A SCIENTIFIC AND INTELLECTUAL HISTORY

History of Diagnostics, Observability, and Debugging in Natural Sciences

From recorded observations and laboratory experiments to distributed observatories and computational evidence

— Natural processes and the artifacts through which we investigate them —

Concept and direction by Dmitry Vostokov, DumpAnalysis.org

Developed with AI assistance (GPT-6 Astra Max and Fable 5.1 Extra) and checked against the sources listed in this article.

Version 5 | September 2026

Historical synthesis with 96 references, chronology, glossary, and evidence taxonomy

Abstract

Natural sciences reconstruct processes that cannot always be observed directly. A spectrum provides information about a distant star. A seismic record constrains the structure of the Earth. A sequence of bands in a laboratory tube helps us distinguish possible mechanisms of DNA replication. In each case, we have an artifact for analysis and a problem of connecting that artifact to the process that produced it. This history follows the development of such connections.

We consider diagnostics, observability, and debugging across physics, astronomy, chemistry, Earth and planetary sciences, and the life sciences. Diagnostics concerns the interpretation of signs and the investigation of possible mechanisms. Observability concerns what can be inferred from the available observations, under a specified model and measurement arrangement. Debugging concerns faults in experiments, instruments, computational models, and analysis procedures. A natural phenomenon need not be defective to require diagnosis. The object of debugging is usually the system through which we investigate it.

The chapters follow recorded observations, controlled experiments, quantitative measurement, spectroscopy, imaging, natural archives, molecular methods, distributed observatories, and computational science. We also consider unsuccessful interpretations and corrected measurements. These cases help us distinguish an anomaly in nature from an anomaly introduced by sampling, instrumentation, computation, or an unsuitable model. A chronology, glossary, and evidence taxonomy provide additional ways to navigate the material.

Throughout this history, we distinguish the recorded observation, the pattern recognized in it, the proposed mechanism, and the experiment or comparison used to check that mechanism. More detailed observations extend the available evidence. Their interpretation still depends on the conditions of acquisition and on the alternatives that the investigation can distinguish.

How to read this history

This is a history of diagnostic capabilities and practices. It is not a general history of every discovery in natural science. We follow developments that changed what investigators could observe, preserve, compare, reconstruct, or correct. The main fields are the physical, chemical, Earth, planetary, and biological sciences. Mathematics and statistics enter as methods used within those fields. Clinical medicine, general software debugging, and engineering maintenance have their own histories and appear here where they contribute directly to scientific investigation.

The vocabulary requires some historical care. An ancient astronomical observer did not use observability in its twentieth-century control-theoretic sense. A nineteenth-century experimentalist did not necessarily call fault isolation debugging. We use these terms as an analytical framework and identify the historical practices to which we apply them. This does not establish that those practices descended from one common diagnostic discipline.

Dates normally identify publications, documented observations, or implemented capabilities. They do not automatically establish priority. A published method may have a long practical prehistory, and similar methods may develop independently. The surviving record also favors institutions that preserved instruments, specimens, notebooks, and correspondence. Technicians, assistants, collectors, local observers, and data curators contributed to the evidence even where the publication record did not give them corresponding visibility.

Numbered references identify original papers, primary texts, institutional collections, and methodological publications. Historical examples describe the evidence and claims of the cited work. Our comparisons with pattern-oriented diagnostics are interpretations developed here. They should be read as proposals for organizing analysis, not as claims that earlier scientists used our pattern names.

Neighboring practices and their primary questions

Table 1. Related practices and the questions they address.

Practice	Primary question	Characteristic evidence
Observation	What was recorded under these conditions?	Notes, specimens, images, counts, spectra, time series
Measurement	What value can be attributed to a defined quantity?	Calibrated responses, standards, uncertainty information
Monitoring	How does a selected condition change over time?	Repeated measurements, baselines, flags, operational records
Diagnostics	Which state, process, or mechanism explains the evidence?	Patterns, comparisons, residuals, controls, competing models
Observability	Which distinctions can this observation arrangement support?	Measurement models, sensor coverage, resolution, identifiability analysis
Experimental debugging	Which part of the investigation produces an unintended result?	Substitution tests, blanks, instrument logs, pipeline versions
Validation	Is the model adequate for its stated scientific use?	Independent observations, predictive checks, uncertainty and sensitivity studies
Historical reconstruction	What past process is consistent with surviving traces?	Strata, fossils, isotopes, archived observations, provenance

These practices overlap. For example, a disagreement between two instruments can initiate a diagnostic investigation of the natural object, a debugging investigation of the instruments, or both. The investigation must preserve enough information to tell these possibilities apart.

Recurring developments

- Observations became records that other investigators could compare and reinterpret.
- Controlled interventions made competing explanations produce different expected results.
- Calibration and uncertainty connected local measurements to wider systems of comparison.
- Instruments converted inaccessible processes into spectra, images, tracks, and time series.
- Natural archives extended investigation into periods before instrumental recording.
- Networks combined observations across locations, wavelengths, and kinds of evidence.
- Computational pipelines became both instruments of discovery and objects of diagnosis.

Contents

Seven parts · 35 chapters · three reference appendices

Table 2. Contents and opening pages.

PART I

Foundations of natural scientific diagnosis

1. Vocabulary and historical scope

Let us first distinguish a natural process from the arrangement used to observe it. A star emits radiation. A telescope and detector collect part of that radiation. An analysis procedure converts the detector output into an estimate of brightness, position, or composition. These are related activities, but they are not the same activity. A surprising result may originate at any of these levels.

We use diagnostic artifact for a preserved observation or derived representation that can be examined during an investigation. A specimen, a photographic plate, a chromatogram, and a simulation residual are different kinds of artifacts. Their interpretation requires different models. A specimen may preserve material from the object of interest. A residual expresses a difference between an observation and a model calculation. The residual is therefore already a relationship between two artifacts.

Diagnostics in natural science does not require a failure. We may diagnose the physical conditions of a nebula, the origin of a mineral, or the process responsible for a change in a population. An anomaly is a departure from an expectation; a defect is a problem relative to an intended function. Confusing these concepts encourages us to treat unfamiliar natural behavior as faulty data, or faulty data as a new natural law.

Observability also has several meanings. In control theory, it concerns whether internal state can be distinguished from outputs under a specified system model. Kalman's 1960 treatment of control systems helped establish the state-space framework in which observability and controllability became precise properties. In scientific practice, we also use the word more broadly for the availability of evidence about otherwise inaccessible processes. The broad use should preserve the question of distinguishability rather than equate observability with collecting more data. [1]

Debugging enters when the experimental or computational arrangement behaves contrary to its intended operation. Examples include an incorrectly calibrated clock, a contaminated reagent, a sign error in a simulation, and a data filter that removes the phenomenon under investigation. A successful correction restores an intended capability. Whether the restored capability supports a scientific explanation remains a separate question.

This distinction gives the history its scope. We follow changes in the relationship between natural processes and the evidence available for their analysis. The growth of a discipline can then be examined through its artifacts and diagnostic methods, without reducing it to a sequence of famous discoveries.

2. Observation records and collective memory

An observation becomes useful beyond its original occasion when enough of its context survives. Babylonian astronomical diaries provide early examples of systematic recording. A tablet in the British Museum records astronomical phenomena for 323–322 BCE. Such records belong to a particular scholarly and cultural practice; their later use as historical evidence does not depend on accepting every original interpretation attached to them. [2]

Chinese records of the guest star of 1054 provide another connection between a historical observation and a physical object studied much later. The event is associated with the supernova remnant now called the Crab Nebula. The old record and the present remnant answer different questions. One constrains the appearance of the event to observers at the time; the other preserves material and radiation from the event's subsequent evolution. Their combination extends the available history. [3]

Observation practices also developed through translation and criticism. Ibn al-Haytham's work on optics examined relationships among light, visual perception, and geometrical arrangements. The first three books of the medieval Latin version of the Book of Optics are available in A. Mark Smith's 2001 critical edition and translation, making it possible to examine this reasoning in detail. We should resist assigning the whole of experimental science a single place or moment of birth. Particular texts establish particular practices, not an exclusive origin for inquiry. [4]

For diagnostic analysis, a record has at least two histories. One is the history of the phenomenon. The other is the history of recording, copying, classifying, preserving, and translating it. A missing entry may indicate that nothing was seen, that observation was impossible, that no record was made, or that the record was lost. These possibilities are especially important when old records are converted into numerical datasets.

The same problem occurs in a modern observatory. A gap in a time series may be caused by weather, maintenance, a communications failure, or a selection rule. We can therefore compare ancient records with modern traces at the level of evidence handling. This is a methodological comparison, not a claim that they share the same semantics.

Collections and archives make observation cumulative. They also determine what can be reexamined. A preserved specimen may permit a measurement unavailable to its collector. An archived image may contain an object that was irrelevant to the original project. Scientific memory includes these unused possibilities, provided that identity and acquisition context remain recoverable.

3. Controlled experiments and diagnostic intervention

Observation alone may leave several mechanisms compatible with the same result. An experiment changes selected conditions so that the mechanisms predict different outcomes. The intervention becomes part of the evidence. We need to know what was changed, what was held sufficiently stable, and what the measurement procedure could detect.

Boyle's investigations with an air pump, published in 1660, made reduced air pressure an experimentally variable condition. An accompanying engraving records the pump and its components. The apparatus did more than reveal an otherwise hidden object. It created an arrangement in which effects could be compared as conditions changed. Leakage, sealing, and the behavior of the pump were consequently relevant to the interpretation of results. [5]

Microscopy introduced a related requirement. Hooke's Micrographia of 1665 joined descriptions and illustrations with an account of observation through instruments. Readers had to consider preparation, illumination, magnification, and the relation between a drawing and the object. The visible image was a product of an optical arrangement as well as a property of the specimen. This relationship persists in later microscopy, although its technical details changed. [6]

By the twentieth century, experimental design had acquired explicit statistical methods for organizing comparison. Fisher's The Design of Experiments, published in 1935, developed randomization, replication, and the analysis of designed experiments. Randomization helps prevent treatment assignment from following an uncontrolled influence. Replication provides information about variation. Blocking organizes comparisons around known sources of heterogeneity. These operations improve what an experiment can distinguish; they do not make every causal assumption unnecessary. [7]

Consider a biological response measured before and after adding a reagent. The difference may arise from the reagent, elapsed time, temperature change, or measurement drift. A suitable control addresses a specified alternative. Another control may be needed for a different alternative. Calling a procedure controlled is therefore less informative than describing which differences the design separates.

This is also the point at which experimental diagnosis resembles debugging. We isolate a condition, substitute a component, or repeat a procedure with one factor changed. However, the intervention may alter the object irreversibly. Preparation can damage a cell, sampling can disturb a habitat, and irradiation can change a specimen. The analysis needs a record of the object's state before intervention and of the transformations introduced during investigation.

4. Measurement standards and uncertainty

Quantification allows investigators to compare observations that were made at different times or places. It also introduces a chain of dependencies. A reported temperature depends on the sensor response, calibration, environment, sampling, and the definition of the quantity being reported. Additional decimal places do not remove these dependencies.

The International Vocabulary of Metrology distinguishes concepts such as measurand, calibration, measurement error, and measurement uncertainty. A measurand is the quantity intended to be measured. Calibration establishes a relationship between reference values and indications under specified conditions. Traceability concerns a documented chain of calibrations, each contributing to uncertainty. These distinctions prevent an instrument reading from being treated as a self-explanatory property of nature. [8]

The Guide to the Expression of Uncertainty in Measurement, first issued in the 1990s and retained here through the 2008 text, provided a common framework for evaluating and expressing uncertainty. It organizes contributions to a result through a measurement model. Repeated observations can reveal variability, while other information is needed for effects that do not vary during the repetition. A stable error may therefore survive many apparently precise measurements. [9]

In diagnostic terms, a calibration record is evidence about the observation process. It can help distinguish change in the natural object from change in the measuring arrangement. The record is especially valuable when a new instrument replaces an old one. Without overlap or another defensible transfer procedure, an apparent discontinuity may be inseparable from the change in equipment.

Detection also needs a decision framework. Currie's 1968 paper on qualitative detection and quantitative determination distinguished questions that are often compressed into one detection limit. Deciding whether a signal is present, specifying the performance of a detection procedure, and obtaining a useful quantitative estimate are related but different tasks. Background variability and the acceptable error probabilities matter to each. [10]

We can see the practical consequence in a trace chemical analysis. An instrument may report a small numerical value even when the procedure cannot reliably distinguish the sample from a blank. Conversely, a signal may be detected without supporting a sufficiently precise concentration estimate. The report should preserve those distinctions. Otherwise, subsequent analysis may convert a weak indication into a categorical claim simply because a number was available.

5. Observability and inverse problems

Many scientific investigations proceed from effects to possible causes. Seismic waves are recorded at the surface, while the structures that affected their propagation lie at depth. A detector records light, while the emitting material may be distant or inaccessible. We call these inverse problems because the investigation works backward through a model relating an object to observable consequences.

The backward step is rarely an exact reversal. Different structures may produce similar observations. Some features may have little influence on the measured signal. Noise can make distinctions unstable. Backus and Gilbert's 1968 analysis of the resolving power of Earth data addressed what information finite observations can provide about internal structure and the scale at which that information is meaningful. Resolution belongs to the observation and inference arrangement, not simply to the number of digits in its output. [11]

Let us consider two layers that produce nearly the same travel-time changes at all available seismic stations. A better numerical optimizer may find a more precise fit, but it cannot create a distinction absent from the observations. Additional stations, different wave types, or independent geological information may change the situation. The diagnostic question is which new evidence would separate the candidate explanations.

This is related to observability in state-space models, but several distinctions remain useful. Observability concerns internal state under specified dynamics and outputs. Parameter identifiability concerns whether model parameters can be distinguished. Practical estimation also depends on noise, sampling, finite duration, and computation. A property established for an ideal model does not guarantee a stable estimate from a real instrument. [1]

Regularization introduces additional restrictions or preferences when observations alone do not determine a useful solution. Smoothness, sparsity, physical bounds, and prior information can reduce instability. The resulting reconstruction depends on those restrictions. A smooth image of the Earth's interior does not establish that every smooth feature was separately resolved by the data.

For diagnostics, this suggests a useful reporting discipline. We can state which features are directly constrained, which depend strongly on assumptions, and which remain undetermined. We can also examine alternative reconstructions that fit the same evidence. An explicit blind spot is part of the result. It identifies where another instrument, experiment, or viewpoint may be required.

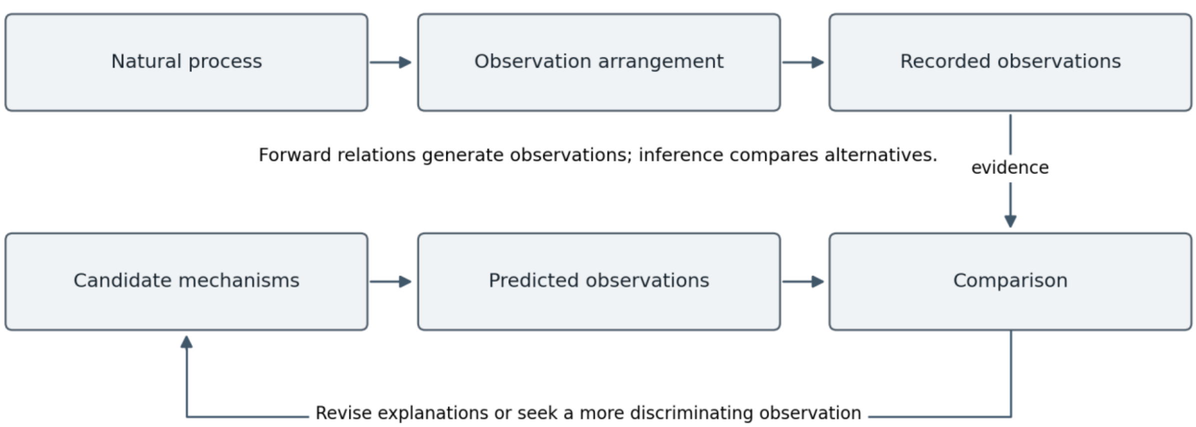

Figure 1. Observations constrain candidate mechanisms through a forward model and comparison. Agreement can leave several alternatives open. Conceptual diagram.

PART II

Physical and astronomical evidence

6. Telescopes photographs and celestial distances

The telescope changed astronomical observation by making previously unresolved features available for inspection. Galileo's Sidereus nuncius of 1610 presented observations of the Moon, stars, and Jupiter's satellites. The repeated positions of the satellites were particularly useful: a sequence of observations could support an account of motion that one isolated view could not establish. The instrument extended vision, while the record made comparison possible. [12]

The evidential problem did not end with seeing more. An apparent point might be a star, a satellite, an optical artifact, or an object below the instrument's reliable resolving power. Repeated observation under changing geometry could distinguish some of these possibilities. The telescope itself had to become a trustworthy part of the explanation.

Photography later allowed astronomical observations to be stored as physical images. Plates could be measured repeatedly and compared across long intervals. In the Harvard collection, the work of observers and measurers produced records of variable stars that supported Henrietta Leavitt's investigation of the relationship between period and brightness. The 1912 circular on 25 variables in the Small Magellanic Cloud gave a concrete observational basis for this relationship. [13]

The diagnostic value lies in the comparison. Stars in the same remote system can be treated as approximately sharing a distance for this purpose, allowing differences in period and apparent brightness to be related. Turning the relationship into an absolute distance scale requires additional calibration. A pattern in a collection and a calibrated measurement method are successive achievements.

Space astrometry extended this history through instruments and analysis procedures designed to infer positions, parallaxes, and motions from large collections of measurements. The 2016 Gaia mission paper describes a spacecraft and observing system whose scientific results depend on an integrated processing arrangement. A catalog entry is consequently a derived artifact with uncertainty and calibration dependencies, rather than a direct transcription of one image. [14]

Across these developments, astronomy moved between individual observations and population comparison. An unusual star may become interpretable only against a catalog; a catalog may conceal a special case when the processing assumes ordinary behavior. Preserving original images and intermediate measurements allows the analyst to move back from the catalog to the evidence that produced it.

7. Spectra as evidence of composition and state

Spectroscopy connected a pattern of light with properties of matter. Kirchhoff and Bunsen's work published in 1860 showed how spectral observations could be used for chemical analysis. Distinctive line positions offered an alternative to identifying a substance only through visible appearance or a sequence of chemical reactions. The same development opened a route from laboratory comparison to the interpretation of light from remote objects. [15]

The correspondence requires conditions. A line can be affected by temperature, pressure, motion, abundance, excitation, and the intervening material. The absence of a conspicuous line need not mean that the corresponding element is absent. An instrument also has its own spectral response and resolution. Two nearby lines may be recorded as one feature.

Cecilia Payne's Stellar Atmospheres of 1925 used the physics of ionization to interpret stellar spectra. The work made clear that the appearance of spectral lines could not be read simply as a direct ranking of elemental abundances. Temperature and ionization altered what became visible. The diagnostic advance was therefore a change in the model connecting the observed pattern to the underlying material. [16]

We can describe this as a distinction between a signature and a mechanism. A signature is a reproducible feature that helps identify a class of observations. A mechanism explains why the feature occurs under specified conditions. A spectral catalog can provide signatures before every relevant mechanism is understood. Later physical theory can explain similarities and differences that previously appeared unrelated.

This also shows why multiple observations are useful only when their relationships are understood. Measuring several lines can constrain more properties than measuring one, but several lines affected by the same unmodeled calibration error are not independent checks of that error. Measurements at different wavelengths or with different physical sensitivities may provide more discriminating evidence.

Spectroscopy became a common diagnostic method across astronomy, chemistry, and materials research. Its history is not a simple transfer of a finished technique from one field to another. Laboratory sources, atomic and molecular theory, detectors, reference spectra, and astronomical questions developed in interaction. The reusable idea is to connect a pattern of response to a model of the process producing that response.

8. Radiation detectors and particle tracks

Some physical processes became observable through the changes they produced in a sensitive material. A photographic emulsion, an ionization chamber, and a cloud chamber do not expose a particle to ordinary vision. They record consequences of its interaction with matter. Interpreting the record requires a model of that interaction.

Geiger and Marsden's 1909 investigation of the diffuse reflection of alpha particles recorded a small population of large deflections. Such observations placed constraints on accounts of scattering in matter and contributed to the experimental setting from which nuclear models of the atom developed. Their diagnostic importance was not that every particle behaved unusually. A relatively rare class of events carried information that the dominant behavior did not provide. [17]

Wilson's 1911 method for making the paths of ionizing particles visible through a gas provided a different kind of artifact. Condensation along ionized paths allowed tracks to be observed and photographed. The track connected a spatial pattern with a transient event, making it possible to examine a residue after the event itself had passed. [18]

However, a track is not a complete narrative of the event. The chamber may be sensitive only during a particular interval. Some interactions leave weak or ambiguous records. Geometry, magnetic fields, and the properties of the recording medium affect what can be reconstructed. A visible trajectory contains evidence of motion, but interpretation may still require assumptions about charge, energy, or particle identity.

Later electronic detectors converted interactions into timed pulses and digital records. This allowed more events to be collected and related across detector components. It also moved part of the selection into hardware and software. Trigger rules decide which events are retained. Reconstruction programs combine signals into candidate tracks or energy deposits. The stored event is therefore already the product of an acquisition policy.

For diagnostics, the history introduces an important distinction between the event population and the recorded population. A detector can be working correctly while systematically excluding a class of events outside its design. An apparent absence must be interpreted through efficiency, acceptance, thresholds, and operating conditions. Investigating the natural process and investigating the detector response are inseparable parts of the same evidential problem.

9. Astronomical anomalies and background signals

An astronomical anomaly often begins as a disagreement with the expected behavior of an instrument or source. Its interpretation may require investigators to examine both. Rejecting every unusual signal as interference would prevent discovery. Accepting every unusual signal as a new object would make instrumental artifacts indistinguishable from astronomy.

Penzias and Wilson's 1965 paper reported an excess antenna temperature at 4080 megacycles per second after considering contributions associated with the instrument and known surroundings. The adjoining paper by Dicke, Peebles, Roll, and Wilkinson discussed cosmic black-body radiation. The observational residual and the cosmological interpretation were distinct contributions that became connected. [19, 20]

A residual is informative because it survives a stated accounting procedure. It is not, by itself, a unique identification of the missing process. Further measurements must investigate properties such as frequency dependence, direction, time variation, and consistency with other consequences of the explanation. This distinction is easy to lose in retrospective accounts that start with the accepted interpretation.

The 1968 paper by Hewish, Bell, Pilkington, Scott, and Collins reported a rapidly pulsating radio source. Repetition and timing provided evidence different from a single bright or noisy observation. A candidate signal could be investigated as an organized sequence, including its relationship to the observing geometry. The resulting artifact was a time pattern whose regularity itself required explanation. [21]

These examples show that background and signal are partly question-dependent categories. An unwanted response in one investigation may become the object of another. The change in classification is warranted by additional evidence, not merely by assigning a more interesting name to the residual.

We can also distinguish the discovery of a candidate from the establishment of a stable interpretation. The first requires noticing something worth investigating. The second requires preserving its context, excluding relevant alternatives, and checking consequences beyond the initial observation. Instruments, patient inspection, theoretical models, and communication between research groups all contribute to this transition. The history of astronomy contains discovery through their composition, rather than through observation or theory acting alone.

10. Large experiments and combined observations

Large experiments make inference a property of an organized system. Detectors, calibration procedures, reconstruction software, simulations, statistical analyses, and collaboration records contribute to the final claim. A single event display may illustrate the investigation, but it cannot represent the full evidential basis.

In 2012, the ATLAS and CMS collaborations reported a new particle in searches for the Higgs boson. Their papers combined evidence from different decay channels and analyzed excesses relative to modeled backgrounds. The experiments used distinct detectors while sharing a collider and a broad theoretical setting. Agreement strengthened the interpretation, but the structure of shared and separate dependencies still mattered. [22, 23]

This is a useful qualification to the phrase independent evidence. Two measurements may have independent detector noise yet share a calibration input, a theoretical cross section, or a software library. Independence must be assessed with respect to the particular error or alternative under investigation.

The first directly detected gravitational-wave event, GW150914, was observed in September 2015 and reported by the LIGO Scientific Collaboration and Virgo Collaboration in 2016. The analysis related signals in two detectors to a model of a binary black-hole merger and considered their statistical and instrumental interpretation. The event date and publication date belong to different stages of the history. [24]

The 2017 observations associated with GW170817 extended the comparison across gravitational waves and electromagnetic observations. A combined investigation connected different kinds of instruments and physical messengers. Each observation constrained a different aspect of the event, while timing and location supported their association. [25]

Combining evidence introduces its own diagnostic work. Clocks must be related, coordinate systems understood, alerts communicated, and follow-up observations distinguished from observations that preceded the alert. A combined timeline must preserve uncertainty rather than turn every timestamp into an exact causal ordering.

These developments made scientific observability more distributed. The relevant system includes several instruments and the infrastructure that relates their outputs. Its diagnostic artifacts include environmental channels, rejected events, calibration versions, and records of analysis decisions. Such information may appear peripheral to a discovery plot, yet it helps determine which explanations the plot can support.

PART III

Chemical and molecular investigations

11. Balances gases and chemical transformations

Chemical diagnosis asks what a material contains and how it changes. Early methods often depended on appearance, heating, precipitation, dissolution, or reactions with selected substances. These procedures generated signs that could be compared with known materials. Their usefulness depended on preparation, purity, and the conditions under which the response was obtained.

The investigation of gases changed what could be included in an experimental account. Material that left a vessel could no longer be treated simply as disappearance. Lavoisier's Elements of Chemistry, translated into English in 1790 from the 1789 French work, organized chemical transformations around quantitative reasoning, the role of gases, and a revised nomenclature. Balances, vessels, and collection arrangements helped connect visible changes with material accounting. [26]

Consider the diagnostic difference between an open and a closed arrangement. If a reaction releases gas from an open vessel, the mass of the material remaining in the vessel can decrease. That observation does not by itself challenge conservation of matter in the chemical process. We first need to specify the system boundary and account for exchanges across it. The same reasoning applies when an environmental sample gains water or loses a volatile constituent before analysis.

Chemical nomenclature also changed the organization of evidence. A name can express a compositional interpretation rather than only a source, appearance, or traditional use. However, a name is still a claim about how the material should be classified. It does not replace an analytical procedure capable of distinguishing it from similar materials.

The history of chemistry therefore joins material transformations with transformations of representation. A visible reaction becomes a measured quantity; a measured quantity becomes evidence for composition; composition is expressed in a notation that permits further comparison. Each step can introduce errors. Incorrect stoichiometric assumptions, incomplete reaction, impurities, or unmeasured losses may produce an apparently coherent account.

From a diagnostic perspective, the balance is useful because it imposes a constraint on an explanation. A proposed transformation has to account for the measured material, within the uncertainty and scope of the experiment. Disagreement can direct investigation toward missing products, losses, contamination, or an incorrect model. The constraint organizes the search without identifying the mechanism on its own.

12. Separation methods and mass spectrometry

A mixture may conceal the properties of its constituents. Separation methods make some of those constituents available for more selective measurement. The diagnostic problem changes from interpreting one combined response to relating several partially separated responses to the original sample.

Martin and Synge's 1941 paper described a form of chromatography employing two liquid phases. James and Martin's 1952 work extended gas-liquid partition chromatography to the separation and estimation of volatile fatty acids. These developments made differential movement through a separation system a reproducible analytical resource. The observed bands or peaks depended on interactions with both mobile and stationary phases. [27, 28]

A retention time is consequently conditional evidence. It depends on the method and operating conditions. Two compounds may overlap, and a peak that appears single may contain more than one constituent. Repeating the separation under different conditions, adding a reference material, or using another detection method can increase discrimination.

Mass spectrometry provides a further kind of separation, based on the behavior of ions according to their mass-to-charge relationship. Ionization and fragmentation connect molecular material with a pattern of detected ions. For large biological molecules, Fenn and colleagues' 1989 paper on electrospray ionization described a method that made mass-spectrometric analysis applicable to large, fragile, polar molecules. This extended the range of possible artifacts and questions. [29]

The sample-to-result path now contains several transformations. Extraction selects material from the original object. Separation changes its distribution in time. Ionization changes its physical state. The detector produces a signal, and software assigns features or candidate identities. A problem at one stage can appear as a chemical result at the end.

For example, a missing peak may indicate an absent compound, poor extraction, suppression of ionization, degradation, or an unsuitable acquisition range. A new peak may reflect contamination or a reaction during preparation. The analysis therefore benefits from procedural blanks, recovery experiments, internal standards, and examination of intermediate results.

Combining chromatography with mass spectrometry adds complementary information. It does not make identity automatic. Library matching, exact mass, fragmentation, and comparison with an authentic reference have different evidential strengths. A useful report states which of these were available. This preserves the distance between a candidate identification and a confirmed assignment.

13. Molecular spectroscopy and magnetic resonance

Molecular spectroscopy investigates how matter interacts with radiation under specified conditions. The response provides information about energy levels, bonding, motion, and local environments. The diagnostic artifact may be a set of peaks, a continuous band, or a response changing with time.

Raman and Krishnan's 1928 report of a new type of secondary radiation described scattered light with a changed frequency. The effect provided another way to investigate molecular properties through the relationship between incident and scattered radiation. Its subsequent analytical use depended on better light sources, optical arrangements, detectors, and interpretation of spectra. [30]

Nuclear magnetic resonance developed through separate experimental approaches in the 1940s. Purcell, Torrey, and Pound reported resonance absorption by nuclear magnetic moments in a solid in 1946. Bloch and his collaborators reported nuclear induction in the same year. These methods made nuclear magnetic behavior experimentally accessible and supported later investigations of molecular environments and dynamics. [31, 32]

An NMR spectrum contains relationships among signals. Their positions, intensities, splitting, and response to experimental conditions can constrain a structural interpretation. However, the appearance also depends on sample composition, solvent, temperature, pulse sequence, relaxation, and processing. A spectrum is therefore a record of a specimen under a particular interrogation, not a context-free molecular picture.

We can use this history to distinguish observation diversity from data volume. Repeating the same acquisition may improve precision. Changing the interrogation can expose a different relationship. In structural analysis, an additional experiment that separates otherwise ambiguous assignments may be more valuable than a much longer acquisition of the original spectrum.

Processing also deserves attention. Baseline correction, phase adjustment, filtering, and peak fitting alter the representation presented to the analyst. These operations may be justified, but they should remain connected to the recorded signal. Otherwise, a processing choice can become an apparently physical feature.

The broader development of molecular spectroscopy thus extended chemical diagnostics in two directions. It provided more selective signatures, and it provided ways of testing relationships among candidate structures. The second capability is particularly important. Recognizing a familiar pattern can start an analysis, while a deliberately chosen additional observation can decide between explanations that share that pattern.

14. Diffraction structures and anomalous order

Diffraction relates the organization of matter to a pattern produced by waves interacting with it. The pattern is not an ordinary enlarged image of the object. Structure must be inferred through a physical and mathematical account of scattering.

The Braggs' 1913 work on the reflection of X-rays by crystals helped establish a relation between measured diffraction and crystal structure. This made regular internal arrangements accessible through their response to X-rays. It also created a class of inverse problems in which the measured pattern constrains possible structures without displaying the structure directly. [33]

Biological molecules extended the difficulty because a structural model must account for large and sometimes flexible arrangements. Franklin and Gosling's 1953 paper on sodium thymonucleate provided X-ray evidence relevant to DNA structure. Watson and Crick's paper in the same issue proposed a molecular structure. The distinction between experimental evidence and structural model remains necessary when explaining their relationship. [34, 35]

A fitted structure can conceal local uncertainty if it is shown only as a finished molecular illustration. The underlying experimental support may be uneven. Alternative conformations, disorder, missing regions, and model assumptions need to remain visible in the analysis. The same caution applies when a computationally predicted structure is presented in the visual language of an experimentally determined one.

Diffraction also produced evidence that challenged classification. Shechtman, Blech, Gratias, and Cahn's 1984 paper reported a metallic phase with long-range orientational order and no translational symmetry. The observed order did not fit the usual periodic crystal description. Investigating the pattern required checking possible explanations while allowing the structural category itself to change. [36]

This is a useful diagnostic case because a classification rule can become an obstacle when it is treated as an infallible property of every possible specimen. The alternatives include a faulty observation, an incorrect assignment, and an incomplete classification. A disagreement does not decide among them automatically.

The history of diffraction therefore includes both increasingly detailed reconstruction within established models and revisions to the models used to describe order. Its artifacts are valuable at several levels: detector images, processed intensities, structural coordinates, fit statistics, and records of refinement. Preserving only the final structure removes opportunities to diagnose how the interpretation was obtained.

15. Reaction rates and transient processes

Knowing the initial and final materials does not fully describe a transformation. Different reaction pathways may produce the same products. Investigating a mechanism requires evidence about rates, intermediates, dependencies, and the order of changes.

Enzyme kinetics provides a familiar example. Michaelis and Menten's 1913 investigation of invertase linked reaction rate with substrate concentration under a specified experimental arrangement. Their work is available in a modern translation with analysis of the original data. Its historical value includes the connection between a carefully controlled experiment and a model used to interpret the response. [37]

A rate curve can constrain a mechanism while remaining compatible with more than one detailed account. Similar curve shapes may arise from different processes. We can improve discrimination by changing concentration, temperature, an inhibitor, or another relevant condition. The additional experiment should be selected because the candidate explanations predict different results.

Transient measurements made shorter-lived states accessible. Zewail's Nobel lecture of 1999 describes the development of femtochemistry and the investigation of molecular changes on femtosecond timescales. Pump-probe arrangements connect controlled excitation with a delayed measurement. Varying the delay provides evidence about changes that are not resolved by a measurement of final products. [38]

Temporal resolution introduces its own limits. The measured response includes the instrument response and the distribution of states being sampled. A very fast process may appear broadened by the apparatus. An apparent delay may include synchronization error. A transient attributed to an intermediate may instead arise from the preparation or detection method.

This is closely related to trace analysis. An ordered sequence provides information unavailable in one final snapshot, but its meaning depends on the clock, the sampling interval, and the event-generation process. A time series acquired by repeated experiments also differs from a continuous record of one individual molecule. The distinction matters when a sequence is used to tell a mechanistic story.

The development of chemical kinetics and time-resolved methods thus changed diagnosis from identifying products to investigating pathways. The additional evidence can eliminate plausible narratives that were indistinguishable at the level of end states. It can also expose several pathways operating together, making a single-mechanism explanation inadequate.

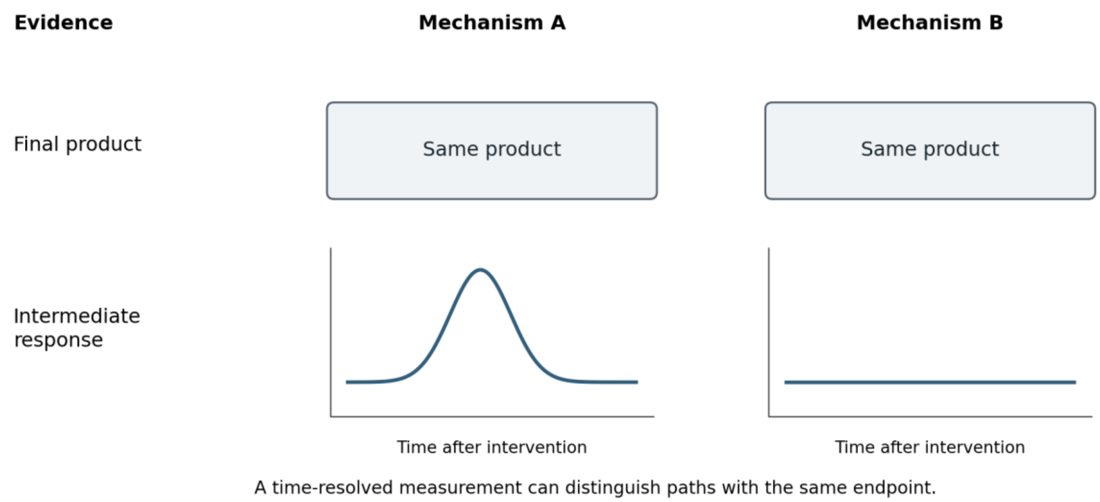

Figure 2. An illustrative pair of mechanisms gives the same final product but different intermediate responses. The curves are conceptual and do not represent experimental data.

PART IV

Earth and planetary systems

16. Strata fossils and geological time

Geology often reconstructs processes from material arrangements that remain after the processes have ended. A layer, an inclusion, a contact, and an erosion surface are spatial evidence with a temporal interpretation. The analyst must establish which relationships survive and which have been altered by later events.

Steno's 1669 work on solids naturally contained within solids provided a systematic treatment of such relationships. His reasoning connected the form and position of material with the order in which it could have been produced. The English translation published in 1671 is preserved in the Smithsonian collection. It offers a primary source for a style of reconstruction in which present structure constrains past sequence. [39]

Hutton's Theory of the Earth, published in the Transactions of the Royal Society of Edinburgh in 1788, developed a history involving processes such as erosion, deposition, and consolidation. The diagnostic move was to relate observable structures to processes operating over extended time. A landscape could then be examined as evidence of transformations rather than as a static arrangement. [40]

William Smith's geological map of 1815 combined observations across a large area. Correlation allowed separated exposures to be interpreted as parts of a wider stratigraphic organization. The map was both a result and an instrument for further investigation: new observations could agree with it, refine it, or expose a local complication. [41]

Radiometric methods added numerical age information. Arnold and Libby's 1949 radiocarbon paper compared age determinations with samples of known age. Calibration later became a continuing research activity rather than a one-time correction. The IntCal20 publication describes a Northern Hemisphere radiocarbon calibration curve extending to 55,000 calendar years before present. Radiocarbon dating and other radiometric systems have different materials, timescales, and assumptions; they should not be treated as interchangeable clocks. [42, 43]

An age estimate also needs an event definition. The age of a mineral, the age of deposition, and the age of a later disturbance may differ. Reworking can place older material in younger deposits. Contamination can mix sources. A date is therefore evidence within a geological account, not a substitute for the account.

Geological diagnosis combines relative sequence, numerical estimates, and physical process. Agreement among these can support a reconstruction. Disagreement may reveal an incorrect correlation, a disturbed sample, or an incomplete history. Preserving the disagreement is often more informative than selecting the date that best fits an expected narrative.

17. Earthquakes the interior and plate motion

Earthquakes generate waves that sample an interior inaccessible to direct inspection at comparable scales. A seismogram records ground motion at a station. Travel times, amplitudes, frequency content, and arrival patterns can then constrain both the source and the material through which the waves traveled.

These two unknowns complicate diagnosis. A difference between records may reflect a different earthquake source, a different propagation path, local ground conditions, or instrument response. Networks provide multiple viewpoints, but the observations must be related through a model and a time system.

Inge Lehmann's 1936 paper P′ used seismic evidence to argue for an inner core. The case illustrates inference from arrivals that were difficult to accommodate within the existing account of the Earth's interior. We should distinguish the historical argument for an inner core from all the properties later attributed to it. Later measurements and models extended the interpretation. [44]

The development of plate tectonics involved a different composition of evidence. Vine and Matthews' 1963 paper interpreted magnetic anomalies over oceanic ridges in relation to spreading and reversals of the Earth's magnetic field. The pattern recorded in crustal magnetization connected spatial arrangement with a history of formation and motion. [45]

Such an interpretation gains strength from relationships with other evidence. Earthquake distributions, geological correlations, ocean-floor structure, and later direct measurements of motion constrain different aspects of the explanation. Repeating one magnetic profile is useful, but it does not provide the same kind of check as evidence with a different acquisition process. The US Geological Survey's account of plate tectonics documents this convergence and its historical development. [46]

Seismic and geodetic monitoring also support the investigation of active volcanic and fault systems. Here the question changes from reconstructing long-term structure to interpreting evolving conditions. A cluster of earthquakes or a deformation signal requires a physical context; it does not uniquely specify a future event. USGS volcano-monitoring practice combines several kinds of measurements because their limitations differ. [47]

Earth science consequently provides a particularly clear separation between detection, diagnosis, and prediction. We may detect motion, diagnose a plausible process, and still have limited ability to predict a consequential transition. A useful diagnostic account states these achievements separately and preserves uncertainty about timing, location, and mechanism.

18. Weather observations and numerical prediction

Weather investigation requires observations that describe a changing spatial field. A pressure reading at one station provides little information about an approaching system unless it can be related to observations elsewhere and to a common time. Networks, standardized instruments, and communication therefore became part of the scientific method.

Bjerknes' 1904 treatment of weather prediction framed the problem in terms of an initial atmospheric state and physical laws governing its development. This formulation made the observational requirement explicit. A model cannot begin from a complete state that has never been measured; the initial condition must be estimated from incomplete and uncertain evidence. [48]

Charney, Fjörtoft, and von Neumann's 1950 paper on numerical integration of the barotropic vorticity equation demonstrated a computational approach to prediction. A simplified dynamical model could be advanced numerically from an analyzed state. The calculation introduced additional questions about approximation, boundary conditions, discretization, and the relation between model variables and observations. [49]

Lorenz's 1963 paper on deterministic nonperiodic flow showed how deterministic equations could produce behavior sensitive to small differences in initial conditions. This provided a precise example of a limit that should not be diagnosed automatically as faulty software. Prediction can diverge because of the dynamics and imperfect initial information even when a program correctly implements its equations. [50]

Operational analysis later developed extensive methods for combining observations with forecasts. In this process, an observation has value through its relationship to an evolving model state. Satellite radiances, surface measurements, and upper-air profiles are different observation types with different error structures. The ERA5 reanalysis is a documented example of a long-period reconstruction produced with a consistent data-assimilation framework, while the observing system itself changes through time. [51]

For diagnosis, a forecast error can originate in the observations, the estimated initial state, the physical model, the numerical method, or the verification procedure. Several may contribute together. Separating them requires more than inspecting the final forecast map.

Weather science thus turned continuing observation into an operational cycle of estimation, prediction, comparison, and revision. The resulting products are histories of model-assisted inference. When we compare two periods, we need to examine changes in observational coverage as well as changes in the atmosphere. A discontinuity in the data supply can leave a structure in the reconstructed history.

19. Climate records and atmospheric diagnosis

Climate investigation joins measurements with very different durations and acquisition methods. Instrumental records describe recent conditions directly through defined measurement procedures. Ice cores and other natural archives preserve material or structural traces from earlier periods. Climate models connect those observations with accounts of physical processes.

Keeling's 1960 paper reported systematic variation in atmospheric carbon dioxide concentration and isotopic abundance. Sustained measurement made seasonal behavior and longer changes available for comparison. The continuity of such a record depends on calibration, sampling decisions, site conditions, and the treatment of interruptions. [52]

Manabe and Wetherald's 1967 study of atmospheric thermal equilibrium investigated the response of a modeled atmosphere with specified relative-humidity conditions. It exemplifies how a controlled computational experiment can examine the consequences of physical assumptions. The model result and the observational record answer related questions, but they are different kinds of evidence. [53]

The Vostok ice-core study published by Petit and colleagues in 1999 related Antarctic climate and atmospheric records over approximately 420,000 years. Interpretation required age models and attention to the distinction between the age of the surrounding ice and the age of trapped air. A core supplies several related records, not a uniformly sampled recording made by an ancient instrument. [54]

Atmospheric ozone provides a complementary case of diagnosis. Molina and Rowland's 1974 paper proposed a stratospheric sink for chlorofluoromethanes involving chlorine-catalyzed ozone destruction. Farman, Gardiner, and Shanklin's 1985 paper reported large seasonal losses of total ozone in Antarctica. These publications belong to a longer investigation of mechanisms and observations; the Antarctic phenomenon required further understanding of polar conditions and chemistry. [55, 56]

Detection of change and attribution of change are separate analytical tasks. A trend can be established before its contributing mechanisms are fully distinguished. Attribution compares explanations, their predicted patterns, and other evidence. The IPCC Sixth Assessment Report's physical-science assessment provides a documented synthesis of this distinction and the evidence used in climate analysis. [57]

The diagnostic contribution of climate science includes methods for comparing imperfect records without pretending they have identical meanings. A proxy needs calibration and a process model. A model needs observational checks suited to its purpose. Agreement is most informative when it occurs across evidence with sufficiently different dependencies and when disagreement is investigated rather than removed by presentation.

20. Oceans remote sensing and planetary evidence

The ocean exposes a coverage problem. Measurements from ships can be detailed but uneven in space and time. A surface observation does not describe the full water column. A sensor can also change response during long deployment, making environmental change difficult to separate from instrumental drift.

The Argo program extended subsurface observation through profiling floats. The 2009 account describes an array and data-management system that grew from regional deployments in 1999 to a global program, reaching its initial target of 3,000 active floats in 2007. The published account also connects observation with quality control and international coordination. [58]

Here the diagnostic artifact is more than a temperature or salinity profile. It includes location, pressure, instrument information, quality flags, and the history of delayed assessment. Reprocessing may improve a result after the initial transmission. A later corrected profile and the earlier operational value should remain distinguishable.

Remote sensing extends this problem to large areas of the Earth. Landsat's record, beginning with the first satellite launched in 1972, made repeated observations of the land surface available over decades. Interpretation depends on spectral response, viewing geometry, atmosphere, cloud masking, calibration, and the relation between measurements from successive instruments. The USGS mission record documents the program and its continuity. [59]

An image classified as vegetation loss or water change is a derived interpretation. The label combines an observation with a method and often a comparison period. Seasonal variation, different illumination, or a sensor transition may produce changes in the image without the same change in the object. Inspection of acquisition metadata can therefore be as important as inspection of the map.

Planetary science adds another constraint: the sample may be remote, unique, or costly to acquire again. InSight seismic observations of Mars supported estimates of crustal structure in 2021. The inference depended on recorded waves and models of the interior, illustrating the extension of terrestrial diagnostic methods to a different planet with much sparser coverage. [60]

Across oceans, land surfaces, and planets, the recurring development is the construction of an observation system around incomplete access. Better coverage changes which processes can be studied. It also requires explicit accounts of missing regions, changing instruments, and the assumptions that connect a measured response with an environmental state.

PART V

Living systems

21. Microscopy cells and living processes

Microscopy changed biological evidence by revealing structures below ordinary visual resolution. Hooke's Micrographia made detailed illustrated observations available to readers in 1665. Leeuwenhoek's later observations extended microscopic investigation to a range of living forms. The Royal Society's records preserve the relationship between observation, correspondence, and attempts to assess reports made through unfamiliar instruments. [6, 61]

An image still requires an account of how it was produced. Sectioning selects a plane. Fixation and staining alter the specimen. Illumination and optical contrast favor some structures over others. A feature that appears sharply bounded in an image may depend partly on the preparation and imaging procedure.

Electron microscopy extended accessible spatial detail through a different interaction between the probe and specimen. Ruska's Nobel lecture describes the development of electron optics and the electron microscope. The method required its own preparation and interpretive practices. Higher resolving power did not remove the need to distinguish biological structure from preparation artifacts. [62]

Fluorescent markers added selectivity. Chalfie and colleagues' 1994 paper showed the use of green fluorescent protein as a marker for gene expression, supporting observation in living organisms. A fluorescent signal could now be associated with a selected biological component through an engineered relationship. That relationship also became something to validate: the marker's presence, maturation, localization, and effects on the system matter to interpretation. [63]

Methods such as photoactivated localization microscopy subsequently used repeated observations of fluorescent molecules to construct images at finer spatial scales. Betzig and colleagues' 2006 paper describes this approach for intracellular proteins. The image is a reconstruction from many observations, with localization and labeling assumptions that differ from those of a conventional micrograph. [64]

These changes extended biological observability from static morphology toward selected processes in living systems. They also changed the diagnostic artifact. The analyst may need the original image sequence, localization estimates, exposure settings, and segmentation results, as well as the final colored image.

We can therefore ask two separate questions. Does the image contain a reproducible feature? Does that feature correspond to the proposed biological structure or process? A successful image-processing procedure answers neither automatically. Controls that examine labeling, preparation, and reconstruction help establish the connection between visible pattern and biological mechanism.

22. Heredity DNA and sequencing

The investigation of heredity combined observations at different levels. Patterns of inheritance concerned organisms and their descendants. Biochemical experiments examined material associated with transformation. Structural studies investigated molecules. Sequencing methods later produced ordered symbolic representations of those molecules.

Avery, MacLeod, and McCarty's 1944 paper investigated the chemical nature of the substance inducing transformation of pneumococcal types. Their experiments associated transforming activity with a DNA fraction. The evidential work included separating material and examining the activity that remained. It addressed a mechanistic question through the composition and treatment of a preparation. [65]

Meselson and Stahl's 1958 experiment used isotopic labeling and density-gradient separation to investigate DNA replication in Escherichia coli. The distribution of DNA after replication distinguished among proposed replication mechanisms. A visible band was useful because the experimental design connected alternative mechanisms to different expected patterns across generations. [66]

Sequencing extended the available representation. Sanger, Nicklen, and Coulson's 1977 method used chain-terminating inhibitors to determine nucleotide sequence. The sequence was inferred from a laboratory separation and its interpretation. It could then be compared, stored, and used as evidence in further investigations. [67]

Amplification increased access to selected sequences. Saiki and colleagues' 1985 paper demonstrated enzymatic amplification of beta-globin genomic sequences. Although its application was diagnostic medicine, the capability became widely useful in biological investigation. Amplification also makes sample identity and contamination especially consequential because small amounts of unintended material may become readily detectable. [68]

Later sequencing platforms increased the scale of collection and moved more interpretation into computation. Reads must be assessed for quality and may be aligned, assembled, classified, or counted. Each operation introduces assumptions about the reference, error model, and possible biological variation. A sequence database is thus a collection of interpreted artifacts with different acquisition histories.

The historical transition is not simply from uncertain visual observations to exact letters. The letters are a useful abstraction that must retain links to their experimental support. For a difficult result, diagnostics may require moving back from an assembled sequence to reads, quality information, sample preparation, and controls. The symbolic representation makes comparison easier while leaving the biological and technical provenance necessary.

23. Physiological mechanisms and perturbation

Living processes can be investigated by controlling one condition and measuring another. The intervention must be understood physically and biologically. Changing an electrical potential, reducing gene expression, and modifying a DNA sequence perturb different parts of a system and support different kinds of inference.

Hodgkin and Huxley's 1952 quantitative account of membrane current connected experimental measurements with a model of conduction and excitation in nerve. Voltage-clamp experiments made it possible to examine currents under controlled membrane potential. The resulting model related measured responses to a mechanism capable of reproducing important aspects of electrical behavior. [69]

The control arrangement is part of the argument. Without it, changes in voltage and current influence one another in ways that complicate interpretation. Experimental control can make a component relationship more observable by restricting the range of concurrent changes.

Genetic perturbation extended this approach to molecular processes. Fire and colleagues' 1998 paper described potent, specific interference by double-stranded RNA in Caenorhabditis elegans. Jinek and colleagues' 2012 biochemical work described a programmable dual-RNA-guided DNA endonuclease associated with bacterial adaptive immunity. These publications concern distinct mechanisms and should not be collapsed into one generic invention of gene manipulation. [70, 71]

For diagnostic reasoning, both illustrate how an intervention can connect a selected molecular change with an observed response. The response still requires interpretation. Incomplete perturbation, unintended targets, compensating pathways, and changes in the experimental environment can affect the result. A lack of visible effect need not establish that the targeted component has no role.

A rescue experiment can provide an additional check by attempting to restore the relevant function. Even then, the route of restoration matters. A compensating mechanism may restore an outcome without reproducing the original process. The experiment must be interpreted at the level of the claim it is designed to test.

Physiology and experimental genetics therefore contribute a rich history of interventions that reveal otherwise hidden relationships. The most useful diagnostic question is not merely whether the outcome changed. We ask which relationship the intervention isolated, what other relationships it changed, and which observations support the proposed mechanism. This prevents a successful manipulation from being treated as a complete explanation of the living system.

24. Ecology evolution and imperfect detection

Natural history assembled observations of organisms, distributions, similarities, and differences. These collections supported classification and historical explanation. Darwin's On the Origin of Species of 1859 connected evidence from several domains with an account of descent and natural selection. The argument relied on relationships among observations rather than on a single decisive laboratory measurement. [72]

Experimental approaches later made particular evolutionary alternatives distinguishable. Luria and Delbrück's 1943 study of bacterial resistance examined variation among cultures to distinguish accounts of when resistance arose. The distribution of outcomes contained information that an average resistance count would have concealed. [73]

Ecological field experiments investigated interactions at larger scales. Paine's 1966 paper on food-web complexity and species diversity used removal of a predator to examine consequences for a community. An intervention in a natural setting could therefore expose dependencies that a species list alone did not establish. The interpretation remained tied to the community and conditions studied. [74]

The structure of replication became an explicit methodological concern. Hurlbert's 1984 analysis of pseudoreplication examined cases in which statistical treatment implied independent replication that the experimental arrangement did not provide. Many measurements from one manipulated unit are not automatically many independent tests of the manipulation. The number of rows in a dataset can therefore overstate the amount of experimental evidence. [75]

Detection presents another difficulty. A species can be present without being observed during a survey. MacKenzie and colleagues' 2002 occupancy work addressed estimation when detection probability is less than one. Repeated visits can provide information about observation as well as occupancy, under the assumptions of the model and study design. [76]

This distinction matters historically because larger surveys and automated sensors can increase the quantity of records without making every location equally observable. Time of day, season, habitat, observer skill, and the recording method affect detections. A map of recorded presence is not automatically a map of complete distribution.

Ecology and evolutionary biology consequently require diagnostic movement between individuals, populations, communities, and historical processes. Each level has its own artifacts and blind spots. A mechanism inferred from one controlled population must be examined before it is used to explain a field pattern. A broad historical pattern, in turn, can suggest experiments that separate mechanisms which remain plausible at the larger scale.

25. Environmental evidence and biological contamination

Environmental investigation often begins with mixed and transported evidence. Water contains material from several sources. Soil integrates deposition, weathering, biological activity, and movement. A biological sample may contain the target organism together with material introduced during collection or processing. The observation has to be related to a source and a history.

Environmental DNA methods provide a clear example. Ficetola and colleagues' 2008 study detected species using DNA obtained from water samples. The method connected biological presence with molecular traces outside an intact, directly observed organism. Such traces extended the possible survey methods, while adding questions about transport, persistence, extraction, and detection. [77]

We should therefore distinguish detecting a sequence from locating a living organism at the exact place and time of sampling. The inference may be justified under particular conditions, but it needs an account of how the trace reached the sample. A positive control, an extraction blank, and repeated sampling address different parts of that account.

Salter and colleagues' 2014 work demonstrated that reagent and laboratory contamination could strongly affect sequence-based microbiome analyses, particularly when sample biomass was low. An observed microbial pattern could consequently reflect the preparation process as well as the original sample. A technically successful sequencing run did not guarantee the biological interpretation of the result. [78]

Bias also arises without contamination. McLaren, Willis, and Callahan's 2019 study analyzed consistent and correctable bias in metagenomic sequencing experiments. Differential efficiencies across the measurement process can alter inferred composition. Reference communities and calibration can help investigate these effects, subject to the scope of the model used. [79]

These examples connect molecular biology with a wider environmental problem. The sample is not the whole system. It is material selected from the system by a procedure. A water concentration depends on flow, location, timing, preservation, and analysis. A sediment layer may combine material from different sources and ages. A biological sequence may arrive through an organism, environmental transport, or a laboratory reagent.

Diagnostic interpretation therefore needs a source history. We can trace the specimen through collection, storage, preparation, measurement, and computation, examining where a candidate signal could have entered or changed. This does not reduce environmental diagnosis to laboratory quality control. It gives the natural interpretation a defensible connection to the physical evidence.

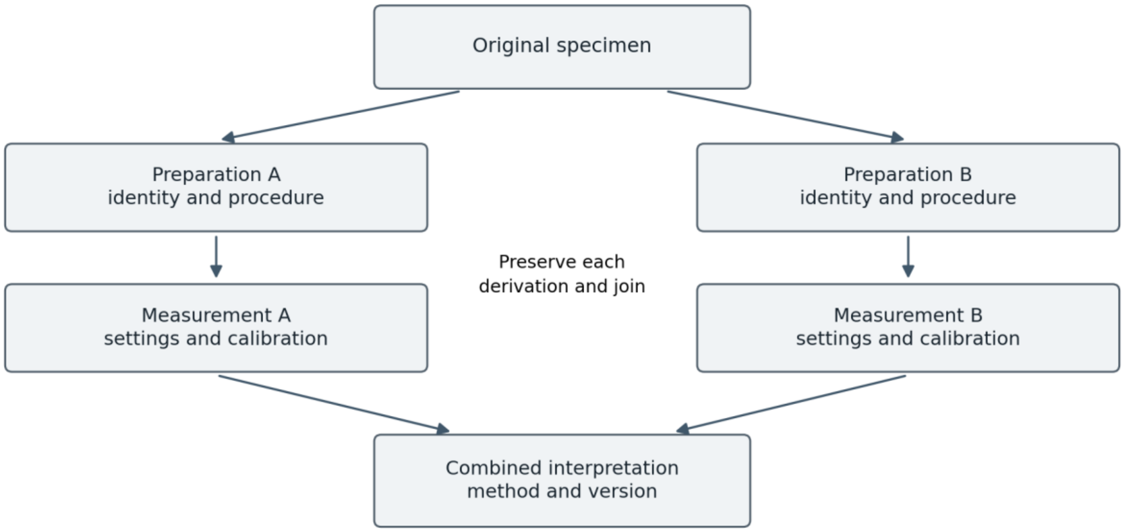

Figure 3. A shared specimen can generate distinct preparation and measurement histories. Combining the results requires preserving each derivation and the relationships used to join them.

PART VI

Evidence infrastructures and computational science

26. Simulation and numerical debugging

Computation became a scientific instrument when numerical models could generate consequences that were difficult to obtain analytically. Weather prediction was one early demonstration. Other fields used simulations to investigate fluids, materials, molecules, populations, and physical systems across inaccessible conditions or timescales.

A simulation introduces several possible objects of diagnosis. The conceptual model may omit a process. The mathematical formulation may not express the intended model. The numerical approximation may be inadequate. The program may implement the approximation incorrectly. Finally, the analysis of the program's output may support a stronger claim than the calculation warrants.

Oberkampf and Trucano's 2002 treatment of verification and validation in computational fluid dynamics distinguishes questions about solving the equations correctly from questions about how adequately the model represents the relevant physical system. Verification can include numerical error assessment and tests against known solutions. Validation requires comparison with physical evidence for an intended use. Neither activity makes the other unnecessary. [80]

Floating-point arithmetic creates an additional level. Goldberg's 1991 account explains why real-number algebra cannot be applied mechanically to finite-precision computation. Rounding, cancellation, and exceptional values can affect results. A numerically unstable calculation may produce a convincing picture without an obvious software crash. [81]

The diagnostic artifacts include convergence studies, conservation checks, intermediate states, residuals, and comparisons between implementations. A smaller time step or finer mesh can help reveal approximation effects. However, agreement between two calculations is weaker when they share the same incorrect boundary condition or source code.

We also need to separate instability in a numerical method from sensitivity in the modeled dynamics. The Lorenz example illustrates why differing trajectories do not automatically establish an implementation error. A useful test may concern distributions, conserved quantities, or another physically meaningful property rather than exact long-term trajectory agreement. [50]

Simulation history consequently extends debugging beyond finding faulty statements. It includes diagnosing relationships among physical assumptions, mathematical models, discretization, implementation, and interpretation. A corrected program may still implement an inadequate scientific model. Conversely, disagreement with an experiment may expose a measurement problem rather than a program defect. The investigation has to move between these possibilities while preserving the evidence generated at each level.

27. Scientific data and reproducible analysis

A scientific result may depend on a long sequence of transformations. Raw measurements are filtered, corrected, joined, normalized, modeled, and plotted. If only the final figure survives, another investigator cannot determine which transformations produced it or examine the effect of changing one step.

Sandve and colleagues' 2013 rules for reproducible computational research address this practical problem through recording analysis steps, preserving exact software versions, and connecting results with the data and parameters used. Reproducibility becomes a property of the recorded workflow, rather than an informal assurance that the analysis could probably be repeated. [82]

The FAIR Guiding Principles published in 2016 organized requirements for data that are findable, accessible, interoperable, and reusable. They emphasize machine-actionable metadata and the relationships needed for reuse. FAIR does not mean that every dataset is unrestricted, nor does conformity by itself establish that a dataset or inference is correct. It improves the conditions under which evidence can be located and understood. [83]

The National Academies' 2019 report distinguishes computational reproducibility, using the same input data and computational steps, from replicability across studies that obtain their own data. These definitions provide a useful convention, although terminology varies between disciplines. Repeating a calculation and independently testing its scientific claim are different achievements. [84]

Data corruption can occur in ordinary tools. Zeeberg and colleagues' 2004 paper documented mistaken conversions of gene identifiers when handling data in spreadsheets. A familiar application could silently reinterpret symbolic identifiers as other data types. The resulting problem was neither a new biological phenomenon nor necessarily a failure of the specialized analysis software. It occurred at a boundary between representations. [85]

For diagnosis, provenance should preserve more than a filename. We need the identity of the input, its version, the transformation, parameters, software environment, and the relation between outputs and inputs. A result can then be followed backward to a suspicious operation and recomputed from an earlier state.

The historical development of reproducible workflows made scientific correction more local. An investigator can ask which conclusion changes when one dataset, calibration, or processing step changes. This supports revision without discarding the entire investigation, provided that the dependencies were preserved in a form that can still be examined.

28. Instrument health and distributed observatories

An instrument observes a natural system while also undergoing changes of its own. Temperature, power, alignment, contamination, mechanical stress, and aging can affect its response. An observatory therefore needs observations of the instrument as well as observations made by the instrument.

The gravitational-wave analysis associated with GW150914 included detailed investigation of possible instrumental and environmental origins. The collaboration's companion paper on transient noise describes the examination of detector behavior and auxiliary information. Such channels helped assess whether a candidate astrophysical event could be explained by a disturbance in the measuring arrangement. [86]

This gives scientific observability a nested structure. One set of measurements concerns the target phenomenon. Another concerns the conditions under which the first set was obtained. A third may concern the software and communications used to collect and process both. A trustworthy final record depends on relationships among these layers.

Ocean floats, remote stations, satellite instruments, and laboratory robots share parts of this problem. Their operating logs may record resets, configuration changes, calibration events, lost packets, or maintenance. Such records are diagnostic artifacts even when they are not classified as scientific data by the main analysis pipeline.

Time synchronization is especially important in distributed systems. An event reported by several instruments may support a common physical explanation only if clock behavior and timing uncertainty are understood. An incorrectly interpreted timestamp can create an impossible sequence or hide a real delay. The OPERA investigation discussed later provides a concrete case in which the timing path itself became an object of analysis. [87]

The historical transition to networked observatories thus enlarged the object that scientists had to understand. The observation system includes instruments, communications, processing, data stores, and maintenance practices. A detector can be healthy locally while the combined product is incorrect because of a unit conversion or an association between the wrong records.

A useful diagnostic design preserves identifiers across these operations. It should be possible to connect a suspicious plotted point with its source measurement, calibration, configuration, and processing history. Without this connection, monitoring can report that all components are operating while leaving the scientific result difficult to explain.

29. Controls blind analysis and independent checks

Investigators can unintentionally adapt procedures to results they expect or prefer. The problem does not require deliberate misconduct. Repeatedly changing selection criteria after seeing the outcome can make a chance pattern appear more convincing than it would under a fixed procedure.

Blind analysis restricts access to selected result information while methods are developed and checked. Klein and Roodman's 2005 review describes blinding practices in experimental nuclear and particle physics. The purpose is to reduce the influence of an anticipated answer on analysis choices. Blinding does not correct a faulty detector or an inadequate model; it addresses a particular pathway by which judgment can affect a result. [88]

Controls also have specific scopes. A reagent blank can reveal contamination introduced by preparation. A positive control can test whether a procedure detects a known target under the control's conditions. A reference material can examine quantitative performance. None automatically establishes that every feature of the real sample is handled correctly.

Reporting standards make these scopes easier to assess. The MIQE guidelines of 2009 specify information needed to evaluate quantitative real-time PCR experiments, including experimental and analytical details. Such guidance treats methods and quality information as part of the scientific result, rather than as optional background to a numerical output. [89]

Independent checking needs similar precision. Another laboratory may reproduce a result using the same commercial reagent and the same analysis software. This is useful evidence, but it may not test an error shared by both. A method based on another physical principle can sometimes provide a stronger check of a particular interpretation.

We can examine independence as a relationship among possible failure mechanisms. If the suspected problem is a clock error, another analysis of the same timestamps may have limited value. If the problem is a statistical implementation, an independent calculation using the same raw observations may be highly informative. The suitable check depends on what it is intended to distinguish.

The history of controls and blinding therefore complements the history of instruments. Better measurement makes weaker signals visible; better investigative design helps prevent the analysis procedure from manufacturing their significance. Both developments are needed when a result is small, unexpected, or highly dependent on selection choices.

30. Automated experiments and machine learning

Automation changes the scale and organization of scientific work. Instruments can select measurements, execute protocols, and collect results without continuous manual operation. Machine learning can classify patterns or predict properties. These capabilities add another set of processes that must remain open to diagnosis.

King and colleagues' 2009 paper on the automation of science described a robot scientist that formulated and tested hypotheses in a specified biological setting. The work linked experimental action with explicit representations of the investigation. Its scope was a defined research task, not an autonomous solution to every part of scientific reasoning. [90]

AlphaFold's 2021 paper demonstrated highly accurate protein-structure prediction in the CASP14 assessment. The advance supplied powerful candidate structures and confidence-related outputs. A prediction remains a model-derived artifact. Its use in a particular mechanistic explanation must consider uncertainty, biological context, and the experimental question. [91]

Machine-learning-based science also introduces opportunities for misleading evaluation. Kapoor and Narayanan's 2023 study examined leakage and reproducibility problems across such research.

Information that should be unavailable at prediction time can enter through preprocessing, dataset construction, or evaluation design. The resulting score may be reproducible while failing to represent the claimed generalization capability. [92]

Let us consider a model trained to distinguish two experimental conditions. If each condition was processed in a separate batch, the model may learn the batch rather than the intended biological difference. A high score alone cannot distinguish these mechanisms. The training and evaluation design must provide an opportunity to separate them.

Automation also affects which evidence is acquired. A system that repeatedly selects the apparently most promising measurements may produce a dataset shaped by its own previous predictions. The policy, its version, and the rejected alternatives become relevant to later interpretation. An adaptive experiment needs a record of why the next observation was chosen.

The diagnostic requirement is consequently to preserve the chain from hypothesis and selection through physical action, measurement, processing, and interpretation. A generated explanation should point to evidence that can be checked. When an automated system proposes a mechanism or another experiment, the scientific value depends on the distinctions that the proposed evidence can actually resolve.

PART VII

Diagnostic cases and recurring principles

31. When an anomaly survives investigation

A successful diagnosis does more than retain an unusual observation. It connects the observation with an explanation that survives relevant checks. The form of the check depends on the object and on the competing explanations.

The cosmic microwave background case joined a persistent radio residual with a cosmological account. The initial excess and the theoretical interpretation were not identical artifacts. Their connection became productive because it supported additional questions about the radiation's properties. We can recognize a general diagnostic transition from unexplained remainder to a phenomenon with independently testable consequences. [19, 20]

The DNA-replication experiment of Meselson and Stahl used a different route. Isotopic labeling and successive replication made candidate mechanisms predict different density patterns. The experiment was informative because it changed the conditions under which the alternatives could be distinguished. The analytical lesson is to seek a discriminating sequence, rather than collect more instances of an observation compatible with every candidate. [66]

The quasicrystal case shows a third possibility. An ordered diffraction pattern survived investigation while the established structural description proved too restrictive. A diagnostic framework must leave room for revising the category used to classify the object. Otherwise, the framework can reject the very evidence needed to improve it. [36]

These cases do not supply one universal recipe for discovery. One involves a residual, another an experimental design, and another a conflict between observation and classification. What they share is the preservation of an evidential relationship while the interpretation is investigated.

We can express that relationship through four questions. What was actually recorded? Which relevant alternatives remain compatible with it? What observation or intervention would separate those alternatives? Which result would require us to revise the proposed mechanism? The questions are useful because they keep the next step connected to the uncertainty that remains.

A historical account written after acceptance can hide the work done by these distinctions. The initial artifact may look obviously significant because we already know its interpretation. Recovering the original alternatives makes the history more instructive for diagnostics. It shows what the investigators could infer at each stage, rather than attributing the final explanation to the first moment of observation.

32. When an anomaly belongs to the measurement process

Some anomalies survive repetition because the same error is repeated. A stable instrument or pipeline effect can produce a highly reproducible result. The investigation must therefore examine how the evidence was generated, not only whether the final number reappears.

In 2011, OPERA reported an apparent anomaly in neutrino flight time for external scrutiny. Investigation identified problems in the timing arrangement, including the optical-fiber connection and an oscillator-related effect. In 2012, measurements by the Gran Sasso experiments were consistent with travel at the speed of light within their uncertainties. The original discrepancy was not evidence for a confirmed violation of relativity. [87, 93]

The BICEP2 and Planck comparison concerned interpretation of a real sky signal. Their joint 2015 analysis showed the importance of polarized Galactic dust and found no statistically significant evidence for primordial tensor modes in the combined analysis. The signal did not need to be a fabricated observation to have an incorrect early interpretation. Foreground and cosmological contributions had to be separated. [94]

Microbiome contamination provides another location in the evidence chain. Material introduced during extraction can be sequenced correctly and classified correctly while being attributed to the wrong biological source. In the study by Salter and colleagues, reagent and laboratory contamination could dominate low-biomass results. The diagnostic failure concerns provenance and interpretation, even when downstream operations execute as designed. [78]

These cases differ in mechanism. OPERA involved timing. The joint polarization analysis addressed source separation. The microbiome example involved introduced material. Calling all of them noise would conceal the distinctions needed to investigate them.

We can organize the search by following the evidence backward. Start with the reported feature, recover the processing steps, identify the source measurements, and examine the conditions under which those measurements were acquired. At each boundary, ask whether the candidate explanation belongs to the natural object or was introduced by the observation process.

Correction is itself a historical achievement. It requires preserved records, accessible methods, and an opportunity for other investigators to examine the claim. A corrected result should retain enough of its earlier history to explain why the interpretation changed. Otherwise, the scientific record loses a valuable example of how diagnostic uncertainty was resolved.

33. Pattern oriented analysis across natural sciences

Our pattern-oriented software diagnostics organizes recurrent structures in artifacts and connects them with possible mechanisms. Theoretical Software Diagnostics develops this approach across collected articles. The trace-analysis reference extends the scope of analysis patterns to traces, logs, text, narratives, and data. We can use that broader scope to compare scientific evidence without assuming that every artifact is literally a software log. [95, 96]

The transfer requires a mapping. In a laboratory record, a sample identifier may organize a sequence of preparation and measurement events. In an observatory, a detector identifier may organize an activity history. These resemble an Adjoint Thread of Activity: records are selected by an attribute that remains fixed while other attributes vary. The chosen attribute and its scientific meaning must be stated.

Other patterns suggest questions rather than ready-made answers. An Unsynchronized Traces interpretation directs attention to clocks. Visibility Limit directs attention to what the acquisition arrangement can reveal. Trace Viewpoints directs attention to the different organization produced by selecting another relation in the same evidence. The pattern is useful when it guides a specific next observation or check. [96]

Table 3. Analysis patterns and proposed applications in natural sciences.

Analysis pattern or concept	Scientific application proposed here	Necessary qualification
Adjoint Thread of Activity	Follow one sample through collection, preparation, measurement, and analysis	A shared identifier must refer to the same physical sample or a documented derivative
Time Delta	Examine delays between stimulus and response	Timing resolution and synchronization constrain the difference
Significant Interval	Compare an interval around a transient or intervention	The interval must not be selected solely because it gives a preferred result
Unsynchronized Traces	Investigate disagreement between instrument timelines	Clock error must be distinguished from a physical propagation delay
Visibility Limit	Identify changes that the acquisition method could miss	Non-detection applies only within stated sensitivity and coverage
Trace Similarity	Compare repeated experiments or observation histories	Similar shape can arise from different mechanisms
Trace Viewpoints	Reorganize observations by instrument, sample, location, or process	A new view changes selection and relationships, not the underlying observations

Consider an illustrative laboratory case. A response appears only in samples measured after an instrument restart. Grouping by biological condition suggests a treatment effect. Grouping the same records by acquisition session reveals a possible instrument dependency. Neither view alone establishes the cause. The second view identifies a comparison that the first obscured: measure suitable samples across restart boundaries while preserving the original treatment assignments.

This is the practical role of a pattern language. It provides a reusable way to notice and communicate an analytical situation. The mechanism still needs field-specific evidence. A resemblance to a software pattern does not validate a physical explanation, and the comparison does not establish historical influence between the disciplines.

34. Mechanisms causal claims and limits of inference

An observed association can support several explanations. Two quantities may influence one another, share a cause, or be related through the procedure used to measure them. A pattern helps organize the problem. A causal claim requires evidence that addresses the relevant alternatives.

Mechanistic explanation adds constraints by specifying how the proposed process could produce the observations. It can connect several artifacts that were initially examined separately. For example, an atmospheric mechanism may need to account for spatial distribution, seasonality, and chemical measurements together. Agreement with one time series is weaker than agreement with several consequences that were not used to select the explanation.

However, a detailed mechanism can also be fitted too flexibly. Additional parameters may reproduce an observation without improving discrimination among explanations. Sensitivity analysis asks how conclusions depend on choices and inputs. Identifiability analysis asks which distinctions the available observations can support. These questions return us to the inverse-problem limitations examined earlier. [11]

Absence requires similar care. A survey may fail to detect a species that is present. An instrument may miss a short event between samples. A preservation process may destroy evidence of a past state. The inference should include the observation mechanism. Repeated ecological surveys provide one formal example of separating occurrence from imperfect detection. [76]

For a historical event that cannot be repeated, we can still compare independent traces and test the processes that connect them with candidate histories. Geology, evolutionary biology, and astronomy frequently work in this way. Experimental repetition of the original event is not the only possible source of evidential constraint.

The strength of a diagnosis is consequently local to a question. Evidence may establish that a measured feature is real while leaving its mechanism uncertain. It may distinguish a class of mechanisms without selecting one detailed model. It may support an explanation in one range of conditions while leaving extrapolation unresolved.

These are useful results when stated precisely. We can report the resolved distinction, the assumptions that support it, and the alternatives still open. A diagnostic narrative becomes misleading when it silently moves from one level of confidence to another. Keeping these levels explicit allows another investigator to continue the analysis from the actual state of knowledge.

35. Continuing development of scientific observability

The developments in this history accumulated rather than replacing one another completely. A modern laboratory can combine microscopy, separation, spectroscopy, and sequencing. A planetary investigation can join images, seismic records, geochemical measurements, and numerical models. Earlier forms of evidence remain useful because the methods answer different questions.

The growing role of computation changes where observability must be designed. A raw measurement can pass through several automated systems before it reaches an investigator. To explain a result, we need access to relevant intermediate representations and the conditions under which they were produced. Preserving only the final output makes some future questions impossible to answer.

The same requirement applies to autonomous experiments. An activity record should connect the scientific objective with the selected action, the physical execution, the acquired data, and the interpretation. Failed operations and rejected observations may be especially informative. A success-only record can conceal the conditions that shaped the result.

One further development is the increasing opportunity to reuse evidence for questions not imagined during acquisition. Historical plates, specimen collections, long-term monitoring, and open datasets all support such reuse. The value depends on metadata, permissions, formats, and preservation of meaning across changes in software and scientific vocabulary. Reuse is a diagnostic capability that must be maintained over time. [83]

Our proposed direction is to make the inferential path easier to inspect. For a substantial claim, the reader should be able to identify its supporting artifacts, the transformations applied to them, the patterns recognized, the competing explanations considered, and the evidence used to distinguish those explanations. This is compatible with concise publication, provided that the supporting record remains available.

Natural science will continue to encounter processes that exceed present measurement and modeling capabilities. The historical response has been to devise new comparisons, interventions, instruments, and representations. Pattern-oriented analysis can help organize this work by preserving recurring diagnostic questions across fields. Its contribution is a common analytical vocabulary whose application remains accountable to the particular natural process.

Conclusion

Diagnostics in natural sciences developed through changes in what could be recorded and in how observations could be connected with explanations. Instruments made inaccessible processes observable through their effects. Experimental design made alternatives distinguishable. Standards and calibration made measurements comparable. Natural archives extended the available history, while computation increased the range of models and evidence that could be examined together.

These developments also produced additional sources of error. Preparation can alter a specimen. A sensor can drift. A processing rule can remove an event. A model can reproduce a pattern for the wrong reason. The observation system therefore became an object of diagnosis alongside the natural system.

We can follow a common analytical movement across the examples: from a problem description to available artifacts, from artifacts to patterns, and from patterns to candidate mechanisms and discriminating checks. The movement is not always linear. A failed check may require another artifact or another viewpoint. A surviving anomaly may require a revised model or classification.

The historical contribution of observability is to make more of this investigation possible. The contribution of debugging is to restore and improve the arrangements through which it is performed. Diagnostics connects the resulting evidence with an explanation whose scope and uncertainty can be examined by others.

Appendix A. Selected chronology

Dates identify the particular records, publications, or capabilities listed. They do not imply that each practice began at that moment. Where an event and its publication have different dates, both are shown when the distinction is relevant.

Table 4. Selected chronology and diagnostic contributions.

Date	Development	Diagnostic contribution
323–322 BCE	Surviving Babylonian astronomical diary [2]	A dated record preserves astronomical observations
Eleventh century	Ibn al-Haytham's optical investigations [4]	Geometry, light, and visual perception are examined together
1054	Chinese observation of the event associated with the Crab Nebula [3]	A historical record constrains a much later study of a remnant
1610	Galileo publishes Sidereus nuncius [12]	Repeated telescopic observations support comparison of celestial structure and motion
1660	Boyle publishes air-pump experiments [5]	Controlled pressure changes expose dependencies of physical effects
1665	Hooke publishes Micrographia [6]	Illustrated microscopic observations become shareable evidence
1669	Steno publishes his Prodromus [39]	Spatial relationships constrain sequences of geological formation
1788	Hutton's Theory of the Earth appears in print [40]	Present structures are related to processes over geological time
1789–1790	Lavoisier's chemistry text and English translation [26]	Quantitative accounting and gas investigation organize chemical explanations
1815	Smith publishes his geological map [41]	Correlated observations form a regional representation of strata
1859	Darwin publishes On the Origin of Species [72]	Multiple evidence classes support a historical and mechanistic biological explanation
1860	Kirchhoff and Bunsen publish spectral chemical analysis [15]	Spectral features become discriminating evidence for composition
1904	Bjerknes formulates a physical prediction problem for weather [48]	Initial-state evidence becomes an explicit requirement of prediction
1909	Geiger and Marsden report alpha-particle reflection [17]	Rare large deflections constrain models of matter
1911	Wilson publishes visible ionization tracks [18]	A transient passage leaves a spatial artifact for examination
1912	Leavitt's variable-star results appear in Harvard Circular 173 [13]	A relation among periods and brightness supports astronomical comparison
1913	The Braggs publish X-ray reflection studies [33]	Diffraction becomes evidence for internal crystal organization
1913	Michaelis and Menten investigate invertase kinetics [37]	A controlled rate response is connected with an explanatory model
1925	Payne publishes Stellar Atmospheres [16]	Ionization physics changes the interpretation of stellar spectra

Date	Development	Diagnostic contribution
1928	Raman and Krishnan report frequency-shifted scattering [30]	A new molecular response becomes available for spectroscopy
1935	Fisher publishes The Design of Experiments [7]	Randomization and replication organize experimental comparisons
1936	Lehmann publishes P’ [44]	Seismic arrivals support inference of an inner core
1941	Martin and Synge publish partition chromatography [27]	Differential partitioning separates mixture components
1943	Luria and Delbrück publish bacterial mutation experiments [73]	Variation among cultures discriminates between accounts of resistance
1944	Avery, MacLeod, and McCarty publish transformation experiments [65]	Biochemical separation links activity to a DNA fraction
1946	Early NMR absorption and induction reports [31, 32]	Nuclear magnetic behavior becomes measurable in bulk matter
1949	Arnold and Libby report radiocarbon age checks [42]	Known-age material tests a method of reconstructing time
1950	Charney, Fjörtoft, and von Neumann publish a numerical weather calculation [49]	Computation implements a simplified dynamical forecast
1952	James and Martin publish gas-liquid chromatography [28]	Gas-phase separation extends analytical capability
1952	Hodgkin and Huxley publish a quantitative nerve model [69]	Controlled electrical measurements constrain a physiological mechanism
1953	DNA diffraction evidence and the double-helix model are published [34, 35]	Structural models are connected with molecular measurements
1958	Meselson and Stahl publish the DNA-replication experiment [66]	Labeling and separation distinguish replication mechanisms
1960	Kalman's control-system formulation and Keeling's atmospheric report [1, 52]	State inference and sustained environmental measurement develop in different settings
1963	Lorenz publishes deterministic nonperiodic flow [50]	Dynamical sensitivity becomes a limit relevant to prediction
1963	Vine and Matthews interpret ocean-ridge magnetic anomalies [45]	A spatial pattern supports a history of crustal formation and motion
1965	Excess antenna-temperature and cosmic-radiation papers appear [19, 20]	An observational residual is connected with cosmological explanation
1966	Paine publishes food-web removal experiments [74]	Field intervention investigates community dependence
1967–1968	Pulsars are detected and a rapidly pulsating source is reported [21]	Timing and repetition become evidence for a new source class
1968	Currie and Backus–Gilbert publish methodological analyses [10, 11]	Detection decisions and inverse resolution receive explicit treatment
1972	Landsat 1 is launched [59]	Repeated satellite land observation begins a long comparative record
1974	Molina and Rowland publish their CFC–ozone mechanism [55]	Chemical reasoning proposes a testable atmospheric process

Date	Development	Diagnostic contribution
1977	Sanger, Nicklen, and Coulson publish chain-termination sequencing [67]	Molecular sequence becomes an ordered analytical artifact
1984	Quasicrystal evidence and the pseudoreplication analysis are published [36, 75]	Structural categories and statistical units receive critical examination
1985	Antarctic ozone losses and an early PCR application are published [56, 68]	Long-term monitoring and molecular amplification extend different diagnostic capabilities
1989	Fenn and colleagues publish electrospray analysis of biomolecules [29]	Mass spectrometry gains access to large, fragile molecules
1994	GFP is demonstrated as a gene-expression marker [63]	Selected biological components can generate observable signals
1998	Fire and colleagues publish RNA interference findings [70]	A molecular perturbation becomes available for functional investigation
1999	Vostok climate history and the beginning of regional Argo arrays [54, 58]	Natural archives and distributed instrumentation extend environmental records
2002	Occupancy estimation with imperfect detection is published [76]	Presence is distinguished statistically from observation
2006	Photoactivated localization microscopy is reported [64]	Many localizations are composed into a structural image
2008–2009	Environmental DNA, MIQE, and robot-scientist publications [77, 89, 90]	Trace material, reporting detail, and automated experimentation receive new methods
2012	Higgs-boson discovery papers and programmable Cas9 biochemistry [22, 23, 71]	Combined event evidence and targeted molecular intervention advance separately
2013–2014	Reproducible-workflow guidance and microbiome-contamination evidence [82, 78]	Computation and sample preparation become explicit diagnostic targets
2015–2016	GW150914 is observed and reported [24]	Correlated interferometer observations support direct gravitational-wave detection
2016	FAIR principles and the Gaia mission description [83, 14]	Data stewardship and integrated catalog production become explicit infrastructures
2017	GW170817 is observed with multiple messengers [25]	Different physical evidence channels are related to one event
2019–2020	Reproducibility report, metagenomic bias study, ERA5, and IntCal20 [84, 79, 51, 43]	Interpretation includes workflow, measurement bias, reconstruction, and calibration
2021	AlphaFold results and InSight crustal constraints [91, 60]	Model-derived structure and sparse planetary measurements expand inference
2023	Analysis of leakage in machine-learning-based science [92]	Evaluation design becomes an explicit object of scientific diagnosis

Appendix B. Glossary

The definitions below describe usage in this history. Some disciplines use the same term more narrowly. The metrological terms should also be read with the International Vocabulary of Metrology and the GUM. [8, 9]

Acquisition. The process by which an observation is selected, captured, and recorded. Its settings help determine what can appear in the evidence.

Adjoint Thread of Activity. In our trace-analysis terminology, a selected sequence organized by an attribute other than the ordinary execution-thread identifier. Scientific analogies require an explicit mapping of that attribute.

Anomaly. A departure from an expectation or reference. It may belong to the natural process, the observation process, or the model used for comparison.

Artifact. A preserved observation or derived representation used in analysis. The word can also mean a spurious effect; the intended meaning must be clear from context.

Attribution. An assessment of the causes or contributions responsible for an observed change, made through comparison of explanations and evidence.

Background. Contributions present alongside a target signal. What counts as background depends on the scientific question.

Baseline. A stated reference condition or record against which change is assessed. A baseline has its own measurement and selection history.

Bias. A systematic tendency introduced by measurement, sampling, estimation, or analysis. Repetition alone need not reveal or remove it.

Blank. A control intended to examine contributions from a specified part of preparation or measurement without the target sample material.

Blind analysis. A procedure that withholds selected result information while analysis choices are developed, reducing dependence on an anticipated answer.

Calibration. Establishing the relationship between reference values and instrument indications under specified conditions, including relevant uncertainty.

Causal mechanism. A proposed process through which conditions and interactions produce an outcome. Evidence must establish its applicability to the case.

Chromatogram. A recorded response from a chromatographic separation, commonly displayed against time or distance.

Confounding. Mixing of influences such that the available comparison does not separate the effect of interest from another cause.

Control. A comparison or procedure designed to address a specified alternative explanation or verify a specified capability.

Data assimilation. Combining observations with a dynamical model to estimate a system state, accounting for assumptions about both.

Debugging. Investigation and correction of unintended behavior in an experimental, instrumental, computational, or analytical arrangement.

Detection. A decision that evidence for a specified signal or condition is present under a stated procedure and error framework.

Detection probability. The probability of recording a target when it is present under specified survey conditions and assumptions.

Diagnostics. Interpretation of evidence to distinguish states, processes, mechanisms, or problems, including assessment of competing explanations.

Drift. A change in response over time that may affect the relation between an instrument indication and the quantity being measured.

Evidence. An observation or artifact considered in relation to a claim. Data become evidence through this stated relationship.

Experimental unit. The unit to which a treatment is independently assigned or applied. It is not necessarily the unit represented by each measurement row.

FAIR. Findable, accessible, interoperable, and reusable. A set of principles for data and metadata stewardship, not a certificate of scientific correctness.

Feature. A selected or computed property of an artifact, such as a spectral peak, arrival time, texture, or sequence motif.

Foreground. A nearer or intervening contribution that must be distinguished from a target, particularly in astronomical observation.

Forward model. A model that predicts observable consequences from a specified state, structure, or set of parameters.

Identifiability. The ability to distinguish model parameters or structures from the observations under stated assumptions.

Independent check. A check whose relevant possible errors do not simply reproduce those of the original result. Independence is specific to a suspected failure mechanism.

Intervention. A deliberate change used to examine the response of a system. Its additional effects are part of the interpretation.

Inverse problem. Inference of a state, structure, or process from observed effects through a model of how those effects arise.

Measurand. The quantity intended to be measured. Its definition includes the object and conditions needed to make the intended quantity clear.

Measurement model. The relationship used to connect observations and other quantities with the reported measurement result.

Monitoring. Repeated observation of selected conditions for awareness of change, often using defined baselines or decision rules.

Natural archive. Material or structure that preserves information about earlier conditions, such as an ice core, sediment, fossil, or mineral.

Negative result. A result that does not support detection or the expected effect under the stated procedure. Its meaning depends on sensitivity and coverage.

Observability. In its formal use, distinguishability of internal state from outputs under a model. Broader scientific usage concerns the evidence available about otherwise inaccessible processes.

Observation model. The relationship between a natural system and what a particular observing arrangement is expected to record.

Occupancy. Presence of a species or other target at a site during a defined period, distinguished from whether a survey detects it.

Pattern. A recurrent organization or relationship recognized in artifacts. It supports analysis but does not uniquely establish a mechanism.

Perturbation. A change introduced to examine a system's response. Small and large perturbations can expose different behavior.

Precision. Closeness among repeated indications or results under specified conditions. High precision can coexist with systematic error.

Provenance. The recorded origin and transformation history of an artifact, including relations to source material, methods, and versions.

Proxy. An observation used to infer another quantity through a relationship that requires calibration or a process model.

Pseudoreplication. Treatment of measurements as independent experimental replication when the design does not provide that independence.

Quality flag. Metadata indicating an assessed condition of a measurement. Its definition and processing version determine what it means.

Reanalysis. A reconstruction using observations and a consistent modeling or assimilation framework across a historical interval.

Regularization. Additional restrictions or preferences used to stabilize or select an inverse solution. Their influence must be included in interpretation.

Replicability. Here, consistency across studies that obtain their own data to address the same scientific question, following the 2019 National Academies convention. [84]

Reproducibility. Here, obtaining consistent computational results with the same inputs and computational steps. It does not alone validate the scientific model. [84]

Residual. A difference between an observation and a model prediction or fitted value. Its structure may reveal an omitted process or an error elsewhere.

Resolution. The scale or degree of separation at which an observation and inference arrangement can distinguish features.

Sampling. Selection of locations, times, objects, or events for observation. The selection process constrains generalization.

Sensitivity analysis. Examination of how a result changes with inputs, assumptions, parameters, or analysis choices.

Signature. A characteristic response used to recognize or distinguish a class of observations under specified conditions.

Traceability. In metrology, a documented calibration chain connecting a result to a reference, with uncertainty contributions. It differs from general workflow provenance.

Trigger. A rule or device that determines when an acquisition occurs or which events are retained.

Uncertainty. A characterization of the dispersion or range of values that can reasonably be attributed under a stated measurement or inferential framework.

Validation. Assessment of adequacy for an intended scientific use through comparison with relevant evidence.

Verification. Assessment that a calculation, implementation, or procedure satisfies its specified formulation; its precise scope should be stated.

Visibility limit. A boundary on what an acquisition or analysis arrangement can reveal, arising from sensitivity, resolution, coverage, or representation.

Appendix C. Evidence taxonomy and blind spots

The same artifact may serve several diagnostic questions. This taxonomy groups evidence by how it is obtained and what relationships it can support. It is an analytical synthesis of the methods discussed in the chapters. A listed blind spot is a question to investigate, not a claim that the evidence type is generally unreliable.

Table 5. Evidence classes, their uses, and common blind spots.

Evidence class	Examples	What it can constrain	Common blind spots
Direct descriptive records	Field notes, sketches, observing diaries	Appearance, sequence, context recognized by the observer	Selective recording, transcription, calendar conversion, missing observations
Physical specimens	Rocks, biological samples, preserved material	Composition, structure, material relationships	Collection bias, alteration, contamination, uncertain identity
Images	Micrographs, photographs, satellite scenes	Spatial organization and change under an imaging model	Preparation, projection, resolution, illumination, reconstruction
Spectra	Atomic, molecular, NMR, mass spectra	Composition, state, motion, local environment	Overlapping responses, matrix effects, calibration, nonunique assignments
Separation records	Chromatograms, gels, density bands	Distinctions among components or labeled populations	Co-elution, incomplete recovery, altered material, indirect identification
Event and track records	Particle tracks, detector events, transient alerts	Interaction geometry, timing, selected event populations	Trigger selection, dead time, efficiency, reconstruction assumptions
Continuous instrument series	Temperature, pressure, gas concentration, light curves	Trends, cycles, transients, associations	Drift, gaps, aliasing, site changes, shared environmental effects
Waveform records	Seismograms, gravitational-wave strain, electrical traces	Propagation, response, source and system behavior	Timing errors, overlapping sources, instrument response, path ambiguity
Natural archives	Ice, sediments, fossils, tree rings, minerals	Past conditions and sequences	Preservation, reworking, age models, proxy interpretation, uneven time resolution
Molecular sequences	Reads, assemblies, marker sequences	Identity, variation, relatedness, selected composition	Contamination, amplification bias, reference bias, assembly errors
Survey records	Presence records, counts, repeated visits	Distribution, occurrence, population comparisons	Imperfect detection, uneven effort, changing definitions, dependence
Intervention results	Removal experiments, perturbations, rescue experiments	Responses to controlled changes and candidate causal relationships	Additional effects, compensation, insufficient controls, limited transferability

Evidence class	Examples	What it can constrain	Common blind spots
Calibration and control evidence	Standards, blanks, reference materials, control runs	Performance of defined parts of the observation procedure	Controls not representative of samples, untested conditions, shared errors
Instrument health records	Environmental channels, resets, service and configuration logs	Changes in the observation arrangement	Missing metadata, clock mismatch, poorly documented maintenance
Computational artifacts	Models, code, parameters, residuals, intermediate arrays	Consequences of a formulation and its implementation	Numerical error, version mismatch, inappropriate assumptions, hidden transformations
Derived products	Catalogs, maps, classifications, reconstructed structures	Organized summaries and model-assisted estimates	Lost provenance, compressed uncertainty, selection rules, reprocessing differences
Cross-method comparisons	Several detectors, physical methods, laboratories, messengers	Agreement across distinct evidential dependencies	Hidden common inputs, common models, correlated errors, mismatched quantities
Negative and rejected evidence	Non-detections, failed runs, excluded events	Limits, acquisition behavior, analysis selection	Unrecorded exclusions, inadequate sensitivity, missing denominators

C1. A practical reading of an artifact

First identify what the artifact actually records. A detector voltage, a count, a classified sequence, and a reconstructed image occupy different positions in the evidence chain. The label on a file may describe the intended interpretation rather than its direct content.

Next recover acquisition context. We need the object or sample identity, location, time basis, instrument settings, preparation, calibration, and relevant environmental conditions. The exact list depends on the claim. More metadata are useful only when their meanings and relationships are preserved.

Then examine transformations. Filtering, background subtraction, normalization, alignment, reconstruction, and classification can all change the visible pattern. We should be able to identify the method and version used and, when necessary, return to an earlier representation.

Finally state the supported distinction. The artifact may distinguish presence from absence, one mechanism from another, or an instrumental explanation from a natural one. If it does not distinguish the alternatives of interest, additional precision in presenting the same artifact will not resolve the problem.

C2. Combining evidence without losing its differences

A collection becomes more informative when relationships among its artifacts are preserved. Common sample identifiers, synchronized time bases, spatial coordinates, and documented derivations provide different kinds of connection. None should be substituted for another without justification.

For example, two measurements made at the same time are not necessarily measurements of the same material. Two files with the same sample name may contain different preparations. Two images with similar structure may have different scales or orientations. The join between records can itself be the location of a diagnostic error.

We can therefore examine both the artifacts and the relations used to combine them. A useful evidence set retains enough information to move between the combined view and the contributing observations. This allows disagreement to be localized and supports further investigation when a new hypothesis changes which relationships matter.

References

References are numbered in order of first citation. Original publication dates are retained; translations, reprints, and archival records are identified where used. Online resources were consulted in September 2026.

[1] Kalman, R. E. (1960). On the general theory of control systems. IFAC Proceedings Volumes, 1(1), 491–502. https://doi.org/10.1016/S1474-6670(17)70094-8

[2] British Museum. Babylonian astronomical diary recording observations for 323–322 BCE. Collection object 1881,0706.403. https://www.britishmuseum.org/collection/object/W_1881-0706-403

[3] NSF NOIRLab. Chinese Astronomers Observe Supernova (video and caption). See also NASA, Time Travel: Observing Cosmic History, on the Chinese records of 1054 and the Crab Nebula.
https://noirlab.edu/public/videos/Chinese-Supernova-CC/
https://science.nasa.gov/mission/hubble/science/science-behind-the-discoveries/time-travel-observing-cosmic-history/

[4] Ibn al-Haytham. Alhacen's Theory of Visual Perception. A. Mark Smith, editor and translator (2001), volume 2, English translation. Transactions of the American Philosophical Society, 91(5). https://www.pennpress.org/9780871699152/alhacens-theory-of-visual-perception-first-three-books-of-alhacens-de-aspectibus-volume-two-english-translation/

[5] Boyle, R. (1660). New Experiments Physico-Mechanicall, Touching the Spring of the Air, and Its Effects. Oxford: H. Hall for T. Robinson. Digitized title pages and apparatus engraving, Science History Institute. https://digital.sciencehistory.org/works/j3860811k

[6] Hooke, R. (1665). Micrographia: Or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses. London: J. Martyn and J. Allestry. Royal Society digitization and collection account. https://royalsociety.org/blog/2020/07/micrographia-online/

[7] Fisher, R. A. (1935). The Design of Experiments. Edinburgh: Oliver and Boyd. https://books.google.com/books/about/The_Design_of_Experiments.html?id=-EsNAQAAIAAJ

[8] Joint Committee for Guides in Metrology (2012). International Vocabulary of Metrology—Basic and General Concepts and Associated Terms. Third edition, 2008 version with minor corrections. JCGM 200:2012. https://doi.org/10.59161/JCGM200-2012

[9] Joint Committee for Guides in Metrology (2008). Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement. JCGM 100:2008; GUM 1995 with minor corrections. https://www.bipm.org/documents/20126/2071204/JCGM1002008E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6

[10] Currie, L. A. (1968). Limits for qualitative detection and quantitative determination. Application to radiochemistry. Analytical Chemistry, 40(3), 586–593. https://doi.org/10.1021/ac60259a007

[11] Backus, G., and Gilbert, F. (1968). The resolving power of gross Earth data. Geophysical Journal of the Royal Astronomical Society, 16(2), 169–205. https://doi.org/10.1111/j.1365-246X.1968.tb00216.x

[12] Galilei, G. (1610). Sidereus nuncius. Venice: Thomas Baglioni. Digitized original, Smithsonian Libraries. https://library.si.edu/digital-library/book/sidereusnuncius00gali

[13] Leavitt, H. S., and Pickering, E. C. (1912). Periods of 25 variable stars in the Small Magellanic Cloud. Harvard College Observatory Circular, 173, 1–3. Leavitt's results communicated by Pickering. https://articles.adsabs.harvard.edu/pdf/1912HarCi.173....1L

[14] Gaia Collaboration, Prusti, T., et al. (2016). The Gaia mission. Astronomy & Astrophysics, 595, A1. https://doi.org/10.1051/0004-6361/201629272

[15] Kirchhoff, G., and Bunsen, R. (1860). Chemische Analyse durch Spectralbeobachtungen. Annalen der Physik und Chemie, 110 (second series), 161–189.
https://doi.org/10.1002/andp.18601860602

[16] Payne, C. H. (1925). Stellar Atmospheres; A Contribution to the Observational Study of High Temperature in the Reversing Layers of Stars. Harvard Observatory Monographs, 1. Cambridge, Massachusetts: Harvard Observatory.
https://articles.adsabs.harvard.edu/pdf/1925HarMo...1.....P

[17] Geiger, H., and Marsden, E. (1909). On a diffuse reflection of the α-particles. Proceedings of the Royal Society of London A, 82, 495–500.
https://doi.org/10.1098/rspa.1909.0054

[18] Wilson, C. T. R. (1911). On a method of making visible the paths of ionising particles through a gas. Proceedings of the Royal Society of London A, 85, 285–288.
https://doi.org/10.1098/rspa.1911.0041

[19] Penzias, A. A., and Wilson, R. W. (1965). A measurement of excess antenna temperature at 4080 Mc/s. The Astrophysical Journal, 142, 419–421.
https://doi.org/10.1086/148307

[20] Dicke, R. H., Peebles, P. J. E., Roll, P. G., and Wilkinson, D. T. (1965). Cosmic black-body radiation. The Astrophysical Journal, 142, 414–419.
https://doi.org/10.1086/148306

[21] Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., and Collins, R. A. (1968). Observation of a rapidly pulsating radio source. Nature, 217, 709–713.
https://doi.org/10.1038/217709a0

[22] ATLAS Collaboration, Aad, G., et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Physics Letters B, 716, 1–29.
https://doi.org/10.1016/j.physletb.2012.08.020
https://arxiv.org/abs/1207.7214

[23] CMS Collaboration, Chatrchyan, S., et al. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Physics Letters B, 716, 30–61.
https://doi.org/10.1016/j.physletb.2012.08.021
https://arxiv.org/abs/1207.7235

[24] LIGO Scientific Collaboration and Virgo Collaboration, Abbott, B. P., et al. (2016). Observation of gravitational waves from a binary black hole merger. Physical Review Letters, 116, 061102.
https://doi.org/10.1103/PhysRevLett.116.061102
https://arxiv.org/abs/1602.03837

[25] Abbott, B. P., et al. (LIGO Scientific Collaboration, Virgo Collaboration, and partner collaborations) (2017). Multi-messenger observations of a binary neutron star merger. The Astrophysical Journal Letters, 848(2), L12.
https://doi.org/10.3847/2041-8213/aa91c9
https://arxiv.org/abs/1710.05833

[26] Lavoisier, A.-L. (1790). Elements of Chemistry, in a New Systematic Order, Containing All the Modern Discoveries. Translated by R. Kerr from the 1789 French work. Edinburgh: William Creech.
https://www.gutenberg.org/ebooks/30775

[27] Martin, A. J. P., and Synge, R. L. M. (1941). A new form of chromatogram employing two liquid phases. 1. A theory of chromatography. 2. Application to the micro-determination of the higher monoamino-acids in proteins. Biochemical Journal, 35(12), 1358–1368.
https://doi.org/10.1042/bj0351358

[28] James, A. T., and Martin, A. J. P. (1952). Gas-liquid partition chromatography: The separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid. Biochemical Journal, 50, 679–690.
https://doi.org/10.1042/bj0500679

[29] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science, 246, 64–71.
https://doi.org/10.1126/science.2675315

[30] Raman, C. V., and Krishnan, K. S. (1928). A new type of secondary radiation. Nature, 121, 501–502.
https://doi.org/10.1038/121501c0

[31] Purcell, E. M., Torrey, H. C., and Pound, R. V. (1946). Resonance absorption by nuclear magnetic moments in a solid. Physical Review, 69, 37–38.
https://doi.org/10.1103/PhysRev.69.37

[32] Bloch, F., Hansen, W. W., and Packard, M. (1946). Nuclear induction. Physical Review, 69, 127.
https://doi.org/10.1103/PhysRev.69.127

[33] Bragg, W. H., and Bragg, W. L. (1913). The reflection of X-rays by crystals. Proceedings of the Royal Society of London A, 88, 428–438.
https://doi.org/10.1098/rspa.1913.0040

[34] Franklin, R. E., and Gosling, R. G. (1953). Molecular configuration in sodium thymonucleate. Nature, 171, 740–741.
https://doi.org/10.1038/171740a0

[35] Watson, J. D., and Crick, F. H. C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. Nature, 171, 737–738.
https://doi.org/10.1038/171737a0

[36] Shechtman, D., Blech, I., Gratias, D., and Cahn, J. W. (1984). Metallic phase with long-range orientational order and no translational symmetry. Physical Review Letters, 53, 1951–1953.
https://doi.org/10.1103/PhysRevLett.53.1951

[37] Johnson, K. A., and Goody, R. S. (2011). The original Michaelis constant: Translation of the 1913 Michaelis–Menten paper. Biochemistry, 50, 8264–8269. Includes the English translation as supplementary material.
https://doi.org/10.1021/bi201284u

[38] Zewail, A. H. (1999). Femtochemistry: Atomic-Scale Dynamics of the Chemical Bond Using Ultrafast Lasers. Nobel Lecture, 8 December 1999.
https://www.nobelprize.org/prizes/chemistry/1999/zewail/lecture/
https://www.nobelprize.org/uploads/2018/06/zewail-lecture.pdf

[39] Steno, N. (1669; English translation 1671). The Prodromus to a Dissertation Concerning Solids Naturally Contained within Solids. Translated by H. Oldenburg. London. Digitized original, Smithsonian Libraries.
https://library.si.edu/digital-library/book/prodromustodisse00sten

[40] Hutton, J. (1788). Theory of the Earth; or an investigation of the laws observable in the composition, dissolution, and restoration of land upon the globe. Transactions of the Royal Society of Edinburgh, 1, 209–304. Archival account by the Royal Society of Edinburgh.
https://doi.org/10.1017/S0080456800029227
https://rse.org.uk/about-us/heritage-collections/archive/the-legacy-of-james-hutton/

[41] Smith, W. (1815). A Delineation of the Strata of England and Wales, with Part of Scotland. London: John Cary. Geological Society of London collection and exhibition.
https://www.geolsoc.org.uk/the-library/online-exhibitions/william-strata-smith/the-map/a-delineation-of-the-strata-of-england-and-wales-with-part-of-scotland/

[42] Arnold, J. R., and Libby, W. F. (1949). Age determinations by radiocarbon content: Checks with samples of known age. Science, 110(2869), 678–680.
https://doi.org/10.1126/science.110.2869.678

[43] Reimer, P. J., et al. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon, 62, 725–757.
https://doi.org/10.1017/RDC.2020.41

[44] Lehmann, I. (1936). P′. Publications du Bureau Central Séismologique International, Série A, Travaux Scientifiques, 14, 87–115. Extracts reproduced with commentary by M. Kölbl-Ebert (2001), Episodes, 24(4), 262–267.
Commentary DOI: https://doi.org/10.18814/epiiugs/2001/v24i4/007
https://courses.seas.harvard.edu/climate/eli/Courses/EPS281r/Sources/Inner-Core/Lehmann-1936-extracts%2Binterpretation.pdf

[45] Vine, F. J., and Matthews, D. H. (1963). Magnetic anomalies over oceanic ridges. Nature, 199, 947–949.
https://doi.org/10.1038/199947a0

[46] Kious, W. J., and Tilling, R. I. (1996). This Dynamic Earth: The Story of Plate Tectonics. U.S. Geological Survey.
https://doi.org/10.3133/7000097

[47] Flinders, A. F., Lowenstern, J. B., Coombs, M. L., and Poland, M. P., eds. (2024). Recommended Capabilities and Instrumentation for Volcano Monitoring in the United States. U.S. Geological Survey Scientific Investigations Report 2024–5062.
https://doi.org/10.3133/sir20245062

[48] Bjerknes, V. (1904). Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. Meteorologische Zeitschrift, 21, 1–7. English translation by E. Volken and S. Brönnimann (2009), The problem of weather prediction, considered from the viewpoints of mechanics and physics. Meteorologische Zeitschrift, 18(6), 663–667.
https://doi.org/10.1127/0941-2948/2009/416

[49] Charney, J. G., Fjörtoft, R., and von Neumann, J. (1950). Numerical integration of the barotropic vorticity equation. Tellus, 2(4), 237–254.
https://doi.org/10.1111/j.2153-3490.1950.tb00336.x

[50] Lorenz, E. N. (1963). Deterministic nonperiodic flow. Journal of the Atmospheric Sciences, 20, 130–141.
https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

[51] Hersbach, H., et al. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049.
https://doi.org/10.1002/qj.3803

[52] Keeling, C. D. (1960). The concentration and isotopic abundances of carbon dioxide in the atmosphere. Tellus, 12(2), 200–203.
https://doi.org/10.1111/j.2153-3490.1960.tb01300.x

[53] Manabe, S., and Wetherald, R. T. (1967). Thermal equilibrium of the atmosphere with a given distribution of relative humidity. Journal of the Atmospheric Sciences, 24, 241–259.
https://doi.org/10.1175/1520-0469(1967)024<0241:TEOTAW>2.0.CO;2

[54] Petit, J. R., et al. (1999). Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. Nature, 399, 429–436.
https://doi.org/10.1038/20859

[55] Molina, M. J., and Rowland, F. S. (1974). Stratospheric sink for chlorofluoromethanes: Chlorine atom-catalysed destruction of ozone. Nature, 249, 810–812.
https://doi.org/10.1038/249810a0

[56] Farman, J. C., Gardiner, B. G., and Shanklin, J. D. (1985). Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. Nature, 315, 207–210.
https://doi.org/10.1038/315207a0

[57] IPCC (2021). Climate Change 2021: The Physical Science Basis. Chapter 3, Human Influence on the Climate System. Working Group I contribution to the Sixth Assessment Report.
https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-3/

[58] Roemmich, D., et al. (2009). The Argo Program: Observing the global ocean with profiling floats. Oceanography, 22(2), 34–43.
https://doi.org/10.5670/oceanog.2009.36

[59] U.S. Geological Survey. Landsat 1. Mission history and instrument record.
https://www.usgs.gov/landsat-missions/landsat-1

[60] Knapmeyer-Endrun, B., et al. (2021). Thickness and structure of the Martian crust from InSight seismic data. Science, 373, 438–443.
https://doi.org/10.1126/science.abf8966

[61] Ferlier, L. (2023, 10 September). Focus on Leeuwenhoek. The Royal Society blog.
https://www.royalsociety.org/blog/2023/09/focus-on-leeuwenhoek/

[62] Ruska, E. (1986). The Development of the Electron Microscope and of Electron Microscopy. Nobel Lecture, 8 December 1986.
https://www.nobelprize.org/prizes/physics/1986/ruska/lecture/

[63] Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. Science, 263, 802–805.
https://doi.org/10.1126/science.8303295

[64] Betzig, E., et al. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. Science, 313, 1642–1645.
https://doi.org/10.1126/science.1127344

[65] Avery, O. T., MacLeod, C. M., and McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. Journal of Experimental Medicine, 79, 137–158.
https://doi.org/10.1084/jem.79.2.137

[66] Meselson, M., and Stahl, F. W. (1958). The replication of DNA in Escherichia coli. Proceedings of the National Academy of Sciences, 44, 671–682.
https://doi.org/10.1073/pnas.44.7.671

[67] Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences, 74, 5463–5467.
https://doi.org/10.1073/pnas.74.12.5463

[68] Saiki, R. K., et al. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science, 230, 1350–1354.
https://doi.org/10.1126/science.2999980

[69] Hodgkin, A. L., and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of Physiology, 117, 500–544.
https://doi.org/10.1113/jphysiol.1952.sp004764

[70] Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature, 391, 806–811.
https://doi.org/10.1038/35888

[71] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 337, 816–821.
https://doi.org/10.1126/science.1225829

[72] Darwin, C. (1859). On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. First edition. London: John Murray.
https://darwin-online.org.uk/converted/published/1859_Origin_F373/1859_Origin_F373.html

[73] Luria, S. E., and Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. Genetics, 28, 491–511.
https://doi.org/10.1093/genetics/28.6.491

[74] Paine, R. T. (1966). Food web complexity and species diversity. The American Naturalist, 100, 65–75.
https://doi.org/10.1086/282400

[75] Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. Ecological Monographs, 54, 187–211.
https://doi.org/10.2307/1942661

[76] MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. Ecology, 83, 2248–2255.
https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

[77] Ficetola, G. F., Miaud, C., Pompanon, F., and Taberlet, P. (2008). Species detection using environmental DNA from water samples. Biology Letters, 4, 423–425.
https://doi.org/10.1098/rsbl.2008.0118

[78] Salter, S. J., et al. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology, 12, 87.
https://doi.org/10.1186/s12915-014-0087-z

[79] McLaren, M. R., Willis, A. D., and Callahan, B. J. (2019). Consistent and correctable bias in metagenomic sequencing experiments. eLife, 8, e46923.
https://doi.org/10.7554/eLife.46923

[80] Oberkampf, W. L., and Trucano, T. G. (2002). Verification and validation in computational fluid dynamics. Progress in Aerospace Sciences, 38, 209–272.
https://doi.org/10.1016/S0376-0421(02)00005-2

[81] Goldberg, D. (1991). What every computer scientist should know about floating-point arithmetic. ACM Computing Surveys, 23(1), 5–48. Edited reprint in the Numerical Computation Guide.
https://doi.org/10.1145/103162.103163
https://docs.oracle.com/cd/E19957-01/806-3568/ncg_goldberg.html

[82] Sandve, G. K., Nekrutenko, A., Taylor, J., and Hovig, E. (2013). Ten simple rules for reproducible computational research. PLOS Computational Biology, 9, e1003285.
https://doi.org/10.1371/journal.pcbi.1003285

[83] Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018.
https://doi.org/10.1038/sdata.2016.18

[84] 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

[85] Zeeberg, B. R., et al. (2004). Mistaken identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics. BMC Bioinformatics, 5, 80.
https://doi.org/10.1186/1471-2105-5-80

[86] LIGO Scientific Collaboration and Virgo Collaboration, Abbott, B. P., et al. (2016). Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. Classical and Quantum Gravity, 33, 134001.
https://doi.org/10.1088/0264-9381/33/13/134001
https://arxiv.org/abs/1602.03844

[87] CERN (2011; updates February and June 2012). OPERA experiment reports anomaly in flight time of neutrinos from CERN to Gran Sasso. Includes the investigation and corrected time-of-flight results.
https://home.cern/opera-experiment-reports-anomaly-in-flight-time-of-neutrinos-from-cern-to-gran-sasso/

[88] Klein, J. R., and Roodman, A. (2005). Blind analysis in nuclear and particle physics. Annual Review of Nuclear and Particle Science, 55, 141–163.
https://doi.org/10.1146/annurev.nucl.55.090704.151521

[89] Bustin, S. A., et al. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry, 55, 611–622.
https://doi.org/10.1373/clinchem.2008.112797

[90] King, R. D., et al. (2009). The automation of science. Science, 324, 85–89.
https://doi.org/10.1126/science.1165620

[91] Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583–589.
https://doi.org/10.1038/s41586-021-03819-2

[92] Kapoor, S., and Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns, 4, 100804.
https://doi.org/10.1016/j.patter.2023.100804

[93] OPERA Collaboration, Adam, T., et al. (2013). Measurement of the neutrino velocity with the OPERA detector in the CNGS beam using the 2012 dedicated data. Journal of High Energy Physics, 2013(1), 153.
https://doi.org/10.1007/JHEP01(2013)153
https://arxiv.org/abs/1212.1276

[94] BICEP2/Keck and Planck Collaborations, Ade, P. A. R., et al. (2015). Joint analysis of BICEP2/Keck Array and Planck data. Physical Review Letters, 114, 101301.
https://doi.org/10.1103/PhysRevLett.114.101301
https://arxiv.org/abs/1502.00612

[95] Vostokov, D. (2024; paperback 2025). Theoretical Software Diagnostics: Collected Articles. Fourth edition, revision 4.01. OpenTask. ISBN 978-1-912636-91-4.
https://dumpanalysis.org/theoretical-software-diagnostics-book

[96] Vostokov, D. (2023; revision 5.06, August 2026). Trace, Log, Text, Narrative, Data: An Analysis Pattern Reference for Information Mining, Diagnostics, Anomaly Detection. Fifth edition. OpenTask. ISBN 978-1-912636-58-7.