– 2 –

Error Itself

The conundrum of defining error • a comedy of errors? • right and wrong • bloopers, blunders and big whoppers • epistemic posture • error and uncertainty • summary

Discussion of error can potentially be quite confusing. Sometimes even paradoxical. For example, error seems to be a lapse in knowledge. But if you *know* something is an error, then it cannot be an error any more, because you know all about it. Perhaps you can only say that once upon a time you were mistaken. Ironically, at that earlier point, you could not possibly have called it an error, because you did not know any better, and were unable to identify it. So when did (or does) the error exist? Puzzling, isn't it?

Consider the case of the planet Pluto. Or *former* planet Pluto? Or the *planetoid/dwarf planet* Pluto, once classified as a planet? Gravitational calculations in the early 1900s indicated that there should be another planet orbiting beyond Neptune. An active search by Clyde Tombaugh revealed in 1930 the presence of such a massive body, which was named Pluto. In 1978, it was even discovered to have its own moon. Not all planets remained planets. In the early 1800s the first asteroid to be found, Ceres, was considered a planet. With the discovery of other asteroids, its status changed. In a similar way, in the 1990s, with more powerful telescopes, more observations and the aid of computers, astronomers found more objects not quite planet size in the outer Solar System. Then more. Meanwhile, the original calculations used to hypothesize Pluto were discredited, as based on incorrect values for Neptune's mass and position. Pluto's status was challenged, then formally demoted by an assembly of astronomers in 2006. Was Pluto never a planet, even though it was called one for decades? When did the error occur? In 1930, when it was "wrongly" designated a planet? Or only in 2006, when its status was changed in retrospect? Was the mistake in the original classification of Pluto (which seemed to follow all the rules), the "failure" to be cautionary about it, the faulty calculations, or perhaps the very concept of a planet, being too inclusive or vague in the early 1900s? Even such a relatively simple case illustrates the puzzle of defining errors. [Weintraub, 2007]

But the apparent paradox in ascertaining when the error occurred also highlights the core problem. How can we trust the justification of scientific claims if those justifications are inherently unreliable? How can scientific claims hold any authority if they are subject to unanticipated change, especially in wholly unpredictable ways? We need to understand the nature of those changes. Why do they occur? *How* do they occur? How were earlier scientists "handicapped" in their abilities to notice the errors (if that is an appropriate expression)? Also, how can scientists deal with the prospect of errors? Are there any methodological tools to limit the frequency or lessen the scope of errors? Can history guide us?

The central problem is thus the creation of knowledge and nature of justification. This field of study is known as *epistemics*. How can there be holes or blind spots in our reasoning? How might our reasoning accordingly be made more complete or secure? How do we develop trustworthy, or reliable, knowledge? In this book, then, I focus not just on the errant scientific claims themselves, but on how the *justification* for such claims can change. Accordingly, one may establish a benchmark definition (at least initially, to guide analysis in this book):

An error is a scientific claim whose justification is later found to *be unwarranted*.

That is, an error is a claim once deemed "reasonable" and later found to be unsupported, or unreasonable. One major question here will be how the interpretation of good reasons can change. Namely, in what ways might our justifications be incomplete or ultimately unwarranted, without our awareness? [see also Hon, 1987, p. 382]

This simple characterization of error does not necessarily exhaust the task of conceptualizing error on a foundational level. A full discussion also involves a hodgepodge of issues. For example, we need to identify common tendencies in how we think about error and how those intuitions shape or bias our thinking. We also need to establish clear language and carefully articulate our basic concepts. That is, although it is not difficult to acknowledge that errors happen all the time, even among famous and undoubtedly competent scientists (Chapter 1), we do not always seem equipped to talk about it clearly. One distraction is the relation between normative and descriptive perspectives (the section below on "Right and Wrong"). At what scale do errors occur? (section on "Bloopers, Blunders and Big Whoppers"). Another key factor is the level of commitment to a particular claim and its justification. Finally, one may consider the distinction between error and uncertainty (simply not knowing at all).

All this reflection on basic conceptualizations might easily impress some as mere academic fussing. The aim, however, is to establish a foundation so that one can more confidently compile a repertoire of general error types and develop a taxonomy of them (Chapter 3) and clearly conceive how to analyze and remedy them (Chapter 4B). Thus, this chapter adopts a more formal, "definitional" posture, largely different from the others. Readers anxious to grapple with the sources of error and their remedy may prefer to forego this chapter and its more academic tone. Others may find that the discussion of "foundational" concepts, while essential, might make more sense only after a more thorough discussion of concrete cases, leading them to engage this chapter only after fuller familiarity with what is at stake.

A Comedy of Errors?

Shakespeare had a good sense of comedy. In one play, he envisions twins separated at an early age (and raised in rival cities) converging in the same place unbeknownst to each other. The scenario is ripe for mistaken identity, and Shakespeare does not disappoint. To compound the laughs, the men are accompanied by their twin servants, likewise separated in infancy. In successive encounters on the street, each brother and servant is taken for the other. The mismatch of assumptions and contexts leads to hilarious misinterpretations. Miscommunications follow. Missives are misdelivered. Their messages are misconstrued. One mistake leads to yet another, and every effort to correct the situation only makes matters worse. Nowadays, it is a stock formula of television sit-coms: mistake compounded upon mistake, all laughable. In the end, each brother discovers the other, lost parents are revealed, the fractured family is reunited, and all becomes happily resolved. Shakespeare called his play, appropriately enough, *A Comedy of Errors*.

For many people — scientists and science spectators alike — errors in science are just like the characters' errors in Shakespeare's play: pleasant diversions for amusement. Silly and whimsical – and due to a blindness that is easily avoided. Philosopher of science Robert Crease contends that scientific errors are just "the stuff of lunch table anecdotes." We laugh at the stories of human foibles or folly. These perspectives are vividly expressed in the titles of several popular books on errors in science: *Scientific Blunders, Brilliant Blunders, Shocking Science,* *Nonsense on Stilts, Follies and Fallacies in Medicine, The Undergrowth of Science, and Discarded Science*.

[Crease 1993, p. 107; Nichols, 1984; Skrabanek & McCormick, 1989; Kelly & Parker, 1996; Youngson, 1998; Gratzer, 2000; Grant, 2006; Pigliucci, 2010; Livio, 2014]

Science is usually pretty serious stuff. So, what makes the errors funny? Errors seem to become entertaining diversions in part by being diversions in another sense, as well. They reflect a *diversion* from the course of "true" science, a detour from the presumed path of progress. In popular lore, scientists presumably follow the scientific method, leading them inexorably to the truth about nature. They are not supposed to make mistakes. When they do, the incongruity, the departure from the expected norm, is amusing. Like an unanticipated pie in the face or a pratfall. Diversion from diversions.

So, for a few moments perhaps, we might divert ourselves with some apparently silly errors. At the same time, such cases often parallel or resonate with more serious cases of error. Our light-hearted "diversion" might also become an informative excursion. One may reflect on the ironies of error, finding a pleasant opportunity to learn more about their implications and meaning.

Consider, for example, the story of paleontologist Edward Cope. In 1868, Cope received a remarkable, complete fossil of a new species, a type of aquatic plesiosaur that he dubbed an elasmosaur. As a major find, it would surely help his early career and launch him into notoriety. He soon sent the reconstruction off for publication and began to prepare the specimen for display. Before the book was printed, though, triumph turned to embarrassment. Colleague and friend Charles Marsh came to see the specimen in 1869, but promptly declared that the vertebrae were reversed and the whole specimen backwards! Apparently, Cope could not even tell which end was up. Cope took offense. Museum curator Joseph Leidy was summoned to adjudicate the dispute. Skull in hand, he went to the "tail" end of the specimen and fit it neatly into the final vertebra. Indeed, a small section of the skull had broken off and was embedded in the rock with the "last" vertebra, confirming the match. In addition, these two end vertebrae were fused, as was typical of the first two neck vertebrae in other plesiosaurs. Marsh was right. Cope scrambled to buy up all the published copies of the error. But Leidy made the mistake public. The conflict opened what soon became a bitter and public professional rivalry between Cope and Marsh, known as the "Bone Wars." Marsh even returned to the episode decades later, profiling the error harshly in the *New York Herald* as a "taste" of Cope's alleged incompetence. Alas, poor Cope! Ironically, Leidy too had made the very same mistake himself on a related species decades earlier, only obvious once Cope's error had been detected. Indeed, Leidy's mistaken reconstruction had served as a model for Cope's own interpretation. Marsh would make his share of mistakes, too. In 1879, in order to display a complete skeleton of his newly discovered *Brontosaurus*, he added a spurious skull, collected four miles away, with no evidence of its relevance. The skull later turned out to be from a type of *Camarasaurus*, but had led to at least one other museum making a similar mistaken reconstruction. The mismatches also contributed to high profile controversies about the uniqueness of *Brontosaurus*, which lingered for over a century. Despite all the mutal allegations, no one seemed invulnerable to error in this case.

[Leidy, 1870; Marsh, 1890; Jaffe, 2000, pp. 12-19; Tschopp, et al., 2015]

A similar error in reconstructing fossils had occurred to Rembrandt Peale in the early 1800s. Peale had excavated a huge unfamiliar creature in the Hudson River Valley. It was elephant-like, but its teeth, with great mounds, did not seem designed for grinding plants. This extraordinary feature was occasion for naming the creature a *mastodon*, meaning "breast-teeth."

Seeing a mastodon tooth, Benjamin Franklin concluded that the creature was carnivorous. The famous British anatomist William Hunter concurred. The mastodon also had monumental curved tusks. When the specimen was first exhibited those tusks emerged from the eye sockets. Whoops. That was soon corrected. But there was still uncertainty. Peale recalled:

> When the skeleton was first erected, I was much at a loss how to dispose of the tusks; their sockets shewed that they grew out forwards, but did not indicate whether they were curved up or down.

Peale imagined the tusks curving upwards into the air and pointing backwards, useless for any function on the ground, then concluded:

> This position was evidently absurd; and there is infinitely more reason in supposing them to have been placed like those of the Walrus and probably for a similar purpose. (Peale, 1803, pp. 51-52)

So the tusks were mounted downwards, completing the image of a great and terrible carnivore. "In the inverted position of the tusks," Peale wrote, "he could have torn an animal to pieces held beneath his foot." A second specimen toured Europe, impressing all, and simultaneously conveying symbolically the new American nation as just as ferocious and grand. Back home at the Philadelphia Museum, the tusks remained in their position until at least 1822. In 1806, the great French anatomist Georges Cuvier weighed in with more comparisons: the mastodon was decidedly herbivorous. The tusks really curved upwards, not down. Oops again.

[Semonin, 2004]

Then there is the story of the "Little Green Men" sending signals to Earth from outer space! Graduate student Jocelyn Bell was helping to pioneer radio astronomy. Since childhood she had loved the stars, but couldn't stay awake at night. By studying radio waves from space, however, she could pursue her passion by working during the day. Her work involved reviewing the daily charts recording the radio waves received by the antennae array. A rather tedious and mundane task — perhaps all too typical of what many graduate students endure. One day, she found some "scruff" in the signal printout. She looked at the records and there it was, the same time the previous day, and the day before. To find out more, they sped up the recorder to spread out the signal and see the details. There were blips at consistent time intervals of just over one second. Well, at that speed, some human source had to be interfering with their measurements. The bane of radio astronomy. So, there was an error to identify and rectify. A closer examination indicated, though, that the signals were aligned with a position of the Earth's rotation, not the human clock. Bell and her advisor checked everything spaceward they could imagine. Rogue satellites? No. Radar bouncing off the moon? No. Reflections from the large metal-faced building next door? No. Something in the equipment itself? Nothing that anybody could find. The possibilities ran dry. Stumped, Bell mused over the absurdity that some intelligent beings in a faraway galaxy were trying to contact Earth with the pulsed signals. Obviously ridiculous, but what else could explain it? So she named the source "little green men," or LGM 1. Soon, Bell found a second pulsed signal coming from another corner of the sky, with a slightly different frequency. She worried. Signals from *two* distant civilizations, each targeting Earth? Eventually, the signals became evidence for a new class of stellar objects named, appropriately, *pulsars*. Bell's advisor, although not Bell (who had done most of the work and interpretation), received a Nobel Prize for the discovery. No little green men, after all.

[Burnell, 1977; Judson, 1980, pp. 91-100]

Another case always good for a chuckle is stumbling over the number of chromosomes in human cells. Today, even many non-scientists know that humans have 23 pairs of chromosomes, for a total of 46. But for decades the number was set at 48. It was published in textbooks. The error evokes a kind of dumbstruck credulity, succinctly expressed in the title of one historical essay: "*Can't anybody count?*" In the late 1800s and for several decades thereafter, the techniques for fixing cells and observing individual chromosomes were (one might generously concede) crude. So the highly variable chromosome counts – from 22 to 52 – might well have been expected. But by the early 1920s, they had refined techniques. They were using fresh tissue samples. They were drawing the chromosomes in particular cells aided by a camera lucida. They could certainly differentiate X and Y chromosomes for sex determination. Most important, the inevitability of individual mistakes was kept in check by agreement among many researchers. The community had reached a stable consensus at 48. Between 1930 and 1950, at least 8 further studies confirmed that number. When the revised count of 46 was introduced in 1956, however, others quickly confirmed the new number. A new consensus, just like that. Apparently, they had finally learned how to count. (I will revisit this case in Chapter 4a, with more details about why.) All that work for such an elementary task: counting chromosomes. Sigh.

[Kottler, 1974; Martin, 2004]

For a final example (here), we may amuse ourselves with the case of the tainted Perrier water. To start, one can imagine what might otherwise have been a routine workday at the Environmental Health Department in Mecklenburg County, North Carolina. The lab there was responsible for testing samples for water quality. Dilute the sample. Run it through the mass spectrometer. Look for any traces of possible organic pollutants. Go on to the next sample. But on January 19, 1990, the novice employee encountered a blip on the screen. It appeared even when the sample was just water. That was odd. An error somewhere, probably. But where? The lab checked the machine. They checked the utensils and then all the sources of potential contamination they could imagine. After a few days, they began to wonder about the water used to dilute the samples. Earlier, they purified their own water. "We used to take deionized water," the laboratory director explained, "boil it, purge it with nitrogen, seal it in a container, then cool it down. But this was fairly time-consuming for the amount we were using." So for five years they had been going to the grocery store and buying bottled water, "two or three bottles at a time." Not just any bottled water, though. Perrier. Premium, "all natural" water. "There were no organics in it, and," the director admitted plainly, "it worked." For a while, at least. Suddenly they were testing the "pure" Perrier itself, like one of their normal unknown samples. Surprise! It had trace quantities of benzene. Well, so much for the hallmark purity of Perrier! In sequel events, the contamination was reported to relevant agencies and Perrier recalled bottles worldwide. They lost sales for several months and ultimately their market edge. They tracked the error to a failure to change their own water filters at a prescribed time, which was easily fixed. Yes, the water renowned as "all-natural" had, from the very beginning, been chemically filtered and artificially carbonated. A little lab in rural North Carolina, eager to lighten their workload, encountered an inconvenient minor error and ended up exposing the pretenses of the world's premier bottled water brand.

[Borman, 1990; Brackett, 1990]

So, errors in science can be funny. One might think of Gaston Lachaille, the aging roué in the musical *Gigi*, memorably portrayed by Maurice Chevalier. In one song, he sits with his former lover, reminiscing about their romantic past. "That carriage ride," he rhapsodizes. "You walked me home," she replies gently. "You lost a glove." "I lost a comb." "Ah, yes!" he continues, "I remember it well." Even errors (as in Shakespeare's comedy, as much as science) can be an occasion for humor.

At the same time, errors – even small or apparently trivial errors – can have profound cultural consequences. The backwards elasmosaur or inverted tusks of mastodons seem harmless, and are funny in part for just that reason. But not all fossil reconstructions are so benign. In 1911, when paleontologist Marcellin Boule analyzed the first relatively complete skeleton of a Neanderthal, he created a visual mis-impression that has largely lasted to this day in popular culture. He interpreted the skull, with a prominent brow ridge and relatively low cranium, as indicative of low intelligence, more akin to apes than humans. That status was also reflected in the posture, a caricature well known today: bent knees, slumped shoulders, and forward head thrust. Boule did not seem to recognize that his specimen's osteoarthritis might have contributed to an atypical posture. Also, he gave the feet opposable toes with no apparent anatomical justification. To finish the picture (literally), the drawing he approved showed the ancient man clutching a club in one hand, a rock in the other. And hairy all over. From head to toe, Neanderthal was a feeble-minded brute. For decades, that image helped shape scientists' interpretations the human evolutionary tree, with progressive intelligence a pervasive theme. In popular culture, the early Neanderthal image still seems to epitomize the "caveman," and with it impressions that early humans (and primates, too) exhibited no behavioral sophistication, culture, or cooperative sociality. Intelligence alone made humans modern. A reconstruction of human ancestors is simultaneously a statement about identity. And the errors in the assumptions that guide its reconstruction can be unduly naturalized, apparently justifying those views inappropriately (for more on naturalizing as an error type, see Chapter 8).

[Smithsonian, 2015; Drell, 2000; Moser, 1992]

Likewise, the story of intelligent "little green men" sending signals to Earth, as a reluctant hypothesis of last resort, is amusing. However, an unqualified claim that organisms are evidence of so called "Intelligent Design," rather than of the non-teleological processes of evolution, is far less amusing. The poverty of natural theology as a scientific explanation was demonstrated in the mid-19th century. Presenting those outmoded ideas today as presumably justified reflects a profound disdain for the system of science, its principles of justification, and the role of expert consensus. Worse, of course, efforts continue to try to wedge those ill-founded claims into public education as embodying the principles of science. What is at stake in accepting these errors is not some jovial tea-time anecdote, but the very cultural respect for scientific knowledge.

Similarly, while the miscount of human chromosomes itself may seem trivial (with no major theoretical misinterpretations), the issue of errors in counting can be complex and have far reaching consequences. For example, ordinary health diagnoses, as well as medical research, still often depend on cell counts – say, how many red blood cells or how many white blood cells of each type. Yet a study many years ago indicated how unreliable human blood counts can be. Conventional counts were compared with those based on photos, where each cell was pierced with a punch to avoid double-counting. Over two-thirds of the time, the differences were unacceptable according to the prevailing quality standards for cross-check agreement. A recent analysis of cell counts of nerve cells is brain tissue noted how many issues of error persist, such as sample preparation and interpreting the cells in three dimensions. Grids are used to organize the cell count, but there are further problems with counting at edges. Some cell counting has been automated, but differentiating cell types or identifying dead cells still often requires human inspection. Efforts continue to find rigorous counting strategies, but each counter still needs to be trained – an inescapable source of potential error. Adopting more computerized methods does not necessarily solve the problem. For many years, the U.S. Environmental Protection Agency was trying to determine the number of deaths associated with certain levels of soot in the air. After five years of generating results, one investigator found an error in a default setting in the statistical program that crunched the raw numbers. That resulted in slightly "overcounting" the relevant mortality rates. Of course, a different error might have resulted in undercounting deaths. Given that the whole regulation of such fine particulate matter as a pollutant is highly contested by industry, such subtle changes in numbers can prove quite important, both in terms of policies and estimated health risks.

[Judson, 1980, pp. 171-72; Guillery, 2002; Kaiser, 2002]

The case of the tainted Perrier is amusing, too, in its many ironies. But not every case of contamination is resolved so quickly, nor with so little wasted scientific work. For example, in 2011 researchers had to scrap two years of work on a potential major breakthrough on chronic fatigue syndrome, or CFS. CFS is an aptly described disease, affecting perhaps as many as 3/4 million people in the U.S. Its cause remains unknown. Nor are there any definitive diagnostic tests or clear effective treatments. So researchers greeted with guarded optimism a report in the prestigious journal *Science* in 2009 that the disease was associated with the XMRV virus in two-thirds of patients (and not in controls). Those who reviewed the original paper for publication expressed strong concerns about possible contamination of blood samples, especially considering that the reported RNA sequence was similar to a cancer virus studied in the same lab. But further work seemed to address the concerns with multiple lines of evidence against contamination. Other labs were eager to capitalize on the promise. However, they were unable to find the virus present. Two years later, one of the original labs confirmed contamination with another virus. Commercial reagents (for PCR), due to their manufacturing process, contained XMRV and mouse DNA. The original results could thus be attributed to an assay that was not specific enough (and to cross-reactivity). The error was not without costs. At least 12 studies had tried to replicate the results, unsuccessfully, costing tens of millions of dollars. The consequences also reached beyond the admittedly risk-prone domain of research. Many patients were given antiviral medications, which generally produce strong side effects. Due to the perceived risk, they were also blocked from donating blood. XMRV has now joined a list of over 7 other viral agents once suspected of causing CFS.

[Alberts, 2012**;** Cohen & Esnerink, 2011; Kaiser, 2011;
Simmons, Glynn, Komaroff, et al., 2011]

So: some cases of scientific error just tickle our funny bone. Yes, we can appreciate the ironies between what was once known and what we now know with a hearty guffaw. Humor can be therapeutic. A healthy check on scientific hubris. The momentary pleasure should, ideally, also help us celebrate our own fallibility. So let us not succumb to scorn or an assumption that only *others* err. When we chuckle at a blooper, we should not miss the opportunity to reflect how the combination of events generated the humor. Good jokes are also occasions for learning.

A proper analysis of error in science, then, is neither fuel for entertainment nor evidence for self-aggrandizement at the apparent folly of others (Gratzer, 2000). Rather, it is an important activities of meta-science, the science *of* science. It is about understanding how science succeeds and fails both. It is about learning to do science better. If we are concerned about the methods for developing reliable, or trustworthy, knowledge, then we need to be keenly aware of when the process fails. A study of error is, ideally, an honest—and possibly humbling— epistemic enterprise.

[Buchwald & Franklin, 2005; Holton, 2005; Hon, Schikore & Steinle, 2009]

Right and Wrong

While some commentators highlight error in science with a light hearted tone, others are far less gracious. The intent behind the humor can sometimes be quite scathing. It can be laced with scorn. Adopting a conspicuously self-righteous tone, they heap indignation on those presumably responsible for the errors. For them, error is an affront to science. It is not harmless at all.

[Examples might include Gratzer, 2000; Helfand, 2015; McIntrye, 2019; Pigliucci, 2010; Youngson, 1998]

Consider the concept of phlogiston. It is not well known today. But in the late 17th century and throughout most of the 18th, phlogiston was integral to explanations of combustion, reduction, calcination and, later, photosynthesis. As described by the 1771 *Encyclopedia Britannica*, it was the principle of fire. It is what makes things combustible. And when things burn, it is released as heat and light. It is also what makes metals metal and able to conduct electricity. What we now call oxidation and reduction reactions all involved the release or absorption of phlogiston. No one hears of the concept today because it has been abandoned. It certainly does appear in chemistry textbooks, even as a simple model. As noted 19th-century chemist William Odling explained, we talk instead about electrons and reduction potential.

[Encyclopedia Britannica, 1771]

But for many, phlogiston is considered an error. Indeed, it is among the most maligned concepts in the history of science. Historical accounts are often permeated with ridicule and blame. John Herschel was especially virulent in his 1830 criticism:

> The phlogistic doctrine impeded the progress of science, as far as science of experiment can be impeded by a false theory, by perplexing its cultivators with the appearance of contradictions, . . . and by involving the subject in a mist of visionary and hypothetical causes in place of true and acting principles.

For Herschel, errors were transgressions, akin to modern-day pseudoscience. But Herschel targeted his criticism at scientists, not merely those interpreting science. In these perspectives, errors—like phlogiston—reflect poor science and, equally, credulous scientists who succumb to some kind of cognitive weakness in ever accepting an erroneous claim. For them, errors can never be reasonable. They implicitly indicate lapses of science. Accordingly, errors reflect "pathological" science. Or they are wholly outside science proper.

[Dolby, 1996; Herschel, 1830, pp. 300-301; Jastrow, 1936; Kohn, 1988; Langmuir, 1989; Rhodes, 1997; Rousseau, 1992; Turro, 1999]

As we will see later (Chapter 6), phlogiston was an effective model, which provided some fruitful insights and predictions. While chemists dramatically reconfigured their conceptions of combustion and fire in the late 18th century, one can still appreciate how, in appropriate contexts, phlogiston is a valuable concept for consolidating the relationships of many energetic phenomena. But that requires delving deeply into the historical perspective, not brandishing retrospect in blind triumph. Blame must be set aside. Normative postures must give way to more descriptive, historical ones. How was the status of the error resolved?

[*Allchin, 1992; 1994; 2013, pp. 184-201; Chang, 2009; Kim, 2008; Scott, 1958; Schufle & Thomas, 1971; Bergman & Tunberg, 1780 (trans. Schufle 1972)]

Understanding the typical response to error may be clarified, perhaps, by articulating the subtle role of language. Consider the ambiguity of meaning in the simple terms "right" and "wrong." They apply to knowledge, of course – our central focus. But they also apply equally to morality and ideology. A statement regarded as objectively *true* (say, validated by science) is "right." At the same time, an action regarded as *good* or *beneficial* is also "right." In either case, the opposite is "wrong" – whether a *false* claim, an *immoral* act OR an *undesirable* outcome. While this book focuses on epistemic errors, the sense that something is "wrong" can easily be taken to imply an ethical or professional judgment as well.

Conventionally, we distinguish sharply between these two realms of reasoning. Ethics and epistemics differ fundamentally in their aims. So (ideally, at least), we do not conflate what *is* with what *ought to be*. Justifying *facts* is not the same as justifying *values* or *norms*. We distinguish between *descriptive* and *normative* ways of thinking. Yet the shared language of "right" and "wrong" opens the way to conflating one with the other. An *analysis* of error can easily be construed as an *evaluation* of error.

Accordingly, error in science is for many, unreflectively, a value judgment. A scientific error implies (foremost through the metaphor of language) a scientist's "wrong" step. It "should not" have happened to a responsible scientist. Any error is often regarded as inherently a misdeed. In general, there seems to be an assumption that error ought not to occur. Recall Benjamin Franklin's assertion (Chap. 1) that truth "does not seem to require so much an active energy, as a passive aptitude of soul in order to encounter it." For him, errors occurred only when personal "absurdities" of the mind opened the way for them. The assumption of error-free science seems more than plain optimism. It is an expectation. Error, for many, is an impropriety, a moral transgression.

Such judgments may easily upstage a fruitful analysis of error. An evaluative perspective easily shifts focus from the error itself to the presumed agent behind the error. One tends to assume that a scientist who errs "should have known better." They become culpable for misleading others. Error is treated as a form of harm. As a result, the search for the causal roots of the error yields to assessments of blame. Simply noting an error in justification (an epistemic issue) can thus be interpreted as an inherent criticism of the person making the claim (a normative issue of accountability). The judgmental posture of whether the implicit intent was "justified" can easily eclipse an analytical posture of just how the error happened, descriptively.

This cognitive tendency (again, based on the mere metaphor of language) is especially noticeable when one alludes to an error of some scientific hero. Admirers typically flock to defend the scientist, as though the attribution was an accusation, intended to besmirch their reputation. They invoke all kinds of extenuating circumstances or mitigating contexts to "excuse" the error. The hallmarks of special pleading are all too evident. And unnecessary. Many assume (again) that error is not supposed to happen to "good" scientists. But, as we saw in the last chapter, it most certainly does. Understanding fully how error shapes science involves getting beyond the preoccupation with value judgments, especially the romanticized image of a pristine, error-free science.

At other times, researchers are instead disparaged for their errors. The primary orientation is still judgment. For example, the scientists were stupid, or stubborn, short-sighted, pig-headed, careless, obstinate, intransigent, self-deluded, overconfident, emotional, feckless, or the like (for example, see Fritze, 2009; Gratzer, 2000; Youngson, 1998; Livio, 2014; see also Chapter 3). These all become faults for which the individual scientist bears responsibility. End of story. No further explanation needed. Such views – and they are common – foster unproductive ridicule and self-righteous humor (as in the examples noted above). They obscure the deeper methodological issues and the human and social contexts in which errors emerge. In this book, the issue of personal responsibility is set aside in favor the epistemic question of how to detect and remedy errors. Blame contributes little to understanding how one might eliminate or reduce errors. The analysis here is intended to be more concretely productive, towards a more effective science.

The judgmental posture can take extreme forms, worthy of note. For example, some interpret any attribution of error as an allegation of fraud or misconduct. For them, that is what error in science signifies. Cases of fabricated data and other forms of misrepresentations do appear in the scientific literature (Broad and Wade, 1982; Bell, 1992; Chevassus-au-Louis, 2019; Kohn, 1988; Judson, 2004; Grant, 2007; Goodstein, 2010). Yes, dishonesty is "wrong." And the false claims are also "wrong." But for an epistemic analysis (the focus in this book), the moral dimension needs to be teased apart as an independent problem. Here, the question is not what motivated the fraud, nor how a perpetrator is to be sanctioned. The epistemic question concerns how others, ironically, came to *accept* the false claim or failed to perceive the deception. That is, the problem of fraud may arise, in part, from the general system that allows other scientists to trust one another. At first, they usually take such fraudulent reports at face value. An epistemic analysis may also probe the ways in which the faulty claims are eventually identified as false – often prior to any exposure of the underlying deception. Fraud does have a role in a study of error in science, but not because of the moral transgression. They are ultimately unjustified claims, like the other forms of error we address. (Fraud is addressed more fully in Chapters 3 and 4B.)

In this book, I focus on the errors themselves, and their causes or contexts, not the personalities of the scientists behind them. Of course, science is a human activity, achieved through human agency. Thus one can always frame any particular error as a human error. Thus, if an experiment was done without a control, one can presumptively trace the error to the blind spot of the investigator who failed to note the need for a control. If an instrument malfunctions (a loose cable, say), should we say that it was the researcher's responsibility to have checked the instrument first, or to have calibrated it properly? I find such reframings unhelpful or misleading. They make undue assumptions about abstract, idealized persons. They do not acknowledge real people and ordinary processes. They do not specify the methodological steps for circumventing the errors. Thus, my focus is on profiling the conditions under which errors emerge, not just what individual humans do.

Some errors certainly involve psychological or social factors (as profiled more fully in the next chapter). However, it seems fruitful to analyze such cases at the general level, in terms of patterns, rather than individual instances. That is, one might interpret the error in terms of the cognitive architecture of the human mind. Or the structure of scientific communities and their communication networks. The focus is on how science works as a system, rather than on the culpability of any one individual. Thus, if a scientist "fails" to abandon a theory when new counterevidence appears, one should not automatically attribute their "error" to a personal lapse in character. Rather, the error might be better understood in terms of typical cognitive processes, such as those that privilege prior conceptions, as well as the biographical contingencies that contributed to that particular individual's initial conceptions. Again, the judgmental perspective, which tends to focus on the positive or negative roles of individuals, is replaced by an analytical perspective, which shifts focus to the systematic features of how science works (or doesn't work).

A final pattern of the judgmental predisposition is to prematurely fault any method associated with a documented error. The assumption is:

right method => right answer

Wrong answers thus inevitably reflect the wrong method. There is no allowance for heeding the "right" methods and reaching a "wrong" conclusion. Indeed, that admission might seem antithetical to the very focus on searching for methodological lessons from error. But the immediate transfer of the judgment from error to one ineffective method is not generally warranted. First, "bad luck" does not seem to be an acceptable reason for error. Even if the development of knowledge involves an element of chance, accident, or contingency. Discovery always involves a bit of blind trial and error at some level. The mandate for accountability (assigning "blame" to some method, if not a person) also reflects the underlying judgmental bias.

Second, conclusions from one case (or, more pointedly, a sample size of one) are especially vulnerable to generalizations. One needs to examine the same method(s) in other cases. One equally needs to contrast those to cases where the targeted method was absent, to ascertain if there really is any causal connection, or whether the problematic factors have been prematurely misidentified.

Third, many methods in science are not intended to be algorithmic. Accordingly, exceptions (errors) do not "disprove" the ultimate value of the method. Methods in science rarely *guarantee* a solution each and every time. More typically, they are *strategies* that prove effective *often enough to warrant adopting them*. Even so, to adapt a phrase, "error happens." We need to embrace a slightly deflationary view of methods: as *heuristics*. They need not be foolproof to be effective tools for guiding investigations. That is, *sometimes*:

right method => wrong answer
wrong method => right answer

Many such historical cases have been explored, in particular, in a recent volume, "Wrong for the Right Reasons" (Buchwald & Franklin, 2005). These exceptions will be explored more fully later. In Chapter 6, I profile cases of errors that proved nonetheless fruitful and in Chapter 7, perspectives that fostered great discoveries while also leading, on other occasions, to errors.

[Wimsatt, 2007]

The unmediated tendency to defend science often results in blaming a scientist for any error. But this harsh judgment fails to appreciate how scientists, like the Nobel Prize winners and other famous scientists in Chapter 1, who proved their mettle in some scientific achievement, could nonetheless feel justified about claims that were *later, in other contexts* deemed unjustified. It is the paradoxical comparison of these two apparently contradictory perspectives that we must understand. For an epistemic analysis of error in science to be effective, the teleological predisposition to judge must be recognized and kept in check. The urge to blame scientists or particular features of their investigation must yield to a deeper quest for explanation.

In this book, the aim is pointedly analytical and descriptive, not judgmental. And for this reason, it is essential to draw on well informed historical perspectives. These revive the "insider's" view, from the vantage point of science-in-the-making, not from the privileged view of retrospect. As exemplified in the case of phlogiston above, historical perspectives and the subtle shifts of evidence and context are critical to understanding error. Exercising this form of perspective-taking is one of the central features in this volume.

Bloopers, Blunders and Big Whoppers

As the many cases above indicate, some errors are bloopers. They are just plain funny. Or trivial. Others, however, can be deemed reprehensible transgressions. Affronts to science itself. When, in fact, do errors truly matter? When does an error become consequential enough to significantly affect scientific practice and warrant our concern? Size or scope of the error would seem to matter. For example, should we really fuss about a typographical mistake or contaminated cell culture when there are big whoppers — rejections of major theories and notable cases of fraud — to address? What about all the cases in between? If a scientist entertains or proposes a tentative hypothesis, but never fully embraces it, is that an error? If she even publishes it, but never defends it as fully demonstrated, is that a genuine error? If the scientist's reasoning is sound, but just one initial assumption later turns out to be misplaced, does that count as a mistaken justification? What about an experimental oversight? If another scientist catches the blunder — say, when failing to replicate an experiment — isn't science working just as it should, hence without real error? What are the appropriate boundaries of error in a fruitful epistemic analysis? Here, I consider the spectrum, from blooopers and blunders to great big whoppers.

Earlier, I characterized error tentatively as "a scientific claim whose justification is later found to be unwarranted." That would seem to apply most notably to scientific revolutions. For example, once upon a time it seemed well established that the Earth was the center of the Solar System, and of the universe, too. In the 16th and 17th centuries that belief dissolved due to the work of Copernicus and Galileo, with monumental consequences for how humans understood their place in the cosmos. In the 18th century, guided by a rigorous principle of the conservation of mass, chemists jettisoned the millennia-old notion that matter was composed of four elements – air, earth, fire, water. Air had many components, and water was composed, ironically (and certainly, counterintuitively), of two gaseous elements, the newly discovered hydrogen and oxygen. Everything was reorganized in a new system of modern elements. In the 19th century, geologists exploded the age of the earth and Darwin and Wallace deciphered the organic history of species, including our own. Early psychologists began to map the material basis of the mind, and doctors set aside the last vestiges of a medicine based on four humors, or animistic fluids. Again, humans overhauled their understanding of who they were and how they fit in the natural, material world. In the 20th century, astronomers redescribed the huge size and age of the universe, geophysicists set whole continents adrift, molecular biologists transformed the views of inheritance and organismal development, and neuroscientists challenged notions of autonomy, free will, and moral responsibility. All these discoveries have been truly monumental. And each has simultaneously entailed abandoning earlier ideas as dramatically ill informed, ill founded, or no longer justified. Science is rife with revolution. And errors on a grand scale. No minor "oops," here. These were Big Whoppers.

What are we to make of these revolutions? Are they really episodes of error? In 1962 historian Thomas Kuhn profoundly challenged our thinking about scientific revolutions and about science, more generally. He triggered many debates – about the nature of progress in science, the rationality of scientists, the ability of scientists to communicate clearly with others, or to inhabit different, "incommensurable" worlds. While some of these debates still linger, they are largely irrelevant to the narrower claims here. Kuhn profiled with astute clarity that scientists did not merely add to previous ideas. They changed them. And in the process they rejected certain older theories and ideas, and even definitions of basic terms, as unsuitable or wrong. For example, the very meaning of 'mass' was not the same for Newton as it was for Aristotle. Aristotle's teleological notion could not function in Newton's causal framework. It had to be discarded. The same pattern occurred when Einstein reconceptualized mass in relativistic terms. Mass was a node warping the gravitational fabric of space-time, not an inherent property denoting the quantity of matter. At least from each new conceptual perspective, the former ideas were *errors*. The former claims were no longer fully justified. Understanding this historical irony in justification is certainly one of the major philosophical puzzles that Kuhn introduced.

[Kuhn, 1970]

How should one approach an error analysis of the Big Whoppers – the Ptolemaic universe, Galenic humoral medicine, the Aristotelean elements, young Earth creationism, vitalism, and the like? For some, the attribution of error may well seem overstated. That is, the historical figures who accepted these views might easily be forgiven for not having the resources to know any better. There may have been insufficient background knowledge. Or lack of instrumental technology for making the relevant observations. For others, these discoveries required great insight that could only wait until the rare genius emerged. On these interpretations, the shortcomings were, historically, unavoidable or inevitable. One cannot truly consider the ancient beliefs "errors." However, at the same time, these reconceptualizations illustrate exactly what is problematic about error. The justification changed. The gestalt of scientific knowledge changed. The change was wholly unexpected. Accordingly, these are among the primary cases that should interest us.

Historians and philosophers of science have justly invested considerable attention to interpreting cases of large-scale conceptual change. Many of the new discoveries were controversial when introduced and numerous studies – historical, philosophical, sociological, cognitive – have detailed how scientists came to accept the unfamiliar and sometimes upsetting new theories. They have dissected how evidence was evaluated and how choices were made between the justifications for alternative theories. They have also delved equally into the circumstances that allowed scientists to form the new ideas, even in the shadow of the earlier ideas that they would ultimately replace. This large body of work forms an indispensable foundation for a comprehensive analysis of errors. What remains unfinished, in particular, is a more thorough analysis of the antiquated justifications (in contrast to understanding why the new theories were accepted). Why were the old theories acceptable at one time ultimately found unacceptable – as incomplete, misleading, or wrong? This invites a sensitive synthesis of history and philosophy.

[for example, Darden, 1991; Donovan, Laudan & Laudan, 1988; Englehardt & Caplan, 1987; Losee, 2005]

Does the philosophical concern for revolutionary reconceptualizations, or "extraordinary" errors, also extend to cases of more "ordinary" science, as well? Consider a handful of cases of misconceptions or errors on a smaller scale — that lasted for a few years rather than centuries. For example, in the early 19th century, William Buckland presented fossils from a cave in northern England as the relics of Noah's flood. Although his work earned him the Copley Medal, the Royal Society's highest honor, evidence from other caves soon indicated that great floods around Europe did not all occur at the same time. Buckland conceded the possibility of a series of floods, many local. As the geological timeline filled in, and the historical timeline expanded, Buckland himself came to regard his original claims as wrong (for a fuller account, see Chapter X). Or consider the claims of James McConnell in the 1960s, whose studies on flatworms seemed to indicate that one could transfer the learning of one organism by feeding their homogenized brains to another worm. Memory seemed to have a molecular basis, an *engram* (apparently encoded in RNA), just like genes passed from generation to generation. But these noteworthy claims unraveled as others failed to reproduce the results in further investigations. In yet another case, in the late 1980s gravitational anomalies were discovered that could only be explained, it seemed, by introducing a fifth fundamental force. Some force of attraction besides the basic four (gravity, electromagnetism, and the strong and weak interactions at the atomic level) seemed to be working at distances of about 100 meters. Further studies aimed to characterize the force more clearly, but as rigor increased, within a decade the bold new idea was put to rest. Finally, consider the chemists' efforts in the 1890s to get a consistent density for nitrogen, the gas that makes up over three-quarters of Earth's atmosphere. Chemical methods of isolating the gas and physical methods of isolating the gas yielded different values. At first the results seemed to be due to some impurity. But when the "impurity" was isolated on its own, it turned out to be a new element, which was named argon. A series of discoveries of new gaseous elements followed: helium, neon, and xenon. But the periodic table had no place for them. The new discoveries exposed a fundamental error in how the periodic table had been assembled, and henceforth it was expanded to include a whole new column: the inert (or "noble") gases. All these cases seem to exemplify how scientists encounter errors at a substantive scale on a regular basis. A concerted analysis might reveal more about how scientists fall into error and eventually resolve them.

[Collins & Pinch, 1993; Franklin, 1993; Friedlander, 1998; Gould, 1985; Guinta, 2001; Hirsch, 1979]

Such cases elicit a variety of responses among scientists. Some admit that researchers pursue mistaken theories and even accept them for short periods, but dismiss such cases as unworthy of special analysis. On principle, if one acknowledges that science progresses ultimately through trial and error, errors seem *inevitable*. Why bother documenting or analyzing the trials that lead nowhere? Dead-ends are not errors exactly. They're just bad luck: inconsequential *blunders*. Indeed, pursuing promising leads always seems well justified, even if in retrospect any particular pathway did not eventually bear fruit. The fifth force and molecular engram concepts were promising ideas that just could not meet the standards of evidence. As for argon, who could anticipate its existence before it was demonstrated? Such cases merely illustrate the Scientific Method at work. No fault, no blame. No potential insights.

Yet such a posture seems unduly preemptive. One has not articulated how the acceptable ideas are ultimately distinguished from the unacceptable. If we do, in fact, blunder about, how do we find the true path? [fn: insufficiency of *The Golem*] Why, or how, are false paths deceptive? This view of mid-scale error does not sufficiently specify the nature of reliable evidence. In particular, it does not address the problem of incomplete versus complete evidence. How exactly can partial evidence be concretely misleading, rather than merely vague or "suggestive"? How can conclusions be *negated* by the *addition* of more information? When, indeed, can evidence be considered complete? When, if ever, is confidence warranted in the conclusions? To what degree? Without some clearer notion of completeness, all scientific claims must be regarded as uniformly "tentative." One could never trust them. The detailed process of ascertaining error, not merely demonstrating more positive claims, matters. (The role of negative knowledge will be elaborated more fully in Chapter 5.)

A second common response to mid-scale error is to fault the researcher. The errors are dismissed as human failings, or *blunders*. For example, McConnell is blamed for poor experimental skills. Buckland succumbed inappropriately to religious bias. The early proponents of the fifth force failed to use sufficiently rigorous experimental controls or to exercise suitable caution. No one should have tried to determine the density of nitrogen without first ensuring the sample's purity. No epistemic analysis is needed because one already knows why the errors occur. Humans make mistakes. Scientists simply need to be more capable and responsible. They need to heed the rules of proper science. That is a problem, but not one that philosophical analysis will enlighten. In this view, cases of error are only useful as cautionary tales, to warn others against sloppiness and self-delusion.

But such a view often prejudices the case. Not all errors in everyday science are personal blunders, due to lack of skill, flimsy reasoning, or professional incompetence. The cases of the Nobel Prize winners (Chapter 1), in particular, should indicate that such a broad judgment is unwarranted. Surely *some* cases of error might fit this explanation. Most others do not. Well informed historical analysis is needed. In particular, any such analysis needs to be sensitive to historical perspective. One needs to appreciate what is reasonable to expect without the privilege of retrospect. Indeed, closer analysis of the four cases above indicates that the casual allegations of incompetence are misplaced. Appraisals made without attention to the historical horizon are thus not helpful. Any methodological lessons must function in context. (For further elaboration on the tendency to judge, see the section below on "Right and Wrong")

A third widespread view is that any error, if it does appear, is necessarily short-lived. Science is *self-correcting*. Even if some researcher is unskilled, for example, or pursues a will-of-the-wisp theory, any such blunder will soon be exposed by other researchers. Peer review, criticism, and replication all guarantee that incorrect claims are identified and discredited. The purported system of "organized skepticism" roots out and eliminates errors, whether major or minor. Indeed, all four cases noted above seem to illustrate how science corrected mid-scale errors. In this view, a study of such errors in science can achieve nothing more than redescribing the inherent process of error-correction already in place.

As I will detail more fully in Chapter 4a, however, historical evidence does not support the image of self-correction. Some errors persist far longer than seems reasonable if there is a genuine "self-correcting" mechanism. Also, some theories once corrected are, ironically, then rejected. Some errors become compounded by fostering other errors before they are corrected.. Other errors are apparently "corrected," but only replaced by sequel errors (the actual fundamental error not being corrected until some time later). Happenstance plays a significant role, in contrast to the methodological purging of error. Replication, peer review and organized skepticism are supposed to expose errors. However, historical analysis indicates that replication (in the sense of checking results) is rare. Nor does peer review uniformly filter out error. Appeals to self-correction have strong rhetorical appeal. They fit a widespread teleological view of inevitable progress. Yet however "obvious" they may seem, they are not historically well founded. Indeed, all these insights about the failure of science to correct itself – contrary to intuitions or folk wisdom – are precisely why an analysis of the "blunders" in everyday practice is so essential to a complete epistemic understanding of science.

Blunders seem to matter just as much as big whoppers, then. What about bloopers? Surely the "oops" moments do not rank as epistemically significant compared to dramatic revolutions and failed theories? Human foibles may be worth a laugh, of course, as illustrated above. But how could they mar or seriously interfere with science?

Again, historical cases can be informative. "Wee" errors are not always so wee. For example, during his renowned voyage on the *Beagle*, Charles Darwin collected a lot of animal specimens. Many were new to science. Such discoveries involve, importantly, documenting where the specimens comes from. When he visited the Galápagos Islands, Darwin took an interest in the tortoises, the iguanas, and the mockingbirds. In addition, on several islands he caught some drab brown birds. Several weeks later, en route to Tahiti, he caught up with his record keeping and catalogued them. But the birds from the various islands were now all mixed together. Locations were left unlabeled.

Several years later, back in London, the ornithologist at the British Museum, John Gould, examined the specimens. In spring of 1837 Gould announced that they were all hitherto unknown finch species. Darwin was then trying to establish himself professionally and was certainly pleased to earn credit for the discoveries. But he was also puzzled. While the birds differed in appearance – one was warbler-like, one used a cactus spine like a woodpecker, others had bulkier beaks and varying body sizes – Gould characterized them all as very closely related species, in allied genera, also similar to species from the South American mainland roughly 1,000 kilometers away. Darwin had already been thinking about the nature of closely allied species and varieties. In the same month, Gould had presented Darwin's specimen of a rhea from South America as a second, new species, and he named it after him: *Rhea darwinii*. In that instance, Darwin had been ruminating on the role of the large river between the two types, perhaps separating one original population in the past. Now the role of geographical separation could be important in the Galápagos case, as well, if each finch came from particular islands. Darwin returned to his records and notes to find them woefully incomplete. Only one distinctive type (with a parrot-like beak) could be placed unambiguously. What a mistake not to have simply labeled the birds when the geographic information was still fresh at hand!

Darwin must have felt some urgency, as he soon consulted three shipmates, seeking information about their own specimens and specifically where they had been collected. He tried to align all the specimen types and known locations. But errors appeared. Ultimately, the location of two specimens were incorrectly assigned due to Darwin's original haphazard grouping of specimens. One species (*Geospiza fortis*) was incorrectly associated with one island, and not with another. Another species (*Cactomis assimilis*) was assigned to only two islands, although one collector's records indicated it had also been collected elsewhere. Other identifications conflict with the distribution patterns as now understood. The errors were eventually published in Darwin's formal account. Ironically, some of those published identifications, inferred solely from *other* specimens, were later used by museum curators to "correct" Darwin's own unlabeled specimens. The original labels of several specimens of other collectors were also "corrected" as well. Through all the labeling mishaps, Darwin became persuaded that the finch species were uniquely associated with certain islands. In this, they would match the mockingbirds, although their number of types was fewer. That all helped critically fuel his thinking about the relationship of geography and the origin of species. Several months later, in beginning a notebook on species transmutation, Darwin made special reference in his private journal to his thoughts at this time and to the Galápagos Archipelago as the origin of all his ideas. At the same time, however, Darwin likely recognized the potential errors involved and, with other vivid examples of geographical isolation available, he ultimately chose not to present the finches in his published work. [Brown, 1995, pp. 358-63; Sulloway, 1982]

Darwin's failure to label all the landmark Galápagos specimens correctly may seem trivial, and slightly amusing given that the finches now bear Darwin's name. But the unreliable labels also had immense theoretical significance. Little, careless errors can have big consequences.

Nor is the significance of (mis)labeling limited to this one case. On his collecting expedition through the Amazon in the 1840s, Alfred Russel Wallace noticed that many species were present on only one side of major rivers, and replaced by similar species on the other side. Wallace was observing the same kind of biogeographical patterns that would help him, like Darwin, discover the origin of new species by evolutionary divergence. When Wallace returned to England, he advised his colleagues of the problem. One needed to correctly label specimens with such details. A label of “Lower Rio Negro,” for example, would not suffice. To avoid errors, one also needed to know which side of the river the specimen represented. This was critical to determining the species’ proper range. Nowadays, biogeography is used to help guide conservation efforts. To decide where to protect a particular species, or to restore lost populations, knowledge of its historic range is fundamental. In a recent survey of the Caribbean frog, *Leptoddactylus fallax*, a team of conservationists encountered numerous mislabeled museum specimens. In one university collection, ten specimens were attributed to the Dominican Republic, but were not the correct species. At another, one attributed to Trinidad was likewise misidentified. One publication confused the Dominican Republic with the Commonwealth of Dominica (a small island nation). Collectively, the errors misrepresented the historical range. If uncorrected, conservationists might have tried to “reintroduce” a species on islands where it had, in fact, never lived. Again, small bloopers can have significant consequences.

[Wallace, 1852; King & Ashmore, 2014]

In February, 1953, James Watson and Francis Crick were busy trying to determine the structure of DNA. They had already decided that it was a helix and that it was a double helix, not a triple helix as Linus Pauling suspected. Four types of nucleotide bases spanned the inside of the molecule in pairs, they assumed. They were trying to fit the bases together. Watson became convinced that each base matched with another of its own type. He was trying to explain the structure to a visiting chemist, Jerry Donohoe, who shared the inner office. Donohoe observed, however, that the form of the bases they were using was misinformed, even though it had come from a well known source. First, there was the issue of how the oxygen was connected on one corner of the molecule. The correct form, Donohoe insisted, was a keto (with a double bonded oxygen), rather than an enol (an oxygen with an additional hydrogen exposed). In another corner position, nitrogen should have been conceived as an amine (NH_2), rather than an imine (NH). These affected the shapes of the molecules and the positions where hydrogens were available for the weak hydrogen bonds that might stabilize the molecule. The like-with-like model was “nonsense.” After a week of arguing, Watson accepted the alternative form. He made new cardboard cut-outs of the revised forms and readily found that they could fit one another elegantly in complementary shaped pairs. Soon thereafter, the pairing gave rise to the realization that the molecule could “unzip” to replicate and that the unambiguous pairing of base shapes would ensure fidelity in building new molecules by adding new bases. The subtle error of chemical form had been minor, surely. But Watson had relied on it initially, and when it was corrected, the now familiar form of DNA rapidly unfolded. Once again, small mistakes can have big consequences.

[McElheny, , pp. 55-56]

As a final example (one of my favorites), consider the international news headline from late 2011 that scientists had discovered neutrinos that traveled faster than the speed of light. Front page news challenged one of the most fundamental principles, or assumptions, of modern physics! One team of researchers was producing neutrinos at a particle accelerator in Switzerland while another observed their appearance at a lab in Italy over 700 kilometers away. When the time was taken into account, the neutrinos seemed to travel very fast indeed. Faster than the speed of light. The claim was presented with some caution, of course. The investigative collaboration openly invited others to find any errors in their results. But the claim was published. As investigations continued, the researchers tracked the result to a loose cable. The time signal from one lab to the other had been faulty due to an improper electrical connection on the back of one of the instruments. Pretty trivial error, if you think about it. Hardly worth our notice. Shrug it off and move on. On the other hand, that error catapulted the false claims to front page news around the globe. At least three subsequent research projects were conducted over the next two years to lay the error to rest, at considerable expense of research time and effort. As we will see in the next chapter, such "trivial" mistakes can have significant consequences, especially as measured in public impact, or beyond the insular world of the scientific community. Such errors matter. And that is a convenient benchmark for knowing what errors to examine.

[Cartlidge, 2012; Zichichi, Sirri & Sioli, 2012]

So, some errors are bloopers or blunders. Others are big whoppers. And some are of intermediate scale. But what seems to matter is not the level of the error — whether it is based on an experimental slip or a major conceptual blind spot leading to a theoretical revolution. Rather, the measure of an error is its consequences. How does an error affect the practice of science? How does it guide work that proves to be unproductive, or costly in terms of the amount of scientific effort eventually needed to find and remedy the error? And how might the claims have adversely guided social policy before the error was detected and resolved?

[Star & Gerson, 1986]

Epistemic Posture

The potential scope of consequences for an error will be greatly shaped by the epistemic posture towards the claim and its justification. In a historical context—at any given historical moment—was it fully accepted, or only adopted provisionally? Was an epistemic commitment made? For epistemic analysis, the relevant benchmark seems to be when the claim is deemed appropriately justified. Thus, in cases where scientists are still explicitly tentative, or hypothetical, there seems to be only acknowledged uncertainty, not error.

However, a posture that *all* claims are "tentative," or subject to change, even when endorsed as justified, is not very helpful. This rhetorical device is really no more than a universal escape clause. A "get out of jail free" card tucked up one's sleeve, to be revealed only when needed. The epistemic commitment is missing. Just as if it was mindfully withheld. Under such circumstances (with no conditions specified), *no* justification can ever be considered complete or acceptable. Thus, when a scientist is aware of particular qualifications, or "known unknowns," they must be explicitly stated. This stipulates the limit of the claim's justification. It identifies the scope of any residual uncertainty.

Of course, there might be varying "levels" of commitment to a claim. This notion is hardly new or unfamiliar. What concerns us here most, however, are the cases when some definitive commitment was made, and *later* the commitment was withdrawn. Something changes. Some assumption becomes suspect, or some new evidence offers fresh context. *The very same claim* is now deemed to be unjustified. At a meta-epistemic level, we should identify just what those changes are in each case and, more generally, what kinds of changes are exhibited. Namely, what are all the possible *sources of error*? Surveying this spectrum of possibilities is the primary aim of the next chapter.

[Latour & Woolgar, 1979]

In articulating a concept (or definition) of error, one should acknowledge that by using a standard of epistemic commitment, uncertain claims are not productively interpreted as errors. So, for example, many critics of evolution had a field day when it was determined that one of Darwin's predictions was "wrong." Recent genetic analysis has indicated that modern chickens evolved from grey-footed jungle fowl. Well, Darwin got that "wrong." He had proposed that the ancestors were *red*-footed jungle fowl, a different species. But it is only an error if we accept that Darwin had presented his claim as definitive and fully justified. He did not. His primary intention was to indicate that domesticated species could be traced to wild progenitors, illustrating descent with modification. So he offered a plausible example, to be investigated for its possible specific merit. Speculation, especially when acknowledged as such, is not error.

[Eriksson et al, 2008]

So, how should we view Linus Pauling's claim in 1953 that DNA was a triple helix? That claim was dramatically eclipsed only a few months later by Watson and Crick, with their double-helix model. Should we view this as a prime example of error in science, as suggested by some? Two different historical commentators draw attention, in particular, to Pauling's professional stature and credibility. He was a Nobel Prize-winning chemist. Earlier he had successfully elucidated the alpha-helical structure of proteins. So even Pauling's opinion would have carried some persuasive weight. Yet while Pauling published the triple-helix idea, he and his co-author never presented it as more than a "promising structure." The evidence was all suggestive. As a claim, it was nowhere near justified, even for the authors. We can easily imagine that Pauling was eager to stake a position that would establish priority of discovery—*if* it turned out to be correct, while protecting himself from embarassment if it later turned out to be "wrong." But there was no substantive epistemic commitment. It was not an *error*. Tentative claims do not reflect errors, precisely because the justification is regarded as incomplete.

[Darden, 1998; Livio, 2014; Pauling & Corey, 1953, p. 84-85]

Similar cautions might be applied to a number of other cases that historians and philosophers have variously portrayed as errors. For example, Francis Crick, co-discover of the DNA model, has been critiqued for proposing that the decoding of DNA's genetic sequence was due to a "comma-free" code. But this was only ever offered as a proposal, with no data, and very little to justify it beyond a perception of theoretical necessity. Einstein's famous cosmological constant is also sometimes portrayed as a blemish on his record of genius, implying an error. But Einstein never defended it with any vigor or rigorous proof. It might have been a bad idea, but not an *error*. In astronomy, Fred Hoyle, famous for interpreting how atomic elements originated in the early universe, is known for rejecting the concept of the Big Bang when everyone else seemed to accept it. By contradicting the consensus of the community—and disagreeing with the theory which ultimately prevailed—Hoyle is considered to have erred. But for Hoyle, and even for most of the community, the justifications for the hypotheses of a stable versus an expanding universe remained indeterminate. The apparent consensus was just a shared best guess. No one knew for sure, and they acknowledged this. Of course, this did not stop them from staking their positions and parading their favored theories. Hoyle's dissent is surely interesting, but due to the acknowledged uncertainty, one cannot consider it an "error" in justification. In all these cases, where epistemic commitment was lacking, there was no formal error that might inform our understanding of a mistaken justification.

[Darden, 1998; Livio, 2014; Maynard Smith, 1999]

Many of these attributions of error seem to focus on accounting for "erroneous" thinking trajectories. They tend to reveal a common assumption (noted above) that proper science should lead scientists "unerringly" along the "correct" path. However (as the examples in Chapter 1 were intended to show), error is inevitably part of the process we call trial and error. Science is better viewed from a larger perspective, as a process of adaptation, which involves both blind variation and selective retention. Claims pursued in the short-term that are not ultimately accepted in the long term are thus not evidence of "bad" science or poor thinking. Rather, they are indicators of the many possible alternatives being productively explored. They express healthy science. *Speculating* creatively, or *entertaining* concepts that are considered only *plausible* or *partly (incompletely) justified* should not be confused with error. This is one reason why the focus in this book shifts from the proposed ideas, en route to a solution, to the *justifications* that foster a sense of epistemic commitment when one has arrived at a, or the, solution.

[Campbell, 1974]

So what marks epistemic commitment in a claim's justification? Three factors may guide analysis. All indicate different levels of commitment. First, does the scientist or lab group in question *pursue* the claim? Are resources invested in investigating it empirically? Are its implications or assumptions explored theoretically? These actions reflect a judgment that the idea is at least plausible, not wholly frivolous. A "tentative" posture, here, is nonetheless a positive stance. That is, it reflects a posture about burden of proof: that the claim can be regarded (albeit modestly) as warranted until further evidence indicates otherwise. Thus, even for something nominally labeled a "hypothesis," an observer can measure concrete commitment in terms of time, lab resources, and other effort. For example, before research granting agencies fund a research project, they typically require some justification that the endeavor itself has merit, or that the target concept is already partly justifed. That is, even an assessment of theory promise involves justification. Accordingly, it is common for scientists who are simply pursuing a hypothesis to say that it "failed" an experimental test, or that it proved "wrong" — indirectly signaling a former expectation and that an epistemic commitment has changed.

[Šešelja & Straßer, 2014; Whitt, 1992]

A second level of epistemic commitment occurs when the claim, along with its justification, are *published*. Using publication as a benchmark expresses an assumption that introducing an idea into a public forum, where the justification is open to scrutiny and criticism, reflects a significant level of confidence about, or commitment to, a claim. The claim moves from the security of the private realm to the vulnerability of the public realm. Publication also takes effort. That reflects a practical level of commitment. And in cases of peer review, publication generally (not always!) reflects that the proposed justification has met at least professional standards of epistemic responsibility—and has not exhibited some glaring reason for being discounted entirely.

Scientists adopt a range of postures towards publication, with some being much bolder (and more willing than others) to present ideas that might eventually be rejected. In addition, standards of quality may vary among different publications. Still, publication is a helpful heuristic, other indications notwithstanding, for supposing that there is some form of epistemic commitment behind presenting the claim formally to others for review. For my analysis, it is of considerable interest when the justification for a published claim—even a cautious, plausible one—is later rejected, or negated by further evidence. We should want to know why.

Another important benchmark is *consensus*. This shifts the level of the claim, and the epistemic commitment embodied in "acceptance," from the individual to the community. Consensus can be notoriously slippery when one tries to establish it formally. Still, if one takes the system of checks and balances within the scientific community as critical to the epistemic process, then consensus among different perspectives reflects a depth of epistemic commitment not available to an individual. What appears in textbooks, as an expression of consensus, matters. These are the kinds of claims that one expects to serve as a stable benchmark in public policy or personal decision-making. Again, if we are interested in science as a source of reliable claims, we should be interested when scientific consensus changes—and more importantly—why.

[Gilbert and Mulkay, 1984; Oreskes, 2019; Ziman, 1968]

As an additional note, I should note that I am focusing on science as a semi-autonomous institution, without fully addressing the more subtle and complex problems of the public understanding of science. Hence, in this book, I am only marginally interested in such cases as Velikovsky's *Worlds in Collision*, modern astrology, homeopathic medicine, anti-vaxxers, or climate change naysayers, the focus of many books on the "errors" of science and pseudoscience. Many books focus on rationality, popular science beliefs, scams, and so on, hoping to equip the citizen or consumer with appropriate analytical tools. Here, I am concerned instead more narrowly with what the community of scientists regards as trustworthy. That expert consensus is typically the baseline against which such views are measured. It is thus important to know when their claims can be trusted and when they might be shaky. When is confidence due, and when is some level of hesitation or critique appropriate? In what ways might the justification behind a scientific consensus be considered incomplete or potentially in error? My analysis of error helps characterize the collective epistemic posture of scientists in such contexts.

[see also Allchin, 2012, 2021; Oreskes, 2019]

Error and Uncertainty

The history of errors in science certainly indicates that science is fallible. Or "tentative," we often hear. The lesson seems to be that any science is inherently incomplete. Errors helps expose, perhaps, the limits of our ability to know. We should have said, perhaps, that our claims were *uncertain*, or susceptible to *error*. That is, the epistemic posture towards *all* scientific claims should be provisional. Errors might foster a profound epistemic humility. Perhaps we should just say that science is inherently *uncertain*? Error and uncertainty thus seem to be intimately related.

But error is not uncertainty. The epistemic posture, as articulated above, differs. Uncertainty is, by its very nature, explicitly acknowledged. Scientists *admit* that they do not know. That is, they explicitly recognize when the available data support the justification of a claim only so far. At that particular moment, they acknowledge, we *cannot* know. They declare the current limit of our ability to decide among a set of alternative conclusions. Further, we can usually articulate the range of possibilities among which we remain unsure. Sometimes we even assign a probability to the degree of confidence in various claims. While it seems impossible to be any more definitive, at least we have a clear sense of the scope of what is not known. In cases of uncertainty, a firm conclusion is not considered fully justified. Or not *yet*, at least. The answer is explicitly *unresolved*.

In cases of error, however, we are initially confident in, and epistemically committed to, our conclusions. Only later do we realize unexpectedly that the justification was imperfect. Further, identifying an error entails articulating exactly *how* the justification was imperfect or incomplete. The structure of the justification changes. Accepted knowledge shifts. That catches us off guard. That upheaval, sometimes a reversal of the original claim, is what makes the study of errors so significant and so compelling. What we once thought we knew, we come to realize, can no longer be trusted. The *justification* can change and with it the scientific verdict (see Chapter 4B).

In cases of uncertainty, by contrast, we never reach that initial conclusion. We do not make an epistemic commitment that places us in the vulnerable position of possibly being "wrong" later. Indeed, uncertainty reflects the acknowledged absence of secure justification. We openly acknowledge an epistemic indeterminacy. In cases of error, the lapse in justification is hidden and goes unrecognized. Indeed, understanding the nature of that invisible/visible duality is precisely the challenge of interpreting cases of error.

While conceptually distinct, error and uncertainty are not always thoroughly distinguished in practice. That confusion may be traced back, at least, to the early history of statistics and the origin of the modern concept of measurement "error." As geographers and astronomers mapped the Earth and sky in the late 18th century, puzzles began to emerge. Astronomers would measure the location of a star in the heavens and later others would check them. But their measurements did not always match. Why? The stars had not changed position. Planets moved, yes. But not stars. Geographers were confronting similar problems. They mapped the location of places on the earth. But multiple measurements of fixed landmarks, such as mountaintops or church steeples, differed. Such features could not shift relative to the landscape. Nor could stars move relative to the heavens. Some of the measurements, they concluded, had to be *wrong*. To correct the apparent errors, the scientists endeavored to improve their instruments. They refined telescope optics. They machined their surveying instruments and scales more carefully. Yet despite their persistent efforts, the inconsistencies continued. How could they identify which measurements were correct, which incorrect?

By monitoring the many implicit errors, measurers ultimately detected a pattern. The measurements tended to cluster, indicating at least some rough concordance. One could discern a scattering around a central value. They decided that the mean value must represent the "true" value. That solved the problem at one level. Simply average the values. But the pattern of distribution was also uniform from case to case. The context of the measurement seemed not to matter. Minor deviations were common, major deviations less so. When they graphed the frequency of the various values, they consistently formed a bell-shaped curve. Errors, apparently, followed a regular pattern, which could be expressed mathematically. Here, ironically arising out of apparent chaos, was one of those striking regularities scientists strive to discover. It was a new law of nature. They called it the *Law of Error*.

It was not long before they characterized the pattern mathematically and discerned its relationship to probability distributions. From that notable insight, the field of mathematical statistics developed. For much of the 19th century, the Law of Error was its foundation. Today we know the historical law of error as a Gaussian distribution. It describes a probabilistic, or random, variation: a form of uncertainty. It is not error. The astronomical and geodesical "mismeasurements" were not wrong per se. They were imprecise. Inherently "fuzzy." That is, they were *not* inaccurate. They *were* indeed the "correct" value, but also inescapably coupled to an additional factor of characteristic variation.. The "law of error" was not really about error, then. Rather, it was about *uncertainty*. Historically, the momentous discovery of the statisticians was, ultimately, to realize this distinction: that measurement uncertainty is not error.

[Hacking, 1990; Porter, 1988; Stigler, 1986]

Ironically, perhaps, the term "error" is still used to describe the *uncertainty* of measured values. A *standard error* denotes the range that encompasses 68.2% of the uncertainty. A *margin of error,* such as reported in public poll results, does the same, typically at a 95% level. When researchers graph their experimental results, they plot mean values and an *error bar* that visualizes a measure of uncertainty in a set of observed values. That error bar is a shorthand version of the bell-curve, or the former Law of Error. Yet the modern term of "measurement error" is, I claim, a misnomer. There is no error involved. The "error" does not refer to a false claim. Nor to any purported failure in justification. As the history of the Law of Error indicates, the "error" refers to *uncertainty*. Ideally, then, experimentalists should modify their terminology. They should refer instead to *measurement uncertainty*. "Error bars" should be labeled, more plainly and honestly, as *uncertainty bars*. They reflect the limited degree of precision in the values represented. Standard error is, more properly, *standard uncertainty*. Margin of error is really *margin of uncertainty*. This simple change in use of terms will, I contend, help promote a clearer focus on error itself, as separate from questions of uncertainty.

A similar confusion of terminology occurs with the general concept of experimental error. Experimentalists now conventionally distinguish two types of "error": systematic error and random error. Systematic errors are, in a sense, the true errors. They encompass all the ways the recorded values or observations may be *uniformly* biased. Many factors presumably affect or distort each measurement equally. Some such factors researchers can be identify and quantify. Hence, they can make corresponding adjustments or corrections to the measurements. For those that remain unknown, we simply use the term experimental error. The sources of error might be incorrect reference values in calculation formulas, properties of the measuring instrument, or underlying assumptions about how the observational set-up relates to the target quantity being measured. That is, all these factors may affect the *accuracy* of the values determined. Once these errors are discovered, they can be corrected. Until then, they remain possible, but ultimately unknown errors.

Not so for the second type of experimental errors, conventionally called random errors. As their name might indicate, they exhibit a probabilistic distribution. Again, the sources cannot be fully specified. But these "errors" are visible, in a sense, in the variation of values. Again, there is observed uncertainty. Some values appear too high, some too low. But they are often symmetrical around the calculated mean value. The effect of the variation is thus not to alter the accuracy of experimental values, but to limit their *precision.*

Again, systematic (true) error relates to accuracy; random "error" to uncertainty and precision. And again, terminology in experimental work can help distinguish between *sources of error* and *sources of uncertainty*.

[Adler, 2004; Hon, 1989; Wilson, 1990, Ch. 9]

The relationship of systematic error and measurement uncertainty is nicely illustrated in the histories of determining the value for many physical constants, such as the charge of the electron, or the atomic mass of chlorine. Consider, for example, the history of the value assigned to the universe's rate of expansion, known as Hubble's constant. In 1928, Edwin Hubble addressed the question of whether the universe is the same size it has always been, or whether it is expanding in size. The expansion would be indicated indirectly by the nature of light from stars in far away galaxies. If the galaxies were moving apart, the light would be slightly redder (imagine the wavelengths stretching and lengthening, changing color, as the galaxy recedes). There was evidence for that.

[For graphical depcitions of the history of various subatomic particle constants, see https://pdg.lbl.gov/2020/reviews/rpp2020-rev-history-plots.pdf]

But what was the *rate* of expansion? Knowing that was important because it would indicate (again, indirectly) the age of the universe. Determining the rate involved observing stars of different ages and comparing how their receding velocities (red shifts) may have changed through time. That required ascertaining the age of stars. The ages, in turn, would be based on the stars' distances from Earth (the farther the star, the longer it took for light to travel to Earth, and thus the older the star). Determining distance is problematic, however. From a central vantage point, it is very difficult to ascertain just how far any star is, especially outside our own galaxy. It depends on relative brightness. If you know the absolute brightness of a star, you can determine its distance; if you know the distance, you can determine its absolute brightness. But from relative (observed) brightness, you can't determine both at the same time. That had left astronomers in ignorance (not even uncertainty!) about the distances to stars outside the Milky Way.

Hubble had solved the dating problem by using a special type of star, known as Cepheid variables. They pulsate, and their period of pulsation is related to their absolute brightness. Cepheids could help establish the distance in galaxies where they were located. And Hubble found many of them. The calculation series fell into place. Hubble calculated the rate of the universe's expansion and, with it, its age.

But one can easily appreciate the many uncertainties in Hubble's value. There is the sensitivity of the telescopes to brightness (especially for the low relative brightness of very distant galaxies). There is the number of distant stars you can find; too few, and you risk an unrepresentative sample. Then there is the distance calibration using the Cepheids, itself an estimate. Indeed, around the same time other astronomers used similar techniques, but arrived at different values. They ranged from 450 to 600 (measured in kilometers per second per megaparsec). One might imagine that one value was correct, and the others in error. But given the acknowledged sources of variation, it seems more appropriate to regard the range of values as just approximate. In the 1930s, the different values for the rate of expansion reflected scientific *uncertainty*, not error in any one value.

But things changed in 1952. With more powerful telescopes available, Walter Baade discovered that the Cepheid variables, once considered a uniform group of stars, actually combined two different sets of stars. Stars that Hubble thought had a certain brightness became *clusters* of stars, each with a lower luminosity. That change cascaded into the determinations of distance. Distance into stellar ages. Stellar ages into the rate of expansion, and the rate of expansion into the age of the universe. Suddenly, the universe seemed twice as old. The new value was notably well outside the range of previous calculations. Baade had exposed a *systematic* error, an undetected assumption about the uniformity of Cepheids.

Then in the late 1950s, again using more powerful telescopes still, Allan Sandage observed that Hubble had misidentified another key group of stars used in the calculations. They were nebulae (formally, HII regions), not stars. The rate of expansion was cut again, to about 75 km/s/Mpc, and the age of the universe nearly doubled again. Another systematic error.

In the past half-century, new calculations have continued to appear. But they now fall within a range of 50 to 100. Different methods for estimating distance have emerged, allowing astronomers to cross-check them. The eponymous Hubble Telescope has also provided more highly resolved images and additional data that help decrease the distribution of measurements. Thus, there is still experimental uncertainty. But as the range of values narrows, there is increasing confidence in the precise values obtained. The history of the Hubble constant thus exhibits both error ("systematic error") and uncertainty ("random error"), and may help clarify the distinction between them.

[Huckra, 1992; 2008; graph: Pritychenko, 2015; Trimble, 1996
see also https://www.cfa.harvard.edu/~dfabricant/huchra/hubble/h1920.jpg]

As the analysis of error unfolds, it will become clear that error and uncertainty do have an important relationship. Uncertainty can emerge more qualitatively. For example, scientists may disagree about the theoretical meaning of the same set of evidence. There are alternative conclusions and the scientific community cannot completely resolve them with the information then available. This is uncertainty *in conceptual interpretation*, not experimental measurements. Such a period of uncertainty often precedes consensus. And, as I will discuss more fully in Chapter 4B, the discovery of an "error" will typically begin by questioning the justification. That shifts the claim to one of uncertainty. It is not quite an error yet, because the status is unknown. With further investigation, the doubt of the claim may be laid to rest and the original claim reaffirmed. Alternatively, knowledge may deepen and the former claim reassigned to the status of "error." At the moment, one may note merely that in addition to being "right" or "wrong" (positive and negative knowledge), there is a third category: uncertainty, or unresolved knowledge (see also Chapter 5). Uncertainty is not the same as error.

Summary

A focus on justification helps indicate why I do not dwell on errors as faulty behaviors of individual scientists. Such actions or decisions are relevant only to the extent that we rely on such individuals to justify our scientific claims. And generally, we do not. We typically link our claims to material evidence, which is accessible (at least in principle) to others. We do not want to rely on "expert judgment" unless a cluster of diverse experts concur. This does not discount the role of experts in having the requisite skills to interpret the material evidence. Nor does it completely eclipse our reliance on some scientists in certain occasions in making "expert judgments." Still, the focus is on the justification through material evidence and demonstration, not through reliance on individual judgment or cognition.

Thus, in a practical sense, an error is a claim that deemed justified at one time or in one context, was deemed unjustified later or in another context. And that shift frames the central puzzle. How are we to interpret a claim that is taken as justified in one context, but not in another context? How do we interpret the relationship of those contexts, and the corresponding justifications? The focus is the justifications.

My aim in this chapter has been to provide a foundation for grounding a discussion by characterizing precisely what I mean—and what I do not mean—by an error in science. First, we must set aside the tendency for moralistic judgments and humorous irony as irrelevant. The primary focus is epistemic. Errors involve changes in the justifications of scientific claims, or in the mappings between the empirical evidence and concepts, as representations. These changes are significant where others depend on scientific claims to be trustworthy and reliable. Because scientific errors involve epistemic commitments, they do not include uncertainties, where the evidence for particular claims is not yet fully articulated or resolved. (Unfortunately, this view contrasts with some common nomenclature about measurements and experimental uncertainties.) Errors occur at many levels, from the apparently trivial to the deeply theoretical. But their significance is best measured by their consequences for scientific practice, especially in terms of the degree of scientific work that must be invested to remedy them. Accordingly, joking about errors, while possibly entertaining or supporting one's sense of superiority, are misplaced. Rather, errors, when addressed seriously from an epistemic perspective, are a valuable source for exploring the nature and limits of justification in science.