Precis of "Did Darwin write the *Origin* backwards?" Philosophical essays on Darwin's theory

Prometheus Books, Amherst, NY, 2011, 230 pp, ISBN 978-1-61614-230-8

Elliott Sober

This book consists of four essays on Darwin's theory of evolution and three postscripts on contemporary evolutionary theory. The main questions that the chapters address are: (1) How are common ancestry and natural selection related to each other in Darwin's theory? (2) What were Darwin's views about group selection and the evolution of altruism? (3) Why did Darwin change his mind about sex ratio, and how are his ideas related to the creationism that preceded him and the evolutionary biology that followed? (4) Do Darwin's theory and modern evolutionary theory obey the principle of methodological naturalism and ought they to do so?

The book is philosophical because it focuses on the logic of arguments. I attend closely to Darwin's texts, but I sometimes use tools that Darwin did not have available. These include ideas from contemporary biology and from probability theory. I see merit in this approach because it helps elucidate Darwin's theory. Darwin pointed to the need for this sort of analysis in the sixth and final edition of the *Origin* when he said the following about his theory:

> It can hardly be supposed that a false theory would explain, in so satisfactory a manner as does the theory of natural selection the several large classes of facts above specified. It has recently been objected that this is an unsafe method of arguing. But it is a method used in judging of the common events of life, and has often been used by the greatest natural philosophers. The undulatory theory of light has thus been arrived at; and the belief in the revolution of the earth on its own axis was until lately supported by hardly any direct evidence (Darwin 1959, p. 748).

E. Sober (✉)
Philosophy Department, University of Wisconsin, Madison, 5185 Helen C. White Hall, Madison, WI 53706, USA
e-mail: ersober@wisc.edu

What catches my eye in this passage is Darwin's claim that his "methods of arguing" are also used in everyday life and in physics. To understand Darwin's theory, one needs to understand patterns of reasoning that are not subject-matter specific.

1 Chapter 1

It is obvious that common ancestry and natural selection are logically independent. I focus on two other relations: How is evidence for each of these related to facts about the other? And is there a causal relation between the two? I am interested in how Darwin answered these questions and in what the right answers are. For Darwin, a trait's providing evidence for common ancestry does not depend on natural selection's having caused the trait to evolve. In fact, it would be better if natural selection wasn't the cause at all. This is what Darwin is saying in the following passage:

> … adaptive 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 (Darwin 1859, p. 427).

Darwin's idea is that adaptive similarities provide little or no evidence for common ancestry, whereas neutral and deleterious similarities provide evidence that is more substantial. I call this Darwin's Principle and analyze it by using an idea that probabilists now call the Law of Likelihood. I note some exceptions to Darwin's Principle, but on the whole Darwin was on to something important.

What about the opposite evidential relation—does evidence for natural selection depend on the fact of common ancestry? Darwin never addressed this question explicitly, but his practice was to use the fact of common ancestry to assess hypotheses concerning how natural selection caused a trait found in a present population to evolve. To carry out this assessment, you need to know something about the state of the lineage way back when. It is the fact of common ancestry that allows you to make inferences about that ancestral condition. I describe several cases in which Darwin uses this method of inference; I argue that there is an implicit appeal to parsimony in his reasoning. Given Darwin's Principle and his dependence on common ancestry to reason about natural selection, I conclude that Darwin thought that common ancestry is "evidentially prior" to natural selection.

Another main topic of Chapter 1 is the causal relation of common ancestry and natural selection. Here it is important to recognize that Darwin's idea of common ancestry is not that there was a single start-up of life on earth. Rather, his idea is that all organisms now on earth trace back to one or a few original progenitors; even if there were thousands of start-ups, only a few have managed to leave descendants that exist today. This thesis is logically stronger than the claim that *some* species have common ancestors, though it is weaker than the modern biologist's commitment to *universal* common ancestor, so I'll call it the thesis of *strong* common ancestry. Darwin's

explanation for strong common ancestry is natural selection. Selection causes extinction, with the result that the tracing-back thesis is true.

Thus, Darwin thought that common ancestry is evidentially prior to natural selection, and that natural selection is causally prior to strong common ancestry. This raises the question that is the book's title. Why did Darwin "front load" natural selection in the *Origin*? I don't mean that he never mentions common ancestry at the start nor natural selection at the end. Rather, I mean that the big picture about natural selection is defended mostly before the big picture of strong common ancestry is presented. I speculate a little about Darwin's motives for using this ordering; however, the main point is not why Darwin put selection first in the *Origin*, but how the parts of his theory fit together.

2 Chapter 2

The hypothesis of group selection came under ferocious attack in the 1960s, but since then the hypothesis has had a comeback. Chapter 2 begins with the 1960s, but the main focus is on Darwin's own ideas. I challenge the popular interpretation of Darwin as a thorough-going individual selectionist. Here again, my focus is on the logic of his thinking. Of particular interest is the puzzle that Darwin found in the existence of sterile workers in the social insects and his disagreement with Wallace about the evolution of hybrid sterility. However, his explanations of human morality and of the honeybee's barbed stinger are interesting as well. He even thought about group selection acting on non-altruistic traits.

3 Chapter 3

Probabilists in the 18th century noticed that the sex ratio in human populations is male-biased at birth and gets more even by the age of marriage. The conclusion was then drawn that sex ratio data underwrite an argument for the existence of God. I place this argument in a probability framework, again using the Law of Likelihood. Darwin read Paley's summary of this design argument while a student at Cambridge. Years later, he transformed it into the selectionist explanation of sex ratio that he proposed in the first edition of *The Descent of Man.* Darwin retracted that explanation in the 2nd edition but never explained why. I speculate about why he changed his mind. I then describe how subsequent evolutionary biologists (Düsing, Fisher, and Hamilton) made progress on the problem. This story furnishes a good example of how evolutionary biology is able to develop testable theories in a way that has eluded creationists.

4 Chapter 4

I begin this chapter by considering Darwin's views about God and religion. I then turn to methodological naturalism. How should that *ism* be understood? Does

Darwin violate methodological naturalism when he discusses God in the *Origin*? I then consider the fact that modern evolutionary biology is mathematical. If numbers are what mathematical Platonists say they are, does this mean that the modern theory violates naturalism? Finally, I take up the question of why science should follow the dictates of methodological naturalism; I criticize some of the answers that others have proposed before suggesting a modest answer of my own.

5 Three postscripts

The first discusses a likelihood justification for using a principle of parsimony in evolutionary biology to infer the chronological order in which two characters evolved in a lineage. The second postscript addresses some modern issues about group and individual selection, including the suggestion that the units of selection problem is to be solved by adopting a convention, not by ascertaining a matter of fact. I embrace pluralism about explanations while rejecting conventionalism about units of selection. I also suggest that conventionalists and their opponents have talked past each other by using different definitions of individual and genic selection. The third postscript argues that the probabilities used in modern evolutionary biology have an objective interpretation and that this follows from the thesis that "micro-probabilities" (e.g., those used in quantum mechanics) are objective.

References

Darwin, C. (1859). *The origin of species*. London: Murray.
Darwin, C. (1959). In M. Peckham (Ed.), *On the origin of species—a variorum edition*. Philadelphia: University of Pennsylvania Press.

Replies to commentators on *Did Darwin Write the Origin Backwards?*

Elliott Sober

Abstract Here I reply to Jean Gayon's, Tim Lewens's, and Samir Okasha's comments on Did Darwin write the Origin backwards? The topics addressed include: (1) Darwin's thinking that common ancestry is "evidentially prior" to natural selection; (2) how Darwin uses phylogenetic trees to test hypotheses concerning natural selection; (3) how group and indivdiual selection should be defined, and how each is related to the concept of adaptation.

I am grateful to Jean Gayon, Tim Lewens, and Samir Okasha for their stimulating reflections on *Did Darwin Write the Origin Backwards?* (hereafter DDWOB). Below are my replies. I am sorry that space limitations prevent me from addressing more of the interesting ideas that the commentators have developed.

1 Reply to Jean Gayon

I am pleased by Gayon's openness to the styles of writing history that I use in DDWOB. As he mentions, historians of science are often much less tolerant. Internalist history of science is now out of fashion. I certainly do not think that the internal logic of Darwin's theory and how Darwin himself understood that logic are the only proper topics of historical interest. But there is a place for them. Many historians have walked away from such topics, not just because they recognize external influences on theorizing, but because they aren't interested in the content of theories at all.

Besides describing different approaches in the history of science, Gayon also describes an idea that he thinks has dominated philosophy of science for some years.

E. Sober (✉)
Philosophy Department, University of Wisconsin, Madison, 5185 Helen C. White Hall, Madison, WI 53706, USA
e-mail: ersober@wisc.edu

He calls it "the 'regionalist turn" and says that its "slogan is: nothing of interest in philosophy of science can be done outside the realm of within-discipline work." Gayon is right that regionalism has been influential, but I suspect that he and I may disagree about whether this influence is 100 % good news. For example, I think there has been valuable work in philosophy of science over the past several decades on scientific realism, reductionism, causality, and scientific inference that has tried to generalize across scientific disciplines. Even though philosophy of biology begins with biology, it often should not end there. The quotation from Darwin at the start of my *Precis* offers one reason why philosophy must sometimes look beyond biology in order to understand biology.

At the start of DDWOB's first chapter, I suggest that neither "evolution by natural selection" nor "descent with modification" is a particularly good characterization of Darwin's theory but that "common ancestry plus natural selection" is better (p. 21). Gayon finds this suggestion "questionable" since it was Darwin himself who used the phrase "descent with modification" and Gayon thinks that Darwin had the hypothesis of strong common ancestry (the thesis that all current life traces back to one or a few original progenitors) in mind when he said "descent with modification." My point in suggesting that "common ancestry" is better than "descent" was to draw attention to the fact that descent with modification (as modern readers understand that phrase) can occur when numerous genealogically unrelated start-ups all have descendants that exist now. There is no strong common ancestry here. I don't see why this clarification is questionable if it helps readers understand how Darwin understood a key element in his own theory.

Gayon thinks the distinction I draw between evidential and causal priority reveals "a real difficulty in Darwin's thinking." I don't think there is a difficulty here, at least not in the sense of revealing a flaw. Darwin had to decide how to order his ideas in the *Origin*, but as I say in Chapter 1, there are many good ways to write a book. The decision he faced concerned rhetoric, not the substance of his theory. In Chapter 1, I offer a simple analogy to explain why the difference between evidential and causal priority is no cause for alarm. Thermometer readings are evidentially prior to temperature, but temperature is causally prior to thermometer readings. You learn about the temperature by looking at the thermometer (and not vice versa) but it is the temperature that causes the thermometer reading (and not vice versa). These twin facts do not constitute a difficulty for the usual theory of how thermometers and temperatures are related.

According to Gayon, the thesis that common ancestry is evidentially prior to natural selection relies on pattern cladistics. Although I think that Darwin believed the evidential priority thesis, I don't think that the thesis commits him to pattern cladism. As for myself, I can say that I am not now, nor have I ever been, a pattern cladist. Pattern cladists think that "trees" can be reconstructed from data where the "trees" they have in mind make no claims about genealogical relationships and inferring these "trees" requires no evolutionary assumptions (Sober 1993). When I talk about trees (without scare quotes), I mean representations of genealogies. Trees are historical hypotheses. One of the main points of my 1988 book (Sober 1988) was that genealogies can be reconstructed from data only when one makes substantive assumptions about the evolutionary process.

Gayon sees a link between pattern cladism and the evidential priority thesis because he misunderstands what the thesis asserts. I explicitly warn against this misinterpretation in DDWOB when I say:

> It is important to distinguish the following two claims: (1) If natural selection favored the evolution of trait T in the lineages leading to species X and Y, then the fact that X and Y both have trait T provides little or no evidence as to whether CA [common ancestry] or SA [separate ancestry] is true. (2) Natural selection favored the evolution of one or more traits in the lineages leading to species X and Y. Darwin's Principle is committed to the first of these, but not to the second. It is in this sense that the inference of common ancestry does not depend on the occurrence of natural selection (pp. 42–43).

Gayon quotes with approval Richards's (2012) remark that "inference of common ancestry depended on inferences about the operation of natural selection. If so, there is no evidential priority here for common ancestry." Richards's phrase "the operation of natural selection" ignores the distinction between (1) and (2). Gayon falls into the same error when he says this:

> If I am right, Darwin's idea, according to Sober, is the following: if non-adaptive characters play a prominent role in Darwin's views about the correct way to discover and corroborate patterns of common ancestry, then this method does not rely on the theory of natural selection, whose main objective is to explain adaptations.

It is not reliance on the *theory* of natural selection that matters, but reliance on the *occurrence* of natural selection.

Gayon thinks I analyze the tree diagram in the *Origin* "as being no more than a pattern of common ancestry," that I do "not consider that Darwin's diagram represents more than a pattern of ancestry." This is not correct. True, I present Darwin's diagram early in Chapter 1 (p. 23) in order to explain Darwin's idea of strong common ancestry and I do not mention there how the figure also represents Darwin's principle of divergence, which is a thesis about natural selection. However, I discuss the principle of divergence later (p. 34) and say that the diagram represents it: "the horizontal axis of his branching diagram … represents dissimilarity; [Darwin] drew the branching to indicate that conspecific varieties have a better chance of persisting if they diverge from the mean of the species, and congeneric species are likewise apt to be more successful if they diverge from the mean of the genus." Gayon and I agree that Darwin's diagram simultaneously represents ideas about genealogy and natural selection.

2 Reply to Tim Lewens

Lewens starts his commentary with a good summary of the main thesis of the first chapter of DDWOB. He avoids two mistakes made by other reviewers—thinking that I say that Darwin's masterpiece is in the wrong order and that I claim that Darwin never mentions natural selection at the end of the *Origin* (Hodge 2012; Ruse

2012). He also sees that my thesis about Darwin's facing a choice in how he would sequence material in the *Origin* isn't based simply on the fact that natural selection and common ancestry are logically independent. I claim that causal priority spoke in favor of front-loading natural selection whereas evidential priority spoke in favor of starting with common ancestry. Each arrangement makes sense.

I argue in Chapter 1 that Darwin uses the assumption of common ancestry to evaluate hypotheses concerning how natural selection influenced the evolution of various traits. Lewens takes issue with my interpretation. Is the naked head of the vulture an adaptation for "wallowing in putridity"? Are the cranial sutures found in mammals adaptations for facilitating live birth? Darwin argues that the answer to both questions is *no*. His reason is that "the head of the clean-feeding male turkey is likewise naked" and that sutures occur in the "skulls of young birds and reptiles, which have only to escape from a broken egg" (Darwin 1959, p. 197). Lewens says that "Darwin's reasoning here is rather sparse" and suggests that "all Darwin does in all these examples is point to the lack of correlation between adaptive pressure and anatomical structure. He does not tell us in these passages … that the shared trait is most likely inherited from a common ancestor." Lewens's phrase "lack of correlation" isn't right (if correlation is understood as a probabilistic idea); I think Lewens means that Darwin's point is that we should hesitate to claim that X causes Y when we find cases in which Y is present without X. This, as Lewens says, has nothing to do with common ancestry.

Earlier in his commentary Lewens advocates using a principle of charity. He says that "if you want to understand what Darwin probably did mean, it is helpful to think about what it would have been sensible for him to have meant." Charity may be an impediment to Lewens's interpretation. You don't throw much doubt on the claim that smoking causes cancer by pointing out that people sometimes get cancer without smoking. Did Darwin stumble over this elementary point?
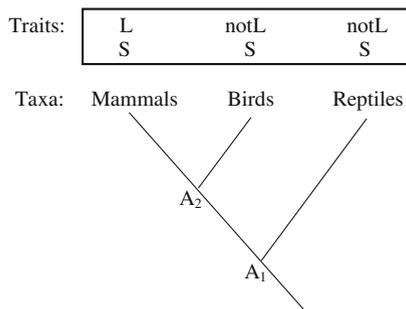
Charity aside, I think that my interpretation is better than the one that Lewens suggests. Given that birds and reptiles have sutures without parturition (they lay eggs and don't do live birth), Darwin says that "we may infer that this structure has arisen from the laws of growth, and has been taken advantage of in the parturition of the higher animals" (Darwin 1959, p. 197). Lewens discussses Darwin's mention of "the laws of growth" by talking about correlations of character, but I would draw attention to another feature of Darwin's sentence. Although the word "then" does not occur there, I think that Darwin means that the sutures evolved at one time and

then were taken advantage of when parturition evolved. In other words, the sutures preceded the parturition.[1] Darwin concludes that parturition didn't cause sutures to evolve, since cause must precede effect. But how does Darwin defend his claim about chronological order? All this makes sense if you interpret Darwin as reasoning about the accompanying phylogenetic tree, which shows the three taxa and their character states; L means live birth and S means skull sutures. There are two common ancestors in the figure, A_1 and A_2. My suggestion is that Darwin uses the characteristics we observe in the three extant taxa to reason that A_1 and A_2 both had sutures and that neither had live birth. Given this, it is false that sutures evolved in the lineage leading to mammals because there was parturition. The sutures were there before live birth evolved.[2]

This reconstruction of Darwin's reasoning is additionally supported by textual evidence that I cited in DDWOB (p. 36). Darwin makes the following suggestion ten pages before he takes up the vulture's naked head and mammalian skull sutures:

> 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 (Darwin 1959, p. 187).

Darwin's using collateral descendants to make inferences about how an organ evolved in a lineage is another instance of tree-thinking.[3]

Lewens next turns to the question of why Darwin organized the *Origin* "cause-first" with the theory of natural selection coming before the defense of strong common ancestry. He suggests that "perhaps the answer lies in the fact that the *Origin*'s argument is an inference to the best explanation," an idea he developed in his illuminating book (Lewens 2007). Lewens's suggestion raises the question of why the use of inference to the best explanation should dictate the *Origin*'s ordering of ideas. After all, strong common ancestry explains the nested similarities that Darwin observed. Why not put that first? Lewens grants that

> It is possible, of course, to lay out an inference to the best explanation by beginning with the effects and then showing how the cause makes sense of them. That said, it would be hard for a reader to appreciate the explanatory relationship between selection and its effects unless one is familiar with what natural selection is supposed to be, and how it is supposed to act, when one comes to review the diverse phenomena Darwin thinks it can account for. That is why … Darwin's book has a cause-first structure.

---

[1] Darwin is explicit about making a claim concerning chronological order in his discussion of bamboo hooks in the same paragraph; I quote this in DDWOB, p. 39.

[2] If this is Darwin's reasoning, he is conforming to an inference principle that biologists would later call cladistic parsimony: the most reasonable estimate of the character states of the ancestors in a phylogenetic tree is the one that minimizes the number of changes that must have occurred in the tree to yield the characteristics observed in the leaves; this is discussed in DDWOB's first postscript.

[3] Darwin's tree-thinking may help explain why he changed his mind about sex ratio; see DDWOB, pp. 105–106.

Lewens's suggestion about Darwin's motivation is reasonable and it does not contradict anything I say in DDWOB.

Lewens connects his suggestion to work by two other scholars. He defends the suggestion by endorsing Hodge's (1977) argument that Darwin was influenced by Herschel's *vera causa* principle. Hodge argues that Darwin organized the *Origin* so that it would demonstrate the existence, competence, and responsibility of natural selection for a wide range of observations. Although I admire Hodge's (1977) work on the structure of what Darwin calls his "one long argument," I think Hodge underestimates the place of common ancestry in Darwin's theory; for Hodge, Darwin's theory is basically the theory of evolution by natural selection. Hodge does not pay much attention to Darwin's argument for strong common ancestry. Strong common ancestry is not an observation that supports the hypothesis of natural selection because it is not an observation at all; it is a hypothesis that Darwin introduces to explain various observations.

Lewens contrasts Hodge's analysis with that of Waters (2003) and notes that Waters argued that Darwin's case for common ancestry in the second half of the *Origin* "remains neutral on the cause of evolutionary change." My interpretation of Darwin's Principle leads me to wince a little at the word "neutral." If the similarity of two species is to be evidence that they have a common ancestor, it matters a lot to Darwin whether the traits evolved because there was natural selection for them in the two linages. In his excellent essay, Waters (2003, p. 123) emphasizes that Darwin's case for common ancestry "does not depend on the premise that natural selection is the underlying cause." That is true, but Darwin says something additional. It isn't just that appeal to natural selection isn't needed. He also says that if natural selection governs the evolution of a similarity, the similarity can't provide substantial evidence for common ancestry. Needing notX goes beyond not needing X. In any event, Lewens thinks that Waters's position is a threat to Hodge's and hence to Lewens's own explanation of why the *Origin* is written selection-first. Hodge sees one long argument, with its separate parts all being defenses of the theory of natural selection. Waters and I think there are two longish arguments— one for natural selection, the other for strong common ancestry.[4] However, I don't think that the two-argument view undermines Lewens's speculation about why Darwin put selection first in the *Origin*. After all, Darwin did think that natural selection explains why strong common ancestry is true.

Hodge (1977), Waters (2003), and Lewens (2007) seek to characterize the arguments in the *Origin* as inferences to the best explanation. I agree, but I think that this tool needs to be developed much more if it is to fully explain the details of Darwin's reasoning. Consider, for example, how inference to the best explanation would apply to Darwin's Principle. Why is common ancestry a much better explanation than separate ancestry of the similarity of two species when the similarity is nonadaptive, but not when the similarity is adaptive? Is this true under some assumptions but not under others? This is where I suggest using the law of

---

[4] I discussed Darwin's Principle, though I didn't use that label, in Sober (1993, pp. 41–42) and in Sober (1999a, p. 265, pp. 273–274).

likelihood as an analytical tool. A similar question arises in connection with Darwin's reasoning about mammalian skull sutures; I discuss a likelihood justification of this argument in DDWOB's first postscript. Whenever the idea of inference to the best explanation is used to clarify a scientific inference, the analysis requires an account of why one proposed explanation is better than another; "better explanation" requires elucidation, as proponents of the theory of inference to the best explanation have emphasized (Lipton 2004). My own view is that explanatoriness is not an epistemologically fundamental category (Roche and Sober 2013).

Darwin pointed to the unifying power of his theory as evidence of its truth. Philosophers of biology have noted this fact but have mostly shied away from the philosophical question of why (or in what circumstances) unification is evidence of truth. To pursue this further question, we need to depart from the regionalism that Gayon describes. The question is philosophical, not historical in a narrow sense, and yet its answer may help us to better understand Darwin's own theory. Part of understanding a theory is seeing how its parts fit together; this is not settled by the fact that the theory's creator *thought* that the parts fit together in a certain way.

Lewens concludes his essay by arguing that "even if, in one sense, selection turned out to be the only cause of trait evolution, this would not undermine our ability to spot common ancestry." I agree. This is why I express reservations about Darwin's Principle in DDWOB (pp. 28–32). When species X and Y each have trait T, the similarity provides meager evidence of common ancestry if a simple sort of selection process is at work in both lineages, but there are more complex types of selection in which the similarity provides substantial evidence. Darwin apparently wasn't thinking of these more complex cases when he formulated his Principle. The passage that Lewens quotes from (Darwin 1959, p. 479) is a good one to think about in this connection. Darwin says that the single "framework of bones" that one finds in several species is evidence that those species have a common ancestor. Yet, it seems that the framework of bones was and is useful. This reasoning can be understood by comparing the following two hypotheses:

(H$_1$) The species had a common ancestor and natural selection caused that ancestor to evolve a single framework of bones. Then there was stabilizing selection for that trait in the lineages descending from that common ancestor

(H$_2$) The species did not have a common ancestor and natural selection caused the separate lineages leading to those species to each evolve some framework of bones or other

H$_1$ makes the shared framework we observe more probable than H$_2$ does, if different arrangements of bones probably would have evolved were H$_2$ true. This likelihood comparison of the two hypotheses resembles what I say in DDWOB (pp. 31–32) about the near-universality of the genetic code.

3 Reply to Samir Okasha

Okasha and I agree that the units of selection problem is a factual problem, not to be settled by adopting a convention. Part of the second postscript in DDWOB is devoted to explaining what is wrong with conventionalism about units of selection;

some of my discussion builds on insightful ideas that Okasha develops in his 2006 book. In that postscript, I describe a simple example concerning the evolution of altruism in order to argue that pluralism about explanation doesn't entail conventionalism about units of selection. I present two explanations that are mutually compatible. The first explains why altruism evolves by describing the opposing forces of group and individual selection; the second explains this outcome by describing the correlation of interactors and the cost/benefit ratio. Both stories are explanatory and causal, even though the second doesn't talk about distinct levels of selection. Contrary to what Okasha says, I do not believe that multi-level selection (MLS) theory, in which group and individual selection are distinguished, is "mandatory" in the sense of furnishing the only causal explanation in town.[5]

Much of Okasha's (2006) book concerns the difference between two modern ways of developing MLS theory. There is the approach that uses the Price equation and the approach that uses the statistical technique known as contextual analysis. Each begins with its own uncontroversial mathematics, decomposing the change in trait frequency due to natural selection into two additive components. The two resulting equations are predictively equivalent in that the change predicted by one will be the same as the change predicted by the other. The Price and contextual "approaches" go beyond this austere mathematics because the Price approach labels the addends in the Price equation "group selection" and "individual selection" while the contextual approach applies these same labels to the two addends in the contextual equation. Now there is conflict; the two approaches sometimes disagree about whether there is group selection and often disagree about the magnitudes of the two selection processes when they agree that both are occurring.[6] But more fundamentally, the two approaches disagree about what group selection *is*. For the contextual approach, "group selection" means that an individual's fitness is affected by the kind of group the individual occupies. For the Price approach, "group selection" means that there is variation in the fitnesses of groups. After weighing the pluses and minuses, Okasha (2006) opts for the contextual approach. I mention these two approaches in a footnote in DDWOB (p. 194) and say that my definitions of group and individual selection are compatible with the Price approach but not with the contextual approach (because the latter says there can be group selection when there is just one group and when groups have identical fitnesses; see Sober 2011). However, I don't discuss the details of either in DDWOB. I avoid that topic because my main goal was to describe how Darwin thought about group and individual selection. I argue that Darwin's thinking is very much grounded in the idea that group selection is selection among groups just as individual selection is selection among individuals in the same group. Notice that this involves a *qualitative* definition of group and individual selection, not a *quantitative* measure of their strengths.

---

[5] I defend pluralism about explanation as an alternative to both reductionism and some forms of anti-reductionism in Sober (1999b), Jackson and Pettit (1995) also are pluralists. David Wilson and I defend MLS theory in *Unto Others* (Sober and Wilson 1998).

[6] When I talk about the two "approaches" in what follows, I mean the two equations where the two addends are interpreted as described. Each approach can of course be supplemented.

I am glad that Okasha agrees that there is strong textual evidence for an MLS interpretation of Darwin's discussion of human morality. However, Okasha expresses reservations about this interpretation in connection with what Darwin says concerning the evolution of hybrid sterility; he thinks that Darwin may here be straying from the MLS framework because Darwin notes that an organism's inability to have fertile offspring when mating with an individual from a different species does not benefit the organism with the inability or its relatives. Okasha suggests that Darwin here may be closer to modern-day kin selection theory than to MLS theory.[7] In fact, Darwin goes beyond this mention of relatives; he says that the inability does not benefit the individual who has it, or its relatives, or "any other individuals of the same variety" (Darwin 1959, pp. 444–445). I don't think his mention of relatives undermines the MLS interpretation of Darwin's analysis, since the standard version of MLS theory (now called MLS1 in the literature) says that the fitness of a group is the expected number offspring organisms that the group produces. When Darwin denied that there could be selection for hybrid sterility, he was implicitly using an MLS1 format.[8] The organism's inability doesn't help the group to be productive any more than it helps the organism itself.

Okasha thinks that the definitions I use of group and individual selection pose a problem for understanding what group and individual adaptation are and for capturing Williams's (1966) distinction between group adaptations and fortuitous group benefits. Okasha is right that on my view a trait can evolve because of group selection and not be a group adaptation, but I don't think this is problematic. Let me explain.

Williams made the important point that adaptation at a level requires selection at that level. That is, if a trait is a group adaptation, it must have evolved because there was group selection; ditto for individual adaptation entailing individual selection (Sober and Wilson 2011). From this it is easy to see why a trait that currently is useful to the group need not be a group adaptation, since it may not have evolved by group selection. For example, suppose a zebra herd benefits when the individuals in it run fast rather than slow, since that reduces the group's risk of being wiped out by lion attacks. Even so, the trait may have evolved purely by individual selection wherein the fast zebras in a herd avoided predators more successfully than the slow ones did.

So if a trait is a group adaptation, group selection must have caused it to evolve. What about the converse? If a trait evolved because there was group selection (in the sense of that term used in DDWOB), does that entail that the trait is a group adaptation? Okasha says that I "imply" in DDWOB that the answer is *yes*. I hope I did not, since I think the answer is *no*. Consider two traits that are perfectly correlated in a metapopulation of groups; suppose that fast zebras have brown eyes and slow zebras have blue, and that running fast is advantageous whereas eye color is neutral. Suppose for the sake of argument that running fast evolved (in part)

---

[7] I discuss what kin selection has to do with genealogical relatives in DDWOB's second postscript.

[8] Gayon (1998, p. 72) discusses Darwin's commitment to "individualism" in connection with this problem.

because it was beneficial to the group. For example, suppose that lions decide which zebra herds to attack by assessing the herd's average running speed, preferring to go after slow herds rather than fast ones. If running fast evolved because there was group selection for that trait, then the trait is a group adaptation. Given the correlation of running speed and eye color, it follows that brown eyes also evolved because there was group selection. However, there was no group selection *for* having brown eyes, so eye color is not a group adaptation. Brown eye color was selected, but there was no selection *for* that trait. Selection-of and selection-for are importantly different concepts (Sober 1984). It is the latter concept that I use to define what an adaptation is (in DDWOB, p. 172 and in Sober 1993, p. 85) and it is selection-for that one needs to use in defining group and individual adaptations (Sober and Wilson 2011). The same argument would apply if we supposed that running fast evolved because there was individual selection for that trait; in that circumstance, individual selection would also cause brown eyes to evolve, but brown eyes would not be an individual adaptation. In summary, a trait can evolve because of group (individual) selection without the trait's being a group (or an individual) adaptation. I don't regard that as a shortcoming of my definitions. I therefore disagree with Okasha's claim that there is a "serious tension" between my "preferred conceptualization of individual versus group selection … and [my] endorsement of Williams' distinction…" By the way, neither the Price approach nor the contextual approach (as defined earlier) distinguishes between two traits that are coextensive.[9]

Okasha (2006, p. 97) says that the "tension" between the Price approach and the contextual approach arises because each satisfies only one of two intuitively plausible requirements: (i) that group selection occurs only if the fitnesses of individuals depend on the kind of group they occupy; (ii) that group selection occurs only if there is variation in the fitnesses of groups. Okasha thinks that the contextual approach satisfies (i) but not (ii) while the Price approach satisfies (ii) but not (i). He is right about the contextual approach, but, as Okasha acknowledges a page later, the Price approach *does* satisfy both requirements if an "interactionist" conception of group is used. The idea is that a bunch of organisms comprises a group only when they affect each other's fitnesses; it isn't enough that they live shoulder-to-shoulder (Sober and Wilson 1998).[10]

Okasha (2006, p. 98) has another objection to the Price approach that he thinks is more serious. He says that the Price approach "always" overstates the strength of

---

[9] A manipulation experiment can break the correlation, yielding data that show that there was selection for running speed, not for eye color. Still, the fact remains that the change-in-trait-frequency equation used by the Price approach for running speed and the Price approach equation for eye color must be identical if the two traits are perfectly correlated; similarly for the two contextual approach equations, one for running speed, the other for eye color. The contextual approach equation can be supplemented by a partial regression treatment of how running speed and eye color each affect an individual's fitness; a similar supplementation is possible for the Price approach equation (Frank 2012, pp. 1013–1014). These supplements go beyond what I am describing as the two "approaches."

[10] Okasha (2006, pp. 45–46) argues that the interactionist concept is sometimes inappropriate, but I find his reasons unpersuasive. I think the interactionist conception is right for MLS1, regardless of whether one is a contextualist or a Pricean; Okasha's point about genealogically defined groups concerns MLS2.

group selection. Here's an example of the kind of point that Okasha is making. He thinks that group selection must be weak when an individual's fitness depends only slightly on the kind of group the individual occupies. This can happen when there is lots of variation in fitness among groups and only a little within them; here the Price approach says that there is strong group selection. Since this assessment uses ideas from contextual analysis, I think Okasha's argument may beg the question.

Even so, I agree with Okasha that the Price approach has some real limitations. For example, Okasha (2006, p. 99) notes that the Price equation requires that the trait value of a group be the average of the trait values of the individuals in the group. This isn't a problem in the example of zebra running speed, but there are plenty of other cases in which the group phenotype of interest isn't an average of individual phenotypes. For example, in Wade's (1976) artificial selection experiments, there was group selection on group census size. The census size of a group isn't the average of the census sizes of the individuals in it. In this example, the Price approach does not generate the wrong conclusion about whether there is group selection; rather, the approach does not apply. The example shows that the approach is incomplete, not that it is false.[11]

MLS theory, as I understand it, is not committed to using the Price approach *or* the contextual approach. My picture is that MLS theory addresses a biological example by identifying traits that are advantageous to individuals within the same group and traits that are advantageous to groups. In both cases, advantageousness is a causal concept, so the causal distinction between group and individual selection is drawn *before* a mathematical model of change in trait frequency is applied. In my 1984 book I distinguish *source laws* from *consequence laws* in the theory of natural selection.[12] The former describe where fitness differences come from; the latter describe the evolutionary consequences of fitness differences. The physical environment and the biology of the organisms in the population induce fitness differences; those fitness differences induce changes in trait frequency. The equations used in the Price and the contextual approaches are both consequence laws and therefore don't capture some of the facts that are described in source laws. Some of the causal differences between group and individual selection are to be found in those source laws.

References

Darwin, C. (1959). *On the origin of species*: *A variorum edition*. In M. Peckham (Ed.), Philadelphia: University of Pennsylvania Press.

Frank, S. (2012). Natural selection IV: The price equation. *Journal of Evolutionary Biology, 25*, 1002–1019.

Gayon, J. (1998). *Darwinism's struggle for survival*. Cambridge: Cambridge University Press.

---

[11] The same point applies to Okasha's (2006, p. 99) observation that the Price approach requires that groups not overlap.

[12] I now think that "models" would have been a better term.

Hodge, J. (1977). The structure and strategy of Darwin's 'Long Argument'. *British Journal for the History of Science, 10*, 237–246.

Hodge, J. (2012). Darwin's Book: *On the Origin of Species. Science & Education*. doi:10.1007/s11191-012-9544-7.

Jackson, F., & Pettit, P. (1995). In defense of explanatory ecumenicism. *Economics and Philosophy, 8*, 1–22.

Lewens, T. (2007). *Darwin*. London: Routledge.

Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). London: Routledge.

Okasha, S. (2006). *Evolution and the levels of selection*. Oxford: Oxford University Press.

Richards, R. (2012). Review of Sober's *Did Darwin Write the* Origin *Backwards? Isis, 103*, 222–223.

Roche, W., & Sober, E. (2013). Explanatoriness is evidentially irrelevant, or inference to the best explanation meets bayesian confirmation theory. *Analysis, 73*, 659–668.

Ruse, M. (2012). Review of Sober's *Did Darwin Write the* Origin *Backwards? History and Philosophy of the Life Sciences, 34*, 481–514.

Sober, E. (1984). *The nature of selection*. Cambridge: MIT Press.

Sober, E. (1988). *Reconstructing the past: Parsimony, evolution, and inference*. Cambridge: MIT Press.

Sober, E. (1993). *Philosophy of biology*. Boulder: Westview Press.

Sober, E. (1999a). Modus Darwin. *Biology and Philosophy, 14*, 253–278.

Sober, E. (1999b). The multiple realizability argument against reductionism. *Philosophy of Science, 66*, 542–564.

Sober, E. (2011). Realism, conventionalism, and causal decomposition in units of selection: Reflections on Okasha's *evolution and the levels of selection. Philosophy and Phenomenological Research, 82*, 221–231.

Sober, E., & Wilson, D. (1998). *Unto Others: The evolution and psychology of unselfish behavior*. Cambridge: Harvard University Press.

Sober, E., & Wilson, D. (2011). *Adaptation and natural selection* revisited. *Journal of Evolutionary Biology, 24*, 462–468.

Wade, M. (1976). Group selection among laboratory populations of *Tribolium. Proceedings of the National Academy of Sciences, 73*, 4604–4607.

Waters, K. (2003). The arguments in the *origin of species*. In J. Hodge & G. Radick (Eds.), *The Cambridge companion to Darwin* (pp. 116–142). Cambridge: Cambridge University Press.

Williams, G. (1966). *Adaptation and natural selection*. Princeton: Princeton University Press.