When is Science "Self-Correcting"?

Douglas Allchin

Abstract. Standard rhetoric portrays science as self-correcting. Here, I explore the nature of "self-correction" as a concept and consider what (if any) mechanism may achieve it. What other views inform the correction or management of error? Analysis of historical cases helps clarify concretely how and when corrections in science occur. We can thus articulate a set of conditions requisite for effective error correction, which indicate that "self-correction" is rarely if ever achieved in the standard use of the term (as a guarantor of reliability).

Keywords: error, self-correction, error correction, peer review, replication, error cascade

1. Introduction

The history of science is littered with abandoned concepts and falsified theories (e.g., Losee, 2015). We are only aware of such errors because other scientists have found and resolved them. Such cases have led some philosophers to speculate about the inevitability of error (the Pessimist's Induction; e.g., Laudan, 1977, 1981; Stanford, 2006). Recently, others have contended that "most published research findings are false" and questioned the foundations for reliability of science (the reproducibility crisis; e.g., Berg, 2019; Ioannidis, 2005, 2012; Pashler & Harris, 2012). At the same time, some interpret the historical remedy of errors as indicating the very opposite: a *self-correcting* process (e.g., Bruner & Holman, 2019; Darden, 1991; Forster, 2002; Mayo, 1996; National Academy of Sciences, 2019; Schimmack, 2019). One view contends that scientists systematically correct each other's blind spots (e.g., Hull, 1988; Longino, 1990; Merton, 1973; Oreskes, 2019; Ziman, 1968, 2000). For some, a process of self-correction helps demarcate science from other intellectual endeavors (e.g., McIntrye, 2019; Zimring, 2019). Hence, in rhetoric that is eagerly echoed by science advocates, science is inherently trustworthy *because* it is self-correcting (e.g., Alberts, et al., 2015; McNutt, 2020; Tharp, 2023). That is, historians, philosophers, sociologists and scientists disagree about the status of errors in science and their correction: the problem to be addressed here.

We need a more fully developed and historically well informed *philosophy of error in science*. What are the many sources of error, and what are the many ways of correcting them? Here, I address some fundamental questions about errors and their correction, stemming from the claims about self-correction. First, in whay ways are errors corrected in practice (as evidenced in concrete history)? Is there (in the spirit of Machamer, Darden & Craver, 2000) a general *mechanism* of systematic (self-)correction? If so, what particular components, processes, activities, inputs and outputs seem essential? Are there identifiable assumptions or boundary conditions for error-correction to be effective? Might error correction itself sometimes be vulnerable to error or susceptible to failure? How? In brief, *when* (historically) is science "self-correcting"?

My analysis here emerges from a larger project on understanding error in science (Allchin, forthcoming). That includes having examined over 200 cases of error in science (from the grand to the seemingly trivial) and documenting their various error types (from observational to conceptual to social-level) — and also clarifying how the errors were resolved in each case. This paper draws from that work.

First, I address several preliminaries about conceptualizing (or defining) error (Section 2) and self-correction and its presumed mechanism (Section 3). Then I consider various recent approaches to error and error correction (Section 4). I next describe a generalized "natural history" of an error based on cases from history, highlighting the various functional stages of the process. These include the emergence of an incongruence (Section 5), noticing it (Section 6), isolating the source of error (or error type) and possibly testing it (Section 7), and vetting the error correction at the community level (Section 8). I conclude with an abstract summary of the process (Section 9) and identify the various contingencies that seem to shape effective error-correction in science: 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. (Thus, while there are many methodologies that, as prophylactic measures, aim to *reduce* the incidence of error, they are not strictly relevant to the cases where errors actually 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 discussion below). The error lies in the mistaken *justification*. Ironically, then, resolved errors constitute a form of knowledge: *negative* knowledge (e.g., Allchin, forthcoming; Gartmeier, et al., 2014).

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 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 (see Allchin, 2001). Counterintuitively, perhaps, correcting error involves *epistemic work* and results in (justified) *negative knowledge.*

3. Conceptualizing Self-Correction and Its Mechanism

How are we to conceive science as *self*-correcting? This can be problematic, as well. For example, what (or who?) is the "self" responsible for the correction? 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*. The subtle shift in subject, however, has important rhetorical overtones for how (and when) science can be viewed as correcting "itself" and thereby earn its cultural authority. We need to disentangle the concepts of self-correction, self-checking, self-policing, and adaptive accommodation, at least.

Canonical views of self-correction may be found in Hesse (1973), Hull (1988) and Merton (1973), among others. For a recent representative articulation 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 a problematic qualification. Assertions about self-correction are typically couched in 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). Namely, the “self-correction” process in this view need not be *systematic*, such that an outcome is guaranteed (causally) or assured within a specified period of time. For this reason alone, perhaps, philosophers need to develop an explicit model or mechanism of self-correction, or abandon such vague notions.

< INSERT TABLE 1 ABOUT HERE >

Such claims, with their ill-defined promise of correction, certainly lose credibility in the light of history (Allchin, 2015b). Namely, some errors persist for years, decades or even centuries (Table 1). Some errors are corrected, but rejected by the consensus (Table 2). Some errors are compounded: leading to other “nested” errors before being corrected themselves (e.g., Millikan’s faulty value for the viscosity of air; Bäcklin, 1929; Kellström, 1936; see also Bassow 1991). Some errors are “corrected,” but only by introducing new errors that have failed to address the underlying source of error, leading to *error cascades* (Section 7 below). Yet other errors are resolved by mere happenstance (Table 3). Historically (as noted in the opening), errors are quite plainly not corrected systematically, by some standard method.

< INSERT TABLE 2 ABOUT HERE >

< INSERT TABLE 3 ABOUT HERE >

Of course scientists can and do correct mistakes. Yes, the baroque fraud of Jan Hendrik Schön was exposed. True, chemists no longer accept Newton’s concept of atomic affinities, and the grand tables of affinities that once adorned lab walls are now remote memories. Just as the remarkable claim, announced in *Science* magazine in, 2009, that chronic fatigue syndrome (CFS) was caused by the XMRV virus has been retracted. And so on (Allchin, 2015b; forthcoming). Yet Schön’s work spawned much futile effort in other labs (Reich, 2009). The concept of affinities lingered for a century, buoyed in part by Newton’s towering reputation (Kim, 2003). Over $12 million was spent on research before finding the contaminated commercial reagent that led to the mistaken claims about CFS (Cohen & Enserink, 2011). In the context of *self*-correction, “eventual” correction alone is irrelevant. What should merit our attention, instead, is the time lag in correction, as well as the epistemic consequences of a yet-uncorrected error. The tabletop cold fusion fiasco was largely resolved in a tidy 39 days, but not before it had spread across the covers of major news magazines and led to Congressional hearings and substantial new research grants (Simon, 2001). How are we to understand or characterize that interim period *before* the error is fully resolved? A promissory note that science “ultimately” corrects itself is conspicuously vague and uninformative. The critical question is not whether scientists correct mistakes (they do so regularly), but when and how: do they do so *systematically* and *in a timely manner*, according to some particular *mechanism* that details *how* and *when* scientists correct errors?

Some observers try to present every instance of error-correction as casual "proof" that science is self-correcting (e.g., Schimmack, 2019). But, of course, this is a recognized fallacy in causal thinking. A retrospective view hardly amounts to a prospective claim that scientists *will* correct errors. There is no control allowing historical comparison. A responsible philosophical account needs to specify the mechanism: the essential entities and how their respective activities are organized to consistently yield a particular outcome.

Given the widespread appeals to *self*-correction in science, one might expect philosophers to have detailed its mechanism. Yet the only model of systematic and timely self-correction available seems to be servomechanisms and other homeostatic regulatory systems. The canonical examples are centrifugal governors on steam engines, anti-aircraft gunnery, or thermostats, all embodying 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 enviromental conditions, they register an "error" and guide the system to its former state of equilibrium: genuine self-correction (Figure 1).

< INSERT FIGURE 1 ABOUT HERE >

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 (e.g., Kuhn, 1970). At best, one can say that there is a temporary "equilibrium" when there is simultaneous coherence between: (a) different sets of findings, (b) theories and data, and (c) different theoretical perspectives. Science must settle for something less than genuine *self*-correction.

A more modest alternative version of the self-correcting or error-correcting thesis asserts that *scientists will revise and update their claims in light of new evidence* (e.g., Bruner & 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 weaker 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 possibility of 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.

Ultimately, a method for *probing* for errors must be considered integral to a fully *self*-

correcting process (e.g., Mayo, 1996; Platt, 1964; Popper, 1963). 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” (1973, p. 264).[1]

A major factor in 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 & Panovsky, 2021). Clarifying how this component of the process is essential is, again, another reason why articulating a general philosophy of error in science is so important.

Lastly, one should bear in mind the standard rhetoric of “*self*-correction” (and its variants) as it appears in public discourse. That is, the philosophical principle is commonly used in appeals defending the authority of science (e.g., Alberts, et al., 2015; McNutt, 2020; Tharp, 2023). The implicit claim is that non-experts can trust science precisely because its institutionalized methods will find and remedy error (as reflected in Romero’s definition, but not in the accommodation version). Hence, they contend, science should be regarded as essentially error-free in most meaningful cultural contexts. At the same time, few science advocates pretend that science is perfect. “Science is tentative,” science educators are habituated to say. So there is an unresolved tension in public rhetoric about science as self-correcting or dependably error-correcting. If science is not *systematically* self-correcting, then it becomes important to clarify, at least, *when or how* errors are indeed corrected.

The pledge of self-correction is also used to defend the autonomy of scientific instutitions. They contend that scientists (not federal official or other “outsiders”) are responsible for (and succeed in) monitoring and sanctioning the conduct (or malpractice) of their peers. Here, we need to distinguish between self-*correction* and self-*policing*. One is an epistemic concept; the other, political.

Conventional views of “self-correction” in science typically cite two mechanisms: (pre-publication) peer review and replication. Peer review is supposed to function like the sensor in a servomechanism (Figure 1), and detect methodological lapses before they are published, presumably protecting the published “scientific record” from error. McIntyre (2019) articulated the widely shared perspective: “Having someone else review work before publication is the best

[1]Recently, Jamieson, et al. (2019) have echoed, without evidence, the standard faith in a “culture of critique.”

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.) 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 immediate remedy). Yet the limits and weaknesses in both these processes have been recognized for decades (e.g., Broad & Wade, 1981). 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 ideal? That is an empirical question, and it had been tested. For example, confirmation bias intrude. 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, errors were not identified uniformly across all reviewers. The editor of the *British Medical Journal* summarized other investigations of the efficacy of peer review: “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 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 (a heuristic against credibility bias) and it does not precipitate any explicit selectivity. Thus, we should refer to this later stage (reflecting Merton’s norms and Oreskes’s comment above) as *critical discourse*, not peer review. The limits of this process will be addressed in Section 8.

Problems also haunt the ideal of replication.[2] Reproducibility is certainly a benchmark for productive science, allowing scientists to build on the stable recurrence of results based on

---

[2]Here, I note the additional problem that philosophers do not seem to agree even on what a replication *is* or what counts as a successful replication.

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, 1981, 170). Sadly, perhaps, replications may easily echo the original error.

On may appreciate the conundrum of replication through the case of 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. *Whose* results were in error? 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. Sequel investigations with well water alone persuaded him that light was the key factor, not plants. 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. Jan van Ingenhousz and others acknowledged that the green scum — algae — was a plant and that both plants and light were required for the restoration of air. Error resolved, but only through a series of controlled experiments that clarified the role of each variable. That is, in his replication efforts, Priestley had mistaken correlation for causation. Replication *simplicter* was uninformative in that regard.

There are plentiful other 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. Errors are not detected merely through replication.

Peer review and replication, touted as methods for reliably filtering errors, fall far short of their normative promise. They do not guarantee “self-correction,” or even error-accommodation.

4. Mechanisms of 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.) Van Ravenzwaaji, 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* or *prophylactic* measures are best viewed as a form of quality control, better characterized philosophically as *self-checking* than as self-correction.

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, theoretical frames, 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.

< INSERT TABLE 4 ABOUT 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 Ravenzwaaji, 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? In this context, a catalog of potential error types noted above functions as a valuable diagnostic reference tool. And given that errors can be slippery and can easily pass unnoticed, a complete catalog needs to be wide-ranging, and not focus on just selected factors perceived to be most problematic.

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 conrete historical evidence). Having clarifed the conceptual context, we may proceed to the core historical analysis, framing the results in philosophical terms. By drawing on a large set of historical cases across multiple disciplines, we can characterize *when* and *how* errors in science have been corrected. These might well form the basis for developing a more *systematic* and *complete* general framework, which might be considered a model, or mechanism, 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? One can theorize abstractly, or one can work empirically, analyzing historical cases. One can simply *document* how scientists have remedied errors effectively in the past. Of course, one can hardly generalize from one case study alone. However, individual instances may provide fruitful insights. A wider sample may be further informative, even if (statistically speaking) it is haphazard (non-random). Important features may resonate among cases, indicating potentially significant patterns. Alas, with no master list of errors in the history of science, one must rely on brute search and the affordances of cumulative historical knowledge.

The pathway to discovering error begins when a problem emerges within the growing body of knowledge: an experimental discrepancy, anomaly or dissonant findings. We may call these *incongruences*. 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 disrupts the coherence of existing knowledge—exposing a flaw somewhere in the justificatory network. But where is usually not immediately obvious.

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 show that the relationship of Boyle’s law broke down at high pressures (say, above ten atmospheres). 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 errors 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, which in turn resulted from 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 then considered inert (nonreactive) for over half a century. Neil Bartlett approached the “fact” with a fresh eye and demonstrated in 1962 that they can indeed form compounds: an incongruence 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 “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, 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 varying estimations of the epistemic severity of any incongruence and/or to divergent interpretations of its 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. 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 and Panofsky, 2021). They may have tried to access a natural phenomenon by means other than the one originally presented. They may have been pursing 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 haphazard. There is no systematic mechanism 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 and *noticing* 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 investiation.

7. Isolating Errors and Differentiating Domains

An incongruence is not yet a clearly delineated error. It only signals indirectly that one exists—somewhere (a variant of the familiar Duhem-Quine dilemma). 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 and Richardson, 1993; Craver and 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* (Figure 2; 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 error of cultural bias.

Similar stories can be told in other cases. For example, anthropologists tried repeatedly to define human uniqueness in terms of tools, finessing their response each time animals were found to exhibit the purportedly exclusive human trait. So, too, for language use (Allchin, 2013, pp. 132-140). In the 1950s-60s biochemists proposed—and rejected—a series of 16 high-energy intermediates of oxidative phosphorylation. Ironically, the very notion of an intermediate molecule with a high-energy bond proved incorrect, and the chemiosmotic theory was ultimately adopted in its place (Allchin, 1997). 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, 1992).

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 (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 affinities, which had dominated chemistry for a century (Kim, 2003).

Science also differs markedly from other self-regulating systems in the initial uncertainty of any proposed error correction. Accordingly, proposed solutions to a problem of incongruence are typically *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. He needed to collect evidence of the consistent correlation of contaminated water and disease, *and simultaneously* to show that the same evidence was incompatible with a miasma explanation (Figure 3) (Johnson, 2006; Snow, 1855).

To “test” an error correction, researchers may gather more evidence toward justifying an alternative, new conclusion. Error *correction* in science thus typically entails *conceptual change*. A revision to the justificatory structure is involved, associated with a broadening or deepening of the evidence. Error correction thus creates (counterintuitively, perhaps) new (albeit negative) knowledge.

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. Even in isolating errors from acknowledged incongruences, science faces numerous uncertainties, thereby mitigating against guaranteed or necessarily timely solutions.

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 of 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 are 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 measured 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 counterbalance 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; Nutrition Source, 2019). Complementary "interests" were essential, here, in disrupting the propagation of an error. Social epistemologists have already noted the importance of diverse perspectives in a healthy epistemic community (e.g., Harding, 1991; 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. 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 more complete general picture of error correction in science (Figure 4). First, new evidence 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 disturbance challenges the coherence of the current knowledge and its justification (Section 5). To lead further to error correction, the problematic finding(s) must then be *noticed* (Section 6). With resources and time available, further work likely ensues. The nature of the incongruence must then be traced (or *isolated*) to a particular (yet unknown) *source of error* or error type in the original justification (Section 7). Typically, additional tests next help confirm or disconfirm the attribution of error. This will correspond to a new alternative interpretation of the evidence. 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.

< INSERT FIGURE 4 ABOUT HERE >

Comparison with Figure 1 may indicate that the servomechanism model, while not appropriate itself as a guide to science, nonetheless contains many corresponding elements. Of course, the parallel is suggestive only and easily overstated.

When correction is viewed as portrayed in Figure 4, one may also see similarities to the general dynamics of an adaptive system, such as the immune responce or natural selection (e.g., Lewontin, 1970; Plotkin, 1997). The comparison helps underscore how a system of correction may be *nonteleological* and indeterminate. That is, it does not return to its original state of justification, where it began, even when coherence between concept and evidence has been reestablished. Adaptive correction and "self-correction" are fundamentally different.

10. When Does Science Correct its Errors?

Errors in science are frequently corrected. But we do not (and cannot) know of errors that have not yet been detected or corrected, leaving an important sampling deficit for any definitive philosophical assessment. 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 5-8). They are not consistently corrected through replication or peer review, as typical rhetoric contends (Section 3). When viewed as an ensemble, the historical cases indicate just how scientific work sometimes corrects errors, while on other occasions it does not (Figure 4). Many conditions, contingencies and qualifications are involved. In practice, science meets these criteria unevenly (Tables 1-3). Namely, science is not "self-correcting" in any guaranteed sense, as typically claimed (Figure 1). The telling question, rather, is: *When* does science correct its errors?

We may now resolve the epistemic incongruence that opened this paper. 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) timely and

(d) probative (e.g., Mayo, 1996), 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, or 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 resolve errors may rely on:

(3) available observational technology

(4) available collateral knowledge, and

(5) available diversity of perspectives.

These last three factors are beyond any systematic control, and thereby seem to establish an ineliminable limit to the timeliness of error-correction in science. On the other hand, these limits 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 provisos in mind*. However, the process of error-correction falls short of a *mechanism*. Given the various contingencies, it has limited normative guidance, other than in conceptualizing the process as a whole. Still, it might guide science journalists, public policy-makers, citizens and consumers, who may all seek to assess the degree to which particular scientifc claims may be considered reliably "settled" by science.

Reflecting on the 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.

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., and 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 on 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 on Science,* 9, 38-59.

Allchin, D. (2007). Teaching science lawlessly. In P. Heering and D. Osewold (Eds.), *Constructing Scientific Understanding through Contextual Teaching*, Berlin: Frank & Timme, 13-31.

Allchin, D. (2012). A comedy of scientific errors. *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*. Dordrecht: Springer, 161-171.

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. *American Biology Teacher* 84:321-323.

Allchin, D. (forthcoming). *Toward a Philosophy of Error in Science*. 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
Bäcklin, E. (1929). Eddington's hypothesis and the electronic charge. *Nature* 123, 409-410.
Baldwin, M. (2019). Peer review. *Encyclopedia of the History of Science*. doi: 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/
Bassow, H. (1991). Interdependency and the importance of errors in chemistry: How the search for a single error led to reexamination of the work of five Nobel Laureates and revised values for certain fundamental constants. *Journal of Chemical Education* 68(4), 273.
Bechtel, W. and 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.
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, 93-9.
Chapman, K. (2019). The element that never was. *Chemistry World* (June 16). 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 submisstions: A cross-disciplinary investigation. *Behavior and Brain Sciences*, 14, 119-186.
Cohen, J., & Enserink, M. (2011). False positive. *Science*, 333 (23 September), 1694-1701.
Catolog of Bias Collaboration. (2024). Catalog of bias. https://catalogofbias.org/ (Accessed December 13, 2024).
Craver, C. and Dan-Cohen, T. (2024). Experimental artifacts. *British Journal for the Philosophy of Science* 35, 381-390.
Craver, C. and Darden, L. (2013.) *In Search of Mechanisms: Discovering Across the Life Sciences.* University of Chicago Press.
Darden, L. (1991). *Theory Change in Science: Strategies from Mendelian Genetics*. Cambridge University Press.
Epstein, M. (1999). The Allison magneto-optic method of chemical analysis. http://mikeepstein.net/path/allison.html
Fee, E. (1979). Nineteenth-century craniology: The study of the female skull. *Bulletin of the History of Medicine* 53, 415-33.
Finn, B. S. (1964). Laplace and the speed of sound. *Isis* 55, 7-19.
Flaherty, C. (Jan. 22, 2020). Beef with Harvard. *Inside Higher Ed*. https://www.insidehighered.com/news/2020/01/23/texas-ams-beef-harvard (accessed December 14, 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.
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. DOI 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.
Guttinger, S. & Love, A. (2019). Characterizing scientific failure. *EMBO Reports* 20, e48765.
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.
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. Boston Studies in the Philosophy of Science, No. 267*. Dordrecht: Springer.
Huckra, J. P. (1992). The Hubble constant. *Science* 256, 321-325.

Huckra, J. P. (2008). The Hubble constant. https://www.cfa.harvard.edu/~dfabricant/huchra/hubble
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), 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.
Kellström, G. (1936). Viscosity of air and the electronic charge. *Physics Review* 50(July), 190.
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.
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.
Laudan, L. (1977). *Progress and Its Problems*. Berkeley: University of California Press.
Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science* 48, 19–49.
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.
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.
Lightman, A. and Gingerich, O. (1992). When do anomalies begin? *Science* 255, 690-695.
Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.
Losee, J. (1972). *A Historical Introduction to the Philosophy of Science*. Oxford 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.
Majumdar, V.D. & Vajifda, M.B. (1938). Coefficient of viscosity of air. *Proceedings of the Indian Academy of Science (Mathematics and Science)* 8, 171-178.
Marshall, B. J. (2001). One hundred years of discovery and rediscovery of *Helicobacter pylori* and its association with peptic ulcer disease. In *Helicobacter pylori: Physiology and Genetics,* ed. H.L.T. Mobley, G.L. Mendz, and S.L. Hazell, ASM 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.
McNutt, M. (2014). Reproducibility. *Science* 343, 229.
McNutt, M. (Dec. 16, 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.
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.
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
Nye, M.J. (1980). N-rays: An episode in the history and psychology of science. *Historical Studies in the Physical Sciences* 11, 125-156.
Pashler, H., & Harris, C. R. (2012). Is the replicability crisis over-blown? Three arguments examined. *Perspectives in Psychological Science*, 7, 531–536.

Peterson, D. and 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.
Rocke, A. (1984). *Chemical Atomism in the Nineteenth Century: From Dalton to Cannizzaro.* Columbus: Ohio State University Press.
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. (2019). Psychological science is self-correcting [blog]. Schimmack, (2019). Psychological science is self-correcting. Accessed December 13, 2024.
Science and Technology Committee. (2011). *Peer Review in Scientific Publications. Eighth Report of Session 2010–12, Vol. I*. London, UK: House of Commons.
Simon, B. (2001). *Undead Science: Science Studies and the Afterlife of Cold Fusion*. Rutgers University Press.
Sismondo, S. (2018). *Ghost Managed Medicine: Big Phrama’s Invisible Hands*. Manchester, UK: Mattering Press.
Slack. F.G. (1935). The magneto~optic 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.
Solomon, M. (2001). *Social Empiricism*. Cambridge, MA: MIT Press.
Soennichsen, J. (2008). *Bretz’s Flood: The Remarkable Story of a Rebel Geologist and the World’s Greatest Flood*. Seattle: Sasquatch Books.
Stanford, P. K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives.* Oxford University 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.
van Ravenzwaaij, D., Bakker, M., Heesen, R., Romero, F., van Dongen, N, et al. (2023). Perspectives on scientific error. Royal Society Open Science 10, 230448.
Vazire, S. & Holcombe, A.O. (2022). What are the self-correcting mechanisms of science? *Journal of General Psychology* 26, 212–223.
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.

Figure 1. A standard servomechanism as a model of genuine self-correction.

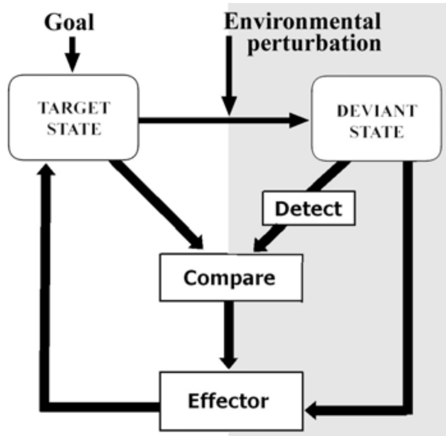

Figure 2. The structure of an error cascade. Errors are identified and abandoned, but not fully isolated to the fundamental error or assumption, and hence leading to a similar subsequent error.

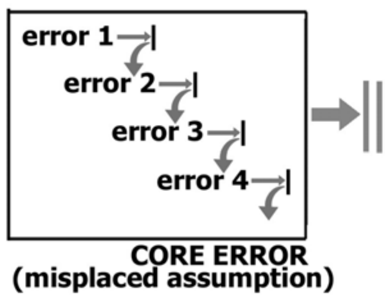

Figure 3. John Snow's strategy for demonstrating water, not miasmas, as the cause of the cholera epidemic in London in the 1850s. He showed that drinking water from the Broad Street pump was consistently associated with cases of cholera (shaded area), while many cases that would have been expected based on airborne miasma instead escaped the disease (Snow, 1855).

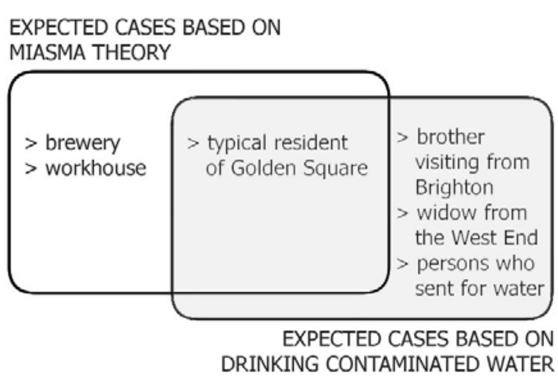

Figure 4. A simplified framework for interpreting the major components of error-correction in science. Note similarities, as well as significant differences, with a truly self-correcting system (Figure 1).

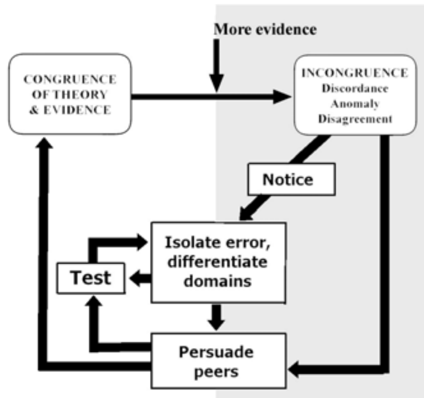

Table 1. Some notable examples of long-standing errors in science.

Errors (chronological by date of discovery)	*Duration of Error (Years)*	*Source*
Ptolemaic/earth-centered solar system (1st century, CE - ~1600)	1700	
Harvey's denial of observable capillaries (1649-1661)	12	Elkana & Goodfield (1968)
California as an island (1622-1701)	79	McLaughlin (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 (1974); LeGrand (1972,1974)
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)	2,000+	
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 (1979)
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	Hein & Hein (1966), 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	see main text
high-energy intermediates of oxidative phosphorylation (1953- ~1977)	24	Allchin (1997)
lungs as evolved from the swim bladder (1859-1978)	119	Liem (1989)
central dogma (ribozymes) (1965-1982)	23	
viceroy as a Batesian mimic of monarch butterflies (1869 -1991)	122	Ritland & Brower (1991)
stress & diet as primary causes of ulcers (mid-1930s-1984)	50	
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), LeGrand (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), Allchin ()
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

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)