

THE IN A HURRY SERIES

Science *in a Hurry*

How humans learned to actually know things

Max Brocklesby

BOOKSINAHURRY.COM

Contents

The Whole Thing in One Page 4

Why You Should Care 6

The Core Ideas 8

How It Actually Works 21

What People Get Wrong 32

Use It 37

Terms 42

Go Deeper 46

Notes and Sources 47

Bibliography 57

The Whole Thing in One Page

Science is usually pictured as a method used by unusually clever people. A question appears. A genius proposes a hypothesis, runs an experiment and discovers a fact. The result enters a textbook, where it sits polished and certain. Almost every part of that picture is wrong.

There was no morning on which humanity woke without science and went to bed with it. The system accumulated unevenly, through institutions built for calendars, trade, war, healing, worship, navigation and prestige. Practices now grouped together once belonged to different worlds and did not share a name.

Humans have always known useful things. Hunters read tracks, farmers learned seasons, healers recognised patterns, sailors read wind and stars. The harder achievement was making knowledge travel beyond the person who had it. Memory bends. Testimony borrows the speaker's authority. Conditions change without announcing themselves. A striking success can hide a hundred failures. Science grew by building ways for a claim to survive separation from its claimant.

Writing put observations outside the head. Long records let Babylonian astronomers find cycles no lifetime could contain. Number, geometry and causal argument made patterns sharper. Instruments extended the senses, but also made calibration necessary. Standards allowed measurements taken in different rooms and centuries to be compared. Experiments created contrasts that rival explanations could not answer equally well. Models joined scattered findings and risked predictions about what should happen next.

None of those devices worked alone. Galileo's telescope needed drawings, dates, repeat observations and readers able to obtain or trust an instrument. Kepler abandoned a circular orbit because Tycho Brahe's observations missed it by eight arcminutes. Robert Boyle's air pump became persuasive through witnessed performances and published reports. Germ theory became powerful when microscopes, cultures, staining, controlled infection

and clinical consequences began pointing together. The 1948 streptomycin trial showed how concealed random allocation and planned comparison could strengthen clinical evidence. In 2012 two immense collaborations using separate detectors and analyses reported the same new particle. In 2019 a planet-sized network of radio telescopes turned recorded signals into an image of the shadow around a black hole.

The schoolbook method is therefore too small. Astronomy cannot rerun a supernova. Geology cannot randomise continents. Evolutionary biology reconstructs processes from surviving traces. Clinical science may manipulate treatments but must protect patients. Each field develops tests suited to what it can observe, compare, disturb or predict. What unites them is organised resistance: claims must encounter records, measurements, alternatives and other people who are permitted to say no.

That last condition changes everything. Journals, laboratories, universities, standards bodies, archives, funding systems and research teams let knowledge outlive an individual. They also create exclusion, hierarchy, metric gaming and error at industrial scale. Science does not correct itself by magic. Correction requires access to methods and data, incentives to challenge, and institutions able to admit that an admired result failed.

Scientific knowledge is neither certainty nor opinion. It is the present survivor of a demanding, uneven and unfinished system for finding error. The system earns trust where its claims have been made public, comparable, vulnerable and repeatedly tested. Where those conditions weaken, the label science cannot rescue them.

That is the book.

Why You Should Care

Pick up a blister pack marked 500 mg paracetamol. The printed number looks like a small fact. You did not identify the molecule, weigh the dose, inspect the factory, audit the clinical trials or watch the regulator test a batch. You cannot. Yet swallowing the tablet need not be blind faith, because the claim on the foil sits at the end of a chain: chemical identification, calibrated balances, manufacturing controls, sampling, records, trials, inspections and penalties for deception. Any link can fail. The strength comes from the links being inspectable by people other than the seller.

Most of what governs a modern life has that shape. You cannot personally verify that a bridge will carry its stated load, that a scan reveals a tumour, that an aircraft wing tolerates fatigue, that a warming trend is global or that a distant galaxy contains a particular element. Direct experience covers almost none of the world on which your decisions depend. You live by testimony. The serious question is not whether to trust, but how trust can be earned without surrendering judgement.

Science is humanity's best developed answer. Its deepest achievement is not a warehouse of correct statements. It is a technology for sharing ignorance. A scientist can state how a result was produced, what was measured, what uncertainty remains, what rival account was considered and what would count against the conclusion. Other people can inspect the route rather than admire the destination. That is why a result can cross languages, governments and generations while remaining open to revision.

Seeing the machinery changes how you read claims. A graph stops being a picture and becomes the end of decisions about instruments, samples, exclusions and scales. A controlled trial stops being a ritual involving two groups and becomes a device for making alternative explanations less plausible. Peer review stops looking like certification and becomes one checkpoint before wider criticism. Consensus stops meaning that everyone voted the same way. It means that several lines of evidence, examined by a

relevant community, have converged strongly enough that dissent now carries a heavier explanatory burden.

It also changes what you ask of experts. Expertise matters because difficult work cannot be reconstructed from first principles every morning. Yet credentials alone do not settle a claim. The useful questions concern the evidence chain, the range over which the result travels, the quality of criticism and whether independent routes arrive at the same place. Trust can therefore be graduated. A finding from one small study deserves less than a result supported across methods, populations and institutions. An expert speaking inside a mature consensus deserves more weight than an expert selling a conclusion outside the relevant field. That is not cynicism. It is calibrated confidence.

There are limits. Evidence cannot choose every goal. A climate model can estimate consequences under stated assumptions; it cannot decide how costs should be shared. A clinical study can compare outcomes; it cannot tell a patient which trade-off matters most. Scientific work carries values in the choice of problems, categories, thresholds and acceptable risks. Pretending otherwise hides judgement rather than removing it.

Science has also been built through empire, exclusion, dangerous labour and military money. It has supplied vaccines and poison gas, sanitation and surveillance, crop yields and weapons. Reliability does not confer innocence. The same system that reveals what can be done cannot absolve anyone from deciding what should be done.

The reason to care is therefore larger than defending science or distrusting it. You need to know what kind of achievement a scientific claim represents, where its authority comes from, and what can break that authority. Once the machinery becomes visible, neither the lab coat nor the conspiracy theory looks quite as impressive.

The Core Ideas

Put Experience Outside the Head

A person can know how to find water, set a broken bone or predict a seasonal wind without possessing science in the modern sense. Skilled experience is real knowledge. Its weakness is transport. The learner must trust the practitioner, the circumstances are hard to reconstruct, failures fade faster than successes, and memory edits the sequence each time it is recalled. What works in one pair of hands may arrive in another as a story with the difficult parts removed.

The first great improvement was external memory. A clay tablet, a ship's log, a case record or a laboratory notebook can preserve an observation after the observer has gone. Babylonian scholars recorded planetary positions, eclipses, prices, river levels and events over generations. Their astronomical work could detect regularities longer than a human life because the institution remembered what no individual could. The record did more than store facts. It allowed observations to be ordered, compared and used to predict what would return.

Recording is not clerical work added after discovery. It changes what can be discovered. A repeated event becomes a pattern only when earlier instances remain available. A slow drift becomes visible only when observations share dates, units and conditions. An anomaly becomes more than a surprise when someone can show that it departs from a preserved sequence. The notebook, archive and database are therefore thinking devices.

They are also selective devices. No observation arrives labelled with everything future readers will need. Someone chooses what to count, which instrument setting to note, how to classify an ambiguous case and whether a failed run deserves space. A temperature without its scale, location, time and instrument can be precise ink and useless evidence. Modern researchers call the surrounding information metadata and the chain of custody provenance. The plain lesson is older: a number detached from

how it was made cannot carry much weight.

Protocols push the record one step further. They describe a procedure closely enough that another skilled person can attempt the same operation. This never removes tacit knowledge. A written recipe does not contain the feel of dough, and a methods section cannot convey every adjustment made at a laboratory bench. Yet explicit procedures expose more of the work to inspection. They turn disagreement from one person's word against another's into questions about materials, sequence, conditions and judgement.

The same move explains why data sharing matters and why it is never enough by itself. A spreadsheet without code, variable definitions, exclusions and collection history may be open while remaining unintelligible. Photographs can preserve appearances yet conceal lens, exposure and processing. Specimens can be re-examined, but only if labels and collection context survive. External memory becomes scientific evidence when it preserves a route back from claim to observation.

Science begins, then, with a form of modesty made material. Do not ask the world to rely on your memory or your sincerity. Leave something that can resist you. The strongest claim is one that another person can inspect after your authority has left the room.

Build a Measurement Chain

Nature does not contain little labels saying 12.4 centimetres, pH 6.8 or 37.2 degrees. Measurement is an organised comparison between an aspect of the world and a defined reference. That comparison runs through concepts, instruments, calibration, units and correction. The displayed number is the last link, not the observation in its pure form.

Take length. A ruler works because its marks are related, through a chain of calibrations, to a shared definition of the metre. The modern SI no longer depends on a treasured metal bar. Its units are defined through fixed values of physical constants and realised by laboratories using specified procedures. A workshop gauge can be checked against a reference, which

is checked against another reference, until the chain reaches a recognised standard. This traceability is what allows parts made in different countries to fit and measurements made decades apart to be compared.

The chain has two distinct virtues. Precision concerns how tightly repeated readings agree. Accuracy concerns how well they represent the quantity intended. A miscalibrated balance can give the same wrong answer all morning. Repetition then demonstrates precision while concealing bias. Random variation can often be reduced by more observations. Systematic error requires a different instrument, a reference material, a redesigned procedure or an independent method.

Instruments are not passive windows. They transform. A telescope gathers and focuses radiation. A microscope stains, illuminates and magnifies a prepared specimen. A particle detector converts an interaction into electrical signals, software classifications and reconstructed tracks. Each transformation increases reach while adding assumptions and possible distortions. Better instruments can reveal a new object, but an apparent object can also be an artefact of the instrument, preparation or algorithm. Scientific seeing therefore includes tests of the apparatus that makes seeing possible.

Measurement also requires deciding what the quantity is. Fever may be represented by body temperature, yet readings differ by site, device and time. Intelligence, poverty, biodiversity and pain are harder because the concept cannot be placed directly on a balance. Researchers create operational measures, then must show that the measure tracks the intended phenomenon rather than a convenient substitute. A number can be impeccably calculated and conceptually wrong.

Reference materials provide another check. A laboratory analysing blood, soil or metal can test a sample whose properties have been independently characterised. Agreement across methods and laboratories then supports the whole chain; disagreement helps locate whether the problem lies in preparation, instrument, calculation or definition. Measurement becomes stronger when it can be approached from more than one direction.

Uncertainty belongs inside the result. It does not mean that nobody knows anything. It states how much variation, calibration error, model dependence or sampling uncertainty remains under specified conditions. Reporting 10.0 plus or minus 0.2 is more informative than reporting 10.000 without a defensible account of the final digits. Scientific confidence grows by locating uncertainty rather than performing certainty.

Standards make knowledge portable, but they can freeze choices. A diagnostic threshold creates comparable categories while placing people on opposite sides of a line that biology may not respect. A survey question allows aggregation while excluding answers it did not offer. Changing a standard can improve validity and break continuity with older data. Measurement is therefore technical and institutional at once.

The useful question behind any scientific number is not merely how large it is. Ask what was compared with what, by which chain, under which definition, and with what uncertainty. Until those links are visible, the number is a promise waiting for its evidence.

Create a Difference That Matters

Observation shows that two things occur together. Science becomes more demanding when it asks whether one made a difference to the other. That question requires an imagined contrast: what would have happened to the same system, at the same time, without the suspected cause? The exact alternative cannot usually be observed. Experimental design builds the closest defensible substitute.

A control group is useful because it experiences much of what the treated group experiences except the feature under investigation. If outcomes differ, the contrast narrows the possible explanations. Yet groups can differ before treatment. People who choose an intervention may be healthier, richer or more motivated. Random allocation does not make groups identical, especially in a small study. It makes assignment independent of known and unknown characteristics in expectation, allowing probability to describe the remaining imbalance.

Blinding attacks a different route of error. A patient who knows a treatment was given may report differently. A clinician who knows may interpret a borderline result differently. A laboratory worker can handle samples with tiny, unintended differences. Masking the assignment where possible reduces the room through which expectations enter. Placebos, sham procedures and automated assessment are tools for particular settings, not badges every experiment must wear.

The decisive feature is discrimination. A good test creates outcomes that rival explanations predict differently. Suppose a drug group improves. The drug may work, but recovery, attention, altered behaviour, selective dropout or biased measurement might produce the pattern. A comparison that holds some of these routes steady makes the drug explanation more credible. Further studies using different designs can close other routes. Causal confidence often comes from a programme of contrasts rather than one immaculate trial.

Timing matters. Choosing the outcome after seeing which one moved converts exploration into a misleading test. Exploratory work is valuable when labelled as such; it discovers patterns and generates questions. Confirmation asks whether a pattern survives a pre-specified or independently repeated test. Confusing the two lets chance discoveries borrow the authority of prediction.

Many sciences cannot manipulate their central objects. Astronomers cannot move galaxies. Geologists cannot rerun an asteroid impact. Evolutionary biologists cannot restart the history of whales. These fields create discriminating comparisons through natural variation, dated layers, independent predictions, simulations and observations across cases. A solar eclipse can place light near the Sun under unusual conditions. A fossil sequence can be tested against anatomy, geology and molecular estimates. The absence of laboratory control does not reduce such work to storytelling. It changes what counts as control.

Effects also have scale. A detectable difference may be too small to matter outside a large sample. An average benefit may conceal harm in one

subgroup and no effect in another. A result in a tightly selected population may fail in ordinary practice. Design must therefore match the claim: mechanism, efficacy under controlled conditions, effectiveness in the world and consequences for different people are separate questions.

Ethics limits design, properly. Researchers cannot assign smoking, famine or radiation exposure to settle a causal question. They must use observational evidence, past accidents, animal or cellular models and cautious inference. Even permissible experiments require consent, safety monitoring and a balance of risks. The cleanest possible contrast is not automatically the right study.

An experiment is not valuable because someone changed a variable. Its value lies in creating a difference that bears on the explanation while blocking credible alternatives. When that cannot be done directly, science earns causal confidence by arranging several imperfect contrasts whose weaknesses do not all point the same way.

Make Explanations Risk Contact with the World

Facts do not organise themselves. A list of planetary positions, infection counts or fossil shapes becomes useful when an explanation connects them and tells us what else to expect. Science needs ideas that go beyond the observations used to build them. That step creates power and danger together.

The vocabulary is often taught as a ladder: hypothesis at the bottom, theory above it, law at the top. Scientific practice does not work that way. A hypothesis is a proposed answer to a question. A model represents selected features of a system. A theory supplies a wider explanatory framework. A law describes a stable relation, often mathematically, without necessarily explaining why it holds. A theory does not graduate into a law after enough applause. Each performs a different job.

Models are deliberately incomplete. A frictionless plane does not exist, yet it isolates a relation that helps explain motion. A climate model divides atmosphere, ocean, ice and land into computable structures. An

epidemiological model may group people by infection status while ignoring much of their individuality. The test is not whether the model copies everything. It is whether the omissions are controlled for the purpose at hand and whether the model succeeds where it claims to travel.

Johannes Kepler gives the cleanest historical anchor. Tycho Brahe's observations of Mars would not fit the circular orbit Kepler wanted. The discrepancy was about eight arcminutes, small enough to dismiss if the theory held more authority than the measurements. Kepler treated it as a debt. He changed the geometry and reached the ellipse. The memorable feature is not that one number proved one orbit. It is that a cherished form lost when a more accurate account had to explain the mismatch.

Risk matters because flexible explanations can survive anything. Astrology can absorb a failed prediction by adding an influence. A conspiracy can treat contrary evidence as proof of concealment. A scientific explanation earns more when it states conditions under which the world could embarrass it. Novel prediction is especially demanding, since the result was not available while the account was being fitted. Retrodiction can also be strong when old evidence is derived without having been built into the model.

Karl Popper made falsifiability famous, but single tests rarely confront a theory alone. Instruments may malfunction. Background assumptions may be wrong. The sample may be unrepresentative. A mature theory supported across many domains is not sensibly discarded at the first anomaly, and a weak theory should not be protected forever by auxiliary excuses. Judgement concerns the whole pattern: severity of tests, quality of measurement, success of alternatives and whether adjustments create new predictions or merely rescue the old claim.

Explanatory depth matters beside fit. A curve can match past points because it contains enough adjustable parts. A stronger account connects the fit to a mechanism or wider structure, then succeeds on cases not used to tune it. Compression is evidence only when it continues to work after leaving the data that inspired it.

Independent convergence is therefore stronger than repetition of one route. Germ theory gained force from microscopy, culturing, transmission studies, sterilisation, epidemiology and treatment. Evolution is supported by fossils, biogeography, comparative anatomy, genetics and observed change. Each line has limits. Their errors would have to cooperate in unlikely ways to produce the same broad structure.

Scientific explanation is conditional without being flimsy. Newtonian mechanics remains excellent within domains where speed, gravity and scale make relativistic or quantum effects negligible. A model can be false as a complete picture and reliable for a stated task. Knowing where it works is part of knowing it. The mature question is not whether science has delivered final truth, but how much contact an explanation has survived, through how many independent routes, and within what range.

Turn Private Doubt into Public Criticism

A solitary thinker can be sceptical and still protect every favourite idea. Science becomes harder when doubt is distributed among people with different interests, skills and loyalties. Public criticism turns personal fallibility into a social resource, provided disagreement can reach the evidence and carry consequences.

Publication was a major change because it fixed a claim in a form that could travel. The first scientific journals of the seventeenth century registered priority, circulated reports and created archives. Henry Oldenburg launched *Philosophical Transactions* in 1665 as editor and publisher, drawing on an international correspondence network. The journal did not contain modern peer review in finished form. It helped establish something more basic: a dated public record to which replies, repetitions and rival reports could attach.

Peer review developed unevenly over later centuries. Its proper job is limited. Editors and referees can identify unclear methods, missing literature, weak analysis or claims that outrun the evidence. They cannot repeat the work, inspect every raw datum or guarantee that an ingenious

error has been found. Journals also select for interest, fit and prestige. A published paper has passed a gate. It has not received a certificate of truth.

The larger review begins after publication. Public claims acquire addresses, dates and authors, which means later criticism can attach to the same object rather than a moving recollection. Corrections can be linked, priority can be disputed against a record, and a result can be placed beside its competitors. Other researchers may reanalyse the data, apply the method elsewhere, build a better instrument or find that the effect disappears under stronger controls. Systematic reviews and evidence syntheses ask what the whole body of studies supports rather than treating the newest paper as a verdict. Criticism works best when negative results can be seen and when methods, code and materials are available to inspect.

Competition helps and harms. Priority can motivate speed, expose claims to rivals and reward successful prediction. It can also encourage secrecy, rushed publication and strategic presentation. Reputation gives skilled researchers the resources to pursue difficult work, then makes their mistakes harder to challenge. Anonymous review can protect a critic and remove useful accountability. Signed review can improve civility and punish junior dissent. Every arrangement trades one vulnerability for another.

Scientific communities have often failed their own ideal. Institutions excluded women, racialised groups, religious minorities and people outside metropolitan centres while using their labour and knowledge. The Royal Society, founded in 1660, did not elect its first women Fellows until 1945. A community that narrows who may speak loses criticism as well as justice. Diversity is not an automatic cure for error, but relevant differences in experience can reveal assumptions a uniform group does not see.

Public criticism also needs standards of conduct. Fraud matters because fabricated data counterfeit the resistance on which trust depends. Honest error is more common and often more instructive. Corrections, retractions and changed conclusions should be signs that the record is being repaired, though the delay and damage still count. A culture that treats every correction as disgrace teaches researchers to defend mistakes.

Objectivity, on this view, does not require objective people. It requires claims to encounter informed resistance beyond the control of their authors. The result is imperfect because communities can share interests and blind spots. Its advantage is structural: an error that survives one person must still survive rivals, instruments, records and future readers.

Build a Community Larger Than Any Knower

The heroic story gives discovery to a name. The working system distributes it across people, objects and institutions. A modern paper may require instruments designed by one team, samples prepared by another, software maintained elsewhere, statistical advice, technical staff, data curators and participants who supplied the bodies or observations. Authorship compresses this network until the knowledge looks lighter than it was.

Large collaborations make the fact impossible to ignore. The ATLAS and CMS results did not depend on any one member personally understanding and checking every detector component, software package and analysis. The Event Horizon Telescope linked radio observatories across Earth, synchronised their recordings with atomic clocks, transported large volumes of data, combined signals and tested imaging methods. Trust operates through modular responsibility: each part has procedures, specialists, cross-checks and interfaces that allow the whole to function without an all-knowing mind.

Small science works the same way at a less theatrical scale. Laboratories depend on glassware, reagents, reference collections, animal care, clean rooms, machine shops, field stations and people who keep instruments stable. Archives preserve specimens and records for questions their collectors did not imagine. Standards bodies allow one laboratory to compare with another. Libraries and databases make old findings retrievable. Maintenance rarely receives the language of discovery, yet a broken freezer can erase years.

Institutions create time. A university post, observatory, grant or public laboratory can support work whose result is uncertain and distant. Teaching

reproduces skills. Journals preserve priority. Professional societies set norms and connect specialists. During the nineteenth century, research laboratories and disciplines expanded, the word scientist entered English, and scientific work became a career for more people. Professionalisation increased competence and continuity while placing access behind credentials, appointments and patronage.

Much of the network depends on tacit skill. An instrument may meet its specification only when an experienced technician hears a change in a pump or recognises a contaminated culture before the assay fails. Moving a method therefore requires people, training and materials as well as text. Replication can expose how much knowledge was local to a laboratory and absent from its paper.

The community has always crossed borders. Greek natural philosophy drew on Egyptian and Babylonian knowledge. Arabic-speaking scholars translated, criticised and extended Greek, Persian and Indian work. Early modern European astronomy depended on instruments, numbers, texts and observations that had travelled through wider networks. Colonial expansion then moved specimens, maps and medicinal knowledge under grossly unequal conditions. Local experts, enslaved people, artisans and collectors often supplied observations while metropolitan institutions controlled naming and credit.

Henrietta Leavitt's work at Harvard shows both division and hierarchy. Examining photographic plates as one of the observatory's women computers, she found the relation between the periods and luminosities of Cepheid variables that later made vast distance measurements possible. The plates had been exposed elsewhere; later astronomers calibrated and applied the relation. The discovery belonged to a chain, yet pay, status and historical memory were distributed sharply differently from scientific importance.

Community also explains why expertise deserves weight. A reliable specialist has learned which measurements are difficult, which apparent anomalies recur, which models fail at the edges and what the literature

already tried. Outsiders can ask sound questions, but reconstructing that judgement takes time. Trust should follow the quality of the community's training, openness, conflict controls and track record rather than a title alone.

Science grew by becoming larger than any knower. That scale makes possible a black-hole image and a global disease trial. It also creates bureaucratic opacity, concentrated funding and dependence on fragile infrastructure. The task is not to return to the lone genius. It is to make the network visible enough that responsibility does not disappear inside it.

Make the Error System Answerable to Itself

Every safeguard can become a target. Once publication brings jobs, grants and status, researchers are rewarded for producing publishable results. Many journals reward clear novelty. Institutions count papers and citations. Sponsors may favour useful conclusions. None of this requires fraud. Ordinary choices about stopping, excluding, analysing and reporting can lean in the rewarded direction while each looks defensible alone.

A flexible analysis is especially dangerous when its flexibility stays hidden. Researchers may try several outcomes, subgroups or statistical models and report the path that produced an attractive result. They may construct the stated hypothesis after seeing the data, making a retrospective story look like a risky prediction. Studies with null results may remain in drawers, so the published record overstates effects. Small biases can accumulate into a literature that appears more settled than the underlying work.

Science began applying its own methods to this machinery under the broad name meta-research. One famous project attempted independent replications of 100 psychology studies published in three journals. Depending on the criterion, the replicated effects were weaker and fewer passed the conventional significance threshold than in the original papers. A later project examined 21 experimental social-science papers from *Nature* and *Science* and also found smaller effects with incomplete replication. These results were serious evidence about particular samples of fields and

journals. They were not a measurement of all science.

The reforms target different failure routes. Better correction begins before a mistake becomes famous: sufficiently large samples, validated measures, realistic uncertainty and analysis plans narrow the room for a flattering accident. After publication, error notices must remain attached to the record so that a corrected paper does not continue circulating as if nothing happened. Preregistration creates a time-stamped account of planned questions and analyses before outcomes are known. Registered Reports ask journals to review the question and design, then offer conditional acceptance before results arrive. Data and code sharing permit reanalysis. Larger collaborative studies can reduce small-sample noise and distribute decisions. Systematic review searches for unpublished and conflicting evidence rather than treating the visible headline as the whole record. Reporting guidelines reveal omissions. None guarantees quality; each moves a choice into a place where it is harder to disguise.

Reproducibility and replicability need declared meanings because fields use the words differently. This book follows the National Academies' useful distinction. Reproducibility asks whether the same data and computational steps yield the same result. Replicability asks whether a new study addressing the same question with new data reaches a consistent conclusion. A result can be computationally reproducible and empirically wrong. A valid effect can fail to replicate when populations or conditions differ. The diagnosis depends on which link failed.

Transparency also has costs. Medical and social data can expose participants. Open materials can enable harassment or extraction without credit. Preregistration can become paperwork performed after the real intellectual choices. Large teams can standardise one flawed design. Metrics created to reward openness can themselves be gamed. Reform must therefore remain experimental rather than harden into another checklist.

The deeper correction concerns incentives. A system that celebrates novel positive findings and treats verification as second-class work cannot rely on

personal virtue to produce a balanced record. Funders, journals and employers must make careful measurement, replication, data stewardship and correction worth doing. Error control has to include the career system through which evidence is made.

This closes the loop. Science began by placing experience into records so it could outlive the observer. At scale, those records became papers, databases and metrics able to amplify error far beyond one observer. The same answer returns at a higher level: expose the procedure, preserve the decisions, create independent checks and let the system encounter evidence about itself. Science corrects when correction has somewhere to stand.

How It Actually Works

Counting what returns

Before writing, knowledge travelled through bodies, tools and teaching. A navigator knew a coastline by practised sequence. A healer recognised a wound by smell and colour. A farmer planted by remembered signs. Such knowledge could be exact, but the record of how it had been tested was usually inseparable from the community carrying it.

Cities created a different memory. In Mesopotamia, writing began in administration and expanded into learned traditions maintained by scribes. Astronomical diaries recorded the Moon, planets, weather, prices and public events across centuries. Later Babylonian scholars used numerical schemes to predict planetary appearances. Their accounts mixed what modern readers separate as astronomy, astrology, mathematics and divination. The categories were theirs, not ours. What matters for this story is the joining of sustained observation, trained calculation and an archive able to outlast its makers.

Other traditions developed their own combinations. Egyptian survey, medicine and calendrical practice joined craft with written rules. Indian mathematical astronomy linked calculation to calendars and planetary

models, transmitting numerals and techniques through Sanskrit, Persian and Arabic networks. Chinese states maintained astronomical offices because calendars, eclipses and celestial signs mattered to government. Long series of observations made novae, comets and sunspots available to later investigators.

None formed an early version of a modern laboratory. They solved different problems under different institutions. Yet each enlarged the span over which experience could be checked. A recurring sky could be compared with a record rather than a grandfather's memory. Knowledge began to acquire a past it could argue with.

The arithmetic also travelled. Mesopotamian place-value calculation and base-sixty divisions left descendants in angular and time measurement. A technique can outlive the cosmology that once gave it meaning. Scientific inheritance often works that way: later users keep an operation because it remains useful while forgetting the institution and questions that produced it.

Argument, proof and causes

Greek thinkers inherited and transformed material from Egypt and western Asia. Their distinctive contribution was not observation from a blank page. It was a public culture of demonstration and dispute that made reasons part of the claim.

In mathematics, proof created conclusions that followed from stated premises rather than the authority of the calculator. In natural philosophy, competing accounts asked what the world was made of, how change occurred and what counted as a cause. Aristotle collected observations, classified animals and built a systematic account of demonstration, explanation and causal knowledge. His framework was powerful enough to organise inquiry for centuries and broad enough to preserve large errors with it.

Greek medicine also moved some illness away from divine punishment towards patterns in bodies, diet, place and season. The surviving Hippocratic writings do not speak with one voice, and their treatments often

failed. Their case histories still show an important turn: symptoms were recorded through time, outcomes mattered, and illness could be described as a natural process.

The gain had a cost. Once an elegant system became a curriculum, later observation could be forced to fit it. Galen's anatomy, based substantially on animals, acquired authority across languages and centuries. Aristotle's physical works became objects of commentary as well as criticism. Reason could challenge tradition, then become tradition itself.

Hellenistic centres joined demonstration to observation and engineering. Euclid organised geometry from definitions and postulates. Archimedes moved between mathematical proof and problems of balance, buoyancy and machines. Ptolemy built an astronomical system able to calculate planetary positions with formidable reach. These achievements were not one method. They showed that mathematical structure could discipline claims about both necessary relations and observed motion.

The Greek inheritance survived because other societies translated, taught and disputed it. The route was never Greece straight to modern Europe. It ran through Hellenistic centres, Syriac scholars, Islamic courts, libraries and observatories, Byzantine copying, Latin translation and many local combinations. Science advanced less like a torch passed intact than like material repeatedly dismantled and rebuilt.

Translation, instruments and mixed traditions

From the eighth century, scholars working in Arabic translated Greek, Persian and Indian texts on mathematics, medicine and astronomy. Translation was productive work. Terms had to be created, arguments repaired and conflicting tables compared. The resulting sciences were neither preserved Greek thought nor one uniform Islamic programme.

Astronomers improved observations and mathematical devices, criticised Ptolemy and built instruments and observatories. Physicians organised clinical experience alongside inherited texts. Algebra developed as a field with its own operations. Ibn al-Haytham's Book of Optics, written in the

eleventh century, combined geometry, controlled arrangements of light and sustained criticism of earlier accounts of vision. He argued that sight depends on light entering the eye, then tested optical relations with apertures, mirrors and refraction. Calling him the inventor of a complete scientific method would replace one origin myth with another. His work shows a mature experimental and mathematical practice inside a longer tradition.

Observatories at Baghdad, Maragha, Samarkand and elsewhere turned patronage into long programmes of instrument building, observation and table making. Astronomers developed geometrical devices that repaired problems in Ptolemy. Historians still debate particular routes of influence on Copernicus, but the larger point is secure: Renaissance astronomy did not restart from an untouched Greek original. It encountered a tradition already criticised and extended.

East Asian knowledge also changed through exchange. Chinese developments in paper, printing, magnetic navigation and gunpowder altered how information, travel and force could be organised. Jesuit scholars later entered Chinese calendrical and astronomical institutions carrying European tables and instruments while learning from Chinese records and officials. Comparisons between predictions became matters of political authority as well as celestial accuracy.

In the Americas, Africa, Asia and the Pacific, European travellers depended on local guides, translators, healers, navigators and collectors. Plants, maps and specimens moved through trade and empire, often stripped of the names of those who supplied the knowledge. Circulation increased the reach of inquiry while conquest decided whose account entered the archive.

Translation into Latin from Arabic and Greek, followed by the spread of paper and print, widened access while changing texts through error and commentary. By 1500, the ingredients later called modern science were widely dispersed: mathematical astronomy, natural histories, medical traditions, instruments, paper, numerical techniques, craft knowledge and institutions of learning. The next transformation rearranged them around

print, new voyages, precision instruments and public claims.

The early modern rearrangement

Copernicus proposed in 1543 that Earth moved around the Sun. His system did not instantly defeat Ptolemy, and its early numerical performance was not uniformly better. It changed the problem by relocating Earth and making a new architecture available for testing.

Tycho Brahe then built an observatory capable of unusually precise naked-eye measurements. Johannes Kepler inherited the Mars observations and spent years trying to fit them to circular motion. The eight-arcminute mismatch would not go away. His ellipse was a victory of disciplined surrender: geometry changed because the record resisted it.

In January 1610 Galileo pointed an improved telescope at Jupiter and watched small lights alter position night after night. Their motion showed that bodies could orbit something other than Earth. The telescope did not speak without interpretation. Critics questioned its distortions, observers needed skill, and celestial objects could not be brought into a room. Galileo answered with dated drawings, demonstrations and instruments that others could try. A new eye required a new argument about seeing.

Anatomy changed through a related collision. Vesalius published *De humani corporis fabrica* in 1543 with detailed illustrations based on human dissection, correcting parts of the Galenic tradition. The book depended on bodies, artists, cutters and printers as well as the anatomist. Direct inspection gained authority because it could be displayed and taught, yet the page still arranged what readers were allowed to see.

Francis Bacon supplied a different part of the rearrangement. He attacked the mind's tendency to seize flattering patterns and urged organised observation, experiment and collective natural histories. Practising investigators did not adopt one Baconian recipe. His lasting contribution was a political and intellectual image of knowledge built through disciplined labour rather than recovered from authority alone.

Printing multiplied diagrams, tables, corrections and priority disputes. Voyages brought new organisms and geographies into European collections while colonial power structured the traffic. Improved clocks, lenses, balances and pumps created phenomena that ordinary senses could not stabilise. By the seventeenth century, a claim increasingly had to state what had been seen, which apparatus had produced it, which procedure had been followed and who had witnessed it.

Societies, journals and public experiment

The air pump made the new arrangement visible. Robert Boyle used it to produce a rarefied space and investigate air's pressure. The machine leaked, required skilled operation and was expensive to copy. An experimental fact therefore depended on witnesses who could attest what occurred inside a troublesome device.

The Royal Society of London, founded in 1660 and chartered in 1662, made such performances part of an organised culture. Its motto, *Nullius in verba*, rejected taking anyone's word as final. In practice the Society relied heavily on credible testimony, social rank, correspondence and trusted instrument makers. The ideal of public checking arrived through the imperfect society available to carry it.

Henry Oldenburg's *Philosophical Transactions* began in 1665. It registered claims, circulated letters and gave observations a dated home. Journals reduced the need for every reader to attend the same demonstration. They also created priority, editorial power and pressure to compress messy work into a persuasive account. Modern external peer review emerged gradually rather than at the journal's birth.

Across Europe, academies, observatories and correspondence networks connected court patronage, craft and natural philosophy. Newton's *Principia* joined terrestrial motion and celestial orbit in one mathematical framework, but the achievement depended on prior observations, mathematical techniques, editors, printers and Edmond Halley's labour and money. The book's authority came from demonstration and predictive reach, not from

Newton having followed a universal method.

Scientific correspondence did work that printed books could not. Letters carried preliminary observations, requests, objections and samples at a speed that let distant investigators coordinate. Oldenburg translated and brokered such exchanges. The network also filtered them through language, reputation and postal access. Public knowledge was expanding, but the public able to make it was still narrow.

Public experiment changed what secrecy meant. Alchemy had long included practical testing, yet its coded traditions and guarded recipes limited collective correction. Open reporting did not end secrecy, private patronage or failed replication. It made disclosure an increasingly defensible norm. Robert Hooke's *Micrographia*, also published in 1665, turned microscopic observations into engraved images that readers could inspect without owning his instrument. The images were crafted representations, not retinal copies, but they widened the witnessing community and made instrument-mediated sight part of public culture. A claim that could not be shown, described or reproduced now carried a visible weakness.

Laboratories, disciplines and professional science

During the eighteenth and nineteenth centuries, inquiry acquired new rooms and careers. Chemistry centred balances, purified materials and controlled reactions. Lavoisier's accounting of mass and new nomenclature helped replace substances known through craft traditions with a common language of composition. Standard names made disagreement easier because researchers could tell whether they were discussing the same thing.

Natural history expanded through museums, voyages and imperial collecting. Classification made comparison possible while metropolitan institutions absorbed specimens and local knowledge from unequal networks. Geologists learned to read layers and processes whose timescales exceeded written history. Darwin assembled natural selection from specimens, breeding, geology, correspondence and decades of revision. No laboratory experiment recreated the history of life. Independent

patterns made the explanation powerful.

Precision industries and state services pulled scientific standards outward. Telegraphy, electrical engineering, navigation and public health required common units, tested materials and trained inspectors. The boundary between science and technology remained porous: instruments built for practical use created research questions, while theories became machines, processes and standards.

Laboratories spread through universities, especially in chemistry and physiology. Students learned by doing under a director, producing both results and trained successors. Governments established surveys, observatories and public-health laboratories. Industry employed chemists and engineers. William Whewell's word scientist appeared in print in 1834 partly because specialised practitioners no longer fit comfortably under natural philosopher.

Germ theory shows the new system working across sites. Microscopes revealed organisms, but seeing a microbe did not establish a cause. Cultures, staining, controlled transmission, sterilisation and clinical change had to connect. Pasteur, Koch, Lister and many less celebrated technicians and patients built different links. Laboratory methods transformed medicine while sometimes forcing complex diseases into a narrow one-pathogen model.

Professional science increased reliability through training, equipment and shared standards. It also drew borders. Women could calculate, illustrate, collect and conduct research while being denied posts or society membership. Colonial subjects supplied data without control over its use. The modern institution made knowledge more durable and exclusion more organised.

Statistics, trials and organised uncertainty

As states counted births, deaths, trade and disease, variation became an object rather than an embarrassment. Probability had begun in problems of games and error. By the nineteenth century it was being used to combine observations, estimate regularities and distinguish signal from scatter. The average person became a statistical construction with administrative power.

Experimental statistics developed around agriculture, biology and industrial quality. Randomisation, blocking and designed comparisons allowed researchers to separate treatment effects from uneven fields and background variation. These tools did not automate judgement. They clarified which uncertainty came from a design and which conclusions the design could support.

Medicine adopted controlled trials unevenly. The 1948 Medical Research Council study of streptomycin for pulmonary tuberculosis became a landmark because treatment allocation was concealed and randomised, comparison was planned, assessment was structured and scarce drug supplies made the question urgent. It also revealed that resistance could emerge. The trial was not the first fair comparison in medicine, and later standards added informed consent, independent ethics review, registration and monitoring.

Statistics created its own temptations. A threshold can help coordinate decisions and then become a ritual separating publishable from invisible results. A model can quantify uncertainty while hiding assumptions about sampling and missing data. Large datasets can produce precise answers to badly framed questions. Statistical evidence is strongest when design, measurement and subject knowledge remain attached to the calculation.

The deeper change was institutional. Uncertainty could now be stated in common forms, audited and combined across studies. Public-health work also showed the value of non-randomised evidence. Mortality registers, maps, outbreak timing and comparisons among water supplies could identify patterns no planned trial could ethically create. The strongest

inference came when administrative records, local investigation and a plausible route of transmission converged.

Clinical evidence moved from the authority of an experienced doctor towards protocols, trials and synthesis. That shift improved the ability to detect benefits and harms, while sometimes undervaluing judgement, individual variation and outcomes that were hard to count.

Big Science and the digital research system

The twentieth century joined science to states, war and industry at unprecedented scale. Radar, nuclear weapons, antibiotics, rockets and computing emerged from large programmes that combined universities, firms and government laboratories. The Manhattan Project proved that coordinated science could transform matter and geopolitics while secrecy blocked the openness often treated as science's safeguard.

After the war, particle accelerators, space missions and genomic projects required machines and teams beyond the reach of a single laboratory. In 2012 the ATLAS and CMS collaborations at CERN independently reported a new particle near 125 gigaelectronvolts, later confirmed as the Higgs boson. The agreement mattered because two detectors, collaborations and analyses confronted the same prediction through partly independent routes.

The Event Horizon Telescope extended the pattern across Earth. In 2017 radio observatories observed the galaxy M87 in coordination, recording signals against precise clocks. Data were transported to correlators because ordinary internet transfer was impractical at that scale. Teams then used separate imaging approaches and extensive tests before the 2019 release. The famous ring was the visible end of metrology, logistics, algorithms and institutional trust.

Computers transformed every field. They enabled simulation, automated instruments, vast databases and analyses no person could perform by hand. They also allowed an error in code or a hidden default to propagate through thousands of results. Digital records can be copied perfectly and become unusable when software, formats or documentation disappear.

Large projects also changed publication. Author lists grew into collaboration rosters, and data policies became part of experimental design. The Human Genome Project used rapid public release to make sequence data a shared platform rather than a private end product. Such rules accelerated reuse while raising familiar questions about credit, consent and who could afford the computing needed to benefit.

The present research system is therefore turning scrutiny towards itself. Replication projects, trial registries, preregistration, Registered Reports, open data, code archives and reporting standards attempt to expose choices once hidden inside polished papers. UNESCO's 2021 Recommendation on Open Science placed access, infrastructure and participation inside an international policy framework. Openness remains uneven and cannot erase privacy, security or cost. The direction is clear: a global evidence system must preserve the route by which its results were made.

How we know

The history of science is reconstructed from surviving instruments, tablets, manuscripts, specimens, images, notebooks, correspondence, publications, institutional records and oral knowledge. Survival is selective. Clay endures better than practice; printed controversy survives better than routine competence; famous authors leave papers while technicians and participants often leave names only in payrolls, labels or margins.

The word science changed meaning, and scientist did not enter English until the nineteenth century. Applying either term to ancient activity can clarify shared practices and conceal the purposes under which people worked. This account therefore follows records, measurement, comparison, explanation and criticism without claiming that a Babylonian scholar or Chinese calendar official was a modern researcher in waiting.

Origin stories are especially unstable. Claims that Greece, Europe or one medieval scholar invented the scientific method usually project a later package onto one contribution. Global histories correct that error but can

flatten important differences by treating every exact craft as the same enterprise. The strongest conclusion is narrower: modern science formed through cumulative and unequal recombination, with the early modern European reorganisation one decisive transformation rather than creation from nothing.

What People Get Wrong

"There is one scientific method"

The familiar classroom diagram runs from question to hypothesis, experiment, analysis and conclusion. It is useful as a first reminder that claims should meet evidence. It becomes false when presented as the operating instructions for science.

A field astronomer cannot control a galaxy, a palaeontologist cannot rerun an extinction and a mathematician proves rather than experiments. Taxonomists compare structures. Epidemiologists combine observation, natural variation and trials. Engineers test designed objects against performance and failure. Even within one laboratory, exploration, measurement, model-building and confirmation rarely occur in a tidy order.

What unites scientific work is not one sequence but a family of constraints: preserve the evidence, make measurements comparable, expose assumptions, create tests that distinguish alternatives, report uncertainty and permit informed criticism. A study can follow the schoolbook cycle and remain weak. Another can depart from it entirely and produce durable knowledge. The method is plural because the world does not offer every question in the same form. The diagram survives because teaching and publication present inquiry after the uncertainty has been removed. A clean report can describe how a claim was tested without describing the wandering route by which anyone thought of it.

"Science was invented once, in Europe"

This story remains persuasive because seventeenth-century Europe did undergo a major transformation. New instruments, print, mathematical physics, societies and public experimentation changed the scale and authority of natural knowledge. Copernicus, Kepler, Galileo, Bacon, Boyle and Newton belong in any serious account.

The error is turning that reorganisation into creation from nothing. Babylonian records fed Greek astronomy. Indian numerals and mathematical astronomy travelled through Arabic scholarship. Islamic astronomers and opticians translated, criticised and extended ancient work. Chinese instruments, paper, printing and long observations altered what could be recorded and transmitted. Early modern European collections relied on navigators, artisans, translators and local experts across expanding commercial and imperial networks.

The correction does not require pretending that all traditions were identical or that modern science existed everywhere in miniature. Distinct institutions pursued distinct aims. The history is one of recombination under unequal power. Europe became a decisive centre, then wrote a genealogy in which everything useful appeared to have been waiting for Europe to discover it. Priority contests make the distortion worse by asking who was first, as though an institution, technique or concept arrived complete in one mind. The more useful question is which earlier materials were recombined, and under whose power.

"A single decisive experiment proves the answer"

Textbooks like clean turning points. One object falls, one telescope points upward, one expedition measures an eclipse, and an old worldview collapses. Such stories give theories a trial scene and evidence a final verdict.

Real tests arrive with supporting assumptions. The instrument must work, the sample must represent the target, the calculation must be sound and rival explanations must be addressed. An unexpected result may expose the

main theory, an auxiliary assumption or an unnoticed fault. A confirming result may fit several accounts. Scientists therefore ask whether the finding survives better measurement, different methods and attempts designed to favour alternatives.

Some experiments are decisive within a restricted dispute. None carries every implication later attached to it. Galileo's moons challenged the claim that all celestial motion centred on Earth, but did not by themselves prove the full Copernican system. The strongest knowledge usually comes from a network of results whose weaknesses differ. A dramatic experiment can turn the argument. It rarely ends the subject. Experiments later called decisive often became so because subsequent work stabilised their instruments, repeated the effect and showed that rival theories could not absorb it. History removes the uncertainty that participants still faced and hands the winning result a spotlight.

"Peer review means the paper is true"

The myth is encouraged by the language of publication. A paper is accepted, passes review and appears in a prestigious journal. News reports then treat that sequence as certification.

Referees usually read a manuscript, not repeat its experiment. They may detect an invalid analysis, missing control or exaggerated conclusion. They may also miss fabricated data, subtle code errors and assumptions outside their own expertise. Reviews disagree. Editors weigh novelty, audience and space as well as reliability. Famous journals can publish weak papers, while strong work can be rejected.

Peer review remains useful because informed criticism before publication often improves a claim and filters material that is plainly unready. Its authority should match its job. Publication says that editors and reviewers judged the work worth entering the record. Confidence should then depend on the design, evidence, transparency, later criticism, independent work and place within the wider literature. The paper is the beginning of public testing, not its successful completion. Prestige should affect attention less than it

often does. A preprint with transparent methods may deserve serious consideration before review, while a celebrated journal article may deserve little confidence after its data, design or replication fails. Venue is evidence about selection, not nature.

"Replication means repeating every detail"

An exact repeat can be valuable when checking whether the reported procedure produces the reported result. Yet copying every visible detail may preserve the original mistake and test only one setting.

A direct replication holds closely to the earlier design. A conceptual replication tests the same claim through a different operation. Reanalysis checks whether rerunning the stated code and analytical workflow on the original dataset reproduces the reported output. New data ask whether the effect appears again. Fields use reproducibility and replicability inconsistently, so the terms should always be defined rather than treated as passwords.

A failure to replicate also needs diagnosis. The original result may have been noise, biased or false. The new study may be underpowered or badly executed. Populations, materials and contexts may differ in ways that reveal a real boundary. The right response is neither automatic dismissal nor automatic rescue. Replication is a comparison between evidence chains. Its value lies in showing which parts travel, which depend on conditions and which cannot be recovered from the published account. Direct and conceptual replications answer different questions and work best together. One tests whether the original route can be recovered; the other tests whether the claimed relation survives after that route changes.

"Objectivity requires scientists without values or bias"

No such people exist. Researchers choose problems, instruments, classifications, thresholds and acceptable risks. They bring training, interests and expectations. A claim that science is objective because its practitioners possess blank minds collapses at the first honest biography.

The stronger ideal is organised objectivity. Records constrain memory. Calibration constrains instruments. Blinding constrains expectation. Rival methods expose shared assumptions. Diverse critics can notice exclusions. Public reasons let others challenge a result without first becoming the author. These devices do not erase judgement. They make some judgements visible and answerable.

Values still enter. Deciding how much evidence is enough to approve a drug, protect a species or evacuate a town cannot be derived from data alone. Costs of false alarms and missed danger matter. The correction is not that every conclusion is political opinion. Evidence can sharply constrain what is plausible. Objectivity means building resistance to personal preference while stating where social choice remains. Scientific images show why trained judgement cannot be removed. A microscope slide or scan requires preparation, selection and interpretation. The answer is not to pretend the image made itself, but to standardise the work, compare readers and preserve the original signal.

"Science always corrects itself"

The phrase sounds reassuring because many famous errors were corrected. It hides the people, time and institutions required. A mistake can persist when data are inaccessible, dissent is costly, journals prefer novelty, sponsors control publication or a field shares the same assumption. Those harmed during the delay do not receive their years back when the textbook changes.

Correction also has no automatic direction. Better evidence may lose to prestige. A retraction may remove a paper while its claim continues to circulate. A null replication may be ignored. Fraud can be exposed by an unusually persistent critic rather than routine safeguards. Scientific communities can defend status as vigorously as any other institution.

The record improves when correction has practical support: accessible methods and data, independent teams, protected criticism, replication funding, clear notices, evidence synthesis and rewards for admitting error.

Preregistration and Registered Reports target particular incentives. They can themselves become ritual. Science is capable of self-correction because it contains tools for exposing error. Capability becomes performance only when people are allowed and encouraged to use them. Speed matters too. A field that eventually changes after decades of preventable harm has corrected its record and failed many people. Self-correction should be judged by detection, response and repair, not by the comforting fact that no error lasts forever.

Use It

Trace the claim to the observation

When a scientific claim reaches you, it has usually crossed several translations. An instrument produced a signal. A researcher classified or transformed it. A statistical model produced an estimate. A paper compressed the estimate. A press release selected a finding. A headline converted it into a sentence about the world.

Work backwards. What was observed directly? Was it a blood marker, a self-report, a satellite reading, an image classification or an outcome that mattered to people? How many transformations separate that observation from the claim? Which link depends on a model or judgement?

This does not mean distrusting every transformation. A raw detector signal is often less meaningful than a carefully reconstructed quantity. The aim is to locate where interpretation entered and whether the route remains inspectable. Beware a claim whose confidence grows as its evidence becomes less visible. The best popular account can usually tell you what was measured, how it connects to the conclusion and what remains indirect.

Inspect the comparison

Causal language should trigger one question: compared with what?

A treatment group improving after treatment proves little by itself. People recover, adapt, change behaviour and report differently. Temperatures

rising in one city do not establish a global cause. A species found near a pollutant may already have occupied unusual habitat. The comparison carries the argument.

Look for the nearest defensible alternative: a control group, the same population before and after a change, a matched region, a natural experiment, a dose pattern, a predicted difference or an independent method. Then ask what could still differ between the cases. Random allocation helps with baseline imbalance. Blinding helps with expectation. Repeated measurement helps with ordinary variation. None repairs the wrong target or a biased outcome measure.

A strong comparison makes rival explanations struggle. A weak one merely decorates the preferred story with numbers. Whenever the word causes appears, the design should be doing visible work.

Follow the measurement chain

A number can look harder than the choices that produced it. Treat every measurement as a chain.

Begin with the concept. What does the study mean by poverty, recovery, biodiversity, intelligence or exposure? Move to the operation. Which questionnaire, sensor, threshold or sample represents that concept? Then inspect calibration, missing data, timing and uncertainty. A highly precise estimate of a poor proxy remains a poor answer.

Comparability deserves special attention. Two unemployment rates may use different definitions. Two historical temperatures may come from different sites or instruments. Two medical studies may measure different outcomes under the same label. Combining incompatible quantities creates a larger dataset and a smaller understanding.

Standards are evidence infrastructure. Shared units, reference materials and reporting rules make aggregation possible. They should still be questioned when a threshold creates a false cliff or a convenient metric displaces the phenomenon of interest. The measurement chain is strong

when each link is suitable and the handovers are documented.

Separate result, explanation and decision

Scientific disputes become confused when three questions are treated as one.

The result asks what the study found: an estimate, pattern or detected signal under stated conditions. The explanation asks what process produced it and whether rival accounts fit. The decision asks what should be done given the evidence, costs, risks and values.

People can agree on the result and dispute the explanation. They can agree on both and choose different policies because they weigh harms differently. A regulator may act before a causal mechanism is complete when delay is dangerous. A clinician and patient may reject the average best option because the individual's priorities differ. None of this makes the evidence arbitrary.

When debate becomes heated, label the level. Is someone challenging the data, the causal model, the range of the claim or the acceptable response? Demanding more science cannot settle a conflict whose remaining issue is moral or political. Equally, invoking values does not erase a physical constraint. Clear boundaries prevent evidence from impersonating policy and preference from impersonating evidence.

Ask how far the claim can travel

Every study has an address. Its participants, instruments, organisms, institutions, climate and period define where the result was made. Generalisation is a further inference, not a free gift included with statistical significance.

Ask what population was sampled and who was absent. Laboratory animals may reveal mechanisms without predicting the size of an effect in humans. University volunteers may respond differently from older adults, children or people under economic strain. A farming intervention that works in one soil and rainfall pattern may fail elsewhere. A social effect can depend on law,

language and expectation.

The right response is not to dismiss local evidence. It is to look for the feature expected to carry the result across settings. A biochemical pathway may be widely shared. A behavioural response may be culturally contingent. Replication across deliberately varied conditions maps the boundary better than repeating one convenient sample.

Claims should expand at the speed of evidence. The more universal the sentence, the more diverse the supporting routes need to be.

Judge the correction machinery

When experts disagree, counting credentials is a poor substitute for examining the system around them.

Look for an accessible record, independent data or methods, active criticism, competing groups and evidence synthesis. Ask whether negative findings can be published and whether conflicts of interest are disclosed. Check whether the claim has survived outside the original team. A mature consensus should be able to state the strongest remaining uncertainty and explain why rejected alternatives now carry less weight.

Also inspect incentives. Is a company testing its own product while controlling the data? Does a journal reward novelty? Would a junior researcher risk a career by challenging the result? Is verification funded? Institutions that make criticism costly can retain impressive procedures while losing their function.

Corrections are informative. A field that issues clear retractions, updates guidelines and revises estimates may look less certain than one that never changes. The comparison should concern how errors are found and repaired, not how rarely fallibility is admitted. Trust the machinery in proportion to its exposure to resistance.

The limits

Scientific practice cannot make every important question scientific. It can estimate consequences, reconstruct causes and test means. It cannot derive a complete account of justice, beauty, dignity or a life worth living. Attempts to measure such things may illuminate part of them and acquire authority over the rest.

Evidence is also limited by access. Some events happen once. Some populations cannot ethically be experimented on. Historical records disappear. Complex systems change while they are studied. A model may be the best available and still leave a wide range of outcomes.

The institutions add their own limits. Funding follows power and commercial value. Categories can carry prejudice. Military and colonial systems have produced knowledge through coercion and extraction. Openness can conflict with privacy and security. Expertise can become insulated, while public pressure can punish unwelcome findings.

Science offers disciplined answers to questions that can be made answerable to evidence. It does not guarantee good questions, fair institutions or wise use. Those remain human work.

The one thing to keep

Keep the chain.

A scientific claim is never strengthened by the label alone. Its authority comes from a connected route: experience placed into a record, a quantity tied to a reference, a comparison built to expose alternatives, an explanation made vulnerable to the world, criticism able to reach the evidence and a community capable of correcting the result.

You will rarely be able to inspect every link. Nobody can. You can ask whether the chain exists, whether relevant specialists can inspect it and where it is weakest. That is enough to distinguish warranted trust from obedience and informed doubt from reflexive denial.

The deepest change in human knowledge was not that people stopped being biased, ambitious or mistaken. They built arrangements in which a claim could leave its maker, encounter resistance and return altered. The final product is not certainty. It is an error that has had fewer places to hide.

Once you see science this way, changing a conclusion under pressure from evidence no longer looks like failure. Refusing to expose the chain does.

Terms

Observation. A recorded encounter with a phenomenon, made through senses, instruments or both. An observation is produced under conditions and classifications; it is not a neutral piece of the world detached from a method.

Data. Representations used as evidence, such as measurements, images, counts, sequences or coded responses. Data are made through selection and processing. Calling them raw does not remove that history.

Metadata. Information needed to interpret data: time, location, instrument settings, units, variable definitions, file versions or collection conditions. Missing metadata can turn an extensive dataset into an unusable one.

Provenance. The documented origin and handling history of a specimen, measurement, image or dataset. Provenance helps reveal contamination, alteration, mislabelling and whether evidence came from the source claimed.

Variable. A feature that can take different values across observations, people, organisms or conditions. Defining and measuring variables well determines what a comparison can detect and what it will miss.

Operational definition. The procedure used to represent a concept in a study. Measuring stress through a questionnaire, hormone or behaviour creates different operational definitions and may answer different questions.

Measurement. A structured comparison between a quantity or property and a reference or rule. Measurement includes definitions, instruments,

calibration and uncertainty; the reported number is the end of that chain.

Calibration. Comparison of an instrument or procedure with a reference to establish how its readings relate to accepted values. Calibration can reveal bias that repeated measurements alone would hide.

Traceability. An unbroken, documented chain connecting a measurement to recognised references through calibrations, each with stated uncertainty. Traceability allows results from different laboratories and times to be compared.

Accuracy. Closeness between a measurement and the value it is intended to represent. Accuracy can be limited by systematic error even when repeated readings agree closely with one another.

Precision. Closeness among repeated measurements under conditions. Precision describes consistency, not correctness. A stable but miscalibrated instrument can produce results that are precise and wrong.

Uncertainty. A quantified or otherwise stated limit on what a result establishes. It may arise from variation, sampling, calibration, model assumptions or missing information. Honest uncertainty strengthens interpretation.

Hypothesis. A proposed answer or relation that can guide investigation. Hypotheses range from tentative working ideas to precise predictions; they do not all begin science and do not mature automatically into theories.

Model. A selective representation of a system, expressed through equations, diagrams, organisms, simulations or physical objects. Models are judged by fitness for purpose, assumptions, range and contact with evidence.

Theory. A connected explanatory framework that organises findings and supports new questions or predictions. In science, theory does not mean a casual guess, nor does it become a law after confirmation.

Law. A compact statement of a regular relation, often mathematical. A law

may describe what happens within a domain without supplying the mechanism, and may require approximation or limiting conditions.

Mechanism. An account of the entities, activities and organisation through which an outcome is produced. Mechanistic evidence can strengthen causal claims, but a plausible mechanism cannot replace evidence that the effect occurs.

Prediction. A consequence expected if an account and its supporting assumptions hold. Predictions made before the result is known are especially demanding, though successful retrospective predictions can also provide strong tests.

Experiment. A designed intervention or arrangement that creates observations bearing differently on rival explanations. Experiments vary widely and need not involve laboratories, white coats or one variable changed at a time.

Control. A comparison condition used to estimate what would have happened without the feature under study. Controls can address specific alternatives, but no single control holds every possible difference steady.

Randomisation. Assignment by a chance procedure, used to prevent systematic selection and support probability-based inference. Randomisation balances characteristics in expectation; it does not guarantee identical groups in one study.

Blinding. Concealing an assignment or information from participants, investigators, assessors or analysts to reduce expectation-driven differences. Who can be blinded depends on the intervention and practical setting.

Confounder. A factor associated with both a suspected cause and an outcome that can create or distort their apparent relation. Design, measurement and analysis can reduce confounding but rarely prove its total absence.

Replication. A new investigation of the same scientific question using new

observations or data. It may follow the original design closely or test the claim through a deliberately different operation or setting.

Reproducibility. In this book, obtaining the same computational result from the same data, code and analytical steps. Other fields reverse the terms reproducibility and replicability, so usage should always be declared.

Peer review. Evaluation of research by relevant specialists, before publication or funding. Review can improve and filter work, but reviewers usually do not repeat the study and cannot certify truth.

Preregistration. A time-stamped record of planned questions, outcomes and analyses created before results are known. It reveals departures from the plan but does not ensure that the plan was wise or complete.

Registered Report. A publication format in which a journal reviews a question and method before results exist and can grant conditional acceptance. It reduces result-based publication decisions while preserving later quality checks.

Meta-analysis. A statistical synthesis of results from multiple studies under stated inclusion and modelling choices. Its value depends on study quality, comparability and the visibility of missing or unpublished evidence.

Scientific consensus. Convergence among relevant specialists after evidence has been tested through multiple routes. Consensus is neither unanimity nor a head count. Its force depends on evidence quality and criticism.

Go Deeper

For the global history. James Poskett, *Horizons: A Global History of Science* (Viking, 2022). Poskett follows modern science from the fifteenth century through networks linking Africa, the Americas, Asia, the Pacific and Europe. It is the best next step for correcting the lone-European-genius story without denying the transformations that occurred in European institutions. The narrative is accessible and full of people, instruments and movement. Its large scale requires selective case studies, so treat it as an organising account rather than the final word on every region. It begins around 1450, which means the ancient and medieval foundations in this book need other histories beside it.

For an original programme. Francis Bacon, *The New Organon*, edited by Lisa Jardine and translated by Michael Silverthorne (Cambridge University Press, 2000). Published in 1620, this is Bacon's attack on premature certainty, inherited authority and the mind's recurring idols. Read selected aphorisms rather than expecting a modern laboratory manual. Bacon is forceful, politically ambitious and often stranger than the simplified schoolbook advocate of induction. The edition's introduction explains what he was trying to build and prevents later scientific practice from being read backwards into every sentence. Keep an eye on his institutional ambition: knowledge is imagined as a coordinated public project, not a private exercise in clever reasoning.

For the philosophy. Peter Godfrey-Smith, *Theory and Reality: An Introduction to the Philosophy of Science*, 2nd edition (University of Chicago Press, 2021). This is the clearest route through logical positivism, Popper, Kuhn, scientific realism, explanation, models and social accounts of knowledge. Godfrey-Smith explains why attractive slogans fail without treating philosophy as a collection of semantic traps. It is readable for a newcomer but denser than this book, and its value lies in seeing rival models receive serious treatment rather than one definition of science declared victorious. No mathematics is required, but the chapters reward slow reading and comparison rather than a single uninterrupted sitting.

For the trust question. Naomi Oreskes, *Why Trust Science?* (Princeton University Press, 2019). Oreskes argues that scientific authority rests on the organised scrutiny of relevant communities, especially where diverse evidence and perspectives converge. The book includes responses that test her case, making the disagreement part of the recommendation. Read it after the historical works, when consensus, expertise and objectivity no longer look like simple appeals to status. It is a defence of warranted trust, not a claim that every institution or published result deserves obedience. The published volume grows from lectures and includes critical responses, so the social account of objectivity is itself exposed to the kind of organised disagreement it recommends.

Notes and Sources

This book treats science as a changing group of practices and institutions rather than a timeless object waiting to be named. The English word *scientist* entered print in the nineteenth century, and earlier actors worked under categories such as astronomy, medicine, natural philosophy, mathematics, natural history, craft and administration. Applying the later label can reveal continuities in record-keeping, measurement, comparison and criticism, but it can also erase differences in purpose, authority and cosmology. Sydney Ross traces the history of the word *scientist*. Steven Shapin, James Poskett and Peter Godfrey-Smith support the book's refusal to identify science with one universal recipe.

The geographical record is unequal. Written archives favour states, courts, temples, universities, observatories, laboratories and literate elites. Craft, oral and Indigenous knowledge often survives through objects, later reports or records made by outsiders, sometimes under colonial rule. The account therefore uses selected traditions to explain mechanisms. It does not treat under-coverage as evidence that exact knowledge was absent elsewhere.

The Whole Thing in One Page and Why You Should Care

The organising claim, that science makes claims capable of surviving separation from their claimants, is a structural synthesis rather than a definition accepted by every historian or philosopher. John Ziman's account of science as public knowledge, Helen Longino's account of critical communities, Naomi Oreskes's account of warranted trust and the National Academies' treatment of reproducibility and replicability provide its main supports. The book does not claim that publicity guarantees truth. Public records can preserve fraud, shared assumptions and institutional power as efficiently as good evidence.

The paracetamol packet is an illustrative evidence chain, not a report about one product or regulator. The relevant point is that a dose claim can be supported by chemical identity, metrology, manufacturing controls, sampling, clinical evidence and legal oversight distributed across institutions. No consumer personally repeats that chain. Trust remains conditional on whether the links exist and can be inspected.

The summaries of Kepler, Boyle, the streptomycin trial, the Higgs results and the Event Horizon Telescope are anchors for different kinds of error control. None is presented as the moment science began. Kepler's Mars discrepancy concerned about eight arcminutes in the observations he inherited from Tycho Brahe. Boyle's air-pump work depended on troublesome apparatus and credible witnessing. The 1948 Medical Research Council study is described as a landmark randomised trial rather than the first controlled comparison in medicine. ATLAS and CMS reported partly independent evidence for a new boson in 2012. The Event Horizon Telescope image was reconstructed from coordinated radio observations, calibration and several imaging pipelines, not taken by one camera.

Records, metadata and provenance

Francesca Rochberg and Eleanor Robson supply the main basis for the account of Mesopotamian scholarly records, numerical astronomy and institutional memory. Babylonian astronomical diaries joined celestial observations with weather, river, price and political information. Modern separations among astronomy, astrology, mathematics and divination should not be projected cleanly onto that work. The claim is that long records made cycles and deviations available across generations, not that Babylonian scholars practised modern science.

Data, metadata and provenance are used in broad contemporary senses. A record becomes more interpretable when it preserves when, where and how it was produced, which definitions and instruments were used, and how a specimen or file was handled. The vocabulary is modern; the underlying problem is old. The book does not imply that every historic notebook met a current data-management standard or that complete procedural description can eliminate tacit skill.

Measurement and standards

The metrology account follows the Bureau International des Poids et Mesures, *The International System of Units*, ninth edition, updated in 2026, and the Joint Committee for Guides in Metrology's *International Vocabulary of Metrology*. The present SI defines units through fixed numerical values of defining constants rather than a single material prototype. National and specialist laboratories realise those definitions through procedures and calibration chains. Traceability means a documented chain of calibrations to references, with uncertainty stated at each link. It does not mean that one physical object travels through every measurement.

The book distinguishes precision from accuracy to expose a common error: repeated agreement cannot identify a stable systematic bias by itself. It also treats measurement as conceptual. Pain, poverty, intelligence and biodiversity require operational choices before an instrument or calculation can return a number. Theodore Porter's history of quantified objectivity and

Lorraine Daston and Peter Galison's history of scientific objectivity support the institutional argument. Standards permit comparison while also fixing categories and thresholds whose consequences may be uneven.

The official BIPM citation used here is the 2026 updated text of the ninth edition, DOI 10.59161/AUEZ1291. This was the newest important current reference checked on 4 September 2026.

Comparison, experiment and causal inference

The account of designed comparison draws on R. A. Fisher's work on experimental design and the later development of randomised clinical trials. Random allocation makes treatment assignment independent of measured and unmeasured characteristics under the randomisation process. It balances groups in expectation, not by guarantee in one realised sample. Blinding, placebos and controls address different routes of error and cannot be applied uniformly across all fields or interventions.

The counterfactual language, what would have happened without the suspected cause, is a general causal model. In practice no design observes the same unit at the same moment under both alternatives. Experiments, natural experiments, longitudinal observations, mechanistic evidence and comparisons across cases each reduce different uncertainties. Astronomy, geology and evolutionary biology are included to prevent laboratory manipulation from becoming the standard by which every science is judged.

The Medical Research Council's 1948 streptomycin study concerned pulmonary tuberculosis. Treatment allocation was randomised through a concealed procedure, assessment was structured and the limited supply of streptomycin shaped the design. The paper also reported the emergence of streptomycin resistance. The study belongs in a longer history of fair tests, allocation methods and trial ethics; it did not create controlled medicine from nothing.

Models, theories, laws and risky explanation

The rejection of a hypothesis-theory-law ladder follows standard philosophy-of-science usage. Hypotheses, models, theories and laws perform different jobs. A model can be deliberately idealised, a law can describe a regularity without supplying its mechanism, and a theory does not become a law after accumulating support. Peter Godfrey-Smith provides the main general treatment.

Kepler's eight-arcminute passage is based on *New Astronomy*, in William H. Donahue's translation. The discrepancy did not allow one datum to prove an ellipse without supporting assumptions. It mattered because Kepler judged Tycho's observations accurate enough that the mismatch could not be dismissed while preserving the circular construction. The body uses the episode to show a model being made answerable to a resistant record, not to present a clean modern hypothesis test.

Karl Popper made falsifiability central to twentieth-century discussion, but the book rejects the crude view that one anomalous result mechanically destroys a theory. Tests involve instruments, background assumptions, sampling and auxiliary models. Mature assessment compares the severity and independence of tests, the performance of alternatives and whether revisions generate new success or merely protect an old claim. Godfrey-Smith's survey and the historical cases supply this narrower treatment.

The claims about germ theory and evolution concern convergence among evidence routes. Microscopy alone did not establish microbial causation, and no laboratory recreated the history of life. Confidence grew because observations, mechanisms, interventions and predictions with different vulnerabilities supported connected explanations. The examples are not claims that every line is independent in a strict statistical sense.

Publication, peer review and organised criticism

The Royal Society was founded in 1660 and received its first royal charter in 1662. *Philosophical Transactions* began under Henry Oldenburg in March 1665. The journal registered claims, circulated reports and preserved dated contributions, but modern external refereeing was not present in complete form at its birth. Melinda Baldwin's history of peer review at *Nature* and Royal Society archival work support the claim that editorial and refereeing practices developed gradually and unevenly.

The motto *Nullius in verba* is not used to claim that members rejected testimony. Robert Boyle's experimental facts depended on skilled operators, witnesses, correspondence and judgements about credibility. Steven Shapin and Simon Schaffer's study of the air pump supplies the main account of experiment as material and social practice. Robert Hooke's *Micrographia* supports the claim that engravings widened access to instrument-mediated observations while remaining crafted representations.

The statement that the Royal Society first elected women Fellows in 1945 follows the Society's institutional history. Marjory Stephenson and Kathleen Lonsdale were elected in March that year. This date illustrates one formal barrier in one institution. It does not stand for the full, geographically varied history of women in science or imply that election ended exclusion.

Peer review is described according to its actual access. Referees usually assess a manuscript and supporting materials rather than repeat the work. Review can identify errors and improve reporting, but publication does not certify truth. Later reanalysis, replication, synthesis and criticism remain necessary. Oreskes, Longino, Ziman, Baldwin and the National Academies support this institutional treatment.

Distributed labour, expertise and credit

The Higgs discussion rests on the independent 2012 ATLAS and CMS papers. ATLAS reported a new particle at about 126 GeV and CMS one at about 125 GeV. Later work established consistency with the Higgs boson. The body does not state an exact author count and does not claim that the two experiments were independent in every respect. They used separate detectors, collaborations and analysis chains while sharing the Large Hadron Collider and much of the theoretical target.

The Event Horizon Telescope account follows its 2019 M87 papers, especially papers I, IV and VI. Observatories used radio interferometry across intercontinental baselines and atomic-clock timing; recorded data were physically transported to correlation centres; independent imaging teams and validation tests contributed to the released image. The ring is therefore evidence produced through a measurement and reconstruction system, not an unmediated photograph.

Henrietta Swan Leavitt's 1912 paper reported that the brighter Small Magellanic Cloud variables had longer periods, building on her earlier catalogue work. Later calibration turned the relation into a distance tool. The body credits the chain rather than making Leavitt solely responsible for the cosmic distance scale. Harvard's plate archive and Leavitt's paper support the description of photographic plates, divided labour and unequal recognition.

William Whewell used *scientist* in print in 1834. Sydney Ross shows that the coinage answered a naming problem created by expanding and increasingly specialised inquiry. The book says the word appeared partly because practitioners no longer fit comfortably under *natural philosopher*. It does not imply that one word caused professionalisation or that research became a modern career everywhere on one date.

Incentives, reproducibility and replication

The Open Science Collaboration attempted replications of 100 experimental and correlational studies from three psychology journals. Replicated effect sizes were, on average, smaller, and fewer replication results met the conventional significance threshold than the original reports. Colin Camerer and colleagues examined 21 experimental social-science papers published in *Nature* and *Science* between 2010 and 2015 and also found smaller effects and incomplete replication. The body keeps the field, journal sample and study count visible and does not turn either project into a census of all science.

The terms reproducibility and replicability vary among disciplines. This book declares the National Academies' 2019 convention: reproducibility is obtaining consistent computational results from the same data, code and methods; replicability is obtaining consistent results in a new study aimed at the same scientific question. A different field may reverse the labels. The scientific issue should be described even when terminology differs.

John Ioannidis, Brian Nosek and colleagues, Christopher Chambers, Marcus Munafò and colleagues, and Paul Smaldino and Richard McElreath support the discussion of publication bias, undisclosed flexibility, Registered Reports, transparency standards and incentives. Each reform attacks a particular failure route. Preregistration can distinguish prior plans from later exploration; it cannot make a poor question good. Registered Reports reduce result-contingent publication decisions; they do not guarantee execution or interpretation. Open data and code can permit checking while creating privacy, security, credit and infrastructure costs.

The UNESCO Recommendation on Open Science was adopted on 23 November 2021. It supplies an international framework covering access, infrastructure, participation and equity. The body does not claim uniform national implementation or use a changeable count of compliant institutions.

The historical sequence

Rochberg and Robson support the Mesopotamian account; Christopher Cullen supports the treatment of ancient Chinese mathematical astronomy; Kim Plofker supports Indian mathematics and astronomy. These traditions are presented as distinct combinations of calculation, observation, state service, ritual and learning. Similar techniques do not establish one shared programme or direct transmission unless historical contact is documented.

G. E. R. Lloyd supports the account of Greek proof, natural philosophy and medical argument, with the warning that Greek inquiry borrowed from and developed beside older traditions. The surviving Hippocratic corpus contains works from several authors and periods. Aristotle systematised demonstration and causal explanation while transmitting errors as well as methods. Euclid, Archimedes and Ptolemy represent different relations among proof, observation and mathematical modelling rather than one Greek method.

A. I. Sabra's edition and translation of Ibn al-Haytham's *Optics*, George Saliba and Ahmad Dallal support the Arabic-language section. Translation created new technical vocabularies and arguments. Ibn al-Haytham's optics joined geometry, experiments and criticism of emission theories of vision, but the book rejects the claim that he invented a complete, timeless scientific method. Observatories and astronomical models developed across several courts and centres. Specific routes from Islamic astronomy to Copernicus remain debated, so the body states only the secure larger claim that Renaissance astronomy encountered Greek material already translated, criticised and extended.

Kapil Raj, Londa Schiebinger and James Poskett support the account of knowledge moving through trade, empire, translation, collection and local expertise. Circulation did not occur on equal terms. Metropolitan institutions often controlled naming, publication and credit while depending on navigators, healers, artisans, collectors, enslaved people and Indigenous experts. This does not reduce every early modern result to empire or imply that borrowed material was unchanged by later work.

Copernicus, Tycho, Kepler, Galileo, Vesalius, Bacon, Boyle, Hooke, Newton and Lavoisier appear because their work reveals changes in mathematical architecture, instrument use, public witnessing, print, standards and laboratory practice. The section rejects a genius procession as a complete history. The early modern European reorganisation was decisive, yet it recombined older and global materials inside new institutions rather than creating exact knowledge from a void.

The accounts of nineteenth-century disciplines, state services, laboratories, museums, public health and industry are compressed. Professionalisation increased training and continuity while concentrating authority and creating new exclusions. Germ theory is treated as a network of microscopy, culturing, transmission, sterilisation, epidemiology and clinical change. The named figures are not credited with the labour of technicians, patients and institutions that made those links usable.

The Human Genome Project's data-sharing passage follows the United States National Human Genome Research Institute's history of the 1996 Bermuda Principles and rapid public release. It is used as one example of data policy becoming part of experimental design, not as proof that all genomics is open or that rapid release resolves consent, credit and computing inequality.

What People Get Wrong and Use It

The seven corrections synthesise claims documented above. The phrase *one scientific method* is rejected without abandoning standards. The global correction does not say that all knowledge traditions were interchangeable. The decisive-experiment correction preserves the possibility that a result can settle a narrow dispute. The peer-review correction preserves review's limited filtering value. The replication correction distinguishes exact repetition, reanalysis, direct replication and conceptual replication. The objectivity correction replaces personal neutrality with organised checks while keeping evidence distinct from preference. The self-correction correction treats correction as an achieved institutional performance, not an automatic property of the label science.

The practical lenses are diagnostic questions, not a validated scoring system. They can expose missing links, weak comparisons, range errors and unsupported decisions. They cannot substitute for field expertise or reconstruct evidence that was never preserved. The distinction among result, explanation and decision is especially important: data can constrain what is happening and what an intervention does without determining every legal, ethical or political choice.

Go Deeper

Publication details for the four recommendations were checked on 4 September 2026. James Poskett's *Horizons* was published by Viking in 2022. The recommended Bacon edition was edited by Lisa Jardine and translated by Michael Silverthorne for Cambridge University Press in 2000. Peter Godfrey-Smith's second edition of *Theory and Reality* was published by the University of Chicago Press in 2021. Naomi Oreskes's *Why Trust Science?* was published by Princeton University Press in 2019 and includes critical responses and a reply.

Bibliography

Primary sources and original evidence

ATLAS Collaboration. "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 (2012): 1-29. DOI: 10.1016/j.physletb.2012.08.020.

Bacon, Francis. *The New Organon*. Edited by Lisa Jardine and translated by Michael Silverthorne. Cambridge: Cambridge University Press, 2000.

Camerer, Colin F., Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. "Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015." *Nature Human Behaviour* 2 (2018): 637-644. DOI: 10.1038/s41562-018-0399-z.

CMS Collaboration. "Observation of a New Boson at a Mass of 125 GeV

with the CMS Experiment at the LHC.” *Physics Letters B* 716 (2012): 30-61. DOI: 10.1016/j.physletb.2012.08.021.

Event Horizon Telescope Collaboration. “First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole.” *The Astrophysical Journal Letters* 875 (2019): L1. DOI: 10.3847/2041-8213/ab0ec7.

Event Horizon Telescope Collaboration. “First M87 Event Horizon Telescope Results. IV. Imaging the Central Supermassive Black Hole.” *The Astrophysical Journal Letters* 875 (2019): L4. DOI: 10.3847/2041-8213/ab0e85.

Event Horizon Telescope Collaboration. “First M87 Event Horizon Telescope Results. VI. The Shadow and Mass of the Central Black Hole.” *The Astrophysical Journal Letters* 875 (2019): L6. DOI: 10.3847/2041-8213/ab1141.

Fisher, R. A. *The Design of Experiments*. Edinburgh: Oliver and Boyd, 1935.

Galilei, Galileo. *Sidereus Nuncius or The Sidereal Messenger*. Translated by Albert Van Helden. Chicago: University of Chicago Press, 1989.

Hooke, Robert. *Micrographia: Or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses*. London: Jo. Martyn and Ja. Allestry, 1665.

Ibn al-Haytham. *The Optics of Ibn al-Haytham: Books I-III, On Direct Vision*. Translated and edited by A. I. Sabra. 2 vols. London: Warburg Institute, 1989.

Kepler, Johannes. *New Astronomy*. Translated by William H. Donahue. Cambridge: Cambridge University Press, 1992.

Leavitt, Henrietta S., and Edward C. Pickering. “Periods of 25 Variable Stars in the Small Magellanic Cloud.” *Harvard College Observatory Circular* 173 (1912): 1-3.

Medical Research Council. “Streptomycin Treatment of Pulmonary

Tuberculosis.” *British Medical Journal* 2, no. 4582 (1948): 769-782. DOI: 10.1136/bmj.2.4582.769.

Open Science Collaboration. “Estimating the Reproducibility of Psychological Science.” *Science* 349, no. 6251 (2015): aac4716. DOI: 10.1126/science.aac4716.

Modern scholarship, methods and institutional sources

Baldwin, Melinda. “Credibility, Peer Review, and Nature, 1945-1990.” *Notes and Records: The Royal Society Journal of the History of Science* 69, no. 3 (2015): 337-352. DOI: 10.1098/rsnr.2015.0029.

Bureau International des Poids et Mesures. *The International System of Units*. 9th ed. Updated 2026. DOI: 10.59161/AUEZ1291.

Chambers, Christopher D. “Registered Reports: A New Publishing Initiative at Cortex.” *Cortex* 49, no. 3 (2013): 609-610. DOI: 10.1016/j.cortex.2012.12.016.

Cullen, Christopher. *Astronomy and Mathematics in Ancient China: The Zhou Bi Suan Jing*. Cambridge: Cambridge University Press, 1996.

Dallal, Ahmad. *Islam, Science, and the Challenge of History*. New Haven: Yale University Press, 2010.

Daston, Lorraine, and Peter Galison. *Objectivity*. New York: Zone Books, 2007.

Dear, Peter. *Discipline and Experience: The Mathematical Way in the Scientific Revolution*. Chicago: University of Chicago Press, 1995.

Douglas, Heather E. *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press, 2009.

Godfrey-Smith, Peter. *Theory and Reality: An Introduction to the Philosophy of Science*. 2nd ed. Chicago: University of Chicago Press, 2021.

Harvard College Observatory. “Henrietta Swan Leavitt.” Plate Stacks

historical research pages. Accessed 4 September 2026.

Ioannidis, John P. A. "Why Most Published Research Findings Are False." *PLOS Medicine* 2, no. 8 (2005): e124. DOI: 10.1371/journal.pmed.0020124.

Joint Committee for Guides in Metrology. *International Vocabulary of Metrology: Basic and General Concepts and Associated Terms*. 3rd ed. JCGM 200:2012. Sèvres: JCGM, 2012.

Lloyd, G. E. R. *Magic, Reason and Experience: Studies in the Origin and Development of Greek Science*. Cambridge: Cambridge University Press, 1979.

Longino, Helen E. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press, 1990.

Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, et al. "A Manifesto for Reproducible Science." *Nature Human Behaviour* 1 (2017): 0021. DOI: 10.1038/s41562-016-0021.

National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: National Academies Press, 2019. DOI: 10.17226/25303.

National Human Genome Research Institute. "Human Genome Project Fact Sheet" and "Human Genome Project Timeline." Institutional history of the Bermuda Principles and rapid data release. Accessed 4 September 2026.

Nosek, Brian A., George Alter, George C. Banks, Denny Borsboom, Sara D. Bowman, Steven J. Breckler, Stuart Buck, et al. "Promoting an Open Research Culture." *Science* 348, no. 6242 (2015): 1422-1425. DOI: 10.1126/science.aab2374.

Nosek, Brian A., Jeffrey R. Spies, and Matt Motyl. "Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth over Publishability." *Perspectives on Psychological Science* 7, no. 6 (2012): 615-631. DOI: 10.1177/1745691612459058.

Oreskes, Naomi. *Why Trust Science?* Princeton: Princeton University Press, 2019.

Plofker, Kim. *Mathematics in India*. Princeton: Princeton University Press, 2009.

Porter, Theodore M. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press, 1995.

Poskett, James. *Horizons: A Global History of Science*. London: Viking, 2022.

Raj, Kapil. *Relocating Modern Science: Circulation and the Construction of Knowledge in South Asia and Europe, 1650-1900*. Basingstoke: Palgrave Macmillan, 2007.

Robson, Eleanor. *Mathematics in Ancient Iraq: A Social History*. Princeton: Princeton University Press, 2008.

Rochberg, Francesca. *Before Nature: Cuneiform Knowledge and the History of Science*. Chicago: University of Chicago Press, 2016.

Ross, Sydney. "Scientist: The Story of a Word." *Annals of Science* 18, no. 2 (1962): 65-85.

Royal Society. Institutional histories of the Society, *Philosophical Transactions*, refereeing and the first women Fellows. Accessed 4 September 2026.

Saliba, George. *Islamic Science and the Making of the European Renaissance*. Cambridge, MA: MIT Press, 2007.

Schiebinger, Londa. *Plants and Empire: Colonial Bioprospecting in the Atlantic World*. Cambridge, MA: Harvard University Press, 2004.

Shapin, Steven. *The Scientific Revolution*. Chicago: University of Chicago Press, 1996.

Shapin, Steven, and Simon Schaffer. *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life*. Princeton: Princeton University Press,

1985.

Smaldino, Paul E., and Richard McElreath. "The Natural Selection of Bad Science." *Royal Society Open Science* 3 (2016): 160384. DOI: 10.1098/rsos.160384.

Stigler, Stephen M. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, MA: Harvard University Press, 1986.

UNESCO. *UNESCO Recommendation on Open Science*. Paris: UNESCO, 2021.

Ziman, John. *Real Science: What It Is, and What It Means*. Cambridge: Cambridge University Press, 2000.