# Evolution and Natural Selection

A LARGE PROPORTION of the philosophy of biology is about evolutionary theory, as this part of biology unifies much of the rest, has a great deal to say about our place in the universe, and gives rise to many puzzles. Evolutionary change occurs at several scales. A standard way this is recognized is with a distinction between *microevolution* and *macroevolution*. Roughly, microevolution is change within a single species, and macroevolution is change in a collection of these units—a collection of species. This terminology makes the divide sound sharp, but rather than a situation where there are two distinct levels in nature, one can continuously "zoom in" and "zoom out" of what is going on in some region of space and time. As we do this, different patterns become visible. At a macroevolutionary scale, we find the "tree of life," a pattern of ancestry and descent linking all species on earth. Zooming in, we find change within the segments or twigs of the tree.

These relationships are represented in a diagram by the biologist Willi Hennig, reproduced in a modified form in Figure 3.1. Three scales are shown at once. At the most coarse-grained level, one species splits into two, giving rise to *phylogenetic* relationships between those species. Zooming in, this event is seen to be composed of many events involving relations between individual organisms, reproducing sexually. Change within each species is microevolutionary change. Zooming in still further, we encounter change within the life of a single organism. Those *ontogenetic* relationships are the subject of developmental biology.

## 3.1. EVOLUTION BY NATURAL SELECTION

In modern biology many concepts are important in explaining change within populations, but the one that generates most
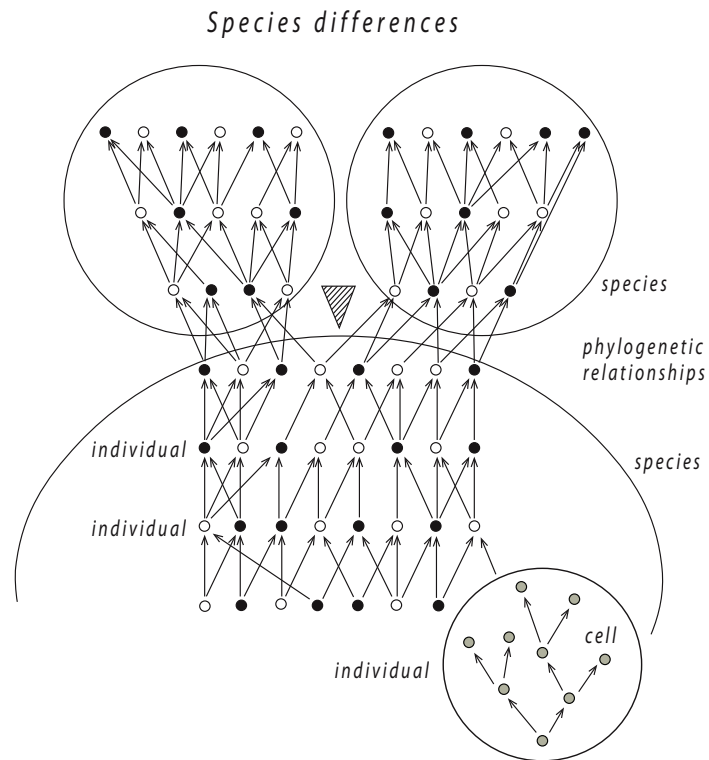
*Species differences*



Figure 3.1. Change is represented at three scales. A species splits into two, a break in a "fabric" of individual organisms tied together by sexual reproduction. Differences in the reproductive success of individuals are seen within the fabric. In Hennig's diagram, change within the life of a single organism, at the lower right, was represented as a sequence of stages. I have replaced the stages with cells, linked by cell division. (Figure adapted from Willi Hennig's *Phylogenetic Systematics* ©1979 by the Board of Trustees of the University of Illinois. Used with permission of the University of Illinois Press.)

controversy is *natural selection*. One of Darwin's breakthroughs was to see that a huge amount can be explained in terms of the repeated action of a simple set of factors. Here is one of his summaries, followed by a passage from the end of *On the Origin of Species*:

Can it, then, be thought improbable . . . that . . . variations useful in some way to each being in the great and

29

complex battle of life, should sometimes occur in the course of thousands of generations? If such do occur, can we doubt (remembering that many more individuals are born than can possibly survive) that individuals having any advantage, however slight, over others, would have the best chance of surviving and of procreating their kind? On the other hand, we may feel sure that any variation in the least degree injurious would be rigidly destroyed. This preservation of favorable variations and the rejection of injurious variations, I call Natural Selection. (1859, pp. 80–81)

Thus, from the war of nature, from famine and death, the most exalted object which we are capable of conceiving, namely, the production of the higher animals, directly follows. (1859, p. 490)

Compare this to a more recent summary by the geneticist Richard Lewontin:

A sufficient mechanism for evolution by natural selection is contained in three propositions:

1. There is variation in morphological, physiological, and behavioral traits among members of a species (the principle of variation).
2. 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).
3. Different variants leave different numbers of offspring either in immediate or remote generations (the principle of differential fitness).

[A]ll three conditions are necessary as well as sufficient conditions for evolution by natural selection. . . . Any trait for which the three principles apply may be expected to evolve. (1985, p. 76)

The two summaries have different forms, as well as using different language. Darwin's summary makes generalizations about actual species. Lewontin's is best read as a conditional statement: *if* a species has the three features he lists, then evolution will occur. Another difference involves Lewontin's second condition. Parents, he says, need to resemble their offspring. They might not resemble them greatly, as long as they resemble them more than they resemble unrelated individuals. In Darwin's summary here (though not in all the summaries he gave) this seems to be taken for granted; favored individuals will "procreate their kind." But it is possible for a useful new trait to arise, help the organisms that bear it, and *not* be inherited, in which case there is no reason for the population to change.

Neither summary says that variation has to appear "randomly." Natural selection can work in a situation where new variations tend in some direction, perhaps even toward useful traits. But new variation *can* be produced in a random, haphazard, or "blind" way, and natural selection will sift the good from the bad. In Lewontin's summary there is no reference to a "battle for life," as in Darwin; whether or not there is a battle, change can occur if some do better than others. Neither summary says anything about genes or other mechanisms for inheritance. That is not surprising in the case of Darwin, but Lewontin, a geneticist, also treats genes as optional. Finally, both summaries make it clear that change is driven by local, short-term advantage, not by any kind of progressive tendency or foresight. Evolution occurs through the accumulation of routine events—births, lives, matings, deaths.[1]

Lewontin's summary gives three conditions and says they are necessary and sufficient for evolution by natural selection. Is it true that *whenever* you have these conditions, a population will change? Not in every case. Once you allow that the pattern of inheritance can be noisy, it is possible for the pattern of inheritance to push in one direction while the fitness differences push

[1] I treat *sexual selection*, in which some individuals have features enabling them to achieve more matings than others, as a kind of natural selection, not as something distinct.

in another, leading to no net change. For example, suppose the taller individuals have slightly *more* offspring than shorter ones, but taller individuals also tend to have slightly *shorter* offspring than themselves while short individuals do not. The two can cancel, leaving the population as it was.[2]

One response to this is to say that any conditional about change (except perhaps in basic physics) includes a *ceteris paribus* clause—a requirement of "other things being equal." Perhaps, but I think something else is being illustrated. There is a trade-off operating. If we make definite assumptions about the pattern of inheritance, it's possible to give definite statements about how differences in reproduction will lead to change. But any description like that will cover only some cases. If we want to say something that captures *all* cases, the summary won't have the same causal transparency.

In chapter 2, I distinguished two kinds of general claims in biology that look to some extent like "laws." There are general statements about actual cases, and conditionals that assert what *would* happen if a certain setup was realized, whether or not this ever happens. We see a distinction of that kind here.[3] Darwin's summary, though expressed using questions, is an attempt to describe facts about actual species. We could modernize it, like this: "In every species on earth, variation continually arises. Some of these new traits tend to be inherited across generations, and some inherited traits are beneficial to survival and reproduction while

---

[2]Here is an example that is about as simple as possible, modified from one by Robert Brandon. Suppose a population has four individuals, two large (L) and two small (S). They reproduce asexually. Two generations, with parent-offspring relations represented with arrows, are pictured here:
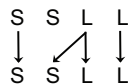


There is variation. Offspring tend, imperfectly, to resemble their parents. There are differences in reproductive success. But the new generation is the same as the old. Although there is heredity in Lewontin's sense, the pattern of inheritance itself pushes from large to small. This cancels the effect of the differences in reproductive success.

[3]Comments by Andreas Keller influenced my discussion in this paragraph.

others are not. In many cases, the traits beneficial to survival and reproduction become more common, while less useful traits are lost. This leads to ongoing change in the features of organisms in all species." It is also possible to look for a conditional: *If* such-and-such conditions hold, *then* a population will change, *guaranteed*. For example, if there is variation in a population, reproduction is asexual and offspring are exact copies of their mothers, everyone lives for the same length of time and reproduces at once, no one enters the population from outside or leaves, and individuals with some traits reproduce more than others, then the population will change. This is a verbal version of a mathematical model called the *replicator dynamics*, described in Box 3.1. This is sometimes seen as a foundational model of evolution (Nowak 2006a), and in a sense it is. But when applied to any real system, the model is an idealization, a deliberate simplification. Part of what Lewontin wanted to do in his summary is recognize that in many cases where the pattern of inheritance is noisy, evolution by natural selection can still occur. When you aim for generality of that kind, covering a wide range of systems, it is hard to make definite predictions. The replicator dynamics, on the other hand, is simple and gives precise predictions, but it is not a very realistic description of actual cases. The trade-offs operating here illustrate some general points made in the previous chapter (§2.3); descriptions that have the "simplicity and power of a general theorem," as Richard Levins put it, tend to be at odds with the "richness and the diversity of living nature."

Many debates about natural selection involve the concept of *fitness*. Evolution by natural selection is often said to be a matter of change due to fitness differences. The ordinary, nontechnical use of the term suggests two things, some sort of *fitted-ness* of an organism to its environment, and a kind of health or vigor. Talk of fitness was introduced to evolutionary theory in the 19th century by Herbert Spencer (1864), with the first of these meanings in mind. The term acquired a more technical role in the 20th century; or rather, it acquired several roles.

Lewontin's summary includes a "principle of differential fitness." But all Lewontin said was that some individuals "leave different numbers of offspring" than others. What if it is an *accident*

that some do better than others? Most evolutionary theorists recognize a distinction between change due to natural selection and change due to "drift"—accidental or random events that involve some individuals reproducing more than others. Lewontin seems to ignore this distinction. Here, in contrast, is a summary of natural selection by Alexander Rosenberg and D. M. Kaplan (2005, with their symbolism reduced a little here).

> *Principle of Natural Selection:* For all reproducing entities $x$ and $y$, all environments $E$, and all generations $n$: if $x$ is fitter than $y$ in environment $E$ at generation $n$, then probably there is some future generation $n'$, after which $x$ has more descendants than $y$.

Rosenberg and Kaplan treat fitness as something that *leads* to reproductive success. Note also that Rosenberg and Kaplan do not mention heritability, so they are focusing on just a part of what is covered by Lewontin.

The term "realized fitness" is often used for the actual reproductive output of an organism or a type of organism. The most influential way of understanding fitness in the other sense, the sense in which fitness explains or gives rise to reproductive success, is to see an organism's fitness as a *propensity* to have a certain number of offspring (Brandon 1978, Mills and Beatty 1979). A propensity is a tendency or disposition that can be described in terms of probabilities. A fair coin has a propensity to come up heads on roughly half the occasions it is tossed, even though it might always come up tails. Similarly, a fitter organism has a propensity to have more offspring than a less fit one. More technically, an organism's fitness can be seen as its *expected* number of offspring, where this expected value is calculated with probabilities that are interpreted as propensities. If an organism has a half chance of having no offspring and a half chance of having ten, its expected number of offspring is five. Very different organisms might have similar propensities to be reproductively successful.

Two kinds of problems arise with this view. First, there are cases where the expected number of offspring is not a good predictor of evolutionary change. I won't discuss those issues

here.[4] The second is that propensities are rather strange features of the world. Any organism has a realized fitness, its actual number of offspring. It has zero, one, ten, or whatever. That outcome is the result of all the actual events in its life, all the causal details. Do we have to believe that behind that number there is some *other* number of offspring that it was "expected" to have, where that number is not merely a reflection of our ignorance of details, but a real feature of the world?

We might believe this, but it surely seems optional from the point of view of evolutionary biology. If someone thinks that realized fitness is the only kind of fitness that makes sense, this person does not have to stop believing in natural selection. The situation, as I see it, is like this. In the Lewontin summary, the Rosenberg/ Kaplan summary, and others, the term "fitness" is applied to different parts of a causal sequence that biologists generally agree about. They agree that organisms live in different environments and have different ways of making a living. They agree that in all these cases, variations arise that in *some* way or other lead to an advantage in survival and reproduction. "Advantage" might be understood in terms of probability, or in some other way. In some cases where a new trait gives the organisms that bear it an advantage in survival and reproduction, those organisms will actually have more offspring. If the trait is heritable, then in many cases the population will change. All that is common ground. Talk of "fitness" is sometimes applied to the possession of a particular structural or behavioral feature that is useful in the case being investigated, sometimes to a propensity to succeed, and sometimes to actual reproductive success. A biologist might be wary of *all* talk about probability when dealing with macroscopic events, thinking that probabilities are just reflections of our ignorance. Indeed, I think it is reasonable to be a bit suspicious of standard distinctions between change due to natural selection and change due to "drift" or "accident." What we call "accidental" and "random" events have ordinary physical causes (unless we are talking about events at the microphysical level, which may be fundamentally indeterministic). Sometimes there is more regularity, more

---

[4]See Gillespie (1977), Sober (2001), Abrams (2009).

of a pattern, in who does well and who does badly, and sometimes there is less. Someone who is skeptical about standard distinctions between selection and drift might want to talk of fitness only in the "realized" sense, as seen in the Lewontin summary.

---

## BOX 3.1. MODELS OF EVOLUTION BY NATURAL SELECTION

The simplest mathematical model of evolution by natural selection is the "replicator dynamics" (Taylor and Jonker 1978, Weibull 1995, Nowak 2006a). Suppose there is a large population containing just two types, $A$ and $B$, with frequencies $p$ and $(1-p)$, respectively. Individuals reproduce asexually and simultaneously, with the parents dying right after reproduction. Then if $W_A$ and $W_B$ are the average numbers of offspring produced respectively by the $A$ and $B$ types, the *new* frequency of the $A$ type after one generation, $p'$, is related to the old frequency by this rule: $p' = pW_A/(pW_A + (1 - p)W_B)$. This model assumes that both types copy themselves exactly when they reproduce, and that other factors such as mutation and migration into the population are absent. It also treats generations as discrete steps. Other versions of the replicator dynamics treat time as continuous, not as a sequence of steps. The case with large and small individuals discussed in note 2 of this chapter where there were fitness differences and heredity but no change does not fit the assumptions of this model, as an $L$ gave rise to a $S$. When applied to almost any real system, even asexual organisms like bacteria, this model is an idealization.

A more general way of representing evolution is with the "Price equation" (Price 1970, 1972, Okasha 2006, Frank 2012). This framework is more complicated than the model above, in part because it approaches populations in a different way, by tracking every individual and describing the statistical relations between "before and after" states. Assume there is an *ancestral* collection and a *descendant* collection of individuals, where all individuals can be described in terms of their value of a quantitative characteristic, $Z$ (which might be size, for example), and assume a relation (usually interpreted as reproduction, though

it can be understood in other ways, including persistence) linking the ancestors to their descendants. The aim is to represent the difference between the descendant and ancestral collections in their average values of $Z$, a difference represented as $\Delta \bar{Z}$. One version of the Price equation is this: $\Delta \bar{Z} = Cov(Z,W) + E_W(\Delta Z)$. Here $Cov(Z,W)$ is the covariance in the ancestral population between each individual's value of $Z$ and their value of $W$, which is the number of descendants that individual is connected to, divided by the average number of descendants that ancestors have. So this first term on the right-hand side, sometimes called the *selection* term, represents the role of fitness differences; do individuals with a high value of $Z$ have more (or fewer) descendants than others? The term $E_W(\Delta Z)$ measures the average change in $Z$ that occurs between ancestors and the descendants they are connected to, where the average is weighted by the relative fitness of each ancestor. This term represents the role of the inheritance system.

This model does not assume copying, and the equation can be applied to sexual reproduction. It is an *abstract* description of evolution, leaving many things out, but not an *idealized* one; it can be applied to real cases without simplifying them. Unlike the replicator dynamics, though, the output of the equation cannot in every case be fed back into the equation as a new *input*, giving a model that applies over many time steps.

The example with fitness differences and heredity but no change in note 2 of this chapter can be described with a Price equation. Think of the large individuals as having the value $Z = 2$ and the small ones as $Z = 1$. The effect of the first term, which represents the effect of differential reproduction, is exactly balanced by the second term, which represents the failure of offspring to resemble their parents. So $\bar{Z}$, the average value of $Z$, remains unchanged. A Price equation can be used to represent evolutionary change at several part-whole levels in a system simultaneously, as the term on the far right-hand side can often be broken down into two terms that represent the roles of fitness differences and inheritance in entities at a lower level.

## 3.2. Origin explanations and distribution explanations

Natural selection is often described as the key to understanding how complex organisms can come to exist as a result of natural processes. But natural selection is also often described as a "filter": once variations have arisen, a few are kept while others are lost. A process of filtering cannot create anything, and assumes the existence of the things being filtered. Is it a mistake to think that selection can have something like a *creative* role in evolution?

The view that Darwin discovered a purely negative factor has been expressed often. An early example is Hugo de Vries, a biologist at the turn of the 20th century who was important in the history of genetics. De Vries noted that "in order to be selected, a change must first have been produced" (1909).

> [Natural selection] is only a sieve, and not a force of nature, not a direct cause of improvement. . . . [W]ith the single steps of evolution it has nothing to do. Only after the step has been taken, the sieve acts, eliminating the unfit. (1906, pp. 6–7)

To look more closely I will introduce some terminology, distinguishing between *origin explanations* and *distribution explanations*.[5] When we give a distribution explanation we *assume* the existence of a set of variants in a population, and explain why they have the distribution they do, or why their distribution has changed. Some variants may be common, while others are rare. Some may have been lost from the population, having been present before. An *origin* explanation, in contrast, is directed on the fact that a population has come to contain individuals of a particular kind *at all*. It does not matter how many there are; the point is just to tell us how there came to be some rather than none. So now we are explaining the original appearance of the variants that are taken for granted when giving a distribution explanation.

[5] This terminology is modified from one used by Karen Neander (1995).

Almost everyone agrees that natural selection can figure in distribution explanations. It initially seems that selection has no role in origin explanations, as selection can sort only things that already exist. This would not mean that evolutionary biology as a whole cannot give origin explanations. They would be given in terms of what we now call "mutation," along with the recombination of characteristics through sex. (De Vries was the person who introduced this modern use of the term "mutation.") Perhaps selection is a distribution-explainer while mutation is an origin-explainer.

I think, in contrast, that selection is essential to many origin explanations, and in a way that does give it a creative role in the evolutionary process. Part of Darwin's achievement was seeing this fact, and he was, as far as we know, the first person who saw it.

Selection is not an immediate, or proximate, cause of a new variant. The most important immediate sources of new variations, again, are mutation and recombination. However, natural selection can reshape a population in a way that makes a given variant *more likely to be produced* by the immediate sources of variation than it otherwise would be. As selection changes the background in which mutation and recombination operate, it changes what those factors can produce.

Suppose we are explaining the evolution of the human eye. Building the genetic basis of the human eye involved bringing together many genes. Consider a collection of genetic material, $X$, that has everything needed, as far as genes go, to make an eye, except for one mutation. So this background $X$ is such that *if* a particular new mutation arises against $X$, it will finalize the evolution of the eye. Initially, $X$ was rare in the population—it was the product of a mutational event that produced $X$ from another precursor, $W$. Selection can make the appearance of the eye more *likely* by making $X$ more *common*. This increases the number of "independent experiments" where a single mutation can give rise to the eye. If $X$ remains rare in the population, then additional mutations are much less likely to produce the eye, as the right mutation has to occur in exactly the right place—in an lineage where $X$ happens to be present. Selection makes the eye accessible to mutation in a way it would not otherwise be.

In that example I told the story working backward from a trait of interest. The process itself runs forward, without foresight, and involves many of these steps. When I call something an "intermediate" or "precursor," these terms apply only in retrospect, and the story can also be told without them. There is a population at time *t*, which contains variation. Some traits are useful to the organisms that bear them and others are not. They are useful for what they do at time *t*, not for what they might lead to later. The useful ones increase and their increase creates many sites at which further new variants arise. *Whatever* is favored at time *t* changes the background in which further mutations appear. Sometimes this process leads nowhere that strikes us as noteworthy, but sometimes it produces eyes and brains.

So selection can have a creative role, even though it is true in every case that "in order to be selected, a change must first have been produced," as de Vries put it.[6] The point can be made even more simply: if you can get to *Y* easily from *X*, but with difficulty from *W*, then you can make *Y* more likely to arise by having lots of *X* around and few *W*, as compared to the situation where you have lots of *W* and few *X*. As Patrick Forber (2005) notes, in a biological context this usually requires that trait *Y* be the product of many genes, or at least a lot of DNA. To the extent that a new trait can arise as a unit through a single change to any background, selection does not make it more likely to appear. But that is not how things are with eyes and brains, whose evolution involved changes to a great deal of DNA.

You might say at this point that it is not *selection itself* that does the originating; that is still due to mutation. Let's then make a three-way comparison, comparing mutation alone, selection alone, and mutation and selection together. Selection alone cannot produce new things, though it can keep the good ones that are already around. Mutation alone can produce new things, but in an indiscriminate way. There is almost no chance of it producing eyes and brains. Selection and mutation together can produce

---

[6]Compare de Vries to Herbert Spencer, in 1864: "To him [Darwin] we owe the discovery that natural selection is capable of *producing* fitness between organisms and their circumstances" (p. 446).

eyes and brains. So you might say it is *only the combination* of selection and mutation that is creative, and that would be fine. It might then be added that selection is *as* creative as mutation is. Perhaps that is exaggerating, though, as there is a tiny chance of mutation alone producing a complex new trait and selection alone cannot do that. And perhaps it is just wrong to say that the parts are "creative" when it is only the combination, selection plus mutation, that plays the crucial role. But the view that selection is only a distribution-explainer while mutation is an origin-explainer is wrong.

As discussed in the first chapter, Darwin had predecessors who glimpsed the idea of variation and selection but did not do much with it. One reason is that their hypotheses did not *iterate* the process. The cosmologies of Empedocles and Lucretius, for example, posited a special period at the beginning of the world in which variation appeared, followed by the culling of monsters. If there is no process where the results of selection feed back on *another* round of variation, there is no role for selection in explaining the origination of new structures.

A difference can also be described between Darwin's work and the "neo-Darwinism" that followed in the 20th century. Darwin's emphasis is on origin explanations. The distribution explanations he gives are simple: a new variant appears, and it either spreads or is lost. The iteration of many of these events explains how new kinds of organisms come to exist. From the 1930s onward, more sophisticated distribution explanations appear, made possible by Mendelian genetics. In writers like Fisher (1930), Haldane (1932), and Wright (1931), we see the idea of a discrete particle, a gene, inherited intact over many generations, coming into new combinations with other genes, and becoming more or less common—perhaps reaching a stable equilibrium frequency—in a gene pool.

I'll make a last point about origin explanations. Selection, I said, can make the evolution of eyes more likely by making eye precursors more common. But "common" is ambiguous—a trait might become more common in *relative* terms or in *absolute* terms. Natural selection is often described in terms of its effects on frequencies. Type *A* is favored by selection if it becomes more common relative to *B*, whether or not there are more *A*s

than there were before. That is fine in the context of distribution explanations, where one might either care or not about absolute numbers. But for selection to make the evolution of the eye more likely than it was before, it has to increase the *absolute* numbers of eye precursors. Then we see that the popular metaphor of selection as a "sieve" or "filter" is not a good one. In most cases of natural selection, some types decrease in numbers, and some types *increase*. Selection filters out some variants and amplifies others.

Suppose we have a population of eye precursors and non-precursors. Something that is not usually acknowledged here is the fact that the evolution of the eye could be made more common either by increasing the numbers of eye precursors alone, or by increasing the numbers of *all* types. In some environments, for a while, this can happen. When rabbits were first brought to Australia, in colonial times, they increased explosively in numbers.[7] The fittest rabbits certainly proliferated, but many of the less fit proliferated too. Eventually a situation is reached where if one type increases in numbers, another has to decrease. The "struggle for life," which Darwin emphasized, becomes relevant. In some modern discussions the struggle for life is treated as an inessential part of Darwin's theory, something that came from the influence of the pessimistic Malthus and Darwin's 19th-century context. But the fact of scarce resources—when it is a fact—ties relative reproductive success and absolute reproductive success together. Selection in Darwin's sense is as much an amplifier as a filter, and it is the amplifying that matters to its creative role.

## 3.3. UNITS OF SELECTION

The theory of evolution by natural selection, in the form discussed so far, is aimed at explaining how change takes place within populations of organisms. Organisms vary, organisms pass on traits, organisms differ in reproductive success, and the population or species changes as a result. But it was quickly seen

[7] There was a lag of about 70 years between their first introduction in 1788 and the explosive increase, which has apparently not been explained.

that a Darwinian pattern of explanation might be applied to other things. This is often explicit in summaries; Rosenberg and Kaplan's principle from section 3.1 is said to apply to all "reproducing systems." Here is another: "Evolution can occur whenever there are units of reproduction that produce other such units which inherit some characteristics of the parent units" (Doebeli and Ispolatov 2010, p. 676).

Some of the applications of this idea are inside biology: evolution by natural selection might operate at many levels in a hierarchy of parts and wholes; it might operate on genes, cells, organisms, groups, and perhaps species. Another set of applications lies outside biology. Natural selection has been seen working on ideas, technologies, economic firms in a market, and patterns of behavior in a culture. This section looks at the biological side (which is continued in chapter 5), and the next section looks at other applications.

Biology in the 20th century developed Darwinism by representing evolutionary change at the level of genes. This sometimes led to the idea that evolutionary change *is* change in frequency of genes in a gene pool. A rigorous version of this view was developed by George Williams (1966). Williams did so as part of a critique of explanations in terms of "the good of the species," and the good of other large units such as ecosystems and populations. Might cooperation, altruism, and restraint evolve because they make whole groups or species better adapted than their selfish rivals? No, said Williams; even if restraint or altruism does make a group "better" in some sense, that will not stop a selfish mutant from *invading* a harmonious group and flourishing at the expense of its well-behaved fellows. The lower-level process of competition within such a group will usually overwhelm any advantage the group might have as a unit.

As well as criticizing explanations in terms of group-level benefit, Williams argued that *all* evolutionary processes, even familiar ones in which organisms compete within a population, can be understood at the genetic level; in every case, one *allele* (alternative form of a gene) increases in frequency because it has some overall or net advantage over rival alleles at its *locus* (its place in the genome), as a consequence of the totality of effects the allele

43

has on cells and organisms that contain it. Richard Dawkins (1976) defended a colorful and grim version of this view, seeing organisms like ourselves as "gigantic lumbering robots" programmed by our genes (p. 21). For Dawkins, all evolution is the result of long struggles between selfish genes. Genes can persist, in the form of copies, while organisms and groups come and go "like clouds in the sky, or dust storms in the desert" (p. 34). Though organisms like ourselves are important parts of the living world, we are not *units of selection*, and whatever evolves is not for *our* evolutionary benefit, but for the benefit of our genes.

One reply to this argument is that although it is generally possible to "track" an evolutionary process by seeing what is happening to the frequencies of genes, it is not possible to *explain* what is happening by staying at the genetic level. Changes to gene frequencies are usually a result of the lives and deaths of whole organisms, and are sometimes affected by competition between larger units such as families and tribes. Most of the time, it is larger entities, like organisms, that natural selection "sees," not genes (Gould 1980, Sober and Lewontin 1982).

It is starting to look like there is an ambiguity of some kind, a failure to separate issues, behind the dispute. Modifying Dawkins's analysis, David Hull (1980) distinguished two senses of "unit of selection." In any evolutionary process, Hull said, two roles are seen. These roles may be occupied by the same things or by different things. First there must be *replicators*, things that are copied reliably over generations. Second there must be *interactors*, things whose activities and interactions with the environment affect which replicators are copied at a higher rate.[8] In the case of evolution in humans, genes are replicators and organisms are interactors. But sometimes groups or even species might be interactors, sometimes cells or genes might be. As for the replicators, these are usually genes, sometimes asexual organisms (for Hull, not Dawkins), and a few other things, but not organisms like us, because sexually reproducing organisms do not *copy* themselves. We pass on genes that are always entering into unique combinations.

[8] This is similar to Dawkins's concept of a *vehicle* (1982).

On this analysis, the people arguing that organisms and groups are important parts of evolutionary processes might be right about their importance as *interactors*, but this does not change the fact that genes are the *replicators*.

This view seems to clear up confusion and was embraced in both biology and philosophy (Lloyd 1988, Sober and Wilson 1998, Gould 2002). I used to think it is helpful, but now I think it is mistaken (though part of this framework will return with a possible new role in chapter 5). The quickest way to see there is something wrong is to look at the Lewontin summary given earlier. This summary had problems of detail, but it describes all that is needed for evolution by natural selection. And in that analysis, there are not two kinds of things, but one: the entities in the population that vary, inherit traits from their parents, and differ in reproductive success. If we have things with *those* properties, that is all that is needed. The passing on of "replicators" is one possible mechanism for inheritance, but it is optional. If genes are being passed from generation to generation, then genes might *themselves* satisfy the three conditions of variation, heritability, and differential reproduction, but to say that is to drop the replicator/interactor distinction and apply the same criteria to organisms, genes, groups, and everything else.

That I think is the right approach, but this is not the end of the units of selection problem; it will return in chapters 5, 6, and 8.

## 3.4. Universal Darwinism

Once the idea of variation and selection snaps into focus, it is tempting to apply it to many systems. The Darwinian dynamic, or something like it, has been seen in scientific change, technological change, individual learning, and elsewhere. One of the quotes I gave earlier summarizing evolution by natural selection— "Evolution can occur whenever there are units of reproduction that produce other such units which inherit some characteristics of the parent units"—is from an article about *religion*; it gives a theory of how religions compete and spread, and not through any kind of "religion gene" but by cultural processes. Are these

analogies superficial, perhaps even mistaken, or do Darwinian ideas help us to understand change in these other systems too?

It is possible to use Darwinism as the basis for a general theory of all change of a certain kind. I'll call this category *adaptive* change—change that involves improvement to the design of a system or its ability to deal with its environment. (The idea of adaptation will be treated more warily in the next chapter.) This broadening of Darwinism can be both explained and motivated by looking at examples. First, here is a theory of learning, in humans and animals, developed especially by Edward Thorndike (1911) and B. F. Skinner (1974): Organisms often produce new behaviors in a haphazard way, trying out new things from time to time. If a behavior has good consequences in a given situation, it is retained. When that situation arises again, the organism is more likely to produce the same behavior.

Second, here is an account of scientific change itself. Karl Popper (1959) thought that change in science occurs by an endless cycle of *conjecture and refutation*. Scientists imaginatively propose new theories, going far beyond the data, and then try to refute these theories by collecting further observational evidence. Conjecture and refutation; trial and error; mutation and selection.

A third example takes us back to biology. How do our bodies learn to recognize invading viruses and bacteria? How does "adaptive immunity" work in organisms like us? Early proposals had it that the body somehow receives the impression of the invader, like the stamping of a form on a wax tablet. Nils Jerne (1955) and Macfarlane Burnet (1958) proposed instead that the immune system uses a mechanism of variation and selection. It produces many different antibodies in a "random" way, and when a cell happens to make antibodies that can bind to an invader, these cells are made to proliferate at the expense of others. Later work on the immune response has complicated this picture, but the basic insight has stood up. The development of the brain in infancy and early childhood is also thought to work by a selection process, though here I think the analogy is becoming weaker: the way to wire up a growing brain is to start with *too many* connections between neurons, and trim away the ones that do not serve a useful role while strengthening the ones that do

(Changeux 1985, Edelman 1987). A more speculative theory in this family is the idea that cultural change, especially improvement in skills and technology, occurs by a process in which new ideas and behaviors pop up from time to time, and some spread because they are *imitated* more than others. I'll look at this one in chapter 8.

Given all this, it is tempting to offer a grand theory: *whenever adaptive change occurs, some process of variation and selection is responsible.* This is a kind of "universal Darwinism." Views along these lines have been developed by the psychologist Donald Campbell (1960), Daniel Dennett, in philosophy (1974), and others.

At one stage in the 20th century it must indeed have seemed that *everything* was turning out to be variation and selection. The argument is harder to make now. The Thorndike-Skinner theory of learning and Popper's theory of science are not widely accepted in their respective fields. Both are oversimplified, but that is not the heart of their problems. They overstate the importance of pure trial and error. Sometimes variation and selection at one time scale builds another system that can adapt to the world in ways that are *not* made up of more variation and selection (Amundson 1989). Evolution by natural selection built our brains, and maybe nothing else could. But once it has done so, our brains can do things that are smarter than just throwing out new behaviors—or beliefs—and seeing if they work. We can engage in logical reasoning and planning (at least some of the time), and shape ideas and behaviors without *exposing* them at every step. Sometimes variation and selection builds more variation and selection, as in the vertebrate immune system, and sometimes it builds something else.

So there is also a more moderate "universal Darwinism": whenever we have a system that can undergo adaptive change, there must be variation and selection *somewhere* in the story, but one variation-and-selection process can build machinery that creates further improvements by working differently. (Richard Dawkins, who coined the term "universal Darwinism," had something like this second view in mind.)

Setting aside the most ambitious views, it is fruitful to keep looking at the relationships between different selection processes. We can start by recognizing a very general category: systems in

which there is variation, and where successful variants become more common or are more likely to be retained. A two-way distinction can then be made within this class. On one side we have cases where the way successful variants are retained is through reproduction *by* those entities. This includes biological evolution. On the other side are cases where the retention of successful variants is done by a more centralized mechanism. Trial and error learning is like this; a successful behavior does not *make more behaviors*. Rather, something in the brain registers the good results that came from a behavior and generates similar behavior on later occasions. To the extent that cultural change in a human society involves the retention and passing on of successful innovations, perhaps a mixture of both modes is present, along with other things. In the case of cultural change and elsewhere, though people often argue about whether such and such a process is or is not "Darwinian," what we find is many differences of degree. Processes can be more or less Darwinian, and can shift, either rapidly or slowly, with respect to this status. Several dimensions are relevant here. Is new variation produced in an undirected way? Does the way that variants are retained allow for the accumulation of small improvements? Do successful variants spread by reproducing, or in some other manner?

Once the power of an iterated variation-and-selection mechanism is seen, it is surprising that the history of theories of this kind goes *from* Darwin's biology *to* other applications. Evolutionary change in biology is slow and inaccessible, and the role of variation and selection there must be detected among a great deal of noise. Darwin was helped by analogies to plant and animal breeding in farming, but in learning by trial and error, or the spread of an invention by imitation, everything happens on a time scale that is much easier to observe. Once attention is drawn to these things, there can be no doubt of the role of *some* kind of variation and selection. But—to pick one lineage as an example—the theories of learning and knowledge in people like Locke, Hume, and Kant missed this idea entirely.[9] There is an alternative

---

[9]Here is a point of near-contact in David Hume, writing about social patterns: "Two men who pull the oars of a boat, do it by an agreement or convention,

history, another possible world, in which Darwin or someone was able to draw on an understanding of variation and selection in a range of more obvious areas and apply it to the less obvious case of biological change. In the actual history, the more difficult application came first, and others came later.

Further reading

For relations between Darwin and modern views, Lewens (2010), Sober (2011); for fitness and drift, Ariew and Lewontin (2004), Walsh et al. (2002), Millstein (2006), McShea and Brandon (2010); for evolutionary explanation, Beatty (2006); for units of selection, Lloyd (2001), Okasha (2006); for general theories of selection, Hull et al. (2001); for other applications of Darwinism, Dennett (1995), Wilson (2002), Hodgson and Knudsen (2010).

---

although they have never given promises to each other. Nor is the rule concerning the stability of possessions the less derived from human conventions, that it arises gradually, and acquires force by a slow progression, *and by our repeated experience of the inconveniences of transgressing it. . . .* In like manner are languages gradually established by human conventions without any promise. In like manner do gold and silver become the common measures of exchange" (1739, bk. 3, pt. 2, §2, italics added).