

WAYS OF KNOWING

General Reader Edition

Twenty-four questions about how knowledge is built

Juan Segurola

Version v1.0 · 22 September 2026

© 2026 Juan Segurola. All Rights Reserved.

How to read this volume

Each chapter separates the central question, what the evidence supports, how the conclusion is built, what remains uncertain and what should not be inferred. The aim is accessibility without erasing the limits of the evidence.

Contents

KNO	Chapter	Page
KNO-001	A Mark on a Stone	4
KNO-002	Counting Before Numbers	6
KNO-003	From Marks to Symbols	8
KNO-004	From Symbols to Writing	10
KNO-005	Oral Language and Transmission	12
KNO-006	Maps and Spatial Representation	14
KNO-007	Calendars and Time Measurement	16
KNO-008	Instruments That Extend the Senses	18
KNO-009	Archives, Libraries and Collective Memory	20
KNO-010	Classification as a Knowledge Tool	22
KNO-011	Measuring Before Theorising	24
KNO-012	Experiments and Variable Control	26
KNO-013	Models, Maps and Reality	28
KNO-014	Causality and Correlation	30
KNO-015	Uncertainty and Probability	32
KNO-016	How Evidence Is Built: Putting the Tools Together	34
KNO-017	Is Mathematics a Language?	36
KNO-018	Images, Diagrams and Visualisation	38
KNO-019	Data, Tables and Records	40
KNO-020	Review, Consensus and Disagreement	42

KNO	Chapter	Page
KNO-021	Open Science and Reproducibility	44
KNO-022	Computation as External Memory	46
KNO-023	How Groups Know More Than Individuals	48
KNO-024	Lost Knowledge and Preservation Bias	50

KNO-001 · A Mark on a Stone

In brief

A groove is first a physical trace. Microscopy, fracture and experiments may show that it was deliberately made, but intention does not by itself establish a symbol. A row of notches may be regular without counting anything; a tally may record units without representing language; writing is a broader conventional system. Ancient shells and ochres show deliberate patterning deep in the Pleistocene, while objects such as the Border Cave and Ishango bones invite—not compel—notational interpretations. Modern cognitive research explains why an external mark can matter: it leaves a stable state that can be revisited, compared, rearranged or shared, reducing what must remain active in memory. Archaeology must independently show that a particular object served such a recurrent task. Shared conventions add another threshold because they let rules travel between people and generations. Preservation then filters the history: durable marked objects are visible, most perishable aids are not. The safest conclusion is modest but consequential: marks created new ways for states to persist, while their functions diversified through many local histories.

Why it matters

External records underpin measurement, accounts, diagrams, archives and science itself. Understanding their beginnings asks what evidence can support claims about minds that left no testimony. If every old line becomes a symbol, the story outruns the objects; if only deciphered writing counts, earlier material aids disappear from history. A calibrated middle course shows both the transformation and its limits. It also teaches a general lesson about evidence: observable traces, reconstructed actions and inferred meanings are different layers, and agreement among conclusions is not independent support when all depend on the same unidentified object.

What we know

Microscopy, chemistry, experiments and context can establish that many ancient marks were deliberately produced and can often reconstruct tools, gestures and sequence. Cognitive experiments show that external states can redistribute memory and comparison. Documented administrative and writing systems demonstrate later conventional uses. These conclusions rest on different evidence and should remain distinct.

What we don't know

For most early objects, we do not know what a mark meant, whether anyone revisited it, whether its rule was shared, or whether its units were numerical. We also cannot measure the vanished record on wood, fibre, skin or temporary surfaces. The earliest surviving example is therefore not necessarily the origin of the practice.

What could change the conclusion

Repeated marked objects from secure contexts, with demonstrable production sequences and stable associations to independently known tasks, could strengthen claims of notation or convention. Direct refits, new dating or trace analyses could also weaken existing interpretations. Exceptional preservation of organic records may reveal routes to external memory that durable collections currently hide.

One thing to remember

A deliberate mark can preserve a state beyond one brain, but symbol, tally, number and writing each require additional evidence.

KNO-002 · Counting Before Numbers

In brief

Quantity can be handled before a language has words for every exact number. An approximate number system distinguishes roughly more from less, while a separate capacity tracks only a few individual items. One-to-one correspondence answers a different question: does each object in one set have a partner in another? A tally can preserve repeated events without naming a cardinal total, and a finger or body sequence can provide an ordered scaffold before a portable numeral exists. Exact counting requires more than reciting a list. The words or signs must have a stable order, each item must receive one term, the final term must state the set's cardinal value, and the system must support extension beyond familiar collections. Clay tokens and numerical tablets show how particular Near Eastern institutions externalised quantities, but they do not establish one universal origin of number or writing. Archaeological marks can reveal manufacture, grouping and repeated sessions; they rarely reveal the counted category or reading rule. The safest history therefore reconstructs tasks and mappings, not a single ladder from notch to arithmetic.

Why it matters

Numbers now look abstract and inevitable because the same written forms move easily between clocks, prices, measurements and equations. Historically, however, those uses depend on several achievements that can come apart: recognising a difference in quantity, pairing sets, preserving a sequence, naming an exact total, grouping units, and operating on written values. Treating them as one invention hides both cognitive diversity and material evidence. Separating them also improves archaeological reasoning. A regular pattern may show controlled action, but the claim that it records seven animals or a lunar phase requires a recoverable mapping, not merely seven visible marks.

What we know

Controlled experiments distinguish approximate magnitude, small-set tracking and learned exact-number practices. Developmental studies show that reciting count words precedes full understanding of cardinality. Cross-cultural studies demonstrate exact nonverbal performance under some conditions and the importance of words or external devices when quantities must be retained and transformed. Archaeological and historical records document material counting and administrative numeral systems in specific contexts.[2,4-10,15-18]

What could change the conclusion

Repeated marked objects from secure contexts, associated with independently known countable processes and governed by the same demonstrable grouping rule, would strengthen early numerical interpretations. New residue, use-wear or refitting evidence could weaken them. Decipherable intermediate records could clarify transitions among tokens, impressed signs and written numerals. Broader cognitive research could also revise how approximate, exact and embodied systems interact.

One thing to remember

Quantity can be managed without numerals, but exact counting requires a learned mapping among units, order and cardinal value; a repeated ancient mark proves none of those mappings by itself.

KNO-003 · From Marks to Symbols

In brief

A scratch is not a symbol merely because it is visible or deliberate. Smoke can indicate fire through a causal relation; a portrait can resemble a person; a pointing gesture can direct attention; a conventional sign can refer even when it neither resembles nor is caused by its referent. These relations can overlap. A map icon may resemble an object and also acquire a conventional meaning. Experimental studies show that people invent signs more successfully when they can interact, test interpretations and build shared precedents. With repeated use, detailed drawings often become simpler and less transparent to outsiders because part of their meaning moves into the users' common history. Cultural transmission then favours forms that can be learned and reproduced, while communication favours distinctions that remain useful. New sign languages show that systematic structure can emerge across cohorts, but they are living languages in specific communities, not replicas of prehistory. Archaeology can identify manufacture, repetition and patterned placement; it cannot usually recover a vanished community's referent or reading rule from form alone. A symbol is best understood as a maintained relationship, not a special kind of mark.

Why it matters

Humans inhabit systems of traffic signs, diagrams, gestures, flags, numerals, interfaces and words. Their physical forms are often trivial compared with the coordination they support. Confusing a visible pattern with a recovered meaning leads to two errors: declaring every repeated ancient mark symbolic, or assuming that arbitrary modern symbols appeared fully formed. A better account asks which relation made a sign interpretable, who shared it, how misunderstanding was repaired, and how the convention survived new users who never met the people who invented it.

What we know

Developmental studies distinguish an object's physical identity from its representational role and document early joint attention and pointing. Interaction experiments show that feedback and common ground can stabilise novel signs. Transmission experiments show that learnability and communicative usefulness reshape systems across learners. Emerging sign languages demonstrate structural change in living communities, while archaeological analysis establishes manufacture and patterning under much stricter limits on meaning.[1-21]

What could change the conclusion

New sites preserving repeated marked objects within tightly resolved activity contexts could strengthen claims of shared convention. Refitting, residue, wear or production-sequence evidence could instead show that apparently similar marks served different tasks. Larger longitudinal studies of new communication systems could separate effects of interaction, network structure and transmission. Experiments that successfully predict which conventions generalise to uninvolved learners would improve the bridge from local grounding to public symbolism.

One thing to remember

A mark becomes a symbol not through its shape, but because a community can repeatedly use, recognise, transmit and correct a relation between that form and what it stands for.

KNO-004 · From Symbols to Writing

In brief

A mark can be deliberate without being a symbol, and a symbol system can be powerful without encoding language. What converts a symbol system into writing, under the definition used here, is phonetic transfer: the discovery that a sign can stand for the sound of its name rather than for its meaning. Before that step a system grows only by adding signs; after it, the inventory stabilises and then contracts while the range of expressible content stops having a limit. Southwest Asia preserves the only stratified sequence in which tokens, sealed envelopes, impressed numerical tablets and proto-cuneiform can all be examined together, although that order records first attestation and not descent. The Indus corpus shows why the threshold matters: the dispute there is not about age or sophistication but about whether the signs encode language at all, and it remains open. Numerical, mathematical and musical notations show that a closed system can be unlimited within its domain and encode no language, not because it fell short but because it was never trying. The Cherokee and Vai syllabaries show that once the idea of writing is available, a working system can be built in a few years.

Why it matters

Writing is routinely treated as the boundary between prehistory and history, so what counts as writing decides which societies are said to have entered the record at all. Reading every repeated mark as script inflates the story past its objects; admitting only deciphered scripts erases the recording systems that carried accounts and quantities without encoding language. The distinction also carries a general lesson about evidence: what a system could do, what it was for, and where it came from are three separate claims, and establishing one of them does not establish the others.

What we know

Stratified fourth-millennium material from southwest Asia preserves the only sequence in which tokens, sealed envelopes, impressed numerical tablets and proto-cuneiform can be examined together, though that order is one of first attestation rather than descent. Documented modern inventions establish the pace: the Cherokee syllabary was in use by the early 1820s and received council recognition in 1821, and the Vai syllabary followed within about a decade. Numerical, mathematical, chemical and musical notations show that a closed system can be unlimited within its domain and still encode no language at all.

What could change the conclusion

A bilingual, a known descendant language or substantially longer texts would move the Indus question from method to evidence. Excavation in a medium that preserves could reveal a long development behind an apparently sudden script, although that bias runs in one direction only and can hide a development that happened rather than manufacture one that did not. Accumulating evidence of long-distance contact would weaken the independence claims, which rest on opportunity and not on demonstration.

One thing to remember

Writing begins when a sign can carry the sound of its name rather than its meaning; every other recording system is an alternative rather than a failed attempt.

KNO-005 · Oral Language and Transmission

In brief

Oral transmission is not one practice but a family of techniques, and they are not comparable technologies. Casual report degrades quickly, because nothing in the practice resists that. Institutionally checked recitation is the strongest oral fidelity regime known. What produces the difference is constraint, not effort. A line of verse in a strong tradition is held at once by metre, by rhyme or alliteration, by melody, by the imagery of its scene, by the formulaic phrasing of its metrical slot and by narrative sequence. Each alone leaves many options; together they often leave one, so a performer who cannot recall a line regenerates one that fits. Three consequences follow: fidelity is domain specific, unconstrained prose is the fragile case, and fidelity and creativity are settings rather than opposites. At one pole the South Slavic singer builds a different performance each time, same plot and different words, and does not experience it as failure. At the other, specialists recite a text in systematically permuted orders, so that a corruption is exposed by disagreement between the recitation forms: redundancy used for error detection, without any theory of information.

Why it matters

A prominent mid twentieth century argument held that alphabetic literacy restructures thought, enabling logic, history and scepticism oral cultures cannot sustain. Comparative work replaced that autonomous model with one in which what literacy does depends on the practices it is embedded in, so no general consequence follows from the technology alone. Treating oral transmission as a stage flatters writing, crediting it with solving a problem the strong oral regimes had already solved, and obscures what writing did change: cost, portability, random access and the recording of the arbitrary. And each loss takes with it whatever was held only in one language and only in memory.

What we know

Cognitive work on strong traditions establishes that stability comes from many individually weak constraints acting at once — metre, melody, formula, imagery and narrative sequence — which together often leave one option where each alone leaves many. Parry and Lord's fieldwork shows a singer producing the same plot at similar length in substantially different words without treating it as failure. Vedic permutation recitation shows a community making error detectable rather than merely improbable.

What could change the conclusion

Independent versions of a tradition recorded before it was disturbed would move the verbatim question from method to evidence. A method able to distinguish an old element from a well integrated new one inside a living tradition would do the same for the age of content. The phylogenetic results are model dependent, and other codings argued in that literature could revise how deep the recoverable signal reaches.

One thing to remember

Strong oral fidelity is produced by constraining what can be said rather than by trying harder to remember it, which is why it favours verse and fails on arbitrary detail.

KNO-006 · Maps and Spatial Representation

In brief

Spatial knowledge, external spatial representation and cartography are three separable things, and an expert navigator may hold the first at a high order while using no graphic artefact during the voyage. Place cells and grid cells are robust findings about neural activity, but they do not establish that human spatial thought takes the form of a graphic map; route knowledge can support successful travel without being a degraded survey map. Marshallese wave charts and the Caroline Islands technique known as etak show that spatial reasoning can be organised around models of ocean behaviour and around reference frames other than the fixed frame of a scaled plan. Geometry then imposes a hard limit: because a sphere and a plane differ in Gaussian curvature, no flat map of a sphere can preserve everything, so a projection encodes a decision about which distortions to accept. Separately, every map selects, so selection, scale, orientation and classification are editorial choices rather than neutral reporting, and Web Mercator shows the geometric constraint persisting in interactive media.

Why it matters

A map can be read as a transparent account of space. A road atlas, a weather chart or a world map on a classroom wall is read as a picture of what is there, and the reading is usually harmless until the map is used to compare sizes, infer distances or settle an argument about the world. Knowing that every flat map of the globe distorts something, and that the kind of distortion it carries follows from a choice rather than a mistake, changes what a reader can responsibly conclude from one. Separating maps from spatial knowledge also changes which spatial systems count as knowledge. ---

What we know

- Grid cells are an established finding from single-unit recording in rats (Hafting et al. 2005). O'Keefe and Nadel proposed that the hippocampus supports an allocentric representation (O'Keefe and Nadel 1978). - Local isometries preserve Gaussian curvature, the Theorema Egregium (do Carmo 1976, p. 234). From it this article deduces that no open region of a sphere maps isometrically onto a plane. - Mercator's navigational value and its later non-navigational use are both documented (Monmonier 2004). - Web Mercator's dominance in online mapping, its technical advantages and its limitations are documented, without one exclusive cause of adoption being established (Battersby et al. 2014). - Comparative volumes cover Islamic, South Asian, East and Southeast Asian, African, American, Arctic, Australian and Pacific traditions in their own settings (Harley and Woodward 1992; 1994; Woodward and Lewis 1998).

What could change the conclusion

Independent contextual evidence of use, rather than resemblance, could settle several disputed identifications in either direction. Behavioural or neural work showing that survey style representation is the default format of human spatial thought, rather than one format among several, would weaken the article's claim that the survey style map is not the natural externalisation. Historical evidence isolating why Web Mercator was adopted would convert a list of plausible factors into a causal account. The geometric constraint of section 5 is not open to this kind of revision.

One thing to remember

No flat map of a sphere preserves every metric property, so the question to ask of one is not whether it distorts but which distortions it accepts.

KNO-007 · Calendars and Time Measurement

In brief

Twelve mean synodic months come to 354.36708 days, 10.87511 short of a tropical year, so a lunar calendar walks through the seasons and the Islamic calendar does that by design rather than by failure. Drop the lunar month and keep the year and the Egyptian civil year follows: twelve months of thirty days plus five epagomenal days, 365 in total, with no intercalated day and a slower drift of its own. Keep both and the price is intercalation, because nineteen tropical years and 235 synodic months differ by 0.08704 days, about 2.09 hours, which is close enough to build a nineteen year cycle on and not close enough to make it exact. The Julian year overshoots by 0.00781 days and the Gregorian by 0.00031, so 1582 reduced a residual rather than removing one, and it revised the lunar tables for Easter as well. Meanwhile the second stopped being astronomical in 1967 and civil time did not.

Why it matters

Read as a ladder, this history becomes a march towards the Gregorian year with everything else a failed attempt at it, and the actual question disappears. A lunar calendar is not an inaccurate solar one, and a 260 day count is not an inaccurate year: thirteen and twenty day cycles have no direct astronomical basis to be inaccurate about. A calendar is also a coordination device, which is why reforming one is never only a calculation, and why the 2022 decision to let atomic and rotational time drift further apart was argued on the risk to critical digital infrastructure.

What we know

For epoch 2000 a standard astronomical reference gives a mean tropical year of 365.24219 days and a mean synodic month of 29.53059 days, and stresses that both are approximate means that change slowly. Everything this article derives from them is its own arithmetic, marked as such: twelve synodic months fall 10.87511 days short of a year, and nineteen tropical years differ from 235 synodic months by 0.08704 days. The Egyptian civil year ran 365 days with no intercalated day, and the Islamic calendar keeps months in relation to lunations without keeping its year with the Sun. Central American thirteen and twenty day cycles have no direct astronomical basis. And the SI second has been atomic since 1967, its present wording fixing the unperturbed caesium 133 ground state hyperfine transition at 9,192,631,770 Hz.

What could change the conclusion

Independent evidence of use, rather than resemblance to a modern calendrical function, would strengthen or break the Warren Field reading, and a published critique addressing it directly would change what this article can say about how contested it is. For the incised bones, a method that separated lunar counting from other tallying and from non notational patterning would settle what the marks alone cannot. Direct reading of the four delegated sources would move them off partial status. And the tolerance the CGPM was directed to propose will replace an open value with a decided one.

One thing to remember

A lunar calendar is not a failed solar one: astronomical periods do not fit exactly, and calendars that track them must choose how to handle the mismatch.

KNO-008 · Instruments That Extend the Senses

In brief

Spectacles aid a sense, a telescope extends one, and a thermometer does neither: it converts a quantity no eye registers into a length the eye can read, and the observer infers. Believing such a reading is believing every link of that chain, and no point along it is settled by looking. Galileo's telescope was accepted through independent replication rather than on sight. Calibration is circular at the outset, because justifying an instrument needs the law that measuring with it is meant to establish; that circle was broken by iteration rather than by decree, and a published critique holds that reasoned argument can break it. Longitude was not solved by one clock: lunar distances were treated as co equal in 1767, rested on dead reckoning between observations, and declined when chronometers grew cheap. Radio astronomy began as an attempt to reduce noise. A gravitational wave detection ends in no sensation at all, and its first signal came from a burst search rather than from template matching. What makes two readings comparable is an institutional chain, and the vocabulary that defines it warns what it does not guarantee.

Why it matters

A reading is offered as if it were an observation, and it is the end of a chain that began where no sense reaches. Taking it as perception hides every place the chain can be wrong: the substance coupled to the quantity, the calibration behind it, the model that turns a correlation into a claim. The distinctions here let a reader ask what validated the instrument, whether agreement among copies is worth anything when they share a design, and whether the story of how a problem was solved is the story the record actually supports.

What we know

Experts at Bologna reported seeing little or nothing through Galileo's telescope in April 1610, the Collegio Romano withheld confirmation until October, and credibility was established by independent replication; the source for this was read only as far as its accessible pages go. The circularity of validating a measuring instrument without the law that justifies it is called the problem of nomic measurement and is attributed to a 2004 study of thermometry, which argues the circle was worn away by iterative self correcting refinement rather than broken by decree. Franklin argues from the gravitational wave episode that reasoned argument broke the regress. The University of Chicago Press 1992 edition titles its chapter 4 with the phrase *Detecting Gravitational Radiation: The Experimenter's Regress*, which the human authority read in the contents and nobody read further. The Board of Longitude treated chronometers and lunar distances as co equal in 1767; the *Endeavour* voyage combined lunar work with dead reckoning, which could not correct for drift or currents. Jansky's result is identified by the observatory's institutional history as radio waves from the central region of the Milky Way, and the primary report records the solar hypothesis it displaced. The first reported direct gravitational wave observation gives a peak strain of 1.0×10^{-21} , and the initial detection was made by a burst search, with PyCBC and GstLAL recovering the event afterwards. Metrological traceability is defined as a property of a measurement result related to a reference through a documented unbroken chain of calibrations.

What could change the conclusion

A reading of Collins's chapter would settle what that work argues about the experimenter's regress, as opposed to that it names it. A source whose text can be read, on the reception of Galileo's telescope, would replace a claim currently resting on a publisher's page and a separate essay. Paginated evidence that lunar distances were or were not the primary method on Cook's voyages would decide a question this article leaves open in both directions. Reading the published 2012 preface would convert a presumption about the vocabulary into a check. And an instrument of an entirely different design agreeing with one of these results would do more for it than another copy of the same design agreeing, which is the point section 2 makes and which applies to this article's own sources.

One thing to remember

An instrument rarely extends a sense; it ends a chain of conversions, and believing the reading means believing every link.

KNO-009 · Archives, Libraries and Collective Memory

In brief

Records, archives and libraries are filed together and run on incompatible principles, so each is silent about something different: an archive about people who generated no records, a library about what never entered the circuits it collects from, a record about whatever its purpose did not require. What enters an archive is decided by appraisal, whose own professional framing is what to save and what to destroy, and the surviving series reads as complete because destruction leaves no gap in it. Trouillot separates four moments where silence enters, which matters because they are four failures with four remedies and better searching fixes only some. Description then governs what can be retrieved, which is a constraint and not a second appraisal, and it can be redone where appraisal cannot. Alexandria is the case where ordinary decay and withdrawn institutional funding suffice to account for the disappearance without a hostile act, which is Bagnall's argument and is given here as his. The digital case makes the same structural point in modern form: a collection is not a thing that persists unless destroyed but a process that stops unless maintained, and the expenditure becomes continuous rather than periodic.

Why it matters

Treating an archive as a container makes its holdings a sample of the past and its gaps an accident. Both are wrong in a way that changes what can be concluded from it. The holdings are the output of a filter that does not describe itself, so from the holdings alone an absence is not evidence of absence and a survival is not evidence of typicality. That is why reading an archive well is not a matter of searching harder, and why a catastrophe, which is easy to narrate, tends to displace a withdrawal of funding, which is not.

What we know

Appraisal is defined by the professional dictionary as identifying materials offered to an archives that have sufficient value to be accessioned, with the debate behind it framed as what to save and what to destroy. Archival theory has described the archivist moving from passive curator to active appraiser across shifting paradigms. Trouillot locates silence at four moments: fact creation, fact assembly, fact retrieval and retrospective significance. Archives have been read as sites of knowledge production and as monuments of states. A general international standard for archival description was issued in a second edition in 2000. And a study of over one million references from more than three and a half million articles published between 1997 and 2012 reports that one in five articles in science, technology and medicine suffers reference rot.

What could change the conclusion

External evidence about an institution would separate destruction from neglect, cost and censorship, which the archive itself cannot. Documented appraisal decisions, in disposal schedules and appraisal reports, recover the fact that a decision was taken even where the destroyed content is unrecoverable. Reading the official ISO and CCSDS registers directly would move those two entries off their inherited status. And resolving the reference disagreement would settle which proposition Bagnall supports.

One thing to remember

An archive's surviving holdings cannot reveal all their own omissions: what survives is evidence about the keeping as much as about the kept.

KNO-010 · Classification as a Knowledge Tool

In brief

Classification looks like description and is not. A scheme is a set of boxes built so that some further operation can run, and its categories were shaped by that operation. Bowker and Star name three properties an ideal classification would have and report that none of the working systems they examined satisfies all three. A scheme that is not complete has to place what its categories did not anticipate, and where a residue sits tells a reader how that scheme handles it. Three operations hide inside the one word: grouping, naming and identifying, which fail differently and are repaired differently. Biology supplies the sharpest case, a species category that is used everywhere and defined by no agreed rule, and a rank system whose levels are not comparable between beetles and primates. Folk schemes are not earlier drafts of scientific ones; they answer a different question. When the things sorted are people, the ordinary properties acquire consequences, and two documented cases show a library scheme carrying its period's assumptions forward and a state scheme deciding what a person was. Revision costs money now and comparability afterwards.

Why it matters

Classifications are not consulted, they are inhabited. A diagnosis code decides what a death is counted as, a shelf mark decides what a reader finds, a population category decides what a state does to a person. Because the categories were shaped by the operation they had to support, a scheme read as a report on how things are divided lends the world a structure that came from the work. That misreading is cheap to make and expensive to hold: it turns a revisable convention into a fact, and it hides the decision that a category records.

What we know

Bowker and Star define a classification system as a set of boxes into which things can be put to then do some kind of work, set out three properties an ideal classification would have, and report that none of the working systems they examined satisfies all three. They observe that the residual phrase runs throughout the ICD but never as a top-level category, and they name torque for the mutual twisting of a biography and a process of metrication. Wells and colleagues report that species remain controversial in definition, delimitation and reality, and propose treating them as a heuristic. Lee and Skinner report no generally accepted criteria for rank; Benton argues that ranks still do usable work. Atran reports Berlin's folk ranks and their partial correspondence with scientific taxa. Thornton traces the Dewey treatment across four editions and 130 years, Posel documents official latitude in South African racial classification, the Seventy-second World Health Assembly adopted ICD-11 in 2019, and Janssen and Kunst found discontinuities remaining in 10.8 per cent of code revisions across six countries.

What could change the conclusion

An operational criterion for rank that held across groups would turn Benton's practical defence into a stronger one, and its continued absence is what Lee and Skinner report. Ethnobiological work on groups whose environments differ sharply from those already studied would test whether the recurrence Atran describes is as general as it looks. Concordance work that separated coding

discontinuity from real change, revision by revision rather than in aggregate, would say how much of a long series is the scheme. And a classification that documented its own residual traffic would let a reader see the strain instead of inferring it.

One thing to remember

A classification is built to make an operation possible, so its categories cannot be read as a neutral picture of the world.

KNO-011 · Measuring Before Theorising

In brief

Two orders both exist and neither is the order this field follows: sometimes theory names a quantity and an experiment goes looking for it, sometimes the numbers accumulate until a theory has to reproduce them. Whether classifying already counts as measuring depends on which definition is in force, and the permissive and the restrictive answer are both named here and neither adopted. An instrument has to be trusted before anything exists that could check it, and in the episodes discussed the circle was not avoided rather than broken. A formula can be written to hit a target that measurement had already set: a communication about long wavelength behaviour came first and the equation followed, which is an order of events and not an account of what caused what. A residual reports the size of a theory's failure and binds only successors covering the same observational domain, at the precision that matters there. In 2019 a measured constant became a defined one, and the defining value has no uncertainty at all; what carries uncertainty now is the realisation of the unit.

Why it matters

A number offered as a measurement does not explain itself. Read without asking what validated the instrument, whether the quantity was theorised first, and whether the figure is a value or a residual, the same number will be made to support conclusions it cannot carry. The distinctions here are what let a reader see that a formula was answerable to data before it meant anything, that a small stable correction can outlast every attempt to explain it away, and that a constant can stop being measured by decision rather than by discovery.

What we know

The permissive definition takes measurement to be the assignment of numerals to objects or events according to rules, and separates four scale types by the empirical operation each supports. The restrictive objection is that under such a view any rule at all already defines a nominal scale, so whether the attribute is quantitative is never put to test. Competing radiation laws differed only at long wavelengths; a residual ray experiment across four cavities found none of them reproducing the data at all temperatures; a communication about the limit at long wavelengths preceded the equation that worked, and the two results were presented at the same meeting. Le Verrier's correction to the secular motion of Mercury's perihelion is recorded by Laskar as roughly thirty eight arcseconds per century from twenty one transit observations, against roughly forty three from the later theory. Resolution 1 defines the kilogram by fixing the Planck constant, the seven defining constants have no uncertainty, and that resolution's own Appendix 2 puts the prototype at 1 kg within the inherited relative standard uncertainty, determined experimentally from then on. Chang's study is used only as its publisher describes it; its body was not read here.

What could change the conclusion

A reading of Chang's argument rather than its publisher's description would settle what that study establishes about breaking the circle. A demonstration that some measurement standard was validated without ever relying on instruments it later judged would narrow the claim section 3 declines to

generalise. On the residual, a successor theory covering the same observational domain and failing to reproduce the figure at the relevant precision would matter; one covering different ground would not. And a metrological decision to unfix a defining constant would return it to being measured.

One thing to remember

Numbers often arrive before the theories that explain them, so ask what a measurement was answerable to before asking whether it is true.

KNO-012 · Experiments and Variable Control

In brief

In an assigned comparative design, an experiment buys one thing: the investigator decides who receives what, so the reason one group got the treatment has a known answer instead of an unknown one. Experiments that assign nothing lie outside what follows. Everything else in the method protects that property. The decision itself is therefore the vulnerable point, and the word randomisation covers two different jobs at it: generating a sequence nobody can predict, and concealing that sequence from whoever enrolls participants. Reporting standards number them separately, and trials judged inadequately concealed report larger effects than trials judged adequately concealed. Concealment stops at the moment of assignment. Blinding is a different instrument, it is not always available, and who was blinded matters more than whether the word appears. Where no one can assign anything, the comparison has to be found instead, and the claim that its variation is as if randomised is an assumption about the world rather than a fact about the design. Hardest of all is a contested result, where judgements about competence and judgements about the answer move together.

Why it matters

The protections discussed here can be the difference between a comparison whose warrant is defensible and one whose apparent result outruns its design. They are also invisible in the result: a trial reports an effect size the same way whether its sequence was concealed or not, and the number that comes out carries no mark of the arrangement that produced it. Reading an experiment well therefore means reading what was done before the data existed, which is the part a published result is least likely to show and most likely to depend on.

What we know

Fisher titled section 9 of *The Design of Experiments* Randomisation; the Physical Basis of the Validity of the Test, and argued there that the simple precaution of randomisation suffices to guarantee the validity of the test of significance. That is a claim about the validity of a test and not about the balance of any particular experiment, and the surrounding context also makes clear that validity is not a claim that randomisation removes every other avoidable source of error. The CONSORT 2010 checklist carries sequence generation, allocation concealment mechanism and blinding as three separately numbered items. Schulz and colleagues, comparing trials by the adequacy of their concealment, report odds ratios exaggerated by 41 per cent for inadequately concealed trials and by 30 per cent for unclearly concealed ones. The Cochrane Handbook says concealment prevents bias before and until assignment and, within the randomised trials the handbook describes, can always be implemented, that blinding prevents bias after assignment and cannot, and that blinding of outcome assessors is considered separately. The Nobel Committee's background document states that many important questions cannot be studied by controlled randomised experiment, and defines a natural experiment. Collins states the regress in a passage Franklin quotes, and Franklin holds it can be broken by reasoned argument.

What could change the conclusion

A comparison that sorted trials by the adequacy of sequence generation rather than of concealment would say something about the half of the pair this evidence leaves untouched. Concealment studies outside the randomised trials the Cochrane Handbook describes would test how far its always extends. For a natural experiment, evidence closing the routes the definition leaves open, selection by a party other than the subjects, anticipation of the event, and systematic differences the event tracks rather than creates, would move the argument towards the procedural warrant it is standing in for.

One thing to remember

In the assigned comparative designs treated here, assignment replaces the unknown reason why these units and not those with a procedure, and the rest defends that substitution.

KNO-013 · Models, Maps and Reality

In brief

A model represents a chosen piece of the world, not the world, so asking whether it is good means asking what it was built for. Leaving something out is not the same as getting it wrong: an abstraction omits, an idealisation distorts on purpose, and only the second asserts a falsehood deliberately. The test is intent, because what a modeller would fix on noticing is an error and what they would defend has to earn its place. One developed position holds that the most unifying laws are the least descriptive, recorded here and not adopted. A constraint from 1966 says generality, retained detail and precision cannot be maximised together in a manageable model; an objection denies that the trade is necessary, not that trades occur. The famous sentence about models being wrong sits in a different document from the one usually cited. Whether a model's parts are real is left open, but fit never accredits itself.

Why it matters

A model in use is not a small copy of anything, and the useful question about one is not whether it is true but what it was built to do and what it gave up doing it. Read without that, a model is either dismissed for the parts it never claimed or trusted for reach it never had. The distinctions here are what let a reader ask which falsehood was chosen, what it bought, and whether a memorable formulation is carrying weight its source does not.

What we know

The reference formulation states that many scientific models are representational and stand in for a selected part or aspect of the world, and that idealised models involve a deliberate simplification or distortion aimed at tractability. Cartwright is reported, from the publisher's description of the book rather than from its argument, as holding that fundamental explanatory laws do not describe regularities in nature while localised laws and theoretical entities may still be read realistically. Levins wrote in 1966 that generality, realism and precision cannot be maximised together in a manageable model, with realism there meaning retained detail. Orzack and Sober reject the necessity of that trade-off. The 1976 paper argues against elaboration and a direct check confirms the famous phrase is absent from it; the literal wording appears in a 1979 technical report. The reference entry gives scientific realism as belief in observable and unobservable content, and van Fraassen's empirical adequacy, quoted within that same entry, as truth about the observable.

What could change the conclusion

A reading of Cartwright's argument rather than its publisher's description would settle whether the position is as stated. A bibliometric study would turn the misattribution from an observation into a measured frequency. On the constraint, a case in which all three properties improve together would not refute Levins as he framed it, but a demonstration that they must trade in some specified class of models would narrow the objection. On realism, an argument that success licenses belief without assuming what it sets out to show would move a question this article leaves open.

One thing to remember

A model is a deliberate misdescription with a job, so ask what job, what was left out, what was distorted on purpose, and what that cost.

KNO-014 · Causality and Correlation

In brief

Association alone does not tell us which causal story generated a pattern. A cause may produce an outcome, a common cause may produce both variables, the outcome may influence the supposed exposure, or selection and measurement may create the dependence. Causal inference therefore begins by defining the intervention or contrast of interest and asking what design makes that contrast identifiable. Randomisation can protect against baseline confounding when properly implemented. Observational studies can also support causal claims, but only through explicit assumptions about time, causal structure, adjustment, measurement and selection. Directed acyclic graphs help state those assumptions; propensity scores address balance in measured covariates; instrumental variables substitute different assumptions; negative controls test some bias pathways; target-trial emulation aligns timing and design. None is a causality machine. The practical discipline is to expose the assumptions that connect an observed association to a counterfactual or interventional claim, then stress-test those assumptions rather than treating statistical precision as causal proof.

Why it matters

Medicine, engineering, economics, climate science and public policy often need to know what would happen if an exposure, treatment, process variable or policy were changed. An association can motivate that question but cannot answer it by itself. The difference matters because prediction and intervention are not the same task. Understanding the inferential bridge from correlation to causation helps readers see why design, timing, causal assumptions and bias diagnostics are part of the evidence rather than technical details added after the result.

What we know

Association and causal effect are distinct targets. Randomisation, causal diagrams, potential outcomes, negative controls, instrumental variables and target-trial emulation each make different parts of the causal problem inspectable. Their evidential force comes from design plus assumptions, not from the statistical estimate alone.

What we don't know

Observed data cannot by themselves establish causal direction, the absence of unmeasured confounding, a valid exclusion restriction or the correctness of an assumed causal graph. Those questions require external knowledge, design information and tests whose scope is itself limited.

What could change the conclusion

Confidence in a causal claim should change if new evidence reveals hidden confounding, a violated time ordering, measurement bias, failed negative controls, poor overlap, an invalid instrument or a target-trial specification that no longer matches the real data-generating process.

One thing to remember

Causal inference begins when an association is linked to a clearly defined intervention question through explicit, testable design and assumptions.

KNO-015 · Uncertainty and Probability

In brief

Uncertainty is not one thing. A measurement can be uncertain because of calibration, repeatability or resolution; an estimate can vary across samples; model parameters can be imprecise; structural choices can change the result; future observations can add prediction uncertainty; and some unknowns lie outside the probability model entirely. Confidence intervals and Bayesian credible intervals answer different questions because they rely on different inferential frameworks. A p-value is not the probability that a hypothesis is true. Propagating uncertainty through a model is necessary when uncertain inputs affect outputs, but propagation cannot represent omitted mechanisms or an inappropriate model structure. Sensitivity analysis asks how conclusions change when assumptions or inputs vary, making hidden dependencies visible. Expert judgement can be structured, documented and calibrated, but it does not become observational data. Good uncertainty practice therefore matches each representation to the claim, separates quantified variability from structural ignorance, and states what the chosen analysis does not cover.

Why it matters

Scientific decisions are often made under uncertainty, but different uncertainties call for different responses. A narrow confidence interval can coexist with model misspecification; a precise instrument can measure the wrong quantity; a probability can be well calculated while excluding the most important unknown. Treating every uncertainty as interchangeable produces false precision and misleading comparisons. Distinguishing what is measured, what is sampled, what is modelled and what remains outside the model is essential for interpreting evidence responsibly and for communicating how strongly a conclusion should guide action.

What we know

Uncertainty can arise from measurement, sampling, parameters, prediction, model structure and judgement, and these sources are not interchangeable. Statistical intervals and probabilities have meanings fixed by their inferential framework, while sensitivity analysis exposes dependence on assumptions.

What we don't know

No finite interval automatically contains every relevant unknown. Structural omissions, unmodelled mechanisms and genuinely unforeseen possibilities can remain outside a probability model even when the arithmetic inside that model is correct.

What could change the conclusion

A conclusion should be revised when calibration changes, new data widen or narrow sampling uncertainty, alternative model structures fit comparably well, sensitivity analyses reveal fragile assumptions, or previously ignored mechanisms become plausible enough to affect the target quantity.

One thing to remember

Represent uncertainty in the form that matches the claim, and state explicitly which important unknowns the calculation does not contain.

KNO-016 · How Evidence Is Built: Putting the Tools Together

In brief

Evidence becomes stronger when the links between claim, observation, design, model and inference are explicit enough to test. A precise claim must be vulnerable to failure. Measurements are produced by instruments and procedures, so their calibration and transformation matter. Study design determines which alternative explanations can be excluded; models connect observations to quantities or mechanisms; causal claims require assumptions beyond association; uncertainty belongs inside the result rather than after it. Replication and reproducibility probe different links in the chain. Triangulation is most informative when methods fail in different ways rather than sharing the same bias. Direct evidence is usually easier to interpret than a long chain of proxies, but indirect evidence can still be strong when each link is independently supported. Transparency makes those links inspectable; peer review and consensus can challenge them but cannot replace them. Evidence is therefore an architecture of support whose strength depends on the weakest relevant links, not simply on how many studies or citations exist.

Why it matters

Scientists rarely observe a conclusion directly. They observe traces, measurements or outcomes and connect them to a claim through design, models, assumptions and background knowledge. When those links remain implicit, disagreement becomes difficult to diagnose and replication failures become hard to interpret. Making the evidential chain explicit clarifies where a result is strong, where it depends on assumptions and what new observation could change it. That matters not only for research practice but also for anyone deciding how much confidence to place in a scientific claim.

What we know

Evidence is built by linking precise claims to observations through measurement, design, models and inferential assumptions. Replication, reproducibility, triangulation, directness and transparency strengthen different links, especially when their failure modes are not shared.

What we don't know

No audit can guarantee that every relevant alternative has been imagined, that every source of bias is independent or that a well-supported model is the only possible explanation. Evidence can therefore become stronger without becoming logically final.

What could change the conclusion

The evidential assessment should change when a key measurement fails calibration, a design permits an alternative explanation, a model prediction breaks, a replication fails under genuinely comparable conditions, or a new method with different error structure contradicts the existing result.

One thing to remember

Strong evidence is not a pile of results; it is a chain of support whose assumptions, error modes and tests can be inspected.

KNO-017 · Is Mathematics a Language?

In brief

Mathematics contains formal languages, but mathematical practice is broader than symbol strings. Formal notation has syntax: rules determine which expressions are well formed and which transformations are permitted. Meaning enters through interpretation, definitions, conventions and connections to mathematical structures or empirical systems. Proof is not merely the presence of symbols; it is a rule-governed argument whose validity depends on the system in which it is read. Notation can still transform cognition by making relations easier to inspect and manipulate. Abstraction lets the same structure apply across very different domains, while calculation can produce correct outputs without supplying an explanation of why a model is appropriate. In science, mathematics can represent relations with unusual precision, yet its success does not by itself decide whether mathematics is discovered, invented or literally a language. The useful conclusion is conditional: language is a powerful analogy for some mathematical components, but not a complete account of mathematical knowledge.

Why it matters

Calling mathematics a language can illuminate how notation, syntax and shared conventions make reasoning portable, but the metaphor can also conceal important differences between formal manipulation, proof, meaning and empirical interpretation. The issue matters because mathematics is central to nearly every quantitative science, and its symbols often look more self-sufficient than they are. Understanding what notation contributes—and what it cannot supply by itself—helps separate mathematical validity from modelling adequacy, calculation from explanation and formal precision from claims about the world.

What we know

Mathematics uses formal languages with syntax and interpretation, and notation can make reasoning more inspectable and computationally tractable. Proof, meaning, modelling and explanation nevertheless depend on practices and interpretations that are not exhausted by symbol strings.

What we don't know

The language analogy does not settle philosophical questions about what mathematics is, why mathematical structures apply so effectively in science or whether mathematical entities are discovered, invented or described in some other way.

What could change the conclusion

The scope of the analogy should be reconsidered if a proposed account explains formal notation while failing to explain proof practice, interpretation, abstraction or applied modelling, or if a competing framework better captures those functions without treating all mathematics as linguistic.

One thing to remember

Mathematics contains languages and uses them powerfully, but mathematical knowledge is broader than the formal symbol systems in which it is written.

KNO-018 · Images, Diagrams and Visualisation

In brief

A visual display can reduce cognitive work because spatial relations can make patterns explicit that prose leaves implicit. But visual channels differ in precision: position on a common scale is generally decoded more accurately than area or angle, and colour scales can create false boundaries or unequal emphasis. Similar summary statistics can conceal very different data structures, which is why plots and numerical summaries complement one another. A graph is still a transformation, not raw reality. Variables, scales, binning, aggregation, projection and framing determine what becomes visible. Scientific images add acquisition, reconstruction, filtering, segmentation and display conversion between the physical system and the published pixels. Processing can reveal useful structure, but it can also erase information or create misleading appearances if choices are hidden. An auditable figure therefore makes source data, transformations, encodings and important omissions inspectable, and uses alternative views when conclusions depend strongly on presentation choices and audiences.

Why it matters

Modern science relies on visual intermediates: microscope images, maps, spectra, simulations, network diagrams, heat maps and statistical plots. A figure may be the first place a researcher sees an outlier, batch effect or nonlinear relation. That makes visualisation part of inquiry, not merely presentation. The same power creates risk: a display can exaggerate weak differences, hide extremes or turn processing decisions into apparently natural features. Treating visualisation as an epistemic instrument makes its transformations subject to the same scrutiny as measurements and models.

What we know

Visual representations transform data into spatial and perceptual encodings. Those choices affect what patterns are easy to detect, how accurately quantities are decoded and which aspects of acquisition or processing remain visible to the reader.

What we don't know

No single visual encoding is optimal for every task, audience or dataset, and an attractive or reproducible rendering does not establish the correctness of the scientific interpretation. Important structure can remain hidden outside the chosen view.

What could change the conclusion

Interpretation should change if alternative scales, encodings, processing steps or views remove the apparent pattern, if provenance reveals unreported transformations, or if a visual feature cannot be reproduced from the stated source data and pipeline.

One thing to remember

A scientific figure is evidence only when its transformation from source data to display is appropriate, inspectable and open to reconstruction.

KNO-019 · Data, Tables and Records

In brief

Data become durable evidence only when their structure and history remain interpretable. Tables make some relationships easy to inspect, but spreadsheet convenience can also hide units, types, formulas or missing-value conventions. Metadata preserve the context needed to understand what values mean. Provenance records how data were produced or transformed. Versioning matters because a citation to “the dataset” is ambiguous if the underlying object changes, while persistent identifiers improve reference stability without replacing documentation. Missing values must be represented explicitly because absence, non-detection and not-applicable are not the same state. “Raw data” are also not automatically untouched: instruments and acquisition systems may already have applied calibration or processing. Preservation therefore succeeds only when future users can reconstruct enough of the chain from observation through processing to analysis. A good record is not merely a file that survives; it is a file whose meaning, lineage and version remain auditable by later investigators.

Why it matters

Modern science depends on records that outlive the moment of observation. Measurements are copied into tables, transformed by software, merged with metadata, deposited in repositories and cited by later work. Each step can preserve information or quietly destroy context. A dataset may be numerically intact yet unusable because variable definitions, units, exclusions or processing history are missing. The problem matters for reproducibility, reanalysis, error correction and long-term preservation. Understanding records as structured, versioned and provenance-bearing objects helps distinguish “data still exist” from the stronger claim that another person can still understand what those data mean and how they became evidence.

What we know

We know that explicit tabular organization can make variables and observations easier to analyse, that spreadsheet conventions can reduce avoidable ambiguity and mixing of raw and derived content, that provenance can represent entities, activities and agents involved in producing data, and that data citation requires specificity sufficient to identify the object or version being cited. NIST RDaF Version 2.0 explicitly treats metadata, provenance, versioning, instruments, processing operations and reproducibility as parts of research-data management. The National Academies likewise connect reproducibility with access to the data and computational information needed to recompute results.

What could change the conclusion

The methodological conclusions here would need revision if evidence showed that the proposed distinctions systematically fail in important domains, if more effective provenance or versioning models replace current practice, or if future infrastructures make some currently necessary contextual records reliably recoverable by other means. Changes in tools would not remove the underlying requirement that a reader be able to identify and interpret the evidential object actually used.

One thing to remember

Data preserve knowledge only when the record preserves enough meaning, provenance and version context to make the values interpretable.

KNO-020 · Review, Consensus and Disagreement

In brief

Peer review, publication, consensus and disagreement are stages or signals within scientific evaluation, not substitutes for the underlying evidence. Reviewers can expose weaknesses before publication, but different reviewers may reasonably emphasize different risks. Publication widens scrutiny rather than ending it. Consensus becomes more informative when it records convergence across evidence, methods and critical examination, while remaining open to revision. A minority view is not disqualified by being a minority, but neither does minority status grant equal evidential support. Productive disagreement identifies where interpretations diverge and what observation, analysis or replication could discriminate between them. Corrections and retractions can reveal weaknesses, but they can also show that the scholarly record remains open to repair. The practical task for a reader is therefore to ask what was reviewed, what evidence supports the claim, where disagreement lies, how independent the supporting routes are, and what could change the evidential balance in scientific practice.

Why it matters

Public discussion often compresses scientific evaluation into slogans: “peer reviewed” is treated as a seal of truth, “scientific consensus” as a vote, or “scientists disagree” as proof that knowledge has collapsed. Those shortcuts obscure the distinction between evidence and the social procedures used to scrutinize it. The distinction matters when papers conflict, expert committees issue assessments, minority interpretations challenge prevailing views or a correction changes a published record. Knowing what review and consensus can and cannot establish helps readers judge controversy without either worshipping authority or mistaking disagreement for epistemic paralysis.

What we know

- Structured peer review can expose problems before publication, but reviewer judgments vary and are not certificates of truth. - Editorial decisions can incorporate conflicting reviewer advice rather than simply count votes. - Reproducibility and replicability extend scrutiny beyond publication and can strengthen, qualify or challenge claims. - Formal scientific consensus processes can combine evidence review, expert deliberation and independent review. - Corrections can preserve an auditable record of errors found after publication.

What could change the conclusion

Evidence that a review model systematically fails to detect important errors, that a consensus process suppresses relevant counterevidence, or that a minority model repeatedly makes better predictions should reduce confidence in the prevailing assessment. Conversely, repeated independent tests that eliminate plausible alternatives should increase it.

One thing to remember

Review, consensus and disagreement organize criticism; their epistemic value depends on the evidence, procedures and discriminating tests they make visible.

KNO-021 · Open Science and Reproducibility

In brief

Open science is not one switch labelled “share everything.” It is a set of controls that make scientific work more inspectable. Reproducibility usually asks whether the same data and computational procedures can yield the same result; replicability asks whether new data collected for the same question produce results that are consistent within the relevant uncertainty. Preregistration can separate predictions from analyses chosen after seeing the data, but it cannot make a weak design strong. Registered reports move part of evaluation before results are known, while open data and code let others inspect transformations, exclusions and implementation. None of these practices guarantees truth. Their value is procedural: they expose more of the path from observation to claim, preserve alternatives that would otherwise disappear, and make correction cheaper. Openness also has limits: privacy, security, consent, intellectual property and the burden of curation can justify controlled rather than universal access. The goal is auditable science, not frictionless disclosure.

Why it matters

Modern science often depends on datasets, scripts, instruments and decisions that are invisible in the final paper. When only the narrative survives, later researchers may be unable to tell whether a disagreement comes from data, preprocessing, statistical choices, software, interpretation or a genuine difference in nature. Open-science practices try to preserve those decision points. They matter because criticism is one of science’s error-correcting mechanisms, but criticism works only on what can be inspected. The same practices also expose a tension: openness can increase auditability while creating new obligations around privacy, governance and long-term stewardship.

What we know

- Reproducibility and replicability test different kinds of consistency and should not be treated as synonyms. - Preregistration can make the timing of hypotheses and analysis choices visible. - Open data, code and provenance can make more of an analysis inspectable, but access without context is insufficient. - Versioning and visible corrections preserve which scientific object a claim refers to.

What we don't know

- How much each open-science intervention improves reliability across different disciplines and study designs. - Which combinations of openness and controlled access best preserve both auditability and legitimate privacy. - How long contemporary research objects remain executable as software and hardware ecosystems change. - How incentive systems alter the real use, rather than nominal adoption, of open practices.

What could change the conclusion

- Large cross-disciplinary evaluations showing that a widely promoted practice adds no auditability or reliability under realistic use. - Evidence that a control intended to reduce

selective reporting systematically creates larger distortions than it prevents. - New privacy-preserving methods that make currently restricted forms of scrutiny safely routine. - Long-term preservation studies showing that current repository and versioning practices fail to keep research objects interpretable.

One thing to remember

Open science does not certify truth; it preserves more of the evidential path so that other people can inspect, reproduce, challenge and correct it.

KNO-022 · Computation as External Memory

In brief

Humans routinely move cognitive work into the environment: we write lists, draw diagrams, set reminders and arrange objects so that the world carries part of the task. Cognitive-offloading research treats this as a general strategy rather than an anomaly. Computers extend the same logic. They can retain exact symbolic states, rerun procedures and transform representations at scales that unaided memory cannot manage. External representations also change the structure of thinking, not merely its storage burden. Search engines and networked information can shift memory toward knowing where information can be found. In science, executable code can preserve a procedure more precisely than prose, especially when documentation and computation are kept together. But computation is not neutral memory. File formats, software versions, interfaces and automation all shape what survives and what becomes easy to do. A computer can preserve an error perfectly. Its epistemic value comes from coupling external memory to provenance, interpretation and checks.

Why it matters

Many scientific results now depend on chains of computation too long to reproduce mentally and too detailed to describe completely in prose. The practical question is no longer whether humans use external memory, but how reliable that external memory is. Scripts, notebooks, databases and version histories can preserve reasoning steps across time and across people. They can also detach users from the assumptions embedded in those steps. Understanding computation as external memory helps separate three things that are often conflated: storing information, executing a procedure and understanding what the procedure means.

What we know

- People use external representations and actions to reduce internal cognitive demand. - External representations can change the structure and cost of reasoning, not only storage requirements. - Expectations of external access can affect what people remember about information and its location. - Executable code can preserve procedures with greater operational specificity than prose alone.

What we don't know

- How different forms of cognitive offloading reshape long-term learning rather than immediate task performance. - Which computational dependencies are necessary to preserve for century-scale executability. - How much automation changes users' ability to detect conceptual rather than implementation errors. - Where responsibility should sit in highly distributed human-computer workflows.

What could change the conclusion

- Evidence that external representations systematically impair reasoning even when task demands and retrieval costs are controlled. - Long-term studies showing that current

reproducibility tools fail to preserve executable scientific workflows despite complete metadata. - Demonstrations that automation errors are reliably caught without independent validation, reducing the need for redundant checks. - New cognitive models that explain offloading effects without any functional role for external memory or representation.

One thing to remember

Computation becomes epistemically powerful when it preserves not only results, but the states, procedures and provenance needed to understand how those results were produced.

KNO-023 · How Groups Know More Than Individuals

In brief

A group can gain epistemic power in several ways. Different members can know different facts; independent estimates can cancel some individual errors; discussion can expose contradictions; specialized roles can preserve knowledge no one person holds completely. Experiments show that groups can outperform their best members on some structured problem-solving tasks. Models by Hong and Page show how diversity of problem-solving approaches can matter under specified assumptions. Yet aggregation can fail. Lorenz and colleagues found that social influence can reduce diversity of estimates without necessarily improving collective accuracy. A group can also amplify shared bias, suppress minority information, follow prestige instead of evidence or converge before alternatives are explored. Collective intelligence therefore requires architecture: mechanisms for eliciting independent information, assigning expertise, recording dissent, integrating contributions and revising decisions. The key question is not whether groups are wiser than individuals in general. It is which communication and aggregation structures preserve the information that individuals collectively possess.

Why it matters

Science, engineering, medicine and public institutions all depend on groups that must combine knowledge distributed across people. No individual reads every paper, performs every measurement or understands every subsystem. The reliability of collective knowledge therefore depends on organizational design as much as on individual expertise. A committee can outperform any member by pooling complementary information, or perform worse by forcing premature agreement. Understanding that difference matters because consensus is often treated as if it were a count of agreeing heads. KNO-023 instead asks what information-processing structure produced the agreement.

What we know

- Groups can outperform their best members on some structured intellectual tasks. - Diverse problem-solving heuristics can improve group search under specified model assumptions. - Social influence can reduce diversity of estimates and undermine some wisdom-of-crowds benefits. - Interaction patterns such as conversational turn-taking are associated with collective performance in laboratory groups.

What we don't know

- How well laboratory measures of collective intelligence generalize to long-lived institutions and high-stakes expert teams. - Which diversity dimensions matter for which kinds of scientific and engineering problems. - How to weight expertise without suppressing informative minority evidence. - How digital communication and AI mediation change independence, conformity and information pooling in human groups.

What could change the conclusion

- Robust evidence that interaction structure adds no predictive value once individual ability is controlled across realistic tasks. - Experiments showing that social influence reliably improves accuracy despite reducing independence across broad domains. - Field studies demonstrating that diversity of relevant problem-solving perspectives systematically harms performance under conditions where integration is effective. - Institutional evidence that explicit dissent-preservation mechanisms reduce rather than improve error detection in comparable decisions.

One thing to remember

Groups know more than individuals only when their structure preserves and combines distributed information better than conformity, shared bias or premature consensus destroy it.

KNO-024 · Lost Knowledge and Preservation Bias

In brief

Records survive through filters. In archaeology, objects move from active use into archaeological context through cultural and natural formation processes, so what is excavated is already transformed. In paleontology, weathering, transport, burial and mineralization make some organisms and environments much more likely to enter the fossil record than others. Archives introduce another layer: people and institutions decide what to document, retain, classify and permit others to access. Digital media do not abolish loss. Vines and colleagues found that the availability of research datasets declined strongly with article age in the sample they studied, illustrating how fragile data become when preservation depends on individual custodians. Digital preservation therefore requires migration, integrity checking, metadata and institutional continuity rather than mere copying. The methodological lesson is simple but demanding: absence in a surviving record is evidence only after the processes that could have removed, hidden or never recorded the item have been considered.

Why it matters

We reason from survivors constantly: fossils, laboratory notebooks, administrative files, archaeological sites, photographs, datasets and source code. Because these survivors are visible, they can feel like the past itself. They are not. Every record has a survival history, and that history can be correlated with the property we want to infer. Large bones preserve differently from fragile tissue; official archives preserve differently from informal speech; maintained repositories preserve differently from forgotten drives. Preservation bias matters whenever the mechanism of survival can distort the apparent frequency, diversity or chronology of what we observe.

What we know

- Archaeological records are shaped by formation processes between use and recovery. - Bone weathering and other taphonomic processes affect which biological traces survive and how they are interpreted. - Archives embody decisions about creation, appraisal, description and access rather than functioning as neutral containers. - Research datasets can become unavailable over time when preservation depends on individual custodians.

What we don't know

- How much undocumented material has already been lost from contemporary digital research ecosystems. - How preservation probabilities vary across underrepresented communities and informal record-keeping systems. - Which digital preservation strategies will remain interpretable over century-scale technological change. - How best to quantify preservation bias when creation, loss and discovery probabilities are all uncertain.

What could change the conclusion

- Empirical records showing that major classes of supposedly fragile evidence preserve at the same rate as durable evidence across relevant contexts. - Archival studies demonstrating

that appraisal and access decisions do not measurably affect later historical inference in comparable collections. - Longitudinal digital studies showing negligible loss of research data without institutional stewardship. - New preservation methods that make format, software and metadata obsolescence effectively irrelevant over long intervals.

One thing to remember

A surviving record is evidence about the past and about the processes that selected what could survive; absence is interpretable only after both are considered.