– 03 –

Sources of Error

Sources of error • error types • observational errors • conceptual errors • discursive errors • error repertoires • error inventory • checklists and error checkpoints • error analytics

Just how do errors in science happen? If we intend to find them and fix them, we should have some idea of what causes them. Not generally or vaguely, but specifically. Experimental scientists are often aware of the common pitfalls in their equipment or experimental design. And they work to minimize the risk of encountering them. They remain on guard to spot them. Researchers refer to them as *sources of error*. In this chapter, I expand this notion from the laboratory to the whole domain of science.

How do scientists themselves interpret the occurrence of error? Sociologists Nigel Gilbert and Michael Mulkay asked a group of scientists in the midst of a controversy why *other* scientists—their opponents—were wrong. The list of reasons is rather colorful. The errors reportedly stemmed from:

succumbing to charisma
"strong personality"
a rhetorical "aura of fact"
"intellectual inertia"
confrontation with "unorthodox" views
being "tenacious"
being "dogmatic"
"pig-headedness"
"fear of losing grants"
dislike
"prejudice"
"subjective bias"
personal rivalry
"emotional involvement"
"threats to status"
"naivety"
"thinking in a woolly fashion"
"sheer stupidity"
an "ostrich approach" (of willfully disregarding the facts)
a whole generation simply "unequal to the task"

"Thinking in a woolly fashion": that's my favorite. Errors seem uniformly attributed to "unscientific" factors that interfere with sound reasoning. At least among those who disagree with you. Ironically, perhaps, few described their own errors in these ways—or even admitted errors. So, there is a tendency to equate error with irrationality, and irrationality with impinging psychological and sociological factors.[1]

[Gilbert and Mulkay, 1984, pp, 49, 65, 66, 71, 79, 81, 93, 96]

Others certainly echo these sentiments. Walter Gratzer in his book *The Undergrowth of Science* ascribes error to self-delusion, or credulity. Or *feckless* credulity. Or *unfathomable* credulity. For him, error is fueled by a fragile ego, ambition, desire for fame and advancement, envy, jealousy, and political ideology. Similarly, Robert Youngson, in his survey of *Scientific Blunders*, lists the roots of error as: carelessness, plain stubborn wrong-headedness, arrogance, willful and culpable ignorance, moral frailty, preoccupation with rewards (fame, status, wealth), and political, religious or other ideologies (at least). In characterizing *The Scientific Attitude*, philosopher Lee McIntyre castigates those with self-delusion or willful ignorance (not just ignorance, no!—*willful* ignorance!), along with the tiresome, gullible, and self-righteous charlatans. One might imagine, again, that "honest error" was impossible: an inherent oxymoron (Ch. 1).

[Gratzer, 2000, pp. 1, 23, 109, 306; McIntyre, pp. 155-166; Youngson, p. xi, xiii]

Unfortunately, these assessments focus on accountability, not on understanding error. As noted in Chapter 2, their primary purpose is to blame. Assign fault. Responsibility. Namely, *culpability*. In many ways, they are no more than declarations that an error has occurred, coupled with a harsh moralistic judgment. Political in spirit, not epistemic. Not diagnostic. And not fruitful for addressing the problem. If our ultimate aim is producing reliable knowledge, we need to focus on identifying the factors that, if addressed, could yield more reliable conclusions. We need an analytical posture.

Consider engineering practice. When a bridge collapses, or a rocket launch fails, the foremost response is to understand *why*, to avert future such calamities. Forensic teams investigate the structural design in detail. They probe the history of the construction, the maintenance, and the repair. They try to isolate the point (or points) of failure. In a similar way, military strategists study historic battles for key factors that might have turned the tide from loss to victory. Business managers study cases of entrepreneurial failure, looking for ill-fated decisions that might have been judiciously averted. Just so for science. If we can identify specifically how errors happen, we are better positioned to avoid them in the future.

[Petroski (1994, 2006)]

The engineering analogy is also useful for considering the comprehensive nature of thinking about error in science. Consider the complexity of trying to land a human on the moon in the 1960s. It was a formidable challenge—of multiple technologies, computing, and sheer project management. Everything had to go right. Quality control was utmost, because any one thing going wrong could jeopardize the whole flight mission. Witness the errors that did occur: three astronauts died, and three others almost did not return to Earth. Knowing about all the possible errors is essential in order to check for them and avoid them. That means assembling an immense catalog of such possible errors. So, too, for science. We might imagine "engineering"

---

[1] These reasons are not atypical to this controversy or field. Consider how another author characterized those who believe various 'pseudohistorical' claims. They may be: naive, biased, prejudiced, cynical, gullible, undiscriminating, unscrupulous, undisciplined, unorthodox, irrational, spiritual, flawed, fallacious, sensationalistic, amusing, quirky, eccentric, crazy, bizarre and embarrassing, pathetic, off-beat, audacious (or "almost unimaginably audacious"), outrageous, rhetorically clever, wild, extremist, over-eager, obsessive, manic, nefarious, reprehensible, and contemptible, not to mention communist and obfuscating (Fritze, 2009, pp. 8, 9, 10, 14, 16, 61, 66, 70, 83, 100, 136, 160, 164, 165, 167, 168, 169, 180, 183, 188, 202, 218, 219, 220, 253, 254).

more reliable scientific practice in part by documenting its many potential errors and developing safeguards against them.

If we apply a wide scope of vision to science, we can find errors in the lab, errors in reasoning about results, errors in peer review and reputational trust, errors in the community's entrenched assumptions or cultural biases. Faulty conclusions may arise from dirty glassware or contaminated reagents, from improperly calibrated instruments, from faulty statistical models, from mistaking correlation for causation, from small unrepresentative samples, from hasty generalization, from theory-laden interpretation of data, from blindness about the gendered nature of one's very questions. All are potential sources of error. All may compromise the reliability of science. At the same time, through careful analysis, each might be remedied. Sources of error, once identified, can be fixed—unlike the vague syndromes of prejudice, tenacity, or naivety, exemplified in the long list above. How would one "unwoolify" woolly thinking? Mapping out the scope and nature of those errors—and providing a useful scheme for interpreting and organizing them—is the aim of this chapter.

[Many sections are excerpted or adapted from Allchin, 2001; 2012sced; 2013tnos; Allchin & Zemplén, 2020; Höttecke & Allchin, 2020]

Error Types

So (again), how do scientists err? How might we begin to characterize the various error types?

Awareness of errors seems as old as science itself. For example, in the early 1600s statesman Francis Bacon proposed a grand image of inductive science. It would accumulate observations and facts, and then find the patterns among them. His vision helped inspire and guide how others built modern science over the next several decades. Yet alongside the ideal, he cautioned his readers that potential errors might lead one astray:

> The illusions and false notions which have already preoccupied the human understanding, and are deeply rooted in it, not only so beset men's minds that they become difficult of access, but even when access is obtained will again meet and trouble us in the instauration of the sciences, unless mankind when forewarned guard themselves with all possible care against them.

Forewarned is forearmed. Bacon described how our thinking can become muddled in several ways. He was essentially identifying psychological pitfalls centuries before there was any formal study of psychology.

Bacon named four general sources of error. He referred to them with fanciful imagery as "phantoms" or "illusions," or occasions where imagination might be taken as reality. They impair our ability to think clearly and effectively. Today, we might characterize them as cognitive biases.*[fn] First, there are illusions based on how humans tend to think (as a species, or "tribe"). We see patterns where none exists. We adhere to first impressions. We let wishes guide our perceptions, or attend unduly to the unusual or unexpected. Thus, some conclusions will be incidental byproducts of our thinking process, not genuine depictions of reality. Second, some errors stem from our personal idiosyncracies, unique backgrounds and tastes: illusions of the individual (or, in Bacon's cryptic metaphor, of a person's "cave" or "den"). As a result, some people focus on similarities, while other emphasize differences; some are oriented to details, others are more holistic; some are preoccupied with the past, some with the future. These dispositions diminish the objectivity of our conclusions. The third type of error is illusions of discourse (or of the "marketplace," where humans gather and exchange thoughts).

Communication itself, along with the ambiguities of language (as well as, perhaps, the use of persuasive rhetoric or "marketing"), can confuse us. They can "lead men away into numberless empty controversies and idle fancies." Finally, there were errors of dogma, or of comprehensive "systems" or overarching ideologies (in Bacon's exotic rhetoric, illusions of "the theater"). Bacon targeted a few cult-like theories familiar in his own time. But we still encounter grandiose theories and fads that tend to receive fervent support. Bacon certainly did not view those grand systems as conducive to good science. So, ultimately for Bacon, thinking could be adversely influenced by inherent cognitive deficits, individual bias, language, and unchecked adherence to programmatic ideas.

[Bacon (1620)]

> [*Note: Bacon's original term, the Latin *idola*, was once commonly translated as "idols." But modern usage of this term proves highly problematic for interpreting Bacon's meaning. *Idola* denotes "phantoms" or "false appearances," without the symbolic, religious, or devotional connotation inherent in the term "idol." Bacon characterized the *idolum* as a "corrupt and ill-ordered predisposition of mind" or "false notion"—hence "illusion," as translated here.]

One can certainly find resonances of Bacon's observations in modern cognitive science, as detailed below. However, Bacon—working before science was organized institutionally as "science"—did not provide many details about his four "illusions." Nor did he articulate strategies for remedying them. Indeed, his categories seem almost as puzzling now as the strange metaphors he used to label them. His obscure organization does not seem very helpful for informing contemporary scientific practice. Rather, they are mostly antiquated curiosities. Still, his effort was an important historical landmark in thinking about types of errors.

By contrast, if one walks into a lab today, one will likely hear researchers talk about systematic versus random errors. Random errors arise from the "noise" in the apparatus, the limits of measurement, and other variations in conditions. These are perhaps best characterized as *uncertainty*, rather than as true errors (Chapter 2). Identifying them in this way allows one to manage them mathematically, through statistical analysis. Systematic errors, on the other hand, are trickier. They arise from flawed theoretical assumptions, poor experimental design, or just plain ignorance of relevant variables. They tend to alter all the results in the same way—not randomly. But by their very nature, these sources of (true) error are unknown, hence very hard to detect. Not surprisingly, they concern experimentalists the most. Unfortunately, there is no commonly accepted or comprehensive list of the systematic errors, so the value of this categorization is limited. (For an illustration of the interplay of random and systematic errors, see the history of the Hubble constant described in Chapter 2.)

There have been various other efforts to conceptualize error types. Each offers a glimpse of the challenge and the issues involved. For example, Giora Hon approaches the problem from the perspective of an experimentalist. He catalogs the errors as one might encounter them sequentially in the course of an investigation. Deborah Mayo approaches the problem as a philosopher of statistics and proposes four categories of *canonical errors*: "(1) mistaking chance effects or spurious correlations for genuine correlations or regularities; (2) mistakes about the quantity or value of a parameter; (3) mistakes about a causal factor; and (4) mistakes about experimental assumptions." But as Mayo herself notes, this short list is hardly exhaustive. It is informative, but also rather ad hoc. Stephan Guttinger and Alan Love, focus on the philosophical context of failures to replicate earlier published experiments. They distinguish between theoretical failures (identified merely as overgeneralization) and methodological failures —

notwithstanding fraud as a possibility. The Catalogue of Bias Collaboration, by comparison, is dedicated to promoting evidence-based medicine, and so address the biases that can sabotage clinical research. They sort the cautions into the stages of a study, from study design to selection of data, to the behaviors of researcher and patient, and finally, reporting.

Kathyrn Schulz, a journalist who ventures into “wrongology,” echoes Bacon’s mentalist approach, and sees errors as based in the senses, mind, or social interactions. Each taxonomy brings some features to light. However, what we need is something that is simultaneously more comprehensive and more detailed than any of these options. We need something that encapsulates the vast reach of epistemic processes in science and provides some framework for organization and for assessing completeness.

[Catalog of Bias Collaboration, 2020; Guttinger & Love, 2019; Hon, 1989; Mayo, 1996, pp. 18, 51 n.17, 150, 453-54; Schulz, 2010]

A simple yet synoptic approach tracks the genesis of scientific claims. Namely, how are the justifications generated, shared, and assessed? What is their *ontogeny*? What ensures reliability at each step, as each may prove important in different cases? This process exhibits a trajectory, of sorts, that begins with the documenting of natural phenomena — yielding “raw data” — and extends through establishing agreement across the community of scientific experts. Intermediate steps involve, for example, properly functioning laboratory equipment, careful observational skills, relevant controls, valid theoretical assumptions, sound reasoning, appropriate statistical models, meaningful analyses and graphical representaions and, eventually, vetting through the contrasting theoretical lenses of other experts. And much more, besides (as described below).

Imagine, then, an epistemic path beginning with the most basic observations or measurements and leading to a generally accepted scientific claim. Simple measurements are first assembled into meaningful graphs. Observations are arranged into significant patterns. Order begins to emerge. But already there are many occasions to check for reliability. Are the samples and reagents free from contamination? Has the instrument been designed properly? Has it been properly calibrated? Has potential observer bias been prevented? Is the experiment designed to control for or monitor possibly confounding variables? If all is secure, one can next compare patterns in parallel sets of data, and apply statistical analysis. But are the statistical measures appropriate to the type of data set? Patterns and numerical trends can then be set in the context of models or theoretical explanations. But have alternative explanations been fully considered? Are there conceptual blind spots—from a researcher’s theoretical framework or cultural context? When significant findings have been established, they may be published—and reach other investigators for criticism or novel development. Yet one may wonder whether peer review has been both suitably critical and fair. Even so, results are gauged against reputations. Has the system of credibility succeeded, or has star-like prestige or a conflict of interest distorted the perceived significance of particular claims? At this stage, critical exchange, when functioning well, isolates errors and promotes their remedy. Perhaps it guides further investigation and the generation of critical missing evidence. There may be reviews of the literature or meta-studies, synthesizing and consolidating information even more. An informal consensus develops. Formalizing that consensus does occur occasionally, through explicit workshops or expert panels assembled to clarify warranted claims for medical practice, say, or for public policy. Scientific knowledge thus emerges from a vast trajectory, from disparate clusters of observations to a consensus of relevant professionals. As a claim unfolds, it reflects longer and longer chains of justification, as the raw evidence is assembled into larger and larger networks. At the same time, it knits together more and more local phenomena and observations into successively more global perspectives and concepts. Quite an extensive process, when examined closely. And any of the successive steps may go awry. The process of justification may falter anywhere along the way. Namely, any single step in this vast ontogeny may be a source of error.

[On networks and chains of transformations, see Latour, 1987]

In some cases, scientific claims become relevant to the actions of citizens, customers, government agencies, or corporate leaders. These claims then continue their trajectory, traveling even farther from the original inscriptions in lab and field notebooks. They may appear in the news, in legislative hearings or courtrooms, in advertising, or in the media. They may be reported again via casual conversations, blogs, webpages, or social media. Epistemic concerns here are no less important. Do the final claims ultimately reflect a professional consensus? Do they reflect expertise? Is the testimony, or the person reporting it, credible? Is it honest, and free from the adverse effects of conflict of interest? The reliability of scientific claims in a cultural setting depends on effective science communicationas much as sound research. Ultimately, the ontogeny of a scientific claim involves a long trajectory — from test tubes to YouTube; from lab bench to judicial bench, from field site to website, from lab book to Facebook. The epistemic concern about preserving the reliability of the claims persists throughout, even beyond the scientific community. No one should understate the challenges of preserving the integrity of scientific information in the public realm. For the analysis in this volume, however, I will focus on the development up through a scientific consensus — that is, what happens largely within the professional scientific community. Still, one might profitably keep the larger context in mind.

Here, then, we have a comprehensive view that can fruitfully guide an analysis of error types. Follow the ontogeny of building and justifying scientific claims. Consider its many individual steps from the original empirical observations. The final claim and its justification will be an abstraction of the series of stages to produce them. Thus, it will be more logical than chronological. But it will still embody the many layers of construction.

Further, one may conveniently intepret this conceptualization in terms of a visual map metaphor. Many philosophers characterize science as "mapping" some aspect of the physical world, whether to characterize its structure abstractly or to shape it technologically. That is, a graph, a claim, a concept, a model, and a theory are forms of representation. Or, literally, a *re*-presentation. Necessarily selective, but corresponding to observed phenomena in stipulated ways. (Note that a map functions through various conventions about how the map represents, so it is not without inherent context.) So: the phenomena are the territory. Scientists "map" that empirical "landscape." But, as noted above, they do this in stages. They develop maps (say, data points), then maps *of* maps (such as graphs), and then perhaps maps of *those* maps (such as correlation coefficients), and more maps (linkages to causal conditions, explanatory frameworks, and so on) all in successive layers. In principle, all the maps are ultimately grounded in, and traceable to, their observational benchmarks (the original measurements or other collected data). Maps may thus be arrayed along a spectrum from *local* to *derived*, reflecting the layers of transformations from the original world. As maps are compounded into meaningful patterns, they usually incorporate or rely on a broader base of observational benchmarks. A map interpretation typically occurs in the context of other observations. Transforming one map into another relies on another body of evidence. Thus, as maps become more derived they also tend to become more *global*. The local-derived scale thus simultaneously embodies a local-global axis, stretching from particular facts to general ideas, or theories.

[Allchin, 1999psa; Bauer, 1992; Giere, 1998, 2006; Hacking 1984, esp. pp. 208-209; Judson, 1980; Latour, 1987, pp. 80-83, 210-57; Turnbull, 1989; van Fraassen, 1980; Ziman 1978]

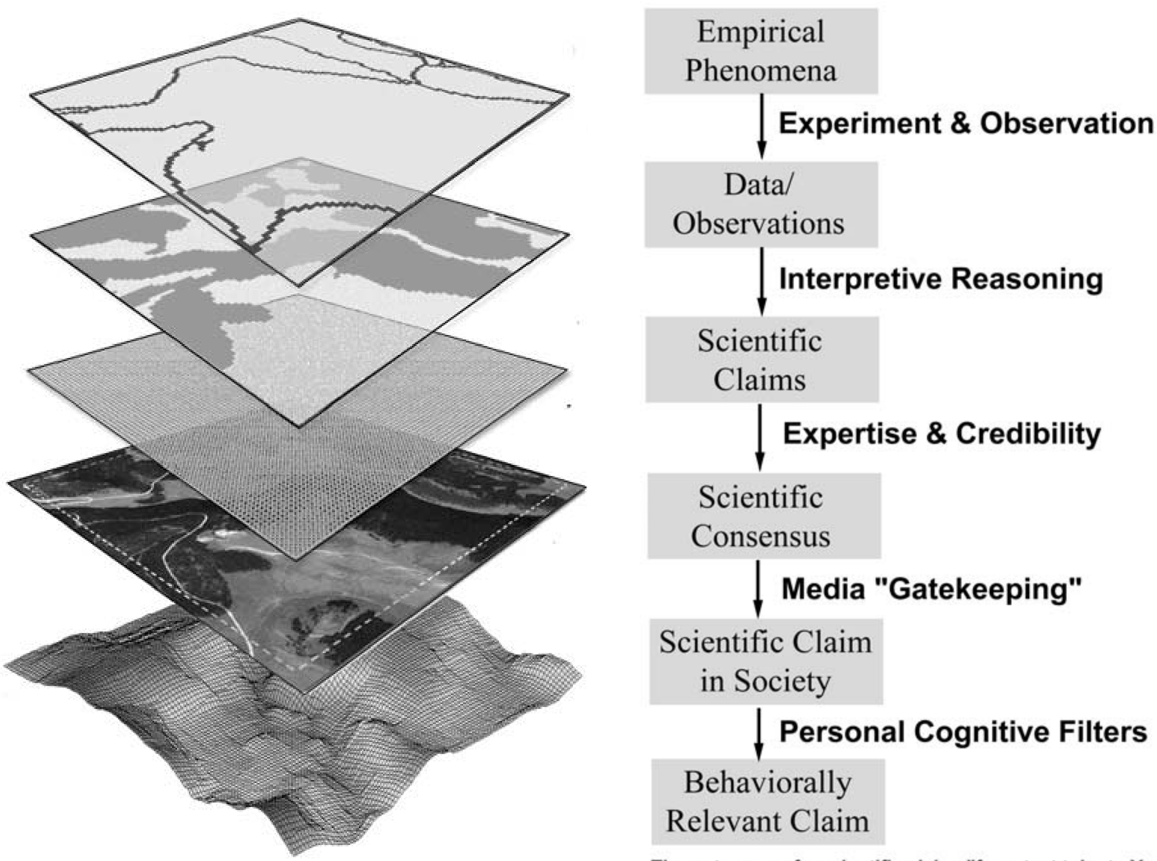
Arranging empirical benchmarks together in a figurative map requires interpretation and, with it, justification. Of course, each step in this process is subject to error. In the map metaphor, an error may thus be characterized as:

a faulty mapping that does not preserve the structure of the world as intended.

That would apply to any process that transforms data or information into another form or expression (from one layer of mapping to the next). Identifying an error indicates that some element of the correspondence process did not work. The map fails. These particular factors are the sources of error that we hope to catalogue here.

Readers familiar with traditional philosophy of science may note that this approach does not adopt as fundamental the justificatory relationship between theory and evidence (or theory and observation). Nor does it focus primarily on logic or other forms of argument. That is, I forego simplified abstract frameworks in favor of the human texture of authentic scientific practice, with all its context and complexities, a view more familiar to historians, sociologists, journalists, and scientists themselves. How is error experienced in a lab or field site? How do scientists address errors in casual conversation, say, or in the pages of *Science* magazine or other journals? I regard those as the relevant contexts and benchmarks (see also the Preface).

One may sort the ontogeny of a scientific claim into three rough stages of mapping: (1) observation, (2) conceptual interpretation, and (3) discoursive assessment. The first is intimately connected to the material processes of the lab or field site: documenting concrete phenomena — observation. The second is largely the cognitive work transforming that empirical information into ideas, typically organized in a publication or “report” for presentation to peers — conceptualization. The last involves the interactions of scientists in a social community of fellow experts, who assess, discuss, revise, and regulate those ideas — discourse. These stages form very general categories for organizing an inventory of error types, or general sources of error, as described in the next three sections.



The ontogeny of a scientific claim, "from test tube to YouTube," "from lab bench to judicial bench," "from field site to website"