THE JOURNAL OF PHILOSOPHY
VOLUME CIV, NO. 10, OCTOBER 2007

CONDITIONS FOR EVOLUTION BY NATURAL SELECTION*

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* (especially the "Recapitulation"), but recent ones are more compact.[1] Perhaps the most commonly cited formulation is due to Richard Lewontin.[2] 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 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. Here I look closely at these verbal summaries, and at how they relate to formal models. Are the summaries merely rough approximations that have no theoretical role of their own? Perhaps they could operate as theoretical state-

*I am grateful to David Haig, Patrick Forber, Samir Okasha, Kim Sterelny, and Michael Weisberg for comments on an earlier draft.

[1] Darwin, *On the Origin of Species by Means of Natural Selection* (Facsimile of 1859 edition, Cambridge: Harvard, 2001).

[2] See Lewontin, "The Units of Selection," *Annual Review of Ecology and Systematics*, 1 (1970): 1–18.

0022-362X/07/0410/489–516

ments in Darwin's time, but have now been superseded by more exact treatments.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*. A third set of cases, 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 formulations can play.

I. STANDARD FORMULATIONS OF THE CONDITIONS

Perhaps the most commonly cited summary of ENS is due to a 1970 discussion by Lewontin.

> 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 (*ibid.*, p. 1).

Though often cited, this formulation is awkward in several ways, and later formulations by Lewontin and others do not follow it 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. Here fitness is not heritable but height is, and that is enough for height to evolve.

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 a 1980 discussion.[3] 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.[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 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.

[3] See Lewontin, "Adaptation," reprinted in Richard Levins and Lewontin, *The Dialectical Biologist* (Cambridge: Harvard, 1985), on p. 76. The main modification is that Lewontin said "and" in line 1, but I assume he meant "or."

[4] See Endler, *Natural Selection in the Wild* (Princeton: University Press, 1986), p. 4.

Lastly, here is an example of a textbook presentation, by Mark Ridley.[5]

> 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 left 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. 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 giving 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 will occur, and the entire process is identified with ENS. The standard summaries explain what ENS is by giving a *recipe* for ENS. But it is also possible to give a summary without giving a recipe. 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.

[5] See Ridley, *Evolution* (Malden, MA: Blackwell, 1996, 2nd ed.), pp. 71–72.

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 later.) Once this is determined to be the aim, there is a further ambiguity, however. The recipes are usually interpreted as saying that whenever we have variation, heritability, and fitness differences with respect to some trait in a population, change ensues. But some of the recipes, including Lewontin's formulations, might be instead read as saying that whenever a population features a *general* tendency to exhibit variation, heritability, and fitness differences, then *some* traits will change. Here I will interpret the recipes in the former, trait-specific way.

I will finish this section with brief comments about topics that are not discussed below. First, a few summaries require that variation be random.[6] I assume that no randomness condition should be included. 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.[7] Third, some discussions of the core of Darwinism focus on *cumulative* selection, but I avoid any such restriction and focus on a more inclusive category.[8] Lastly, there is an alternative foundational description of ENS that uses the idea of a "replicator." But when intended as a fully general description that can function as an alternative to the standard summaries, this approach fails.[9] 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/

[6] See, for example, Joseph Fracchia and Lewontin, "Does Culture Evolve?" *History and Theory*, XXXVIII (1999): 52–78.

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

[8] See Kim Sterelny and Paul Griffiths, *Sex and Death: An Introduction to Philosophy of Biology* (Chicago: University Press, 1999), chapter 2.

[9] For expositions of this approach, see Richard Dawkins, *The Selfish Gene* (New York: Oxford, 1976), and David Hull, "Individuality and Selection," *Annual Review of Ecology and Systematics*, XI (1980): 311–32. For the criticisms summarized in this paragraph, see my "The Replicator in Retrospect," *Biology and Philosophy*, XV (2000): 403–23.

offspring similarity 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 than 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. As a consequence, replicators are not necessary for ENS.

II. BIRTHS AND DEATHS

I now turn to cases that cause problems for the standard summaries. The first is not a puzzle case *per se*, but a routine phenomenon whose proper categorization is unclear, and that introduces more important cases.

> *Case 1, 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*. The higher death rate in *B* is due to their inferior ability to fight off disease.

No one reproduces, so the Lewontin 1980 and Ridley summaries preclude this from being a case of ENS. Endler's formulation and Lewontin's 1970 account allow that survivorship differences alone count as fitness differences, but if culling alone counts for ENS then heritability is not strictly necessary.

A natural response is to deny that this is a case of ENS. Pure culling is *part* of a process of ENS, but not sufficient alone. This response is largely right, but it is worth going through some arguments that can be made on each side.

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 originally to Theodosius Dobzhansky). If so, change in gene frequencies due to culling on the basis of phenotype should be enough for ENS. In reply, it may be argued that this shows a misunderstanding of the idea of "change in gene frequencies." The standard criterion for evolution 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 1 reproduce.

But this reply leads to trouble—to the unraveling, in effect, of an idealization that lies in the background of many discussions. We are

supposed to wait until *who* reproduces? All of the population? Only some of them? All of those who will reproduce eventually?

To air these issues more fully, I introduce Case 2.

> *Case 2, 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*.[10]

The Lewontin 1980 and Ridley summaries clearly exclude this case. (Endler will be discussed below.) Lewontin and Ridley explicitly treat fitness as the *number* of offspring produced by an individual (or the number produced on average by individuals of a given type).[11] All the individuals in Case 2 have two offspring—eventually—and the "eventually" versus "quickly" distinction has no place in these summaries.

This is not to say that Case 2 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.[12] I will sketch the simplest kind of analysis that would be given.

The crucial difference is that we now think of reproduction as occurring in time. For simplicity I will treat time in a discrete way,

[10] 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.

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

[12] For helpful summaries, see James Crow, *Basic Concepts in Population, Quantitative, and Evolutionary Genetics* (New York: Freeman, 1986), chapter 6; and Jonathan Roughgarden, *Theory of Population Genetics and Evolutionary Ecology: An Introduction* (Upper Saddle River, NJ: Prentice Hall, 1979), chapter 18. For a detailed account, see Brian Charlesworth, *Evolution in Age-Structured Populations* (New York: Cambridge, 2004).

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 first think of the *A*s and *B*s as forming two subpopulations. Each subpopulation will reach a stable distribution of ages 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 separately for each type.

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

Let us assume the following $l(x)$ and $m(x)$ schedules, which satisfy the description given for Case 2 above,

Type *A*: $l(1) = 1$; $l(2) = 0$, $l(3) = 0\ldots$

$m(1) = 2$; $m(2) = 0$, $m(3) = 0\ldots$

Type *B*: $l(1) = 1$; $l(2) = 1$, $l(3) = 0\ldots$

$m(1) = 0$; $m(2) = 2$, $m(3) = 0\ldots$

We find that $\lambda_A = 2$ and $\lambda_B = \sqrt{2}$. These numbers can be used to get a representation of the rate at which the frequency of *A* will grow relative to *B*, in the total population.

So it is possible in Case 2 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 *nonoverlapping* 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 has also been operating, often unacknowledged, in much of the literature on the "propensity view of fitness."[13]

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

[13] See Susan Mills and John Beatty, "The Propensity Interpretation of Fitness," *Philosophy of Science*, XLVI (1979): 263–88.

of the real system, and may also be taken (in many cases) to be approximately true of the more complex real system.[14] The present example involves a special kind of idealization, however. Some organisms *do* have nonoverlapping 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 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. And the simplest "discrete generation" cases can indeed 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.[15]

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

[14] See Martin Thompson-Jones, "Idealization and Abstraction: A Framework," in Thompson-Jones and Nancy Cartwright, eds., *Idealization XII: Correcting the Model* (Amsterdam: Rodopi, 2005), pp. 173–217; and my "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* (forthcoming).

[15] See Andre Ariew and Lewontin, "The Confusions of Fitness," *British Journal for the Philosophy of Science*, LV (2004): 347–63.

model with rates of increase, and these two models converge a lot of the time, especially when selection is weak.[16] 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 always "driven" by something like the familiar fitness differences invoked by Lewontin and Ridley.[17]

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 via idealization 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. In retrospect, we can see that this is what Endler's formulation was doing. Endler's summary is expressed *as if* it is describing a recipe with the kind of predictive features seen in Lewontin's, 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 something 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.

So the Lewontin/Ridley approach is to idealize, while the Endler approach is to avoid idealization at the expense of predictive "punch." And what is then surprising is that none of Ridley, Lewontin, or Endler *say* that this is what they are doing. Ridley and Lewontin do not confess to idealization, and Endler hangs onto the language of definite pre-

[16] See Crow, *op. cit.*, p. 174.

[17] There are other measures besides the ones discussed here. But Charlesworth's authoritative survey concludes that in an age-structured sexual population "no single parameter can be regarded as the fitness of a genotype with arbitrary selection intensities" (*op. cit.*, p. 136).

diction even when the details of his formulation let much of the predictive air out.

I will make a few more comments on the relation between the idealizing and nonidealizing approaches, and then move to a new set of problem cases. First, 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. Then it is easy to say that Case 1, pure culling, does not count as ENS. But once we are thinking of an age-structured population, there is no nonarbitrary minimal unit. Then we just have shorter and longer intervals of time. (The Price equation approach to modeling evolution, which will be discussed below, embodies such a picture.) 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. That is not itself a problem. Modelers like to think in terms of minimal units of change, but that is purely for convenience. There are ways to avoid wrangling about the status of Case 1 within such a picture. One can say that the *paradigm* cases of ENS include reproduction and extend over many typical generation lengths for the organisms in question. If we ask what is the *minimal fragment* of such a process that counts as ENS, the question can 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."

We have seen in this section that the standard summaries often engage in idealization, and that avoiding idealization trades off against predictive power. Further, the idealizations are present even when the language used by an author does not acknowledge idealization, but seems intended to establish as literal and direct a mode of description as possible.

III. 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). Here I use the term "heritability" to refer to a family of 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 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.[18]

It is common to distinguish several different senses of heritability, which arise within two frameworks or approaches.[19] 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 second approach will be used here, as it is important not to assume that standard genetic mechanisms of inheritance are present in all cases. Heritability will be initially understood as the slope of the linear regression of offspring character on parental character.[20] 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 selection itself.

[18] See John Heywood, "An Exact Form of the Breeder's Equation for the Evolution of a Quantitative Trait under Natural Selection," *Evolution*, LIX (2005): 2287–98.

[19] See A. Jacquard, "Heritability: One Word, Three Concepts," *Biometrics*, XXXIX (1983): 465–77, and S. Downes, "Heredity and Heritability," *Stanford Encyclopedia of Philosophy* (http://plato.stanford.edu/entries/heredity/).

[20] 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.

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 most of this section I assume asexual reproduction, so each individual has just one parent. Heritability is measured as the slope of the line that gives the best prediction of Z' values from Z values.[21] Absolute fitness is symbolized with W.

The first case is introduced in an unpublished manuscript by Robert Brandon.[22]

> *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.[23] The population's composition is unchanged across generations, despite variation, fitness differences, and a heritability of 0.74.

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

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

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

[21] 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 conspicuous, the remainder of this section will use a discrete-generations assumption of the kind discussed in the previous section.

[22] Brandon, "Inheritance Biases and the Insufficiency of Darwin's Three Conditions" (manuscript).

[23] Though Z' is generally a mean in this paper, this table should be read so that the top two $Z = 2$ individuals have only offspring whose Z value is 2.

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 problem case involving heritability:

> *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 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 highly heritable. A numerical example is given in Table 2.

Table 2. Numerical Example for Case 4

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	7	1.571
Mean	1.5	2	1.5 (weighted by size of offspring classes)

Once again, the three-part recipe is not sufficient for change. (If individual 6 produces 4 offspring of $Z=2$ and 3 of $Z=1$, we get the same population statistics back with a larger population size.) 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.[24] 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.[25]

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 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)

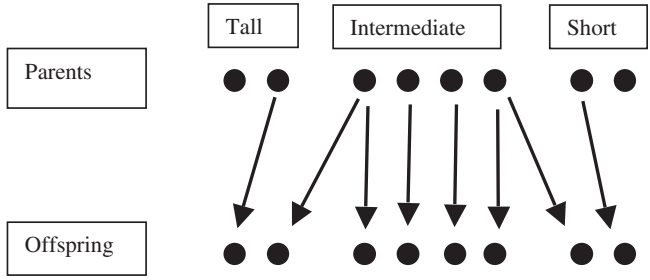

Figure 1: Stabilizing selection in an asexual population.[26]

[24] 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 if 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 here. We could only count the parents that do reproduce, but ignore how much they reproduce. Or we could weight each Z_i and Z'_i pair by the fitness of *i*. See Heywood (*op. cit.*) for an endorsement of fitness-weighted heritabilities.

[25] 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 treatment of the problem, as discussed below.

[26] If you are like me, you will see a slight optical illusion here, where the lower "flanking" individuals seem pulled in towards the central group. Note that there is not supposed to be phenotypic variation within the three classes here.

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.

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 tendency for short individuals to produce short individuals and tall to produce tall, hence heritability of phenotype, 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 of an extreme example is given in Figure 2.

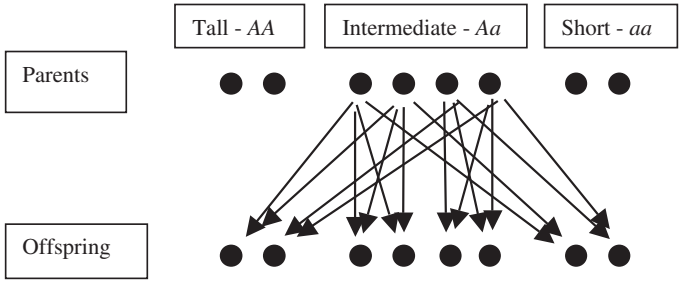

Figure 2: A superior intermediate phenotype produced by a heterozygote (Case 6).

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 II. First, 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, and he notes these phenomena in his 1970 discussion. However, both fitness and phenotype are heritable in Case 5.[27] Second, Endler's qualifications in his clause E5 are probably intended to handle these sorts of cases.

So the use of heritability in specifications of conditions sufficient 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 II 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

[27] Here I assume that the same pattern of reproduction continues in the new generation.

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 II 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 raised here are best handled by leaning on this idea. 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.

IV. 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. Samir Okasha, in an important recent discussion, has claimed that, in the light of Price, the status of the Lewontin conditions must be revised.[28] My discussion here draws extensively on Okasha's treatment.

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

$$\Delta\overline{Z} = Cov(w,Z) + E_w(\Delta Z) \tag{2}$$

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. Fitness, symbolized by w, is now a form of relative fitness, in which absolute fitness is divided by mean fitness in the population. 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 fitness-weighted average of the ΔZ values.[29]

[28] Okasha, *Evolution and the Levels of Selection* (New York: Oxford, 2005), chapter 1.

[29] Derivations of the equation are given in Steven A. Frank, *Foundations of Social Evolution* (Princeton: University Press, 1998), and Okasha, *op. cit.*

$Cov(w, Z)$ is zero in the cases of stabilizing selection (Cases 5 and 6) discussed in the previous section. Although there were fitness differences in those cases, there was no overall tendency for either high or low values of Z to have high fitness.

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:

$$\Delta\overline{Z} = b_{w,Z}\,Var(Z) + E_w(\Delta Z) \tag{3}$$

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 measure. 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's treatment:

$$\Delta\overline{Z} = hCov(w,Z) + Cov(w,e) + a + \overline{Z}(h - 1) \tag{4}$$

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 its value of Z'_i. So we see via Price that heritability in this familiar sense is only one of *three* factors that matter, concerning inheritance. Case (3) above works via the role of a, and Case 4 works because of the role of $Cov(w, e)$.[30]

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 (reflected in the weighting of the average) from the second term on the right hand side, the one representing the role of the inheritance system. That yields this formula:[31]

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

[30] And also a nonzero intercept a.

[31] This formula is used in Sean Rice, *Evolutionary Theory: Mathematical and Conceptual Foundations* (Sunderland: Sinauer, 2004), and Heywood (*op. cit.*).

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…" (*op. cit.*, 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 nonzero" 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 underspecified. 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.

Second, 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:

$$\Delta\bar{Z} = Cov(w,Z) + Cov(w,\Delta Z) + E(\Delta Z) \quad (6)$$

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 do not think either (2) or (5) is superior to the other, in a principled sense.[32]

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. 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 discrete types without well-defined distances between them. These cases can be handled within the quantitative framework via some shoe-horning, but there are probably better ways to represent them.[33] Once again, it is easy to find ourselves analyzing a particular subset of the cases and then treating the outcome of the analysis as applying straightforwardly to all.

[32] Okasha also gives an argument based on counterfactuals, but I think it has similar problems. He argues that standard ways of assessing counterfactuals imply that $E(\Delta Z)$ expresses the change in Z that would have occurred if selection was wholly absent, and this shows that $Cov(w, Z')$ measures "the difference made" by selection. But if we consider the other counterfactual, imagining inheritance bias wholly absent, we do *not* find that $Cov(w, Z')$ measures the change that would then result. So again, I do not think that equation (5) gives us a superior breakdown.

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 counteracted by $Cov(w, Z')$.

[33] This is in effect what mutation parameters do.

Second, we should note one other fact about the treatment of heritability in terms of regression. A regression can have any value between plus and minus infinity. On the "fraction of variance" approach, heritabilities are between zero and one (inclusive). If we assume standard genetic mechanisms, heritabilities measured by regressions lie between zero and one and the two frameworks coincide. But if our aim is a treatment that covers all possible inheritance systems, we have to deal with the possibility, at least in principle, of regressions of offspring on parent that are negative or greater than one. In effect, we see that many discussions of heritability have assimilated two different phenomena, parent/offspring *similarity* and parent/offspring *predictability*. The former is a special case of the latter.

This shows, I think, a slight 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 that parents must "resemble" offspring.[34] But 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.[35] Cultural inheritance might be a domain where such parent/offspring anti-correlation may be common. Different aspects of the theory pull us in different ways with respect to the categorization of such cases. They look Darwinian when we are thinking about the abstract idea of a populational response to fitness differences; they look "anti-Darwinian" when we have an eye on the importance of cumulative selection and adaptation.

V. DRIFT AND CORRELATED RESPONSE

In this section I turn to 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.

[34] Endler (*op. cit.*) is an exception.

[35] See Jacquard, *op. cit.*, and M. Blute, "The Evolution of Replication" (forthcoming). Jacquard 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.

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 change in *Z* seems to be described by 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.[36] 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 in 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, which I will take as representative of a family of such approaches (*op. cit.*, chapter 1). Okasha again draws on the Price equation. He notes that in principle

[36] This fact has been noted frequently in discussions.

the fitness differences used in the Price equation 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 (absolute) 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.[37]

$$W_i = W_i^* + \delta_i, \tag{7}$$

This breakdown is introduced by Okasha into a Price equation. Change in mean phenotype can be represented (ignoring transmission bias) as a sum of $Cov(W^*,Z)$ and $Cov(\delta,Z)$. He claims this "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" (*op. cit.*, p. 33, and see his equation 1.4). This breakdown also answers 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.[38]

I doubt that the probabilistic breakdown achieves Okasha's goal, however. The breakdown in terms of *P*, if it is available, tells us about the extent to which realized fitnesses conform to expected values. But an expected value can itself be produced by accident. It can obtain without having a basis in the causal factors that *make* W^* the expected value. 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.

Once this problem with the "deviation from expectation" approach is clear, we seem pushed back towards the simple causal criterion

[37] $W^*_i = \Sigma_j P_{ij} j$. Here the *j*'s are numbers of offspring, and P_{ij} is the probability of *i* having *j* offspring.

[38] For the skeptical arguments, see Mohan Matthen and Andre Ariew, "Two Ways of Thinking about Fitness and Natural Selection," this JOURNAL, XLIX, 2 (February 2002): 53–83.

introduced above. I suggest that this is 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 discussed 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 include 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 $\mathrm{Cov}(Z', w)$ 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 do not think it is necessary at this point.[39] What is needed for present purposes is a way of distinguishing

[39] Here a manipulability approach might be used; see Judea Pearl, *Causality: Models, Reasoning, and Inference* (New York: Cambridge, 1998), and James Woodward, *Making Things Happen: A Theory of Causal Explanation* (New York: Oxford, 2003).

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.

So 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 (and presently the only available way) of expressing a crucial requirement in a fully general form.

VI. CONCLUSION

The range of problem cases discussed in this paper is not complete, but is sufficient to motivate some fairly definite conclusions.[40]

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 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.[41] This sort of discus-

[40] 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 (*op. cit.*). I should add that I *do not* think the summaries need to be narrowed in order to deal with spurious cases of change at higher or lower levels than the "real" level at which evolution is occurring, but that claim will not be defended here.

[41] See Philip Kitcher, "Explanatory Unification and the Causal Structure of the World," in Kitcher and Wesley Salmon, eds., *Minnesota Studies in the Philosophy of Science, Volume XIII* (Minneapolis: Minnesota UP, 1989), pp. 410–505.

sion can also devolve into terminological dispute of the unproductive kind. As argued in section II, we should resist the temptation to impose overly sharp borders.

So consider the following summary:

(S) Evolution by natural selection is change in a population due to:

(i) variation in the characteristics of members of the population,
(ii) which causes different rates of reproduction, and
(iii) which is inherited.

This summary allows that change may occur because of the inheritance system alone, or by various other mechanisms; ENS is change over and above that resulting from these other factors. The reference to "rates" is intended to measure output with respect to time, but this need not be the particular "growth rate" parameter discussed in section II. In a related vein, I assume that inheritance involves parent/offspring similarity (as opposed to any systematic relationship), but it is not assumed that a regression is always the appropriate measure of this similarity.

This summary does not include within ENS change to due to differences in survival that are not reflected in reproductive differences. Those include short-term change due to simple culling, and effects on a population due to some individuals living longer than others after all their reproduction is completed. This issue does not arise when discrete generations are assumed. But when discrete generations are not assumed, it becomes necessary to work out the relation between change due to survival differences *per se*, and change due to differences in reproduction. Ridley's summary rules out change due to survival differences *per se*. But as we saw, he makes a tacit idealizing assumption of discrete generations. Endler rules these cases in, but does so by giving two descriptions of change, E4 and E5, where E4 handles change due to differential survival and E5 handles the "response" across generations. So inheritance matters only to the change in E5. In the summary above I treat only change due to reproductive differences as ENS, but that is a decision of the kind described above—a decision about how broadly to extend the borders of an explanatory category in a short summary, given that there is an admitted grey area.

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 excessive idealizations that lead to a loss of contact with important cases. It is also possible to

stop trying to give such recipes in verbal form, trusting instead to a collection of equations (the Price equation, the one-locus diploid model, the replicator dynamics, the breeder's equation). But as emphasized in section v, a verbal summary does a better job with the causal component of the idea of selection than existing 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's 1980 formulation:

(IR) The following conditions are sufficient for evolution of trait *Z* by natural selection in a 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.

PETER GODFREY-SMITH

Harvard University