Studies in History and Philosophy of Science 119 (2026) 102216

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa

When is science "self-correcting"?

Douglas Allchin

Minnesota Center for the Philosophy of Science, University of Minnesota, Minneapolis, MN, 55455, USA

ARTICLE INFO

Keywords:
Error
Self-correction
Error correction
Peer review
Replication
Self-policing
Accommodation
Error cascade

ABSTRACT

Standard rhetoric portrays science as self-correcting. Analysis of numerous historical cases helps clarify concretely when and how corrections in science occur, and when they do not. We may thereby articulate a general set of conditions necessary (although not necessarily sufficient) for *when* science may be considered self-correcting.

1. Introduction

Science progresses in part by discovering its own earlier errors and replacing them with new interpretations based on deeper evidence and fuller justification, sometimes precipitating episodes of revolutionary conceptual change (e.g., Kuhn, 1970). But just how do scientists err? How do they then find and resolve the errors? These important questions have helped spur the emergence of a new philosophy of error in science (e.g., Allchin, 2026; Buchwald & Franklin, 2005; Holton & Mack, 2005; Karakas & Tuboly, 2026; Mayo, 1996). Here, I am interested primarily in how historical study can inform our understanding of how and when errors are corrected in actual scientific practice (in contrast to abstract arguments or norms about how science "should," ideally, function).

Of course, philosophers have a long tradition of regarding science as "self-correcting" (e.g., Bruner & Holman, 2019; Darden, 1991; Forster, 2002; Mayo, 1996; National Academies of Sciences, Engineering and Medicine, 2019; Vorland et al., 2020). For example, Herbert Feigl noted in 1949 that "the self-corrective aspect of science has rightly been stressed as its most important characteristic" (Sismondo, 2025). For some, self-correction helps demarcate science from other intellectual endeavors (e.g., McIntyre, 2019; Zimring, 2019) — a benchmark that implies that correction is systematic, or based on some clearly articulated mechanism. This confidence is eagerly echoed by editors of *Science* and other science advocates, who often portray self-correction as the basis for public trust in science — further implying that correction is timely, admitting no sustained lapses or gaps of uncertainty (e.g., Alberts et al., 2015; McNutt, 2020; Tharp, 2023). Given the ubiquity of this posture, as well as its central importance, it seems appropriate to clarify what is meant by "self-correcting" and what, if any, conditions may affect its effective functioning.

Recently, the canonical view of self-correction has been sharply criticized. For example, failures to replicate several landmark studies have shaken faith in the reliability of the published literature, with Ioannidis (at least) contending boldly that "most published research findings are false." Journals and blogs brim with fretting over a reproducibility crisis (e.g., Berg, 2019; Ioannidis, 2005, quote on p. 1; 2012; Pashler & Harris, 2012). The processes of peer review and replication — conventionally regarded as the stopgaps against error — seem limited or in disrepair (e.g., Baldwin, 2019; Broad & Wade, 1982; National Academies of Sciences, Engineering and Medicine, 2019; Science and Technology Committee, 2011; Tennant & Ross-Hellauer, 2020). Even science journalists now report that "science is broken" (e.g., Engber, 2017). While many errors are surely corrected, we need more concrete insight into how — and precisely when — correction indeed occurs.

A historical perspective provides more challenges. The history of science is littered with abandoned concepts and falsified theories—chemical affinities, mesosomes, black bile, ether, hysteria, coronium, the planet Vulcan, a universal genetic code, the initial values for e or H_0, caloric, geological racism, and more (e.g., Buchwald & Franklin, 2005; Allchin, 2026; Holton & Mack, 2005; Losee, 2005). True, each documented error is now ironic witness to correction of some kind (e.g., Schimmack, 2019). But that may easily blind us to the discomforting status of unknown errors still waiting to be discovered, *not yet* "self-corrected" (the hidden contrast cases). Indeed, some historical

E-mail address: allchindouglas@gmail.com.

https://doi.org/10.1016/j.shpsa.2026.102216
Received 30 November 2025; Received in revised form 10 May 2026; Accepted 22 August 2026
0039-3681/© 2026 The Author. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

errors lingered for years, decades, or even centuries (Allchin, 2015; see Table 1 and discussion in Section 3). What should a well-informed notion of "self-correction" say about *timeliness* (e.g., Lindsay, 2011; Rohrer et al., 2021)?

In some cases, errors have been "corrected," only to yield to other errors because scientists failed to effectively isolate the underlying source of error. For example, anthropologists tried repeatedly to define human uniqueness in terms of tools, finessing their response each time that naturalists found animals that exhibited the purportedly exclusive human trait. So, too, for language use (Allchin, 2013, pp. 132-140). In the 1950s-60s biochemists proposed—and rejected in turn—a series of 16 high-energy intermediates of oxidative phosphorylation. Yet the very notion of an intermediate molecule with a high-energy bond itself proved incorrect, with the chemiosmotic theory ultimately adopted in its place (Allchin, 1997). Such *error cascades* (Allchin, 2015b) seem to indicate a contingent process. If we acknowledge that there is no singular "scientific method"—and that discovery is never guaranteed—should we be similarly skeptical about the promises of a corresponding "self-correcting" method? For instance, many historical errors were famously remedied only by sheer happenstance (Table 2). History invites us to consider: should a "self-correcting" process be *systematic* and/or governed by an explicit and *reliable mechanism* or causal process (*sensu* Machamer et al., 2000; e.g., Vazire & Holcombe, 2022)?

In yet other cases—many familiar in informal disciplinary folklore—errors are corrected, only to be rejected by the scientific community (Table 3; Allchin, 2015b). For example, Amodeo Avogadro successfully solved the errors manifest in conflicting molecular weights early in the 19th century by conceptualizing diatomic elements and the law that now bears his name. But Berzelius, Dalton and others at the time dismissed it as inherently implausible and the solution languished for over half a century (Levere, 2001, pp. 110-116; Rocke, 1984, pp. 101-107, 294-299). These cases challenge the conventional wisdom that *peer review* is effective in guaranteeing self-correction. Sometimes, credibility bias or communal confirmation bias at the social level can be its own source of error.

How, then, are we to address these contrasting views about self-correction? Yes, corrections surely do occur. But significantly, the historical and contemporary evidence raises questions about whether corrections are (a) timely, (b) systematic, (c) consistently reliable (mechanistic or causal), and/or (d) resolved through peer review (Allchin, 2015b). Is science truly "self-correcting" in any meaningful sense? What does "self-correction" in science ultimately mean?

Competing blog headlines and paper titles attest to an ongoing and testy back-and-forth about whether science is (or is not) self-correcting. Here, I hope to avoid such contentiousness by reframing the debate and asking instead (in a more particular and tractable way): "*When* is science 'self-correcting'?" This orientation is reflected, for example, in the NanoBubbles (2021) project, which asks specifically, "How, when and

Table 1
Some notable examples of long-standing errors in science.

Errors (chronological by date of discovery)	*Duration of Error (Years)*	*Source*
Harvey's denial of observable capillaries (1649-1661)	12	Elkana and Goodfield (1968)
California as an island (1622-1701)	79	McLaughlin & Mayo, (1995)
Newton's speed of sound formula (1687-1802)	115	Finn (1964)
particle theory of light (1704-1840s)	140	Cantor (1975)
oxygen theory of acidity (1778-1810)	32	Crosland (1973), Le Grand (1972), Le Grand (1974)
affinity theory of chemical composition (1718-c.1811)	93	Lehman (2010)
caloric theory: heat as a substance (1777- mid-1840s)	65	Chang (2009), Fox (1971)
humoral medicine (4th century BCE - mid/late-19th century)	2000+	
electricity as a fluid (or *two* fluids) (~1732 - ~1897)	165	Heilbron (1979)
Kelvin's thermodynamic limit to the age of the Earth (1862 -1903)	41	Hallam (1989), Burchfield (1990)
Newtonian mechanics (for relativistic velocities and subatomic masses) (1687-1905)	218	
tetranucleotide structure of DNA (1902/1909–1949)	~40–47	Judson (1980)
human chromosome number as 48, not 46 (~1923-1956)	33	Martin (2004), Kottler (1974)
synthesis of proteins by reverse proteases (~1933–1961)	~30	
noble gases as inert (non-reactive) (1920-1962)	42	Gay (1977)
stress as the primary cause of ulcers (mid-1930s-1984)	~50	Thagard (1999)
central dogma (reverse transcriptase) (1958-1970)	12	
dinosaurs as (exclusively) "cold-blooded" (1824-1968)	144	
mesosomes as bacterial structures (1953- ~1976)	23	Allchin (2022)
high-energy intermediates of oxidative phosphorylation (1953-~1977)	24	Allchin (1997)
lungs as evolved from the swim bladder (1859-1978)	119	Liem (1988)
central dogma (ribozymes) (1965-1982)	23	
viceroy as a Batesian mimic of monarch butterflies (1869 -1991)	122	Ritland and Brower (1991)
stress & diet as primary causes of ulcers (mid-1930s-1984)	50	Thagard (1999)
central dogma (prions) (1965-~1997)	32	Liu (2011)
Pluto as a planet (1930-2006)	76	Weintraub (2007)

Table 2
Cases of scientific corrections initially rejected by the scientific consensus.

Rejected Error-Correction	*Duration of Error (Years)*	*Source*
Young's correction to Newton's corpuscular theory of light (1801-mid-1840s)	40	Cantor (1975)
Avogadro's correction to atomic theory & atomic wgt. measurements (1811-1860)	49	Rocke (1984), Levere (2001)
John Snow's correction to the miasma theory of cholera (1849-1866)	17	Eyler (2001); Johnson (2006)
Semmelweis's correction theory of puerperal fever (1847- 1870s)	23+	Carter (1983)
DNA as hereditary material [Avery's "transforming factor"] (1944-1951)	7	
Wegener's correction to static Earth crust (~1922-1965)	43	Glen (2002), Le Grand (1988), Hallam (1989), Oreskes (1999)
chemiosmotic hypothesis of oxidative phosphorylation (1961-1974)	13	Allchin (1997)
"warm-blooded" dinosaurs (1968-~1975)	7	
Scablands mega-flood (1923-~1965)	42	Soennichsen (2008)
rejection of dark matter (Zwicky) (1933-late 1980s)	~50	
rejection of meteorite impact hypothesis of mass extinction(s) (1980-1991)	11	Glen (1994, 2002)
rejection (in the U.S.) of acupuncture as an analgesic (1972-1997)	25	Bowers (1979)
rejection of prions as correction to central dogma (1982-1997)	15	Liu (2011)

Table 3
Some notable error corrections based on happenstance (or "chance" discoveries).

Error Discovered	Context of Identified Error	Context of Investigation
lunar craters & sunspots lead to correction of error of sun and moon surfaces as perfect and unblemished; discovery of Jupiter's moons leads to correction of Earth as uniquely having moons; both challenge Ptolemaic system	astronomy	optics/lenses; telescope as a new optical instrument
observed capillaries leads to correction of Harvey's denial of them	blood circulation	fine structure of the lungs (Malpighi)
adiabatic phenomenon lead to correction of Newton's formula for the speed of sound	speed of sound/ physical constants	nature of heat and temperature
precise measurements of density of nitrogen leads to new element and major correction of the periodic table	periodic table/nature of atomic elements	fundamental constant (density/atomic weights)
multiple "polar-wandering" curves indicate continents drift, reviving rejected theory	Earth geomorphology and dynamics	history of Earth's magnetic field
reverse transciptase leads to correction of central dogma of molecular biology	molecular biology from DNA to proteins	RNA viruses and cancer
discovery or iridium layer leads to meteorite impact hypothesis of mass extinction, correcting both adaptive theories and predominant notions of gradualism	paleobiology and mass extinctions	geological dating
discovery of prions, further upsets the central dogma	central dogma of molecular biology	rare human and animals diseases

why does science fail to correct itself?" Similarly, Vazire and Holcombe (2022) ask, "*Where* are the self-correcting mechanisms in science?" (my emphasis) and then propose concrete "observable self-correction indicators" as independent variables that each foster error correction. Likewise, Ulrich Schimmack (2014) has endeavored to quantify varying degrees of credibility in a field-specific "replicability index.". All pose questions about the assumptions, institutional contexts, or boundary conditions for error correction to be effective. Here, my interests extend well beyond peer review or replication, standardly viewed as the relevant scope for thinking about error correction. In addition, I frame the question primarily in terms of concrete history, not abstract, universal philosophical principles — although the aim is still to establish informative general insights. Again: *When* is science "self-correcting"?—or, more plainly, *when* are errors in science corrected, *and when does the process falter (and how)*?

Such an inquiry is prospectively quite ambitious: encompassing the whole history of modern science. An exhaustive study would be formidable methodologically, not the least because we have no comprehensive list of the errors that might be relevant (what even counts as a meaningful error?). How can one ever hope for a systematic or representative sample? The whole project seems mired with potential error and bias (perhaps like any account of the history of science?). An authoritative historical study of error or error correction seems nigh impossible.[1]

At the same time, the aim here is far more modest. I am interested merely in identifying relevant significant variables: a set of necessary — but hardly fully sufficient — conditions for when corrections occur. These can be inventoried through analysis of individual cases. That is, if an error was not corrected immediately, what factor(s) seems to have made a difference, which emerged only later? How? Here, any historical case can be informative, provided one can document: (1) the original source of error, and (2) the dynamics of finding and resolving that error, over what time, and under what circumstances. (For example, did peer review matter? Was there a "failed" replication? If so, how was the specific error isolated?) Namely, one simply needs a set of case analyses with the appropriate historical details (see the cited references for each case below). More cases, more data, greater scope. Here, I draw from my experience collecting informative cases over several decades. The sample is (statistically speaking) haphazard: a sample of convenience. But informative, nonetheless — in part, because many factors have been encountered on multiple occasions, indicating some level of importance (see Sections 9–10). More historical work may surely reveal more relevant contingencies. Let this be an initial effort at least, subject (of course) to further evidence and subsequent revision. The Reader with further concerns about methodology may consult the discussion in the appendix.

First, I address several preliminaries about conceptualizing (or defining) error and various interpretations of "self-correction" (Sections 2–3). Then I consider various approaches to error and error correction (Section 4). In the next four sections, I describe a hybridized general history of error correction, highlighting various functional stages. These include the emergence of an incongruence (Section 5), noticing the problem (Section 6), isolating the source of error (identifying the error type) and possibly testing it (Section 7), and vetting the resolution at the community level (Section 8). I conclude with a skeletal model of the process (Section 9) and articulate the various contingencies that seem to shape effective error-correction in science across its various stages: namely, *when* science is self-correcting (Section 10).

2. Conceptualizing error

What is error itself? Even here, philosophers and others express divergent views. In this section, I simply clarify my use of terms. (Some readers may prefer to jump ahead to Section 3 or 4, and return to this section only as needed).

For the purposes of the analysis here, an error is *a scientific claim whose justification is later found to be unwarranted*. The concept is necessarily diachronic. Namely, errors are errors precisely because their flaws go unnoticed. Accordingly, their status as errors can only be known and attributed in retrospect. The focus is on the justification and the negation of the original claim, typically accompanied with an alternative (replacement) claim (see also Stegenga, 2024, 2026). (Thus, while there are many methodologies that aim to *reduce* the incidence of error through checking for sources of error, they are not strictly relevant to the cases where errors ultimately occurred and, hence, needed *correction*.)

An error (here) is not necessarily any false claim or belief. A claim once considered true and later deemed false, is an error (e.g., caloric, nebulium, mesosomes) because the consensual justification shifted. So, too, for a claim once rejected as false and later discovered to be true (e.g., Avogadro's hypothesis; continental drift; see also Table 2). The error lies in the mistaken *justification*. Ironically, then, resolved errors constitute a form of knowledge: *negative* knowledge (e.g., Allchin, 2026; Gartmeier, et al., 2008).

Epistemic posture, or commitment, matters. A hypothesis or a speculative (or "tentative") claim, explicitly acknowledged as uncertain, does not yield an error. (In a sense, the basic element of belief has been withheld.) Still, a partial justification that was once accepted as appropriately warranted and then later dismissed, is an error.

Note that this conceptualization is *epistemic* and *descriptive*, eschewing any normative methodological judgment. All too often, "error" seems to connote a *mistake* (in the sense of methodological lapse or "wrong" action in lieu of an acknowledged "right" one), thereby implying (explicitly or not) incompetence, negligence or other culpable

[1] This would apply equally to those who contend (without systematic evidence) that science is self-correcting.

behavior (e.g., Hon, 1995). I focus exclusively on claims and their justification, not on any scientist's behavior ("right" or "wrong"), nor on deviation from any prescribed methodological norm. Scientific malpractice and misconduct (or other evaluative notions of "failure") are distinct, epistemically, from errors (Anderson, 2022).

Note, too, that an error is distinct from a *source of error* (in the sense of the term commonplace in experimentalists' discourse). A source of error is an instance of a general error type (e.g., a contaminated reagents, biased samples, confirmation bias, heuristic gaps). An effect (error) should obviously not be conflated with its cause (the source of error, or error type). (On error types, see Allchin, 2001).

A scientist's recognition that something is "amiss" is not (yet) an error, although it strongly indicates a latent error, still to be isolated and fully identified (e.g., Hon et al., 2009). For example, a failed replication is not an error. Not yet. The error could be in the methods or reasoning of the original experiment, or it could be in the replication effort (e.g., the ironies in the case of Joseph Priestley and the restoration of air; see Section 3). Thus, incongruences of various sorts — a discordance among observational results, a theoretically anomalous finding, or a conceptual disagreement or ambiguity — are not errors. They are merely indicators that further work is needed to identify a prospective error (see also Sections 5 and 7 below).

The concept of correcting an error can be just as slippery as the concept of error itself. Here, an error correction is a resolution to one of the incongruences just noted, wherein an error is traced (or "isolated") via further evidence to a particular source of error in the justification (and perhaps tested and confirmed), thereby accounting for the incongruence and indicating the "correct" way to interpret the evidence. Through correction, the problematic claim becomes *known* as a definite error (again, knowledge in a negative sense). This view of error couples conventional conceptual perspectives with the experimental and social dimensions of scientific practice into one synthetic framework (Allchin, 2001). Counterintuitively, perhaps, correcting error involves *epistemic work* and results in (justified) *negative knowledge.*

3. Conceptualizing self-correction

How are we to conceive the term "self-correcting"? This can be problematic, as well. Here, we need to disentangle the concepts of self-correction, self-checking, self-policing, and adaptive accommodation, at least.

The classic model of a self-correcting system is a servomechanism. Canonical examples include centrifugal governors on steam engines, anti-aircraft gunnery, and thermostats. All embody the principle of negative feedback. Other familiar examples include homeostatic systems in organisms, an automobile's cruise-control, or a market-governed equilibrium processes in economics. When disturbed by environmental conditions, they register an "error" or deviation, and guide the system to its target (former) state. Each has an explicit causal *mechanism.* Its components standardly include an assigned setpoint, a sensor, a comparator, and an effector. Together, they determine a *systematic* process which can ensure correction in a *predictable* and *timely* manner. These features (also noted in the Introduction) seem requisite to the concept of genuine self-correction, construed causally.

Of course, this is a conspicuously poor model for science. Science has no "setpoint" or external *telos* defining the ultimate state of the system. There is no convenient touchstone or independent assay for "truth" or reliable knowledge. Indeed, science regularly experiences conceptual change, with new theories and concepts sometimes displacing old ones. At best, one can say that there is a temporary "equilibrium" when there is simultaneous coherence: (a) between different sets of findings, (b) between theories and data, and (c) between different theoretical perspectives. Science, it seems, may need to settle for something less than genuine self-correction.

One challenge for articulating self-correction in science is defining the "self" responsible for the correction (e.g., Sismondo, 2025). In what sense can science, as either an abstract method or concrete institution, be said to have causal agency or intention? Perhaps, as noted by Oreskes (2019), echoing the principles of social epistemology, one should say not so much that science corrects *itself*, as that scientists correct *each other*. That is, the hope is that scientists compensate for each other's blind spots (e.g., Hull, 1988; Longino, 1990; Merton, 1973; Oreskes, 2019; Ziman, 1968, 2000). The shift in subject may seem subtle; however, it has important overtones for how (and when) science can be viewed as correcting "itself," without anthropomorphizing science.

Conventional views of self-correction in science may be found in Hull (1988) and Merton (1973), among others. For a recent representative expression of the "self-correcting thesis" (SCT), one may turn to Romero (2016): "In the long run, the scientific method will refute false theories and find closer approximations to true theories" (p. 56). This is a causal claim: science *will* correct itself. Yet Romero's inclusion of the phrase "in the long run" is problematic. Assertions about self-correction are typically couched in such provisos, qualified by such terms as "eventually," "ultimately," "sooner or later," "in time" or some such disclaimer (e.g., McIntyre, 2019, pp. 134, 138, 171, 175; Zimring, 2019, pp. 79, 124, 229, 233, 234). "Eventually," I contend, is vague and functionally hollow. Ultimately, conceptualizations of this type (and they are widespread) acknowledge the ubiquity of contingencies in error correction (as noted briefly in the Introduction). That is, the strong causal claim seems unsupported.

Here, we may briefly consider the two mechanisms frequently cited as responsible for effecting self-correction in science: (pre-publication) peer review and replication. Do they provide a complete causal account? First, peer review is supposed to function much like the sensor in a servomechanism: detecting methodological lapses before they are published. Presumably, this protects the published "scientific record" from error, as expressed by McIntyre (2019): "Having someone else review work before publication is the best possible way to catch errors and keep them from infecting the work of others." It is "a bedrock principle of science" (pp. 99, 101). (One may note, however, that the inherent aim here is more quality-control than error-correction — a gate-keeping strategy, not a remedy.) If the peer review fails as a conceptual filter, efforts to replicate the original experiment are expected to fail, thereby exposing error empirically and (presumably) making clear the particular source of error and its remedy.

The limits of both these processes have been recognized for decades (e.g., Broad & Wade, 1982). Expert panels have now expressed doubt whether either of these do (or can) bring attention to significant *errors*, in contrast to recognizing known missteps, or mistakes (e.g., National Academies of Sciences, Engineering and Medicine, 2019; Science and Technology Committee, 2011). These deficits deserve far fuller historical treatment than can be accommodated here. But a few key points should be mentioned.

First, does the reality of peer review live up to the widely touted normative ideal? This is an empirical question, and it has been tested. For example, in a 1977 study, 75 psychologists, whose theoretical positions were a matter of record, were sent the same bogus manuscript with a glaring error. 71% of the critically-stanced reviewers found the error; only 25% of the sympathetic ones did (Mahoney, 1977). Namely, confirmation bias intrudes and mistakes were not identified uniformly across all reviewers. The editor of the *British Medical Journal* summarized other investigations: "At the BMJ we did several studies where we inserted major errors into papers that we then sent to many reviewers. Nobody ever spotted all of the errors. Some reviewers did not spot any, and most reviewers spotted only about a quarter" (Smith, 2006, p. 179). Other blind studies have obtained similar results (e.g., Armstrong, 1997; Cicchetti, 1991; Rothwell & Martyn, 2000). A House of Commons committee found, nonetheless, that peer review generally improves the quality, transparency, and readability of a manuscript. While many authors complain about the system, most also agree that the quality of their own published paper was improved as a result of review. But in terms of correcting individual errors, it is a crude filter: a social-level

heuristic. "Peer review and refereeing are at best coarse screens." They are not the "infallible, fine discrimination systems" that many suppose (or want?) them to be (Science and Technology Committee, 2011, pp. 9, 15). In contrast to the widespread folklore about its efficacy, peer review is inherently limited in regulating error (Baldwin, 2019; Tennant & Ross-Hellauer, 2020).

Peer review is also sometimes interpreted to include, more broadly, all subsequent peer assessment. However, the secondary process is not blinded and does not precipitate any explicit selectivity. Thus, we should refer to this later stage (reflecting Merton's norms and Oreskes' comment above) as *critical discourse*, not peer review. The limits of this process will be addressed further in Section 8.

Problems also haunt the ideal of replication (e.g., Feest, 2019). Reproducibility is certainly a benchmark for productive science, allowing scientists to build on the stable recurrence of results based on clearly specified protocols. Yet the success of a replication, alone, does not indicate how to interpret the results. Only that they are repeatable. Some results may be misleading, based on artifacts of the investigative or observational process (Craver & Dan-Cohen, 2024). Artifacts may be a significant source of error. And, as Walter Gilbert once cautioned, "artifacts are very, very reproducible" (Judson, 1980, p. 170). Sadly, perhaps, replications may easily echo the original error. There are plentiful historical examples of artifacts and misinterpretations being quite reproducible. Dozens of researchers attested to producing and observing N-rays (Nye, 1980). Likewise for Allison's magneto-optic effect (Epstein, 1999; Slack, 1935). "Anomalous water" (polywater) was reproduced by many chemists (Franks, 1981). Microscopists generated mesosomes (spiral membranous structures in bacteria) for 23 years before deciding that they were not authentic structures in native bacteria (Allchin, 2022). Even today, mesosomes are perfectly reproducible.

Even some early efforts at replicating Pons and Fleischmann's bogus tabletop fusion proved "successful." No such success occurred at Caltech, because they were trying to do the experiment as it "should" have been done. Not until they abandoned efforts at rigorous protocol and, ironically, tried to literally *replicate* the original apparatus exactly, were they able to achieve similar results—which now they could attribute to misplaced thermometers and the failure to circulate the fluid in the battery cell (Barnes, 1989; Lewis et al., 1989; Smith, 1989). Debunking the pretensions of cold fusion required both producing the error—a strict replication—*and comparing it to other results through controlled experiments*. In short, errors can be replicated, with no telltale signs of an experimental artifact, or error. *Replications do not necessarily reveal errors*, especially if the original errors are themselves replicated.

Nor does a "failed" replication necessarily signal or pinpoint an error. For illustration, consider Joseph Priestley's investigations on the restoration of air by plants (Allchin, 2012). Originally, he documented that a sprig of mint and other plants could yield a form of air that allowed candles to burn and mice to breathe longer. His work was recognized by the Copley Medal. Others tried to replicate his results, yet many failed. This only showed a discordance of results. Namely, *whose* results were in error? Replication alone provided no clues. Which version was "correct"? That required further investigation. Errors do not announce themselves, and "failed" replications rarely reveal what is "wrong," only that something seems awry (see Section 5 below on incongruences).

Consider the sequel. Years later Priestley himself, now working in a new home and laboratory, also failed to replicate his own findings. He traced the difference to the absence of a window: the availability of light. Priestley turned his attention to the new variable. New experiments, now focused on well water alone, persuaded him that light was the key factor, not plants. Ironically, that itself was an error. Priestley had noticed the production of green scum in his vessels, but considered it a byproduct of the restored air. But here, Priestley had mistaken correlation for causation. Jan van Ingenhousz and others recognized that the green scum — algae — was a plant and that both plants and light were required for the restoration of air. The error was "eventually" resolved, but not through simple replications. Isolating the error required a series of controlled experiments that (as in the cold fusion case) clarified the role of each variable (see Section 7 below on isolating errors.) Priestley's original results and interpretation had been correct all along. The replication failures had been misleading.

In summary, peer review and replication, frequently regarded as the chief methods or mechanisms for correcting errors, fall far short of their normative promise. They do not systematically guarantee "self-correction." How, then, are errors in science corrected (as they frequently are)? We need a more expansive view, which can be developed in part by exploring historical errors more fully (Sections 5,6,7,8).

Consider, next, a more modest version of the self-correcting or error-correcting thesis, which asserts that *scientists will revise and update their claims in light of new evidence* (e.g., Brunauer & Holcombe, 2019; Darden, 1991; McIntyre, 2019). One may call this (by contrast) *the accommodation thesis*. Namely, science *accommodates* itself to changes in the available evidence. Notably, however, *this conception does not specify how evidence contrary to the original claims may arise*. Thus, to the degree that errors reflect conceptual or methodological blind spots, the accommodation version sidesteps the critical promise at the core of self-correction. Namely, how do scientists initially become aware of their blind spots and the possible flaws in their already-accepted justifications (see Section 5)? Resolving anomalies (or discordant results or alternative theories) — *once they emerge* — is not inconsequential (e.g. Darden, 1991), but surely falls short of the stronger causal claim inherent in the "self-correcting" label. Namely, accommodation occurs only when relevant new and apparently contrary evidence first presents itself.

Focusing on accommodation alone leaves unanswered the critical question about how any such contrary evidence may arise (Section 5). Ultimately, for science to be fully *self*-correcting, it must include practices that effectively *probe* for errors and reveal possible contradictory evidence (e.g., Mayo, 1996; see also Platt's [1964] "strong inference"; or Popper's [1963] "severe tests"). Merton (1973) expressed it as "institutional vigilance" — an "unending exchange of critical judgment." It would be expressed as "vigorous but cognitively disciplined mutual checking and rechecking." Namely, science (ideally) "affords both commitment and reward for finding where others have erred or have stepped before tracking down the implications of their results or have passed over in their work what there is to be seen by the fresh eye of others" (pp. 101, 339). Merton, for his part, seemed persuaded that scientists are imbued with the requisite motivation: "a latent questioning of certain bases of established routine, authority, vested procedures and the realm of the 'sacred' generally" (Merton, 1973, p. 264; see also Jamieson et al., 2019, on a "culture of critique.").

A major factor in conceptualizing error correction, then, is individual motivation. Hull (1988) echoed Merton's sentiments in describing the need for rewards and incentives for challenging existing claims, which he attributed to the drive for professional credit (see also Latour & Woolgar, 1979). However, whether any such rewards or incentives actually work is an empirical question and an object of current debate (e.g., Peterson & Panofsky, 2021). Clarifying how this component of the process is essential is, again, another reason why a general philosophy of error in science is so important.

Lastly, "self-correcting" is sometimes used to refer to the community's effective management of fraud and other questionable research practices *occurring among its own members* —more about maintaining professional standards of conduct (by scientists) than remedying scientific error (e.g., Markman, 2010). Here, appeals to "self-correction" were used in the 1980s largely to defend the autonomy of scientific institutions in cases of lapses of research *ethics*. For example, scientists declared (then and now) that they (alone) should be responsible for detecting and sanctioning any malpractice by their peers—and that they are successful in doing so. In these arguments, they are less interested in the epistemic logistics of correcting claims than in the power struggle over who is entitled to enforce accountability (the "self" being scientists themselves rather than federal officials or "outsiders"). That is, the

Mertonian norm of (epistemic) "self-correction" was used rhetorically for other (political) purposes. Here, we need to distinguish between self-*correcting* and self-*policing* dimensions, even if the processes overlap.

4. Conceptualizing error-correction and error management

Recent concerns under the banner of "the reproducibility crisis" have extended well beyond treating peer review or replication as stopgaps, to other forms of error-correction and error management. Rather than bemoan inadequacies, many practitioners are now posing probative questions about how and "where" errors occur, how they are remedied (when they are), and how scientific practices might thus be improved.

Many commentators have focused on the *sources* of error, ostensibly aiming to *reduce* the incidence of error. They extend earlier work on documenting and organizing the broad spectrum of error types (e.g., Allchin, 2001; Catalog of Bias Collaboration, 2024; Sackett, 1979). The strategy seems to be to *guard against* errors ever being published. That is, they are intended to function as "gate-keeping" processes, or *filters*. Such projects (e.g., Ioannidis, 2012) target, for example: publication bias; selective reporting of results; fraud; questionable research practices; and underpowered studies. (Note the emphasis is largely on adhering to or enforcing known methodological principles—e.g., Hon's focus on "mistakes".) Van Ravenzwaaij, et al., (2023, p. 12) organized their analysis of error correction in terms of the typical research cycle, from grant application to publication — proposing steps that would anticipate errors prior to their occurrence, toward combating "methodological myopia" and "cargo cult inference." They suggest a grant lottery (to avoid biases in funding); stronger statistical education (to reduce statistical mistakes); blind data analyses via outsourcing (to alleviate confirmation bias); and standardized reporting practices.

While developing new methods (and inculcating effective old methods) to address sources of error before publication should be encouraged, of course, such *preventative* measures are best viewed as a form of quality control, better characterized philosophically as *self-checking* than as self-correction. Again, for "correction" to occur, there must have been an actual error first: a mistaken justification widely accepted as warranted (only to be discounted *later*). Self-checking is forward looking; self-correction, retrospective and remedial.

Reducing or limiting error is fruitful. Yet the core issue of self-correction is catching the errors that arise inadvertently in an admittedly imperfect (sometimes even biased) process. A proper *corrective* mechanism will address claims *earlier* accepted as justified (e.g., caloric, affinities, black bile). The history of errors in science informs us that we cannot escape the many sources of error altogether. Heuristic gaps, limited sample size, simplified theoretical models, and biased perspectives are integral to how science functions. Prophylaxis cannot fully eliminate the need for error *correction*. That *remedial* process is my central focus here.

Thus, many commentators focus instead on features of the process of recovering from errors once problems arise. Currently, these chiefly revolve around critical discourse: (post-publication) peer review and replication (e.g., National Academies of Sciences, Engineering and Medicine, 2019). Table 4 presents a sampling of factors that have been identified variously as "impediments" (Ioannidis, 2012), reasons for replication failures (Guttinger & Love, 2019), or breakdowns in fragile assumptions about social context (Romero, 2016). Vazire and Holcombe (2022) have sought to formulate "observable self-correction indicators" (OSCI), falling in two major categories: (1) transparency (open materials, methods, data and code; preregistration; Mertonian universalism); and (2) critical appraisal (pre- and post-publication peer review; automated error detection; computational and empirical reproducibility checks; detection of errors and biases in the literature; highly limiting predictions fostered by strong theory; and diversity of critical perspectives). Van Ravenzwaaij, et al., (2023, p. 12) echoed those concerns in advocating (a) a system to publish as-you-go, and (b) expansion and formalization of post-publication peer review. The lists vary widely. Still, one finds overlap in what is deemed significant. While all these isolated factors contribute to our understanding, they need to be integrated in a comprehensive, historically informed framework. We need to conceptualize the *whole process* of error correction. How are overlooked or hidden sources of error discovered, identified, confirmed and then widely acknowledged?

Table 4
Various factors viewed as problems in error correction.

• lack of replication (including editorial and reviewer biases against publishing such efforts) • lack of public access to data and analyses (basic background for peer analysis)	Ioannidis (2012)
• lack of motivation to replicate (either in diagnostic or integrative mode)	Peterson and Panofsky (2021); (see also Hull, 1988)
• fraud • methodological failures (faulty materials, flawed technique, design flaws) • overgeneralization	Guttinger and Love (2019)
• not everything is published (thus available for review) • resources are limited (preventing unfettered replication) • research exhibits direction bias	Romero (2016, p. 60)

To review: we need to mindfully distinguish between self-correcting, self-checking, self-policing, error correction, and accommodation to new evidence. We need to acknowledge, too, the critical difference between philosophical norms (what science should be, in abstract idealizations) and what science is in actual practice (based on concrete historical evidence). Having clarified the conceptual context, we may proceed to the core historical analysis, framing the results in philosophical terms. We can characterize *when* and *how* errors in science have been corrected (having drawn on a large set of historical cases across multiple disciplines, illustrative cases cited individually below—see methodological reflections in Section 1.) These might well form the basis for developing a more *systematic* and *complete* general framework for error-correction — an alternative to the rhetoric of replication and peer review or to some implied scientific servomechanism.

5. Emergence of incongruence

How, then, are errors corrected—concretely, as demonstrated in historical cases, not as imagined abstractly or stipulated theoretically in philosophical ideals?

The pathway to discovering error appears to begin when a problem emerges within the growing body of knowledge, when different research trajectories yield apparently irreconciliable findings. We may call these *incongruences* (Allchin, 2026). An incongruence may be either:

(a) a discordance between two sets of empirical evidence or experimental results,
(b) a theoretical anomaly, or
(c) an interpretive ambiguity (often manifested as a theoretical disagreement).

Essentially, new evidence challenges the coherence of existing knowledge—exposing a flaw somewhere in the justificatory network. But where is usually not immediately obvious (a variant of the familiar Duhem-Quine dilemma).

The emergence of incongruences seems highly contingent historically. Discrepancies may occur at the lab bench, during ordinary day-to-day work. Or investigators may be alerted by colleagues in conversation or correspondence. They may be highlighted during peer review. They may be encountered by other investigators investigating the same phenomenon, with or without any formal intent to replicate. Encountering the relevance of other (presumably unrelated) phenomena may be pure happenstance (or "chance" or accident, in the vernacular). There seems to be no predictable pattern.

New evidence, likewise, may be introduced by a variety of routes: new instrumental technology, the independent development of collateral knowledge, or new perceptual perspectives from individuals with different backgrounds.

For example, Boyle's law was labeled a law in part because it was presumed to be universal. All the exceptions had yet to be encountered. Only with the development of new instrument technology—exceptionally tall columns of mercury and other equipment for producing high pressures— were Victor Regnault, Emile Amagat and Louis Cailletet able to learn that the relationship of Boyle's law broke down at high pressures (say, above 10 atm). Only after the development of cryogenics was Thomas Andrews able to observe the deviations from Boyle's law at extremely low temperatures. Other apparatus was needed to discern the effects of low volumes and of different gases (van der Waals constants). Here, it was the domain, or scope, of Boyle's law that needed correction. But the original errors of overgeneralization were only recognized when new technology, in the form of suitable instruments, emerged to investigate the relevant domain (Allchin, 2007).

Another factor is collateral knowledge. The periodic table was formed and persisted without Group 18 (the noble gases) for three decades. A problematic incongruence only emerged with the discovery of argon and, subsequently, neon, helium, and xenon. That is, the periodic table was originally conceived without full knowledge of all the elements. That blind spot was not obvious at first, and not corrected by any systematic study. Rather, the existence of argon emerged by "accident," based on Lord Rayleigh's relatively pedestrian — and totally unrelated — study of the atomic weight of nitrogen (Giunta, 2001).

Sometimes, all that is needed is a new perspective to transform the obvious into the questionable. For example, the noble gases were first considered inert (nonreactive). Neil Bartlett approached the "fact" with a fresh eye after a half-century and demonstrated in 1962 that they can indeed form compounds: an anomaly to the notion that their filled orbitals fully inhibited reactivity. Indeed, some such reactions proceed rapidly at room temperature, without any need to "force" them through extreme conditions (Gay, 1977).

Any of these factors — new technology, new collateral knowledge, or new conceptual perspectives — may, on different occasions, be relevant to the emergence of an incongruence. Understanding the contingent nature of these events is critical, as it implies that expectation of *systematic* error-correction in science (the persistent and ever-hopeful vision of timely "self-correction") is probably misplaced.

6. Noticing

Errors in science do not announce themselves. That is, in large part, why they are errors. The lapses in the justification have escaped the scientists' notice. The mere presence of an incongruence is no guarantee that any researcher will *notice* it or regard it as worth pursuing.

Most importantly, pre-existing conceptual perspectives and familiar ("proven") methodologies form powerful cognitive blinkers. Confirmation bias is ubiquitous (Nickerson, 1998; Sunderland, 1992). For example, Lightman and Gingerich (1992) contended, based on historical analysis, that anomalies seem to appear only when there is an *alternative* perspective under which some observations can be seen to be anomalous for some *other* theory (not merely as unsolved puzzles) (Kuhn, 1970). That may be an overstatement. Still, the interpretation of anomalies in many cases seems to be theory-dependent (Allchin, 2015a), leading to divergent interpretations of their meaning or import.

Confirmation bias may thus sustain blind spots and help eclipse the perception of errors. Consider, for example, the case of Christian Eijkman in his investigations of the cause of beriberi in Java in the 1890s. Germ theory was still very new and revolutionizing thinking about disease. Beriberi was frequent in institutions, and seemed likely to be contagious. That was Eijkman's working assumption. Yet he and his colleagues were unable to isolate any disease-causing microbe. That did not dissuade him. Later, through a series of happenstances, Eijkman traced the disease in chickens to a polished rice diet, and its cure to consuming the rice polishings. This might have signaled Eijkman to the role of a nutrient deficiency. That was surely the interpretation of his successor in Java, Gerrit Grijns. But Eijkman continued to interpret the findings in terms of germ theory, microbial toxins and anti-toxins. The error completely escaped Eijkman's notice (Carpenter, 2000). Some errors seem to hide in plain view, undetected. Error-correction is contingent on perceiving the incongruence appropriately.

Nevertheless, researchers can and do encounter incongruences. Yes, researchers eager to capitalize on new findings may endeavor to replicate some findings, as a basis for further research, only to be stymied by different results (e.g., Peterson & Panofsky, 2021). They may have tried to access a natural phenomenon by means other than the one originally presented. They may have been pursuing independent research questions that yielded apparently conflicting results. *Or* promising initial results may have been taken to the next level of rigor, only to yield problematic results (e.g., the oops-Leon particle; Lederman, 1997). *Or* one's findings may simply not fit comfortably with clear theoretical expectations (themselves strongly supported by other evidence). Other researchers may have a different background, which yields an alternative interpretation of the same evidence (hence, a vital role for mavericks and iconoclasts; e.g., Harman & Dietrich, 2008, 2013). However, all these encounters seem happenstance. There is no systematic mechanism or uniformly reliable method at the social level of science, by which every publication is rigorously or exhaustively examined for any potential incongruence. Nor is it clear how one would incentivize such diffuse and largely speculative research. Not to mention finding the funding and extra labor. Resources limit investigative potential (e.g., Romero, 2016).

Correcting errors in science thus seems conditional upon encountering, *noticing* and *pursuing* incongruences. The role of replications (either diagnostic or integrative)—which are not and perhaps cannot be guaranteed—can easily be overstated. The context-dependent nature of perceptions and investigative choices certainly underscores further the contingencies of launching an error-correcting investigation—correction perhaps, but without a reliable or dependable *mechanism* for "self-correction."

7. Isolating errors and differentiating domains

An incongruence is not yet a clearly delineated error. It only signals indirectly that one exists—somewhere. Thus, a major factor in correcting errors in science is *isolating* the source of error to a particular flawed element of the once-secure justification. Resolving that uncertainty involves further research, not mere reference to an itemized catalog of known mistakes. Error correction thus requires epistemic *work*. In some cases, anomalies may lead to *de novo* theory-change (e.g., Darden, 1991). In other cases, the process may involve articulating the proper scope (or domain) of the concept(s) in question (e.g., Boyle's law). In some cases, one differentiates the domains of once-overlapping concepts (e.g., Allchin, 1997). Investigators must be prepared to trace the incongruence to any error type—whether observational, conceptual, or social (discoursive) (Allchin, 2001).

Thus, when researchers do encounter an incongruence, their focus often shifts to resolving the newly identified uncertainty. The original research question becomes *displaced* by a focus on the observational and/or theoretical mismatch instead (see, e.g., Rheinberger, 1997). The new task aims to *isolate* the flawed element in the justification for one of the extant claims. Investigators may deploy many of the same strategies used for decomposing and localizing functions in complex systems (e.g., Bechtel & Richardson, 1993; Craver & Darden, 2013; Darden, 1991; Wimsatt, 2007). The same applies to articulating the scope or domain of a concept, resolving it into finer parts to isolate the incongruence/source of error in some dimension of context (an "external" variable).

However, reorienting research still does not guarantee an immediate

or effective solution, anymore than in any other research enterprise in science. Many limiting factors may apply. First, confirmation bias—or worse, *communal* confirmation bias—may misdirect critical thinking precisely where it is needed most. Thus, the error may be inappropriately localized to a factor other than the ultimate source of error. One error is merely substituted with another. This may happen several times in succession, leading to a telltale *error cascade* (Allchin, 2015b).

For example, anthropologists and craniologists in the 19th century sought to rank intelligence of different groups based on measuring skulls: an apparently plausible inference from structure to function. The most obvious measurement, the sheer size of the cranium, however, yielded an intuitive incongruence, aptly referred to as "the elephant problem." Namely, the largest skulls belonged to elephants and whales. But no human seemed willing to concede that these animals were more intelligent than themselves. So, heeding the "evidence," that measure was scrapped. A sequel measure focused on the ratio of skull size to body size. However, this implied that many women would be smarter than men—a conclusion also rejected by the male researchers. So that was set aside. Georges Cuvier proposed the ratio of cranial bones to facial bones, presumably reflecting a proportionately larger brain. Yet on this measure, birds, anteaters and bear-rats all ranked above humans. Another error jettisoned. And so on for cranial height, facial angle, prognathism, orthognathism, and others (Fee, 1979). The history exhibited a conspicuous error cascade: a succession of rejected claims. The foundational error emerged at the end of the century: the researchers had been trying to naturalize social ranking by appeal to biological criteria. The work was mired in sloppy statistics (for example, using group averages to justify comparisons of individuals from each group) and the whole enterprise was shaped by gender and racial bias. Yes, the error was corrected — "eventually." But only in a roundabout way, after a decades-long error cascade of successive failures, each "correction" reflecting the same core source of error: cultural bias. Error cascades are possible because science does not have a systematic or algorithmic way to isolate errors or, more generally, to isolate causes to a particular component or activity of a system (Bechtel & Richardson, 1993).

Error correction can also be stymied by a lack of relevant collateral (or background) knowledge. For example, Newton's formula for the speed of sound (presented in the *Principia*) generated a theoretical value that deviated from measured values by 15%. Newton tried to finesse the anomaly, referring to the "crassitude of the particles of the air" and "vapours floating in the air" (*Principia*, Book II, Proposition 49). Still, no solution. By the turn of the 19th century it had become, according to Kuhn (Kuhn, 1958), "one of the scandals of physical science for more than a century and had repeatedly though fruitlessly drawn the attention of Europe's outstanding theoretical scientists, including Euler and Lagrange" (p. 137). The ultimate solution appeared only after the discovery of adiabatic phenomena, and Laplace's realization that they were relevant to the short-term compression and expansion of sound waves. To isolate Newton's error, one needed to be aware of this additional behavior of gases. Without that knowledge, the anomaly lingered unresolved—in this case, for 115 years (Finn, 1964).

Error correction, like scientific discovery, may also depend on technology. Only with the emergence of the Voltaic pile and batteries, and the subsequent discovery of electrolysis, would chemists begin to conceive of atoms in terms of electrical charge. And only then might the electrochemical dualism of Jacob Berzelius provide a viable alternative to Newton's erroneous concept of chemical affinities, which had dominated chemistry for a century (Kim, 2003).

Science also differs markedly from other canonical self-correcting systems in that the initial incongruence does not unambigously signal the requisite correction. Any proposed solution is *uncertain*. It must typically be *tested*. A hypothetical solution, like any concept in science, needs further evidence as confirmation. One equally needs to exclude the possible role of confounding variables or discount alternative interpretations (Priestley's case). Thus, it was not enough for John Snow to (correctly) propose water as the cause of the cholera outbreaks in London in the 1850s. While he needed to collect evidence of the consistent correlation of contaminated water and disease, he needed to show *simultaneously* that the same evidence was incompatible with a miasma explanation (Johnson, 2006; Snow, 1855).

In all these ways—the potential role of confirmation bias and error cascades, the limits of technology, the limits of collateral knowledge, and the custom of testing proposed corrections — error correction in science is much more complex than either standard self-correcting systems or trying to leverage peer review or replication alone. When isolating errors from incongruences, science faces uncertainties. Even *accommodating* surprising new evidence cannot be guaranteed to be timely or methodologically assured (by some systematic mechanism).

8. Persuading peers

Finally, integral to the process of error correction and conceptual change is shifting the consensus of the relevant experts at the social level. Error correction, like other forms of scientific discovery, involves critical discourse and negotiating one's way through contrasting interpretations of the evidence until residual uncertainties are largely resolved.

Thus, even when an incongruence is noticed (heralding a hidden error) and even when the error is isolated and tested, the process of error correction in science is still not complete. Scientists must also persuade their peers. Scientific knowledge is characterized by a critical consensus across a diversity of relevant expert perspectives, and here that includes affirming the "negative" knowledge of errors.

One might well imagine that, once an error was found and corrected, other scientists would avidly welcome the news and the remedy. Not always so. One can cite numerous cases where an error correction was, ironically, rejected by the community, sometimes for decades (Table 2).

For example, in the 1930s gastric ulcers became associated with diet and stress as sources of excess acidity. Excess acidity was indeed a symptom, but not necessarily the primary cause (once again, conflating correlation and causation). When antibiotics became available after World War II, they proved effective—not merely in treating, but also apparently in curing ulcers. Evidence on their effectiveness was published in 1951, 1953, 1955, 1958, and 1959 —all disregarded. In the 1960s, a Greek physician treated thousands of ulcer patients successfully with antibiotics. *Communal confirmation bias*, it seems, had set in (Kidd & Modlin, 1998; Marshall, 2001; Rigas, et al., 1999; Thagard, 1999). Even in the early 1980s, before Barry Marshall and Robin Warren published their now landmark clinical study in *The Lancet*, there were plentiful comments disparaging their views about bacterial infection as a cause. For decades, the medical community did not respond to the available evidence for solving an error.

One could equally mention J. Harlan Bretz and his interpretation of the remarkable geological formations of the Washington Scablands region. His evidence for a large-scale catastrophic geological event was summarily dismissed by a panel of esteemed geologists, none of whom had visited the region or reviewed Bretz's evidence. Their comments simply propounded the impossibility, blindly parroting the widely adopted doctrine of (unqualified) uniformitarianism. They were wrong. And it took four decades for other geologists to visit the Scablands, witness the evidence for themselves, and acknowledge that Bretz has been right all along (Soennichsen, 2008).

The process of community review and acceptance of new ideas relies significantly on effective communication and critical discourse. The identification of an error and the evidence behind it must be in the public sphere and open to critique. That is, skepticism—manifested *socially* as contrasting perspectives distributed across a scientific community—must be *organized*, as noted by Robert Merton (1973). Merton postulated the requisite social structure in his now well known scientific norms. They are basically rules for effective critical discourse, and remain as valid today (as norms) as when he formulated them. What remains open is the degree to which contemporary scientific practices

fulfill or approximate these norms.

First, the norm of *universalism* dictates that all voices should be afforded, by default, equal opportunity. Namely, there should be no *credibility bias*. Of course, credibility is an important heuristic in daily practice, allowing scientists to exercise calibrated trust in lieu of checking every individual claim themselves (Allchin, 1999). That system of trust and judgments of credibility (or reputation), however, can (like any heuristic) sometimes go awry. For example, Amodeo Avogadro addressed the problem of discordant atomic weights for many gaseous elements, an incongruence that plagued chemists in the early 1800s. Avogadro claimed that equal volumes of gases contained equal numbers of particles (whether atomic or molecular). That implied, however, that some gases were diatomic. That ran afoul of Jacob Berzelius's view of electrochemical attraction in molecular composition and John Dalton's view that atoms could not combine with themselves. The towering stature of both chemists eclipsed Avogadro, whose idea was given little credence, even as a proposal for discussion. Avogadro's notion would not be "rescued" until decades later, when his critics had passed away, and someone else took up the banner of diatomic elements (Levere, 2001, pp. 110-116; Rocke, 1984, pp. 101-107, 294-299).

Credibility bias—or judgments of trust—apply equally to cases of fraud. Some scientists are simply afforded too much trust, and outright lies and fabricated data go unrecognized, not probed by secondary research (initially). Some observers tend to place all the blame on the person committing the fraud. But that was not the conclusion of a committee that reviewed the failed claims of element 118 in 1999. The error was traced to Victor Ninov, who had been hired as the expert computer analyst and had manipulated a data file. He was dismissed, of course. But that was largely an *institutional* judgment, based on professional malpractice (Anderson, 2022). *Epistemically*, the committee had equally stern words for everyone else. They should have caught the error earlier by having a second person check the data analysis. Namely, they ostensibly attributed the error, in part, to the team's misplaced trust in Ninov, not to Ninov's own falsified data (Chapman, 2019). So too, perhaps, for other cases of fraud, from Piltdown Man to Henrik Schön to Ranga Dias? One may certainly try to reduce the incidence of fraud through moral training and (dis)incentives. However, from an error-correction perspective, one needs to consider the heuristic function of credibility—and how to manage occasions when it fails.

Another element of Mertonian discourse is *communalism*: a spirit of the collective enterprise and the sharing of results, methods, samples, computer code, and so on. This has been the focus of the Open Science Movement and is widely discussed now, sometimes under the label of "transparency," so little more needs to be added here. One should merely note that one should view these conditions as an additional qualification to an effective error-correction system.

The third Mertonian norm relevant to error is *disinterestedness*. Errors and misleading claims arising from industry-sponsored research plague modern science (e.g., McGarity & Wagner, 2008; Michaels, 2020; Sismondo, 2018). We may not be able to eliminate conflict of interest, but the social system can be organized to detect its adverse influence and neutralize it. For example, one study on the healthiness of a meat diet received swift criticism from medical researchers who, after describing its errors, also noted that key authors were affiliated with the meat industry and their institution had recently received a large endowment for a "Beef Cattle Academy" (Flaherty, 2020; The Nutrition Source, 2019). Complementary "interests" were essential, here, as a counterbalance, alerting others to the error. Social epistemologists have already noted the importance of diverse perspectives in a healthy epistemic community (e.g., Harding, 1990; Solomon, 2001). Here, one may merely note that this is a basic part of a Mertonian system for regulating the errors that arise from the biases introduced by particular interests.

Error correction requires community consensus, just as much as other scientific conclusions do. Numerous elements of social organization that shape the critical discourse of science thus form further contingencies for the dynamics of correction.

9. A general framework

The various elements described in the last four sections may now be stitched together into a general (even if perhaps unfinished) picture of error correction in science (Fig. 1). First, new evidence from ongoing research generates an *incongruence*: either (a) a discordance between two sets of empirical evidence or experimental findings, (b) a theoretical anomaly, or (c) an interpretive ambiguity (often manifested as a theoretical disagreement). The intersection of inconsistent conclusions challenges the coherence of the current knowledge and indicates a yet-to-be-identified error in the network of justifications (Section 5). To lead further to error correction, the problematic finding(s) must then be *noticed* (Section 6). With resources, time, and motivation available, further work at accommodation likely ensues. The nature of the incongruence must then be traced (or *isolated*) to a particular *source of error* or error type in the original justification (Section 7). Typically, additional tests help confirm or disconfirm the attribution of error. This will correspond to a new interpretation of the evidence, with an alternate justificatory framework. Finally, the justification for the new (replacement) claim, as for any scientific claim, must pass muster across a critical scientific community (Section 8). Each of these steps requires work — epistemic work — and each is critically contingent upon certain circumstantial contextual factors. This would be a prospective basic model of error correction.

When correction is viewed as portrayed in Fig. 1, one may see similarities to the general dynamics of an adaptive system, such as the immune response or natural selection (e.g., Lewontin, 1970; Plotkin, 1997), sometimes viewed as "corrective" in nature. The comparison helps underscore how a system of correction may be nonteleological and indeterminate. That is, unlike a thermostat and other similar systems (Section 3), science does not return to its original state, even when a coherence between concept and evidence has been reestablished. The process is permeated with unpredictable contingencies. The comparison here helps underscore how a system of adaptive accommodation (Section 3) may be effective, yet also *nonteleological* and indeterminate. Adaptive accommodation (responsive correction) and systematic, timely and mechanistic "self-correction" are fundamentally different.

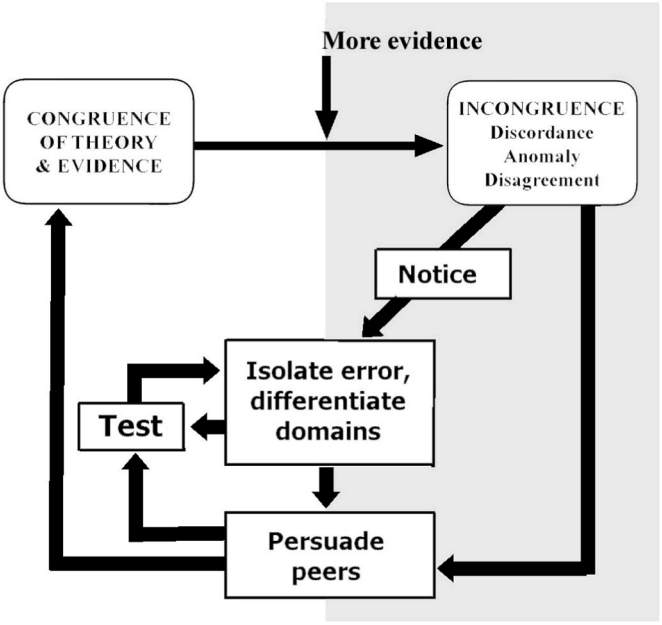


Fig. 1. A simplified framework for interpreting the major components of error-correction in science.

10. When does science correct its errors?

Errors in science are frequently corrected. But a wide-ranging historical survey of errors in science indicates that errors arise from multiple causes, and are resolved in multiple ways, not systematically (Sections 5,6,7,8). They are not consistently corrected through replication or peer review, as much rhetoric contends (Section 3). When viewed as an ensemble, the historical cases indicate just how scientific work sometimes corrects errors, while on other occasions that work falters (Fig. 1). Many conditions, contingencies and qualifications are involved. In practice, science meets these criteria unevenly (Table 1, Table 2, Table 3). Namely, science is not "self-correcting" in any guaranteed sense, as typically claimed. The telling question, rather, is: *When* does science correct its errors?

We may now resolve the title question. The analysis here supports the conclusion that correction in science is fundamentally contingent on:

(1) active (ongoing) research
    Active research, in turn, depends (materially) on:
    (a) funding and resources, and
    (b) personal motivations and orientations.
    Moreover, the research must be:
    (c) probative, not merely diagnostic.
(2) Mertonian discourse
    Mertonian discourse, as a form of socially organized skepticism, is characterized by:
    (a) universalism (equal opportunity access to publication; no credibility bias);
    (b) communalism ("open science") and
    (c) disinterestedness (controls for conflicts of interest).

In addition, the emergence or development of deeper levels of evidence that help identify and/or resolve errors may rely on:

(3) new observational technology,
(4) new collateral knowledge, and
(5) new diversity of perspectives.

These last three factors are beyond any systematic control, and thereby seem to establish an ineliminable limit to any guarantee of timely error-correction in science.

On the other hand, all these conditions hardly discount the importance of error correction when it does occur. Trust in a particular scientific claim may justifiably rest on the cogency and completeness of the available evidence, *keeping these various contingencies in mind*. Nevertheless, the process of error-correction seems to fall short of a systematic *mechanism*. Still, the basic model (Fig. 1) and the various conditions (1–5), might guide science journalists, public policy-makers, citizens and consumers, who may all want to assess the degree to which particular scientific claims may be considered reliably "settled" by science.

Reflecting on the correction process and the parameters for its effectiveness, one may well wonder whether discovering errors is really any different from other forms of discovery in science. Isolating and confirming errors, ironically, also involves justificatory processes. Error correction thereby yields "negative" knowledge. In both cases, well formed evidence is central. Both exhibit periodic progress, but without any systematic guarantee of timely success.

Conflicts of interests

The Author declares no conflicts of interest.

Appendix : Methodology

As noted in the Introduction, a historical study of this scope is fraught with potential bias or error. For example, how would one characterize a null hypothesis that could be rigorously "tested"? Here, I address several fundamental concerns for the methodologically-minded reader.

I should reiterate that this study focuses specifically on the title question about the nature of "self-correction." It distills several significant stages in the process, allowing us to characterize a generic "natural history" of an error (and introduce examples that can help vividly illustrate the role of each stage). My analysis indicates that an account of systematic error correction (or a mechanism that reliably finds and resolves errors) is unlikely. Thus, any philosophical vision that we should expect a universal process (of necessary and sufficient conditions) — and that we should thus search for or expect evidence of such — seems misplaced.

First, *what counts an error?* As defined in Section 2, an error is "a scientific claim whose justification is later found to be unwarranted" — namely, an instance that does indeed require correction (not merely caution). The identification is necessarily retrospective. Claims that never seemed justified at the outset are not errors. They are acknowledged uncertainties. Uncertainties are not errors. Nor are failures errors: they are simply expectations (not formally justified) that were not met. This clarification constrains the relevance of any case.

What counts as a meaningful error? Because the focus is on justification, and the factors involved in justification are multitudinous, errors may range widely, from the apparently trivial — contaminants, missed aliquots, mismatched fossil joints — to major conceptual upheavals — heat as a substance, electricity as a fluid, continents as globally fixed. They range from the observational, through the conceptual, to social-level. For fuller discussion of the spectrum of error types, see Allchin (2001, 2026). That is, the sources of error that yield false claims are many and varied. Correction involves a change in the justification, and thus the scale of the error is secondary. (Further research may or may not reveal a correlation between the scale of an error and the duration until it is found and resolved.)

What counts as a useful historical case of error? As noted in Section 1, analysis depends on: (1) identifying the source of error, and (2) tracing the historical trajectory from discovery of the error to its resolution through additional evidence. Documentation may be provided by original publications, other historical documents, oral histories, testimony, news reports, and so on (the standard sources that inform any historian or journalist). There are surely many cases of error for which we do not have adequate documentation for proper analysis.

What counts as an episode of genuine self-correction? Public rhetoric typically portrays self-correction as what guarantees the trustworthiness of current scientific claims. This notion reflects an assumption that self-correction is systematic and timely (Section 1). But not all cases exhibit these features (e.g., Tables 1–3): we must differentiate. Nor all corrections are "self"-corrections. For example, cases of accommodation to new evidence exhibit correction, but because they do not involve *finding* the error, they do not constitute self-correction. Likewise, episodes of prophylactic self-checking are not self-correction (which involve having first accepted an errant justification). Clarifying the concept of self-correction (Sections 3-4) limits the relevant historical cases by controlling for confusions about the term "self-correction."

How does one control for bias in a historical sample? Because there is no comprehensive list of errors in the history of science, it is impossible to stipulate what would constitute a fair (representative) or random sample. As noted in Section 1, the sample here is "haphazard." Over roughly two decades, I collected sources by (a) delving into familiar examples from historical texts and journal articles; (b) asking scientists and historians that I encountered during that period to describe their favorite example (including their own errors); (c) habitually perusing *Science* magazine, attentive to news stories and retractions; (d) noticing cases mentioned in the conventional news media; (e) following through on

casual references in books or papers on science. Because I have taught history of science (to non-majors) and "nature of science" (to science teachers), I have enjoyed wide exposure to a variety of historical cases. Interested in comprehensive coverage, I have collected cases from all disciplines, so I hope to have minimized the risk of disciplinary bias. Because the aim has been to understand error in all its manifestations, I hope to have minimized theoretical bias (e.g., a prior theory of error that would likely guide selective collecting). By including all levels of concern (as noted above), I hope to have escaped philosophical bias (limited to logical fallacies, say, or fraud, or just experimental methods). The catalog of cases in my "sample" is the product of happenstance. That is, I cannot account for cases I did not include. If there is a noticeable bias (for example, in Tables 1–3), I invite others to identify the bias, so that (in the spirit of "self-correction") it can be addressed.

The cases used for illustration were selected for their value in highlighting the relevant feature. They are not the only relevant cases, nor even the most typical (or "representative" of other similar cases). That is, they function as illustrations, not as models or generalized exemplars. The depth of available cases allowed for such discretionary selectivity. So: these examples are, in a sense, "cherry-picked." The reader will have to decide if they are sufficient or definitive historical evidence to justify the minimalist existence-claims for the correcting factor associated with each.

What is an appropriate or informative sample size? Because there is no reference population, it is quite impossible to generate a statistical model or to calculate statistical power as a measure of adequate sample size. An alternative method (used in exploratory studies of species richness, for example) is to continue sampling until no new variables appear, indicating that the cumulative list of relevant factors seems saturated. This study draws on nearly 200 cases as a resource. Further, the recurrence of certain elements determining when (or how) scientists corrected earlier errors begins to indicate patterns, not merely scattered instances. These are the factors described in Sections 5-8. (Still, exceptional cases may indicate dimensions that are relevant in practice historically, even if only rarely.) I have limited the presentation here to a few illustrative examples, rather than give an exhaustive account of all the cases. A quantitative analysis would be premature, in my view. If there are important omissions (based on incomplete sampling), I welcome further research that will help reveal the deficits and extend or qualify the observations noted here.

How can one establish meaningful insights from an incomplete sample? How, indeed. Again, the analysis here is not a universal account. It identifies several contingencies (Sections 9-10) that help characterize when (and how) errors in science are corrected in practice — and when (concretely) that process seems to fail. Like all claims, the conclusions are subject to accord with further evidence.

Data availability

No data was used for the research described in the article.

References

Alberts, B., Cicerone, R. J., Fienberg, S. E., Kamb, A., McNutt, M., Nerem, R. M., Schekman, R., Shiffrin, R., Stodden, V., Suresh, S., Zuber, M. T., Pope, B. K., & Jamieson, K. H. (2015). Self-correction in science at work. *Science, 348*, 1420–1422.

Allchin, D. (1997). A 20th-century phlogiston: Constructing error and differentiating domains. *Perspectives in Science, 5*, 81–127.

Allchin, D. (1999). Do we see through a social microscope?: Credibility as a vicarious selector. *Philosophy of Science, 60*(Proceedings), S287–S298.

Allchin, D. (2001). Error types. *Perspectives in Science, 9*, 38–59.

Allchin, D. (2007). Teaching science lawlessly. In P. Heering, & D. Osewold (Eds.), *Constructing scientific understanding through contextual teaching* (pp. 13–31). Berlin: Frank & Timme.

Allchin, D. (2012). A comedy of scientific errors. *The American Biology Teacher, 74*, 530–531.

Allchin, D. (2013). *Sacred bovines*. Oxford University Press.

Allchin, D. (2015a). Context-dependent anomalies and strategies for resolving disagreement: A case in empirical philosophy of science. In S. Wagenknecht, N. J. Nersessian, & H. Andersen (Eds.), *Empirical philosophy of science: Introducing qualitative methods into philosophy of science* (pp. 161–171). Dordrecht: Springer.

Allchin, D. (2015b). Correcting the "self-correcting" mythos of science. *Filosofia e História da Biologia, 10*, 19–35.

Allchin, D. (2022). The very reproducible (but illusory) mesosome. *The American Biology Teacher, 84*, 321–323.

Allchin, D. (2026). *Toward a philosophy of error in science*. New York: Oxford University Press.

Anderson, H. (2022). Philosophy of scientific malpractice. *SATS, 22*, 135–148.

Armstrong, J. S. (1997). Peer review for journals: Evidence on quality control, fairness, and innovation. *Science and Engineering Ethics, 3*(1), 63–84.

Baldwin, M. (2019). Peer review. *Encyclopedia of the History of Science*. https://doi.org/10.34758/8f53-jd27

Barnes, C. A. (1989). *Interview by Heidi Aspaturian, Pasadena, California, June 13 and 26, 1989. Oral History Project*. California Institute of Technology Archives. https://oralhistories.library.caltech.edu/212/.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton University Press.

Berg, J. (2019). Replication challenges. *Science, 365*, 957.

Bowers, J. L. (1979). Reception of acupuncture by the scientific community: From scorn to a degree of interest. In S. H. Mauskopf (Ed.), *The reception of unconventional science* (pp. 91–104). Boulder: Westview Press.

Broad, W., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science*. New York: Simon and Schuster.

Bruner, J. P., & Holman, B. (2019). Self-correction in science: Meta-analysis, bias and social structure. *Studies In History and Philosophy of Science Part A, 78*, 93–99.

Buchwald, J. Z., & Franklin, A. (2005). *Wrong for the right reasons*. Dordrecht: Springer.

Burchfield, J. D. (1990). *Lord Kelvin and the Age of the Earth*. Chicago: University of Chicago Press.

Cantor, G. (1975). The reception of the wave theory of light in Britain: A case study illustrating the role of methodology in scientific debate. *Historical Studies in the Physical Sciences, 6*, 109–132.

Carpenter, K. J. (2000). *Beriberi, White rice and vitamin B: A disease, a cause and a cure*. Berkeley, CA: University of California Press.

Carter, K. C. (1983). Translator's introduction. In *The etiology, concept, and prophylaxis of childbed fever* (pp. 3–59). Madison: University of Wisconsin Press.

Catalog of Bias Collaboration. (2024). Catalog of bias. https://catalogofbias.org/. (Accessed 13 December 2024).

Chang, H. (2009). We have never been Whiggish (about phlogiston). *Centaurus, 51*, 239–264.

Chapman, K. (2019). The element that never was. *Chemistry World*. https://www.chemistryworld.com/features/victor-ninov-and-the-element-that-never-was/3010596.article.

Cicchetti, D. A. (1991). The reliability of peer review for manuscript and grant submissions: A cross-disciplinary investigation. *Behavior and Brain Sciences, 14*, 119–186.

Craver, C., & Dan-Cohen, T. (2024). Experimental artifacts. *British Journal for the Philosophy of Science, 35*, 381–390.

Craver, C., & Darden, L. (2013). In *In search of mechanisms: Discovering across the life sciences*. University of Chicago Press.

Crosland, M. (1973). Lavoisier's theory of acidity. *Isis, 64*, 306–325.

Darden, L. (1991). *Theory change in science: Strategies from Mendelian genetics*. Cambridge University Press.

Elkana, Y., & Goodfield, J. (1968). Harvey and the problem of the "capillaries.". *Isis, 59*, 61–73.

Engber, D. (2017). Is science broken? Or is it self-correcting? *Slate*. https://slate.com/technology/2017/08/science-is-not-self-correcting-science-is-broken.html.

Epstein, M. (1999). *The Allison magneto-optic method of chemical analysis*. http://mikeepstein.net/path/allison.html.

Eyler, J. M. (2001). The changing assessments of John Snow's and William Farr's cholera studies. *Sozial- und Präventivmedizin, 46*, 225–232.

Fee, E. (1979). Nineteenth-century craniology: The study of the female skull. *Bulletin of the History of Medicine, 53*, 415–433.

Feest, U. (2019). Why replication is overrated. *Philosophy of Science, 86*, 895–905.

Finn, B. S. (1964). Laplace and the speed of sound. *Isis, 55*, 7–19.

Flaherty, C. (2020). *Beef with Harvard*. Inside Higher Ed. https://www.insidehighered.com/news/2020/01/23/texas-ams-beef-harvard. (Accessed 14 December 2024).

Forster, P. (2002). Scientific inquiry as a self-correcting process. In M. Bergman, & J. Queiroz (Eds.), *The commens encyclopedia: The digital encyclopedia of peirce studies*. New Edition. Pub. 140509-1757a. Retrieved from http://www.commens.org/encyclopedia/article/forster-paul-scientific-inquiry-self-correcting-process-0.

Fox, R. (1971). *The caloric theory of gases: From lavoisier to regnault*. Oxford, UK: Clarendon.

Franks, F. (1981). *Polywater*. Cambridge, MA: MIT Press.

Gartmeier, M., Bauer, J., Gruber, H., & Heid, H. (2008). Negative knowledge: Understanding professional learning and expertise. *Vocations and Learning, 1*, 87–103. https://doi.org/10.1007/s12186-008-9006-1

Gay, H. (1977). Noble gas compounds: A case study of scientific conservatism and opportunism. *Studies in History and Philosophy of Science, 8*, 61–70.

Giunta, C. J. (2001). Argon and the periodic system: The piece that would not fit. *Foundations of Chemistry, 3*, 105–128.

Glen, W. (Ed.). (1994). *The mass-extinction debates: How science works in a crisis*. Stanford, CA: Stanford University Press.

Glen, W. C. (2002). A triptych to Serendip: Prematurity and resistance to discovery in the earth sciences. In E. B. Hook (Ed.), *Prematurity in scientific discovery* (pp. 92–108). Berkeley: University of California Press.

Guttinger, S., & Love, A. (2019). Characterizing scientific failure. *EMBO Reports, 20*, Article e48765.
Hallam, A. (1989). *Great geological controversies* (2d ed.). Oxford: Oxford University Press.
Harding, S. (1990). *Whose science? Whose knowledge?* Cornell University Press.
Harman, O., & Dietrich, M. R. (2008). *Rebels, mavericks, and heretics in biology*. New Haven: Yale University Press.
Harman, O., & Dietrich, M. R. (2013). *Outsider scientists*. University of Chicago Press.
Heilbron, J. L. (1979). *Electricity in the 17th and 18th centuries*. Berkeley: University of California Press.
Holton, G., & Mack, A. (Eds.). (2005). Errors: Consequences of big mistakes in the natural and social sciences. *Social Research, 71*(1).
Hon, G. (1995). Going wrong: To make a mistake, to fall into an error. *The Review of Metaphysics, 49*, 3–20.
Hon, G., Schickore, J., & Steinle, F. (2009). Going amiss in experimental research. In *Boston studies in the philosophy of science, No. 267*. Dordrecht: Springer.
Hull, D. (1988). *Science as a process*. Chicago: University of Chicago Press.
Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), Article e124.
Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science, 7*, 645–654.
Jamieson, K. H., McNutt, M., Kiermerc, V., & Sever, R. (2019). Signaling the trustworthiness of science. *PNAS, 116*, 19231–19236.
Johnson, S. (2006). *The Ghost map*. New York: Riverhead Books.
Judson, H. F. (1980). *The search for solutions*. New York: Holt, Rinehart, Winston.
Scientific mistakes and mistaken science. In Karakas, Alexandra, & Tuboly, Adam Tamas (Eds.), *Studies in History and Philosophy of Science*, (2026). Cham: Springer Nature Switzerland, ISBN 9783032259240. https://doi.org/10.1007/978-3-032-25925-7. https://link.springer.com/10.1007/978-3-032-25925-7.
Kidd, M., & Modlin, I. M. (1998). A century of *Helicobacter pylori*: Paradigms lost – Paradigms regained. *Digestion, 59*, 1–15.
Kim, M. G. (2003). *Affinity, that elusive dream*. MIT Press.
Kottler, M. J. (1974). From 48 to 46: Cytological technique, preconceptions, and the counting of human chromosomes. *Bulletin of the History of Medicine, 48*, 465–502.
Kuhn, Thomas S. (1958). The Caloric theory of adiabatic compression. *Isis, 49*(2), 132–140. https://doi.org/10.1086/348664. https://www.journals.uchicago.edu/doi/10.1086/348664
Kuhn, T. S. (1970). *The structure of scientific revolutions* (2d ed.). University of Chicago Press.
Latour, B., & Woolgar, S. (1979). *Laboratory life: The construction of scientific facts*. Princeton: Princeton University Press.
Le Grand, H. E. (1972). Lavoisier's oxygen theory of acidity. *Annals of Science, 29*, 1–18.
Le Grand, H. E. (1974). Ideas on the composition of muriatic acid and their relevance to the oxygen theory of acidity. *Annals of Science, 31*, 213–225.
Le Grand, H. E. (1988). *Drifting continents and shifting theories*. Cambridge: Cambridge University Press.
Lederman, L. (1997). The discovery of the Upsilon, bottom quark, and B mesons. In L. Hoddeson, L. Brown, M. Riordan, & M. Dresden (Eds.), *The rise of the standard model* (pp. 101–113). Cambrdige University Press.
Lehman, C. (2010). Innovation in chemistry courses in France in the mid-eighteenth century: Experiments and affinities. *Ambix, 57*, 3–26.
Levere, T. H. (2001). *Transforming matter*. Baltimore: Johns Hopkins University Press.
Lewis, N. S., Barnes, C. A., Heben, M. J., et al. (1989). Searches for low-temperature nuclear fusion of deuterium in palladium. *Nature, 340*(6234), 525–530.
Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics, 1*, 1–18.
Liem, K. F. (1988). Form and function of lungs: The evolution of air breathing mechanisms. *American Zoologist, 28*, 739–759.
Lightman, A., & Gingerich, O. (1992). When do anomalies begin? *Science, 255*, 690–695.
Lindsay, R. A. (2011). *Science is self-correcting … sort of. Amherst, NY*. Center for Inquiry. https://centerforinquiry.org/blog/science_is_self-correcting_-_sort_of/. (Accessed 29 April 2026).
Liu, P. A. (2011). *Creating controversy: Science writers, corporate funders, and non-expert scientists in the debate over prions (1982-1997) [Ph.D. diss.]*. Toronto: University of Toronto.
Longino, H. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.
Losee, J. (2005). *Theories on the scrap heap: Scientists and philosophers on the falsification, rejection and replacement of theories*. Pittsburgh, PA: University of Pittsburgh Press.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.
Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1*, 161–175.
Markman, A. (2010). Why science is self-correcting. *Psychology Today*. https://www.psychologytoday.com/us/blog/ulterior-motives/201008/why-science-is-self-correcting.
Marshall, B. J. (2001). One hundred years of discovery and rediscovery of *Helicobacter pylori* and its association with peptic ulcer disease. In H. L. T. Mobley, G. L. Mendz, & S. L. Hazell (Eds.), *Helicobacter pylori: Physiology and Genetics*. ASM Press.
Martin, A. (2004). Can't any body count?: Counting as an epistemic theme in the history of human chromosomes. *Social Studies of Science, 34*, 923–948.
Deborah G., Mayo, Error and the growth of experimental knowledge, 1996, 10.7208/chicago/9780226511993.001.0001, 9780226511986, http://www.bibliovault.org/BV.landing.epl?ISBN=9780226511986, University of Chicago Press.
McGarity, T. O., & Wagner, W. E. (2008). *Bending science: How special interests corrupt public health research*. Cambridge, MA: Harvard University Press.
McIntyre, L. (2019). *The skeptical attitude*. Cambridge, MA: MIT Press.
California Map Society, McLaughlin, G., & Mayo, N. H. (1995). *The mapping of California as an island : an illustrated checklist* (1st ed.). California Map Society.
McNutt, M. (2020). Self-correcting by design. *Harvard Data Science Review*. #2.4.
Merton, R. K. (1973). *The sociology of science*. Chicago, IL: University of Chicago Press.
Michaels, D. (2020). *The triumph of doubt*. Oxford University Press.
NanoBubbles. (2021). NanoBubbles: How, when and why does science fail to correct itself? *Radboud Universiteit*. https://www.ru.nl/en/research/research-projects/nanobubbles-how-when-and-why-does-science-fail-to-correct-itself. (Accessed 10 May 2026).
National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25303
Nickerson, R. S. (1998). Confirmation bias; a ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–200.
Nye, M. J. (1980). N-rays: An episode in the history and psychology of science. *Historical Studies in the Physical Sciences, 11*, 125–156.
Oreskes, N. (1999). *The rejection of Continental drift*. New York: Oxford University Press.
Oreskes, N. (2019). *Why trust science?* Princeton: Princeton University Press.
Pashler, H., & Harris, C. R. (2012). Is the replicability crisis over-blown? Three arguments examined. *Perspectives on Psychological Science, 7*, 531–536.
Peterson, D., & Panofsky, A. (2021). Self-correction in science: The diagnostic and integrative motives for replication. *Social Studies of Science, 51*(4), 583–605.
Platt, J. R. (1964). Strong inference. *Science, 146*, 347–353.
Plotkin, H. (1997). *Darwin machines and the Nature of knowledge*. Harvard University Press.
Popper, K. R. (1963). *Conjectures and refutations. The growth of scientific knowledge*. New York: Routledge & Kegan Paul.
Rheinberger, H.-J. (1997). *Towards a history of epistemic things*. Stanford University Press.
Rigas, B., Feretis, C., & Papavassiliou, E. D. (1999). John Lykoudis: An unappreciated discoverer of the cause and treatment of peptic ulcer disease. *Lancet, 354*, 1634–1635.
Ritland, D., & Brower, L. P. (1991). The viceroy butterfly is not a Batesian mimic. *Nature, 350*, 497–498.
Rocke, A. (1984). *Chemical atomism in the nineteenth century: From Dalton to cannizzaro*. Columbus: Ohio State University Press.
Rohrer, J. M., Tierney, W., Uhlmann, E. L., DeBruine, L. M., Heyman, T., Jones, B., et al. (2021). Putting the self in self-correction: Findings from the loss-of-confidence Project. *Perspectives on Psychological Science, 16*, 1255–1269. https://doi.org/10.1177/1745691620964106
Romero, F. (2016). Can the behavioral sciences self-correct? A social epistemic study. *Studies In History and Philosophy of Science Part A, 60*, 55–69.
Rothwell, P. M., & Martyn, C. N. (2000). Reproducibility of peer review in clinical neuroscience: Is agreement between reviewers any better than would be expected by chance alone? *Brain, 123*, 1964–1969.
Sackett, D. L. (1979). Bias in analytical research. *Chronic Diseases, 32*, 51–63.
Schimmack, U. (2014). Replicability-index. http://replicationindex.com. (Accessed 10 May 2026).
Schimmack, U. (2019). Psychological science is self-correcting [blog]. https://replicationindex.com/2019/03/04/psychological-science-is-self-correcting. (Accessed 2 May 2025).
Science and Technology Committee. (2011). Peer review in scientific publications. In *Eighth report of session 2010–12, I*. London, UK: House of Commons.
Sismondo, S. (2018). *Ghost managed medicine: Big pharma's invisible hands*. Manchester, UK: Mattering Press.
Sismondo, S. (2025). Cultures of trial and error: Self, corrections and science in conflict-laden science. *Center of Trial and Error*. https://blog.trialanderror.org/self-correction-and-science-in-conflict-laden-science#references.
Slack, F. G. (1935). The magnetõoptic method of chemical analysis. *Journal of the Franklin Institute, 218*(4), 445–462.
Smith, D. L. (1989). Quest for fusion. *Engineering and Science [CalTech Office of Public Relations], 52*(4), 2–15.
Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*, 178–182.
Snow, J. (1855). *On the mode of communication of cholera* (2nd ed.). London: John Churchill. Facsimile edition available online at: https://archive.org/details/b28985266.
Soennichsen, J. (2008). *Bretz's flood: The remarkable story of a rebel geologist and the world's greatest flood*. Seattle: Sasquatch Books.
Solomon, M. (2001). *Social empiricism*. Cambridge, MA: MIT Press.
Stegenga, J. (2024). Justifying scientific progress. *Philosophy of Science, 91*, 543–560.
Stegenga, J. (2026). *Heart of science: A philosophy of scientific inquiry*. Chicago: University of Chicago Press.
Sunderland, S. (1992). *Irrationality*. New Brunswick, NJ: Rutgers University Press.

Tennant, J. P., & Ross-Hellauer, T. (2020). The limitations to our understanding of peer review. *Research Integrity and Peer Review, 5*(#6).
Thagard, P. (1999). *How scientists explain disease*. Princeton, NJ: Princeton University Press.
Tharp, H. (2023). It matters who does science. *Science, 380*, 873.
The Nutrition Source. (2019). New 'guidelines' say continue red meat consumption patterns, but recommendations contradict evidence. *Harvard School of Public Health*. https://www.hsph.harvard.edu/nutritionsource/2019/09/30/flawed-guidelines-red-processed-meat.
van Ravenzwaaij, D., Bakker, M., Heesen, R., Romero, F., van Dongen, N., et al. (2023). Perspectives on scientific error. *Royal Society Open Science, 10*, Article 230448.
Vazire, S., & Holcombe, A. O. (2022). Where are the self-correcting mechanisms of science? *The Journal of General Psychology, 26*, 212–223.
Vorland, C. J., Brown, A. W., Ejima, K., Mayo-Wilson, E., Valdez, D., & Allison, D. B. (2020). Toward fulfilling the aspirational goal of science as self-correcting: A call for editorial courage and diligence for error correction. *European Journal of Clinical Investigation, 50*(2), Article e13190. https://doi.org/10.1111/eci.13190
Weintraub, D. A. (2007). *Is Pluto A planet? A historical journey through the solar system*. Princeton: Princeton University Press.
Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University Press.
Ziman, J. (1968). *Public knowledge: The social dimension of knowledge*. Cambridge: Cambridge University Press.
Ziman, J. (2000). *Real science: What it Is, and what it means*. Cambridge University Press.
Zimring, J. (2019). *What science is and how it really works*. Cambridge University Press.