giving a recipe for change, especially if the recipe is a compact one with clearly separable ingredients. But summaries that satisfy this goal are not able to handle all cases. Other problems we will encounter have no analogues in the previous section.

The role of heredity in ENS is often described using metaphors. If a population has variation and fitness differences, there will be no change if the population is not disposed to "respond" to selection, and that requires that parental characters are "transmitted" across generations. Lewontin's summaries use the concept of heritability to capture this extra ingredient more precisely. Summaries of this kind also shadow a formal representation of evolution known as the "breeder's equation," which treats change in the

mean value of some characteristic in a population as the product of a term representing variation in fitness and a term representing heritability (see Heywood 2005, and below).

It is common to distinguish several different senses of heritability (Jacquard 1983, Downes 2004), which arise within two frameworks or approaches. One approach, which I will call the "fraction of variance" approach, is based on a causal model of inheritance that assumes the presence of genes or something similar to them. Heritability is measured as the genetic variance divided by the total phenotypic variance in the population. The other, the "regression approach" is more abstract, independent of any causal model of inheritance, and aimed simply at the representation of predictability relations between parents and offspring. This more abstract approach will be assumed here, as it is important not to assume that standard genetic mechanisms of inheritance are present in all cases. Specifically, heritability will be initially understood as the slope of the linear regression of offspring character on parental character.[7] This approach to the "third ingredient" nicely meets the goals discussed above. Heritability in this sense is a single number, representing the tendency of population to "respond" to selection. Further, it is usually understood and measured independently of fitness differences, so it operates as a distinct ingredient in the recipe, not something conceptually entangled with the role of selection itself.

I now turn to some problem cases. The symbolism used is as follows. In the population of parents, individual i has phenotypic value Z_i. This is a quantitatively varying character, such as height. The average phenotypic value that the offspring of individual i have, or would have if there were any, is Z'_i. In the remainder of this section I assume asexual reproduction, so each individual has just one parent. The heritability is measured as the slope of the line $b_{Z',Z}$ that gives the best prediction of Z' values from Z values.[8] It is also important that the only measure of change being considered is change in the mean value of Z in the population; we ignore various processes that count by most standards as genuine change, but that do not affect the mean. (Many or all could be accommodated by changing our choice of Z, the characteristic analyzed.)

The first case is introduced in an unpublished manuscript by Robert Brandon.

Case 3, Biased Inheritance: A population varies with respect to Z (Brandon's example is wealth). Z is heritable, and positively associated with fitness. But there is also a tendency for offspring to have a lower *Z* value than their parents. A numerical example is given in Table 1. The mean in this population before and after selection is 1.67, despite variation, fitness differences, and a heritability of 0.74.

Individual	**Z**	**W**	**Z'**
1	2	2	2
2	2	2	2
3	2	1	1.67
4	2	1	1.67
5	1.67	1	1.67
6	1.67	1	1.67
7	1.67	1	1.33
8	1.67	1	1.33
9	1.33	1	1.33
10	1.33	1	1.33
11	1.33	0	1.33
12	1.33	0	1.33

Table 1: Numerical example of biased inheritance, after Brandon.

Brandon argues that this shows that Lewontin-style summaries assume the absence of *bias* in the inheritance system. Although it was said above that heritability is identified with the slope of a regression line predicting *Z'* from *Z*, a regression analysis gives us two parameters, the slope of the line and the intercept with the vertical axis. Here the intercept is negative, and the influence of this factor exactly counteracts the evolutionary change that would have been predicted from the fitness differences and the heritability alone. So if heritability is understood as a regression slope, then at least one extra parameter needs to be taken into account when predicting change.

I now introduce a second case:

Case 4, Heritability Fails in the Fit: An asexual population contains variation in height. There is a positive covariance between height and fitness. There is a positive covariance between parental height and offspring height. But there is no change in mean height across generations. This is because although taller individuals have more offspring on average, and taller individuals have taller offspring on average, the taller

individuals with the high fitness are not the *same* tall individuals as those that have taller offspring. The high-fitness tall individuals are not the tall-offspring tall individuals. The mean value of Z is unchanged across generations, though there are fitness differences and Z is heritable. A numerical example is given in Table 2.

Individual	Z	W	Z'
1	1	1	1
2	1	1	1
3	1	1	1
4	2	1	2
5	2	1	2
6	2	5	1.6
Mean	1.5	5/3	1.5 (weighted by size of offspring classes)

Table 2: Numerical Example for Case 4

Once again, the three-part recipe is not sufficient for change.[9] The key point is obvious in retrospect. If we calculate heritability from the entire parental population, the heritability can be affected by individuals who make little or no contribution to the next generation. So there *has* to be some error in a prediction that is made using heritability in this way, unless the pattern of inheritance is the same across the whole population.

This, like Brandon's case, is a toy example. But both cases illustrate phenomena that are real possibilities. One is a directional tendency in the mechanisms producing departure of offspring from parental phenotype. The other is a mixed underlying basis for the inheritance of a single trait, so that the fittest individuals do not have the same inheritance patterns as those seen for other values of the trait. In each case, a fine-grained analysis removes any appearance of paradox, but that does not invalidate the fact that the population-level criteria used for predicting change have problems.

The obvious response to Case 4 is to understand heritability in a way that takes into account fitness differences in the parental generation. We do not need to modify the formulation of the recipe itself, but just the interpretation of a key component. This response is reasonable (see also Heywood 2005 for an endorsement of fitness-weighted

heritability measures).[10] This has consequences for how a recipe for evolutionary change is understood, however. The original aim was to think of heritability as dispositional property of the parent population that exists independently of the pattern of fitness differences. Whether or not heritability should be called a "cause" of anything, the aim was to treat it as a distinct ingredient in a breakdown of explanatory factors. If heritability properties are treated as dependent on fitness, we have logically "entangled" two ingredients in the breakdown of factors.[11]

Yet another set of problem cases involve the interaction of heredity with *stabilizing selection*, selection that acts to maintain an intermediate value of some trait in the population. A simple asexual case is as follows.

> **Case 5: Stabilizing Selection in an Asexual Population** (see Figure 1). The population contains short (Z=1), intermediate (Z=2), and tall (Z=3) individuals. Intermediate individuals are fitter than either extreme. Heritability as measured by parent-offspring regression is one, but the parental and offspring generations are phenotypically identical. Stabilizing selection is exactly compensated by a dispersing tendency in inheritance, seen in the phenotypically intermediate individuals.

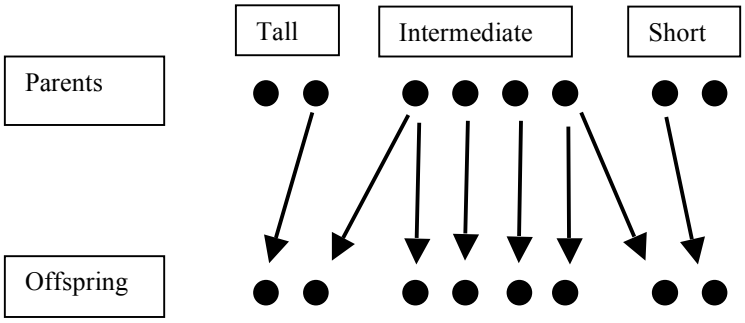

Figure 1: Stabilizing selection in an asexual population.[12]

This case mirrors a category of cases involving sexual reproduction, featuring heterozygote superiority with respect to fitness but not with respect to phenotype. That is, an intermediate phenotype is favored by selection and is produced by a heterozygote at one locus (genotype *Aa*), resulting in a stable equilibrium of gene frequencies. There is a

net tendency for short individuals to produce short individuals, and tall to produce tall – heritability may be as high as one – even when the population is in the equilibrium state. And there are fitness differences between individuals in this equilibrium state. This I will call **Case 6**. A graphic representation is given in Figure 2.

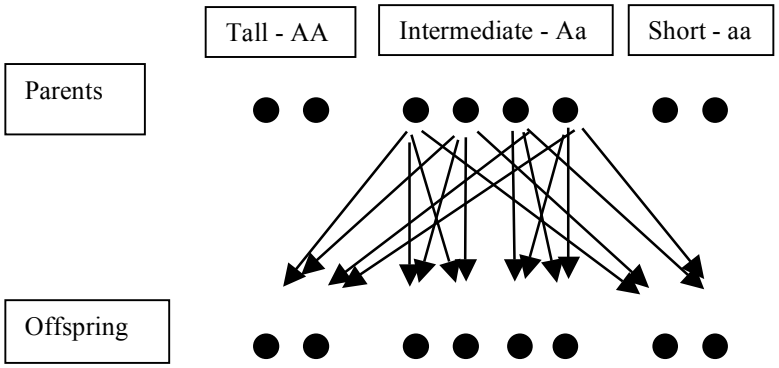

Figure 2: A superior intermediate phenotype produced by a heterozygote (Case 6)

The message of these cases is that heritability is quite a blunt instrument, even when our sole aim is to say whether change will occur or not. These cases also highlight a fact about stabilizing selection. In talking about selection, it is often said that fitness differences in the population must be "systematic." Stabilizing selection involves systematic fitness differences in one sense, but not in another sense. These are cases where we can say in independent terms which phenotypes are the fit ones, and may be able to give ecological reasons why they are fit. A golden mean principle may be applicable, for example. But in another sense, the fitness differences in these cases are not "systematic" because there is no overall tendency for higher values of Z to be fitter, or less fit, than low values. This will be discussed again in the next section.

This case also casts some light on the recipes quoted in section 2. Lewontin's 1970 recipe required that *fitness* be heritable, not the phenotypic trait evolving. In some simple cases this seems to give the wrong answer, and Lewontin did not use this

condition in his 1980 and later formulations. But in Case 6, fitness is not heritable even though phenotype is. So Lewontin's 1970 recipe has no problem in that case.[13] However, both fitness and phenotype are heritable in the asexual Case 5. In addition, Endler's qualification regarding equilibria in his clause E5 is probably intended to handle these sorts of cases.

So the use of heritability in recipes for change must be seen as involving an approximation. Positive heritability is compatible with inheritance biases that can cancel change. If heritability is measured without regard for fitness, it is affected by misleading information. And the pattern of heredity can interact with stabilizing selection in such a way that despite high heritability and fitness differences, the same population is restored across generations.

In some (though not all) of the cases discussed in this section, the problem arises because we stipulate a pattern of heredity that is liable to produce change *on its own*, and selection exactly cancels that tendency, yielding no net change. Such cases might reasonably be regarded as special. If so, Case 6, which does not have this feature, becomes an important one.

In this section I have used the term *approximation* to describe the role of heritability. In the previous section I discussed *idealization*. How are these phenomena related?

When idealizing in the manner discussed in the previous section, we imagine a *structural modification* to the systems we are interested in, in order to make them easier to describe. In these cases involving heritability, that is not what is going on (or at least, not with respect to the features emphasized in this section). Rather, we ignore some possibilities, and are also content with an analysis that makes predictions that are *largely* accurate the rest of the time. We are not imagining structural modifications, but merely allowing our descriptions to have a "loose fit" to real-world phenomena. We *could* describe the situation as one in which we are "idealizing away from" certain possibilities, but the specific type of imaginative act that was uncovered in Section 3 is not found here.

Though an idealized description is only straightforwardly true of an imagined relative of a real system, it will often be approximately true of the more complex real-world system. So idealization can yield approximate truth. But not all approximately true

descriptions involve idealization. If I say someone is 6 feet tall when they are only 5 feet 11 inches, this will count as approximately true in many contexts, but I am not idealizing in the sense discussed above.

In sum: the problems involving age-structure in Section 3 show the role of idealization in standard summaries of ENS, while the problems involving heritability show a role for approximation in a more general sense that need not be understood in terms of idealization.

The cases in this section also raise the possibility of appealing to *ceteris paribus* clauses in making sense of the recipes. Certainly we must assume that something like a *ceteris paribus* clause is in the background in all claims about the conditions sufficient for evolutionary change to occur. A variety of strange events outside the usual domain of evolutionary biology could intervene to prevent change. Though I accept that minimal role for a *ceteris paribus* assumption, I doubt that the specific problems raises in the last sections are best handled by leaning on this idea. This is because we are not confronting problems that involve breakdowns of normal conditions or processes. Rather, they are ordinary biological complexities that are conspiring to make it hard to say what we want to say in a compact way.

5. Heritability and the Price Equation

In this section I discuss the relation between the traditional summaries and the "Price equation" framework for the abstract representation of evolutionary change. Okasha, in an important recent discussion, has claimed that, in the light of Price, the status of the Lewontin conditions must be revised (2007, Chapter 1). My discussion here draws extensively on Okasha's treatment.

A simple form of the Price equation for representing change is as follows:

(2) $$\overline{W}\Delta\overline{Z} = Cov(W,Z) + E(W\Delta Z)$$

As above, Z is a quantitative character (such as height), and $\overline{Z}$ is its mean in the parental generation. The term $\Delta\overline{Z}$ is defined as $\overline{Z}_o$ - $\overline{Z}$, where $\overline{Z}_o$ is the mean in the next generation.

W represents absolute fitness, the number of offspring an individual has. So each individual i in the parental generation is characterized by its Z_i and W_i, its phenotype and its fitness, and also by Z'_i, the average Z value of its offspring. An individual is also characterized by its ΔZ_i, which is its value of Z'_i-Z_i. Then $Cov(W, Z)$ is the covariance in the population between Z and fitness. $E(W\Delta Z)$ is the average of the ΔZ values weighted by the fitnesses of each individual. $\overline{W}$ is mean fitness in the population.[14]

$Cov(W, Z)$ is zero in the cases of stabilizing selection (Cases 5 and 6) discussed in the previous section. Although there were fitness differences in that case, there was no overall tendency for either high or low values of Z to have high fitness. So we seem to have evaded one problem already.

Initially, it seems that there is also a simple link between this equation and a Lewontin-style recipe. Evolutionary change has been broken down into a term that represents the role of fitness differences, $Cov(W, Z)$, and a term that represents the net role of inheritance, expressed as a correction for "transmission bias." The link can be made more explicit by unpacking the covariance term into the regression of fitness on character ($b_{W,Z}$), and the variance of Z:

$$(3) \qquad \overline{W}\Delta\overline{Z} = b_{w,z}Var(Z) + E(W\Delta Z)$$

So we have terms representing variation, fitness differences associated with phenotype, and the role of inheritance. But the term $E(W\Delta Z)$ is far from a standard heritability concept. To locate the role of heritability in the usual sense in equation (2), we must break it down in a more complicated way. Combining several equations from Okasha:

$$(4) \qquad \overline{W}\Delta\overline{Z} = hb_{W,Z}Var(Z) + Cov(W,e) + a + \overline{Z}(h-1)$$

Here, h is the regression slope of Z' on Z, without fitness weighting. The intercept of that regression is a. $Cov(W, e)$ is the covariance between the fitness of each individual i and the "residual" or error for that individual (e_i) when using the regression line $Z'_i = hZ_i + a$ to predict the value of Z'_i. So we see via Price that heritability in its usual sense is only

one of *three* factors that matter, concerning inheritance. Brandon's case above works via the role of *a*, and my Case 4 works because of the role of *Cov*(*W*, *e*).[15]

Okasha argues, more strongly, that in the light of the Price equation, the Lewontin conditions are structurally problematic. His argument is as follows. It is desirable to re-write the Price equation in a way that removes the role of fitness differences from the second term on the right hand side, the one representing the role of the inheritance system. That yields this formula (Rice 2004, Heywood 2005):

(5) $\overline{W}\Delta\overline{Z} = Cov(W, Z') + E(\Delta Z)$

But once we have this version of the equation, we see that "what is *really* required for there to be evolution by natural selection is for *Cov*(*W*, *Z′*) to be non-zero, that is, for an entity's fitness to correlate with the average character of its *offspring*. This is the fundamental condition...." (p.37, some symbols changed). That is what tells us when there will be change over and above that produced by the inheritance system alone. The Lewontin recipe, in this analysis, tries to capture this "fundamental condition" with two others, that *Cov*(*W*, *Z*) and *Cov*(*Z*, *Z′*) both be nonzero. But *Cov*(*W*, *Z′*) cannot be determined from these. In effect, the Lewontin conditions treat covariance as a transitive relation, which it is not.

Though this analysis is very illuminating, I do not accept the conclusions Okasha draws. I first introduce a problem case. If "what is *really* required for there to be evolution by natural selection is for *Cov*(*W*, *Z′*) to be non-zero" then this is a case of evolution by natural selection:

> **Case 7, *Cov*(*W*, *Z′*) Positive with No Variation in Z:** Some individuals in the parental generation have more offspring than others, and the offspring of these individuals have higher values of Z than the mean value in the parental population. The mean value of Z is then higher in the offspring generation. But the reproductively successful individuals did not *themselves* have high values of Z. In fact, everyone in the parental generation had the same value of Z.

This certainly does not look like a case of ENS, at least with respect to *Z*. However, it is also under-specified. Why did the high *Z′* individuals have more offspring, if there was

no variation in Z? Given the story told, there seem to be two options. One is pure accident, and the other is hitchhiking; the individuals with high values of Z' were favored by selection on some other trait. Both of these cases must be distinguished from natural selection acting on Z itself. So far, the Lewontin recipes (although not the Endler or Ridley recipes) have said nothing to help us distinguish selection from random drift or hitchhiking. But as I will argue in the next section, the way to make sense of these distinctions is *not* to take Z, the parental phenotype, out of the picture, as $Cov(W, Z')$ does.

Secondly, Okasha's claim that $Cov(W, Z')$ is the fundamental condition is based on the claim that equation (5) has a kind of primacy in its representation of the factors responsible for change, in particular a superiority over equation (2). This can be questioned on independent grounds. Okasha's aim is to treat the second term on the right hand side of the equation, which handles the role of "transmission bias," in a way independent of the effect of fitness differences. That is apparently a desirable separation. However, the removal of fitness differences from this term introduces an effect of the inheritance system on the representation of the role of fitness differences, in the new first term on the right hand side. Consequently, I do not agree that equation (5) gives us a true separation of the role of fitness and inheritance.

Another way to look at it is to compare yet another Pricean breakdown of change:

$$(6) \qquad \bar{W}\Delta\bar{Z} = Cov(W, Z) + Cov(W, \Delta Z) + E(\Delta Z)$$

The first term on the right hand side represents fitness differences only, the far right term concerns the inheritance system only, and the middle term combines both. That middle term can either be assimilated into the $Cov(W, Z)$ term, yielding equation (5), or assimilated into the $E(\Delta Z)$ term, yielding equation (2). Either way, we get a mixture of the role of inheritance and fitness somewhere. So I don't think either (2) or (5) is superior to the other, in a principled sense.[16]

In sum, I accept that the Pricean analysis is very informative about the status and workings of the traditional recipes, as Okasha claims. It helps us understand the respects in which those recipes are approximations. In effect, the traditional recipes embody a

different formal model of evolution, the "breeder's equation," $\Delta \overline{Z} = hs$, where h is heritability and s is a term representing fitness differences. But the breeder's equation is an approximation (Heywood 2005), so we can expect the verbal formulas that shadow it to be approximations as well. However, I am not convinced by Okasha's arguments about the need for a *restructuring* of summaries of ENS.

I close the section with two more general remarks on heredity. First, throughout this section I have assumed that we are dealing with a trait that varies quantitatively. This is needed for the measurement of variances and covariances. But other traits are more naturally represented as occurring in many discrete types without well-defined distances between them. These cases can be handled within the quantitative framework via some shoe-horning, but it is important to keep an eye on the fact that, once again, we tend to slip naturally into analyzing a particular subset of the cases and then treating the outcome of the analysis as applying straightforwardly to all.

Secondly, we should note one other fact about the treatment of heritability in terms of regression. In principle, a regression can have any value between plus and minus infinity. On the "fraction of variance" approach to heritability, heritabilities are between zero and one (inclusive). If we assume standard genetic mechanisms, heritabilities measured by regressions are constrained to lie between zero and one and the two frameworks coincide. But if our aim is a general treatment that covers all possible inheritance systems, then we see that (i) the regression approach to heritability is the appropriate one, where it is applicable, and (ii) there is no reason in principle why the regressions could not be negative or greater than one. In effect, we see that standard treatments of heritability have assimilated two different phenomena, parent/offspring *similarity* and parent/offspring *predictability*. The former is a special case of the latter. If our aim is to treat the heritability concept as measuring the evolutionary "response" to selection, then we see that there will be some response as long as parent and offspring have *some* systematic relationship, whether this is one of systematic similarity or systematic dissimilarity (see also Jacquard 1983, Blute forthcoming).[17] Cultural inheritance might be a domain where such parent/offspring anti-correlation may be common.

This shows, I think, a slight breakdown or rupture in our usual picture of the relationships between key theoretical concepts. Summaries of Darwinism routinely say that ENS requires that "like must produce like," or there must be a "resemblance" between parent and offspring.[18] But both the recent formal development of the theory, and, indeed, a direct application of its underlying logic, show us that parent/offspring predictability is the crucial relationship here, not parent/offspring similarity. If summaries of ENS require that "like produces like," they rule out cases where there are both fitness differences and an evolutionary "response" to them, via a systematic relation between parents and offspring that involves a negative regression slope. I am not saying that such cases *must* be understood as ENS; perhaps they should be put in a distinct category. At present it is not clear how best to categorize them at all.

6. Drift and Correlated Response

In this section I turn to what may seem the most obvious problem with some standard summaries: the problem of distinguishing ENS from change occurring via reproductive differences that arise by various kinds of accident. Though this problem may initially seem the most difficult to fix, it will receive a comparatively simple treatment here.

I organize the discussion with two cases handled simultaneously.

> **Case 8, Accident:** Individuals' values of *Z* have no causal role in survival and reproduction. But individuals with higher values of *Z* have more offspring purely by accident. *Z* is heritable. The mean value of *Z* increases.

> **Case 9, Correlated Response:** Individuals with higher *Z* values also have higher values of *X*. *Z* is causally inert, but high values of *X* are advantageous, and individuals with high values of *X* consequently have more offspring. *X* and *Z* are both heritable. Mean values of *X* and *Z* both increase.

In Case 9, the problem is what we say about *Z*. Trait *X* evolves by natural selection, clearly, but *Z* seems to fit some of the standard summaries as well. So Case 9 raises the possibility that a summary needs to be expressed in terms of criteria for evolution *of some particular trait* by natural selection.

Of the summaries quoted earlier, Lewontin's are the most susceptible to these problems. Both Endler and Ridley make explicit gestures towards ruling out reproductive differences that arise by accident. In effect, both require that differences in reproductive output have a *systematic* relation to parental phenotype (clauses E2 and R4). In almost all summaries and sketches of ENS, in fact, one can sense a desire to say *something* that rules out accident, accompanied by uncertainty over the right way to do this. Thus we see, as in Endler, Ridley, and perhaps Lewontin 1970, various kinds of modal or causal loading of the language used to describe fitness differences and their relation to phenotype.

A natural first response to these problems, and one that would deal with both pure accident and the hitchhiking phenomena of Case 9, is to add a requirement that there be a *causal link* between the phenotypic variation and the reproductive differences cited in the summary, in order for a given trait to evolve by natural selection. Specifically, we might require that the phenotypic variation be partly causally responsible for the differences in reproductive output.

The main shortcoming of this approach is that it seems coarse-grained and crude, especially given the large body of theory on the relation between selection and random drift. All we have here is a binary distinction: either the phenotypic variation played *some* causal role, perhaps a minor one, or it did not. One might hope for a more fine-grained treatment.

This hope is seen in Okasha's treatment of the issue (2007), which I will take as representative of a family of such approaches. Okasha again draws on the Price equation. He notes that in principle the fitness differences used in the Price equation (the differences between W_i's) could be due to chance rather than selection, but argues that the Price formalism can be used to distinguish the roles of the two factors. Suppose there is a probability distribution, P, that assigns probabilities of various levels of reproductive success to an organism of a given kind in a given environment. Each organism i's realized fitness W_i can be expressed as a sum of its *expected* fitness in that environment, W_i^* calculated using P, and a deviation δ_i from that expectation.[19]

(7) $$W_i = W_i^* + \delta_i,$$

This breakdown can be introduced into the Price equation:

(6) $$\overline{W}\Delta\overline{Z} = Cov(W^*, Z') + Cov(\delta, Z') + E(\Delta Z)$$

Okasha claims that this move "partitions the total change into a component due to selection on *Z* and a component due to random drift;..... In principle, that is, if we could discover the probability distribution *P*, we could determine whether the overall change is the result of chance, natural selection, or a combination of the two" (p. 33). This also enables us to answer recent skeptics about the distinction between selection and drift, Okasha says, because it gives us a common currency, units of *Z*, in which the contributions of each can be compared (for the skeptical arguments, see Matthen and Ariew 2002).

I doubt that this probabilistic breakdown achieves Okasha's goal, however. The breakdown in terms of *P*, if it is available at all, tells us about the extent to which realized fitnesses conform to expected values. But an expected value can itself be produced *by accident*. The degree of match between an expected and an actual value cannot itself tell us *how* the actual value was produced. And the question of whether an outcome was due to selection or mere accident is a question of exactly this kind, a question about how a set of realized fitness values were produced.

More precisely, if the probability distributions that Okasha uses here were available, then it might be a *necessary* condition for the absence of drift that the realized fitnesses are identical to the expected fitnesses. But it is not a sufficient condition, because an expected value can obtain by accident. It can obtain without having a basis in the causal processes that *make W** the expected value.

Once this problem with the "deviation from expectation" approach is clear, we seem pushed back towards the simple causal criterion introduced above. I suggest that this is, in fact, an adequate treatment of both drift and correlated response. For evolution by natural selection to act on *Z*, we require (among other things) that the fitness differences affecting the population have a partial causal basis in variation in *Z*. In the case of drift (Case 7), that requirement is not met because the fitness differences were accidental, at least with respect to the role of *Z*. In the case of correlated response, the association

between *Z* and *X* may or may not be accidental, but variation in Z played no causal role in the fitness differences that produce change.

This shows that the verbal summaries have a very different role in the context of this last family of problems, when compared to their role in earlier sections of this paper. Whereas in the earlier sections, the verbal summaries seemed destined to be no more than approximations of more formal treatments, a verbal formulation of the kind endorsed here is the most natural and appropriate way of capturing the causal requirement that handles the problems of drift and correlated response. This shows also that standard recipes such as Lewontin's were not mistaken to express their criteria in terms of a pair of conditions, one on heritability and one on the covariance between fitness and parental phenotype, *Cov*(*Z*, *W*). Okasha, as discussed earlier, suggests that the key requirement can be expressed in terms of Cov(*Z'*, *W*) alone. But moving to a criterion of that kind makes it impossible to capture the causal role of parental phenotype, as it "jumps over" the parental phenotype.

Consequently, a uniform treatment of the last three cases (7, 8 and 9) can be given. In each case, there is no natural selection on *Z* because differences in *Z* had no causal role in producing fitness differences. In Case 9, we do have change in the population due to natural selection on *X*, so there is *some* natural selection. Some may think that the standard summaries are only designed to say when there is natural selection present at all; on that view, Case 9 is not a problem. In any case, the need to deal with the distinction between selection and random drift (Case 8) is enough on its own to motivate a causal requirement.

It might be argued at this point that the problem has not been solved until a more exact specification has been given of the relevant kind of causal link between phenotype and fitness. I accept that a more precise treatment of causation would be desirable, in general and in this specific context, but I don't think it is necessary at this point.[20] What is needed for present purposes is a way of distinguishing a broad and heterogeneous category of changes due to ENS, from changes that might look similar but are due to accident. Knowledge of causal facts of the relevant kind might be difficult, but those are the facts that mark the crucial difference. Once we know that differences in *Z* were partly

responsible for differences in fitness, we know the case is not a pure case of random drift, and that is the distinction that has to be made for present purposes.

This causal constraint also deals with yet another problem case, presented here as a simplification of one due to David Haig (personal communication), developed in the context of modeling birth weight.

> **Case 10, Environment as Common Cause of Fitness and Phenotype.** Suppose that individual *i*'s phenotype is determined by the function $Z_i = G_i + E_i$, where the variable *G* represents genotype and can be in three states (1, 2, or 3), and *E* represents environment, also with three states (1, 2, 3), that are equally probable. There is perfect asexual inheritance of *G*. Fitness, however, is determined by the function $W_i = E_i$, and hence is independent of *G*. Then *Z* will be heritable and correlated with fitness, but all the fitness differences among individuals are due to the environment, and are uncorrelated with genotype.[21] As a consequence, there will be no evolutionary change despite heritability and fitness differences.

Although fitness differences are non-accidental here, they are not causally dependent on phenotype itself. Instead, environment is a common cause of both fitness and phenotype. Whereas in Cases 7, 8, and 9, there was change across generations that appparently is not due to natural selection on Z, here there is no change, even though Z is heritable and correlated with fitness differences. In that respect, this last case resembles Case 4, in which we had fitness differences and heritability, but "heritability failed in the fit." This case also resembles Case 3, the case of transmission bias, because this is a case where if we consider only the fit individuals within each genotypic class, we find their offspring are biased downwards with respect to both phenotype and fitness. The case also has a similarity with Case 6, because both are cases in which fitness itself is not heritable (as was required by Lewontin 1970). So this last case is multi-faceted, but one simple response to it is to note that it is a case where fitness has no causal dependence on phenotype, and in this context that fact makes a dynamical, as well as interpretive, difference.[22]

The main point of this section is that the verbal summaries play a very different role here than they did above. In earlier sections, they struggled to accommodate all the cases, and had to be seen as embodying idealizations or approximations. In this last

section, it is verbal formulations that give us the most natural way of expressing a crucial requirement in a fully general form.

7. Conclusion

The range of problem cases discussed in this paper is not complete, but is sufficient to motivate some fairly definite conclusions.[23]

The project of giving a *summary* of ENS should be kept distinct from the project of giving a *recipe* for change in a population. Both projects are worthwhile, but they are somewhat different. Problems arise when a single formulation is intended to do both. A recipe that is simple and straightforwardly predictive will not capture all the cases. A summary that covers all cases will not give simple conditions causally sufficient for change.

Once these goals are separated, we also see that the procedures and problems faced in each project are different. Suppose first that our aim is to give a summary characterization of ENS that will capture all genuine cases. Then we are able to *assume* the presence of change, and our aim is to say which changes *count as* ENS. Our aim is to mark the boundaries of a category that has a particular explanatory role. The decisions that must be made will often concern how broad the category should be, and how we should manage trade-offs between simplicity and scope – the kind of trade-off emphasized in some unificationist approaches to explanation (Kitcher 1989). This sort of discussion can also devolve into terminological dispute of the unproductive kind. Should pure culling be regarded as ENS? As argued in Section 2, we should resist the temptation to impose overly sharp borders.

So consider the following summary:

(S1) Evolution by natural selection is change in a population due to:
(i) variation among members of the population
(ii) which causes different rates of reproduction, and
(iii) which is heritable.

Here I treat heritability as any systematic relation (not necessarily resemblance) between parent and offspring. The reference to "rates" in clause (ii) is intended to measure reproductive output with respect to time, hence avoiding an idealization to discrete generations. Unlike the summaries discussed in Section 2, this one includes explicit reference to a causal relation between variation and differential reproduction. And the summary is structured to allow that change may also occur because of the inheritance system alone, or by culling alone. The summary says that ENS is change *over and above* that resulting from these other factors.

A distinct project is trying to give a *recipe* for change that captures the core features of ENS in a compact and transparent way. Now we can help ourselves to idealizations of various kinds – we can assume discrete generations and asexual reproduction, if we want. We may also explicitly allow approximation. The risk now becomes not a collapse into terminological wrangling, but an embrace of idealizations that lead to loss of contact with important cases. It is also possible to stop trying to give such recipes in verbal form, trusting instead in a collection of equations (the Price equation, the one-locus diploid model, the replicator dynamics, the breeder's equation). But as emphasized in Section 6, a verbal summary does a better job with the basic distinction between selection and drift than these formalisms do, and verbal recipes will also be practically useful in contexts where it is necessary to avoid technicality. Some recipe-makers may also want to use an equation *plus* a verbal commentary as their preferred form of representation.

One idealized recipe can be constructed via modification of Lewontin (1980):

> (IR) The following conditions are sufficient for evolution of trait *Z* by natural selection in a large population with discrete generations:
> (i) There is variation in *Z*,
> (ii) There is a covariance between Z and the number of offspring left by individuals, where this covariance is partly due to the causal role of Z, and
> (iii) The variation is heritable, and inherited without directional bias.

Here we indicate a key idealization and rule out problems of inheritance bias explicitly. We can assume that heritability is understood in a fitness-weighted way. The requirement of covariance between Z and fitness rules out problems with stabilizing selection. We could also make the recipe more exact by stipulating an absence of migration. Some possibilities might be seen as captured by a tacit *ceteris paribus* clause. But once we are in the domain of idealized recipes, there is no need to capture all cases, and no need to settle on a single recipe for all purposes and contexts.

Although these idealized recipes do not substitute for a summary of what ENS is, they can be very illuminating. They give a compact and causally informative representation of some core cases of ENS, and provide a basis for the analysis of more complex ones. Problems only arise when the idealizations that have been made are forgotten or denied.

* * *

Acknowledgment: ***

References

Ariew, A. and R. Lewontin (2004). "The Confusions of Fitness," *British Journal for the Philosophy of Science* 55: 347-363.

Blute, M. (forthcoming). "The Evolution of Replication."

Brandon, R. (unpublished). "Inheritance Biases and the Insufficiency of Darwin's Three Conditions."

Charlesworth, B. (2004). *Evolution in Age-Structured Populations*. Cambridge: Cambridge University Press.

Crow, J. (1986). *Basic Concepts in Population, Quantitative, and Evolutionary Genetics*. New York: Freeman.

Darwin, C. (1859/2001). *On the Origin of Species*. (Facsimile of first edition). Cambridge MA: Harvard University Press.

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

Downes, S. (2004). "Heredity and Heritability." *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/heredity/>

Endler, J. (1986). *Natural Selection in the Wild*. Princeton: Princeton University Press.

Fisher, R. A. (1930). *Genetical Theory of Natural Selection*. Oxford: Oxford University Press.

Fracchia, J. and R. Lewontin (1999). "Does Culture Evolve?" *History and Theory* 38: 52-78.

Frank, S. (1998). *Foundations of Social Evolution*. Princeton: Princeton University Press.

Godfrey-Smith, P. (2000). "The Replicator in Retrospect." *Biology and Philosophy* 15: 403-423.

Godfrey-Smith, P. (forthcoming). "Abstractions, Idealizations, and Evolutionary Biology." To appear in A. Barberousse, M. Morange, and T. Pradeu (eds.), *Mapping the Future of Biology: Evolving Concepts and Theories*. Cambridge: Cambridge University Press.

Heywood, J. (2005). "An Exact Form of the Breeder's Equation for the Evolution of a Quantitative Trait Under Natural Selection." *Evolution* 59: 2287-2298.

Hull, D. (1980). "Individuality and Selection." *Annual Review of Ecology and Systematics* 11: 311-32.

Jacquard, A. (1983). "Heritability: One Word, Three Concepts." *Biometrics* 39: 465-477.

Kitcher, P. S. (1989). "Explanatory Unification and the Causal Structure of the World." In P. Kitcher & W. Salmon (Eds.), *Minnesota Studies in the Philosophy of Science Volume XIII*. Minneapolis: University of Minnesota Press, pp. 410-505.

Levins, R. "The Strategy of Model-Building in Population Biology." *American Scientist* 54: 421-31.

Lewontin, R. (1970). "The Units of Selection." *Annual Review of Ecology and Systematics* 1: 1-18.

Lewontin, R. (1980). "Adaptation." Reprinted in R. Levins and R. Lewontin, *The Dialectical Biologist*. Cambridge: Harvard University Press, pp. 65-84.

Matthen, M. and A. Ariew (2002). "Two Ways of Thinking About Fitness and Natural Selection." *Journal of Philosophy* 49: 53-83.

Mills, S. and J. Beatty (1979). "The Propensity Interpretation of Fitness." *Philosophy of Science* 46: 263-288.

Okasha, S. (2007). *Evolution and the Levels of Selection.* Oxford: Oxford University Press.

Pearl, J. (1998). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Rice, S. (2004). *Evolutionary Theory*: *Mathematical and Conceptual Foundations.* Sunderland: Sinauer.

Ridley, M. (1996). *Evolution* (2nd edition). Oxford: Blackwell.

Roughgarden, J. (1979). *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. Upper Saddle River: Prentice Hall.

Sterelny, K. and P. Griffiths (1999). *Sex and Death: An Introduction to the Philosophy of Biology*. Chicago: University of Chicago Press.

Thompson-Jones, M. (2005). "Idealization and Abstraction: A Framework." In M. Thompson-Jones and N. Cartwright (eds.), *Idealization XII: Correcting the Model*. Amsterdam: Rodopi, pp. 173-217.

Weisberg, W. (2003). *When Less is More*. PhD Dissertation, Department of Philosophy, Stanford University.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Notes

[1] Lewontin said "and" in L1, but I assume he meant "or."

[2] Many cases of "natural" selection, whether humans are involved or not, feature a key role for preference and choice as a causal factor.

[3] The reference to "remote generations" in the Lewontin formula does not help. All individuals have the same number of grand-offspring and great-grand-offspring as each other, though some individuals take longer than others to achieve them.

[4] Lewontin's 1970 recipe uses the term "rate" which is ambiguous with respect to units, but certainly may include measures per unit time.

[5] As Charlesworth summarizes, in an age-structured sexual population "no single parameter can be regarded as the fitness of a genotype with arbitrary selection intensities" (1994, p. 136).

[6] Those who think culling should count as ENS might note that "change in gene frequencies" is supposed to be sufficient for evolution in general. That is a textbook criterion (due initially to Dobzhansky). If so, change in gene frequencies due to culling on the basis of phenotype should be enough for ENS. In reply, it will be argued that this shows a misunderstanding of the idea of "change in gene frequencies." This is supposed to be understood as change in gene frequencies *across generations*. So to see whether or not there is ENS, we must wait until the individuals in Case 2 reproduce. But that reply leads to the unraveling of the idealization. We are supposed to wait until *who* reproduces? All of the population? Only some of them? All of those who will reproduce eventually?

[7] Within the fraction of variance approach, both "narrow" and "broad" sense heritabilities can be distinguished. When both frameworks (fraction of variance, and regression) are applicable, the sense of heritability measured as the regression of offspring on parent corresponds to the narrow sense.

[8] The regression line is the predictor of Z' from Z that minimizes the squared distances of the data points from the line, and is calculated as $Cov(Z', Z)/Var(Z)$. As will become not just apparent but conspicuous, the remainder of this section will use a discrete-generations assumption of the kind discussed in the previous section.

[9] If individual 6 produces 3 offspring of Z=2 and 2 of Z=1, we get all the same population statistics back with larger population size only.

[10] I modified Brandon's numerical example to reduce, but not eliminate, the role of these factors. Individuals 11 and 12 in the chart do not reproduce at all, but they are associated dispositionally with a Z' of 1.33. This makes sense of heritability is seen as a dispositional property of the parents independent of fitness, but, of course, the basis of such dispositions may be controversial. And we see that there are various ways of thinking of heritability. We could only count that parents that do reproduce, but ignore how much they reproduce. Or we could give weight each Z_i and Z'_i pair by the fitness of i.

[11] This entanglement is already present in the fact that *variation* must exist in the population if there are fitness differences and/or heritability. The ingredients that it would be attractive to keep separate are heritability and fitness, though. This goal exerts real influence, for example, on Okasha's 2007 treatment of the problem, as discussed below.

[12] If you are like me, you'll see a slight optical illusion here, where the lower "flanking" individuals seem pulled in towards the central group.

[13] In this respect Lewontin's formulation shadows not so much the Breeder's equation, but Fisher's fundamental theorem (Fisher 1930).

[14] Derivations of the equation are given in Frank (1998) and Okasha (2007).

[15] And also a non-zero intercept a.

[16] Suppose that only one individual in the parental generation exhibits a tendency to biased transmission. But that individual has zero fitness. Then we can say *either* that this was a case where transmission bias played no role (talking the language of equation 2), *or* that this was a case where transmission bias was present in the system but was cancelled by $Cov(W, Z')$.

[17] Jacquard (1983) makes the conceptual separation, and gives separate measures of parent/offspring *resemblance* (k) and parent/offspring *determination* (D), where k is the regression slope and D compares the average variance within offspring classes to the overall variance in the population. He then says that, even in a model without mechanistic assumptions, the two will be closely related, and in fact that $D=k^2$. But this result assumes that the variances of the parental and offspring generations are equal, which Jacquard notes but downplays.

[18] The Lewontin 1980 and Ridley summaries are examples; but Endler's, interestingly, is not.

[19] $W^*_i = \sum_j P_{ij} j$. Where the j's are numbers of offspring. So P_{ij} is the probability of i having j offspring.

[20] Here a manipulability approach might be the most promising (Pearl 1998, Woodward 2003), especially in the light of Case 10 below.

[21] The heritability (not weighted by fitness) will depend on the frequencies of the genotypes in the parental generation. If the genotypes are equally frequent, then the heritability as measured by regression of Z' on $Z = 0.5$.

[22] If genotype covaried with environment for some reason, then there would be evolutionary change, though not due to selection on Z. Here Endler's E3 once again contains a relevant qualification.

[23] I have not discussed problems arising from the role of *variance* in fitness in predicting change, and the occasional need to track *grand-offspring* rather than offspring when measuring fitness. For discussion of both, see Ariew and Lewontin (2004).