

Chapter 2

SCIENCE AWRY

If you google “How old is the Earth?” the first answer you will be offered is 4.543 billion years old. This is the accepted scientific value based on radiometric dating of asteroids and lunar materials. You will find it affirmed if you visit the web page of NASA, the US Geological Survey, or the Encyclopedia Britannica. It has been more or less in place for half a century and most educated Americans accept it as factual. It is what any mainstream earth science professor or teacher would teach, and what you will find in any college earth science textbook. However, if you scroll down on your computer you will also find:

How old is the earth?—creation.com

*creation.com/how-**old-is-the-earth***

Creation Ministries International

The answer offered by Creation Ministries International, based upon biblical exegesis, is about six thousand years. If we were to judge a knowledge claim by the antiquity of its provenance, we would have to judge this claim to be more stable than the accepted scientific one, because it has been around since the mid-seventeenth century. Similarly, if we were to define authority as the ability to drive out competing claims, then the authority of the scientific value is clearly by no means total. This is not merely the case for the age of the Earth. If we look for answers about climate change, the safety of vaccinations, whether plate tectonics is an accurate and adequate theory of global

tectonics, and if drinking water fluoridation prevents cavities, we will find many claims competing for our attention.

Some of these claims are simply unscientific—which is to say not based on vetted evidence—while others have been shown by evidence to be false. Yet they persist. Indeed, the fragile status of facts—both scientific and social—is now so widely acknowledged that the Oxford English Dictionaries declared the 2016 word of the year to be “post-truth.”¹ Comedian Stephen Colbert complained that this was a rip-off of his earlier neologism “truthiness.”²

The inclination of some religious believers to distrust scientific findings is neither new nor unstudied. Scholars have amply described and attempted to account for religiously motivated dissent from scientific theories of evolution from Darwin to Dawkins. But rejection of scientific claims is not restricted to matters of theological concern; people reject scientific conclusions for a host of reasons. Clearly, the establishment of scientific claims qua science does not entail the acceptance of those claims by people outside the scientific community. On the contrary, a “post-truth” world is one in which the fundamental assumptions of scientific inquiry—including its capacity to yield objective, trustworthy knowledge—have been called into question.

Some scholars, most notably Bruno Latour and Sheila Jasanoff, have argued that scientific knowledge is co-produced by scientists and society, in which case truthiness might be viewed as a normal state of affairs.³ A co-produced claim, in their view, is one on which both scientists and society have converged, and it is this convergence—rather than empirical reality or even empirical support—that grants stability to the claim. Until this scientific and social convergence occurs, disputation is inevitable, and not just about values but also about facts.

As an empirical matter, this is clearly so. But the concept of co-production begs the question of what it means for a claim to be scientific and whether factual claims should fairly be understood as distinct from other types of claims. It also begs the question of whether we are justified in rejecting (or at least suspending judgement on) a claim that scientists consider settled when other members of society have demurred. The theory of co-production begs the question of whether scientific claims made by scientific experts merit trust.⁴

Latour has argued that scientific claims are *performances* about the natural world, and that scientists have been successful at “performing the world we live in.”⁵ By this he (presumably) means that scientists have achieved substantial social authority and are broadly accepted as our leading societal experts on “matters of fact.”⁶ (They perform and we applaud.) He also suggests (presumably ruefully) that natural scientists are “better equipped at performing the world we live in than [social scientists] have been at deconstructing it.”⁷ But he may be overestimating the success (performative or otherwise) of the natural sciences, given the large numbers of Americans who doubt many important claims of contemporary science (I restrict myself to the United States here, but similar claims could be made about other countries, such as the HIV-AIDS link in parts of Africa).

If we define success in terms of cultural authority, the success of science is clearly not only incomplete but at the moment looking rather shaky. Large numbers of our fellow citizens—including the current president and vice president of the United States—doubt and in some cases actively challenge scientific conclusions about vaccines, evolution, climate change, and even the harms of tobacco. These challenges cannot be dismissed as “scientific illiteracy.” Studies show that in the United States, among Democrats and independent voters, higher levels of

education are correlated with higher levels of acceptance of scientific claims, but among Republicans the opposite is true: The more educated Republicans are, the more likely they are to doubt or reject scientific claims about anthropogenic climate change. This indicates not a lack of knowledge but the effects of ideological motivation, interpreted self-interest, and the power of competing beliefs.⁸

And, as we saw in chapter 1, there is a deeper problem, one that transcends our particular political moment and varying cultural conditions. Even if we accept contemporary scientific claims as true or likely to be true, history demonstrates that the process of transformative interrogation will sometimes lead to the overturning of well-established claims. William James argued more than a century ago that experience has a “way of boiling over, and making us correct our present formulas.”⁹ He astutely pointed out that what we label as “‘absolutely’ true, meaning what no further experience will ever alter, is that ideal vanishing point toward which we imagine that all our temporary truths will someday converge . . . Meanwhile we live today by what truth we can get today, and be ready tomorrow to call it falsehood.”¹⁰ This was also Karl Popper’s point where he argued for the provisional character of all scientific knowledge.

The overturning of claims is not arbitrary; it is related to experience and observation. But why we should accept any contemporary claim if we know that it may in the future be overturned? One might point out that incomplete and even inaccurate knowledge may still be useful and reliable for certain purposes: the Ptolemaic system of astronomy was used to make accurate predictions of eclipses, and airplanes were flying before aeronautical engineers had an accurate theory of lift.¹¹ That scientific knowledge may be partial or incomplete—or that old

theories get replaced by new ones—is not ipso facto a refutation of science in general. On the contrary, it may be read as proof of the progress of science, particularly when in hindsight we can look back on the older theories and understand how and why they worked. (Newtonian mechanics still works when the objects under consideration are not moving very quickly.) But if our knowledge is overturned wholesale—if it is deemed in hindsight to have been wholly incorrect—that calls into question whether we can trust current scientific knowledge when we need to make decisions.¹²

Climate skeptics sometimes raise this point. In public lectures on climate science, I have been asked: “Scientists are always getting it wrong, so why should we believe them about climate change?” The “it” that scientists are allegedly getting wrong is rarely specified, and when I ask my interlocutor what he has in mind, usually there is no specific answer. When there is, most often it is the changing and seemingly contradictory recommendations of nutritionists. There are many reasons why nutritional information in recent years has been a moving target, and why nutrition seems to be a dismal science. These include the role of the mass media in publicizing novel but unconfirmed findings; the misuse of statistics by ill-trained scientists; the problems of small sample size and the difficulty of undertaking a controlled study of people’s eating habits (see Krosnick, this volume); and the influence of the food industry in funding distracting research on the relative harms of sugar and fat.¹³ (Elsewhere I have written on the potential adverse effects of industry funding of science when the desired outcomes are clear and biasing.¹⁴) But even if nutritional science is atypical, or even if it is typical but the sources of confusion in it can be identified and addressed, the skeptical challenge is epistemologically legitimate.

If scientists sometimes get things wrong—and of course they do—then how do we know they are not wrong now? Can we trust the current state of knowledge?

In this chapter, I set aside the issues of corruption, media misrepresentation, and inadequate statistical training to look at a problem that I think is more vexing, and certainly more challenging epistemically. It is the problem of science gone awry of its own accord. There are numerous examples in the history of science of scientists coming to conclusions that were later overturned, and many of those episodes have to do neither with religious commitments, nor overt political pressures, nor commercial corruption.¹⁵ This has been the central question guiding much of my research career: How are we to evaluate the truth claims of science when we know that these claims may in the future be overturned?

Elsewhere I have called this problem the instability of scientific truth.¹⁶ In the 1980s, philosopher Larry Laudan called it the pessimistic meta-induction of the history of science.¹⁷ He observed (as have many others) that the history of science offers many examples of scientific “truths” that were later viewed as misconceptions. Conversely, ideas rejected in the past have sometimes been rescued from their epistemological dustbins, brushed off, polished up, and accepted into the halls of respectable science. The retrieval of continental drift theory and its incorporation into plate tectonics—the topic of my first book—is a case in point.¹⁸ As I wrote in 1999 when discussing that retrieval: “History is littered with the discarded beliefs of yesterday and the present is populated by epistemic resurrections.” Given the perishability of past scientific knowledge, how are we to evaluate the aspirations of contemporary scientific claims to legitimacy and even permanence?¹⁹ For even if some truths of science prove to be permanent, we have no way of

knowing which ones those will be. We simply do not know which of our current truths will stay and which will go.²⁰ How, therefore, can we warrant relying on current knowledge to make decisions, particularly when the issues at stake are socially or politically sensitive, economically consequential, or deeply personal?²¹

In this chapter, I consider some examples in which scientists clearly went astray. The examples are drawn either from my own prior research and that of my students, or from historical examples that I have come to know well through three decades of teaching. Can we learn from these examples? Do they have anything in common? Might they help us answer the question of *ex ante* trust, by helping us to recognize cases where it may be appropriate to be skeptical, to reserve judgment, or to ask with good reason for more research?

I do not claim that these examples are representative, only that they are interesting and informative. They all come from the late nineteenth century onwards, because in my experience many scientists discount anything older on the grounds that we are smarter now, have better tools, or subject our claims to more comprehensive peer review.²² Of course, no two historical cases are the same. Each of the examples I will present is complex, with more than one possible interpretation of how and why scientists took the positions they did. These cases do not define a “set.” But they do have one crucial element in common: each of them includes red flags that were evident at the time.

Example 1: The Limited Energy Theory

In 1873, Edward H. Clarke (1820–77), an American physician and Harvard Medical School professor, argued against the higher education of women on the grounds that it would adversely affect their fertility.²³ Specifically, he argued that the demands of higher education would cause their ovaries and uteri to shrink. In the words of Victorian scholars Elaine and English Showalter, “Higher education,” Clarke believed, “was destroying the reproductive functions of American women by overworking them at a critical time in their physiological development.”²⁴

Clarke presented his conclusion as a hypothetic-deductive consequence of the theory of thermodynamics, specifically the first law: conservation of energy. Developed in the 1850s particularly by Rudolf Clausius, the first law of thermodynamics states that energy can be transformed or transferred but it cannot be created or destroyed. Therefore, the total amount of energy available in any closed system is constant. It stood to reason, Clarke argued, that activities that directed energy toward one organ or physiological system, such as the brain or nervous system, necessarily diverted it from another, such as the uterus or endocrine system. Clarke labeled his concept “The Limited Energy Theory.”²⁵

Scientists were inspired to consider the implications of thermodynamics in diverse domains, and Clarke’s title might suggest he was applying energy conservation to a range of biological or medical questions.²⁶ But not so. For Clarke, the problem of limited energy was specifically female, i.e., female capacity. In his 1873 book, *Sex in Education; or, a Fair Chance for Girls*, Clarke applied the first law to argue that the body contained a

finite amount of energy and therefore “energy consumed by one organ would be necessarily taken away from another.”²⁷ But his was not a general theory of biology, it was a specific theory of reproduction. Reproduction, he (and others) believed, was unique, an “extraordinary task” requiring a “rapid expenditure of force.”²⁸ The key claim, then, was that energy spent on studies would damage women’s reproductive capacities. “A girl cannot spend more than four, or in occasional instances, five hours of force daily upon her studies” without risking damage, and once every four weeks she should have a complete rest from studies of any kind.²⁹ One might suppose that, on this theory, too much time or effort spent on any activity, including perhaps housework or child-rearing, might similarly affect women’s fertility, but Dr. Clarke did not pursue that question. His concern was the potential effects of strenuous higher education.

In 1873, thermodynamics was a relatively new science, and Clarke presented his work as an exciting application of this important development. His book was widely read: *Sex in Education* enjoyed nineteen editions; over twelve thousand copies were printed in the three decades after its release. Historians have credited it with playing a significant role in undermining public support for educational and professional opportunities for women at that time; one contemporary commentator predicted that the book would “nip co-education in the bud.”³⁰

Clarke’s argument was primarily aimed at co-education—that women could not withstand the rigors of a system of higher education designed for men—but it was also used against rigorous intellectual training for women of any sort, particularly that being conceptualized at the women’s colleges that were being founded around that time, such as Smith (founded in 1871), Wellesley (1875), Radcliffe (1879), and Bryn Mawr (1885). Higher education for women was problematic, Clarke and his followers

insisted, unless it was specifically designed to take account of women's "limited energy."³¹ M. Carey Thomas, the first dean and second president of Bryn Mawr College, recalled that in the early years of the college, "we did not know when we began whether women's health could stand the strain of education." Early advocates of higher education for women were "haunted," she reflected, "by the clanging chains of that gloomy little specter, Dr. Edward H. Clarke's *Sex in Education*."³²

Clarke's theory was also linked to emerging eugenic arguments (of which we will shortly say more). Like many elite white men in the late nineteenth and early twentieth centuries, Clarke feared the combination of women abandoning domestic responsibilities and the declining birth rate among native-born white women would be disastrous to the existing social order. He spoke for many when he fearfully predicted that "the race will be propagated from its inferior classes," and exhorted readers to "secure the survival and propagation of the fittest" by keeping women home, uneducated and child-rearing.³³ Perhaps for this reason his work was heralded by many male medical colleagues, who often shared these fears. One of these was Dr. Oliver Wendell Holmes, dean of the Harvard Medical School (and father of the future Supreme Court justice, who later defended the legality of eugenic sterilization in the infamous case of *Buck v. Bell*).³⁴ Holmes publicly expressed his "hearty concurrence with the views of Doctor Clarke."³⁵

Clarke offered seven cases of young women who pursued traditionally male educational or work environments and experienced a variety of disorders, from menstrual pain and headaches to mental illness. His prescription to these women—and therefore to women in general—was to refrain from mental and physical effort, particularly during and after menstruation. Clarke did not attempt to measure or quantify the energy transfer

among the body's organs, nor did he theorize the mechanism by which energy was selectively distributed to some parts of the body rather than others.³⁶ Rather, he asserted that his conclusion was a "deductive consequence from general scientific principles [i.e., the first law] using auxiliary assumptions." In this sense, his approach was similar to others at that time, such as social Darwinists, who also attempted to apply theories developed in the biological domain to problems in social worlds.

In hindsight it does not take much effort to identify the ways in which Clarke embedded prevailing gender prejudice and racial anxiety into his theory. But that risks historical anachronism. If our concern is how to identify problematic science, not in hindsight, but in our own time, then we must ask the question: Did anyone at the time object? The answer is yes. Feminists in the late nineteenth century found Clarke's agenda transparent and his non-empirical methodology ripe for attack. His leading critic within the medical community was Dr. Mary Putnam Jacobi, a professor of medicine at Columbia and the author of over a hundred medical papers.

Jacobi signposted the gender politics inside Clarke's theory, writing that the popularity of his work could be attributed to "many interests besides those of scientific truth. The public cares little about science, except insofar as its conclusions can be made to intervene in behalf of some moral, religious or social controversy."³⁷ She also identified its empirical inadequacy, based as it was on only seven women. As we saw in chapter 1, drawing deductive consequences from theory is part of accepted scientific methodology, but only part: deductive consequences have to be tested by reference to empirical evidence. And Clarke, Jacobi noted, didn't have much.

In 1877 she published a study of her own, *The Question of Rest for Women during Menstruation*, in which she sampled 268

women “who ranged in health, and education and professional status.” (She also allowed the women to self-report their status, in contrast to Clarke who used his own interpretations of their symptoms.) Jacobi presented her data in a series of thirty-four tables examining the relationship between multiple variables, such as rest, exercise, and education.³⁸ She found 59% of women reported no suffering or only slight or occasional suffering from menstruation. Physiologically, she noted that there was “nothing in the nature of menstruation to imply the necessity, or even the desirability, of rest,” particularly when the women’s diets were normal. She supported this conclusion with a thorough literature review on menstruation and nutrition, as well as laboratory experiments on nutrition and the menstrual cycle.³⁹ Her research earned Harvard’s Boylston Medical Prize. But it had little effect on Clarke or his male medical colleagues. In 1907 Dr. G. Stanley Hall wrote in his widely read work *Adolescence*, “it is, to say the very least, not yet proven that higher education of women is not injurious to their health.”⁴⁰ Clarke’s theory was viewed as sufficiently established as to place the burden of proof on those who claimed that higher education for women was fine.⁴¹

Example 2: The Rejection of Continental Drift

In the 1920s and ’30s, American earth scientists rejected a claim that forty years later was accepted as fact.⁴² This was the claim that the continents were not fixed, but moved horizontally across the surface of the Earth; that these movements explained many aspects of geological history; and that the interactions of moving continents explained crucial geological features, such as the

distribution of volcanoes and earthquakes. This concept came to be known as continental drift. Alfred Wegener, the prominent and respected geophysicist who proposed it, compiled a large body of empirical evidence drawn from existing geological literature.

While continental drift was not accepted at the time, there was broad consensus that existing theories were inadequate and an alternative explanation of the facts of geological history was needed. When the reality of drifting continents was accepted in the 1960s, in part based on these facts (as well as new ones that had come to light), some scientists were embarrassed to acknowledge that, not very long before, their community had rejected continental drift. In response, some suggested that continental drift had been rejected in the 1920s for lack of a mechanism to explain it. This was a plausible notion, and it was enshrined in many textbooks and even repeated by some historians and philosophers of science.⁴³ But it wasn't true. Several credible mechanisms had been offered at the time. None of these mechanisms was flawless—newly introduced theories rarely are—but scientists at the time had vigorous discussions about them, and some thought the mechanism issue had been resolved. American geologist Chester Longwell, for example, wrote that a model involving convection currents in the mantle—an idea that in the 1960s would be accepted as part of plate tectonics—was “a beautiful theory” that would be “epoch-making.”⁴⁴

If geologists had plausible mechanisms to explain continental drift, including ones that were later accepted, then why did they reject the theory? A telling element in this story was that American geologists were far more hostile to the theory than their European or British colleagues. Many continental Europeans accepted that pieces of the Earth's crust had moved over substantial horizontal distances; this was evident in the Swiss

Alps. Some British geologists also cautiously entertained the theory; in the 1950s and '60s many British school children learned about continental drift in their O- and A-level geology courses. But this was not the case in the United States: American geologists did not just reject the idea, they accused Wegener of *bad science*. This offers a rare opportunity to explore how scientists decide what constitutes good or bad science.

In debates over the theory, many American geologists explicitly raised methodological objections. In particular, they objected to the fact that Wegener had presented his theory in hypothetico-deductive form, which they considered to be a form of bias. Good science, they held, was inductive. Observation should precede theory and not the other way around. Edward Berry, a paleontologist at Johns Hopkins University, put it this way:

My principal objection to the Wegener hypothesis rests on the author's method. This, in my opinion, is not scientific, but takes the familiar course of an initial idea, a selective search through the literature for corroborative evidence, ignoring most of the facts that are opposed to the idea, and ending in a state of auto-intoxication in which the subjective idea comes to be considered an objective fact.

Bailey Willis, chairman of the geology department at Stanford University and president of the Seismological Society of America, felt the books were “written by an advocate rather than an impartial investigator.” Joseph Singewald, a geology professor at Johns Hopkins University, claimed Wegener “set out to prove the theory . . . rather than to test it.”⁴⁵ Harry Fielding Reid, a founder of modern seismology, argued that the proper method of (all) science was induction. In 1922, he wrote a review of the English translation of Wegener's *Origin of Continents and Oceans* in which

he described continental drift as a member of a species of failed hypotheses based on hypothetico-deductive reasoning.

There have been many attempts to deduce the characteristics of the Earth from a hypothesis, but they have all failed . . . [Continental drift] is another of the same type. . . . Science has developed by the painstaking comparison of observations and through close induction, by taking one short step backward to their cause; not by first guessing at the cause and then deducing the phenomena.⁴⁶

It has sometimes been suggested that comments such as these reflect an American rejection of theory in general. But American geologists did not reject theory *per se*. Many of them were actively involved in theory development in other domains. But they did have particular ideas about how scientific theories should be developed and defended. Scientific theory, they believed, should be developed inductively and defended modestly.

American geologists were suspicious of theoretical systems that claimed universal applicability and of the individuals who expounded them. One example was the “Neptunist” school, developed in the eighteenth century by Abraham Werner, which held that geological strata could be understood as the evolving deposits of a gradually receding universal ocean.⁴⁷ For many American geologists, Neptunism epitomized the type of grandiosity, operating under an authoritarian leader, that Americans discerned throughout European science. On a trip to Europe, Bailey Willis met Pierre Termier, director of the French Cartographic Service, who was known for his theory of *grande nappes*—the concept that large portions of the European Alps could be understood as mega-folds, created when a portion of continental crust was displaced over great lateral distances. Willis

lamented that Termier was “an *authority*,” whose theory young geologists in France “cannot decline to accept.”⁴⁸

The tone with which Willis discussed Termier explains what might otherwise be perplexing in this case: Scientists are supposed to be authorities, but the concern here is that this can slide into arrogance and dogmatism. It can slide into intellectual *authoritarianism*; Termier’s authoritarian status could make it difficult for others to question his theory. The spirit of critical inquiry would be suppressed and scientific progress would be impeded, because no one would feel free to challenge or improve upon the idea.

The American preference for inductive methodology was thus linked by its advocates to American political ideals of pluralism, egalitarianism, open-mindedness, and democracy. They believed that Termier’s approach was *typically* European—that European science, like European culture, tended toward the anti-democratic. American geologists thus explicitly linked their inductive methodology to American democracy and culture, arguing that the inductive method was the appropriate one for America because it refused to grant a privileged position to any theory and therefore any theorist. Deduction was consistent with autocratic European ways of thinking and acting; induction was consistent with democratic American ways of thinking and acting. Their methodological preferences were grounded in their political ideals.

This anti-authoritarian attitude was foregrounded by scientists who propounded the “method of multiple working hypotheses.” Popularized by the University of Chicago geologist Thomas Chrowder Chamberlin, the method was an explicit methodological prescription for geological fieldwork. According to it, the geologist should not go into the field to test a hypothesis, but should first observe, and then begin to formulate

explanations through a “prism of conceptual receptivity that refracted multiple explanatory options.”⁴⁹ That meant developing a set of “working hypotheses” and keeping all of them in mind as work progressed. Chamberlin compared this to being a good parent, who should not allow any one child to become a favorite. A good scientist was fair and equitable to all his working hypotheses, just as a good father loved all his sons. (Chamberlin did not discuss daughters.)

The method was also a useful reminder that in complex geological problems the idea of a single cause was often wrong: many geological phenomena were the result of diverse processes working together. It was not a matter of either/or but rather both/and; the method of multiple working hypotheses helped geologists to keep this in mind. Chamberlin thought that the bitter divisiveness that had characterized many debates in nineteenth-century geology had arisen because one side had fixed on one cause and the other side on another, rather than accepting that the right answer might be a bit of both.⁵⁰ Scientists should be investigators, not advocates. Chamberlin encapsulated this idea in a paper called *Investigation vs. Propagandism*.⁵¹

At the University of Chicago, Chamberlin designed the graduate curriculum in geology specifically to train students to be “individual and independent, not [merely] following of previous lines of thought ending in a predetermined result”—his gloss of the European method. He also warned against the British system of empiricism, which he believed was “not the proper control and utilization of theoretical effort but its suppression.”⁵² (Chamberlin was thinking specifically of Charles Lyell’s denunciations of high theory.) The method of multiple working hypotheses was the via media between dogmatic theory and empiricist extremism that would help in the future to avoid divisive battles and factionalism.

One might wonder if this was just talk, but geologists' field notebooks and classroom notes from the period show that the method of multiple working hypotheses was practiced. Observations were segregated from interpretation, and geologists frequently listed various possible interpretations that occurred to them. One example is Harvard geologist Reginald Daly, an early advocate of continental drift. His field notebooks show how he enacted Chamberlin's prescription: in these notebooks he would record his observations on the left side of his notebook and, on the facing page, list a variety of possible interpretations of them. Reading Daly's field notes, one is reminded of Richard Hofstadter's famous claim that in the United States "a preference for hard work [was considered] better and more practical than commitments to broad and divisive abstractions."⁵³ It was not that American scientists were opposed to abstraction; it was that they were seeking a nondivisive approach to it. In the 1940s when Harvard professor Marland Billings taught global tectonics, he offered his students for their consideration no less than nineteen different theories of mountain-making, declining to say in class which one he preferred.⁵⁴ In this context, we can understand why American geologists reacted negatively to Wegener's work: He presented continental drift as a grand, unifying theory with the available evidence taken as confirmatory. For Americans, this was bad scientific method. It was deductive, it was authoritative, and it violated the principle of multiple working hypotheses. It was exactly what they expected from a European who *wanted* to be an authority.⁵⁵

Americans, however, had become dogmatic in their anti-dogmatism, because in rejecting Wegener's theory on *methodological* grounds, they dismissed a *substantial body of evidence that in other contexts they accepted as correct*. Many of Wegener's harshest critics acknowledged this point, as when Yale geologist

Charles Schuchert allowed that the super continent of Gondwana “was a fact” that he “had to get rid of.”⁵⁶ (Schuchert’s solution was the ad hoc theory of “land bridges” to account for the paleontological evidence, but which failed to explain the correspondences in stratigraphy, which others sedulously analyzed.) In later years, geologists would acknowledge that the evidence that Wegener had marshalled was substantively correct.

Example 3: Eugenics

The history of eugenics is far more complex than the two examples we have just examined, in part because it involved a wide range of participants, many of whom were not scientists (including US president Teddy Roosevelt), and the values and motivations that informed it were extremely diverse. Perhaps for this reason some historians have been reluctant to draw conclusions from what nearly all agree is a troubling chapter in the history of science. But it has been used explicitly by climate change deniers to claim that because scientists were once wrong about eugenics, they may be wrong now about climate change.⁵⁷ For this reason, I think the subject cannot be ignored, and because of its complexity I grant it more space than the two examples we have just considered.

As is widely known, many scientists in the early twentieth century believed that genes controlled a wide range of phenotypic traits, including a long list of undesirable or questionable behaviors and afflictions, including prostitution, alcoholism, unemployment, mental illness, “feeble-mindedness,” shiftlessness, the tendency toward criminality, and even *thalassophilia* (love of the sea) as indicated by the tendency to join the US Navy or Merchant Marine. This viewpoint was the basis for the social

movement *eugenics*: a variety of social practices intended to improve the quality of the American (or English, German, Scandinavian, or New Zealand) people, practices that in hindsight most of us view with dismay, outrage, even horror. These practices were discussed either under the affirmative rubrics of “race betterment” and “improvement,” or the negative rubrics of preventing “racial degeneration” and “race suicide.”⁵⁸ The ultimate expression of these views in Nazi Germany is well known. Less well known is that in the United States, eugenic practices included the forced sterilization of tens of thousands of US citizens (and principally targeting the disabled), a practice upheld in the *Buck v. Bell* decision, wherein Supreme Court justice Oliver Wendell Holmes, Jr., upheld the rights of states to “protect” themselves from “vicious protoplasm.”⁵⁹

The plaintiff in *Buck v. Bell* was a young woman, Carrie Buck, who had been sterilized after giving birth after being raped. State experts in Virginia testified that Carrie, her mother, and her child were all “feeble-minded”; this was used to warrant Carrie’s sterilization to ensure that no further offspring would be produced. Justice Holmes encapsulated the decision in his memorable conclusion: “Three generations of imbeciles are enough.”⁶⁰ Eugenic sterilization laws in the United States helped to inspire comparable laws in Nazi Germany, used to sterilize mentally ill patients and others deemed to be a threat to German blood; after World War II eugenics was largely discredited because of its relation to Nazi ideology and practices.⁶¹

We might be tempted to dismiss eugenics as a political misuse of science, insofar as it was promoted and applied by people who were not scientists, such as President Roosevelt or Adolf Hitler, or by men who worked in eugenics but were not trained in genetics, such as the superintendent of the Eugenics Record Office, Harry Laughlin, who testified in the US Congress on

behalf of eugenic-based immigration restrictions.⁶² But that only gets us so far, insofar as eugenics was developed and promoted to a significant extent by biologists, and by researchers who came to be known as “eugenicists.” Moreover, like Clarke’s Limited Energy Theory, eugenics was presented as a deduction from accepted theory, in this case Charles Darwin’s theory of evolution by natural selection. If, as Darwin argued, traits were passed down from parent to offspring, and fitness was increased by the differential reproduction and survival of fit individuals, then it stood to reason that the human race could be improved through conscious selection. Darwin had developed his theory of natural selection in part by observing selective breeding by pigeon fanciers: breeding was the deliberate and conscious selection of individuals with desirable traits to reproduce and the culling of those with undesirable ones. If breeders improved their pigeons, dogs, cattle, and sheep through selection, was it not obvious that the same should be done for humans? Should we not pay at least as much attention to the quality of our human offspring as of our sheep? And was it therefore not equally obvious that society should take steps to encourage the fit to reproduce and discourage the unfit? This latter question was famously posed by Thomas Malthus in the eighteenth century, who argued against forms of charity that might encourage the poor to have more children, and who, through his arguments about the inexorable mathematics of reproduction, inspired Darwin.⁶³

The founder of “scientific” eugenics is generally taken to be Darwin’s cousin Francis Galton (1822–1911) and many elements in Darwin’s work seemed to support the view that the laws of selection that operate in nature must also operate in human society. In *The Descent of Man* (1871), for example, Darwin made clear that he believed that natural selection applied to men as well as beasts, and he argued that some human social practices, such

as primogeniture, were maladaptive. It was not a stretch to read Darwin as suggesting that human laws and social practice should be adjusted to account for natural law.

For humans, Galton argued, the most important trait was intelligence, and so Galton undertook the study of intelligence and heredity. Many physical traits, such as height, hair, skin and eye color, and even overall appearance, seemed to be largely inherited, but was intelligence? In his 1892 work *Hereditary Genius*, Galton concluded that it was. Analyzing the family trees of “distinguished men” of Europe, he found that a disproportionate number came from wealthy or otherwise notable families. While he recognized that “distinction” was not the same as “intelligence,” Galton used it as a proxy. Finding that distinctions of all types—political, economic, artistic—did cluster, he concluded that character traits ran in families just as physical ones did.⁶⁴

Galton, however, observed one crucial difficulty, what he called the *law of reversion to the mediocre*: that the offspring of distinguished parents tended to be more mediocre—that is to say, more average—than the parents. He illustrated this with height: tall couples gave birth, on average, to children who were not as tall as they. In an early insight into what we would now call population genetics, Galton reasoned that the children were inheriting traits not only from their parents, but also their grandparents and great-grandparents, i.e., their entire family tree. The same would be true of any trait, including intelligence.

Galton’s conclusions regarding the prospects for overall improvement of the human race were thus pessimistic, because if offspring inherited from their entire family tree then improvements would take many generations to achieve. Pigeon fanciers and dog breeders did not achieve their results in a single generation, but by patient selection over years and decades, and for human breeding this was unrealistic. Galton did suggest vaguely

that “steps” should be taken to encourage the “best” to procreate more and the “worst” to procreate less so as “to improve the racial qualities of future generations” and avoid “racial degeneration.”⁶⁵ Such non-coercive encouragement came to be known as “positive eugenics.” But Galton was not optimistic that a program of eugenics could be readily or reasonably achieved. The law of reversion to the mediocre seemed to undermine such aspirations. Others, however, insisted not only that human improvement through breeding could be achieved, but that it needed to be.

Today, the idea of “racial degeneration” is impossible to separate from its Nazi associations, but in the early twentieth century the threat was keenly felt—at least by many white men—as real and present, and the eugenic ideal was taken up by physicians, scientists, intellectuals, and political leaders. In the United States, besides Teddy Roosevelt, another prominent eugenicist was the conservationist Madison Grant, a founder of the Save-the-Redwoods League, trustee of the American Museum of Natural History, and author of the popular book *The Passing of the Great Race* (1916).⁶⁶ This was the Nordic “race”—what we might now call white Anglo-Saxons—which Grant believed was threatened by the weaker “races” of Jews, southern Europeans, and Negroes. These latter groups, he argued, should be isolated in ghettos and prevented from interbreeding with men and women of northern European descent. Grant’s arguments played a role in the Johnson-Reed Immigration Restriction Act of 1924, which limited immigration from southern and eastern Europe to no more than 2% of the US population as measured in the 1890 census and completely eliminated immigration from Asia.⁶⁷ Stephen Jay Gould characterized *The Passing of the Great Race* as the most influential work of scientific racism ever published in America; historian Jonathan Spiro notes that

it was widely embraced in Nazi Germany, including by Hitler, who wrote to Grant saying, “The book is my Bible.”⁶⁸

Eugenics as a social movement grew dramatically in the years 1910–20, with a proliferation of books and articles on race and fitness, nearly all of which were framed as applications of biological science. As Grant crisply put it, “The laws of nature require the obliteration of the unfit.”⁶⁹ The rediscovery in 1900 by Hugo de Vries and colleagues of the work of Gregor Mendel, and the support it seemed to give—indeed, the proof, in some eyes—for the hard inheritance of characteristics, was important to the surge of support for eugenics, as Mendel’s findings seemed to rule out Lamarckian notions that individual improvement could be effected via environmental improvement.⁷⁰

In the United States, the locus of scientific eugenics was the Eugenics Record Office (ERO), founded in 1910 at Cold Spring Harbor, Long Island, and later incorporated as a department within the Carnegie Institution of Washington Station for Experimental Evolution.⁷¹ Its director was Charles Davenport, a professor of biology at the University of Chicago and pioneer in biometrics. In founding the ERO, Davenport declared in language that Oliver Wendell Holmes, Jr., would echo, “Society needs to protect itself; as it claims the right to deprive the murderer of his life so also it may annihilate the hideous serpent of hopelessly vicious protoplasm.”⁷²

One could not do experiments on humans as Mendel did on peas, but one could collect data, and Davenport launched a major study on “heredity in relation to eugenics.”⁷³ The goal was to establish the scientific basis of human inheritance through study of family histories; the methodology was to hire field workers to interview families about their histories. (In this regard, the activities at the ERO were quite different from the work of the biologists at the adjacent experimental station.) Trained field

workers asked questions about such behaviors as alcoholism, prostitution, gambling, promiscuity, and criminality; physical “defects” including hermaphroditism, cleft palate, and polydactyly; illnesses such as hemophilia and tuberculosis; mental “defect” such as “feeble-mindedness,” schizophrenia, and other forms of mental illness; and the general category of social attainment and accomplishment.

Between 1911 and 1924, 250 field workers, mostly women, trained at the ERO and were sent out to collect these data. The answers were recorded on index cards. The field workers found that these traits often did run in families. Davenport therefore concluded that social remedies were needed to prevent reproduction by parents carrying “undesirable” traits, and he became an advocate of “segregation”—to keep the mentally and physically ill in home and asylums where they could not breed—and sterilization—to ensure that the unfit, both incarcerated and at large—would not reproduce.

His deputy, Harry Laughlin, used the ERO results to promulgate “Model Sterilization Laws,” and to testify in Congress to the desirability of restricting immigration from southern and central Europe. ERO data demonstrated, he claimed, that immigrants were more likely than native-born Americans to commit crimes, and that this tendency toward criminality was inherited. In 1924, the US Congress passed the Johnson-Reed Act, which severely restricted immigration along eugenic lines.⁷⁴

In the 1930s, thirty-two states in the Union passed sterilization laws, and at least thirty thousand US citizens were sterilized, mostly without informed consent and sometimes without their knowledge.⁷⁵

Laughlin was a hero to many Nazis. In 1936, he was awarded an honorary degree from the University of Heidelberg for his

work on the “science of racial cleansing.” It has been said that the Nazis based their own sterilization laws on the model laws developed by Laughlin at ERO.⁷⁶ At Nuremberg, one proffered defense was that Nazi laws were based on what Americans had advocated.

As I have already noted, eugenics was complicated. Historian Daniel Kevles has argued that eugenics had several principal components, intermixed in various ways:⁷⁷

- *Social control* of reproduction, either through control of marriage or isolation in asylums, jails, and other institutions;
- *Natalism*. Encouraging large families among the “fit” (generally understood to be wealthy and white) and discouraging of reproduction among the “unfit” (everyone else);
- *Malthusianism*. Discouraging social welfare programs, including universal education, minimum wage laws, and public health measures intended to reduce infant mortality on the grounds that they ran against the natural laws that would otherwise weed out the unfit. Eugenicists also discouraged birth control, assuming that those who should use it would not and those who shouldn’t use it would;
- *Hereditarianism and anti-environmentalism*. Rejecting the role of environment and locating the cause of social position and behavioral traits exclusively or nearly exclusively in physical inheritance; and
- *Racial anxiety*. Fearing that breeding of the unfit, coupled with immigration, was polluting or diluting the racial identity of the country, leading to “national” or “racial” deterioration, with those two terms and concepts often used interchangeably.

To this list we may add

- *Gender anxiety.* Eugenic arguments were often coupled to arguments against women's participation in the work force and the promotion of a constricted role centered on home and family.⁷⁸

Most of these elements—racial anxiety, gender anxiety, natalism—are not scientific values, which raises the question: What exactly was the role of science and scientists in the eugenics movement?

It is sometimes claimed that there was a scientific consensus supporting eugenics, and therefore we are justified in disbelieving or rejecting contemporary matters about which there is a scientific consensus. The novelist Michael Crichton, for example, used this argument to try to discredit climate science, likening contemporary calls for action to prevent anthropogenic climate change to earlier calls to prevent race suicide.⁷⁹ Both, he suggested, were politics masquerading as science.

The fact that scientists may have been wrong about some matter in the past in no way tells us whether they are right or wrong about some wholly unrelated matter today, but Crichton's argument does remind us that scientists have not always been on the side of the angels. Insofar as eugenics, like the Limited Energy Theory, was conceptualized and justified as a logical deduction from scientific theory, we cannot simply explain it away as a "misuse" or "misapplication" of science. So was there a scientific consensus on eugenics? The short answer is no.⁸⁰ Prominent social scientists and geneticists objected to eugenic claims. As historian Garland Allen has put it, "It was not the case that nearly everyone in the early twentieth century accepted eugenic conclusions."⁸¹

Social scientists made a complaint that is obvious in retrospect and often invoked today in nature-nurture debates: that many

of the ills recorded by field workers could be explained by bad nutrition, bad education, lack of linguistic skills, and/or bad luck. It was possible that genetics explained adverse outcomes, but so could many other things. The observation of an adverse outcome was no proof of the genetic theory of its causation.

Many poor whites in the 1910s and '20s were immigrants who faced numerous obstacles, including overt discrimination in employment and lack of adequate health care. Reformers pointed to the many immigrant children who had “improved themselves” with the help of education and other social programs, demonstrating that social reforms if seriously pursued could work to improve outcomes. The German-Jewish immigrant anthropologist Franz Boas, in particular, argued that while traits like hair and eye color might be wholly inherited—and there was scientific evidence from laboratory and breeding studies to suggest that this was so—other matters were not easily so reduced. Height, one of Francis Galton’s favorite topics of study, was a case in point. A person’s stature, Boas remarked, is partly inherited, but “is also greatly influenced by more or less favorable conditions during the period of growth.”⁸² Insufficient science had been done to understand the interplay between physical and social factors in determining developmental outcomes, and in the absence of understanding this interplay it was wrong to assume that any complex trait was controlled by genetics, and certainly wrong to assume that it was wholly so.

Boas particularly objected to claims regarding the hereditary character of intelligence. IQ tests had not been shown to measure anything meaningful, and there was no evidence of racially specific hereditary mental or behavioral traits in blacks, immigrants, or any other group.⁸³ We could observe disparate outcomes, but we had no independent evidence of the causes of those outcomes. On the contrary, there was evidence of social

causes: Boas's student Margaret Mead had shown in her 1924 master's thesis that the scores on IQ tests of the children of Italian immigrants varied according to family social status, length of time in the United States, and whether English was spoken at home.⁸⁴

Mead's discussion of Italian immigrants is an important reminder that, while the language of eugenics was that of "racial degeneration," eugenics in America was concerned both with issues of race (as we understand the term today) and with gradations of European ethnicity, both of which were tied to class.⁸⁵ The threat was understood to be to the "Nordic race"—the peoples of northern European descent—from both European and non-European sources, and so a major focus of eugenic study and target of eugenic practice was poor whites. In the United States, that largely meant immigrants, but in the United Kingdom it meant the working class. For this reason, it is perhaps not surprising that another group of scientists who objected to eugenics were socialists, including the British geneticists J.B.S. Haldane, J. D. Bernal, and Julian Huxley, and the American socialist Herman Muller.⁸⁶

Professor of genetics and biometry at University College London, J.B.S. Haldane was the son of the famed Oxford physiologist John Scott Haldane, a socialist who pioneered the study of occupational hazards and originated the practice of bringing canaries into coalmines to monitor air quality.⁸⁷ Initially Haldane sympathized with aspects of eugenics—in college he joined the Oxford Eugenics Society—but he was soon offended by its evident sociopolitical prejudices, particularly its class bias.

Haldane highlighted the thin empirical basis for eugenic claims, particularly given that the mechanisms of inheritance, particularly of complex traits, were only just now coming into scientific focus. Too little was known about heredity to justify

any eugenics program, and “many of the deeds done in America in the name of eugenics are about as much justified by science as were the proceedings of the inquisition by the gospels.” He opposed all sterilization programs, including voluntary ones, on the grounds that “any legislation which does not purport to apply, and is not actually applied (a very different thing) to all social classes alike, will probably be unjustly applied to the poor.” He also insisted in the value and dignity of the working class: “A man who can look after pigs or do any other steady work has a value to society and . . . we have no right whatever to prevent him from reproducing his like.”⁸⁸

Haldane did not believe that “the theory of absolute racial equality” was necessarily correct, but he thought that any actual difference—either of type or degree—would be difficult to establish objectively. The best one might hope for would be to establish difference between populations—as Galton had done—but that would not tell you anything meaningful about the characteristics, much less the social value, of any individual. Perhaps reflecting on his acquaintance with the great American actor and singer Paul Robeson, he insisted that “it is quite certain that some negroes are intellectually superior to most Englishmen.”⁸⁹

Herman Muller, who shared the Nobel Prize for his work demonstrating that x-rays could induce heritable genetic changes in fruit flies, also objected to eugenics. Muller is a complex case, insofar as he did not doubt that the human race could in principle be improved through eugenic practices. Nor did he doubt that ideally it should be. But equally firmly he believed that improvement would never happen equitably under capitalism.

Muller was the principal author of the 1939 *Geneticists' Manifesto*, signed by twenty-two American and British scientists (as well as the historian of science Joseph Needham), in response

to a request from the Science Service to answer the query, “How could the world’s population be improved most effectively genetically?”⁹⁰ Muller and his colleagues rejected the premise that this question could be answered biologically. They began by insisting that the question “raises far broader problems than the purely biological ones, problems which the biologist unavoidably encounters as soon as he tries to get the principles of his own special field into practice.”⁹¹ In other words, this was not only or even primarily a biological question:

For the effective genetic improvement of mankind is dependent upon major changes in social conditions, and correlative changes in human attitudes. In the first place, there can be no valid basis for estimating and comparing the intrinsic worth of different individuals, without economic and social conditions which provide approximately equal opportunities for all members of society instead of stratifying them from birth into classes with widely different privileges.⁹²

These men were not unilaterally opposed to efforts to make genetic improvements to the human race. Even after revelations of Nazi atrocities, Muller continued to support the idea of deliberate human improvement, arguing in 1954 that “the fact that the so-called eugenics of the past was so mistaken . . . is no more argument against eugenics as a general proposition than say the failure of democracy in ancient Greece is a valid argument against democracy in general.”⁹³ (This would have been an interesting response to Michael Crichton.) But Muller and his colleagues rejected much if not all of the evidence being invoked to support eugenic claims, because reigning accounts assumed a level social playing field that patently did not exist.

Existing studies assumed that observed differences were genetic: in effect assuming the thing they were intended to prove.

The authors of *Geneticists' Manifesto* accepted that there were both genetic and environmental aspects of intelligence, behavior, social accomplishments, and many other things—as indeed most scientists accept today. But existing studies failed to identify the relative contributions of social and genetic elements in human characteristics.

Before people in general, or the state which is supposed to represent them, can be relied upon to adopt rational policies for the guidance of their reproduction, there will have to be . . . a far wider spread of knowledge of biological principles and recognition of the truth that both environment and heredity constitute dominating and inescapable complementary factors in human well-being.⁹⁴

No real advance could be made, they held, without “the removal of race prejudices and of the unscientific doctrine that good or bad genes are the monopoly of particular peoples or of persons with features of a given kind,” and this would not occur until “the conditions which make for war and economic exploitation have been eliminated [through] some effective sort of federation of the whole world based on the common interests of its peoples.” Capitalist societies manifestly did not provide “approximately equal opportunities for all.” Eugenics could not work under capitalism: the lower classes would always be targeted.

A level playing field would only be a start, moreover, because it was unreasonable to expect any parent to worry about the state of future generations unless they were first “extended adequate economic, medical, educational and other aids in the bearing and rearing of children.” It was also unreasonable to expect intelligent women to abandon their personal interests and aspirations on behalf of improving the population at large; the scientists therefore suggested the need for social policies to

ensure that a woman's "reproductive duties do not interfere too greatly with her opportunities to participate in the life and work of the community at large." This meant that workplaces needed to be "adapted to the needs of parents and especially mothers," and that towns and community services needed to be reshaped "with the good of children as one of their main objectives." It also meant that women needed access to safe and effective birth control: "A . . . prerequisite for effective genetic improvement is the legalization, the universal dissemination, and the further development through scientific investigation, of ever more efficacious means of birth control . . . that can be put into effect at all stages of the reproductive process," including voluntary sterilization and abortion.⁹⁵

Finally, these geneticists noted, to improve the world through selection would require agreement about what constituted improvement, something that was by no means apparent, particularly if the goals of selection were social ones. In their view, the most important genetic characteristics one might want to try to foster would be those for health, for the "complex called intelligence," and "for those temperamental qualities which favour fellow-feeling and social behavior, rather than those (to-day most esteemed by many) which make for personal 'success,' as success is usually understood at present."⁹⁶ So despite expressing in-principle support for eugenic ideals, they opposed eugenic proposals in practice. The prerequisite for improving the quality of the world's population was improving the social conditions of the world.⁹⁷

The socialist geneticists' opposition to eugenics was rooted in their politics, but one did not have to be a socialist (or social scientist) to recognize flaws in eugenic research. In particular, many geneticists pointed out the fallacy of conflating genes with outcomes. Garland Allen has stressed that the great British

statistician Karl Pearson, who was a eugenicist, strongly criticized the work of the ERO as “carelessly and sloppily conceived and executed, and lack[ing] any semblance of normal scientific rigor.”⁹⁸

Herbert Spencer Jennings was an American geneticist at Johns Hopkins University known for his 1930 book, *The Biological Basis of Human Nature*.⁹⁹ While the title might suggest an argument for genetic determinism, the book presented the scientific case for the interaction of genes and environment. Against the genetic determinists, Jennings wrote:

A given civilization is the outgrowth of the interaction of the genetic constitutions present in the population, with the environment—including knowledge, inventions, traditions—of that population. By changes in the latter set of factors enormous differences have in the past been made in the cultural system. . . . No cultural system is the outgrowth of genetic constitution alone.¹⁰⁰

And against the environmental determinists:

[The environmental determinist argues that] by subjection to adequately diverse environments, diverse training and instruction, any of [a group of people] can be made . . . into “doctor, lawyer, merchant, chief” . . . Biology has no proper quarrel with such an assertion. What an enlightened view of biology would add . . . is this: While any of the normal individuals, taken early and properly guided, could be made into physicians, it would take different treatment to accomplish that end in the different individuals.¹⁰¹

Eugenicists had committed numerous logical and methodological fallacies, including being overly influenced by implicit assumptions (“underlying . . . but never stated”), ignoring evidence that did not support their positions, and persisting in

“mistaken conclusions after the discovery that they are mistakes.”¹⁰² Jennings was particularly critical of what he called “the fallacy of non-experimental judgments,” noting that precisely because nearly everyone had an opinion on heredity and evolution, it was “essential to set aside prior views and build one’s opinions on the basis of experimental evidence.” But this was just what most eugenicists failed to do; they ran with their priors and ignored disconfirming evidence.¹⁰³ Jennings also noted the widespread use of what today we would call the fallacy of the excluded middle—to assume that because some traits have been shown to be inherited, all traits are inherited, and vice versa regarding the environment—and, in language reminiscent of T. C. Chamberlin, “The fallacy of attributing to one cause what is due to many causes.”¹⁰⁴

For Jennings it was obvious that the answer to the nature/nurture debate was both/and. He made the point in a 1924 article by analogy to material objects:

What happens in any object—a piece of steel, a piece of ice, a machine, an organism—depends on the one hand upon the material of which it is composed [and] on the other hand upon the conditions in which it is found. Under the same conditions objects of different material behave diversely; under diverse conditions objects of the same material behave diversely. . . . Neither the material constitution alone, nor the conditions alone, will account for any event whatever; it is always the combination that has to be considered.

And so it was for organisms. “The individual is produced by the interaction of genes and environmental conditions; so that the same set of genes may yield diverse characteristics under diverse environments.” Eugenics was doomed to fail, because “behavior is bound to be relative to environment, it cannot be dealt

with as dependent on genes alone. A given set of genes may result under one environment in criminality; under another in the career of a useful citizen.”¹⁰⁵

Jennings is but one example; if space permitted we could easily multiply his critique. The Nobel Laureate T. H. Morgan, famous for his work on the genetics of fruit flies, stressed in the 1920s that the problems eugenicists proposed to repair would likely be more quickly remedied through social reform than through selective breeding.¹⁰⁶ Many non-scientists also raised methodological and moral objections.¹⁰⁷ (And there were objections raised in other countries that I have not considered here.)¹⁰⁸ The important point here is that eugenics as a political movement in important ways conflicted with scientific understanding, and it is simply not correct to say that there was a scientific consensus on eugenics.¹⁰⁹

Now let us consider an example where there was a consensus, but one that ignored or at least discounted important, significant evidence.

Example 4: Hormonal Birth Control and Depression

Many women have had the experience of becoming depressed or melancholic on taking the contraceptive pill, many doctors are aware of their patients' experience, and many scientific studies have affirmed this link. Indeed, some of the earliest studies of the effects of the Pill in the late 1950s noted side effects including “crying spells” and “irritability,” and the package insert that now comes with it states that one of the known side effects is “mental depression” (see fig. 1).

ADVERSE REACTIONS

An increased risk of the following serious adverse reactions has been associated with the use of oral contraceptives (see WARNINGS section).

- Thrombophlebitis and venous thrombosis with or without embolism
- Arterial thromboembolism
- Pulmonary embolism
- Myocardial infarction
- Cerebral hemorrhage
- Cerebral thrombosis
- Hypertension
- Gallbladder disease
- Hepatic adenomas or benign liver tumors

There is evidence of an association between the following conditions and the use of oral contraceptives:

- Mesenteric thrombosis
- Retinal thrombosis

The following adverse reactions have been reported in patients receiving oral contraceptives and are believed to be drug-related:

- Nausea
- Vomiting
- Gastrointestinal symptoms (such as abdominal cramps and bloating)
- Breakthrough bleeding
- Spotting
- Change in menstrual flow
- Amenorrhea
- Temporary infertility after discontinuation of treatment
- Edema
- Melasma which may persist
- Breast changes: tenderness, enlargement, secretion
- Change in weight (increase or decrease)
- Change in cervical erosion and secretion
- Diminution in lactation when given immediately postpartum
- Cholestatic jaundice
- Migraine
- Allergic reaction, including rash, urticaria, and angioedema
- Mental depression
- Reduced tolerance to carbohydrates
- Vaginal candidiasis
- Change in corneal curvature (steepening)
- Intolerance to contact lenses

FIGURE 1. Detail of package insert for the oral, hormonal contraceptive ORTHO TRI-CYCLEN® Lo Tablets (norgestimate/ethinyl estradiol) showing “mental depression” among the list of potential adverse reactions, which are “believed to be drug-related.”

Recently, there was a flurry of media attention about a new study demonstrating that the Pill can cause depression.¹¹⁰ Physicians lauded the study, and the media presented the result as a novel finding.¹¹¹ My own daughter, however, asked me on the day the coverage hit the media: How is this *news*? She knew that the Pill could cause depression, because I had told her so.

I have no history of depression—no family history of depression or mental illness of any sort—but when I was in my mid-twenties, I experienced a sudden and peculiar bout of extreme melancholy. I lost my energy for daily tasks, lost interest in my work, and, after about six weeks, found myself having trouble getting out of bed. And yet, in other respects my life was going well. I was in my second year of graduate school, had done very well in my first year, was working on an exciting project for which I had adequate funding, and had met a very nice man who would soon become my husband (and to whom I’ve now been married for more than thirty years).

I went to counseling at a campus health center, and I was lucky. The female counselor asked me straight away: Are you on the Pill? The answer was yes. I explained that I had recently returned from Australia, and because Australia at that time had free health insurance, including prescription drugs, I had bought a year’s worth before I left. But the particular formulation that I had been prescribed in Australia was not available in the United States, so when the year was up I had to switch to another type. That had occurred two months before. The onset of my depression began shortly after I had started this new form of the Pill. The therapist told me that the type of pill I was now on—a combination formulation—was well known to be more likely to cause depression than some other options. I stopped the drug immediately and my recovery began nearly as immediately. Within a few weeks I was back to my normal self, I

thanked the therapist, and went on to a successful academic career and life.

My experience can be dismissed as “just an anecdote,” but I prefer to view it as a clinical study in which $n = 1$. The more important point is that many women have had such experiences and reported them to their physicians and therapists. The website Healthline.com, which claims to be the “fastest growing consumer health information site,” notes that “depression is the most common reason women stop using birth control pills.”¹¹² Moreover, like me, many women have bounced back to normal when they stopped taking the Pill or switched to other formulations. And these case reports have spurred numerous scientific studies. As one physician recently wrote, “decades of reports of mood changes associated with these hormone medications have spurred multiple research studies.” So my daughter was correct to ask: how was this new study news?

One answer was offered by Monique Tello, a practicing MD, MPH, who writes for the *Harvard Gazette*: “The study of over a million Danish women over age 14, using hard data like diagnosis codes and prescription records, strongly suggests that there is an increased risk of depression associated with *all* types of hormonal contraception.” Previous studies, in contrast, were all “of poor quality, relying on iffy methods like self-reporting, recall, and insufficient numbers of subjects.” The authors of the new study concluded that previously it had been “impossible to draw any firm conclusions from the research on this subject.”¹¹³

It is hard to argue with a study of over one million women. It is also hard to argue with any study done in Denmark, which has a national health care database covering every Danish citizen and thus allows researchers to correct for sampling biases and other confounding effects. It is thanks to Denmark that we can say with confidence that children who are fully vaccinated according to

prevailing public health recommendations do not suffer autism at greater rates than those who are not.¹¹⁴ So, three cheers for Denmark. Three cheers, as well, for this big, new convincing study. But note the explanation of why it took so long to come to this point: the lack of “hard data like diagnosis codes and prescription records.” Previous studies, we are told, relied on “iffy methods like self-reporting, recall, and insufficient numbers of subjects.”¹¹⁵

The term “hard data” should be a red flag, because the history and sociology of science show that there are no hard data. Facts are “hardened” through persuasion and their use. Moreover, remarks of this type raise the question of why some forms of data are considered hard and others are not. Just look at what is being considered hard data here: diagnosis codes and prescription records. Many people would say hard data are quantitative data, but neither of these constitutes a measurement: they are the subjective judgments of practitioners and the drugs they choose to prescribe in response to those judgments.¹¹⁶ Moreover, there is a substantial literature on misdiagnosis in medicine, and on the distorting effects of pharmaceutical industry advertising and marketing on prescribing practices.¹¹⁷ Given what we know about medical practice and its history, the idea that diagnosis codes and prescription records should be taken as hard facts seems almost satirical.

But it gets worse: the study authors accepted the reports of doctors—their diagnosis codes and prescription records—as facts, whereas the reports of female patients were dismissed unreliable—in Tello’s words: “iffy.” Bias—either against women or against patients—is clearly on display. But here is the key point: the conclusion of the Denmark study is the *same* as all those iffy, self-reports from female patients. If the new study is correct, then the allegedly iffy self-reports were correct all along.

These self-reports involved millions of women, too. The Pill has been on the market in the United States and Europe since the early 1960s. According to the CDC, during period 2006–10, over ten million American women took the Pill.¹¹⁸ According to the World Health Organization, over one hundred million are currently taking it worldwide.¹¹⁹ While self-reporting does not offer a good basis for an accurate quantitative assessment of risk of depression caused by hormonal contraception, it surely offers important qualitative evidence. It seems extremely unlikely that all the women who reported mood changes while on the Pill were simply confused or making it up.

In fact, the connection between hormonal birth control and depression has been known almost as long as the Pill has been on the market. In 1969, feminist journalist Barbara Seaman published *The Doctor's Case against the Pill*, a book that helped to launch the women's health movement. Seaman's book made women and doctors around the country aware of the serious health risks of the Pill as it was then formulated, and led to congressional hearings resulting in the first package insert to warn against risks involved with a prescription medication. Chapter 15 of her book was entitled "Depression and the Pill," and it began:

Psychiatrists were among the first doctors to persuade their own wives to stop using birth control pills. Finely tuned to emotional feedback, they did not take long to notice certain adverse reactions in their wives and daughters, patients and friends. The effects that were the most obvious ranged from suicidal and even murderous tendencies to increased irritability and tearfulness. . . . A few pill-users have become so hostile, suspicious and delusional that they have seriously thought of murdering—or have actually attempted to murder—their own husbands and children. Others attempt to commit suicide and some have succeeded.¹²⁰

Within a few years of the Pill coming on the market, adverse mental health effects had been widely reported. A 1968 study in the United Kingdom looked at 797 women who took oral contraceptives; many reported emotional side effects and two committed suicide.¹²¹ By 1969 British researchers had found that one in three Pill users experienced personality changes; three in fifty who were studied became suicidal. A US study by researchers at the University of North Carolina School of Medicine found that 34% of otherwise healthy young women reported depression after starting on the Pill. These studies did not include control groups, but one in Sweden compared two groups of postpartum women, matched with respect to social background, previous history of depression, and other factors. It found significantly higher rates of psychiatric symptoms in women who went on the Pill after giving birth than in the group who used other forms of birth control.

We cannot judge from Seaman's account how good any of these studies were; her point was that to the extent that scientists had examined the question, they had found evidence to support women's accounts, accounts that formed the emotional heart of Seaman's story. She told of women who became agitated and disorganized; who experienced panic attacks in movie theatres; who set "accidental" fires; who found themselves weeping uncontrollably for no apparent reason; and who felt themselves to be on the verge of breakdown. Some of these women may have been depressed for other reasons, but Seaman supported their accounts with testimony from psychiatrists. Women's stories formed the emotional center of the book, but doctor's stories provided the intellectual center. It was not the *patient's* case against the Pill, but the *doctor's* case.

With respect to the mental health effects, the key doctors were psychiatrists, to whom women had gone for help after

becoming depressed on the Pill, or who noticed changes in their wives and friends and in patients they had been seeing for some time and knew well. Seaman quoted one Manhattan psychiatrist describing the resistance he initially encountered from other doctors:

My fights with the gynecologists began in 1963 [three years after Enovid, the first oral contraceptive, was approved.]¹²² I'd been seeing one patient twice a week for two years. . . . She was tough as nails . . . Her father had been an alcoholic. She'd fought her way to the top as a fashion model. . . . She's one of the most sensible patients I ever had. Exploitative? Yes. Neurotic? A little. Depressed? Never. Eight days after this patient went on the pill, she arrived for her appointment and wept through the whole session. The same thing happened the next time and the next . . . She talked about "giving up" and "ending it all." I suggested that she get off the pill. We'd see what happened then. She did. The next time I saw her she was her old self. But then came the first in a series of calls from her gynecologist. In essence what he had to say was, "You stick to your own unraveling or whatever it is you do, and let me take care of *my* knitting. Birth control is not a psychiatrists' province."¹²³

(Eventually, gynecologists would accept there was a pattern; this particular gynecologist was convinced by the psychiatrist and began to send him patients for Pill-induced depression.)

Other psychiatrists told similar stories. Patients they had known for years were suddenly different; or patients were sent by their families because of sudden, frightening changes. The Atlanta physician John R. McCain presented a paper at the New England Obstetrical and Gynecological Society warning that the mental health effects of the Pill were "among the complications which seem to have the most serious potential

danger.”¹²⁴ The good news, many doctors noted, was that when women went off the Pill, the abatement of symptoms was as fast as the onset. This, of course, was further evidence that the Pill was a factor in their condition.

In many of these stories, women noted that their mood swings and depression were similar to what they had experienced when they were pregnant or just after giving birth—and doctors had rarely doubted that hormones had something to do with those experiences! Among the various stories recounted in the book, my personal favorite is this one: “When I was on the pill,” one psychiatrist’s wife reported, “I hardly ever got off the couch except to slap one of the children.”¹²⁵

In the years that followed, scientists and physicians undertook studies of the mental health effects of the Pill. But considering how many women have taken the Pill, the total number of studies is startlingly modest. A quick PubMed search in 2016 on hormonal birth control and depression/mood or psychological disorders/libido changes found twenty-seven papers. This may be an underestimate—other key words or phrases might have turned up more, and mood changes may also have been detected in studies concentrating on other things—but compare this to another issue that I have studied: climate change. In my 2004 study of climate science, I used a sample of just under one thousand articles to estimate the state of scientific opinion.¹²⁶ That sample came from a population that was estimated to be over ten thousand papers. Since that time, at least that many more have been published.¹²⁷ Given that over one hundred million women are on the Pill today, doesn’t it seem troubling that there are so few studies on something that was recognized as a potentially serious problem more than fifty years ago?

Mood changes are admittedly a difficult thing to study and almost impossible to quantify. Feelings are, by definition,

subjective, and depression cannot be measured in the sort of way that cholesterol or high blood pressure can be. But consider this: in 2016, a clinical trial of a male hormonal contraceptive injection in 320 men was *abandoned* after the men taking part reported increased incidences of adverse effects, including changes in libido and mood disorders. In fact, more than 20% reported mood disorders. One man developed severe depression; another tried to commit suicide. Because of the adverse effects, the trial was halted—even though the rate of pregnancy suppression was more than 98%. The researchers reported:

The study regimen led to near-complete and reversible suppression of spermatogenesis. The contraceptive efficacy was relatively good compared with other reversible methods available for men. The frequencies of mild to moderate mood disorders were relatively high.¹²⁸

The male hormonal contraceptive injection was shown to work as well as the Pill, yet the clinical trial was stopped because of adverse effects, one of which was a dramatic increase in mood disorders.¹²⁹ If you are wondering how the researchers measured this, the answer is: self-reporting.

This result could have been predicted, not only because similar effects were seen in women, but because there is a mechanism that explains why hormonal contraceptives have this effect. It is the link between reproductive hormones and serotonin.

Low levels of serotonin, a neurotransmitter in the brain, have been linked to depression. High levels of estrogen, as in first-generation [oral contraceptives], and progestin, as in some progestin-only contraceptives, have been shown to lower the brain serotonin levels by increasing the concentration of a brain enzyme that reduces serotonin.¹³⁰

The converse is also true: anti-depressant drugs that target serotonin uptake, such as Prozac and Zoloft, are known to have an adverse effect on libido.¹³¹ They can also cause erectile dysfunction and anorgasmia; one study in the 1990s found that 45% of female patients on SSRIs (selective serotonin reuptake inhibitors) experienced drug-induced sexual dysfunction and some studies suggest even higher rates.¹³² This occurs because drugs that stimulate serotonin uptake can interfere with the uptake of hormones involved in sexual desire and reproduction, like dopamine.¹³³ In other words, the issue cuts both ways: drugs that are or target hormones involved in sex can cause depression; drugs that treat depression can affect the hormones involved in sex.

We have known for fifty years that the Pill can cause mood disorders in women. We know that drugs that treat mood disorders can affect hormones involved with libido, and scientists know at least one mechanism by which this occurs. And a recent study was stopped because hormonal contraceptive caused mood disorders in male subjects. A reasonable person might therefore ask: what was left to be established? Or as my daughter put it, why was the finding that the Pill causes depression in women viewed as *news*?

Let us return to the Denmark study. It did not find that previous studies of oral contraceptives had shown that hormonal contraception did *not* cause mood changes. Rather, it concluded that “inconsistent research methods and lack of uniform assessments [made] it difficult to make strong conclusions about which . . . users are at risk for adverse mood effects.”¹³⁴ In other words, it suggested that until now, we didn’t know enough to draw a firm conclusion.

These researchers took the conventional approach of assuming no effect and requiring statistical proof at a specific

significance level to say that an effect had been detected—and was therefore *known*. So did the various studies that preceded them. There's nothing particularly shocking about this; it is common statistical practice. But it says, in effect, that if evidence is not available that meets that standard, we must conclude that our results are inconclusive—or in lay terms, that we just don't know.

There are two problems with this approach. The first, which is a general one, is that a negative finding is often taken as indicating “no effect,” when in fact it simply means that the researchers have not been able to detect the effect, at least not at a level that achieves statistical significance. (Many negative studies actually do see effects, but not ones that pass the bar of statistical significance at the 95% level.)¹³⁵ It is the classic conflation of absence of evidence with evidence of absence, and it can lead to false negative conclusions. Still, if enough good studies are done that consistently fail to find an effect (or one really large one with great statistical power), we might fairly conclude that the effect really isn't there.

But what if there is evidence from non-statistical sources, such as patient reports, that there may well be an effect? What if there is a theoretical reason (as there is here) to think that an effect is in fact *likely*? In that case, why are we assuming that there is none? Why are researchers playing dumb? If we know or have reason to suspect that something is a risk, it may be warranted to flip the null and use a default assumption of “effect” rather than “no effect,” or to accept a lower level of statistical significance. (This has sometimes been done, as when the Environmental Protection Agency accepted some studies of the impacts of secondhand smoke at a confidence level of 90% rather than 95% on the grounds that the same chemicals that were known to cause cancer in primary smoke were also present in secondhand

smoke.)¹³⁶ After decades of case reports, and with a mechanism to explain why it might be so, researchers should have accepted the null hypothesis that the Pill could cause depression and sought statistical evidence to disprove that hypothesis.

The second problem relates to how we think about causation. The classic argument that correlation is not causation is misleading. What we should say is that correlation is not *necessarily* causation. Many things are correlated that are not causally related. But if we have an observed correlation between two phenomena, and we are aware of a mechanism that explains how one of them can be caused by the other, and if that mechanism is known to be present, then the logical conclusion is the observed correlation *is* caused by the known mechanism. Under these conditions, correlation *is* causation. Or at least, it is likely to be.

A classic example is the correlation between shark attacks and ice cream sales. Statisticians love to use this as an example to prove how correlations can be misleading: both are related to warm weather, when people swim in the ocean and eat ice cream. Neither one causes the other. But what if we had independent evidence that the smell of ice cream attracted sharks? Then it might be the case that there was a causal relation. Now suppose that the correlation did not achieve statistical significance at the 95% level. Would we conclude that there was no relation between the ice cream and the attacks? Under currently prevailing norms, we would. And we would be wrong. We need to pay attention to mechanisms.¹³⁷

Consider another example. When the United States reduced the speed limit on interstate highways to fifty-five miles per hour, traffic fatalities dropped dramatically. The motivation for the speed limit change was to save fuel, not lives, so one might initially suppose that this correlation was just coincidental. In fact,

driving at lower speeds reduces the chance of an accident and the likelihood that any accident that occurs will be fatal. Because we understand this, we rightly conclude that lowering the speed limit caused a decrease in traffic fatalities.

Playing dumb makes sense when we have no reason to suspect that phenomena are linked, or have affirmative reason to suppose they are not. If we knew nothing about hormones and mental health we might have rightly said that we needed more evidence to conclude that the Pill might cause depression. But we know that hormones affect brain chemistry. This is one reason why manufacturers have worked to decrease estrogen levels in oral contraceptives.

Women have always known that we sometimes get moody and depressed right before our periods. Indeed, popular lore makes us unreliable—as scientists, as political leaders, as CEOs—*because* of this. Stereotypes typically draw the wrong conclusions from the evidence on which they are based, but that in and of itself is not a refutation of the evidence. Hormones affect our moods. This is true for men and women.

Doctors who have not warned their patients of this risk during the past thirty years have been ignoring evidence. Public health officials who have downplayed the risk by discounting evidence—in this case, reams of it collected over more than three decades—because it did not meet certain methodological preferences have done women a grave disservice. Had doctors and public health professionals paid more attention to “iffy” case reports instead of discounting them, they would not simply have come to a better conclusion, epistemologically. They would have done their jobs better—and served their patients well—by not discounting a real and troubling side effect of an otherwise desirable medication. Given that the Pill has been implicated in suicidal ideation, they might even have saved lives.

Example 5: Dental Floss

My final case involves a very grave public health issue: dental floss.

Many people have recently heard that flossing your teeth doesn't do you any good. In August 2016 there was a flurry of coverage saying so. The *New York Times* asked, "Feeling Guilty about Not Flossing? Maybe There Is No Need."¹³⁸ The *Los Angeles Times* reassured its readers that if they didn't floss, they needn't feel bad because it probably doesn't work anyway.¹³⁹ So did *Mother Jones*, which ran the headline, "Guilty No More: Flossing Doesn't Work."¹⁴⁰ *Newsweek* asked, "Has the Flossing Myth Been Shattered?"¹⁴¹

These various reports were based on an article by the Associated Press (AP) that claimed that there is "little proof that flossing works." The AP quoted National Institutes of Health dentist Tim Iafolla, acknowledging "that if the highest standards of science were applied in keeping with the flossing reviews of the past decade, 'then it would be appropriate to drop the floss guidelines.'" ¹⁴² The *Chicago Tribune* linked this latest reversal in scientific fortune to previous (alleged) reversals on salt and fat.¹⁴³ Evidently, we can add dental floss to the list of issues on which scientists have "got it wrong."

It was not just the alleged lack of evidence that caught reporters' attention; there was also a suggestion of incompetence or even malfeasance. The *New York Times* suggested that the federal government may have violated the law that stipulates that federal dietary guidelines must be based on scientific evidence. So too the AP: "The federal government has recommended flossing since 1979, first in a surgeon general's report and later in the Dietary Guidelines for Americans issued every five years. The

guidelines must be based on scientific evidence, under the law.”¹⁴⁴ *The Week* ran the story under the headline “Everything You Believed about Flossing Is a Lie.”¹⁴⁵ The *Detroit News* referred to the defenders of floss as the “floss-industrial complex.”¹⁴⁶ One website called it “The Great Dental Floss Scam.”¹⁴⁷

Many reports contained an element of *schadenfreude*: some journalists seemed practically gleeful that journalists had one-upped scientists.¹⁴⁸ WRVO, an NPR affiliate in Oswego, New York, ran the story under the headline “How a Journalist Debunked a Decades-Old Health Tip.”¹⁴⁹ The report claimed that the story began when AP reporter Jeff Donn learned from his “son’s orthodontist . . . that there was in fact no good evidence that dental floss helps prevent cavities and gum disease.”¹⁵⁰ Poynter.org labeled the story “How a Reporter Took Down Flossing.”¹⁵¹ A website promoting collective consciousness and natural living ran the story under the headline “The Deceit of the Dental Health Industry,” stating that “flossing has been shown to be almost useless in terms of its purported benefits” and suggesting that “most of your oral health is determined by your diet and nutrition.”¹⁵²

On the face of it, this certainly appears to be a case of scientists having “got it wrong.” Dentists and public health officials, including those in positions of governmental authority, have been instructing us about something that now we are told is not the case. We have wasted time and money on something useless. And this bears directly on the issue of trust, because if scientists have been wrong for decades about dental floss—as well as perhaps fat and sugar—then what else have they been wrong about? Will they tell us next that it is all right to smoke? Or that climate change is a hoax? Scientists might be tempted to respond that the fracas over flossing is just the messy work of science correcting itself, as a major scientific study revealed the weaknesses

in previous work. But that is *not* what transpired. In fact, this is not a case of scientists getting it wrong at all. It's a case of journalists getting it wrong, and scientists getting blamed.

The “scientific” finding was not a scientific finding at all, but the result of an investigation by one reporter for the AP.¹⁵³ The source of the media story was the media itself. According to their own reporting, which they filed under the rubric “The Big Story,” the AP “looked at the most rigorous research conducted over the past decade, focusing on 25 studies that generally compared the use of a toothbrush with the combination of toothbrushes and floss. The findings? The evidence for flossing is ‘weak, very unreliable,’ of ‘very low’ quality, and carries ‘a moderate to large potential for bias.’ ‘The majority of available studies fail to demonstrate that flossing is generally effective in plaque removal,’ said one review conducted last year. Another academic review, completed in 2015, cites ‘inconsistent/weak evidence’ for flossing and a ‘lack of efficacy.’”¹⁵⁴

The *New York Times*, seemingly staying close to the facts, informed readers that “A review of 12 randomized controlled trials published in The Cochrane Database of Systematic Reviews in 2011 found only ‘very unreliable’ evidence that flossing might reduce plaque after one and three months. Researchers could not find any studies on the effectiveness of flossing combined with brushing for cavity prevention.” (We will address the distinction between gum health and cavity prevention in a moment.) But, as the *Times* rightly noted, that study was done in 2011, so how and why did this become a story in 2016?

According to the AP, their investigation of the matter was triggered by a decision by the US government to drop flossing from federal dietary guidelines (and not by Jeff Donn’s conversation with his son’s orthodontist, raising further questions about the whole story). This led them to ask the question: “What were

those guidelines based on in the first place?”¹⁵⁵ While it was later revealed that the guideline change was the result of a decision to focus the dietary guidelines on diet—i.e., food—rather than other health practices, the cat was out of the bag.¹⁵⁶ The “finding” that flossing does not work was all over the news. As for Donn, he was quoted in a subsequent interview: “I think the best science indicates that [by flossing] I’m not doing anything beneficial for my health.”¹⁵⁷

Let us step back from the media coverage to ask: what scientific evidence exists to support or refute the claim that dental floss is of value? Donn is neither a scientist nor a dentist, and in fact his claim is not correct. The available science does *not* indicate that by flossing we are “not doing anything beneficial” for our health.

The most well-known and respected source of information on the state of the art in biomedicine is the Cochrane group, a non-profit collaboration that bills itself as “representing an international gold standard for high quality, trusted information.” The collaboration claims thirty-seven thousand participants from more than 130 countries who “work together to produce credible, accessible health information that is free from commercial sponsorship and other conflicts of interest.”¹⁵⁸ As the *New York Times* correctly reported, in 2011, the collaboration issued a report from its oral health group reviewing existing clinical trials examining the benefits of regular use of dental floss.¹⁵⁹

The report was based on a review of twelve trials, with 582 subjects in flossing-plus-toothbrushing groups and 501 participants in toothbrushing-alone groups. The report summary reads as follows:

There is some evidence from twelve studies that flossing in addition to tooth-brushing reduces gingivitis compared to

tooth-brushing alone. There is weak, very unreliable evidence from 10 studies that flossing plus tooth-brushing may be associated with a small reduction in plaque at 1 and 3 months. No studies reported the effectiveness of flossing plus tooth-brushing for preventing dental caries [tooth decay].¹⁶⁰

That part of the summary was reported, wholly or in part, in many of the media reports. But the report also said:

Flossing plus tooth-brushing showed a statistically significant benefit compared to tooth-brushing in reducing gingivitis at the three time points studied [although the effect size was small].¹⁶¹ The 1-month estimate translates to a 0.13 point reduction on a 0 to 3 point scale for . . . gingivitis . . . and the 3 and 6 month results translate to 0.20 and 0.09 reductions on the same scale.¹⁶²

This additional information refutes much of the media presentation. The crux of the news coverage was that many existing studies are weak, involving small numbers of people or very short periods of time. That is true. But it is not the same as demonstrating that flossing has no benefit. On the contrary, if the Cochrane review is correct, these studies indicate that, over the time period of the study, small but statistically significant reduction of gingivitis was observed in patients who flossed along with brushing.

The Cochrane review also considered evidence that flossing may help reduce plaque, which is associated with cavities as well as other matters. On this, they concluded that

Overall there is weak, very unreliable evidence which suggests that flossing plus tooth-brushing may be associated with a small reduction in plaque at 1 or 3 months. None of the included trials reported data for the outcomes of caries, calculus, clinical attachment loss, or quality of life.

Here we can identify one source of difficulty and potential misunderstanding: a number of different questions are being conflated, including whether flossing improves your life. Let us concentrate on the two main issues, as reported on both by the Cochranes and the news media: plaque and gingivitis. Plaque matters because it can lead to dental caries, and gingivitis matters because it is the first stage of periodontal disease, which can lead to tooth loss later in life. More than 70% of Americans over 65 have some form of periodontitis, which is *always* preceded by gingivitis.¹⁶³ If flossing reduces gingivitis, then it is likely that flossing reduces periodontal disease. Periodontal disease has been linked to serious illness, including increased risk of cancer and Alzheimer's disease.¹⁶⁴

Dental floss defenders made this point. What the Cochranes concluded was not that flossing doesn't help, but that we don't have sufficient studies of high enough quality pursued over sufficiently long periods to demonstrate that it does help. The American Academy of Periodontology pointed out that "the current evidence fell short because researchers had not been able to include enough participants or 'examine gum health over a significant amount of time.'" Dr. Philippe Hujoel, a professor of oral health sciences at the University of Washington, Seattle, called it "very surprising" that "we don't have the . . . randomized clinical trials to show [flossing is] effective," given how widespread the belief is that flossing does help.¹⁶⁵

But it is so surprising? Perhaps not. What we learned in 2016 was that we didn't have the long-term, randomized clinical trials that would be necessary to prove the benefits of dental floss according to prevailing medical standards. It's not that hard to understand why, in a world of cancer, heart disease, opioid abuse, and the continued use of tobacco products, such studies have not been done. It's not egregious that researchers have focused their

attention on matters that appear to be more serious. What is egregious is that in the absence of evidence that meets the “gold” standard of the randomized clinical trial, people have concluded that there is no evidence at all. That is both false and illogical.¹⁶⁶

Moreover, the gold standard of clinical trials is not just the *randomized* trial, but the *double-blind* randomized trial, and it is impossible to do a double-blind trial of dental floss. (This difficulty also plagues studies of nutrition, exercise, yoga, meditation, acupuncture, surgery, and any number of interventions of which the subject is necessarily aware.) Any study of floss usage will also require self-reporting, which, as we have seen, is disparaged. Moreover, if you believe that long-term flossing can prevent tooth loss in old age, it would be unethical to ask a control group to refrain from flossing for what would have to be the better part of their lives. The sort of study that would be required to convince those who subscribe to the “gold standard” is both impossible and arguably unethical to perform.¹⁶⁷

Donn interpreted his findings to say that existing studies show no long-term benefits even when floss is used properly; once again we are observing the fallacy of equating absence of evidence with evidence of absence.¹⁶⁸ None of these studies was long enough to demonstrate long-term benefits. Dunn was also quoted as saying that there was “no good evidence.” Whether this is correct depends on your definition of “good,” but clearly there *is* evidence that flossing may have benefit.

In the aftermath of the negative media coverage, dentists who support flossing appealed to clinical experience. Several articles quoted dentists, professors of dentistry, and deans of dental schools affirming that clinical practice reveals that those who floss have healthier teeth and gums than those who don’t. Some dentists went so far as to suggest that they can tell who among their patients is lying about their flossing habits simply by

observing the conditions of their gums. (This reminds us of another reason why a good clinical trial would be hard to do: people lie about flossing. One study concluded that one in four Americans who claimed to floss regularly was fibbing.)¹⁶⁹ And then there is the experience of patients—which is to say, all of us. Many of us have noticed that when we floss regularly our gums don't bleed, and bleeding gums can be a sign of early periodontal disease. The dean of the dental school at the University of Detroit, Mercy, used this clinical and patient experience to suggest why high-quality trials had never been done: "They don't do research on things that are common knowledge."¹⁷⁰

How can we reconcile the experience of dentists and patients with the lack of high quality, long-term epidemiological evidence? We could dismiss these observations as correlation but not causation, but we could also view the experience of dentists and patients as a form of observation that confirms the hypothesis that flossing helps prevent gum disease. In other words, as in the case of the Pill, we can accept the experience of patients and clinicians as evidence, even if the explanation for that evidence is not fully clear. Put another way, we can reject the rejection of this evidence as "merely" anecdotal, and insist that these are *case reports*, and *n* is far greater than 1. Moreover, as with the Pill, we can consider mechanism.¹⁷¹ There is in fact good reason to think that dental floss is likely to be beneficial—that these correlations are in fact causally related—because it removes plaque and tartar that can contribute to gum disease, which, over time, can lead to tooth loss. Just as there is a known mechanism that links estrogen to serotonin and mood control, there is a known mechanism by which flossing is expected to prevent tooth loss.

This was explained by Dr. Sebastian G. Ciancio, the chairman of the department of periodontology at the University at Buffalo: "Gum inflammation progresses to periodontitis, which is bone

loss, so the logic is if we can reduce gingivitis, we'll reduce the progression to bone loss." But severe periodontal disease may take five to twenty years to develop, so this effect cannot be demonstrated in a clinical trial that lasts only weeks or months. Dr. Wayne Aldredge, president of the American Academy of Periodontology put it this way: "It's a very insidious, slow, bone-melting disease. . . . You don't know if you'll develop periodontal disease, and you can find out too late."¹⁷² In short, the "gold standard" of the randomized clinical trial is unable to reveal the benefits that periodontists predict. The clinical trials that have been undertaken were not the right tools for addressing that question.

The term "gold standard" should remind us that there are silver and bronze standards, too—or at least there should be. As Nancy Cartwright and Jeremy Hardie have argued, the ideal of a uniform gold standard is misguided: No one would use gold for household pipes; it is too expensive. Nor would we use gold for cooking knives: it is too soft.¹⁷³ The best tool depends on the job, and that applies to intellectual jobs as well as industrial and household ones.

What would be the right tool to investigate dental floss? One might be a different sort of clinical trial. The American Dental Association notes that disappointing results might be the result of poor flossing, which, they noted, is a "technique sensitive intervention."¹⁷⁴ The *New York Times* concluded: "So maybe perfect flossing is effective. But scientists would be hard put to find anyone to test that theory."¹⁷⁵ With due respect, that is an ill-informed remark, because scientists *have* tested that theory. The clinical trials reviewed by the Cochranes did not examine the impact of flossing technique, but a review of six trials in which professionals flossed the teeth of children on school days for almost two years, saw a *40% reduction in the risk of cavities*.¹⁷⁶

That is a huge effect. So consider this alternative headline: “A New Job Opportunity: Science Shows the Need for Professional Flossers.” Imagine the social change that might have ensued and the employment opportunities created. On our way to work, instead of stopping at Peets or Starbucks for a quick latte or a Drybar for a blow-out, we could stop at a flossing bar for a five-minute professional floss.

What Does It Take to Produce Reliable Knowledge?

There are many ways in which scientists can fail to live up to their own standards, as well as ways in which the standards they set can be unhelpful, incomplete, inadequate, or inappropriate to a particular situation. Still, I believe there are some themes that we may glean from these diverse cases. They are: (1) consensus, (2) method, (3) evidence, (4) values, and (5) humility.

Consensus

In chapter 1, we saw that historians, philosophers, and sociologists have come to focus on scientific consensus because there is no independent measure of what scientific knowledge is. We cannot identify science by any unique method. We can only identify claims as being scientific based on their provenance, that is to say, based on the way they were established and by whom. Scientific facts are claims about which scientists have come to agreement.

Some skeptics have used this argument to try to discredit contemporary science, claiming that there was a consensus

supporting eugenics or rejecting continental drift.¹⁷⁷ This, they argue, proves that scientific consensus is an insufficient basis to command our trust, a faulty foundation for decision-making. But these claims are misplaced: scientists did *not* have a consensus about eugenics or continental drift. Social scientists, socialist geneticists, and some mainstream geneticists critiqued eugenics; the rejection of continental drift was a distinctly American affair. (Europeans for the most part withheld judgment, which is a different thing.) Nor was there a consensus over the Limited Energy Theory, the Pill, or dental floss. Gynecologists liked the Pill for its efficacy; psychiatrists were concerned about its psychological health side effects. Short-term epidemiological studies fail to find strong evidence for beneficial effects of flossing, but nearly all clinicians observe benefits. And leading women physicians pointed out the obvious flaws in the Limited Energy Theory.

A key finding from historical inquiry into these episodes, then, is that in all of these cases there was significant, important, and *empirically informed dissent within the scientific community*. When we see disputes within scientific communities across geographic, disciplinary, or other gaps, this should command our attention. Debates may arise between different types of scientific experts examining a common topic—psychiatrists and gynecologists—or between different types of people—male doctors and female ones—or between scientists in the same field bringing different background assumptions and values. These debates occur because different groups of scientists are emphasizing different bodies of evidence, highlighting different aspects of those bodies of evidence, or bringing different values and background assumptions into the interpretation of evidence.

Scientific consensus is hard to come by. This is an underappreciated fact. Therefore, in any debate, it is crucially important

that we evaluate whether an expert consensus prevails or not. In 2004 I wrote a paper asking: Is there a scientific consensus on anthropogenic climate change? I had discovered that no one had analyzed the scientific literature with this question in mind and it seemed to me that any discussion of a mooted question should begin with an analysis of this sort.

In a recent issue of the *Hedgehog Review*, the editors wrote that “when we hear conflicting scientific pronouncements being issued on almost any subject (climate change, diet, vaccination) . . . it is not hard to see why science, and particularly scientific authority, has become the target of heated contestation and debate.”¹⁷⁸ This claim is wrong on two counts. First, it has cause and effect backward. These issues are contested because various groups—the tobacco and fossil fuels industries, advocates of deregulation, parents of autistic children who feel inadequately supported, some evangelical Christians—are unhappy with scientific authority. Some of them *want* science to be devalued.

Because science has challenged their interests or beliefs, they challenge science. Contestation is the outcome of a conflict about authority. Second, these are not conflicting *scientific* pronouncements. On most of the scientific issues that are highly contested in American culture—evolution, vaccine safety, climate change—there is a scientific consensus. What is lacking is cultural acceptance by parties who have found a way to challenge the science. This is the source of the contestation, not conflicting positions within the scientific community. Political and cultural debate is by no means illegitimate, but political debate masquerading as science is dishonest. It has led to the sort of confusion displayed by the editors of *Hedgehog Review* and many others.

Consensus analysis of peer-reviewed literature (as I have done) is a means to determine whether scientists agree. If they do, then we can take the next step to identify who is contesting their findings, and why. In our book, *Merchants of Doubt*, Erik Conway and I were able to show that climate science was being contested by the fossil fuel industry, whose economic interests were threatened, and by Libertarian think tanks and conservative scientists whose political beliefs were challenged. Rather than admit this, they challenged the science as a means to protect their economic interests and political commitments.

If there is informed dissent within the scientific community, more (scientific) research may well be needed. If, however, the dissent is emanating from outside the relevant expert scientific community, then we have a different issue at stake. In the latter case, more scientific research is unlikely to settle the matter, because non-scientific objections are not driven by scientific considerations and therefore will not be resolved by more scientific information.

This is not to say that non-scientific objections are invalid, but only that they should not be confused with “scientific pronouncements.” There can be important moral objections to social programs based on science, even if the underlying science is legitimate. And, as the contraceptive pill case illustrates, relevant information can emerge from outside specialist communities. My intent in presenting the Pill case was not to say that patients were necessarily correct, but rather that they had relevant information that should not have been disparaged simply because it came in the form of self-reporting.

How do we judge if non-experts have relevant, useful, and accurate information? This is not an easy question to answer. We have clear markers of scientific training and expertise: higher education, membership in scientific and learned societies,

records of publication and research grants, H-indices, awards and prizes, and the like. Scientists know who their scientific colleagues are and what their track records look like. Scientists (for the most part) know which journals have rigorous peer review and which do not.

Judging information from outside the expert world, however, is a different and trickier matter.

Scholars have identified several categories worthy of attention. One is other professionals who have relevant information. This could include nurses and midwives, for example, who have direct contact with patients and may differ from physicians on questions such as pain management.¹⁷⁹ A second category is people who may not have professional training, but whose daily experiences may lead them to relevant knowledge and understandings, such as farmers and fishermen.¹⁸⁰ We might say that these people have daily “on the ground” experience, and therefore may see things that scientific experts, for whatever reason, have missed. (Earth scientists call this “ground truth,” in this case referring to what geologists on the ground see and therefore know about, as compared with evidence, for example, from satellite remote sensing.) As Brian Wynne has stressed, the non-expert world is not “epistemically vacuous.”¹⁸¹

A third category is what Marjorie Garber has called “amateur professionals.”¹⁸² These are people—perhaps independent scholars or scholars from other fields—who have educated themselves on a particular subject. Developing expertise outside of conventional avenues of credentialism is certainly possible (although if a scholar from one field moves into another, they can establish credentials by publishing). A fourth category is citizen scientists: people who earn their living in other ways, but participate in science out of love or interest. In some domains—astronomy, entomology, ornithology, and the search for

extraterrestrial life—citizen scientists have played significant roles in observing things that professionals do not have the time, money, or human resources to track.

People in all these categories may have knowledge relevant to a particular scientific question. They have a known relation to their object of study and basis for claiming a role in scientific conversations bearing on that study. Where their experience and expertise overlap with scientific expertise, we should pay attention and not automatically discount what they have to say, nor assume that their claims are necessarily in conflict with those of scientific experts.¹⁸³ Often expert and lay perspectives can be reconciled or seen as complementary. But, again to draw on Wynne, we should not misunderstand a claim for recognition of these knowledge categories as a claim for their intellectual superiority or equivalence.¹⁸⁴ Just because someone is close to an issue does not mean he or she understands it; conventional notions of objectivity assume distance for just this reason. Parents of autistic children will have detailed knowledge of their children's conditions, but this does not mean that they are in a position to judge what caused it.¹⁸⁵

Respecting professional diversity and lay expertise is also a different matter from heeding “dissent” from people with no credible claim to expertise—celebrities, K-Street lobbyists, or the op-ed writers of the *Wall Street Journal* or the *New York Times*. When people without relevant expertise criticize science, we should consider the possibility that something fishy may be going on. If people are attacking science, there is something at stake, but it is not necessarily something scientific. Indeed, it is probably not.

An abundant literature now documents how various parties have tried to create the impression of scientific uncertainty and debate as a means to block public policy that conflicts with their

political, economic, and ideological interests.¹⁸⁶ But these are not the only reasons that people attack science, insist there is no consensus, or promote alternative theories. People attack science to get attention, to sell alternative therapies, or because they are frustrated that science doesn't have an answer to a problem that affects them.¹⁸⁷ But it is a relatively simple matter to distinguish between scientific debate and other stuff: Scientific debate takes place within the halls of science and on the pages of academic journals; other stuff takes place in other places. Political debate takes place on the op-ed pages of newspapers. Grievances can be aired anywhere. Sadness, isolation, and frustration make people lash out. But if, like the editors of *Hedgehog Review*, we mischaracterize political debate, industry shilling, or social disaffection as scientific controversy, then our attempts to remedy the situation will almost certainly fail.

Method

In the episodes we have been discussing, problems arose because scientists discounted evidence that failed to meet their methodological preferences. In the early twentieth century, geologists rejected continental drift, because it did not fit their inductive methodological standards. Charles Davenport was attracted to eugenics in part because he wanted to make biology more rigorous by making it more quantitative. In the cases of dental flossing and the Pill, scientists discounted clinical evidence because of a lack of robust epidemiological data. This last point is particularly important, because in the contemporary world, we have come to rely on statistical analysis to a degree that has led many people to ignore important evidence, including the evidence of everyday experience that hormones affect our moods

and flossing makes our gums less bloody. This doesn't mean that everyday experience is superior to statistics; it is not. Good statistical studies are an essential part of modern science. It just means that statistics, like any tool, don't work well in all cases and conditions and like any tool can be used well or badly (Krosnick, this volume).

A focus on one method above all others is a kind of fetish. These cases suggest that some of the historical examples of "science gone awry" arose from what I designate *methodological fetishism*. These are situations where investigators privileged a particular method and ignored or discounted evidence obtained by other methods, which, if heeded, could have changed their minds.

Experience and observation come in many forms. A good deal of evidence is imperfect, but that is no reason to ignore it. It is foolish to discount evidence that comes in messy forms simply because they are messy, particularly when the preferred methodological standard is difficult to meet or unsuitable to the question at hand. Randomized double-blind trials are powerful when they can be done, but when they cannot we should not throw our hands up and suggest we know nothing. There is no way to know how a drug makes people feel without asking about their feelings. There is no way to do a double-blind trial of flossing or nutrition. Imperfect information is still information.

When we have independent information about causes and mechanisms—such as knowing that flossing reduces gingivitis, that hormonal contraception can affect serotonin receptors (and vice versa: that anti-depressants that target serotonin uptake can affect hormones), or that greenhouse gases alter the radiative balance of the planet—this information is crucial to helping us evaluate claims when our statistical information is noisy,

inadequate, or incomplete. Mechanisms matter. When we know something about relevant mechanisms, there is no reason to play dumb.¹⁸⁸

Evidence

It seems obvious to say, but scientific theories should be based on evidence. However, in two of the cases here, we saw scientists making affirmative claims on the basis of little or scant evidence. Dr. Edward Clarke built an ambitious and socially consequential theory about female capacity on the basis of seven patients. Critics at the time noticed not only that his data base was scant, but also that it was biased: his patients were all young women who had come to him suffering anxiety, backache, headache, and anemia, and who he described as pursuing educational or professional goals in a “man’s way.”¹⁸⁹ (This included an actress and a bookkeeper; only one was actually a student in a woman’s college.)

In hindsight it is more than obvious that the symptoms he described—headache, backache, anxiety—could have any one of a number of causes. They are also afflictions that often occur in men, yet Clarke offered no evidence that these ills were more common in women, or more common among women who were educated than in those who were not. He presented his theory in the framework of hypothetico-deductivism, yet he failed to pursue the required next step: to determine if his deduction were true. Most conspicuously, he provided no evidence that these women’s reproductive systems were weakened or that their fertility had been decreased. When women physicians and educators pointed out these flaws, Clarke ignored them. His theory was

elegant, but could only be sustained by ignoring evidence available to him at that time.

Values

The role of values in science is a much-mooted issue, and the stories told here show how easily prevailing social prejudices may be instantiated into scientific theory. Scientists have not always been on the side of the angels. Anyone who values science must acknowledge this.

The traditional impulse of scientists has been to say that in cases such as eugenics, science was “distorted” by values. But historians of science, particularly but not only feminists, have noted the ways in which values are broadly infused into scientific life and not always in adverse ways. It is true that racial and ethnic prejudice infused eugenic thinking, and the sexism in Edward Clarke’s work is not difficult to discern. But values also played a role in the critiques of those theories. Socialist values were crucial to some geneticists’ critique of eugenic thinking; feminist values informed Mary Putnam Jacobi’s identification of the theoretical and empirical inadequacies of the Limited Energy Theory. Barbara Seaman was a journalist, not a scientist, but her feminist values motivated her to follow up on the “anecdotes” she had heard, to seek out the doctors who could confirm the substance in these stories, and to highlight information that some doctors were discounting.

This, it seems to me, is the most important argument for diversity in science, and for diversity in intellectual life in general. A homogenous community will be hard-pressed to realize which of its assumptions are warranted by evidence and which

are not. After all, just as it is hard to hear your own accent, it is hard to identify prejudices that you share. A community with diverse values is more likely to identify and challenge prejudicial beliefs embedded in, or masquerading as, scientific theory.

Critics of efforts to make science more diverse sometimes insist that the only relevant standard in science is “excellence.”¹⁹⁰ Science, they insist, is a meritocracy in which demographic considerations are misplaced. These critics seem to think that calls for diversity are *merely* political; that there is no intellectual value in building diverse communities. The stories told here refute that idea. They suggest that diversity can result in a more rigorous intellectual outcome by fostering critical interrogations that reveal embedded social prejudice.

Admittedly, this claim cannot be proved, because in science we have no independent metric to judge epistemic success. We cannot stand apart from our truth claims and independently determine if they are true; nor can we compare the “truth-production” of more and less diverse communities. But in a domain where there are metrics of success—namely, business—rigorous studies have demonstrated that diverse teams yield better outcomes, in terms both of qualitative values such as creativity and quantitative outcomes such as sales. If we know that diversity is beneficial in the commercial workplace, why would we not presume that it would be beneficial in the intellectual workplace as well? Moreover, we saw in chapter 1 that there is an epistemological basis for presuming that diversity does benefit science. The examples presented in this chapter support that claim. Thus we may conclude that scientific communities that are “politically correct”—in the sense of taking seriously the value of diversity—are more likely to yield work that is scientifically correct.

Considering the role of values also helps explain what we could call the *misapplication of theory* and the *asymmetry of application*. In hindsight, there is an obvious theoretical flaw in Clarke's work: while presented as an application of thermodynamics, it was actually a misapplication of the theory because conservation of energy applies to closed systems. The human body is not a closed system: it is sustained and supported through nutrition. Life is possible *because* organisms are not closed systems, so Clarke's use of thermodynamics was logically fallacious. It was also asymmetrical, because for some odd reason it only applied to women. Admittedly, Clarke had an explanation for this: he suggested that the female contribution to reproduction was uniquely demanding, and he did allow the possibility that overexertion could be harmful to both boys and men as well. Yet, while stressing the claim that if a woman was educated, her uterus would shrink, he evidently never paused to ask: if men were educated, what part of their anatomy would shrink?

Eugenicists likewise applied their theories asymmetrically. As Muller and Haldane stressed, the target of their attention was the working class. There were drunkards, gamblers, and layabouts among the wealthy, yet few eugenicists advocated sterilization of underperforming rich white men.

Humility

If the history of science teaches anything, it is humility. Smart, hard-working, and well-intentioned scientists in the past have drawn conclusions that we now view as incorrect. They have allowed crude social prejudice to inform their scientific thinking. They have ignored or neglected evidence that was readily

available. They have become fetishists about method. And they have successfully persuaded their colleagues to take positions that in hindsight we see as incorrect, immoral, or both.

Many of the scientists in these stories were driven by a genuine desire to do good: to promote an effective means of birth control, for example, or protect women from something they honestly believed would harm them. But their failings are a reminder that anyone engaged in scientific work should strive to develop a healthy sense of self-skepticism. Edward Clarke was a supremely confident man. So was Charles Davenport. So were many of the early advocates of the contraceptive pill. Wegener's critics accused him of "auto-intoxication," and I dare say we have all encountered scientists who are overly enamored of themselves. It seems to me that individual scientists, if they care about truth, should be mindful of this problem and not ride roughshod over their colleagues.

If the social view of science is correct, however, then it may not matter too much if a particular individual is auto-intoxicated. Inevitably there will be arrogant individuals in science, but so long as the community is diverse and alternative views are available, and so long as the community as a whole finds the means for all its members to be heard, things are likely to go well. Nonetheless, collectively scientists should still bear in mind that—whatever conclusions they come to and however they come to them—even with the best practices and the best of intentions, there is always the possibility of being wrong, and sometimes seriously so.

Conclusion: Science as a Form of Pascal's Wager

In evaluating a scientific claim that has social, political, or personal consequences there is one more question that needs to be considered: What are the stakes of being wrong in either direction? What is the risk of accepting a claim that turns out to be false versus the risk of rejecting a claim that turns out to be true?

Knowing there is a risk of depression, if a healthy woman decides to take the Pill she can quickly stop taking it if the risk materializes. Pill-induced depression generally clears up quickly, so for many women the risk is modest and worth taking. Similarly, dental floss is cheap and only takes a few minutes a day to use. If it turns out to have little benefit, little has been lost. But some issues are not so easily resolved.

Consider anthropogenic climate change. Despite fifty years of sustained scientific work, communicated in tens of thousands of peer-reviewed scientific papers and many hundreds of governmental and nongovernmental reports, many people in the United States are still skeptical of the reality of climate change and the human role in it. The president has doubted it, as have members of Congress, business leaders, and the editorial page of the *Wall Street Journal*. Rejecting centuries of well-established physical theory and reams of empirical evidence regarding matters such as sea level rise and the intensification of extreme weather events, others have suggested that while anthropogenic climate change might be a real thing, it is inconsequential and might even be beneficial.¹⁹¹

As a historian of science, mindful of the Limited Energy Theory and eugenics and the history of hormonal

contraception—mindful of the difficulties of evaluating dental floss—and above all mindful of the political ideals that geologists brought to bear in evaluating continental drift—I have never *assumed* that trust in science is always or even usually warranted. I have always felt that it is fair to ask: What is the basis for any scientific claim? *Should* we trust scientists?

We cannot eliminate the role of trust in science, but scientists should not expect us to accept their claims solely on trust. Scientists must be prepared to explain the basis of their claims and be open to the possibility that they might be wrongly dismissing or discounting evidence. If someone—be it a fellow scientist, an amateur professional, a journalist, or an informed citizen—has a credible case that evidence is being discounted or weighed asymmetrically, this should concern us. Scientists need to remain open to the possibility that they have made a mistake or missed something significant.¹⁹² The key point is that the basis for our trust is not in scientists—as wise or upright individuals—but in science as a social process that rigorously vets claims.

This does not mean that scientists must spend time and energy continuing to prove and reprove conclusions that have already been established beyond a reasonable doubt, nor refuting claims that have been refuted. As Thomas Kuhn argued more than half a century ago, to the extent that science can be said to progress, it is because scientists have mechanisms by which they reach agreement and *then move on*. Perhaps the most salient aspect of the continental drift debate is that it *was* reopened, which occurred when a new generation of scientists developed new lines of pertinent evidence.¹⁹³

We can reframe this problem in terms of Pascal's Wager. No matter how well-established scientific knowledge is—no matter how strong the expert consensus—there will always be residual

uncertainty. For this reason, if our scientific knowledge is being challenged (for whatever reason), we might take a lead from Pascal and ask: What are the relative risks of ignoring scientific claims that turn out to be true versus acting on claims that turn out to be false?¹⁹⁴ The risks of not flossing are real, but not inordinate. The risks of not acting on the scientific evidence of climate change are inordinate.¹⁹⁵

Admittedly, the advocates of eugenic social policies considered the risks of not implementing eugenic social policies to be extremely high. That, of course, was their interpretation of the scientific evidence. But as we have seen, there was no consensus on that *evidence*. So we are back to the importance of consensus. If we can demonstrate that there is no consensus among relevant experts, then it becomes clear that we have a weak basis for public policy. This is the reason why the tobacco industry tried for so long to claim that the science regarding the harms of tobacco was unsettled: if it really had been, then they might have been right to insist that tobacco control was premature.¹⁹⁶ Similarly, if there were no scientific consensus about anthropogenic climate change, then the fossil fuel industry and Libertarian think tanks might be right to ask for more research. This is why consensus studies are relevant and important: Knowing there is a consensus does not tell us what to do about a problem like climate change, but it does tell us that we almost certainly have a problem.¹⁹⁷

If we can establish that there is a consensus of relevant experts, then what? Can we be confident in accepting their conclusions and using them to make decisions? My answer is a qualified yes. Yes, *if* the community is working as it ideally should. That is a substantial qualification. As Brian Wynne has put it, if we are to respect and trust science, then “it becomes evident why the quality of its institutional forms—of organization, control and

social relations—is not just an optional embellishment of science in public life, but an essential component of critical social and cultural evaluation.”¹⁹⁸

The history of science shows that there is no guarantee that the ideals of an open, diverse community, participating in transformative interrogation, will be achieved. Often it will not be (although the consequences of failing to meet this ideal may not always be profound or even significant). Historian Laura Stark notes that the National Bioethics Advisory Commission recommends that one-quarter of the members of the boards that review human subjects research should not be affiliated with the institution at which the research is being done, but this goal is rarely achieved.¹⁹⁹

How do we determine if a scientific community is sufficiently diverse, self-critical, and open to alternatives, particularly in the early stages of investigations when it is important not to close off avenues prematurely? How do we evaluate the quality of its institutional forms? We must examine each case on an individual basis. Many scientists were wrong about continental drift, but that does not mean that a different group of scientists are wrong today about climate change. They may be or they may not be. We cannot assert either position *a priori*.

If we can establish that there is a consensus among the community of qualified experts, then we may also want to ask

- Do the individuals in the community bring to bear different perspectives? Do they represent a range of perspectives in terms of ideas, theoretical commitments, methodological preferences, and personal values?
- Have different methods been applied and diverse lines of evidence considered?

- Has there been ample opportunity for dissenting views to be heard, considered, and weighed?
- Is the community open to new information and able to be self-critical?
- Is the community demographically diverse: in terms of age, gender, race, ethnicity, sexuality, country of origin, and the like?

This latter point needs further explication. Scientific training is intended to eliminate personal bias, but all the available evidence suggests that it does not and probably cannot. Diversity is a means to correct for the inevitability of personal bias. But what is the argument for *demographic* diversity? Isn't the point really the need for *perspectival* diversity?

The best answer to this question is that demographic diversity is a proxy for perspectival diversity, or, better, a means to that end. A group of white, middle-aged, heterosexual men may have diverse views on many issues, but they may also have blind spots, for example, with respect to gender or sexuality. Adding women or queer individuals to the group can be a way of introducing perspectives that would otherwise be missed.

This is the essential point of standpoint epistemology, raised particularly by the philosopher Sandra Harding (chapter 1). Our perspectives depend to a great extent on our life experience, so a community of all men—or all women for that matter—is likely to have a narrower range of experience and therefore a narrower range of perspectives than a mixed one. Evidence from the commercial world supports this point. Studies of gender diversity in the workplace show that adding women in leadership positions increases company profitability—but only up to a point. That point is about 60%. If a company's leadership becomes all or nearly all female, then the “diversity bonus” begins to decline, as indeed, if the argument here is correct, it should.²⁰⁰

It may not always be easy to answer the questions posed above in the affirmative, but it is often obvious if the answer to any of them is negative. Moreover, we may (and likely will!) identify individuals in the community who are arrogant, closed-minded, and self-important, but on the social view of epistemology the behavior of any particular individual is not what matters. What matters is that the group as a whole includes enough diversity and maintains sufficient channels for open discussion that new evidence and new ideas have a fair chance of a fair hearing.

The philosopher Heather Douglas has argued that when the consequences of our scientific conclusions are non-epistemic—i.e., when they are moral, ethical, political, or economic—it is almost inevitable that our values will creep into our judgments of evidence.²⁰¹ (Liberals, for example, may have been quicker to accept the scientific evidence of climate change, because they were more comfortable with its implied consequences for government intervention in the marketplace.) Therefore, the more socially sensitive the issue being examined, the more urgent it is that the community examining it be open and diverse.

But sometimes an issue that appears to be purely epistemic isn't, and scientists may claim they are evaluating an issue on solely epistemic grounds even when they are not.²⁰² This suggests that no matter what the topic, it behooves the scientific community to pay attention to diversity and openness in its ranks and to remain open to new ideas, particularly when they are supported by empirical evidence or novel theoretical concepts. It means, for example, that in considering dissenting views in grant proposals or peer-reviewed papers, it is probably better to err on the side of tolerance than critique. Many scientists consider it extremely important to be intellectually tough, if not actually rough, but sometimes toughness can have the unintended effect of shutting down colleagues, particularly those who are young,

shy, or inexperienced. It is important to be tough, but it may be more important to be *open*.

In chapter 1, I argued that the advocates of the Extended Evolutionary Synthesis are receiving a thorough hearing, if not always a polite one. Something similar happened to Alfred Wegener. He was not a neglected genius: his papers were published in peer-reviewed journals and his work had a hearing—albeit not always a gracious one. The socialist opponents of eugenics likewise got their manifesto published in *Nature*.²⁰³ None of these dissenters was “shut down” by the scientific hierarchies of their day.

Fast forward to the present: Many AIDS researchers lament Peter H. Duesberg, the University of California molecular biologist who does not accept that AIDS is caused by a virus. By his own account he has “challenged the virus-AIDS hypothesis in the pages of such journals as *Cancer Research*, *Lancet*, *Proceedings of the National Academy of Sciences*, *Science*, *Nature*, *Journal of AIDS*, *AIDs Forschung*, *Biomedicine and Pharmacotherapeutics*, *New England Journal of Medicine* and *Research in Immunology*.”²⁰⁴ Whether he is right or wrong, tolerated or vilified, the fact is that he has had a hearing, and at the highest level of American and international science. His colleagues have not shut him down; they have published his work and considered his arguments. But they remained unconvinced.²⁰⁵ There is a difference between being suppressed and losing a debate. Sometimes a “skeptic” is just a sore loser.