

# Did Darwin write the *Origin* backwards?

Elliott Sober<sup>1</sup>

Philosophy Department, University of Wisconsin, Madison, WI 53706

After clarifying how Darwin understood natural selection and common ancestry, I consider how the two concepts are related in his theory. I argue that common ancestry has evidential priority. Arguments about natural selection often make use of the assumption of common ancestry, whereas arguments for common ancestry do not require the assumption that natural selection has been at work. In fact, Darwin held that the key evidence for common ancestry comes from characters whose evolution is *not* caused by natural selection. This raises the question of why Darwin puts natural selection first and foremost in the *Origin*.

common ancestry | evidence | likelihood | natural selection

## What is Darwin's Theory?

To characterize Darwin's theory, what could be more natural than to cite the title that Darwin gave to his own book (1)? How could this formulation lead us astray? In fact, there is trouble here, and it is of Darwin's own making. Although Darwin (ref. 1, p. 1) says that the origin of species is the "mystery of mysteries" that he proposes to solve, his solution of the problem is in some ways a dissolution. I say this because Darwin had doubts about the species category; he regarded the difference between species and varieties as arbitrary. When 2 populations split from a common ancestor and diverge from each other under the influence of different selection pressures, they begin as 2 populations from the same variety, then they become 2 varieties of the same species, and finally they reach the point where they count as different species. It is convenience, not fact, that leads us to classify different degrees of divergence in different ways (ref. 1, pp. 48–52). This vague boundary between variety and species is no reason to deny the existence of individual species, nor did Darwin do so (2, 36). This is the lesson we learn from other vague concepts – from rich and poor, hairy and bald, tall and short; a vague boundary does not entail that no one is rich, or hairy, or tall. Even so, "species" is not the central concept in Darwin's theory. True, the process he describes produces species, but it produces traits and taxa at all levels of organization. For these reasons, Darwin's theory is better described as "the origin of diversity by means of natural selection."

Darwin's concept of natural selection has several noteworthy features. Although the *Origin* introduced the idea of natural selection by first describing artificial selection, Darwin hastened to emphasize that natural selection is not an agent who intentionally chooses. When cold climate causes polar bears to evolve longer fur, the weather is not an intelligent designer who wants polar bears to change. The weather kills some bears while allowing others to survive, but the weather does not need to have a mind to do this. It is in this sense that natural selection is a mindless process (for a different assessment, see ref. 3). So concerned was Darwin to emphasize this point that, in the 5th edition of the *Origin*, he followed Alfred Russel Wallace's advice and used Herbert Spencer's phrase "the survival of the fittest" to characterize his theory (4). Darwin hoped this new label would make it harder for readers to misunderstand his theory.

Another important feature of Darwin's concept is that the direction in which selection causes populations to evolve depends on accidents of the environment. There is no inherent tendency for life to grow bigger or faster or harder or slimier or smarter. Everything depends on which traits do a better job of allowing organisms to survive and reproduce in their environments. This is the vital contrast that separates Darwin from Lamarck, who saw

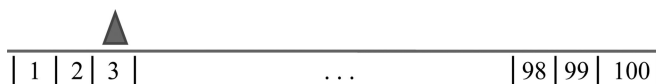


Fig. 1. A random walk on a line with 100 locations. Unless the marker is at the left-most or the right-most location, it moves 1 space to the right if the tossed coin lands heads and 1 space to the left if the coin lands tails.

evolution as leading lineages to move through a preprogrammed sequence of steps, from simple to complex. Of course, if life starts simple, evolution by natural selection will lead the average complexity of the biota to increase. However, that is not because the "laws of motion" of natural selection inherently favor complexity. Parasites evolve from free-living ancestors, and the effect is often a move toward greater simplicity, with parasites losing organs and abilities possessed by their ancestors (ref. 1, p. 148). Complexity increases from life's beginning because of the initial conditions, not the laws. This is analogous to the random walk depicted in Fig. 1. A marker on a line changes position as a result of a coin toss. If the coin lands heads, you move the marker one space to the right; if the coin lands tails, you move the marker one space to the left. These are the rules of change unless the marker happens to be at the left-most or the right-most points. If the coin lands tails when the marker is at the extreme left, you simply toss again. Suppose the game begins with the marker placed at the left-most point on the line. Where do you expect the marker to be after 5 or 50 or 500 coin tosses? Probably not at square one. The line in this game represents complexity, with 1 being the least complex and 100 the most. Selection can be indifferent to simplicity versus complexity and yet evolution by natural selection can be expected to manifest a net increase in complexity (5).

A third important feature of Darwin's concept is that selection acts on "random" variation. This is a loaded word, apt to mislead. Darwin says in the *Origin* (ref. 1, p. 131) that "random" just means that the cause of a new variant's appearance in a population is unknown. However, "random" for Darwin was more than a confession of ignorance. What he meant was that variations do not occur because they would be useful to the organism in which they occur. In *The Variation of Animals and Plants under Domestication*, Darwin explains his point in terms of a beautiful analogy:

Let an architect be compelled to build an edifice with uncut stones, fallen from a precipice. The shape of each fragment may be called accidental; yet the shape of each has been determined by the force of gravity, the nature of the rock, and the slope of the precipice,—events and circumstances all of which depend on natural laws; but there is no relation between these laws and the purpose for which each fragment is used by the builder. In the same manner the variations of each creature are determined by fixed and

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "In the Light of Evolution III: Two Centuries of Darwin," held January 16–17, 2009, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and audio files of most presentations are available on the NAS web site at [www.nasonline.org/Sackler-Darwin](http://www.nasonline.org/Sackler-Darwin).

Author contributions: E.S. designed research, performed research, and wrote the paper.

The author declares no conflict of interest.

This article is a PNAS Direct Submission.

A Spanish translation of this article appears in *Teorema*, XXVIII/2, 2009, pp. 45–69. [BIBLID 0210-1602 (2009) 28:2; pp. 45–69].

<sup>1</sup>E-mail: [ersoer@wisc.edu](mailto:ersoer@wisc.edu).

immutable laws; but these bear no relation to the living structure which is slowly built up through the power of natural selection, whether this be natural or artificial selection.

Ref. 6, p. 236.

A fourth important feature concerns the level at which Darwin took natural selection to act. In almost all of the examples that Darwin discusses, traits are said to be selected because they help the individual organisms that possess them to survive and reproduce. Tigers have sharp teeth because tigers with sharp teeth do better than tigers with dull teeth. The reason the trait evolved is not that sharp teeth help the species to avoid extinction or somehow keep the ecosystem from collapsing. In examples of this sort, Darwin embraces what biologists now call “individual” selection. An exception to this pattern of thinking occurs when Darwin considers the evolution of human morality. Why do human beings often sacrifice their welfare for the good of the group? This is how Darwin sets the problem in the *Descent of Man*:

It is extremely doubtful whether the offspring of the more sympathetic and benevolent parents, or of those which were the most faithful to their comrades, would be reared in greater number than the children of selfish and treacherous parents of the same tribe. He who was ready to sacrifice his life, as many a savage has been, rather than betray his comrades, would often leave no offspring to inherit his noble nature. The bravest men, who were always willing to come to the front in war, and who freely risked their lives for others, would on an average perish in larger number than other men.

Ref. 7, p. 163.

Then he proposes his solution:

It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an increase in the number of well-endowed men and an advancement in the standard of morality will certainly give an immense advantage to one tribe over another. A tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection. At all times throughout the world tribes have supplanted other tribes; and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase.

Ref. 7, p. 166.

Here, Darwin invokes the hypothesis of group selection. When groups compete, characteristics that are deleterious to the individuals who have them can evolve because they are good for the group in which they occur. Biologists now call such traits “altruistic.” For Darwin, natural selection can involve both individual and group selection.

Darwin discusses 2 examples of altruism in the *Origin*—the barbed stinger of honey bees and the sterility of workers found in many species of social insect (ref. 1, pp. 202 and 236). Both traits are deleterious to the individuals that have them. Bees that sting intruders to the nest eviscerate themselves; sterile workers have a reproductive success of zero. In each case, Darwin explains the trait’s evolution by pointing out that it is advantageous to the community. Some modern commentators interpret Darwin’s discussion of these traits as anticipating the idea of kin selection, which they view as a type of individual, not group, selection (8). Others regard kin selection as a kind of group selection and so regard

Darwin’s theorizing about barbed stingers and worker sterility as following the same pattern he later used to think about human morality (9). For those who prefer the former interpretation, there is an interesting interpretive question: Why did Darwin embrace group selection to account for human morality, but decline to do so in connection with the stinger and the sterility?

Regardless of how one interprets this small handful of examples, it is clear that Darwin invoked group selection hypotheses only rarely. Was this because he thought that group selection occurs more rarely and is a less important cause of evolution than individual selection? In the first edition of the *Origin*, Darwin (ref. 1, p. 87) does make a general comment about selection’s effect on traits that are good for the group. He says that “in social animals it will adapt the structure of each individual for the benefit of the community; if each in consequence profits by the selected change.” This is not an endorsement of group selection, since traits that are good for the group can evolve by individual selection if they also happen to be good for the individuals who have them. However, in the 6th edition of 1872, Darwin revised this sentence to read: “in social animals it [selection] will adapt the structure of each individual for the benefit of the community; if the community in consequence profits by the selected change (ref. 4, p. 172).” This is an endorsement of the general role played by group selection.

The last facet of Darwin’s concept of natural selection that I want to mention concerns his comment in the *Origin* (ref. 1, p. 6) that selection is “the main but not the exclusive cause” of evolution. One part of this pronouncement is clearer than the other. The idea that selection is not the exclusive cause of evolution just means that there are other causes. Darwin (ref. 1, pp. 134–139) allows for the Lamarckian mechanism of “use and disuse,” the inheritance by offspring of traits (phenotypes, in modern parlance) because they were acquired by their parents and turned out to be useful. A standard example is the blacksmith’s growing big muscles because of his work and then transmitting these big muscles to his children, who develop those muscles without needing to do what their father did to get them. Darwin also had the idea that descendants retain the traits that their ancestors had, sometimes even though these traits are no longer favored by selection. This is the idea of “ancestral influence”; it explains many rudimentary features (ref. 1, pp. 199, 416, and 450–456) as vestiges of a bygone age; for example, this is why human beings have tail bones and why human fetuses have gill slits (ref. 1, p. 191). Darwin also discusses correlation of characters as a cause of evolution. If a trait favored by selection is correlated with a trait that is neutral or even deleterious, the latter may evolve by piggybacking on the former (ref. 1, pp. 143–147). To use a modern example, our blood is red, not because the color promotes survival and reproduction, but because hemoglobin is red, and hemoglobin was selected for its ability to transport oxygen. Darwin discusses other nonselective causes of evolution, but the point is clear—he denied that selection is the only cause of evolution.

Unfortunately, Darwin does not explain what he meant by saying that selection is the “main.” I take it that if selection is the main cause, then it is the most important cause. This might mean that selection is the most frequent—that selection is implicated in the evolution of more traits in more populations than any other cause. Or it could mean that selection is more powerful than the other causes that affect the evolution of a given trait. Here we must consider the different causes that influence a trait’s evolution in a population and then imagine how the outcome would have been different if selection had been absent and the other causes present, and how the outcome would have been different if selection had been present and one of the other causes absent (and do this for each of the other causes). Important causes are big difference makers while causes of modest importance make only a small difference in the outcome. Applying this format for separating more important from less to Darwin’s theory (and to the evolu-

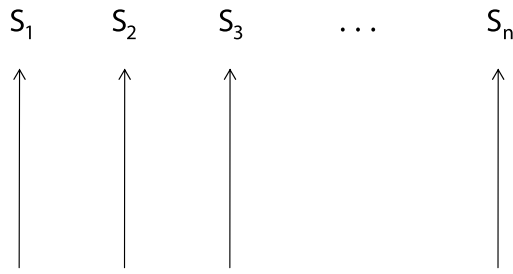


Fig. 2. A set of genealogically unrelated lineages, each evolving by natural selection. This is not Darwin's theory.

tionary theory of the present) is an interesting exercise, but it cannot be pursued here.

Given his statement that selection is the main cause of evolution, how are we to interpret the following comment in the *Origin* concerning the importance of ancestral influence: "the chief part of the organization of every being is simply due to inheritance; and consequently, although each being assuredly is well fitted for its place in nature, many structures now have no direct relation to the habits of life of each species (ref. 1, p. 199)?" If ancestral influence has played so large a role, how can natural selection have been the main cause? Perhaps Darwin should simply have said that selection has been very important. Was it needlessly audacious for Darwin to put selection at the top of a list whose members he had no reason to think he could completely foresee?

### Common Ancestry

With these caveats about natural selection duly noted, is "evolution by natural selection" a good characterization of Darwin's theory? The answer is emphatically no, as can be seen by considering Fig. 2. Darwin's theory gives the concept of common ancestry a central place. The phrase "evolution by natural selection" does not capture this idea, nor does "descent with modification" (10, 11). Instead of describing Darwin's theory as evolution *by* natural selection, the theory is better described as common ancestry plus natural selection. This is not a trivial correction, since the idea of common ancestry plays a central role in the big picture that Darwin painted, or so I will argue.

How much common ancestry did Darwin embrace? In the last paragraph of the *Origin*, where Darwin waxes poetic in his description of the "grandeur in this view of life" (ref. 1, p. 490), he says that,

in the beginning, "life was breathed into a few forms, or into one." A few pages earlier, he is less cautious:

I believe that animals have descended from at most only four or five progenitors, and plants from an equal or lesser number. Analogy would lead me one step further, namely to the belief that all animals and plants have descended from some one prototype. But analogy may be a deceitful guide. Nevertheless all living things have much in common, in their chemical composition, their germinal vesicles, their cellular structure, and their laws of growth and reproduction. We see this even in so trifling a circumstance as that the same poison often similarly affects plants and animals; or that the poison secreted by the gall-fly produces monstrous growths on the wild rose or oak-tree. Therefore I should infer from analogy that probably all organic beings which have ever lived on this earth have descended from some one primordial form, into which life was first breathed.

Ref. 1, p. 484.

Both these passages may suggest that Darwin's view was that there was one start-up of life from nonliving materials, or just a few of them. However, this is not what his theory really says. In the fifth edition of the *Origin*, Darwin adds the following remark:

No doubt it is possible, as Mr. G. H. Lewes has urged, that at the first commencement of life many different forms were evolved; but if so, we may conclude that only a few have left modified descendants.

ref. 4, p. 753.

Darwin was not changing his mind here, but was merely clarifying what he intended all along. The idea was already in the first edition of the *Origin*—not in words, but in a picture. In fact, it was in the book's only picture, shown in Fig. 3. Darwin's view about common ancestry concerns tracing-back, not the number of start-ups. Perhaps life started up one time or many; this may be unknowable and, in any event, was not something that Darwin thought he knew. Darwin's claim is that all of the life that exists now, and all of the fossils that are around now too, trace back to one or a few original progenitors.

Tracing back to a single common ancestor does not entail that there was exactly one start-up. Nor does it entail that all but one of the start-ups failed to have descendants that exist now. If a genealogy is strictly tree-like (with branches splitting but never joining), all but one start-up must go extinct if all current life is to

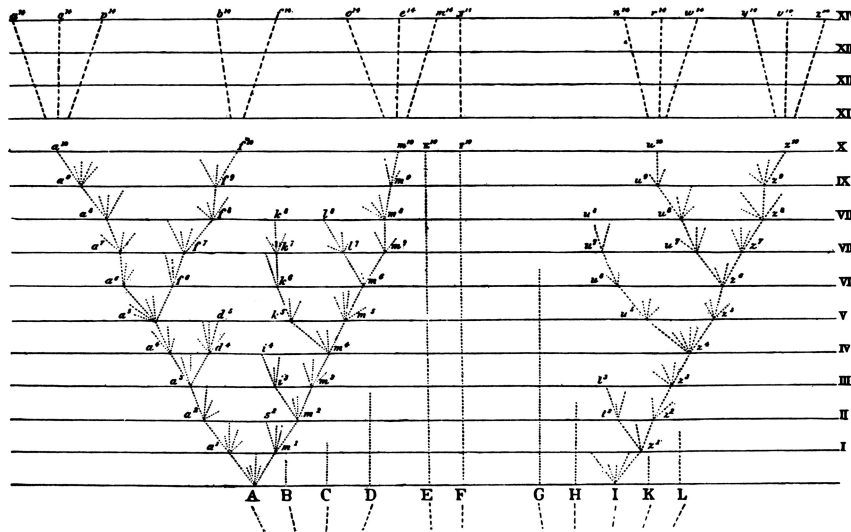
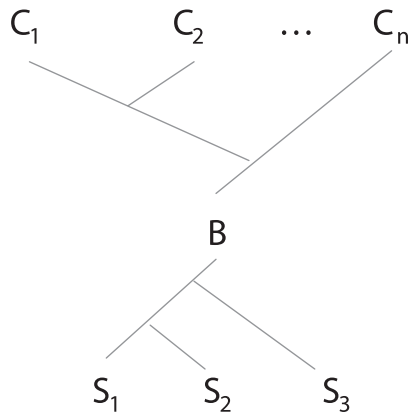


Fig. 3. The only diagram in the *Origin*.





**Fig. 4.** Life with multiple start-ups and a bottleneck. In this example, all current life ( $C_1, C_2, \dots, C_n$ ) traces back to a single common ancestor, but this does not require that 2 of the 3 start-ups ( $S_1, S_2, S_3$ ) fail to have descendants now. Reticulation leading to a bottleneck (B) is the reason why.

trace back to a single common ancestor. However, if there are reticulations (with branches joining and splitting), this need not be so (12). This point is illustrated in Fig. 4.

One of the main objections to Darwin's theory, both when the *Origin* was published and in the minds of many present-day Creationists, is the idea that species (or "fundamental kinds" of organism) are separated from each other by walls. No one doubted, then or now, that natural selection can cause small changes within existing species. The question was whether the process Darwin described can bring about large changes. Maybe a species can be pushed only so far. Darwin was an extrapolationist, inspired by the geological gradualism of Charles Lyell. Darwin reasoned that if artificial selection has achieved what it did in the brief span of time with which plant and animal breeders have had to work, then natural selection can bring about changes that are far more profound since it has operated over the far larger reaches of time that have been available since life began on an ancient earth. Darwin extrapolated from small to large; many of his critics refused to follow him here. If we focus just on natural selection, it is hard to see why Darwin had the more compelling case. However, if we set natural selection aside and consider instead the idea of common ancestry, the picture changes. Darwin thought he had strong evidence for common ancestry. This is enough to show that insuperable species boundaries (and insuperable boundaries between "kinds") are a myth; if different species have a common ancestor, the lineages involved faced no such walls in their evolution. And the case for common ancestry does not depend on natural selection at all.

### Darwin's Principle

Darwin tells us in the *Origin* that when it comes to finding evidence for common ancestry, the adaptive features that provide evidence for natural selection are precisely where one ought not to look:

[A]daptive characters, although of the utmost importance to the welfare of the being, are almost valueless to the systematist. For animals belonging to two most distinct lines of descent, may readily become adapted to similar conditions, and thus assume a close external resemblance; but such resemblances will not reveal – will rather tend to conceal their blood-relationship to their proper lines of descent.

Ref. 1, p. 427.

Two of the facts mentioned earlier – that humans and monkeys have tailbones, and that human fetuses and fish have gill slits—are evidence for common ancestry precisely because tailbones and gill slits are useless in humans. Contrast this with the

torpedo shape that sharks and dolphins share; this similarity is useful in both groups. One might expect natural selection to cause this trait to evolve in large aquatic predators whether or not they have a common ancestor. This is why the adaptive similarity is almost valueless to the systematist.

Let's distinguish the two parts in this idea and give it a name:

**Darwin's Principle.** Adaptive similarities provide almost no evidence for common ancestry while similarities that are useless or deleterious provide strong evidence for common ancestry.

Darwin's Principle can be justified in terms of something deeper. The Principle is an application of an idea about probabilistic reasoning called

**The Law of Likelihood.** Observation  $O$  favors hypothesis  $H_1$  over hypothesis  $H_2$  precisely when  $\Pr(O | H_1) > \Pr(O | H_2)$ . And the strength of the favoring relation is to be measured by the likelihood ratio  $\Pr(O | H_1) / \Pr(O | H_2)$  (13).

The expression " $\Pr(O | H)$ " means "the probability of  $O$ , given  $H$ ." R.A. Fisher chose to call this quantity the likelihood of the hypothesis. This was an unfortunate choice of terminology, but it has stuck. In ordinary English, "likelihood" and "probability" are synonyms, but the Law of Likelihood concerns the likelihood of  $H$ ,  $\Pr(O | H)$ , not its probability,  $\Pr(H | O)$ . These can have different values. And anti-Bayesians maintain that  $\Pr(H | O)$  often has no objective meaning at all, while  $\Pr(O | H)$  does (14).

Notice that the Law of Likelihood has 2 parts, one qualitative, the other quantitative. To see the intuitive plausibility of the qualitative part of the law, consider an example that has nothing to do with common and separate ancestry. Suppose you draw some balls from an urn of unknown composition. You draw 100 times, with replacement, and find that 81 of the draws are green. What does this evidence tell you about the following 2 hypotheses?

- $H_1$ : Exactly 80% of the balls in the urn are green.
- $H_2$ : Exactly 10% of the balls in the urn are green.

It seems obvious that the evidence favors the first hypothesis over the second, and the Law of Likelihood explains why. The observations would be more surprising if  $H_2$  were true than they'd be if  $H_1$  were true. I will not try to motivate the quantitative part of the Law of Likelihood, except to note that  $\Pr(O | H_1)$  and  $\Pr(O | H_2)$  are both small in this example. The likelihood difference, therefore, is tiny, far smaller than the difference there would be if you had drawn just one ball from the urn and it was green. However, the likelihood ratio for the 100 draws is far larger than the ratio for the one. This is a point in favor of using the ratio measure.

How does the Law of Likelihood bear on Darwin's Principle? Let  $X$  and  $Y$  be 2 species (or organisms) that both have trait  $T$ . This is our observation. We wish to know what this observation says about the common ancestry (CA) and the separate ancestry (SA) hypotheses. Darwin's principle is correct to the extent that

$$\Pr(X \text{ and } Y \text{ have trait } T | \text{CA}) / \Pr(X \text{ and } Y \text{ have trait } T | \text{SA}) \approx 1$$

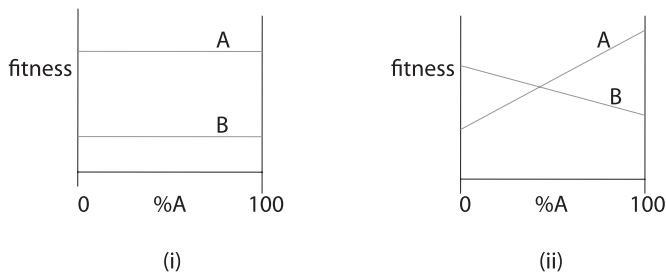
when  $T$  is adaptive for both  $X$  and  $Y$ .

$$\Pr(X \text{ and } Y \text{ have trait } T | \text{CA}) / \Pr(X \text{ and } Y \text{ have trait } T | \text{SA}) \gg 1$$

when  $T$  is not adaptive for both  $X$  and  $Y$ .

The torpedo shape of sharks and dolphins involves a likelihood ratio that is close to one; the tailbones of humans and monkeys and the gill slits of human fetuses and fish involve likelihood ratios that are much larger than unity.

Darwin's Principle applies outside of biology, both in other sciences and in everyday life. For example, suppose 2 students in a philosophy class submit essays on an assigned topic that are word-for-word identical (15). The common cause hypothesis says



**Fig. 5.** Two fitness functions for the traits A and B. When 2 populations each exhibit 100% A, this is not strong evidence that they have a common ancestor if the fitness function is the one shown in (i); the evidence for common ancestry is stronger if the fitness function is the one shown in (ii).

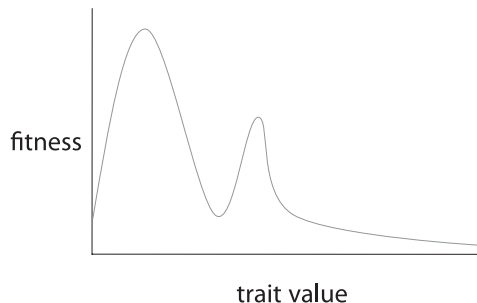
that the students plagiarized from the same source (a file they found on the Internet, perhaps). The separate cause hypothesis says that the students worked separately and independently. The matching is more probable under the first hypothesis than it is under the second. And the kinds of matching features that provide strong evidence for a common cause and the kinds that provide only weak evidence or none at all are the ones that Darwin's principle describes. That both essays use nouns is not worth much. In contrast, that both misspell the same words in the same way is more telling. And what should we make of both essays quoting the same passage from Darwin? It matters if the passage is relevant to the assigned topic.

If Darwin's Principle is to be understood in terms of the Law of Likelihood, there is an important part of his theory that fails to conform to the dictates of hypothetico-deductivism, which some see as Darwin's key methodological innovation (16). This methodology says that theories are tested by deducing observational predictions from them. However, if hypotheses merely confer non-extreme probabilities on observational outcomes, the relationship of hypothesis to observation is not deductive. It is not true that human beings and monkeys must both have tail bones if they share a common ancestor and it is not true that they can not both have tail bones if they do not share a common ancestor. What is true is that the probability of this similarity is greater under the common ancestry hypothesis.

### Exceptions to Darwin's Principle

Although Darwin's Principle is often correct, the two parts of the principle are each sometimes mistaken. Let's take the second part first, the one about neutral or deleterious characters. If a drift process goes on long enough, the resulting character states of the descendants  $X$  and  $Y$  will have about the same probability, regardless of whether the common ancestry or the separate ancestry hypothesis is true. In a drift process as well as in others, time is a destroyer of information about ancestry. A second counterexample to the second half of Darwin's Principle may be found in characteristics that confer no advantage or disadvantage but are correlated with ones that do. These are the features that now are called "spandrels" (17). As mentioned earlier, having red blood confers no advantage, but having hemoglobin does, and the redness is a consequence of the hemoglobin. If hemoglobin is widespread because of its adaptive advantage, 2 species having red blood will not provide strong evidence for common ancestry.

The other side of Darwin's Principle has exceptions as well; there are adaptive similarities that sometimes provide substantial evidence of common ancestry. There are 2 cases in which this is true. The first simply involves lots of data. Suppose we know of  $n$  adaptive similarities that unite species  $X$  and  $Y$ . Each of them may provide only negligible evidence favoring common ancestry over separate ancestry. However, put them together and the likelihood ratio may be substantially greater than unity. This will happen if the different features ( $T_1, T_2, \dots, T_n$ ) are independent of each other, conditional on each of the 2 genealogical hypotheses:



**Fig. 6.** The fitnesses of different trait values of a quantitative character. There are 2 adaptive peaks.

$$\frac{\Pr(T_1 \& T_2 \& \dots \& T_n | CA)}{\Pr(T_1 \& T_2 \& \dots \& T_n | SA)} = \frac{\Pr(T_1 | CA)}{\Pr(T_1 | SA)} \times \frac{\Pr(T_2 | CA)}{\Pr(T_2 | SA)} \times \dots \times \frac{\Pr(T_n | CA)}{\Pr(T_n | SA)}$$

If each term on the right hand side has a value just a bit larger than unity, their product will have a value that is much larger than unity. This point might underlie the thought that complex adaptations can provide substantial evidence of common ancestry even if simple ones do not.\*

The second context in which Darwin is wrong to dismiss adaptive similarities is a bit less obvious. Consider the 2 fitness functions shown in Fig. 5. Each describes how an individual's fitness depends on whether it has trait A or trait B. In Fig. 5i, A is always fitter than B, regardless of the frequency of trait A in the population; in Fig. 5ii A is fitter than B when A is common, but the reverse is true when A is rare. Now suppose you encounter 2 populations that both have trait A at 100%. Is this evidence that the two populations trace back to a common ancestor? Darwin's principle seems right in connection with the fitness function in Fig. 5i; you'd expect A to evolve to fixation, whether or not the two populations share a common ancestor. The inferential situation with respect to Fig. 5ii is different. When there is selection favoring the majority trait, a population will evolve to 100% A or to 100% B depending on what the trait's starting frequency is. In what state do the lineages leading to the two observed populations begin? Suppose that all starting frequencies have the same probability. Then the probability that a lineage starts with A in the majority is 1/2. If the two populations have a common ancestor, the probability of them both exhibiting 100% A is  $\approx 1/2$ . If the separate ancestry hypothesis is true, the probability that both lineages will have 100% A is approximately  $(1/2)(1/2) = 1/4$ . So the common ancestry hypothesis has twice the likelihood as the separate ancestry hypothesis. The likelihood ratio will be bigger if it is very improbable that a lineage will start with A in the majority. If the probability of this is  $p$ , then the likelihood ratio of the two hypotheses is approximately  $p/p^2 = 1/p$ . If  $p$  is small, the evidence favoring common ancestry is very strong (ref. 13, chapter 4)†.

What is true for frequency dependent selection for the majority trait also is true when there is frequency independent selection with multiple peaks of the sort depicted in Fig. 6. A population that starts with a given average trait value will evolve toward a local adaptive peak and then selection will serve to keep the population at that equilibrium value. The larger the population is, the harder it is for the population to traverse a valley and evolve from one peak to

\*Perfect independence is not essential here; the weight of evidence grows if the separate traits have some degree of conditional independence.

†Consider the effect of population size on this inference problem. In Fig. 5i, the bigger the population, the more valueless the observation is that the two populations are each 100% A; in Fig. 5ii, the reverse is true.

another. If 2 populations are at the same adaptive peak, this is evidence that they share a common ancestor. The higher the peaks and the wider the valleys, the more strongly their similarity favors common ancestry over separate ancestry.

These last two cases, in which adaptive similarities provide strong evidence for common ancestry, are not just abstract possibilities. They underlie the reasoning that leads biologists to cite the near-universality of the genetic code as evidence that all current life traces back to a single common ancestor (18). This is an important part of the reason that most biologists would now regard Darwin's "one or a few" original progenitors as too cautious. An organism with a given genetic code will usually have its viability drastically decline if its code changes to one that is "nearby" in the space of possible changes. And if the organism is at least partly sexual, its ability to produce viable fertile offspring will be impaired if its code changes to one not shared by conspecifics. So there is both a frequency-independent and a frequency-dependent effect. As long as there are multiple codes that each would work, a shared code is evidence for common ancestry. And the more such codes there are, the stronger the evidence that the near-universality of the code provides for common ancestry. This point holds even if the shared code we observe in the life around us turns out to be optimal.

Although Darwin's Principle is overstated, a rational kernel can be extracted. Nonadaptive characters often provide strong evidence for common ancestry. And adaptive characters often provide little or no evidence for common ancestry.

### How Common Ancestry and Natural Selection Are Related in Darwin's theory

Darwin (ref. 1, p. 459) says that the *Origin* is "one long argument," and scholars have puzzled over what his argument is. Thinking about this requires that a question about logic be separated from a question about rhetoric. There is the logical structure of his theory and its relation to the evidence he musters. However, there is also the question of how Darwin chooses to present that body of theory and evidence. Why did Darwin organize the book as he did? He front-loads his discussion of natural selection and lets his full argument for common ancestry emerge only later, and in a somewhat fragmented form. Inspired by John Herschel's ideas on *vera causae* (19), he starts with artificial selection; this is a context in which selection has been observed. From this he extrapolates to natural selection, where selection must usually be inferred, and argues that selection is competent to produce the traits we now observe in nature and that it has actually done so (20–22). Darwin could have begun with common ancestry and still pursued this Herschelian strategy. The exposition would start with observed cases of common ancestry (in human family trees and in the ones recorded by plant and animal breeders), with conjectured instances of common ancestry developed subsequently, the argument culminating with his conclusion that all life traces back to one or a few original progenitors. Darwin does do some of this in the book's beginning. In the Introduction he says that species belonging to the same genus have a common ancestor. And in the first chapter, on artificial selection, he argues that all varieties of domesticated pigeons descended from the rock dove. Still, the big picture, wherein all current life traces back to one or a few start-ups, is mostly developed at the end of the book. On the whole, it is natural selection that comes first.

Four years after the *Origin*'s publication, Darwin wrote to Asa Gray about his priorities; he says that "personally, of course, I care much about Natural Selection, but that seems to me utterly unimportant, compared with the question of Creation or Modification" (23). Why, then, did Darwin give selection top billing in the *Origin*? Perhaps he thought that this was his theory's more novel element. Or perhaps he chose this ordering to recapitulate his own intellectual odyssey in which selection came into focus before common ancestry (24, 25). Or maybe he realized that if he began with the grand idea of common ancestry, readers would immedi-

ately contemplate the genealogical connection of human beings to monkeys, a subject that he very much wanted to avoid.

There are other explanations to consider that are more rooted in the details of what Darwin says in the *Origin*. Perhaps he placed natural selection at center stage because he thought that selection is more important than common ancestry. This seems to be the point he is making in the following passage:

It is generally acknowledged that all organic beings have been formed on 2 great laws: Unity of Type, and the Conditions of Existence. By unity of type is meant that fundamental agreement in structure which we see in organic beings of the same class, and which is quite independent of their habits of life. On my theory, unity of type is explained by unity of descent. The expression of conditions of existence, so often insisted on by the illustrious Cuvier, is fully embraced by the principle of natural selection. For natural selection acts by either now adapting the varying parts of each being to its organic and inorganic conditions of life; or by having adapted them during past periods of time: the adaptations being aided in many cases by the increased use or disuse of parts, being affected by the direct action of the external conditions of life, and subjected in all cases to the several laws of growth and variation. Hence, in fact, the law of the Conditions of Existence is the higher law; as it includes, through the inheritance of former variations and adaptations, that of Unity of Type.

Ref. 1, p. 206.

We can understand this passage by thinking about its application to the example of human and monkey tail bones. Human beings have tail bones because the trait was present in the common ancestor that human beings share with monkeys, not because the trait is adaptive for humans. Darwin adds to this the further thought that the trait occurs in the common ancestor because it was adaptive for that common ancestor.

Darwin's point in this passage is not specifically about the importance of common ancestry; his thought applies equally to evolution in a single lineage. Consider a lineage of polar bears extending from an ancestral population *A* to a current population *C*. Suppose *C*'s fur length is closer to *A*'s than it is to the fur length that would be optimal for *C* to have under current conditions. For example, suppose the trait values (in some unit of length) are  $A = 40$  and  $C = 50$  and that the optimum for *C* is  $O_c = 100$ . Arguably ancestral influence has had a bigger effect on trait evolution than natural selection has had because 50 is closer to 40 than it is to 100. Darwin is saying that even when ancestral influence has been very strong (as in the hypothetical case we are considering), it still is true that the ancestor has its trait value because of natural selection. How he knows this is not so clear. If a descendant can have a trait value that is far from optimal, why can't an ancestor? After all, that ancestor itself had an ancestor and the problem recurs. In any event, to show that selection had a stronger effect than ancestral influence on *C*, it does no good to show that selection had a strong influence on *A*.

A better answer to the question of why Darwin put selection first in the *Origin* is provided by his thought that selection explains branching. This is the point of his Principle of Divergence (26). A single population of generalists will often be driven by selection to evolve into 2 populations of specialists, which become increasingly different as each accumulates adaptations specifically suited to its unique way of life. Here, Darwin was influenced by the increasing specialization and division of labor that he saw in the British economy of his time. The idea that selection leads to branching may have been Darwin's reason for putting selection before common ancestry, but it is important to recall our earlier distinction between tracing back and number of start-ups. That selection leads to branching does not entail that all of current life traces back to one



or a few original progenitors.<sup>‡</sup> Here we may add Darwin's idea that selection also leads lineages to go extinct. Selection causes branching *and* extinction, which means that selection does explain why the life around us traces back to one or a few original progenitors. Selection is the source of what the single figure in the *Origin* depicts. So there is a logical reason why selection should come first—it has causal priority.

Darwin faced a choice. Selection has causal priority; common ancestry has evidential priority. What should the order of exposition be? For some authors, the problem does not arise. Consider, for example, the relation of axioms to theorems in Euclidean geometry. If the axioms make the theorems true and the axioms are intuitively obvious while the theorems become obvious only when we see how they are related to the axioms, then the axioms have a “causal” and an evidential priority. But when the causal and the evidential orderings differ, which should be followed? There is no right or wrong here. Darwin led with the part of his theory that has causal priority, but he could have done otherwise. There are many good ways to write a book.

This duality of causal and evidential orderings is hardly unique to Darwin's theory. Consider the relation of temperature and thermometer readings. We know what the temperature is by looking at the thermometer, but it is the temperature that causes the thermometer reading, not vice versa. We often know about causes by looking at their effects. Even so, there is a special feature of the relationship between common ancestry and natural selection in Darwin's theory. Natural selection and common ancestry fit together, but only if selection has not been all-powerful. If all traits evolve because there is selection for them, Darwin's Principle will conclude that we have little or no evidence for common ancestry. What is needed is that selection causes branching and extinction but that some traits persist in lineages for nonadaptive reasons. Darwin's claim that selection is not the exclusive cause of evolution plays an essential role in allowing him to develop his evidence for common ancestry. His conjunction—common ancestry and natural selection—would be unknowable, according to Darwin's Principle, if the second conjunct described the only cause of trait evolution.

In broad outline, the evidential structure of Darwin's argument for his theory of common ancestry plus natural selection goes like this:

1. The argument for common ancestry. Here neutral and deleterious traits (vestigial organs, embryology, biogeography) do the main work.
2. It follows from (1) that populations have evolved across species boundaries.
3. The argument that natural selection is an important part of the explanation of many adaptive traits. Here artificial selection and the Malthusian argument for the power of selection are important, as are Darwin's many examples of adaptive traits in nature.

The expository order in the *Origin* has (3) first, and then (1), with (2) more or less implied.

It is not just that common ancestry answers the question addressed in (2). In addition, common ancestry can be used to answer questions about natural selection. Consider what Darwin says about the vertebrate eye, which was William Paley's most famous example of a complex adaptive feature that, he thought, cries out for explanation in terms of intelligent design (27). Darwin discusses this in chapter 6 of the *Origin*, which he called “Difficulties for the Theory.” He begins by noting that if the different eye designs found in nature can be arrayed in a graded sequence, from simpler and cruder to more complex and more adaptive, that this will be the

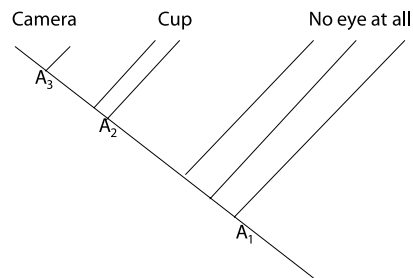


Fig. 7. A simplified example in which the camera eye, the cup eye, and the complete absence of an eye are distributed across the tips of a phylogenetic tree. Parsimony considerations dictate that the best reconstruction of ancestral character states is that  $A_1$  had no eyes,  $A_2$  had a cup eye, and  $A_3$  had a camera eye.

beginning of an argument that the trait evolved by natural selection. But then he adds:

In looking for the gradations by which an organ in any species has been perfected, we ought to look exclusively to its lineal ancestors; but this is scarcely ever possible, and we are forced in each case to look to species of the same group, that is to the collateral descendants from the same original parent-form.

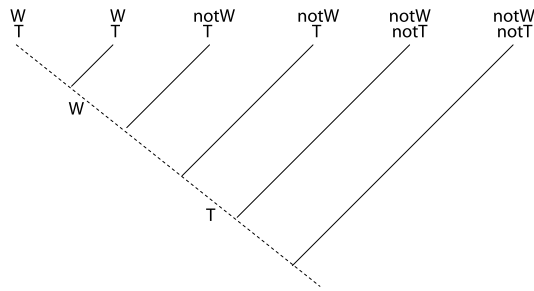
Ref. 1, p. 187.

If it is the lineal ancestors of present day vertebrates that matter to understanding how natural selection has produced the vertebrate eye, why look at current organisms that are not vertebrates? If Darwin's modest goal were to argue merely that it is possible that the vertebrate eye evolved through a series of simpler eye designs, then seeing the different eye designs found in collateral descendants would be relevant because this would allow one to imagine a sequence of steps that might have been taken in the lineal ancestors. I think Darwin wanted to draw a stronger conclusion and contemporary biologists certainly do. They want to argue that the designs found in collateral descendants provide evidence about the designs found in lineal ancestors. But why should the one be relevant to the other? Darwin's language of lineages reveals the reason why. It is common ancestry that makes the characteristics of nonvertebrates that are alive now relevant to inferring the characteristics of the lineal ancestors of present day vertebrates.

A simplified example of the kind of inference problem that Darwin faced is depicted in Fig. 7. As we move from current vertebrates and their camera eyes back through their lineal ancestors, Darwin thought we'd find cup eyes and then no eyes at all. Why think that this is the most plausible assignment of character states to ancestors? Many contemporary biologists would answer by appealing to parsimony. A different assignment of character states to ancestors would be less parsimonious in the sense that it would require more changes in character state in the tree's interior. The thought that this is the right way to make inferences about the historical process does not require an a priori commitment to evolution's always moving from simple to complex. Rather, Darwin's idea about the history of eye designs is a plausible reconstruction, given an independently justified phylogeny, if parsimony makes for plausibility.

The use of parsimony to reconstruct ancestral character states is intuitively attractive. If 2 or more descendants have a given trait, it seems natural to infer that the trait was present in their most recent common ancestor. But what is the logical justification of this inference from present to past? Cladists influenced by Will Hennig (28) have sought to justify the principle in terms of Popperian ideas about falsifiability (29, 30) and explanatory power (31). At about the same time that Hennig's work was translated into English, Anthony Edwards and Luigi Cavalli-Sforza (32), students of R. A. Fisher, proposed a principle of minimum evolution as a heuristic

<sup>‡</sup>Although tree diagrams like Fig. 3 often are used to portray both divergence and genealogy, the two concepts are importantly different. Organisms can be identical and still trace back to a common ancestor; in this case, there is no divergence, since the variance in each generation is always zero. And lineages can diverge from each other even if they have no common ancestry.



**Fig. 8.** The use of parsimony to reconstruct the character states of ancestors in the lineage leading to modern land vertebrates (which is represented by the dashed line). Given the tree shown and the character states ( $W$  = walking and  $T$  = tetrapody) at the tips, the inference is that tetrapody evolved before walking.

device for inferring phylogenies, one that they thought was justified to the extent that it coincides with the dictates of likelihood. The justification of parsimony, and the question of whether there are better methods of phylogenetic inference, is a subject of continuing investigation (ref. 14 chapters 3 and 4 and ref. 33).

The Darwinian reconstruction of the history of eye evolution uses the fact of common ancestry to infer the states of lineal ancestors from the states of collateral descendants. Parsimony considerations, applied to an independently attested phylogeny, also play an important role in testing hypotheses about natural selection. Consider, for example, the hypothesis that land vertebrates evolved 4 limbs to help them walk on dry land. Biologists reject this hypothesis because they think that the morphological trait was present in the lineage before vertebrates came out of the water. Why do they think this? Again, the traits of collateral descendants allow one to infer the traits of lineal ancestors. Fig. 8 provides a simple example of this “chronological test” of adaptive hypotheses. We infer from current organisms (and from fossils) that there were ancestors of land vertebrates that had 4 limbs before vertebrates came up on dry land. Tetrapody evolved before walking in the vertebrate line.

As far as I know, Darwin does not explicitly describe the use of parsimony to infer ancestral character states; however, he does deploy this pattern of reasoning in applying his theory to examples. Consider his comment in the *Origin* concerning why mammals in utero have skull sutures that allow them to pass through the birth canal:

The sutures in the skulls of young mammals have been advanced as a beautiful adaptation for aiding parturition, and no doubt they facilitate, or may be indispensable for this act; but as sutures occur in the skulls of young birds and reptiles, which have only to escape from a broken egg, we may infer that this structure has arisen from the laws of growth, and has been taken advantage of in the parturition of the higher animals.

Ref. 1, p. 197.

The sutures predate mammalian parturition; we know this because the sutures, but not the parturition, are found in contemporary birds and reptiles. Darwin is explicit, here and elsewhere, that what makes a trait currently useful may differ from what made the trait initially evolve. It is common ancestry that permits him to say something more—that the reason a trait initially evolved actually differs from the reason the trait is now useful.

For Darwinians, a lineage is like a mineshaft that extends from the surface to deep in the earth, with multiple portholes connecting surface to shaft at varying depths. By peering into these portholes, we obtain fallible guidance about what is happening in the shaft; the more portholes there are, the more evidence we can obtain. Common ancestry is not an unrelated add-on that supplements Darwin’s hypothesis of natural selection; rather, common ancestry provides a framework within which hypotheses about natural selection can be tested. It is because of common ancestry that facts about the history of natural selection become knowable.

### Tree-Thinking

Tree-thinking is central to reasoning about natural selection, both for Darwin and for modern biology (34, 35). The reverse dependence is not part of the Darwinian framework, as we learn from Darwin’s Principle. You do not need to assume that natural selection has been at work to argue for common ancestry; in fact, what Darwin thinks you need to defend hypotheses of common ancestry are traits whose presence cannot be attributed to natural selection. This is the evidential asymmetry that separates common ancestry from natural selection in his theory. So, did Darwin write the *Origin* backwards? The book is in the right causal order; but evidentially, it is backwards.

**ACKNOWLEDGMENTS.** I thank Jason Alexander, Francisco Ayala, David Baum, Luc Bovens, James Crow, Daniel Dennett, Marc Ereshefsky, Joshua Filler, Jonathan Hodge, Ronald Numbers, Robert Richards, Michael Ruse, Silvan Schweber, and Mark Taylor for useful discussion.

- Darwin C (1859) *On the Origin of Species by Means of Natural Selection* (London, Murray).
- Sloan P (2009) In *The Cambridge Companion to the ‘Origin of Species’*, eds Ruse M, Richards R (Cambridge, Cambridge Univ Press); pp 67–86.
- Richards R (2009) Darwin’s place in the history of thought: A reevaluation. *Proc Natl Acad Sci USA* 106:10056–10060.
- Darwin C (1959) *On the Origin of Species—a Variorum Edition*, ed Peckham M (Univ of Pennsylvania Press, Philadelphia), p 164.
- Sober E (1994) In *Creative Evolution?!*, ed Campbell J (Jones and Bartlett, Boston), pp 19–33.
- Darwin C (1868) *The Variation of Animals and Plants under Domestication* (D. Appleton, New York), 2nd Ed.
- Darwin C (1871) *The Descent of Man and Selection in Relation to Sex* (London, Murray).
- Ruse M (1980) Charles Darwin and group selection. *Ann Sci* 37:615–630.
- Sober E, Wilson, D (1998) *Unto Others—the Evolution and Psychology of Unselfish Behavior* (Harvard Univ Press, Cambridge, MA).
- Mayr E (1985) *The Darwinian Heritage*, ed Kohn D (Princeton Univ Press, Princeton), pp 755–772.
- Sober E, Orzack S (2003) Common ancestry and natural selection *Bri J Phil Sci* 54:423–437.
- Sober E, Steel M (2002) Testing the hypothesis of common ancestry. *J Theor Biol* 218:395–408.
- Hacking I (1965) *The Logic of Statistical Inference* (Cambridge Univ Press, Cambridge, UK).
- Sober E (2008) *Evidence and Evolution—the Logic Behind the Science* (Cambridge Univ Press, Cambridge, UK), Chap 1.
- Salmon W (1984): *Scientific Explanation and the Causal Structure of the World* (Princeton Univ Press, Princeton).
- Ghiselin M (1969) *The Triumph of the Darwinian Method* (Univ of California Press, Berkeley).
- Gould S, Lewontin R (1979) The spandrels of San Marco and the panglossian paradigm—A critique of the adaptationist paradigm. *Proc R Soc London Ser B* 205:581–598.
- Knight R, Freeland S, Landweber L (2001) Rewiring the keyboard—evolvability of the genetic code. *Nat Rev Genet* 2:49–58.
- Herschel J (1830) *A Preliminary Discourse on the Study of Natural Philosophy* (Longman, London).
- Hodge M-J-S (1977) The structure and strategy of Darwin’s “long argument.” *Br J Phil Sci* 10: 237–246.
- Ruse M (1979) *The Darwinian Revolution* (Univ of Chicago Press, Chicago).
- Waters, C-K (2003) The arguments in the *Origin of Species*” eds Hodge J, Radick G, *The Cambridge Companion to Darwin* (Cambridge, Cambridge Univ Press), pp 116–142.
- Darwin C (1887) *The Life and Letters of Charles Darwin*, ed Darwin F (Murray, London), Vol 2, pp 163–164.
- Ospovat D (1981) *The Development of Darwin’s Theory—Natural History, Natural Theology, and Natural Selection 1838–1859* (Cambridge Univ Press, Cambridge, UK)
- Schweber S (1988) In *The Darwinian Heritage*, ed Kohn D (Princeton Univ Press, Princeton), pp 35–70.
- Kohn D (2009) *Darwin’s keystone—the principle of divergence*, eds Ruse M, Richards R, *The Cambridge Companion to the ‘Origin of Species,’* (Cambridge, Cambridge Univ Press), pp 87–108.
- Paley W (1802) *Natural Theology, or, Evidences of the Existence and Attributes of the Deity, Collected from the Appearances of Nature* (Rivington, London).
- Hennig W (1966) *Phylogenetic Systematics* (Univ of Illinois Press, Urbana).
- Eldredge N, Cracraft J (1980) *Phylogenetic Patterns and the Evolutionary Process* (Columbia Univ Press, New York)
- Wiley E (1981) *Phylogenetics—the Theory and Practice of Phylogenetic Systematics* (Wiley, New York).
- Farris J-S (1983) In *Advances in Cladistics—Proceedings of the 2nd Annual Meeting of the Willi Hennig Society*, eds Platnick N, Funk V (Columbia Univ Press, New York), pp 7–36.
- Edwards A-W-F, Cavalli-Sforza L (1964) In *Phenetic and Phylogenetic Classification*, eds Heywood V, McNeill J (New York Systematics Association, New York), Vol 6, pp 67–76.
- Sober E (1988) *Reconstructing the Past—Parsimony, Evolution, and Inference* (MIT Press, Cambridge, MA).
- O’Hara R (1998) Population thinking and tree thinking in systematics. *Zool Scripta* 26: 323–329
- Baum D, Offner S (2008) Phylogenies and tree thinking. *Am Biol Teacher* 70:222–229.
- Ereshefsky E (2009) Darwin’s solution to the species problem. *Synthese*, in press.