

Teacher’s Guide: Examining Misunderstandings of Evolution through the History of Biology

Charles H. Pence*

Version 1.0: December 20, 2020

Contents

Introduction	1
1. Charles Darwin and the “Perfection” of Organisms	3
2. Darwin, Sir John Herschel, and Scientific Reasoning	10
3. Lamarck and the Inheritance of Acquired Characters	16
4. Orthogenesis and “Runaway Evolution”	22
5. W. F. R. Weldon, Genes, and Traits	27
Acknowledgments	32

Introduction

Students often enter our classrooms with a variety of fairly simple misunderstandings of evolutionary theory, absorbed from popular culture or as collateral damage from the battle between science and religion. Sometimes, dispelling these misconceptions can be a significant challenge, and can leave students feeling as if they’ve failed to “get” something obvious.

But when we turn to the history of biology, what we find is that many of the same kinds of errors that our students express are precisely those that characterized the period surrounding Darwin’s introduction of evolution. To put it simply, evolution is a difficult idea to master, and our students are making not silly or stupid mistakes, but the very same mistakes made by professional, practicing scientists! My aim here is to provide you with a handful of pre-packaged examples of how to integrate instances of these misunderstandings into your classroom, including relevant primary source excerpts.

This series of lessons is designed to address five such cases using real historical episodes – simultaneously introducing content knowledge about evolutionary theory, historical context, as well as engaging questions in the nature of science, now increasingly important in contemporary science-education standards. The broad format of each lesson follows the example found in Douglas Allchin’s *Teaching the Nature of Science*:

*Université catholique de Louvain, Institut supérieur de philosophie; charles@charlespence.net

Perspectives & Resources, which masterfully details some of the ways in which material in the history and philosophy of science can shed light on NOS education.

In short, then, these lessons cover the following student misunderstandings:

- The history of evolution is a steady progress toward *more perfect* or *more complex* organisms, with humans as the pinnacle
 - We’ll look at *Darwin’s own* struggle with exactly this question – while he repeatedly claimed that we ought not describe evolutionary advance in terms of progress, he nonetheless regularly slipped back into progress-based language.
- Evolutionary theory isn’t *truly scientific*, because it doesn’t use *the scientific method*
 - This was a critique raised against Darwin himself, notably by the nineteenth-century philosopher *John Herschel*.
- Organisms evolve by *striving for success*, and pass the results of that striving to their offspring
 - This *Lamarckian* view was advocated by a number of biologists in the late-nineteenth and early-twentieth centuries.
- Evolution isn’t always about adaptation, sometimes it can *get stuck*, creating non-adaptive features
 - Perhaps more a historical curiosity than a common misunderstanding, the theory of *orthogenesis* proposed precisely this, which proves to be an interesting way to examine evolutionary adaptations and non-adaptive characters.
- Mendel’s work in genetics means that *traits are “binary,”* either “on” or “off” as we might draw in a Punnett square
 - Even the very traits that Mendel himself studied in peas are more complex than they first appear, and *W. F. R. Weldon* argued early in the days of genetics that environmental influence is important.

In the lessons that follow, you will find historical vignettes, with images and suggested readings. Those vignettes are punctuated by “THINK” questions designed to encourage reflection on nature-of-science themes, and conclude with an explicit NOS reflection question to support direct discussion of major NOS topics. In the rest of this guide for instructors, each THINK question is briefly described to give teachers the material they need to advance student discussions.

As Allchin describes the type of material that follows:

Allchin (2013)

The primary purpose of the THINK questions is for students to develop scientific thinking skills and to *reflect explicitly* on the nature of science. The questions are open ended. The notes here are only guides about the possible diversity of responses.

In many cases, there is actual history as a benchmark (which can be shared after the students' own work), but by no means does it indicate an exclusively correct answer. Accordingly, the teacher may strive to avoid overt clues, fishing for answers, or implying that a particular response is expected or considered "more right." Again, the case study should illustrate the partly blind process of science-in-the-making. To help promote thinking skills, the teacher should encourage (and reward) thoughtful responses, well-articulated reasoning, and respectful dialogue among students with different ideas or perspectives.

Student inquiry often leads to requests for more information, and the teacher equipped with a deeper perspective is better prepared to provide guidance. The additional information below addresses these pedagogical demands, while also allowing the teacher to extend discussion, once students have found what they consider to be good solutions to the questions in the text.

Finally, I should note that this material forms a part of a project that also led to the publication of a book, *The Rise of Chance in Evolutionary Theory: A Pompous Parade of Arithmetic*. This work, a scholarly monograph focused more narrowly on the development of the methods of statistics and concepts of chance in evolutionary theory, would primarily be illuminating for the fifth lesson.

1. Charles Darwin and the "Perfection" of Organisms

THINK[1]: Evolution and Controversy / Social Context

Why might arguing for the view that species have changed over time have been seen as controversial, unusual, or problematic? What kinds of scientific, religious, social, or cultural views could you see it as threatening?

This warm-up question should be fairly familiar to students, since they are likely well aware that evolutionary theory has been controversial since its introduction. Answers that invoke religious connections are likely. The challenge is to get students to start thinking historically, about why this would have been controversial *in Darwin's day*. This could include, for instance: discussions of the status of the life sciences in the mid-nineteenth century (extinction has only recently been discovered, cellular biology is in its infancy and molecular biology does not yet exist); cultural ideas about an "ordered" or "perfect" world; the social structures that supported widespread slavery and colonialism; etc. In short, the goal is to get students to begin to think about the making of science as a socially and culturally situated, historical act.

THINK[2]: Racism, Evolution, and Social Responsibility

Why would this ancient theory so easily support a racist world-view? How could you detect other instances of racist theories in science? More generally, what obligations do scientists have to ensure that their theories can't be used to support harmful social outcomes?

The first level of this question is an empirical one: students should easily see that any theory that ranks all the organisms on earth into a single scale can very easily be modified to rank particular humans differently. The more interesting side of the question is the next one. What kinds of strategies might we use to detect the influence of racist beliefs on science? Here we might profitably talk about the need to cultivate diversity in the scientific community – not only for the sake of fairness for groups that might be kept out of science by discrimination, but, more profoundly, with the aim of incorporating the perspectives of such groups into scientific practice so that the community as a whole has a better chance to detect the influence of bias on our theories.

The final question concerns the general social responsibility of scientists. While there are some guidelines, expressed in places like the codes of conduct adopted by scientific professional societies, there is surprisingly little consensus about what kinds of duties scientists owe society at large. What sorts of consequences of their research should scientists be expected to foresee? At what point do potential ethical consequences mean that research should not be undertaken? What is the ethical value of “scientific freedom,” and how can it be balanced with impacts on the other communities that might be negatively affected by the results of scientific research? These questions are too large to resolve in a single discussion, but all scientific work relies, at least tacitly, on the adoption of answers to questions like these.

THINK[3]: Alternative Explanations / Designing Experiments

In Darwin's day, there were two alternative ways to understand the difference between humans and bacteria. One would describe humans as “higher” than bacteria, on the basis of looking at their apparent complexity. The other would describe humans and bacteria as having evolved for the same amount of time, perhaps at different speeds (where humans evolved faster).

Do you think that this is only a conceptual difference, or could we collect data or perform experiments that might let us tell which of these explanations is right? If so, what would those data or experiments look like? More generally, how should we think about the relationship between conceptual change and experiment in science?

It should be entirely possible, at this point, to pursue an evaluation of these two alternative explanations for apparent “progress” in evolution that paints either one as more successful and truer to the available data; scientists at the time were assuredly split on the issue.

The deeper question here concerns experimental design and the relationship between hypotheses and data. How could we build an experimental regime that could test the difference between these two hypotheses? We might look at shared features of

humans and bacteria, whether cellular or genetic. We could try to more clearly define “complexity,” so that we better understood what it was that we were trying to compare between humans and bacteria in the first place. We could also investigate rates of evolution in different lineages, to see if there was support for the hypothesis that humans were evolving more quickly than bacteria.

Finally, consider the overall relationship between conceptual change and scientific experiment. How are new theoretical innovations introduced? What does it mean if we develop a new conceptual way of understanding a part of science, to which we don’t yet have much experimental access? Arguably, this is what took place when Darwin introduced evolution by natural selection. How can we be sure that our concepts really map onto the outside world? What kinds of checks or comparisons could we do that don’t involve controlled experiment? Darwin thought (see especially the next lesson) that we could build structures of analogies, creating broad-scale explanations that demonstrated that natural selection would unify and explain a number of facts that, at the time, were only understood as the results of disparate and disunified causes. Is this good enough as a justification for introducing natural selection, or no?

THINK[4]: Scooping and Scientific Prestige / Social Class

Darwin was apparently very worried about what we now call being “scooped” – someone else getting credit for a scientific idea that you were really the first person to think up. Why might this kind of prestige be important for scientists? Should scientists be motivated by this kind of social credit, or is it harmful to the scientific process? Is it relevant to the story that Darwin was wealthy and well-connected, while Wallace was middle-class and not part of the traditional “scientific establishment?”

Students may not be aware that “scooping” is still a problem in contemporary science. This question allows them to explore the kinds of career, prestige, and social influences that might structure the lives of today’s scientists. Being a scientist is, after all, still a job, and scientists need grants and promotions in order to advance.

Whether and when this kind of pressure is beneficial to science is a different sort of question. On the one hand, the pressure to produce quality results drives the production of scientific knowledge, and so we might think that adding at least some such external forces to the generation of science is a net good. On the other hand, these pressures encourage scientists to “game the system” in a variety of ways, from taking a single result and publishing it in multiple, tiny parts (“salami science”) to trying to artificially increase numerical measures of research productivity (like the infamous “h-index”). This kind of pressure also exacerbates class differences in scientific training and productivity. Even in the nineteenth century, scientists who could afford to – as Darwin could, for instance – ship their specimens around the world could ensure their success in ways that others couldn’t; Wallace was forced to watch from a lifeboat as an entire trip’s worth of specimens burned aboard his ship. These effects continue today, as access to childcare or travel funding continues to impact career success.

THINK[5]: Scientific Networks and Communities

One thing that clearly helped Darwin here was the fact that he had important and influential friends who could quickly arrange a meeting at one of the most important scientific societies in the world. Major developments in science often involve not only the empirical or theoretical results, but also the social structures that you need in order to be able to distribute and publicize those results within the broader community. If Wallace had needed to do the same thing, without this kind of network, how could he have shared his results? What might you be able to do to bring attention to scientific results today that wouldn't have been possible for scientists in the nineteenth century? Are we better off now than we were then, or not?

Scientific knowledge is, of course, not only about experiments and theories – science is a social process, and if insights are not distributed to and then taken up by the broader scientific community, they aren't good for anything after all. Both in Darwin's day and our own, a significant part of this social networking was about face-to-face connection and personal relationships. For someone like Wallace, who was something of an outsider to this community, this kind of social capital was harder to come by. In today's digital research environment, with preprints and articles now (usually) readily available online, discussed on social media, and shared over e-mail, these barriers may have been reduced, or at least their shape is likely to have changed. Whether or not this has constituted an improvement (one might think of the spread of both accurate and inaccurate science communication online) is a matter for debate.

THINK[6]: Science and Colonialism

The example of Australia points to the importance of connections between science and colonialism, especially throughout the nineteenth century. Why do you think that scientists might have been particularly interested in what was happening in the colonies? How might this exposure have changed our understanding of the world? How might it have been harmful to the people living in the colonies?

The same goes for military expansion. Darwin's trip around the world happened on the H.M.S. Beagle, a British navy ship in charge of surveying the coastline of South America to produce high-quality maps. What other connections can you think of between military power and scientific discovery? How might these links have altered the shape of the science that was produced?

From the British colonization of Australia and India to the Dutch colonization of Southeast Asia, the flow of material goods, specimens of plants and animals, and indigenous knowledge back to the colonial centers of Western Europe was a crucial part of the development of the Scientific Revolution. It allowed for an expanded knowledge of the flora and fauna of the world – think of the first Western explorers to grapple with the dramatically different creatures of South America or Australia – as well as knowledge of local medical practices, geology, archaeology, and anthropology. Of course, such colonialism and military expansion was still colonialism and military expansion, and hence often had drastic consequences for indigenous people, whose contributions to scientific knowledge were, at best, often dismissed or ignored. It is also worth considering whether the very kind of science done on these voyages would

have been different as a result – do the products of a military voyage look different from those of an independent scientific trip?

THINK[7]: Finding the Problem / Evidence, Theories, and Concepts

What is the problem that keeps us from being able to judge this global sense of “higher” and “lower,” on Darwin’s view? Is it just that we don’t have access to enough data, or that our evidence is incomplete? Do we need a new theory to be able to understand it? Is it a problem with our concepts?

An important part of understanding a problem that was raised for historical scientists is to clarify just what exactly was at issue. We can see several potential difficulties here in Darwin’s hesitation about progress. First, we might be worried that we just don’t have enough data; since we can’t really understand “evolutionary distance” in Darwin’s day, we don’t have the kinds of measurements connecting species that we would need to understand increases in evolutionary complexity. We could also think that more complete evidence, for instance in the fossil record, would give us a better understanding of just how species like *Typhlops* actually evolved. More profoundly, though, we might worry that we have conceptual work to do in order to understand what’s happening here. What do we really mean by evolutionary complexity? How should we actually understand “progress?” If this is what’s at stake, then no further amount of data is going to enable us to resolve the question; we need theoretical advances in order to proceed.

THINK[8]: Understanding Historical Contradiction

What should we do when we think we have found a case of confusion or contradiction in the works of a scientist? How might these kinds of contradictions cause problems for the scientific theories that the author was hoping to defend? On the other hand, how might this kind of contradiction or tension serve as a useful aid for the generation of new knowledge?

Even in the case of scientists as renowned as Darwin, careful historical analysis will almost always reveal theoretical problems, empirical failings, or outright confusion. What should be the impact of these episodes on our understanding of these historical figures? Emphasizing the humanity and fallibility of historical “great minds” is vital for making the scientific process seem more approachable. Further, these moments in the history of science can often serve as productive locations of friction – they can point us toward places where particularly important theoretical or empirical facts are in play, improving our analysis and understanding of the relevant science. In addition, we can think of this kind of tension as setting up the sorts of problems and doubts that are needed to drive scientific knowledge forward. Spotting the holes and weaknesses present in science – and of course, they always exist! – can be extremely helpful for encouraging today’s practitioners to advance the frontier of knowledge.

THINK[9]: Alternative Explanations / Cross-Species Comparisons

If we wanted to compare the case of a bird's flight with that of a flying fish, what kinds of things would we need to keep in mind? What would you need to know about the lives of a flying fish and of a bird to make such a comparison? Do you think that the comparison even makes sense? What would the comparison help us learn about the structure of evolutionary theory?

Cross-species comparisons, especially the farther apart those species are phylogenetically, make for a particularly interesting challenge in understanding evolutionary theory. These comparisons are often extremely appealing – for example, the evolution of eyesight has taken place at least four or five times at places radically distant from one another in the tree of life. These comparisons can, however, be somewhat problematic. One might assume that something as complex as an eye would have to be a case of homology (evolution via common descent) rather than analogy (convergent evolution of the same part multiple times), but it turns out that intuition is sometimes misleading. Eyes, for instance, almost certainly have multiple, independent origins – but these multiple origins rely on a number of the same genetic parts, recruited independently for their utility in developing eyes! Thus, while these kinds of comparisons might still make sense, there will remain many alternative explanations in play, and deciding between them may require a large amount of data. We might learn, if we're successful, information about the evolutionary relationships between distant clades, common ancestors, and the structure of the tree of life, but we'll need to be sure that we have properly established the kinds of evidence required to support our claims.

THINK[10]: Notions of Progress

How is this concept both like and unlike our traditional understanding of “progress?” Many in the nineteenth century, when Darwin introduced evolution, were worried about the overall direction in which their culture was headed. What would they have thought about the impact of evolution on their view of the world and their place in it? How might this, in turn, have affected what kind of theory scientists like Darwin would have tried to develop?

Progress might be cashed out in a variety of different ways. We might imagine increases in complexity of structure, or in the way that organisms developed. We might focus on only one character or kind of character – improvement in speed or locomotion, improvement in eyesight, or, perhaps most common historically, improvement in cognition or mental capacities. Any one of these choices might result in a profoundly different sorting of organisms over the history of life. This, then, intersects with cultural notions of progress – the nineteenth century saw dramatic advances in industrialization, agriculture, and quality of life, at least for those members of society who were sufficiently wealthy and well-connected. These measures of societal impact could well push us in different directions than “biological” progress. This tension between “social Darwinism” and “biological Darwinism” is important to underline for students, who may well be exposed to naive forms of social Darwinism as it is sometimes described in popular culture.

THINK[11]: Evolution and Our Place in the World

Returning to one of the questions from early in the reading, why would a view like this possibly have been scary at the time – that is, why would Darwin have wanted to reassure his readers? Do you think it’s still troubling for readers today, or are we used to the idea of evolution? What kinds of relevant cultural and societal changes have taken place since that could be important to answering this question?

Many stories about the place of humans in the world, of course, have us as comfortably “on top” of creation; the Christian creation story, for instance, talks about our “dominion” over the world. Biological evolution threatens to upend this, and thus would be, at least for this reason, potentially frightening. More broadly, the idea that humans aren’t particularly special animals could be seen as generally upsetting. We now think that there are no major qualitative differences between us and other creatures – there is a difference in degree between our mental capacity and that of dolphins or crows, perhaps, but there is probably not a difference in kind. This sort of continuity also disturbs our social as well as moral separation from the rest of the animal kingdom. Social changes in the time since Darwin have seen a decrease in the tension between religious belief and evolutionary theory and a rise in ethical treatment of animals, both of which seem to make culture at large more accommodating to the impact of these once-shocking facts about evolutionary theory.

THINK[12]: Summary Evaluation

After everything you have now read, do you think that we should talk about evolution in terms of progress, or not? As we have seen, there is still a sense in which evolution improves organisms, and it is undeniable that advanced features now exist that once did not. Is this enough to support a progressive understanding of evolution, or do you think that the arguments against progress are more compelling? What consequences would this have on the kind of science that you would do, if you were studying evolution professionally?

This summary question is just intended to give students a chance to form an opinion about the overall empirical question that’s discussed throughout this lesson. It remains an open issue among biologists, so there’s no sense in which they might arrive at the “wrong answer.” In particular, considering whether or not evolutionary progress is something worth studying allows students to think about how they might design large-scale research programs, and how those research questions would be influenced by their more general conceptual perspective.

THINK: NOS Reflection Questions

Explicitly reflecting on these themes serves partly to help the students in reviewing the material that we have just covered. But research also indicates that if students do not *explicitly* consider the impact of what they’ve read on NOS issues, they will often fail to see these links. It is therefore crucial that the lesson closes with time for this detailed reflection. The following NOS themes are listed, with links back to earlier

THINK questions:

- nature of scientific credibility (THINK 4, 5, 8)
- social responsibility of scientists (THINK 2, 6, 11)
- role of cultural beliefs in science (THINK 1, 10, 11)
- role of racial/class bias (THINK 2, 4)
- alternative explanations for phenomena (THINK 3, 7, 9, 12)
- scientific collaboration and competition (THINK 4, 5)

2. Darwin, Sir John Herschel, and Scientific Reasoning

THINK[1]: The Scientific Method

Does this match your idea of “the scientific method?” Can you think of elements of scientific practice that might be missing from this account, or fields of science whose work couldn’t be described in this way? More broadly, do you think that all parts of science use one, single “scientific method,” or not?

Lots of students will probably already have learned the classic “three-panel science-fair project board” view of the Scientific Method. Hopefully they’ll also have had a chance to think about how various parts of science don’t actually map onto that naive picture of how research is done!

The goal for this question is to compare this traditional picture of the scientific method with the “inductivist” position described here and common in the nineteenth century, as well as other examples that students might be able to bring to the table. Current philosophy of science has settled on a broad consensus that scientific methodology is too pluralist to be captured by any single such method. Good contrasts for discussion (about which we’ll see more in the 7th question, below) might include thinking about paleontology or astronomy as sciences that are often incapable of performing controlled experiments, as well as natural history (or even bird-watching) as purely observational. Students could also consider engineering or medicine, where we have instrumental in addition to epistemic goals that our knowledge is supposed to support.

THINK[2]: Agriculture and Evolution / Interdisciplinarity

How would data that we acquired from agricultural breeding be relevant to answering these kinds of questions? How is it similar and different from data that we would collect from natural organisms? What kinds of problems might you run into if you tried to use data derived in one context to support a theory derived in a very different context?

Scientific theories often need to be reinforced by, or even created from, data that weren’t “gathered for purpose” – whatever observations we might be able to cobble together sometimes have to suffice. In the early history of evolutionary theory and genetics, this meant a strong link with agriculture. The nineteenth century saw a massive effort to create stable, reproducible breeds of plants and animals, and this meant that there was a huge amount of potential information about the variations that animals would experience when bred over the long term. Some such data was

correlated to environmental changes, changes in food or climate (as, for instance, varieties from warm climates were brought to Northern Europe for cultivation), while others derived from extensive programs in hybridization and the efforts to produce novel varieties (this tradition was also the one that led to Mendel's work, about which more in the final lesson).

This means, of course, that we have to figure out how to use such data – agriculture is not evolution, and the conditions under which agricultural breeds grow and are selected are extreme special cases by comparison to evolution in general. We thus have to look out for problems with our data. Without realizing it, agriculturalists have likely focused only on a very small number of traits of interest. Breeding – even for our oldest varieties – has only gone on for a small number of generations. Similar kinds of concerns can be generalized for the use of interdisciplinary data across the sciences.

THINK[3]: Evidential Standards

Do you think it would be better to be too strict with our standards for scientific investigation, or more relaxed? What kinds of problems might arise in each case, and how could we balance or possibly correct them?

This is a hot-button issue in today's sciences: the "replication crisis" is sometimes phrased as being directly a question of evidential standards. While the common p -value used in the social or human sciences is 0.05, the detection of the Higgs boson by CERN's Large Hadron Collider used a 5σ standard for statistical significance, equivalent to a p -value of around 3×10^{-7} . That's a dramatic difference in standards!

Of course, these standards have consequences; it would be nearly impossible to obtain that kind of precision for any research that required measuring living things! The major worry about over-strict standards, then, is making scientific investigation impossibly stringent, ballooning our rate of false negatives, and stagnating the process of science. On the other hand, the consequences of lax standards have also become increasingly visible: a proliferation of non-reproducible results, declining public confidence that the scientific process is generating genuine knowledge, and wasted effort as scientists pursue blind alleys. Balancing these is a careful and difficult question of career incentives, social pressure, and tailoring our science to the kinds of data available.

THINK[4]: Disciplinary Specialization

We don't see very many scientists today that are famous for doing so many different kinds of things. The idea of being a "polymath," someone who is good at many different fields of knowledge, has steadily become less possible. What reasons can you think of that might be driving this kind of disciplinary specialization? Is it a good thing that scientists today are more specialized than they were a century ago? What would be the advantages and disadvantages?

The onward process of disciplinary specialization in the sciences has proceeded without restraint for several centuries now, as scientific fields steadily gained their inde-

pendence from “natural philosophy,” and proceeded to divide into their own finer and finer specialities. There are many good reasons for this – perhaps foremost among them the fact that scientists need to be familiar with significantly more detailed theoretical and experimental work now than they once did. That said, there has recently been a significant push to encourage interdisciplinary and transdisciplinary work, framed by the vague sense that this hyper-specialization has led researchers to miss certain kinds of important connections between fields, novel solutions to problems, and so forth. Both of these aspects are already reflected to some degree in our science teaching, and students should thus have plenty of experience to reflect upon!

THINK[5]: Old Evidence and New Evidence

When we propose a new scientific theory, we will have some set of evidence that we already know, which has led us to think about that theory in the first place. After we propose it, we will discover new evidence, things that we didn't already know about before the theory was proposed. This difference between old evidence and new evidence has often been taken by philosophers and scientists to be very important to understanding whether a theory might in fact be accurate.

What do you think are the important differences between old evidence and new evidence? Why might you say that new evidence would be better than old, or vice versa? Is there a difference between a theory's being “consistent with” old evidence, and its “explaining” new evidence?

Philosophers of science have extensively debated whether or not there is a difference between old evidence and new evidence. On the one hand, evidence that we already had at the time a theory was created might be seen as “baked into” that theory, in the sense that the theory could have been tailored to accommodate it. If that's the case, then going on to say that the evidence “confirms” the theory would only be true in a trivial, uninteresting sense. But then again, sometimes theories can be created without thinking directly about some piece of old evidence, and the problem of trivial explanation doesn't arise. Einstein's theory of general relativity, for instance, was taken to be supported by the fact that it explained the procession of the perihelion of Mercury – even though this fact had been known for a long time, Einstein wasn't *thinking about it* when he derived general relativity.

This is also related to the question of “consilience” of evidence. Does a theory – like Darwin's! – become more plausible if it provides an explanation for large numbers of different kinds of facts that had previously been thought to merit different explanations? Many have thought that the answer to this question is yes, but it's interesting to think about what kind of confirmation this is. Why would we think that explaining lots of different kinds of old evidence with a single theory would constitute a vote in its favor?

THINK[6]: Life Sciences vs. Physical Sciences

One important difference between Herschel and Darwin is simply a question of the fields in which each worked: Herschel is primarily a physicist or engineer, and

Darwin is a life scientist. Can you think of reasons that we might need different standards for what counts as “good science” in biology versus in physics? What about in the social sciences? Do these standards mean that some of these fields are “more scientific” than others, or not?

This topic was broached to some degree in question 3 above, and a piece of it will be discussed again in the following question. The idea that “harder” and “softer” sciences can be separated by degree of consistency, rigor, or “how scientific” they are is fairly common culturally. That said, once we begin to unpack questions about evidential standards, methodologies, and types of argument, we can see that these questions are much less straightforward. To declare the physical sciences “more scientific” than the life sciences involves making a fairly large set of choices about what kinds of method, data, or inference are more or less valid – choices that seem much less obvious than they might when they are packaged as part of a general stereotype.

THINK[7]: Experimental Sciences vs. Historical Sciences / Learning by Making

Another distinction that might help us explain the difference between Herschel and Darwin’s views of science is the contrast between science which proceeds mainly by performing controlled experiments, and science that primarily involves observation of phenomena in the world around us. What kinds of differences in data or theory might arise as a result? What kinds of scientific procedure might be possible in one case but not possible in the other?

A similar kind of claim in the philosophy of science is sometimes described as the importance of “learning by doing” or “learning by making.” Do we know more about a scientific system when we can create one ourselves, or can we learn all there is to know by observing systems “in the wild?”

Recent philosophy of science has underlined a distinction between the classic “experimental sciences” and what are often called either “observational” or “historical sciences.” While in physics, if we want to generate more data to test a hypothesis about the dynamics of a system we can go perform more controlled experiments, this possibility is often unavailable to us in sciences like evolutionary theory, geology, or astronomy. We can’t go create a second Earth and re-run the history of life on it in order to see how things might go differently (Stephen Jay Gould called this the idea of “replaying life’s tape”). That must mean that when we’re trying to explain historical events, like the extinction of the dinosaurs, we need different kinds of evidence, different kinds of inferences, even different kinds of science than we would when trying to understand particle physics.

The same goes for the distinction between science and engineering, often caricatured by engineers as the importance of “learning by making.” Recent work in synthetic biology, for instance, has echoed this idea that we don’t really understand what’s going on in a system unless we know how to build one ourselves – so we won’t be able to really understand life until we can build a cell from scratch. There’s something to this claim, insofar as we would undoubtedly obtain all kinds of important insights from building such an artificial cell, but how to explain that intuition in more detail

becomes challenging very quickly.

THINK[8]: Response to Criticism

How might you respond to these criticisms, if you were Darwin? Does he need to make changes in his theory in order to account for Herschel's view?

Student responses here could vary dramatically. Darwin could have dug in his heels, arguing that his treatment of variation – insofar as we could still be confident that variation *did indeed* occur – was sufficient to provide the premise he needed for his understanding of natural selection. He could have gone away to try to derive the law of variation himself, or even accepted divine direction for variations as Herschel indicated. What he in fact did was resist at the level of his approach to the philosophy of science. Recall that one of the weaknesses of the naive inductive approach with which this lesson began was that it couldn't explain contemporary theories of light. Darwin wrote in correspondence that he thought he was doing with natural selection exactly what theorists of light had done – proposing a novel hypothesis for how organic change might have come about, then seeing what its consequences looked like and what kinds of natural phenomena it was able to explain if it were true. In that sense, Darwin thinks he's doing something that Herschel should have agreed with, even in the absence of the kind of "complete" theory that Herschel was demanding.

THINK[9]: Black Boxes / Resolving Dispute about Standards

We might rephrase this part of the disagreement between Herschel and Darwin as follows. Darwin believes that, because there is currently no theory that describes how variation works in the wild, he can treat the way that variations appear as a "black box." We have enough evidence, he thinks, to believe that the black box works, even if we don't know how. Herschel, on the contrary, thinks that we have to open up the black box, and provide an explanation for what's going on inside, for how variations are actually created.

How do you think we could resolve this dispute between the two? More generally, if we have a debate between scientists not over the empirical facts of the matter, but over what kinds of standards we should accept for a scientific explanation, to what kinds of resources should we turn to solve it? What do you think would be a "good argument" from either Herschel or Darwin for their view? Who should get to decide what those standards for explanations are?

Important not only for Darwin, but also for our understanding of the replication crisis, disputes about meta-level standards in the sciences are difficult to resolve. Here we have Herschel arguing that Darwin has illegitimately "black-boxed" a piece of theory that we should have to open in order to really understand what's going on in the history of life. It doesn't seem that more data would be able to resolve this dispute – not all scientific arguments can be answered by collecting more data! Rather, they need to argue in terms of their respective understandings of what good science would look like, or what the acceptable methodologies would be for engaging in scientific discovery. That's much more difficult, and starts to look more like doing

the philosophy of science than doing science itself. Thinking as well about the way in which such an argument would end can allow students to consider how these are socially mediated arguments – the question of which such standard to accept is, in the end, a community decision that cannot be made by a single individual in isolation.

THINK[10]: Genetics and Evolution / Reductionism

What exactly do you think we would get by adding genetics to evolution? One answer might be that we would understand what was happening at a “lower level” (that is, the biochemistry) as a way to explain what was going on at the “higher level” (that is, organisms). There is a powerful trend toward this sort of *reductionism* in the sciences – the idea that explanations that work at lower levels are “better” than those that work at higher levels.

Do you think this kind of reductionism is correct, or not? What advantages and disadvantages are produced by an explanation in terms of chemistry, as opposed to a biological explanation? What might this mean for how we think about the relationships between scientific fields?

Empirical research indicates that people, as a rule, tend to find reductionist explanations for phenomena more satisfying – given a choice between an explanation of the same phenomenon at the ecological level and one at the biochemical level, people will tend to prefer the biochemical. This has been a pronounced trend not only in the general public, but also in at least some parts of the sciences as well, and has been resisted by scientists attempting to defend, for instance, the autonomy of biology from underlying claims in physics. Unpacking just what the advantages of these lower-level explanations are supposed to be, however, is more difficult. Are they taken to be based on more secure knowledge, less likely to be mistaken? Better supported, theoretically or experimentally? This kind of reflection can help to round out students’ thoughts in questions 2 and 6 about the relationships between scientific fields.

THINK[11]: Deduction, Induction, Abduction

The difference here between *kinds* of arguments is an old one in the study of philosophy. Some arguments, like in mathematics or logic, are *deductive* – if their premises are true, their conclusions *have* to be true. Some, like Herschel’s approach to science, are *inductive* – we move from a large collection of data to a general theory about that data. Some, finally, are *abductive* – we support some explanation of our phenomena because it is much more compelling than the alternatives.

Can you think of examples of arguments in science that use each of these three kinds of argument? Do some of them seem more powerful than others? Some will be available and unavailable in different fields at different times, depending on the evidence we have. Does this correspond to a judgment about the overall “quality” of those sciences?

These three different kinds of argument are each extensively used in science. Deductive argument underlies mathematical or logical reasoning, and hence is the kind of

inference that grounds the use of a mathematical law in physics to derive the trajectory of a particular particle. If the initial conditions are specified, the resulting trajectory that one calculates is necessarily true – there’s no possibility that things could go otherwise. For an inductive argument, on the other hand, which generates a theory from a large body of observational data, it always remains possible that one future observation could overturn the theory that we’d constructed. To take a classic example, if every swan we see is white, and we thus form a theory that all swans are white, all it takes is a single Australian black swan to overturn our theory. Abductive inference is also not guaranteed; the best explanation we have available could always be modified, not only by the generation of more data, but by the generation of a newer, better explanation for the same phenomena. All three of these kinds of argument, however, interact in the structure of scientific explanations, and generate different kinds of relationships between the data that we collect and the scientific knowledge that results.

THINK: NOS Reflection Questions

Explicitly reflecting on these themes serves partly to help the students in reviewing the material that we have just covered. But research also indicates that if students do not *explicitly* consider the impact of what they’ve read on NOS issues, they will often fail to see these links. It is therefore crucial that the lesson closes with time for this detailed reflection. The following NOS themes are listed, with links back to earlier THINK questions:

- completeness of evidence (THINK 3, 9)
- robustness (agreement among different types of data) (THINK 2, 5, 7)
- evidential relevance (empiricism) (THINK 1, 3, 5, 7)
- consilience with established evidence (THINK 5, 9, 11)
- role of analogy, interdisciplinary thinking (THINK 2, 4, 6, 7)
- forms of persuasion (THINK 8, 10, 11)
- response to criticism (THINK 8, 9)

3. Lamarck and the Inheritance of Acquired Characters

THINK[1]: Acquired Characters / Everyday Experience

For many scientists over hundreds of years, this conclusion seemed obvious; yet for us now it might seem strange. What do you think about it? More broadly, when is our everyday experience useful as a source of scientific knowledge, and when might it mislead us? How can we tell the difference?

Students are probably now used to the idea that changes that are passed on to offspring are only those that are encoded genetically. So it may be hard for them to see that Lamarckism could have been a particularly attractive way of understanding the natural world. It’s worth taking some time, then, to examine the intuition that everyday experience would show us that people can make heritable changes in physical and mental characters over the course of their lives. Combine this with the fact that children often go into the same careers as their parents, and one can start to understand why Darwin contemplated in a notebook whether or not blacksmiths would have

children with unusually muscular arms.

The general question of how we should test this everyday experience – whether by controlled experiment, by collating it with our existing theories, or in some other way – is much more open to interpretation!

THINK[2]: Nationalism and Science

Lamarck is also interesting because of how he relates scientific pride and national pride. For many years in France, it was thought that Lamarck had largely invented the theory of evolution, and Darwin had in some sense stolen the credit. What's different between Darwin's view and the picture of Lamarck's theory illustrated here? How might the question of inventing scientific theories become a matter of national pride or a dispute between countries?

The differences between Lamarck's picture and Darwin's should be clear enough. Lamarck explicitly rejects the idea of common descent – new organisms are appearing all around us all the time, and today's bacteria have simply not been around for as long as today's vertebrates. Lamarck also had no conception of natural selection – organisms' responses to their environment sufficed to drive apparently adaptive changes.

The nationalism connection is more interesting. Scientific prestige – both then and now! – is certainly a kind of currency on the world stage. (Think of the perennial worrying by some figures in Western Europe and North America that countries there are “losing their edge” to China or India.) There can thus be scientific consequences to diplomatic arguments. Scientists looking to play up France's scientific supremacy over England's in the late-nineteenth century, and even up until the mid-twentieth century, continued to push the idea, first that there was nothing novel in Darwin's theory, and then that in fact what was novel in Darwin's theory was wrong, adopting a fully fledged neo-Lamarckism.

THINK[3]: Physical Evidence and Biological Theory

We saw that an important part of evaluating evolutionary theory was the calculation of the age of the earth that had been derived in physics (and, later, the discovery of radioactivity). How do you think the knowledge generated in physics might be related to biological claims? When would these be useful? Can you think of circumstances where appeals to other sciences might be unhelpful or problematic?

At least in Darwin's day, there was the impression that Kelvin's derivation of the age of the earth – based on the hypothesis that it had begun as a fully molten sphere of rock that had been steadily cooling in deep space ever since – would trump Darwin's calculations for it based upon either geology or evolutionary concerns. In some sense, this is unsurprising, as the scientific claims that supported the age of the earth (primarily thermodynamics) came from sciences that were much more mature than Darwin's early-stage proposal of evolution. The question for students to consider is whether this kind of relationship should hold generally. Do physical sciences (or

“lower-level” sciences, following on the discussion of reductionism in the last lesson) always take precedence over the life sciences? It’s not clear that this is the case in current scientific practice, where dating of strata can rely on a complicated combination of physical carbon dating, geological analysis of the rocks themselves, and paleontological assessment of the fossils found there.

THINK[4]: Randomness and Chance in Evolution

The idea of “random variation” has always been difficult to interpret in evolutionary theory. By it, Darwin means that variations are not biased in the direction of what the organism “wants,” but rather occur without regard to whether they will be helpful or harmful. Natural selection, on the contrary, is not at all random, in its ability to drive populations toward increased fitness.

What kinds of misunderstandings do you think could arise from the description of variation as “random?” What difference is made by the presence of natural selection within the evolutionary process? In general, what might be some differences between scientific theories that describe their results as probabilities versus those, like Newtonian physics, that describe precisely what will happen?

As the question states, the sense of “random” needed for evolutionary variations is only, essentially, the opposite of neo-Lamarckism: variation arises without having any “bias” in the direction of traits that might be helpful or useful for the organism (things that it “wants”). Natural selection’s push of populations in the direction of increased fitness is the very opposite of random. That said, misconceptions abound, particularly as creationists continue to describe evolutionary theory as entailing that evolutionary change proceeds entirely randomly.

The primary philosophical upshot to consider in this context is that evolutionary theory is assuredly notable for offering only probabilistic predictions. First, the relevant variations might not in fact arise to enable an organism to adapt in a particular manner. Second, even if they were to arise, natural selection offers us no guarantees: fitter organisms are only *more likely* to outcompete the less fit. This is obviously different from the case in classical physics, where outcomes are known with complete certainty – roll the ball down this inclined plane, and it *will* have the given velocity at the bottom. Quantum mechanics, of course, returns us to the realm of probabilities. This is yet another fairly stark difference in types of scientific knowledge, each arising from mature and well confirmed theories in different domains.

THINK[5]: Fossils and Evolution

The fossil record is an important source of data for our understanding of evolution, but it is also quite strange. What do you think might be some advantages and disadvantages of using fossil data to support a conjecture in evolution? How might we correct some of those disadvantages with other kinds of experiments to produce a more robust hypothesis?

The potential problems with the fossil record are numerous – fossilization is a tricky physical process that, even in the best case, only happens to some parts of some kinds

of organisms in some locations on the earth. The fossil record is difficult and expensive to explore, and we only know about fossils that we have invested significant time in excavating, preparing, classifying, and analyzing. We are also (of course) incapable of generating more fossil data when we want it – the record is a static artifact.

That said, we are now more and more able to connect fossil data with various kinds of experiments and simulations. Computer analysis can reconstruct the forces acting on organisms, or the kinds of musculature that would be required for giant sauropod dinosaurs to be able to walk. Comparative analysis between fossil and living organisms can help us see how we ought to understand those fossils we do find. And at least occasionally, fossilized specimens can be studied using techniques from molecular biology. All of these, however, can be interpreted as ways that we might attempt to work around the peculiar status of evolutionary theory, which, in part, has to reconstruct the history of life from whatever data we're able to find.

THINK[6]: Persuasion / Credibility

Hyatt is openly making an appeal to persuasion here. In placing himself in opposition to Darwin, he explicitly lists the names of authorities – Lamarck himself, as well as Cope and Ryder – who agree with him, calling them “some of our leading scientific men.”

What do you think this says about Hyatt's theory and his position within the scientific establishment? How should we understand these arguments that appeal directly to the authority of other scientists? When do you think that they might be valid?

Contemporary scientific writing rarely contains appeals to credibility or personal credentials that are quite this explicit, but this kind of direct posturing was fairly common until relatively recently. More importantly, it absolutely still takes place – only now, it's expressed in a relatively more coded fashion (for instance, in lists of citations) instead of overtly.

Hyatt would have never thought that such persuasion was explicitly necessary if he didn't think that he was something of a “scientific outsider,” needing to demonstrate the power of his theory to a largely skeptical scientific community. But that doesn't mean that these kinds of appeals have to be groundless. Showing that there is an alternative to an existing, dominant theory can sometimes be important – especially if, as was the case in Hyatt's day, that dominance was by no means complete. Part of the process of doing science is itself this process of winning adherents to a new theory.

THINK[7]: Alternative Explanations

This leaves us three alternatives in play for the explanation of these non-adaptive characters. One is a classic, Lamarckian force of upward progress caused by struggle with the environment (as supported also by Hyatt). One is an unknown force driving organisms in particular directions, but without any invocation of adap-

tation (as supported by Cope). Lastly we have Darwin, who would presumably have appealed to natural selection to explain these changes.

How should we compare these possibilities? What kinds of data does each explanation have to support it? What features of the natural world does each leave unexplained or unconsidered? Which do you think would have made the most sense had you been asked to choose in the late-19th century?

In essence, this is a question asking students to consider how appealing they might have found each of three different partial explanations for a phenomenon. Each seems to have – at least, according to its supporters – data which would support it (Hyatt’s struggles against the environment, Cope’s progress in the fossil record, and Darwin’s cases of natural selection). Each is also relatively limited, in that (even for natural selection) there were a number of cases at the time that could not be adequately explained on each theory. One might foster student debate here, trying to get the students to consider what evidence would have been decisive had they been choosing between the three in the late-nineteenth century.

THINK[8]: Contradictory Data / How Theories End

Imagine that you were a neo-Lamarckian, and you were presented with the following experimental result. Rats have been trained to solve a maze for multiple generations in the lab, and in spite of this, their descendants don’t wind up solving the maze any faster than their ancestors. What kinds of responses might you have? How could you still defend your theory? More broadly, when do you think a scientist would *have* to admit defeat, to concede that their theory had failed? How much evidence, and evidence of what kinds, would be required?

A classic result in the philosophy of science, derived independently by Pierre Duhem and W.V.O. Quine, argues that we never have enough data to *conclusively* choose between theories. No matter how apparently fatal a counterexample to a scientific theory might be, we could always choose to modify other hypotheses in order to save the theory (up to and including extreme cases, like arguing that the results were faked, or that the experimenter was hallucinating). There’s no *logical* point at which we are *forced* to say that a scientific theory has failed.

This is clear in the case of our putative neo-Lamarckian example here as well. We have an instance of organisms struggling with their environment – rats that have to solve a maze to get food – and hence exactly the kind of thing that a Lamarckian process should improve over the long run. But the experimental data show that there’s in fact no change. There are a myriad ways to dig in one’s heels and save the theory: claim that maze-solving was in fact not the kind of thing that Lamarckism would improve (but, then, what would be?), claim that the lab experiment had not been carried out for long enough, or that the experimenters weren’t measuring “success” in the right way (in fact, they were checking how long it took the rats to solve the maze), etc. What kinds of data would it take to really consider the theory refuted, then?

THINK[9]: History of Science for Science

We not infrequently see examples like this, where theories that were thought to be long since disproven gain a new life as they are reinterpreted in light of fresh data. What do you think this might tell us about the relationship between the history of science and the practice of science today? Should we encourage scientists to learn more about history, or would that be a waste of time?

This question is inspired by the work of Hasok Chang, who has argued that one of the principal uses for the history and philosophy of science should be in expanding the conceptual space available for scientific theorizing. Taking the works of past scientists seriously can let us see avenues – which were once thought to be extremely promising by professional scientists, not just cranks – that have been closed off for so long that they have ceased to be evaluated, even if new data or new perspectives could help us understand them in new ways. Chang argues that precisely this kind of thing has taken place in chemistry, for instance, where historical approaches to measurement can shed new and exciting light on contemporary work. This must, of course, be counterbalanced against the drive for specialization and content knowledge all too familiar to science educators.

THINK[10]: Scientific Terminology

Do you think it is helpful to refer to contemporary results in molecular biology in terms of old theoretical names like “neo-Lamarckism?” Give some reasons that this identification might be both confusing and helpful.

There is currently a heated debate in molecular biology over the status of the term “Lamarckism.” While some authors have argued that these methylation results (and others) constitute something very much like the inheritance of acquired characters, others have countered that reintroducing a term that had described a cluster of views for some fifty years, and which included a great number of claims *besides* the inheritance of acquired characters, adds more conceptual baggage than it does clarification. There’s thus a discussion to be had here about the role of scientific terminology in general. Are we intending only to be maximally accurate, or is the invocation of Lamarckism (and hence a connection to an older biological tradition) serving a different kind of role, such as generating intuitions about how we think these systems might work?

THINK: NOS Reflection Questions

Explicitly reflecting on these themes serves partly to help the students in reviewing the material that we have just covered. But research also indicates that if students do not *explicitly* consider the impact of what they’ve read on NOS issues, they will often fail to see these links. It is therefore crucial that the lesson closes with time for this detailed reflection. The following NOS themes are listed, with links back to earlier THINK questions:

- evidential relevance (empiricism) (THINK 1, 5)
- role of cultural beliefs (national identity) (THINK 2)
- interdisciplinary thinking (THINK 3, 5, 9)

- role of probability in inference (THINK 4)
- role of imagination and creative synthesis (THINK 5, 7, 10)
- resolving disagreement (THINK 3, 6, 8)

4. Orthogenesis and “Runaway Evolution”

THINK[1]: Power of Natural Selection

Can you think of some characters of organisms that might be a problem for this view – characters that don’t seem as though they were produced by natural selection? How might Weismann respond to your examples?

One of the main thrusts of this lesson is to get students to think about the nature and evolution of apparently non-adaptive characters, so the first question is designed to start off by thinking of examples. Closest to home, they have probably heard that the function of the human appendix is hotly debated; wisdom teeth are an even better example of a trait that we retain that is almost entirely non-adaptive. One could also think about instances of what is now called *exaptation* – that is, when a trait is initially selected for one purpose but later co-opted for an entirely different one. We think that this was the case, for instance, for the wings of birds, which were initially designed for dinosaur temperature regulation, only later became useful for a sort of gliding, and finally were converted into organs for flight. In this sense, the obvious selective explanation (wings were built for flight) proves to be false.

THINK[2]: Testing Orthogenesis / Designing Experiments

Imagine that you were asked in 1898 to evaluate Eimer’s theory. What kinds of experiments or observations might you think of as a way to test some of his claims? (Imagine that you both had access to the fossil record and the ability to breed whatever organisms you wanted in whatever kinds of conditions you could think of.)

If Eimer is right, and evolution occasionally gets stuck in certain kinds of patterns, this process should be accessible to experimental testing. We could, for example, try to find an organism that was already showing some linear trend in its development, and see if we could get it to vary away from that path. If evolution is really “stuck” as a result of orthogenesis, all our efforts to move the organism from that trend should fail. We also could think about trying to create the conditions for orthogenesis and seeing if it began to occur in another organism – for instance, isolating all of the organisms that express the early stages of some kind of linear trend, and seeing if breeding them together would create a population in which that trend would continue.

THINK[3]: The Creativity of Selection

We often talk about natural selection as *creating* adaptations. If we want to understand Eimer’s objection that it cannot, we should first think about what this means. What do you think we mean when we say this? How do natural selection

and variation work together to build new features of organisms?

Evolutionary theorists and philosophers even today struggle to explain the precise sense in which natural selection “creates” adaptations. Darwin certainly thought that it did. He believed that, in essence, natural selection would *always* have enough variations upon which to act that any kind of adaptation that was selectively favored would arise in a population – selection would thus “create” adaptations. Several Darwinians after Darwin, on the other hand, restricted selection’s action, arguing that selection was only useful for *eliminating* variations that appear and are harmful to organisms. This would mean that a theory of how variation arises (which Darwin, as we have already seen in the second course, did not have) would become just as important as selection itself, if not more so. This debate has continued right down to the present, with the two sides remaining those arguing for the creativity of selection versus the creativity of mutation.

THINK[4]: Uses for Antlers / Alternative Explanations

What purposes can you think of for which an elk might have used these giant antlers? How other than natural selection do you think they might have come about? How would you compare the various explanations that you’ve thought of?

Again, here, the goal is to get the students to anticipate a bit the work of the scientists which we’ll unpack in a moment. The goal should be to encourage thinking about how complex the various explanations for these antlers might be – this is really a non-obvious case to explain. The natural instinct will probably be to see them as useful for attack or defense, but then their size and weight are problematic. Other alternatives are found in the text that follows.

THINK[5]: Sexual Selection / Gender Roles

Darwin’s way of thinking about sexual selection has been criticized for just recapitulating 19th-century gender roles, with angry, violent males that fight over the coy, choosy females. Is this reasonable, or just the illegitimate copying of social values into the scientific realm? Would there be a way to have a theory of sexual selection that didn’t have these problems, or not? And more generally, how can we be on the alert that we aren’t just interpreting our scientific data in the light of what seems reasonable or acceptable in our society?

This, too, is a debate that continues right down to the present day. Darwin himself only offered examples of sexual selection that fairly tightly conformed to Victorian gender roles. One could imagine the possibility of other examples that didn’t – for instance, where males were selected for demonstrating abilities that might correlate with child-rearing – but these have, in point of fact, been fairly rare. There has thus been a call by some evolutionary biologists to stop using theories of sexual selection entirely, as they are simply too tied to their gender-discriminatory roots to be scientifically useful.

At the more general level, it’s worthwhile to consider the relationship between our scientific theories and our social and cultural values. Since scientific theories are made

by humans, living in particular cultures and times, it is unreasonable to demand that they be entirely free of connections to the values that produced them. But we should scrutinize those value-connections, and try to understand how they could lead our interpretations of data or the scientific knowledge we produce astray.

THINK[6]: Critiquing Correlation of Growth

Can you think of problems with this explanation? If you were one of Darwin's opponents, how could you criticize this?

While there is nothing necessarily problematic about a correlation-of-growth explanation for the antlers of the elk, there are a whole host of reasons that a motivated objector in the nineteenth century would have been able to disagree – here are a few. One might claim that citing the “correlation of growth” is only a way to *describe* the phenomenon at issue, not to *explain* it. Without some kind of understanding of why the growth of these parts might be correlated, we have no more real knowledge of the process than we did when we started. A correlation-based explanation also requires that we have an account of the selective process that is driving it. In this case, the entire basis of the correlational account is derived from the plausibility of selective pressure in favor of larger general body size. If we aren't convinced by the latter, the former never gets off the ground.

Put most succinctly, Darwin's opponents would have been legitimately able to demand that he offer a more detailed account of the process that led to the increase in antler size – this was by no means a straightforward explanation of the phenomenon at issue.

THINK[7]: Evidential Standards

How much evidence, and of what kind, should we have in order to support a scientific theory? Both Eimer's orthogenesis and Darwin's correlation of growth appeal to processes that we don't fully understand, and both believe that they have some empirical evidence to support those processes. How might you argue in favor of the correlation of growth, or orthogenesis, with only the *effects* of that process as your data?

The question of evidential standards is present throughout these readings, not least because scientists found themselves without much useful evidence to resolve these sorts of arguments about early evolutionary theory. As was briefly considered in the last question, orthogenesis and the correlation of growth in fact both appeal to a process whose details we do not fully understand. Both theories argue that they can find examples of that process operating in nature, and also that they can offer a vague outline of how it might work. For the correlation of growth, this relies on factors during an organism's development, while for orthogenesis it involves the generation of variations.

Arguments for either process will necessarily be rather nuanced. Advocates of Darwin's correlation of growth might point to the fact that we can see certain deformities in development which go on to produce correlated problems in other organs. We could

experiment with embryos, for example, to determine how correlation might really work in particular cases. Orthogenesis advocates could note that their explanation is *simpler* than that offered by the correlation of growth, because one single process is posited which should explain every instance of these kinds of non-adaptive traits, while the correlation of growth needs to be spelled out for every case in which it putatively occurs.

THINK[8]: Simplicity

We often see appeals to the simplicity of a scientific hypothesis, also called “Ockham’s razor” after the medieval philosopher William of Ockham (1285–1347). Simpler scientific hypotheses are thus taken to better match the world. Do you think this is a good principle for scientific reasoning? What kinds of problems might it lead to, and how could we avoid them?

While direct appeals to simplicity are somewhat rare in the life sciences, they are extremely common in the physical sciences – where, for instance, general relativity is taken to receive a significant increase in likelihood because of the simple manner in which it explains gravitational phenomena. The kind of simplicity at work here is related. If we don’t need to *add* a new kind of cause to our view of evolution in order to understand how it is that some particular phenomenon arises, then we shouldn’t.

There’s no consensus about whether or not this is actually a good way to reason about scientific theories. On the one hand, simpler theories are certainly *more useful* for us, as it makes it easier for us to comprehend and apply them in particular cases. On the other hand, it seems like there is no necessary reason to think that the world actually is simple. Knowledge claims that are supported by a number of different theoretical paths, each contributing an independent reason to think that the given claim is true, might be better grounded than those that are only derived from one theory, however elegant and simple it might appear.

THINK[9]: Alternative Explanations

Gould is implicitly considering the following three alternatives for how this evolutionary change might have occurred. First, it might be that the antlers were evolving independently of body size – that is, that antlers and size aren’t actually correlated at all. If they are, there are two further possible explanations. It could be that while both body size and antler size were going up, natural selection was responsible for both. Or, lastly, it could be that natural selection was *only* responsible for the increase in body size, and the antler size increase was entirely “by accident.”

What kind of data would you collect in order to tell the difference between these three hypotheses? How would you know that you had proven one and disproven the others?

The logical structure of Gould’s argument here is a bit complex, and hence worth unpacking for the students. He notes that we didn’t actually have any data confirming the correlation between body size and antler size in the first place – thus the

entire question of allometry might have been moot from the very beginning, if this correlation failed to hold.

If it does hold, then we still haven't established what is actually being altered by natural selection. It's possible that only the body size is being increased. In this case, the explanation for the antler increase really would be the correlation of growth; antlers grew as a side-effect of the increase in body size. But in order to claim that, we have to rule out a third possibility, that neither Darwin nor Eimer ever really considered – that natural selection was in fact increasing *both* body size *and* antler size at the same time.

To test these hypotheses, at the very least you would need to gather the data about the correlation between antler and body size, in as many fossil specimens as you could. If the correlation was established, then, you would need to turn to evaluating different selective hypotheses for why larger body size, antler size, or both might be encouraged by natural selection. What data would be required for each of these would differ depending upon the hypothesis that was to be tested. A sexual-selection explanation, for instance, is *extremely* hard to test using the fossil record – how can we know whether or not female Irish elk found large antlers attractive? Correlation of growth might be tested by seeing how antlers develop in juvenile specimens, if we have any, or by comparing fossil specimens to the growth and development of contemporary elk or deer.

THINK[10]: Scientific Theory and Policy

What do you think is the importance of scientific theory for social or governmental policy? That is, what kinds of discoveries in science might be important for the ways in which we structure our social life? How could science be abused when it makes the jump to policy? What responsibilities do scientists have if they think their work might be used in this way?

The relationship between Darwin's actual theory of evolution and social Darwinism is complex enough to merit multiple books of explanation. Rather than focusing on that here, this question pushes the broader consequences of using science for building social policy, and the moral responsibility on the part of scientists that this might entail. Some still argue – though this perspective is increasingly rare – that scientists in fact have no such social responsibility. Scientific knowledge is morally neutral, they claim, and the uses to which it is put are the responsibility of politicians. This, however, is a fairly dubious claim. There is no obvious reason that scientists – unlike every other member of society – should somehow not be morally responsible for what they produce. From cases as apparently benign as sociological or demographic research (e.g., “algorithmic” predictions of crime that turn out to only be predictions of race) to those as extreme as the production of nuclear weapons, the fact that certain kinds of knowledge are produced *changes* the moral and policy landscape in ways that scientists are responsible for considering when they engage in research. (While it would be a bit far afield for my discussion here, this is also clearly evident in climate change research.)

How to correctly deal with these responsibilities, on the other hand, is a more complex question. Just as no one in everyday life is *free* from responsibility for their actions, neither are people held responsible for *all of* the consequences of their actions, however remote. In many cases, these consequences involve social and political decisions, of the sort that should probably *not* be made by one individual person – they require public, democratic deliberation of a sort with which scientists are not often familiar. How to engage the public and policy-makers in these kinds of debates is one of the major challenges facing science in the twenty-first century.

THINK: NOS Reflection Questions

Explicitly reflecting on these themes serves partly to help the students in reviewing the material that we have just covered. But research also indicates that if students do not *explicitly* consider the impact of what they've read on NOS issues, they will often fail to see these links. It is therefore crucial that the lesson closes with time for this detailed reflection. The following NOS themes are listed, with links back to earlier THINK questions:

- alternative explanations (THINK 1, 3, 4, 6, 9)
- evidential relevance (empiricism) (THINK 2, 7, 8)
- social responsibility of scientists (THINK 5, 10)
- conceptual change (THINK 3, 8)
- role of gender bias (THINK 5)
- uncertainty (THINK 6, 7)
- role of systematic study (vs. anecdote) (THINK 2, 7)

5. W. F. R. Weldon, Genes, and Traits

THINK[1]: Background Knowledge on Mendel

What do you already know about Mendel and his work? How was he presented in your textbook?

It's almost certain that the canned history of evolution in at least some of your students' textbooks would have presented a story including the idea that Darwin initially described natural selection, and Mendel's experiments added the knowledge to Darwin's view necessary to produce contemporary genetics. Elaborating this background level of student knowledge is important to situate the remainder of this lesson's discussions.

THINK[2]: Genetics without DNA

It is important to remember that, when Mendel's work was first published and evaluated, we had not yet discovered the fundamentals of cellular biology, biochemistry, or DNA. What kind of data would you be able to accumulate, then, and how could you use it to fashion scientific theory? What would have been the limits of the kind of theories you could construct on this basis?

Again important for altering student expectations of the science of the late-nineteenth

century, it's worth spending some time considering what it would be like to explore genetics without understanding the fundamental biochemical nature of DNA. By the 1890s, it was fairly clear that there was some kind of physical basis for inherited traits in cells, and it was also becoming increasingly likely that this physical basis was connected in some way to chromosomes. But the entirety of the first half-century of genetic research was performed without knowing anything about nucleic acids, and in particular without understanding the metaphor of a piece of DNA "coding for" the production of proteins.

Understanding these processes of inheritance would thus have required that we make inferences only from phenotypic data – simply looking at how many organisms had a given trait in a particular generation, and how many organisms had that same trait in the next generation. We might do this at the population level (generating frequency distributions of traits, as we'll see Weldon do later), or we could track the offspring of individual parents (as Mendel did with his pea plants). Each of these would have their own limitations and would only be able to support certain kinds of theoretical conclusions.

THINK[3]: Mathematics in Biology

Darwin had introduced his entire theory of evolution without, essentially, any math at all. How do you think the addition of mathematical methods to biology could have been seen by some biologists as a welcome addition to biological practice, and by others as a way in which to add all kinds of unnecessary complexity and a bad idea?

Mathematical techniques are often a double-edged sword. On the one hand, they can offer significant increases in precision and clarity in areas where this can be badly needed. Thinking about change in evolving populations as a statistical question was an extremely important advance that made possible significant developments in modern genetics. On the other hand, that clarity comes at a cost of higher barriers to entry and higher complexity. Many students are attracted to the life sciences precisely because they take themselves (whether this is true or not is another question) to be "bad at math." The same was true in the nineteenth century. A number of biologists who had made their living for decades engaging in qualitative study (for instance, of the morphological characteristics of organisms) saw no need to build competence in radically different and newly developed mathematical tools to theorize about evolution.

THINK[4]: Objectivity / Persuasion

Including this color photograph in his journal article was both difficult and expensive at the time. Weldon was thus clearly attempting to persuade his readers in a particular kind of way. How might a photograph have been more persuasive than just a description? What might Weldon have been trying to show his readers about himself or his scientific process by including it?

It was precisely during the late nineteenth century that, as has been argued by the

historians of science Lorraine Daston and Peter Galison, a new concept of “objectivity” began to take over in the sciences. If we could take photographs of scientific phenomena, then we could remove the observer entirely from scientific knowledge generation – a sort of “self-discipline” that would, people increasingly thought, lead to the construction of science that was more “true to nature.”

Weldon thus was working at the forefront of techniques of scientific persuasion, trying to show his readers that we should believe that Weldon hasn’t doctored or cherry-picked this data. (He also offered to send collections of exactly these peas for free to anyone who wanted them, until his supply ran out.) If students guess that, in part, he wanted to do this because he expected criticism, they are exactly right – while the lesson doesn’t have the space to go into it, Weldon was embroiled in a bitter controversy over the right way to incorporate results like Mendel’s into evolutionary theory.

THINK[5]: Variable Data / Uncertainty

If the data here are as uncertain and variable as Weldon says they are, we have an interesting challenge that we have to solve: how do we do science, in spite of these problems with our data? How can we avoid just interpreting the data we have in terms of our prior beliefs (confirmation bias)? What kinds of techniques can you think of that could help us overcome this problem?

There are a number of ways that we might decide to deal with the data shown by these peas, assuming for the sake of argument that Weldon has interpreted them correctly. We might need to engage in better kinds of classification – perhaps these are in fact multiple different types of peas that have been lumped together, despite the fact that Weldon has purchased them from agricultural suppliers who swear that they are a single “variety.” Perhaps they have been grown in dissimilar conditions, such that if we improved our experimental technique we would be able to improve the quality of our data.

Those methods all assume that we could eliminate the uncertainty if we were dedicated enough in trying to do so. If this were in fact false, we’d have a different set of problems. Data analysis methods (especially contemporary uses of statistics, which wouldn’t have been available at the time) could help. But we would constantly be under threat of reinterpreting our data in accordance with what we hoped we would find there. While the lesson lacks space to go into this question, there is even some evidence that Mendel himself might have faked, or at least charitably interpreted, some of his data – the famed geneticist R.A. Fisher would later argue that the correspondence between Mendel’s data and the $\frac{3}{4}$ ratio for the offspring of hybrids was, statistically speaking, too good to be true.

THINK[6]: Reversion / Alternative Explanations

What kinds of possible explanations can you think of for how this phenomenon might happen?

There are many ways that we might think about reversions taking place. Darwin and a

number of his immediate successors in the evolutionary biology community thought that this had to mean that particular hereditary elements must have been conserved for a long time from ancestors to descendants – that is, the elements that make a pigeon look like a street pigeon were still there for all those intervening generations in the fancy pigeons, just somehow in a “deactivated” state. Some argued that the data supporting these reversions was in fact bad – how could you be sure that there hadn’t been an errant cross-breeding between a wild-type bird and your fancy birds? And as early genetics began to take hold, the idea became current that perhaps reversions were indicative of a hybridization, not of one or two characters as in Mendel’s peas, but of dozens, all happening to realign in precisely the way that they did in a wild-type pigeon.

THINK[7]: Continuous and Discontinuous Characters

If you believed that most characters were actually transmitted like those in Mendel’s peas, what other kind of explanation could you give for those characters like height, which seem to be distributed in a very different way? What sorts of evidence would we need to decide between Weldon’s explanation and yours?

As normally presented, Mendel’s way of understanding the transmission of characters would only be applicable for discontinuous characters like the color of peas or human eye color. How, then, could the transmission of individual alleles lead to continuous characters like height?

When Mendel’s theory was initially elaborated by the early geneticists, this problem was often ignored; it was thought that if we only better understood the way in which height was inherited, we might discover that in fact it’s not continuously variable after all, but is indeed transmitted as alleles for some small number of genes. This view was quickly abandoned, however, as this new interpretation never panned out. An important early theoretical breakthrough occurred when scientists realized that a large number of binary characters could produce an extremely large number of values for a possible character outcome, and thus a large number of “traditional” Mendelian characters could produce the appearance of continuous variation. If we imagine that height were controlled by twelve on/off genes, that would give us $2^{12} = 4096$ possible outcomes for height, which would be practically indistinguishable from continuous variation with slight error. Examples of such characters were soon found in a number of plants.

THINK[8]: Scientific Realism

One of the questions at play in Weldon’s response has to do with the purpose of scientific theory. Why do we build scientific theories in the first place? Are we trying to use them just to make predictions about the world around us, or are our scientific theories supposed to tell us what the world is made out of and how those parts work together? Do you think one of those gives us “better” science? Which one?

This question touches on a classic problem in the philosophy of science: the debate over *scientific realism*. According to the traditional, realist view, the goal of science is to create *exactly true* theories about the world. Of course, we know that we won't succeed (for we know that our current theories are probably false, even if they're better than all the theories that have gone before), but we will at least steadily be increasing in *approximation* to the truth. When science tells us, for instance, that atoms are made of electrons, protons, and neutrons, we should take that claim seriously, even if the only way we have to "see" those particles is by the aid of machines that themselves are part and parcel of that same scientific theory.

Anti-realists, on the other hand, argue that what science is really for is the prediction and control of the world around us – we want to be able to build airplanes and cell phones, and science is useful only insofar as it helps us do that kind of thing. When science says that "electrons really exist," it's fine to act as though that's true, but we don't have any reason to *really* believe that, because that's just not what science is *about*. If those particles weren't *really* there, and we still could use the theory as if they were, it just wouldn't matter.

Weldon is trying to push a kind of realist claim against the Mendelians. According to Weldon, while Mendel's exact, clean, 3-to-1 ratios might be useful for certain kinds of prediction or control (like the breeding of peas), and hence satisfy a minimal, anti-realist criterion, they aren't *really* telling us what's going on under the surface, nor offering a realist explanation of the way that traits are actually transmitted.

THINK[9]: "Genes for X" in the Media

Have you ever seen examples like this from the media in your own experience? For instance, say that someone claimed to have found the "gene for obesity?" How might you critically evaluate that kind of a claim? What sorts of further evidence do you think we'd need to know whether or not it was really true?

Students will likely be able to come up with any number of examples of this kind of talk from their own experience, or even from their own textbooks – "genes for" talk is unfortunately relatively ubiquitous. Critical evaluation of such claims is all about obtaining other kinds of data. Genes may only be activated in certain kinds of environments, and the data gathered on them might coincidentally only have been gathered within a single environment. (This is especially problematic in humans, who tend to live in environments that are in some senses extremely diverse while in others extremely similar.)

Further, these kinds of assertions are sometimes only supported by a statistical correlation, without any kind of knowledge of how the supposed effect might actually be the responsibility of the gene that was identified. Further molecular research could, for instance, help us understand whether this is a genuine correlation – in the sense that we can understand why the gene's impact would contribute to the character of interest.

THINK[10]: Genes and Environment

Imagine that we had two competing explanations for a single phenomenon – one genetic and one environmental. How might we compare and contrast them? What would you want to know to be able to choose between them?

The answers to this question are related to those for the last one, though the goal here is rather to push students to consider how they themselves might gather evidence about both the genetic and the environmental causes for a particular character, if we believed that each were present. For the case of the environment, we would need to search for correlations between environments and individual outcomes, and then try to understand how those environmental influences might have led to the development of the character involved. See the last question for more details about how we might do this in the case of genetic causes.

THINK: NOS Reflection Questions

Explicitly reflecting on these themes serves partly to help the students in reviewing the material that we have just covered. But research also indicates that if students do not *explicitly* consider the impact of what they've read on NOS issues, they will often fail to see these links. It is therefore crucial that the lesson closes with time for this detailed reflection. The following NOS themes are listed, with links back to earlier THINK questions:

- robustness (agreement among many different types of data) (THINK 2, 4, 7, 10)
- confirmation bias (THINK 5, 9)
- interdisciplinary thinking (THINK 3)
- new techniques and their validation (THINK 3)
- forms of persuasion (THINK 4, 8, 10)
- error and uncertainty (THINK 4, 5)
- credibility of news media (THINK 9, 10)

Acknowledgments

Many thanks to [John S. Wilkins](#) for help fact-checking the historical claims throughout the lessons. I owe much of the general structure of Lesson 5 to Greg Radick's approach to the Weldon-Mendel relationship, though I don't presume that he would agree with everything in my presentation there. Some of the material here dates from discussions that I had about these questions with my colleagues and friends [Greg Macklem](#) and [Erik Peterson](#), building on work that we'd done for a presentation at the [NSTA's 2017 New Orleans meeting](#). Finally, I owe some of my perspective on these issues to my years teaching in the [LSU GeauxTeach program in the College of Science](#).

This material was prepared in part with funding from the US National Science Foundation, under HPS Scholars Award #1826784.