

WAYS OF KNOWING

Scientific Edition

Twenty-four studies on evidence, representation, measurement and knowledge

Juan Segurola

Version v1.0 · 22 September 2026

© 2026 Juan Segurola. All Rights Reserved.

Contents

KNO	Chapter	Page
KNO-001	A Mark on a Stone	5
KNO-002	Counting Before Numbers	13
KNO-003	From Marks to Symbols	20
KNO-004	From Symbols to Writing	28
KNO-005	Oral Language and Transmission	38
KNO-006	Maps and Spatial Representation	47
KNO-007	Calendars and Time Measurement	56
KNO-008	Instruments That Extend the Senses	64
KNO-009	Archives, Libraries and Collective Memory	73
KNO-010	Classification as a Knowledge Tool	81
KNO-011	Measuring Before Theorising	90
KNO-012	Experiments and Variable Control	98
KNO-013	Models, Maps and Reality	106
KNO-014	Causality and Correlation	114
KNO-015	Uncertainty and Probability	125
KNO-016	How Evidence Is Built: Putting the Tools Together	134
KNO-017	Is Mathematics a Language?	143
KNO-018	Images, Diagrams and Visualisation	151
KNO-019	Data, Tables and Records	158
KNO-020	Review, Consensus and Disagreement	167
KNO-021	Open Science and Reproducibility	176
KNO-022	Computation as External Memory	183
KNO-023	How Groups Know More Than Individuals	189

KNO	Chapter	Page
KNO-024	Lost Knowledge and Preservation Bias	196

Preface

How knowledge leaves traces

Knowledge rarely arrives as direct access to the thing we want to know. We work from traces: marks on stone, spoken traditions, maps, calendars, instrument readings, archives, tables, experiments, models and preserved datasets. Each trace carries information, but each also has a history of production, selection and loss. The central problem of this volume is therefore not only what we know, but how a claim becomes warranted.

The twenty-four KNO studies follow that problem from early external marks and counting practices through writing, spatial representation and measurement, then into causality, uncertainty, modelling, visualisation, records, review, reproducibility, computation and collective intelligence. The final chapters turn the method back onto the archive itself: what disappears, what survives, and how preservation can bias the picture reconstructed from what remains.

The chapters are arranged as a coherent route, but they are not a ladder of inevitable progress. Different communities solved different informational problems with different tools. A later technique is not automatically a better answer to an earlier question, and a surviving object is not automatically the first instance of a practice. Throughout the series, observation is kept separate from interpretation, bibliographic identity from claim support, and a useful model from the reality it represents.

This scientific edition preserves the technical arguments, evidential limits and references of the KNO corpus while presenting them as a single volume. The recurring discipline is simple to state and difficult to maintain: ask what was observed, what had to be inferred, which assumptions connect the two, what independent checks exist, and what evidence would force the conclusion to change.

Read in sequence, KNO is a study of the machinery of knowing. Read selectively, each chapter remains an independent examination of one part of that machinery. In both cases, uncertainty is treated as information about the reach of the evidence, not as permission to treat every explanation as equally plausible.

KNO-001 · A Mark on a Stone

KNO · Ways of Knowing

Juan Segurola

v0.2 · EN

All Rights Reserved

1. A Scratch Is Not Yet a Symbol

The Question

A groove can be measured, photographed and dated. Its meaning cannot be lifted from the surface in the same way. What chain of evidence allows archaeology to move from a physical trace to an intentional mark, and then—more cautiously—to a possible sign? A line on a durable object begins as an observable modification, not as a decoded message. Archaeologists can establish material, geometry, superposition and manufacture before they address purpose. That order matters because an intentional cut may have been practical, decorative, imitative, playful or communicative; survival alone does not choose among those possibilities.[1,3] At Trinil in Java, a freshwater shell from the Homo erectus assemblage bears a geometric engraving made before fossilisation. The associated shell context is hundreds of thousands of years old. The study establishes deliberate manufacture more strongly than it establishes function: its authors explicitly state that the engraving's purpose or meaning cannot be assessed.[1] Middle Stone Age ochres from Blombos Cave and Klein Kliphuis carry deliberately incised patterns. Detailed studies distinguish patterned engraving from grinding or scraping used to process pigment. Yet even a repeated cross-hatched design does not contain its own dictionary. Calling it symbolic is an interpretation supported by pattern and context, not a translation of recovered semantic content.[2,3,5] A silcrete flake from the approximately 73,000-year-old levels at Blombos preserves intersecting red lines. Microscopy and chemical analysis support deliberate application with an ochre crayon, and the lines continued beyond the surviving edge. This identifies a drawing action and part of a larger design; it still does not recover what the design meant to its maker. [4] The inference therefore has layers. Surface analysis can show that a mark is ancient; fracture and sequence can show when it was made; experiments can identify a tool and gesture; recurrence can support deliberate patterning. Intention, notation and shared meaning sit progressively farther from direct observation and require progressively more contextual evidence.[1,3,4]

Where That Leaves Us

The oldest durable marks enlarge the history of deliberate pattern-making, but they do not come with captions. Archaeology is strongest when it reports the sequence of inferences rather than collapsing them: object, modification, deliberate gesture, recurrent form, possible convention. The mark is evidence; the recovered meaning is usually the open question.

2. Marks, Tallies and Numbers Are Different Claims

The Question

Several parallel notches look countable to a modern eye. But countable is not the same as counted, and counted is not the same as numerical notation. Which observable features can distinguish decoration, grip, ownership, sequential action and a record of discrete events? Incisions can result from

cutting, scraping, sharpening, gripping, butchery or deliberate marking. Microscopic profiles, stroke direction, superposition and tool changes can exclude some alternatives. They identify how a sequence was produced more securely than why. A regular row is therefore evidence of organised action before it is evidence of counting.[3,9] Experiments on the Zaskalnaya raven bone indicate that two notches may have been added to make an existing sequence look more regular. That result supports attention to visual spacing. It does not demonstrate that each notch stood for one animal, day or quantity, and it offers no recovered rule for reading the sequence.[8,9] A notched bone from Border Cave, about 44,000 years old, contains groups made at different times and has been interpreted as notational. Sequential production makes an external record plausible. The recorded category is unknown, however, and the object does not by itself reveal whether the units were numerical, calendrical, social or practical.[7,9] Archaeological proposals for the emergence of number symbols use testable stages: perceptual grouping, one-to-one marks, accumulation across sessions, conventional forms and eventually operations on signs. This is a model for organising evidence, not an excavated universal sequence. Different communities could externalise quantity through different materials and routes.[9] The Ishango bone concentrates the temptation to overread. It is a small object with a quartz point and grouped notches, found near Lake Edward and usually dated to roughly twenty thousand years ago. Many arithmetic and calendrical interpretations have been proposed. The notches are real; a uniquely demonstrated decoding is not.[10]

Where That Leaves Us

A row of notches can be intentional without being a tally, and a tally can track items without constituting writing. The strongest analyses test manufacture, sequence and grouping before proposing content. Number enters the argument only when evidence supports a mapping between marks and units, not when modern observers can simply count the marks.

3. How Archaeology Tests Intention

Microscopy turns a mark into a sequence of physical events. Investigators examine groove cross-sections, shoulders, internal striations, pigment particles, entry and exit points, and crossings between lines. Those features can distinguish a dragged point from deposited pigment and can reconstruct which stroke preceded another without assigning a message to either.[3,4] Experimental replication supplies comparison rather than reenactment. Researchers use candidate tools and gestures on similar materials, then compare resulting traces with archaeological ones. A close match can support a manufacturing hypothesis and estimate skill or effort. It cannot establish that the experimenter's motive was the ancient maker's motive.[3,8] Context asks whether the marked object belongs where investigators think it does. Stratigraphy, refitting, dating, surface alteration and association with other artefacts constrain age and behaviour. A striking object without secure provenance carries less historical information than a modest one excavated within a documented sequence.[1,2,5] Confidence grows when independent observations converge: the mark predates burial, its strokes were deliberately sequenced, experiments reproduce its morphology, and comparable patterns recur in related contexts. These signals are not fully independent if they derive from the same misidentified object, so identity and provenance checks must precede interpretation.[3,4,9] Good analysis also preserves uncertain cases. The Blombos ochre study separated definite engravings, ambiguous incisions and traces better explained by processing. More generally, Pleistocene behavioural evidence accumulates unevenly rather than arriving as one sudden package. Classification with explicit uncertainty is more informative than forcing every mark into a story of cognitive revolution.[3,6]

Where That Leaves Us

Archaeology cannot interview the maker, but it can discipline the questions asked of the object. Microscopy reconstructs gestures, experiments compare causes, and context secures time and association. Meaning remains the most distant inference. Uncertainty is not failure here; it is the recorded distance between a surviving trace and a vanished act.

4. External Memory Changes the Task

The Question

Remembering internally and consulting a stable mark can produce the same answer, but they are not the same cognitive operation. What changes when part of a task is held outside the nervous system, where it can be inspected, rearranged and shared? External representations change a task by making some states perceptually available. A written intermediate result, arranged token or marked position need not be maintained entirely in working memory. The environment does not solve the problem alone; it redistributes storage, search and comparison between a person and a structured surface. [15,18] In Hutchins's cockpit study, remembered speeds were distributed across people, spoken procedures, instruments and movable markers. No single component contained the whole memory in isolation. The case demonstrates how a coordinated system can preserve task-relevant state, while providing an analogy—not direct evidence—for much older practices of marking and checking.[16] Some actions are epistemic: they are performed to reveal, simplify or stabilise information rather than to advance directly toward a physical goal. Rotating a piece to recognise it or arranging counters to compare groups can reduce mental search. A mark can serve this role even if it is temporary and understood by only one person.[17,18] The extended-mind thesis goes further, arguing that a reliably integrated external resource may count as part of a cognitive process. That is a philosophical position about where cognition ends, not an archaeological measurement. KNO-001 needs only the narrower empirical point: external records can alter how remembering and reasoning are performed.[19] For prehistory, the safe inference is conditional. If marks were revisited, matched one-to-one with events or accumulated across sessions, they could have reduced demands on unaided recall. Microscopic sequence may support repeated production, but the cognitive role remains a model until contextual evidence links the object to a recurrent task.[9,15,18]

Where That Leaves Us

An external mark does not think, but it can transform what a thinker must keep active. Stable states can be revisited, compared and corrected; arrangements can make relations visible. Modern cognitive studies establish those mechanisms. Applying them to prehistory requires a separate archaeological argument that the marks were actually used in a recurrent task.

5. From Private Trace to Shared Convention

A person can use a mark without anyone else knowing its rule. Such a private trace already extends persistence across a delay, but it dies as a usable record when the rule is forgotten. Public notation requires more: a community must stabilise how forms are produced, distinguished and connected to things or actions.[15,18] Recurrence can support convention, but resemblance alone is not enough. Simple grids, parallel lines and crosses can arise independently because human hands and available surfaces constrain design. Stronger evidence would combine repeated form with shared manufacture, placement, chronology and context, while still acknowledging that similar signs can carry different

meanings.[3,4,5] Cumulative cultural evolution involves modifications that are retained and build on earlier ones, although scholars distinguish several forms and mechanisms. Durable records can assist accumulation by preserving outcomes for later learners. They are neither necessary in every case nor sufficient: copying a persistent error can accumulate just as effectively as copying an improvement.[25] External symbolic storage has been proposed as a major transition in human cognitive-cultural history because records can outlast individual memory and support instruction, retrieval and revision. The broad mechanism is plausible. Any fixed stage sequence, however, compresses diverse histories and should be tested against archaeological chronologies rather than treated as a universal ladder.[20,25] There was probably no instant at which one scratch converted an isolated mind into a cumulative culture. Patterning, teaching, material aids, convention and institutions leave different traces and need not emerge together. A better history tracks capacities and practices separately, then asks when particular combinations became stable enough to change transmission.[6,20,25]

Where That Leaves Us

Persistence becomes socially powerful when a rule survives along with the trace. Then another person can retrieve, compare or correct what the maker recorded. Convention is therefore not decoration added to memory; it is the coordination that makes an external state portable across minds. Archaeology must demonstrate that coordination rather than assume it from geometry.

6. Counting Objects, Accounts and Writing

The Question

A token can stand for a sheep, a notch for a delivery and a written sign for a spoken word. These systems can interact historically, but they solve different representational problems. What evidence justifies moving from quantity to account and from account to language? Humans and other animals can discriminate approximate numerosities, while small exact sets and learned symbolic numbers involve partly different systems. Neural studies identify populations tuned to quantity, but a sense of more or less is not an inscription. Exact, indefinitely extensible number systems require learned practices that connect forms, words or gestures to ordered values.[11,14] Experiments with Pirahã participants found accurate performance for some exact matching tasks but increasing error when quantities had to be retained or transformed without exact number words. The result supports number language as a cognitive technology for stable exact representation. It does not show that speakers lack numerical thought or general intelligence.[12] Indigenous Australian children speaking languages with restricted number vocabularies nevertheless performed well on nonverbal numerical tasks. Read with the Pirahã results, this blocks a simple claim that number words create all numerical capacity. Language and material signs can stabilise exact operations without being the source of every quantitative discrimination.[13] In the ancient Near East, clay tokens, sealings and numerical tablets form an influential reconstruction of administrative recording before and alongside early writing. Proto-cuneiform documents from the late fourth millennium BCE are heavily numerical and administrative. The proposed token-to-sign continuity is powerful for this regional case, but details and wider universality remain debated.[21,22] Writing is not identical to every durable sign system. It conventionally represents language across a sufficiently broad range, whereas tallies, ownership marks and numerical notations may encode restricted domains. Comparative histories identify several independent inventions of writing. There is no need to force all of them through one chain beginning with tally marks.[21,23]

Where That Leaves Us

Quantity can be perceived without symbols, exact values can be stabilised by words or marks, accounts can organise restricted domains, and writing can represent language. These capacities overlap but are not interchangeable. Their histories should be reconstructed locally, from documented transitions, rather than arranged in a universal staircase from notch to number to text.

7. What Survives Is Not What Existed

Bone and shell enter the archaeological record through selective pathways. Weathering cracks, abrasion, trampling, chemical alteration, transport and breakage remove objects and modify surfaces. Taphonomic analysis can identify some of these processes. It also shows why the surviving sample is not a neutral census of what people made or marked.[24] Organic records are especially vulnerable. Wood, bark, fibre, skin and soft plant materials can carry marks or arrangements but usually decay outside exceptional conditions. Border Cave's organic artefacts are informative partly because such preservation is unusual. Stone and bone may dominate our evidence without having dominated ancient representational practice.[7,24] A durable engraved object is more likely to survive, be recognised and be published than an unmarked or perishable aid. Frequency in museum collections therefore combines ancient behaviour, preservation, excavation, identification and research attention. Apparent bursts of marking may reflect some mixture of real change and improved visibility.[1,2,3,24] Absence of marks in a layer is strongest when excavation, preservation and sample size make absence detectable. Otherwise it may indicate no practice, a different material, a different context of use or simple loss. Archaeological silence is evidence only after the conditions that could produce silence have been modelled.[6,24] Archaeology, cognitive science and neuroscience answer different questions. Traces constrain what happened to objects; experiments reveal capacities and task mechanisms; neural data describe living systems. They can inform one another, but a modern capacity does not date its origin, and an ancient mark does not disclose the complete cognition that produced it.[9,11,15,24]

Where That Leaves Us

The archive is filtered before excavation begins. Durable materials, stable contexts and recognisable patterns are overrepresented; perishable aids and failed experiments mostly disappear. Any history of external memory must therefore explain both the marks that survived and the processes that selected them. A gap in the record is not automatically a gap in capacity.

8. A Mark Changes What Can Persist

Once a state is materially stabilised, it can persist while attention moves elsewhere. The maker can return to it; another person may inspect it; later action can begin from an earlier result. This is the minimal transformation external memory offers, independent of whether the mark is artistic, numerical or linguistic.[15,18] A persistent record also permits operations unavailable to a vanishing thought: side-by-side comparison, reordering, annotation and correction. Those affordances can make discrepancies visible and distribute complex work across time. They do not guarantee truth. A durable false entry can coordinate later action just as effectively as an accurate one.[15,17,18] When records acquire shared rules, copying practices and institutions, they can connect people who never meet. Accounts coordinate stores and obligations; texts preserve procedures and arguments; archives allow later comparison. These are historically documented expansions of persistence. They should not be projected backward onto every Palaeolithic line.[20,21,23,25] No known object deserves to be called the single first external memory. Earlier traces may be lost, and surviving candidates instantiate different behaviours. The

historical question is therefore plural: when did particular communities begin revisiting marks, mapping them to units, sharing their rules and building later records from them?[1,6,9,20] The deepest change is not that stone began to contain thought. It is that a human action could leave a structured state available after the action ended. From that modest capacity, under particular conventions and institutions, came many forms of tallying, diagramming, accounting and writing—related technologies with distinct histories rather than one inevitable ladder.[9,15,23,25]

Where That Leaves Us

A mark matters because it can remain available after the gesture and perhaps after the maker. That persistence creates opportunities for checking, sharing and accumulation, but each requires additional evidence and social practice. The scientifically honest story is not a leap from scratch to civilisation; it is a branching history of external supports for thought.

What We Know, and How

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 Still 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 Our Minds

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.

If You Remember Only One Thing

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

Related Articles

None assigned in Master Index v2.50.

1

[1] Josephine C. A. Joordens; Francesco d'Errico; Frank P. Wesselingh; Stephen Munro; John de Vos; Jakob Wallinga; Christina Ankjærgaard; Tony Reimann; Jan R. Wijbrans; Klaudia F. Kuiper; Herman J. Mûcher; Hélène Coqueugniot; Vincent Prié; Ineke Joosten; Bertil van Os; Anne S. Schulp; Michel Panuel; Victoria van der Haas; Wim Lustenhouwer; John J. G. Reijmer; Wil Roebroeks. (2015). Homo erectus at Trinil on Java used shells for tool production and engraving. *Nature* 518, 228–231. DOI: 10.1038/nature13962. <https://doi.org/10.1038/nature13962> [2] Christopher S. Henshilwood; Francesco d'Errico; Royden Yates; Zenobia Jacobs; Chantal Tribolo; Geoff A. T. Duller; Norbert Mercier; Judith C. Sealy; Hélène Valladas; Ian Watts; Ann G. Wintle. (2002). Emergence of modern human behavior: Middle Stone Age engravings from South Africa. *Science* 295, 1278–1280. DOI: 10.1126/science.1067575. <https://doi.org/10.1126/science.1067575> [3] Christopher S. Henshilwood; Francesco d'Errico; Ian Watts. (2009). Engraved ochres from the Middle Stone Age levels at Blombos Cave, South Africa. *Journal of Human Evolution* 57, 27–47. DOI: 10.1016/j.jhevol.2009.01.005. <https://doi.org/10.1016/j.jhevol.2009.01.005> [4] Christopher S. Henshilwood; Francesco d'Errico; Karen L. van Niekerk; Laure Dayet; Alain Queffelec; Luca Pollarolo. (2018). An abstract drawing from the 73,000-year-old levels at Blombos Cave, South Africa. *Nature* 562, 115–118. DOI: 10.1038/s41586-018-0514-3. <https://doi.org/10.1038/s41586-018-0514-3> [5] Alex Mackay; Aara Welz. (2008). Engraved ochre from a Middle Stone Age context at Klein Kliphuis in the Western Cape of South Africa. *Journal of Archaeological Science* 35, 1521–1532. DOI: 10.1016/j.jas.2007.10.015. <https://doi.org/10.1016/j.jas.2007.10.015> [6] Sally McBrearty; Alison S. Brooks. (2000). The revolution that wasn't: a new interpretation of the origin of modern human behavior. *Journal of Human Evolution* 39, 453–563. DOI: 10.1006/jhev.2000.0435. <https://doi.org/10.1006/jhev.2000.0435> [7] Francesco d'Errico; Lucinda Backwell; Paola Villa; Ilaria Degano; Jeannette J. Lucejko; Marion K. Bamford; Thomas F. G. Higham; Maria Perla Colombini; Peter B. Beaumont. (2012). Early evidence of San material culture represented by organic artifacts from Border Cave, South Africa. *Proceedings of the National Academy of Sciences* 109, 13214–13219. DOI: 10.1073/pnas.1204213109. <https://doi.org/10.1073/pnas.1204213109> [8] Ana Majkić; Sarah Evans; Vadim Stepanchuk; Alexander Tsvelikh; Francesco d'Errico. (2017). A decorated raven bone from the Zaskalnaya VI (Kolosovskaya) Neanderthal site, Crimea. *PLOS ONE* 12, e0173435. DOI: 10.1371/journal.pone.0173435. <https://doi.org/10.1371/journal.pone.0173435> [9] Francesco d'Errico; Luc Doyon; Ivan Colagè; Alain Queffelec; Emma Le Vraux; Giacomo Giacobini; Bernard Vandermeersch; Bruno Maureille. (2018). From number sense to number symbols: an archaeological perspective. *Philosophical Transactions of the Royal Society B* 373, 20160518. DOI: 10.1098/rstb.2016.0518. <https://doi.org/10.1098/rstb.2016.0518> [10] Royal Belgian Institute of Natural Sciences. (Updated record accessed 2026). The Ishango Bone. Official collection/exhibition record. <https://www.naturalsciences.be/en/museum/exhibitions-activities/exhibitions/250-years-of-natural-sciences/the-ishango-bone> [11] Lisa Feigenson; Stanislas Dehaene; Elizabeth Spelke. (2004). Core systems of number. *Trends in Cognitive Sciences* 8, 307–314. DOI: 10.1016/j.tics.2004.05.002. <https://doi.org/10.1016/j.tics.2004.05.002> [12] Michael C. Frank; Daniel L. Everett; Evelina Fedorenko; Edward Gibson. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition* 108, 819–824. DOI: 10.1016/j.cognition.2008.04.007. <https://doi.org/10.1016/j.cognition.2008.04.007> [13] Brian Butterworth; Robert Reeve; Fiona Reynolds; Delyth Lloyd. (2008). Numerical thought with and without words: Evidence from indigenous Australian children. *Proceedings of the National Academy of Sciences* 105, 13179–13184. DOI: 10.1073/pnas.0806045105. <https://doi.org/10.1073/pnas.0806045105> [14] Andreas Nieder. (2016). The neuronal code for number. *Nature Reviews Neuroscience* 17, 366–382. DOI: 10.1038/nrn.2016.40. <https://doi.org/10.1038/nrn.2016.40> [15] Jiajie Zhang; Donald A. Norman. (1994). Representations in distributed cognitive tasks. *Cognitive Science* 18, 87–122. DOI: 10.1207/s15516709cog1801_3. https://doi.org/10.1207/s15516709cog1801_3 [16] Edwin Hutchins. (1995). How a cockpit remembers its speeds. *Cognitive*

Science 19, 265–288. DOI: 10.1207/s15516709cog1903_1. https://doi.org/10.1207/s15516709cog1903_1 [17] David Kirsh; Paul Maglio. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science* 18, 513–549. DOI: 10.1207/s15516709cog1804_1. https://doi.org/10.1207/s15516709cog1804_1 [18] David Kirsh. (2010). Thinking with external representations. *AI & Society* 25, 441–454. DOI: 10.1007/s00146-010-0272-8. <https://doi.org/10.1007/s00146-010-0272-8> [19] Andy Clark; David Chalmers. (1998). The extended mind. *Analysis* 58, 7–19. DOI: 10.1093/analys/58.1.7. <https://doi.org/10.1093/analys/58.1.7> [20] Merlin Donald. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press. <https://www.hup.harvard.edu/books/9780674644847> [21] Hans J. Nissen; Peter Damerow; Robert K. Englund. (1993). *Archaic Bookkeeping: Early Writing and Techniques of Economic Administration in the Ancient Near East*. University of Chicago Press. <https://cdli.earth/publications/1657148> [22] Denise Schmandt-Besserat. (1992). *Before Writing, Volume I: From Counting to Cuneiform*. University of Texas Press. <https://utpress.utexas.edu/9781477325766/> [23] Christopher Woods; Emily Teeter; Geoff Emberling (editors). (2010). *Visible Language: Inventions of Writing in the Ancient Middle East and Beyond*. Oriental Institute Museum Publications 32. <https://isac.uchicago.edu/research/publications/oimp/oimp-32-visible-language-inventions-writing-ancient-middle-east-and> [24] Anna K. Behrensmeyer. (1978). Taphonomic and ecologic information from bone weathering. *Paleobiology* 4, 150–162. DOI: 10.1017/S0094837300005820. <https://doi.org/10.1017/S0094837300005820> [25] Alex Mesoudi; Alex Thornton. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B* 285, 20180712. DOI: 10.1098/rspb.2018.0712. <https://doi.org/10.1098/rspb.2018.0712>

KNO-002 · Counting Before Numbers

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

1. Quantity Before an Exact Number

The Question

If a person can choose the fuller basket without saying how many objects either basket contains, what numerical knowledge is present, and what is still absent? Humans, infants and many non-human animals can compare approximate numerosities. Performance depends on ratio: distinguishing eight from sixteen is easier than eight from ten. This approximate number system represents magnitude with limited precision; it does not by itself generate an exact symbol for every positive integer.[2,3] A second system can track a small number of individual objects with high precision, usually within a narrow capacity limit. Following one, two or three items is not the same operation as estimating a large crowd, and neither system alone provides the successor rule that makes the counting numbers indefinitely extensible.[2] Experiments with Mundurucu participants showed competent approximate addition and comparison despite a limited exact-number lexicon. Studies with Pirahã participants likewise found that performance remained strong when items could be matched directly, but accuracy fell when exact quantities had to be stored, transformed or reproduced across a delay. These results show that exact number words are tools for stabilising some tasks, not prerequisites for every perception of quantity.[4-6] The experiments do not rank languages or peoples by cognitive sophistication. They isolate task demands. When a collection remains visible, direct matching can carry the solution; when it disappears, a conventional word, gesture or material arrangement may preserve an exact state. The contrast is between representational resources, not between minds with and without quantity.[5-7]

Where That Leaves Us

Approximate magnitude, small-set tracking and exact symbolic number are related but distinct, and neither core system supplies the successor rule that makes counting indefinitely extensible. Before asking when people had numbers, a stronger question is which quantitative task they could solve, at what precision, and with which internal or external resources. The contrast these experiments isolate is between representational resources, not between minds.

2. One-to-One Correspondence Is Not Yet a Numeral

The Question

Can two sets be shown equal without identifying their cardinal number? One-to-one correspondence pairs every member of one collection with one member of another. A herder can give one pebble to each animal leaving an enclosure and remove one when each returns. If no pebble remains and no animal is unmatched, the sets correspond even if no word equivalent to “thirty-seven” was used.[1,13] This procedure preserves equality and detects surplus or absence. It does not necessarily represent the total independently of the objects or explain that all arrangements with the same cardinality share one abstract number. The relation may remain practical and local: one marker per animal, vessel or

obligation.[13,14] A tally introduces persistence. If one mark is deliberately mapped to one event, the record can be revisited after the event has passed. Yet repeated marks are not automatically tallies. Archaeology must establish intentional production and a plausible unit mapping before it can treat regularity as quantitative notation.[1] Even a genuine tally need not be a numeral system. Its marks may be interchangeable accumulations rather than named signs with conventional values. The tally answers “how many units have occurred here?” by allowing recounting; a numeral such as 37 encodes a value within a structured notation that can travel between contexts.[18]

Where That Leaves Us

Matching, tallying and naming an exact cardinal solve different problems. One-to-one correspondence may secure equivalence without a number word; a tally may preserve units without a portable numeral; a numeral system makes values conventional and reusable. Repeated marks are not automatically tallies, and deliberate production and a plausible unit mapping must be established before regularity can be read as quantitative notation.

3. Fingers, Bodies and Grouping

Body-based counting systems turn a stable anatomy into an ordered route. Fingers are especially convenient because they are visible, movable and naturally grouped, but documented systems also traverse wrists, elbows, shoulders, the head and the opposite side of the body. The sequence can serve as an externalised order even when its “signs” are enacted rather than detached objects.[11,12] Finger counting is not uniform. Cultures differ in starting hand, direction, finger order and whether a gesture denotes a cardinal value, an ordinal position or an action in a calculation. Those differences can influence numerical processing, which makes the hand a learned representational tool rather than a universal code waiting to be read.[11] Grouping reduces the burden of long one-by-one sequences. Collections can be organised in twos, fives, tens, twenties or other bundles; a higher-order unit then stands for several lower-order units. Bases often reflect body practices or material devices, but no single base follows inevitably from human anatomy.[18,20] An ordered count list becomes an exact-number technology only when learners coordinate several principles. Each item must receive one term, the stable order must be preserved, irrelevant changes of arrangement must not alter the total, and the final word must be understood as the cardinal value of the whole set. Children can recite a count list before they understand this cardinal principle.[9,10] Once the successor relation is understood, adding one item connects a set to the next number in the sequence. That insight supports generalisation beyond memorised small values. The history of a society's number practices cannot be inferred directly from child development, but developmental experiments clarify which conceptual relations a mature counting system must coordinate.[9,10]

Where That Leaves Us

The body can supply order, memory and grouping without dictating one universal notation, and no single base follows inevitably from human anatomy. Exact counting emerges from learned coordination among a stable sequence, individual items and cardinal totals, not from the mere availability of ten fingers. Developmental evidence clarifies which relations a mature system must coordinate; it does not reconstruct any society's history.

4. Tokens, Accounts and Institutions

Material counters can preserve quantity while freeing the represented objects to move elsewhere. A token may be handled, grouped, transferred or sealed; its usefulness depends on a rule connecting the object to a category, obligation or commodity. Material stability makes checking possible, but the rule remains social.[13,16] In southwest Asia, clay tokens occur long before the earliest written tablets. Denise Schmandt-Besserat proposed a long development from tokens to sealed envelopes, impressed signs and writing. The reconstruction is influential because it connects material forms with changing administrative needs, but the degree of direct continuity and its explanatory reach remain debated.[16] Late fourth-millennium BCE proto-cuneiform tablets provide firmer evidence for institutional numerical recording. Many record quantities and commodities using several metrological and numerical sign systems. A sign's value could depend on what was being counted or measured, so the same graphic form did not necessarily function as one context-free abstract number.[15,17] These tablets were not simply arithmetic notebooks. They belonged to administrative practices involving trained personnel, conventions, archives and recurrent transactions. Institution and notation co-developed: a durable sign system made accounts comparable, while repeated accounting supplied the settings in which signs could become standardised.[15] The regional sequence should not be inflated into a universal law. Other societies used knots, sticks, body sequences, oral formulae, written numerals and mixed devices. Similar tasks can recruit different media, and similar media can serve different tasks.[14,18]

Where That Leaves Us

The Near Eastern evidence documents one especially rich transition from material accounting to written numerical administration. It demonstrates what tokens and tablets could do in that history; it does not prove that every number system or every writing tradition began the same way. Institution and notation developed together there, and the degree of direct continuity from tokens to script remains debated rather than settled.

5. From Quantities to Numerals and Place Value

A numeral is a conventional sign or sign sequence for a number. Its efficiency depends on the rules of the notation: which forms are basic, how repetition works, whether position changes value, whether zero is available, and how units of different order are combined.[18] Additive systems repeat or combine value signs; multiplicative-additive systems express products of units; positional systems assign value partly through place. These are analytical categories, not evolutionary ranks. A notation can be highly effective for its uses without converging on modern decimal place value.[18] Place value compresses representation because a small set of digits can express arbitrarily large numbers when positions stand for powers of a base. It also creates a problem: an empty position must be distinguished from no notation at all. Placeholder marks and numerical zero have related but not identical histories.[18,20] Written numerals make values portable across objects and occasions. “37” can describe animals, days or lengths because the sign sequence is abstracted from the counted category. But practical systems often retain units, measures and commodity signs, reminding us that abstraction is achieved by convention rather than by erasing context automatically.[15,18] No single notation is the inevitable endpoint of counting. Roman numerals, Mesopotamian sexagesimal systems, Chinese counting rods, Mesoamerican vigesimal notation and Indo-Arabic decimal place value organise values differently. Comparison reveals multiple solutions to representation, calculation and transmission.[18]

Where That Leaves Us

Numerals are engineered conventions. Their forms encode choices about grouping, order, position and absence, and the categories that describe them are analytical rather than evolutionary ranks. Place value is powerful, but it is one historical solution within a diverse field, not the definition of numerical thought itself, and its abstraction is achieved by convention rather than by erasing context.

6. Measurement Is Not the Same as Counting

Counting assigns a number to discrete units under a rule for what counts as one. Measurement compares a continuous magnitude with a chosen unit. Five vessels can be counted; five litres result from repeatedly relating volume to a standard.[15,18] The distinction is not absolute at the procedural level because measurement often uses counting: one counts unit iterations, subdivisions or instrument marks. Conceptually, however, the operations differ. A counted total depends on individuating objects; a measured value depends on a unit, calibration and permissible precision.[15] Units are social and technical achievements. A cubit tied to a body reference becomes more reproducible when material standards and institutional procedures stabilise it. Standardisation allows results to travel, while also embedding authority in who maintains the reference and resolves discrepancies.[15] Measurement values can be fractional or approximate even when the counted instrument divisions are exact. Reporting a length therefore requires more than a numeral: it requires a quantity kind, a unit, a method and an uncertainty appropriate to the comparison.[18] Confusing counting with measurement can distort archaeology. A set of grouped signs may record discrete objects, grain capacity, labour time or an accounting conversion. The marks alone do not decide which quantity was represented.[15,17]

Where That Leaves Us

Counting concerns discrete units; measurement concerns comparison with a standard, and units are social and technical achievements rather than given facts. Historical practices intertwine them, but interpreting a record requires identifying the unit system and the represented magnitude, not merely reading a visible number. Grouped signs may stand for objects, capacity, labour time or an accounting conversion, and the notation by itself settles none of those readings.

7. What Archaeological Marks Can and Cannot Establish

Archaeological analysis can often establish material, manufacture, stroke order, spacing and whether groups were added in separate sessions. Those observations may support deliberate accumulation. They do not recover the counted category, the spoken labels or the rule connecting a mark to a unit.[1] Regularity is compatible with several purposes: decoration, grip, practice, ownership, rhythm, game, schedule or tally. Numerical interpretation becomes stronger when independent evidence links repeated marks to a recurrent countable process, when group boundaries fit a demonstrated rule, or when related objects preserve the same convention.[1] Modern observers can always count ancient marks. That fact belongs to the observer's notation, not automatically to the maker's practice. Claims about primes, calendars or arithmetic patterns are especially vulnerable to selection after the fact because many groupings can be discovered in a finite set of incisions.[1,14] Ethnographic analogy can generate hypotheses about how bodies, objects or marks support quantity. It cannot assign an ancient object to a living community's practice without historical connection. Similar solutions may arise independently, while superficially similar objects may participate in different conventions.[12-14] Preservation also filters the record. Perishable cords, wooden sticks, fibres, gestures and spoken

sequences are less likely to survive than bone or stone. The oldest surviving tally-like object is therefore not necessarily the first quantitative aid, and the absence of a durable notation does not demonstrate the absence of exact practice.[13,14]

Where That Leaves Us

The secure archaeological sequence is trace, manufacture, recurrence and context. Quantity, unit and meaning enter only when additional evidence supports the mapping, and the ability of a modern observer to count ancient marks belongs to that observer's notation rather than to the maker's practice. Preservation filters what survives, so the oldest durable tally-like object need not be the earliest quantitative aid.

8. Multiple Routes, Not One Staircase

Counting is best understood as a family of coordinated practices. Approximation compares magnitudes; small-set tracking follows individuals; correspondence matches collections; tallies preserve events; words stabilise ordered values; grouping creates higher units; numerals make values portable; measurement relates magnitudes to standards.[2,9,13,18] These capacities can reinforce one another without appearing in a fixed order. A community may possess exact oral counting with little durable notation, sophisticated measurement with specialised numerals, or material accounting tied closely to commodities. Historical change depends on tasks, teaching, media and institutions.[14,15,18] External devices do not merely store an already complete abstract number concept. By making groupings and correspondences visible, they can change which comparisons and operations are easy to learn and perform. At the same time, no object generates a numerical rule without users who maintain the mapping.[13,19] The diversity of number systems is therefore evidence, not noise. Different bases, body sequences, classifiers, sign values and calculation media reveal that numerical abstraction is constructed through repeatable relations rather than copied from one universal notation.[11,12,18] The most defensible origin story has no single starting object. Biological capacities for quantity are old; exact symbolic systems are cultural; surviving material records are selective; and locally documented transitions differ. The history of number is the history of how communities made quantitative relations stable enough to check, share and extend.[1,2,14]

Where That Leaves Us

There was no necessary march from notch to numeral to modern arithmetic. There were recurrent problems of matching, memory, grouping, exchange and measurement, solved through different combinations of bodies, language, materials and institutions. The diversity of number systems is evidence rather than noise, and the most defensible origin story has no single starting object.

What We Know, and How

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 We Still Don't Know

For most early marks, we do not know whether they represented quantity, which units they tracked, whether their rule was shared, or whether spoken number words accompanied them. We also cannot reconstruct the lost majority of perishable devices and gestures. Nor does present cognitive diversity give a direct chronology for when particular prehistoric communities coordinated exact words, objects and signs.

What Could Change Our Minds

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.

If You Remember Only One Thing

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.

Related Articles

KNO-001 · A Mark on a Stone KNO-003 · From Marks to Symbols KNO-004 · From Symbols to Writing KNO-007 · Calendars and Time Measurement KNO-011 · Measuring Before Theorising

References

- [1] Francesco d'Errico; Luc Doyon; Ivan Colagè; Alain Queffelec; Emma Le Vraux; Giacomo Giacobini; Bernard Vandermeersch; Bruno Maureille. (2018). From number sense to number symbols: an archaeological perspective. *Philosophical Transactions of the Royal Society B* 373, 20160518. DOI: 10.1098/rstb.2016.0518. <https://doi.org/10.1098/rstb.2016.0518>
- [2] Lisa Feigenson; Stanislas Dehaene; Elizabeth Spelke. (2004). Core systems of number. *Trends in Cognitive Sciences* 8, 307-314. DOI: 10.1016/j.tics.2004.05.002. <https://doi.org/10.1016/j.tics.2004.05.002>
- [3] Andreas Nieder. (2016). The neuronal code for number. *Nature Reviews Neuroscience* 17, 366-382. DOI: 10.1038/nrn.2016.40. <https://doi.org/10.1038/nrn.2016.40>
- [4] Peter Gordon. (2004). Numerical cognition without words: Evidence from Amazonia. *Science* 306, 496-499. DOI: 10.1126/science.1094492. <https://doi.org/10.1126/science.1094492>
- [5] Pierre Pica; Cathy Lemer; Véronique Izard; Stanislas Dehaene. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306, 499-503. DOI: 10.1126/science.1102085. <https://doi.org/10.1126/science.1102085>
- [6] Michael C. Frank; Daniel L. Everett; Evelina Fedorenko; Edward Gibson. (2008). Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition* 108, 819-824. DOI: 10.1016/j.cognition.2008.04.007. <https://doi.org/10.1016/j.cognition.2008.04.007>
- [7] Brian Butterworth; Robert Reeve; Fiona Reynolds; Delyth Lloyd. (2008). Numerical thought with and without words: Evidence from indigenous Australian children. *Proceedings of the National Academy of Sciences* 105, 13179-13184. DOI: 10.1073/pnas.0806045105. <https://doi.org/10.1073/pnas.0806045105>
- [8] Elizabet Spaepen; Marie Coppola; Elizabeth S. Spelke; Susan E. Carey; Susan Goldin-Meadow. (2011). Number without a language model. *Proceedings of the*

National Academy of Sciences 108, 3163-3168. DOI: 10.1073/pnas.1015975108. <https://doi.org/10.1073/pnas.1015975108> [9] Karen Wynn. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology* 24, 220-251. DOI: 10.1016/0010-0285(92)90008-P. [https://doi.org/10.1016/0010-0285\(92\)90008-P](https://doi.org/10.1016/0010-0285(92)90008-P) [10] Barbara W. Sarnecka; Susan Carey. (2008). How counting represents number: What children must learn and when they learn it. *Cognition* 108, 662-674. DOI: 10.1016/j.cognition.2008.05.007. <https://doi.org/10.1016/j.cognition.2008.05.007> [11] Andrea Bender; Sieghard Beller. (2012). Nature and culture of finger counting: diversity and representational effects of an embodied cognitive tool. *Cognition* 124, 156-182. DOI: 10.1016/j.cognition.2012.05.005. <https://doi.org/10.1016/j.cognition.2012.05.005> [12] Geoffrey B. Saxe. (1981). Body parts as numerals: a developmental analysis of numeration among the Oksapmin in Papua New Guinea. *Child Development* 52, 306-316. DOI: 10.2307/1129244. <https://doi.org/10.2307/1129244> [13] Karenleigh A. Overmann. (2016). The role of materiality in numerical cognition. *Quaternary International* 405, 42-51. DOI: 10.1016/j.quaint.2015.05.026. <https://doi.org/10.1016/j.quaint.2015.05.026> [14] Karenleigh A. Overmann. (2018). Constructing a concept of number. *Journal of Numerical Cognition* 4, 464-493. DOI: 10.5964/jnc.v4i2.161. <https://doi.org/10.5964/jnc.v4i2.161> [15] Hans J. Nissen; Peter Damerow; Robert K. Englund. (1993). *Archaic Bookkeeping: Early Writing and Techniques of Economic Administration in the Ancient Near East*. University of Chicago Press. <https://cdli.earth/publications/1657148> [16] Denise Schmandt-Besserat. (1992). *Before Writing, Volume I: From Counting to Cuneiform*. University of Texas Press. <https://utpress.utexas.edu/9781477325766/> [17] Christopher Woods; Emily Teeter; Geoff Emberling (editors). (2010). *Visible Language: Inventions of Writing in the Ancient Middle East and Beyond*. Oriental Institute Museum Publications 32. <https://isac.uchicago.edu/research/publications/oimp/oimp-32-visible-language-inventions-writing-ancient-middle-east-and> [18] Stephen Chrisomalis. (2010). *Numerical Notation: A Comparative History*. Cambridge University Press. DOI: 10.1017/CBO9780511676062. <https://doi.org/10.1017/CBO9780511676062> [19] Jiajie Zhang; Donald A. Norman. (1994). Representations in distributed cognitive tasks. *Cognitive Science* 18, 87-122. DOI: 10.1207/s15516709cog1801_3. https://doi.org/10.1207/s15516709cog1801_3 [20] Karl Menninger. (1992). *Number Words and Number Symbols: A Cultural History of Numbers*. Dover Publications.

KNO-003 · From Marks to Symbols

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

1. A Mark Is an Event Before It Is a Sign

The Question

What must be added to a physical difference before that difference represents something for someone? A mark is first a material event: pigment deposited, a surface incised, a branch broken, a hand moved. Its geometry and manufacturing sequence may be observable even when its purpose is unknown. Calling it a sign adds a functional claim: the difference was used or interpreted as standing in a relevant relation to something else.[18,19] That relation may exist without deliberate communication. Footprints can inform a tracker because they were caused by an animal's passage. Dark clouds can guide action because they correlate with weather. Such cues carry information for an observer, but the animal and cloud need not have produced them to alter another person's beliefs.[21] A signal is usually narrower: a behaviour or artifact whose production participates in communication. Human signals commonly depend on an audience's recognition that the producer is trying to make something manifest. Yet intentional production alone does not guarantee success. A private mark whose rule nobody else can infer remains an attempted signal, not an established public symbol.[16,17] The same form can change status across situations. A red patch may be a stain, an index of injury, an intentionally displayed warning or a conventional emblem. The physical mark does not contain one permanent semiotic category. Classification depends on the relation, task and interpretive community.[9,16,21]

Where That Leaves Us

Material description comes first. Representation begins only when evidence supports a relation among a form, an object or state of affairs, and an interpreter. Deliberateness, informativeness and conventionality must be tested separately, because a mark made on purpose whose rule nobody else can infer remains an attempted signal rather than an established public symbol. The physical form carries no permanent semiotic category.

2. Resemblance Opens a Door but Does Not Fix Meaning

The Question

If a sign looks like its referent, has its meaning already been secured? Iconicity is a perceived resemblance or structural correspondence between a sign and some aspect of what it denotes. A drawing can preserve outline; a gesture can imitate motion; a vocalisation can vary speed, duration or pitch in ways listeners map onto an event. Iconicity can make a novel sign easier to infer, especially before a convention exists.[9-11] Resemblance is selective rather than complete. A sketch of a bird may highlight wings, beak, flight or size. Which features count depends on the contrast set, the medium and the observer's prior knowledge. Even an apparently transparent icon therefore relies on choices about what similarity matters.[9] Experiments show that participants can create non-linguistic vocalisations that naive listeners interpret above chance, and that repeated interaction can turn initially iconic forms

into shared signs. These results demonstrate available mechanisms under controlled conditions. They do not show that every spoken word began as imitation or that iconicity follows one universal decline. [10,11] Languages combine iconicity, arbitrariness and systematicity. A sign may resemble its referent, be conventional in its community, and pattern with other signs at the same time. These properties are dimensions, not mutually exclusive boxes, and conventionalisation need not erase iconic motivation.[9]

Where That Leaves Us

Iconicity can bootstrap understanding because resemblance constrains a first guess. It cannot determine a unique interpretation, and it does not remove the need for shared attention, feedback and learned convention. Resemblance is selective, so which features count is itself a choice, and the experiments demonstrate available mechanisms under controlled conditions rather than showing that every word began as imitation.

3. Traces, Pointing and Indexical Grounding

A causal trace connects sign and source through production: smoke comes from combustion, a footprint from contact, a scar from an earlier event. An indexical gesture connects interpretation to the current situation: pointing directs another person's attention toward a location, object or event. Neither relation is equivalent to resemblance.[16,17] Human pointing is especially powerful because it operates within joint attention. Infants begin to use and understand pointing not only to request objects but also to share attention and, in some experiments, to provide information another person lacks. The gesture's meaning depends on a shared scene and on inferences about why this object is being indicated now. [16,17] The visible direction of a point underdetermines its message. Pointing at a cup might mean "take it", "that one is mine", "look at the crack" or "put it there". Gaze, timing, prior discourse and the participants' common task narrow the possibilities. Indexical anchoring reduces uncertainty; it does not eliminate interpretation.[16] Material signs can also inherit indexical force from placement. A marker beside a path, an object sealed with an account, or a sign fixed to a doorway can be understood partly through where it occurs. Moving the same form to another setting may change or destroy its function even if its shape is preserved.[18,19] This explains why context loss is so damaging in archaeology. Provenance, association, orientation, wear and repetition can show how a marked object participated in activity. Removing a motif from that network and comparing only its outline sacrifices much of the evidence that could distinguish decoration, ownership, counting, instruction or display.[18,19]

Where That Leaves Us

Indices recruit causal history and present context. They can anchor a sign to the world, but their interpretation still depends on what observers know about the situation and one another's purposes. Placement carries part of that force, so the same form moved to another setting may lose its function with its shape intact, which is why lost provenance costs an archaeological reading so much.

4. Intention and Shared Attention

Communicative intention is recursive. A producer may act so that an audience recognises the attempt to direct attention or understanding, and the audience interprets the act partly by recognising that attempt. Research on shared intentionality treats joint attention, role coordination and motive reading as central resources in human cultural cognition.[16] This does not make communication a matter of mind reading without evidence. People use gaze, bodily orientation, conventional frames, previous turns and expected relevance to infer what an act is doing. Misunderstanding remains possible, and public

responses reveal whether interpretations converge.[4,16] Children's symbolic development shows that understanding a symbol involves a dual relation. A model, picture or word is an object in its own right and also a representation of something else. Young children can focus on the concrete object so strongly that the representational relation becomes difficult, while experience with pictures, language and pretense supports increasingly flexible use.[15] Intentional communication and symbolic understanding therefore should not be collapsed. Someone may intentionally point without using an arbitrary conventional symbol; someone may competently read a conventional emblem without knowing who produced the token in front of them. Intention concerns an act; convention concerns a repeatable social rule.[15-17]

Where That Leaves Us

Shared attention creates a scene in which a sign can be interpreted and corrected. It helps establish meaning, but stable symbols require a history that extends beyond one successful act. Intention and convention should not be collapsed: intention concerns an act, convention concerns a repeatable social rule, and either can be present in a community without the other.

5. Interaction Turns Guesses into Conventions

Laboratory experimental semiotics places participants in tasks where ordinary language is restricted and new communication systems must be built. These studies do not recreate language origins. They expose, under controlled conditions, how invention, feedback, alignment and repeated use can transform signs.[1-3] In graphical communication experiments, partners initially produced relatively elaborate drawings. Through interaction, successful forms became simpler and more schematic. Mere repetition was not enough: feedback and coordination mattered, and outsiders who lacked the interaction history often found the reduced signs difficult to interpret.[2] This pattern shows how information can shift location. At first, much of the clue lies visibly in the sign's resemblance. Later, a shorter mark works because users remember earlier successful uses. Efficiency for insiders can therefore reduce immediate transparency for newcomers.[2,3] Conversation research describes a related process of grounding. Referring expressions are jointly refined until participants have enough evidence for the current purpose. Partners can form conceptual pacts, temporarily favouring one description for an object because it belongs to their shared interaction history.[4,5] A convention need not be explicitly negotiated or formally declared. It can emerge from repeated coordination when a community expects a form to be used and interpreted in a particular way, observes departures and repairs failures. What matters is not unanimous private association but sufficiently stable public practice.[1,4]

Where That Leaves Us

Symbols can grow from interaction because feedback selects workable mappings and shared history supports compression. Their apparent arbitrariness is often the residue of a path that outsiders did not witness. Repetition alone does not produce that compression, coordination does, and the efficiency insiders gain is paid for in transparency for anyone who arrives later.

6. Transmission Reshapes What Interaction Creates

A sign that works for one pair still faces another test: can people who did not participate in its invention learn and use it? Cultural transmission introduces a bottleneck. Learners receive limited examples, reconstruct a system and pass on their own productions rather than a perfect copy of every precedent.

[6-8] Iterated-learning experiments show that artificial languages can become more structured and easier to learn across transmission chains. Learnability alone can also produce degenerate systems that collapse distinctions. When transmission is combined with communicative use, pressure for compression is balanced by pressure to keep relevant meanings distinguishable.[6,7] Experiments with silent gesture similarly find that interaction and transmission make different contributions. Improvised pantomimes can become more regular and systematic when successive learners also communicate with one another. The resulting changes are population-level outcomes, not evidence that any participant designed the final system in advance.[20] Cumulative cultural evolution can preserve useful modifications and allow later users to build on them. Laboratory tasks show that social learning may generate convergence and adaptive complexity, but the outcome depends on task structure, available feedback and what is transmitted. These models identify mechanisms, not a universal timetable for symbolic evolution.[8] Transmission also creates normativity. A learner can produce a comprehensible variant and still be corrected because the community recognises a usual form. Once a precedent is teachable, repeatable and policed, symbols acquire histories of right and wrong use that exceed individual intention.[3,6]

Where That Leaves Us

Interaction establishes local common ground; transmission tests whether it can survive absent founders. Symbol systems become durable when they remain both learnable and discriminating under repeated reconstruction, since learnability by itself can collapse the distinctions a system needs. These are population-level outcomes rather than anyone's design, and they identify mechanisms rather than a timetable.

7. New Languages Reveal Structure Emerging

Homesign systems created by deaf individuals without access to a conventional sign language demonstrate that gesture can develop language-like organisation without a full linguistic model. These systems are not merely copies of surrounding speakers' gestures, although they arise within rich social environments and remain shaped by interaction.[12] Nicaraguan Sign Language developed as deaf children entered communities where they could communicate with peers. Comparative cohort research found increasing segmentation and systematic expression of motion events, showing that children were not only acquiring a finished language but also transforming a developing one.[13] Al-Sayyid Bedouin Sign Language provides another community case. Research documented systematic grammatical organisation in a relatively young language. This is one trajectory rather than a simple sequence toward a predetermined endpoint; different social networks, cohort structures and contact conditions can support different trajectories.[14] These cases establish that conventional linguistic structure can emerge within observable communities. They do not provide transparent fossils of the first human language, nor do they imply that mature sign languages are more pictorial or less complex than spoken languages. Modality changes available resources, not the status of a system as language.[12-14] The strongest inference is mechanistic: repeated use among interacting learners can create stable contrasts and combinatorial organisation. The historical origin of symbolic communication remains a broader question involving capacities, ecologies and cultural histories not reproduced by any one modern community.[13,14,20]

Where That Leaves Us

Emerging languages let researchers observe convention and structure taking shape. They illuminate processes of formation while retaining their own histories and refusing the role of living stand-ins for prehistory. Different social networks and cohort structures support different trajectories, and modality changes the resources available to a system rather than its status as a language.

8. What Ancient Marks Can and Cannot Tell Us

Engraved ochre from Blombos Cave demonstrates deliberate production of geometric patterns in a secure Middle Stone Age context. Microscopic and technological analyses can address how lines were made and whether actions formed an organised sequence. The surviving object does not disclose a spoken label or unique message.[18] Broader archaeological studies can compare motifs, manufacturing gestures, recurrence and spatial organisation. Repeated conventions across objects and contexts strengthen a claim that marks participated in shared practice. Even then, "symbolic" can name several propositions, from socially recognisable display to reference to an absent category, and each requires its own support.[18,19] Visual similarity across distant times or regions is weak evidence of identical meaning. Simple crosses, grids and parallel lines are easy to produce independently. A persuasive historical connection needs contextual continuity, shared production rules, transmission evidence or associated practices, not resemblance alone.[18,19] Absence is equally difficult to interpret. Gesture, body paint, fibre, wood, sand drawings and speech often disappear, while stone and bone survive selectively. The earliest durable mark is not necessarily the earliest symbol, and a community without preserved graphics may have possessed complex conventional communication.[19] Claims should therefore climb an evidential ladder: mark; deliberate manufacture; patterned recurrence; socially transmitted convention; referential function; specific recovered meaning. Archaeology often supports the first three and sometimes the fourth. The last two usually require an inscriptional, ethnographic or otherwise independent bridge.[18,19]

Where That Leaves Us

Ancient marks can establish sophisticated making and recurrent practice without yielding a dictionary. The honest reconstruction preserves the distance between patterned behaviour and decoded meaning. Symbolic names several different propositions, each needing its own support, and because durable media survive selectively the earliest preserved mark need not be the earliest symbol.

9. No Single Staircase from Image to Word

It is tempting to arrange traces, icons, indices and arbitrary symbols as successive evolutionary steps. Actual systems mix them. A pointing diagram may combine resemblance, spatial indexing and learned notation; a spoken word may be historically arbitrary yet contain sound symbolism; a conventional gesture may remain partly iconic.[9] Nor does simplification always mean progress toward abstraction. Signs can become shorter for practiced users, more elaborate when audiences diversify, or more iconic when a new contrast must be introduced. Change responds to communicative ecology, memory, medium and institutions.[2,3,7] Material persistence alters the scale of coordination. A durable mark can be inspected after its maker leaves, copied by people who never met the inventor and embedded in places or objects that constrain interpretation. Persistence enables wider conventions, but it also separates signs from the immediate feedback that originally grounded them.[1,18] The origin of symbols is therefore not one moment when marks acquired meaning. It is the repeated achievement of

communities that align attention, test mappings, remember successful uses, teach newcomers and maintain distinctions. Forms matter because they make some inferences and reproductions easier; conventions matter because form alone never fixes the relation.[1-3,6]

Where That Leaves Us

The passage from marks to symbols is a change in coordinated practice, not a predictable sequence of shapes. Resemblance, context, intention, interaction, memory and transmission contribute in different combinations. Simplification is not progress toward abstraction, and durable media widen coordination while separating signs from the immediate feedback that first grounded them.

What We Know, and How

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 We Still Don't Know

We do not know one sequence by which human symbolic communication began, how much early convention depended on gesture, voice or perishable media, or when particular marks first referred to absent entities rather than organising immediate action. We also lack a general measure of when a local coordination becomes a community-wide norm. Modern experiments and emerging languages reveal mechanisms, but no surviving dataset directly observes the first transition.

What Could Change Our Minds

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.

If You Remember Only One Thing

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.

Related Articles

KNO-001 · A Mark on a Stone KNO-002 · Counting Before Numbers KNO-004 · From Symbols to Writing KNO-005 · Oral Language and Transmission KNO-009 · Archives, Libraries and Collective Memory

References

- [1] Bruno Galantucci. (2005). An Experimental Study of the Emergence of Human Communication Systems. *Cognitive Science* 29(5), 737-767. DOI: 10.1207/s15516709cog0000_34. https://doi.org/10.1207/s15516709cog0000_34 [2] Simon Garrod; Nicolas Fay; John Lee; Jon Oberlander; Tracy MacLeod. (2007). Foundations of Representation: Where Might Graphical Symbol Systems Come From? *Cognitive Science* 31(6), 961-987. DOI: 10.1080/03640210701703659. <https://doi.org/10.1080/03640210701703659> [3] Nicolas Fay; Simon Garrod; Leo Roberts; Nik Swoboda. (2010). The Interactive Evolution of Human Communication Systems. *Cognitive Science* 34(3), 351-386. DOI: 10.1111/j.1551-6709.2009.01090.x. <https://doi.org/10.1111/j.1551-6709.2009.01090.x> [4] Herbert H. Clark; Deanna Wilkes-Gibbs. (1986). Referring as a collaborative process. *Cognition* 22(1), 1-39. DOI: 10.1016/0010-0277(86)90010-7. [https://doi.org/10.1016/0010-0277\(86\)90010-7](https://doi.org/10.1016/0010-0277(86)90010-7) [5] Susan E. Brennan; Herbert H. Clark. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(6), 1482-1493. DOI: 10.1037/0278-7393.22.6.1482. <https://doi.org/10.1037/0278-7393.22.6.1482> [6] Simon Kirby; Hannah Cornish; Kenny Smith. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences* 105(31), 10681-10686. DOI: 10.1073/pnas.0707835105. <https://doi.org/10.1073/pnas.0707835105> [7] Simon Kirby; Monica Tamariz; Hannah Cornish; Kenny Smith. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141, 87-102. DOI: 10.1016/j.cognition.2015.03.016. <https://doi.org/10.1016/j.cognition.2015.03.016> [8] Christine A. Caldwell; Ailsa E. Millen. (2008). Studying cumulative cultural evolution in the laboratory. *Philosophical Transactions of the Royal Society B* 363(1509), 3529-3539. DOI: 10.1098/rstb.2008.0133. <https://doi.org/10.1098/rstb.2008.0133> [9] Mark Dingemanse; Damián E. Blasi; Gary Lupyan; Morten H. Christiansen; Padraic Monaghan. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences* 19(10), 603-615. DOI: 10.1016/j.tics.2015.07.013. <https://doi.org/10.1016/j.tics.2015.07.013> [10] Marcus Perlman; Rick Dale; Gary Lupyan. (2015). Iconicity can ground the creation of vocal symbols. *Royal Society Open Science* 2(8), 150152. DOI: 10.1098/rsos.150152. <https://doi.org/10.1098/rsos.150152> [11] Marcus Perlman; Gary Lupyan. (2018). People Can Create Iconic Vocalizations to Communicate Various Meanings to Naïve Listeners. *Scientific Reports* 8, 2634. DOI: 10.1038/s41598-018-20961-6. <https://doi.org/10.1038/s41598-018-20961-6> [12] Susan Goldin-Meadow; Heidi Feldman. (1977). The Development of Language-Like Communication Without a Language Model. *Science* 197(4301), 401-403. DOI: 10.1126/science.877567. <https://doi.org/10.1126/science.877567> [13] Ann Senghas; Sotaro Kita; Asli Özyürek. (2004). Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua. *Science* 305(5691), 1779-1782. DOI: 10.1126/science.1100199. <https://doi.org/10.1126/science.1100199> [14] Wendy Sandler; Irit Meir; Carol Padden; Mark Aronoff. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences* 102(7), 2661-2665. DOI: 10.1073/pnas.0405448102. <https://doi.org/10.1073/pnas.0405448102> [15] Judy S. DeLoache. (2004). Becoming symbol-minded. *Trends in Cognitive Sciences* 8(2), 66-70. DOI: 10.1016/j.tics.2003.12.004. <https://doi.org/10.1016/j.tics.2003.12.004> [16] Michael Tomasello; Malinda Carpenter; Josep Call; Tanya Behne; Henrike Moll. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28(5), 675-691. DOI: 10.1017/S0140525X05000129. <https://doi.org/10.1017/S0140525X05000129> [17] Ulf Liszkowski; Malinda Carpenter; Tricia Striano; Michael Tomasello. (2006). 12- and 18-Month-Olds Point to Provide Information for Others. *Journal of Cognition and Development* 7(2), 173-187. DOI: 10.1207/s15327647jcd0702_2. https://doi.org/10.1207/s15327647jcd0702_2 [18] Christopher S. Henshilwood; Francesco d'Errico; Royden Yates; Zenobia Jacobs; Chantal Tribolo; Geoff A. T. Duller; Norbert Mercier; Judith C. Sealy; Hélène Valladas; Ian Watts; Ann G. Wintle. (2002). Emergence of Modern Human Behavior: Middle Stone Age Engravings from South Africa. *Science* 295(5558), 1278-1280. DOI:

10.1126/science.1067575. <https://doi.org/10.1126/science.1067575> [19] Francesco d'Errico; Luc Doyon; Ivan Colagè; Alain Queffelec; Emma Le Vraux; Giacomo Giacobini; Bernard Vandermeersch; Bruno Maureille. (2018). From number sense to number symbols: an archaeological perspective. *Philosophical Transactions of the Royal Society B* 373(1740), 20160518. DOI: 10.1098/rstb.2016.0518. <https://doi.org/10.1098/rstb.2016.0518> [20] Yasamin Motamedi; Marieke Schouwstra; Kenny Smith; Jennifer Culbertson; Simon Kirby. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition* 192, 103964. DOI: 10.1016/j.cognition.2019.05.001. <https://doi.org/10.1016/j.cognition.2019.05.001> [21] Terrence W. Deacon. (1997). *The Symbolic Species: The Co-evolution of Language and the Brain*. W. W. Norton & Company. <https://www.norton.com/books/9780393317541>

KNO-004 · From Symbols to Writing

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

Article Abstract

Ask when writing began and the answer depends entirely on what is being counted. Notched bone, potters' marks, accounting tokens, knotted cords and this sentence have all been called writing. KNO-004 separates six things that the single word collapses: mark, symbol, notation, proto-writing, partial writing and full writing. It treats them as analytic categories rather than rungs of a ladder, because the record shows societies crossing them in different orders, stopping at different points, and building powerful recording systems that never became writing and never tried to. The threshold it uses is glottography, the encoding of a specific utterance in a specific language such that a reader recovers wording rather than gist. It examines four traditions usually called independent inventions, one case still genuinely disputed, and the inventions documented in historical time, and closes with what the evidence cannot settle.

In 60 Seconds

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 This Question 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.

1. The problem with the word "writing"

Ask when writing began and the answer depends entirely on what is being counted. Marks on bone from the Upper Palaeolithic, tallies, potters' marks, seal impressions, accounting tokens, Andean knotted cords, mathematical notation and this sentence have all been called writing by someone. They are not the same kind of thing, and the differences are not matters of degree.

The confusion is old and it has a cause. The word carries two distinct ideas at once. One is the general idea of durable graphic communication, of making a mark that carries meaning to someone absent in space or time. The other is the specific idea of encoding language, of a system that captures a particular utterance so completely that a competent reader can reproduce it. Almost every dispute about the origins of writing turns out, on inspection, to be a dispute about which of these two is meant.

This article uses the second sense for the word writing and gives the first sense its own vocabulary. That is a stipulation, not a discovery, and it is worth stating plainly rather than smuggling in.

Where That Leaves Us

Almost every dispute about the origins of writing is a dispute about which of two ideas the word carries: durable graphic communication, or the encoding of a particular utterance. This article uses the second and gives the first its own vocabulary. That is a stipulation and not a discovery, which is why it is stated plainly, and it is what makes the distinctions that follow necessary rather than tidy.

2. Six categories, not six stages

The following distinctions are analytic. They classify what a system does, not when it appeared or what it turned into.

Mark. A deliberate graphic trace whose meaning, if any, is not recoverable from the trace itself. A notch, a stroke, a scratch. Marks may be counters, ownership signs, decoration, or nothing at all. The interpretive problem is that a mark and a sign look identical when the context is lost.

Symbol. A graphic form conventionally associated with a referent, where the association is shared across a community rather than private. A symbol carries meaning without carrying language. A crossed circle can mean a place, a god, a clan or a commodity, and its users can read it whatever language they speak, **so long as they share the convention**. What crosses a language boundary is the convention and not the sign, and a sign whose convention is not shared carries nothing.

Notation. A closed system of symbols with rules of combination, restricted to one domain. Numerical notation is the clearest case: a small inventory, strict syntax, unlimited productivity within its domain, and no capacity to say anything outside it. Notation is not defective writing. It is a different instrument, and it can outperform writing at what it does. The comparative history of numerical systems shows notation developing along its own constrained pathways, in societies with and without writing (Chrisomalis 2010).

Proto-writing. A symbol system with enough conventional signs, enough regularity and enough combinatorial behaviour to record structured content, but without a demonstrable link to the sounds or grammar of a specific language. Proto-writing can record what was owed, to whom and how much. It cannot record why, or what someone said about it.

Partial writing. A system that does encode language, but only within a bounded domain and only for a restricted range of utterances. Early administrative scripts are the type case: they encode names, quantities and commodity terms in a real language, with real phonetic signs, and nothing else. A reader recovers the wording of a receipt and could not use the same system to record a complaint.

Full writing. A system capable of encoding any utterance in its language, including utterances the system was not designed for. This is the property that matters and it is the one that makes decipherment possible in principle, because an open system leaves grammatical traces: function words, affixes, agreement, sign frequencies that behave like a language rather than like a list.

The reason for insisting on six categories rather than a line is that the record does not run along the line. Numerical notation and full writing coexist and always have. Proto-writing sometimes precedes writing in the same tradition and sometimes has no descendants. Partial writing sometimes expands into full writing over centuries and sometimes does not expand at all. And the categories are properties of systems in use, so the same signs can belong to different categories in different hands.

Where That Leaves Us

Mark, symbol, notation, proto-writing, partial writing and full writing classify what a system does, not when it appeared or what it later became. A symbol carries meaning without carrying language, and what crosses a language boundary is the convention rather than the sign. Notation is not defective writing but a different instrument, and reading any of these as a rung converts a classification into a history it cannot support.

3. The threshold: phonetic transfer

What converts a symbol system into a writing system is the discovery that a sign can stand for the sound of its name rather than for its meaning. Once a sign for a reed can be used to write any syllable that sounds like the word for reed, the inventory stops being a list of things and becomes a way of spelling. This is usually called the rebus principle, and it is the mechanism visible in the independently attested transitions this article examines. **Whether every such transition anywhere took that route is not established here**, and the cases below are read as cases rather than as instances of a rule.

Its importance is structural rather than clever. Before phonetic transfer, the system can only grow by adding signs, and it can only express what it has signs for. After phonetic transfer, the system can express anything sayable, including proper names, foreign words, abstractions and grammatical endings, with an inventory that stops growing and begins to shrink. The signature of the crossing is therefore not a new sign but a new economy: sign inventories that stabilise, then contract, while the range of expressible content expands without limit.

This gives a workable indicator, and an indicator is not a test. A system whose sign count keeps rising with the range of content is unlikely to be glottographic. A system whose sign count falls while its content broadens is behaving as a glottographic system does, **which is evidence and not a demonstration**: classifying a system needs converging evidence, from sign frequencies, from grammatical traces and from the range of content actually recorded, and the sign count alone settles nothing.

Where That Leaves Us

Phonetic transfer is the mechanism this article follows: a sign able to stand for the sound of its name stops being an entry in a list and becomes a way of spelling. Its signature is an economy rather than a shape, with inventories that stabilise and then contract while expressible content stops having a limit. That yields a workable indicator and not a test, so classification still needs converging evidence.

4. Four traditions usually treated as independent

Four traditions are conventionally described as inventing writing without a model. The word usually is doing real work in that sentence, and section 7 returns to it.

4.1 Southwest Asia

Mesopotamia offers the only case where something close to the whole sequence is visible in stratified material. The fourth millennium yields clay tokens, sealed clay envelopes, impressed numerical tablets and, in the late fourth millennium, proto-cuneiform tablets from Uruk, followed over several centuries by cuneiform capable of recording Sumerian and later Akkadian in full. **The order in that list is the order of first attestation and not a line of descent.** The media overlap in use, several of them are in circulation at once, and which of them is ancestral to which is disputed rather than settled.

The influential proposal that the clay tokens are the direct ancestors of the signs, with three dimensional counters flattening into two dimensional impressions and then into drawn characters, was set out at length by Schmandt Besserat (1992). It has never commanded consensus. Reviewers questioned both the geographic argument and the proliferation of token shapes on which the sign correspondences depend (Van De Mieroop 1994). The honest summary is that tokens and early tablets belong to one accounting world and that a specific genealogy from token shape to sign shape remains contested rather than established.

Recent work has widened the pool of possible ancestors beyond tokens. A study of late pre literate cylinder seal imagery identifies motifs, including a fringed cloth and a netted vessel, that recur as proto cuneiform signs used for the same commodities, arguing that seal iconography and emerging script were parts of one clay based communication practice rather than a source and a derivative (Kelley, Cartolano and Ferrara 2025). This does not resolve the token question. It reframes it, by removing the assumption that the ancestry has to be single.

Two features of proto cuneiform deserve emphasis because they are what the six categories are for. Its known texts are administrative, they show no syntax, and no reading of them as sentences in a language has been established. By the definitions above it falls on the proto writing side, or at most the partial writing side, **and that placement is a reading of the surviving corpus rather than a property demonstrated of the system.** Several further centuries separate it from cuneiform recording Sumerian in full.

4.2 Egypt

The tomb labelled U j at Abydos yielded roughly 160 small incised bone and ivory labels which the excavator identified as the earliest Egyptian writing, from the late fourth millennium (Dreyer 1998). Radiocarbon work on the Umm el Qaab necropolis provides the chronological frame for the sequence of royal tombs there (Görsdorf, Dreyer and Hartung 1998), and the imported jars from the tomb have been reassessed petrographically and contextually more recently (Ownby, Hartung and Sowada 2022).

The Egyptian case is important for a reason that has nothing to do with priority. The earliest labels already use phonetic signs. No long apprenticeship is visible in the surviving material, and no stage of purely semantic Egyptian symbols slowly acquiring sound values is attested. That stage may have happened on materials that do not survive; it may not have happened; or it may have happened on surviving materials that have not been recognised as such. **The alternatives are not exhausted by the first two**, and none of them makes the independence of the Egyptian invention a settled fact either way.

4.3 China

The oracle bone inscriptions of the late Shang capital at Anyang are already a mature writing system when they appear. They encode a real language, with grammatical particles and productive phonetic elements, and they can be read. Radiocarbon dating of oracle bones places the reign of Wu Ding, from which the earliest large corpus comes, at roughly the thirteenth century BCE, with the Shang ending around the mid eleventh (Liu et al. 2021).

The problem is the gap behind them. Neolithic incised signs are known from much earlier sites, including tortoise shell and bone marks from Jiahu in the seventh millennium BCE, which were published under a question mark rather than a claim (Li et al. 2003). Nothing demonstrates continuity between those marks and Shang writing, and several thousand years separate them. The Chinese case is therefore the clearest illustration of a general problem: a mature script with no visible ancestry tells us that the ancestry was written on something that rots.

4.4 Mesoamerica

Mesoamerica supplies the strongest case for genuinely independent invention among those discussed here, because no plausible contact route to Old World scripts has been proposed. **That is an argument from the absence of a proposed route**, which is the form section 8 warns is the weakest available, and it is offered as such. It also supplies the most instructive dispute about evidence.

The Cascajal Block, a serpentine slab bearing sixty two signs, was published as the oldest writing in the New World and assigned to the San Lorenzo phase ending around 900 BCE (Rodríguez Martínez et al. 2006). The block was not excavated. Road builders recovered it from a debris pile in the late 1990s, and the publication states plainly that it is divorced from its stratigraphic context and dated by association with surrounding material and by iconographic analysis. Whatever the block is, its date rests on inference from an unstratified find, and that should be said whenever it is cited.

Zapotec carved stones from the Valley of Oaxaca provide a better anchored early case, with Monument 3 at San José Mogote bearing a calendrical day name beside a captive figure (Marcus 1976). Maya writing is fully glottographic and decipherable, and a day sign recovered from sealed contexts at San Bartolo pushes attested Maya calendrical writing back to roughly the third to second century BCE (Stuart et al. 2022), following the earlier publication of hieroglyphic text from the same site (Saturno, Stuart and Beltrán 2006).

Where That Leaves Us

The four traditions are conventionally called independent inventions, and the word conventionally carries the weight. Mesopotamia's sequence is an order of first attestation rather than a line of descent, the token-to-sign genealogy has never commanded consensus, and the Cascajal Block is dated by association from an unstratified find. Mesoamerican independence rests on no contact route having been proposed, which is the weakest form available.

5. The case that shows why the threshold matters

The Indus Valley signs are the best available test of the framework, precisely because their status is unresolved.

One position holds that the corpus is not writing at all. The inscriptions are extremely short, averaging under five signs with the longest known at seventeen; a quarter of the sign inventory occurs only once; the sign repetition patterns do not behave as phonetic scripts do; and the archaeological correlates of manuscript production are absent. On this reading the signs are a non linguistic emblem system comparable to Near Eastern seal and badge conventions (Farmer, Sproat and Witzel 2004).

The opposing position applies statistical measures of sequence structure. Conditional entropy computed over the Indus corpus falls closer to that of known languages than to that of non linguistic sign systems, which the authors offer as evidence, explicitly not proof, that the signs encode language (Rao et al. 2009). The method was then contested on its own ground, with a wider battery of statistical tests reported as failing to separate linguistic from non linguistic systems reliably (Sproat 2014).

Two things follow. First, the dispute is not about how old the Indus signs are or how sophisticated the society was. It is about whether a symbol system crossed the glottographic threshold, which is exactly the distinction section 2 sets out. Second, statistical structure alone has so far not settled it. Short texts carry little signal, and the argument from sequence statistics has proved reversible.

Where That Leaves Us

The Indus corpus is the best available test of the framework precisely because its status is unresolved. Short inscriptions, a large one-off sign inventory and absent manuscript correlates support a non-linguistic reading; conditional entropy places the corpus nearer known languages; a wider battery of tests was then reported as failing to separate the two classes reliably. The dispute is about glottography, not about age or sophistication.

6. Systems that were never trying to become writing

Treating writing as the destination makes every other recording system look like an attempt that fell short. This is a serious historiographic error and it distorts what those systems were for.

Numerical notation is the plainest counterexample. It is closed, precise, unlimited within its domain and completely unable to encode language, and it has never been under pressure to change (Chrisomalis 2010). Mathematical, chemical and musical notations behave the same way. They are semasiographic by design: they represent content directly rather than through language, which is why they cross language boundaries without translation.

Mesoamerican and Andean recording systems have been analysed on the same terms, as literacies in their own right rather than as approximations to alphabetic writing (Boone and Mignolo 1994). The Inka khipu, a system of knotted cords, is the hardest case in that group. Its accounting function is not in doubt; the proposal that it also encodes non numerical content through a binary combinatorial coding is a live hypothesis rather than a demonstrated fact (Urton 2003).

The general point is that semasiography is not a stage. It is a permanent alternative, it coexists with writing everywhere writing exists, and in several domains it is the better tool.

Where That Leaves Us

Numerical, mathematical, chemical and musical notations are closed, precise, unlimited within their domains and unable to encode language, and they have never been under pressure to change.

Mesoamerican and Andean systems have been analysed as literacies in their own right rather than as approximations. The khipu's accounting function is not in doubt while its proposed non-numerical coding remains a hypothesis. Semasiography is a permanent alternative, not a stage.

7. How fast the step can be: the documented inventions

The strongest evidence against a slow universal ladder comes from cases observed in historical time.

The Cherokee syllabary was devised by Sequoyah and was in use by the early 1820s, receiving Cherokee council recognition in 1821 (Cushman 2011). It is a complete syllabary for a real language, produced without a long developmental sequence, by someone who knew that writing existed among neighbours. Archaeological work has been brought to bear on the question of its independence (Weeks and Tankersley 2011).

The Vai syllabary of Liberia followed within about a decade, and the question of whether it was influenced by Cherokee through Liberian settlers has been examined in detail rather than assumed either way (Tuchscherer and Hair 2002).

These cases isolate the variable that the archaeological cases cannot. What takes centuries is not the invention of writing. It is the invention of the idea that language can be written at all. In the two cases documented in historical time, once that idea was available, even at second hand and without knowledge of any particular script, a functioning system was built within a few years. **Two cases are what this supports**, and they show that the step can be fast rather than that it always is. This is stimulus diffusion, and it means that independence claims for the ancient cases require an argument about what their inventors could have known, not merely a demonstration that their signs look unlike anyone else's.

Where That Leaves Us

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, its possible Cherokee influence examined rather than assumed. These cases isolate what the archaeological ones cannot: the slow part is not building a script but the arrival of the idea that language can be written. Two cases support that the step can be fast.

8. What the evidence cannot settle

Four limits deserve to be stated rather than managed.

Preservation. Early writing survives on clay, stone, bone and shell. It was also written on papyrus, bark, bamboo, wood, cloth and skin, which do not survive in most climates. An apparently sudden appearance of a mature script is compatible with a long invisible development, and further excavation in a medium that does not preserve cannot rule that out. **Compatible is not the same as likely**, and the bias runs in one direction only: it can hide a development that happened and cannot manufacture one that did not.

Negative claims about independence. Showing that two scripts differ does not show that one was invented without knowledge of the other, because stimulus diffusion transmits the idea and not the forms. Independence is therefore an argument from opportunity and contact, and those arguments get weaker as evidence of long distance contact accumulates.

Undeciphered corpora. For a script with no bilingual, no known descendant language and short texts, the question of whether it is writing may not be answerable with present methods. That is a statement about method, not a verdict on the society.

The categories themselves. The six terms in section 2 are instruments, not natural kinds. A different partition would produce different boundary cases. What the partition should not do is convert into a chronology, and the main claim of this article is that the standard chronological reading of those terms is the error to avoid.

Where That Leaves Us

Four limits are stated rather than managed. Preservation can hide a development that happened and cannot manufacture one that did not. Showing that two scripts differ does not show independent invention, because stimulus diffusion transmits the idea and not the forms. An undeciphered corpus may be unanswerable by present methods, which is a statement about method. And the six terms are instruments rather than natural kinds.

9. Conclusion

Writing did not emerge once, and it did not emerge by a single route. Symbol systems became writing when signs began to carry sound as well as sense, and that step was taken at different times, for different reasons, in traditions that reached it from different starting points. Some traditions took it early and expanded slowly. One appears to have taken it immediately. Others built recording systems of great power and never took it, because their purposes did not require it.

The useful question is not when writing began. It is what a given system could do, for whom, and what it could not do. Answering that requires the distinctions, and it forbids the ladder.

Where That Leaves Us

Writing did not emerge once or by a single route. Signs began to carry sound as well as sense at different times and for different reasons, in traditions reaching it from different starting places, while others built powerful recording systems and never took the step, because their purposes did not require it. The useful question is what a given system could do, for whom, and what it could not.

What We Know, and How

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 We Still Don't Know

Whether the Indus signs encode language is unresolved: entropy measures placed them nearer known languages, that method was contested on its own ground, and no bilingual exists. Whether clay tokens are the direct ancestors of cuneiform signs has never commanded consensus. Whether every transition to writing took the phonetic-transfer route is not established here. Whether the Inka khipu encodes non-numerical content remains a live hypothesis rather than a demonstrated fact.

What Could Change Our Minds

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.

If You Remember Only One Thing

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.

Related Articles

- KNO-002 · *Counting Before Numbers*
- KNO-003 · *From Marks to Symbols*
- KNO-005 · *Oral Language and Transmission*

References

Boone, Elizabeth Hill, y Walter D. Mignolo (eds.). 1994. *Writing Without Words: Alternative Literacies in Mesoamerica and the Andes*. Durham: Duke University Press. ISBN 978-0-8223-1377-9 (tapa dura); 978-0-8223-1388-5 (rústica). Chrisomalis, Stephen. 2010. *Numerical Notation: A Comparative History*. Cambridge: Cambridge University Press. ISBN 978-0-521-87818-0. Cushman, Ellen. 2011. «The Cherokee Syllabary: A Writing System In Its Own Right». *Written Communication* 28 (3): 255–281. DOI 10.1177/0741088311410172. Daniels, Peter T., y William Bright (eds.). 1996. *The World's Writing Systems*. New York y Oxford: Oxford University Press. ISBN 978-0-19-507993-7. Dreyer, Günter. 1998. *Umm el-Qaab I: Das prädynastische Königsgrab U-j und seine frühen Schriftzeugnisse*. Deutsches Archäologisches Institut, Abteilung Kairo, Archäologische Veröffentlichungen 86. Mainz: Verlag Philipp von Zabern. ISBN 978-3-8053-2486-1. Farmer, Steve, Richard Sproat, y Michael Witzel. 2004. «The Collapse of the Indus-Script Thesis: The Myth of a Literate Harappan Civilization». *Electronic Journal of Vedic Studies* 11 (2): 19–57. URL estable: <https://safarmer.com/fsw2.pdf> Ferrara, Silvia, Barbara Montecchi, y Miguel Valério (eds.). 2024. *Writing from Invention to Decipherment*. Oxford: Oxford University Press. ISBN 978-0-19-890874-6. DOI 10.1093/oso/9780198908746.001.0001. Görsdorf, Jochen, Günter Dreyer, y Ulrich Hartung. 1998. «New 14C Dating of the Archaic Royal Necropolis Umm el-Qaab at Abydos (Egypt)». *Radiocarbon* 40 (2): 641–647. DOI 10.1017/S0033822200018579. Houston, Stephen D. (ed.). 2004. *The First Writing: Script Invention as History and Process*.

Cambridge: Cambridge University Press. ISBN 978-0-521-83861-0. Kelley, Kathryn, Mattia Cartolano, y Silvia Ferrara. 2025. «Seals and signs: tracing the origins of writing in ancient South-west Asia». *Antiquity* 99 (403): 64–82. DOI 10.15184/aqy.2024.165. Li, Xueqin, Garman Harbottle, Juzhong Zhang, y Changsui Wang. 2003. «The earliest writing? Sign use in the seventh millennium BC at Jiahu, Henan Province, China». *Antiquity* 77 (295): 31–44. DOI 10.1017/S0003598X00061329. Liu, Kexin, Xiaohong Wu, Zhiyu Guo, Sixun Yuan, Xingfang Ding, Dongpo Fu, y Yan Pan. 2021. «Radiocarbon Dating of Oracle Bones of Late Shang Period in Ancient China». *Radiocarbon* 63 (1): 155–175. DOI 10.1017/RDC.2020.90. Marcus, Joyce. 1976. «The Origins of Mesoamerican Writing». *Annual Review of Anthropology* 5: 35–67. DOI 10.1146/annurev.an.05.100176.000343. Ownby, Mary F., Ulrich Hartung, y Karin Sowada. 2022. «The jars from Tomb U-j – a further petrographic and contextual reassessment». *Ägypten und Levante* 32: 231–268. DOI 10.1553/AEundL32s231. Rao, Rajesh P. N., Nisha Yadav, Mayank N. Vahia, Hrishikesh Joglekar, R. Adhikari, e Iravatham Mahadevan. 2009. «Entropic Evidence for Linguistic Structure in the Indus Script». *Science* 324: 1165. DOI 10.1126/science.1170391. Rodríguez Martínez, Ma. del Carmen, Ponciano Ortiz Ceballos, Michael D. Coe, Richard A. Diehl, Stephen D. Houston, Karl A. Taube, y Alfredo Delgado Calderón. 2006. «Oldest Writing in the New World». *Science* 313: 1610–1614. DOI 10.1126/science.1131492. Saturno, William A., David Stuart, y Boris Beltrán. 2006. «Early Maya Writing at San Bartolo, Guatemala». *Science* 311 (5765): 1281–1283. DOI 10.1126/science.1121745. Schmandt-Besserat, Denise. 1992. *Before Writing, Vol. I: From Counting to Cuneiform*. Austin: University of Texas Press. ISBN 978-0-292-70783-2. Sproat, Richard. 2014. «A statistical comparison of written language and nonlinguistic symbol systems». *Language* 90 (2): 457–481. DOI 10.1353/lan.2014.0031. Stuart, David, Heather Hurst, Boris Beltrán, y William Saturno. 2022. «An early Maya calendar record from San Bartolo, Guatemala». *Science Advances* 8 (15): eabl9290. DOI 10.1126/sciadv.abl9290. Tuchscherer, Konrad, y P. E. H. Hair. 2002. «Cherokee and West Africa: Examining the Origins of the Vai Script». *History in Africa* 29: 427–486. DOI 10.2307/3172173. Urton, Gary. 2003. *Signs of the Inka Khipu: Binary Coding in the Andean Knotted-String Records*. Austin: University of Texas Press. ISBN 978-0-292-78540-3. Van De Mieroop, Marc. 1994. Reseña de *Before Writing*, vol. I, de Denise Schmandt-Besserat. *The Classical World* 87 (6): 501. Weeks, Rex, y Kenneth Tankersley. 2011. «Talking leaves and rocks that teach: the archaeological discovery of Sequoyah's oldest written record». *Antiquity* 85 (329): 978–993. DOI 10.1017/S0003598X00068435.

KNO-005 · Oral Language and Transmission

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

Article Abstract

Ask how people remembered before writing and three assumptions arrive with the question: that remembering was the difficulty, that writing was the solution, that oral transmission is one thing. KNO-005 asks a narrower question instead: given a body of knowledge, a community and no graphic support, what determines how much survives a generation, and at what cost? It distinguishes six modes whose fidelity differs sharply, from casual report to institutionally checked recitation, treated as categories rather than stages. It examines the two poles: the singer who composes instead of memorising, and the community that recites in permuted orders so that corruption becomes audible. It weighs what long chains demonstrably carry against the conditions under which a deep time claim is testable, and revisits the great divide thesis. It closes on the asymmetry: a tradition never recorded leaves nothing to excavate.

In 60 Seconds

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 This Question 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.

1. A badly posed question

How did people remember things before writing is a question that has an answer, and the answer is misleading, because the question smuggles in three assumptions. It assumes that remembering was the difficulty. It assumes that writing was the solution to that difficulty. And it assumes that oral transmission is a single thing about which one general statement can be made.

None of the three survives inspection. Oral societies did not principally struggle to retain knowledge; many of them retained specific bodies of knowledge with high fidelity, by the mechanisms section 3 describes. The earliest uses of writing that survive are largely administrative rather than mnemonic, which is a statement about what survives and not a single motive for its adoption. And the practices lumped together as oral tradition differ from one another more than any of them differs from writing.

The better question is narrower and answerable. Given a body of knowledge, a community and no graphic support, what mechanisms determine how much of it survives a generation, and what do those mechanisms cost.

Where That Leaves Us

The familiar question assumes that remembering was the difficulty, that writing solved it, and that oral transmission is a single thing. None of the three holds. What the earliest surviving writing shows is a fact about what survives rather than a motive for adoption. A narrower question replaces it, and answering that one starts by refusing to treat oral transmission as one practice.

2. Six modes, not one

As in KNO-004, the divisions below classify what a practice does. They are not stages and they do not form a sequence.

Casual report. One person tells another what happened. Fidelity tends to be low and to degrade quickly, because nothing in the practice is arranged to resist that; where a casual account happens to be repeated under constraint it behaves differently. This is the mode that transmission chain experiments in psychology usually study, and it is the mode least like the traditions that carry knowledge over centuries.

Narrative tradition. Stories transmitted across generations with a stable plot skeleton and unstable surface. What persists is structure, character function and sequence. Wording does not persist and is not expected to. Comparative phylogenetic work on folktale corpora treats exactly this signal, using the stable skeleton to reconstruct descent, and finds recoverable lineage across very long chains (Graça da Silva and Tehrani 2016).

Formulaic composition in performance. The mode described by Parry and Lord for South Slavic epic and applied to Homeric material. The singer does not reproduce a text. He builds one, live, from a repertoire of metrically shaped formulas and recurrent themes, and builds a different one on the next occasion (Lord 2019). The same song, on this account, is a type rather than a text, and asking whether a performance is accurate is asking the wrong question of the practice.

Constrained poetic transmission. Verse held stable by simultaneous constraints: metre, rhyme, alliteration, melody, imagery and formulaic phrasing acting at once. The constraints do not help the singer try harder. They cut down the space of things that could go in a slot until very few candidates fit, which is why partial recall reconstructs rather than fails (Rubin 1995).

Ritual verbatim transmission. Speech whose value depends on exact wording, transmitted with the explicit aim of zero change. Fidelity here is a social requirement enforced by training and by sanction, not a property of the material.

Institutionally checked transmission. Verbatim transmission with an added apparatus for detecting corruption, taught to specialists over years, in which the text is recited in deliberately altered orders so that any error becomes audible. This is the strongest oral fidelity regime known and it is discussed in section 4.

The reason for six categories again is that a single word conceals the differences that matter. Casual report and checked recitation are both oral, and they are not comparable technologies.

Where That Leaves Us

Six modes are separated by what a practice does, not by where it stands in a sequence: casual report, narrative tradition, composition in performance, constrained verse, ritual verbatim speech and institutionally checked recitation. The divisions are classifications and not stages, and the two extremes are both oral without being comparable technologies. What they do not yet explain is why the strong cases hold together at all.

3. Fidelity is built from constraint, not effort

The central finding of the cognitive work on oral traditions is that stability comes from the simultaneous action of multiple constraints, each individually weak.

A line of verse in a strong tradition is constrained by its metre, by its position in a rhyme or alliteration scheme, by the melodic line it is sung to, by the imagery of the scene it belongs to, by the formulaic phrasing available for that metrical slot, and by the narrative sequence. Any one of these leaves many options. All of them together often leave one. A performer who cannot recall a line does not draw a blank; the constraints regenerate a line that fits, and the regenerated line is frequently the original or close to it (Rubin 1995).

Three consequences follow, and they explain most of the apparent paradoxes in the field.

First, **fidelity is domain specific**. The same community can transmit a genealogy with near perfect accuracy and a piece of practical instruction badly, because one is embedded in a constraint structure and the other is not.

Second, **unconstrained prose is the fragile case**. It has almost none of the machinery above, which is why traditions that must be preserved tend to be versified, sung, or bound to fixed formulas even when their content is not poetic. Prose that is held by other constraints, such as a fixed liturgical order, is not fragile in the same way.

Third, **fidelity and creativity are not opposites but different settings**. Composition in performance sacrifices verbatim stability and buys unlimited productivity. Checked recitation sacrifices productivity and buys near verbatim stability. Both are oral, and a society may run both at once for different bodies of content.

Where That Leaves Us

Stability comes from many individually weak constraints acting at once, which together often leave one option where each alone leaves many, so a performer who cannot recall a line regenerates one that fits. The regenerated line is frequently the original, which is not the same as always. Fidelity is therefore domain specific, unconstrained prose is the fragile case, and the clearest cases of each setting are the two poles.

4. The two poles

4.1 The singer who does not memorise

Parry and Lord's fieldwork established that the South Slavic epic singer, asked to repeat a song, produces a performance of similar length and identical plot with substantially different wording, and does not experience this as a failure. He is not reciting; he is composing within a system (Lord 2019).

The result reframes a great deal. It means that the question of whether an oral tradition preserves an original text may be empty for entire genres, because there was no original text. It also means that stability at the level of plot, character and theme is compatible with instability at the level of words, so a tradition can be highly conservative about what happens and indifferent about how it is said.

4.2 The community that builds a checksum

At the opposite pole are transmission regimes designed to make error detectable rather than merely improbable. The Vedic recitation tradition of South Asia is the standard example. Alongside continuous recitation, specialists learn to recite the same text in systematically permuted orders, word by word in isolation, then in overlapping pairs, then in more elaborate interleaved patterns. Because a corruption changes what the permuted forms produce, disagreement between the recitation forms exposes it.

The structural point is worth stating in general terms. This is redundancy used for error detection: the same content encoded several ways, so that a fault in one encoding is visible against the others. It is the same principle that underlies checksums in data transmission, arrived at without any theory of information and maintained by institutional training over a very long period.

The caution attached to this example belongs in the text and not only in a footnote. This article describes the technique because it is well attested in the specialist literature on South Asian recitation; the specific sources listed in the accompanying bibliography have been verified as publications but their treatment of these particular recitation forms has not been confirmed directly, and that is recorded in the uncertainty list.

Where That Leaves Us

At one pole the singer composes within a system, producing the same plot in different words without treating that as failure. At the other, recitation in permuted orders turns a corruption into disagreement between forms, which is redundancy used for error detection. The sources listed here are verified as publications, not as treating these particular forms. What such chains carry across very long spans is a separate question.

5. Depth: what long chains can carry

Two lines of evidence bear on how much survives very long oral transmission, and they support different degrees of confidence.

Reconstructible lineage. Phylogenetic comparison of folktale corpora treats tales as heritable traits and reconstructs descent from their distribution across languages, recovering signal that the authors interpret as reaching Proto Indo European depth (Graça da Silva and Tehrani 2016). **The result is model dependent:** it follows from the phylogenetic method and the tale coding those authors adopt, and other codings and models are argued in that literature. What it supports is that some structural features survive extremely long chains. It does not establish that any particular content in any particular tale is old.

Correspondence with independently dated events. A stronger and more contested claim is that specific oral traditions encode information about events datable by other means. Aboriginal Australian traditions describing the inundation of coastal land have been compared with the post glacial sea level curve and argued to preserve knowledge of coastlines drowned more than seven thousand years ago (Nunn and Reid 2016). **That is a published hypothesis and it is reported here as one,** not as an established correspondence.

The methodological problem with claims of this type is general and was set out before this case: deep time oral claims are structurally difficult both to disprove and to believe, because the evidence that would confirm them and the evidence that would refute them are often the same evidence read differently (Henige 2009). No published rebuttal specifically addressing the Australian coastal argument was located during the preparation of this article, which is a statement about this search and not an endorsement.

Recent work extends the approach to volcanism and to geological features, again reading oral traditions against independently dated events (Franks, Nunn and McCallum 2024; Franks et al. 2025). The field is small and its authorship overlaps, which is itself relevant when weighing how independent the confirmations are.

The defensible position is narrow. Long oral chains demonstrably carry structural signal. Claims that a particular tradition encodes a particular dated event are testable only where an independent date exists, and their strength depends entirely on how tightly the tradition constrains the event, which is usually not very tightly.

Where That Leaves Us

Long chains demonstrably carry structural signal, and that is the whole of what the phylogenetic result supports: it is model dependent and establishes nothing about the age of any particular content. A claim that a tradition encodes a dated event is testable only where an independent date exists, and this literature is small with overlapping authorship. A stronger claim about writing itself follows.

6. The great divide, and the criticism it drew

A prominent mid twentieth century argument held that alphabetic literacy restructures thought, enabling logic, history, scepticism and cumulative critique that oral cultures cannot sustain (Goody and Watt 1963). A related account described orality and literacy as distinct modes of consciousness, with writing restructuring the mind that acquires it (Ong 2013).

The thesis drew sustained criticism from comparative work. The principal critique replaced the autonomous model, in which literacy is a uniform technology with uniform cognitive effects, with an ideological model, in which what literacy does depends on the social practices it is embedded in, so that different literacies in different institutions produce different effects and no general consequence follows from the technology alone (Street 1984).

This matters for the series and not only for the topic. The great divide has the same shape as the error KNO-004 identified in the study of writing systems, in a different domain: a real distinction is converted into a ladder, one side is treated as the destination, and everything on the other side is redescribed as a deficit. The correction is the same in both cases. Ask what a system does and for whom, and stop asking what it was on the way to becoming.

Where That Leaves Us

The argument that alphabetic literacy restructures thought was met by comparative work replacing the autonomous model with one in which what literacy does depends on the practices it is embedded in, so that no general consequence follows from the technology alone. That refutes a general claim rather than showing literacy changes nothing. Refusing the deficit framing does not dispose of the real costs.

7. Cost, and what the mode cannot do

Refusing the deficit framing does not mean pretending there are no limits.

Cost in people. Strong oral fidelity is expensive. It requires specialists, years of training, institutional continuity and social arrangements that reward exact transmission. When those arrangements fail, the capability disappears within a generation. The costs of the two are of different kinds and this article does not compare them per unit of content preserved, which would require a measure neither tradition supplies.

Retrieval is usually sequential. Content is retrieved by working through the structure that holds it, which is efficient for whole units and poor for arbitrary lookup. Traditions that build indexing devices into the structure, such as fixed enumerations or place sequences, reduce that cost without escaping it.

Poor for the arbitrary. The constraint machinery of section 3 works because content has structure. Long lists of unmotivated particulars, exact quantities and arbitrary strings are what it stabilises worst, and the traditions that do carry such material tend to have given it an artificial structure first. It is not a coincidence that the earliest writing recorded quantities and names.

Loss is usually irreversible. This is the asymmetry that distinguishes oral from graphic transmission most sharply. A lost manuscript may be recovered, copied elsewhere, or excavated. A tradition that was never recorded leaves nothing when its last speakers stop; where recordings or transcriptions were made, what survives is a record of the tradition rather than the tradition. Language endangerment is therefore not only a linguistic concern: network analysis of endangered languages identifies geographic concentrations where loss is structurally driven by contact patterns rather than distributed evenly (Lee, Siew and Ng 2022), and each loss takes with it whatever was held only in that language and only in memory.

Where That Leaves Us

Strong oral fidelity is expensive in specialists, training and institutional continuity, and the capability disappears within a generation when those fail. Retrieval runs sequentially, arbitrary particulars are what it stabilises worst, and loss is usually irreversible. The article does not compare the two modes per unit of content preserved, because neither supplies that measure. Some of this the evidence cannot settle at all.

8. What the evidence cannot settle

Fidelity before observation. Measured fidelity in oral traditions comes from traditions observed by researchers, which are traditions that survived to be observed. Nothing corrects for the practices that failed, so estimates of what oral transmission can achieve are biased upward by survivorship.

Age of any particular content. Structural methods date lineages, not utterances. A tale type may be ancient while every word in a recorded version is recent, and no method currently distinguishes an old element from a well integrated new one within a living tradition.

The verbatim question for undocumented traditions. Whether a tradition transmitted verbatim can only be established where multiple independent versions were recorded before the tradition was disturbed. For most of the historical record this condition is not met.

The categories. As in KNO-004, the six modes in section 2 are analytic instruments. Real practices mix them, and a single community can move content between modes deliberately.

Where That Leaves Us

Four limits stand. Measured fidelity comes only from traditions that survived to be observed, so the estimates are biased upward. Structural methods date lineages and not utterances. Verbatim transmission is demonstrable only where independent versions were recorded before the tradition was disturbed. The six modes are instruments that real practices mix. What survives those limits is what the article will conclude.

9. Conclusion

Oral transmission is not what preceded writing. It is a family of techniques for holding knowledge in a population, ranging from practices with almost no error control to practices with error control that writing did not improve upon for a very long time. Its strong forms achieve their fidelity by constraining what can be said, which is why they favour verse, formula and song, and why they fail on arbitrary detail.

Treating it as a stage produces two errors at once. It flatters writing, by crediting it with solving a problem that the strong oral regimes had already solved for their own content, and it obscures what writing actually changed, which was cost, portability, random access and the ability to record the arbitrary. Those are large changes. They are not the same as making thought possible.

Where That Leaves Us

Oral transmission is a family of techniques rather than a stage before writing, and its strong forms buy fidelity by constraining what can be said. Treating it as a stage flatters writing and obscures what writing did change: cost, portability, random access and the arbitrary. Those are large changes and not the same as making thought possible. What is established, and how, follows.

What We Know, and How

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 We Still Don't Know

Measured fidelity comes only from traditions that survived to be observed, and nothing corrects for the practices that failed, so estimates of what oral transmission can achieve are biased upward. Structural methods date lineages rather than utterances, so a tale type may be ancient while every recorded word is recent. Whether a given tradition transmitted verbatim can be settled only where independent versions were recorded before it was disturbed.

What Could Change Our Minds

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.

If You Remember Only One Thing

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.

Related Articles

- KNO-003 · *From Marks to Symbols*
- KNO-004 · *From Symbols to Writing*

References

Franks, Leigh, Patrick D. Nunn, y Adrian McCallum. 2024. «The selective geography of volcanism in oral traditions». *Quaternary Research* 122: 1–17. DOI 10.1017/qua.2024.29. Franks, Leigh, Patrick D. Nunn, Adrian McCallum, y Gertz, Janine. 2025. «Stories in Stone». *Geoheritage* 17 (2): art. 73. DOI

10.1007/s12371-025-01116-2. Goody, Jack, y Ian Watt. 1963. «The Consequences of Literacy». *Comparative Studies in Society and History* 5 (3): 304–345. DOI 10.1017/S0010417500001730. Graça da Silva, Sara, y Jamshid J. Tehrani. 2016. «Comparative phylogenetic analyses uncover the ancient roots of Indo-European folktales». *Royal Society Open Science* 3 (1): art. 150645. DOI 10.1098/rsos.150645. Henige, David. 2009. «Impossible to Disprove Yet Impossible to Believe: The Unforgiving Epistemology of Deep-Time Oral Tradition». *History in Africa* 36: 127–234. DOI 10.1353/hia.2010.0014. Lee, Nala H., Cynthia S. Q. Siew, y Nadine H. N. Ng. 2022. «The network nature of language endangerment hotspots». *Scientific Reports* 12: art. 10803. DOI 10.1038/s41598-022-14479-1. Lord, Albert B. 2019. *The Singer of Tales*. Ed. David F. Elmer. 3.^a ed. Cambridge, MA: Harvard University Press. ISBN 978-0-674-97573-6. Primera edición 1960. Nunn, Patrick D., y Nicholas J. Reid. 2016. «Aboriginal Memories of Inundation of the Australian Coast Dating from More than 7000 Years Ago». *Australian Geographer* 47 (1). DOI 10.1080/00049182.2015.1077539. Ong, Walter J. 2013. *Orality and Literacy: The Technologizing of the Word*. 30.^o aniversario, con dos ensayos de John Hartley. London: Routledge. ISBN 978-0-415-53838-1. Rubin, David C. 1995. *Memory in Oral Traditions: The Cognitive Psychology of Epic, Ballads, and Counting-out Rhymes*. New York: Oxford University Press. ISBN 978-0-19-508211-1. Rústica 1997, ISBN 978-0-19-512032-5. Staal, Frits. 1986. *The Fidelity of Oral Tradition and the Origins of Science*. Amsterdam: North-Holland / Koninklijke Nederlandse Akademie van Wetenschappen. ISBN 0-444-85669-2. Street, Brian V. 1984. *Literacy in Theory and Practice*. Cambridge Studies in Oral and Literate Culture 9. Cambridge: Cambridge University Press. ISBN 0-521-24785-3.

KNO-006 · Maps and Spatial Representation

KNO · Ways of Knowing

Juan Segurola

v0.7 · EN

All Rights Reserved

Ways of Knowing

Epistemic profile: Current scientific interpretation **Licence at publication:** All Rights Reserved

Bilingual title: Maps and Spatial Representation / Mapas y representación espacial

Article Abstract

What is a map, and how much of human spatial knowledge does the word actually cover? This article separates three things that are often collapsed into one: the capacity to orient and navigate, the external artefact that stands for spatial relations, and the practices called cartography. It asks what neural findings about place and grid cells do and do not establish about the format of spatial thought. It examines wave charts and moving reference frames, the disputed prehistoric candidates, and the geometric constraint that forces every flat map of a sphere to distort something. It then asks what follows for reading maps as arguments once distortion and selection are unavoidable, why the dominance of one projection online does not dissolve the constraint, and which of the questions raised the evidence assembled here cannot settle.

In 60 Seconds

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 This Question 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.

1. Three things that are not the same

Spatial knowledge is the capacity to orient, locate or navigate. An external spatial representation is an artefact that stands for spatial relations. Cartography is the study and practice of making and using maps under explicit or implicit conventions, with histories in many societies. These three are related but not identical.

An expert navigator crossing open ocean without instruments has spatial knowledge of a high order and may use no graphic artefact during the voyage. A Marshallese training chart can foreground swell behaviour rather than act as a scaled plan. A modern road atlas is an external representation whose user may know little of the terrain it covers. Nothing forces the three capacities to coincide.

The article uses *map* broadly for an external artefact that represents spatial relations. *Projected map* refers more narrowly to a representation that transforms locations from a curved surface to a plane. Neither term is reserved for a European tradition.

Where That Leaves Us

The three can come apart, as the examples show, so nothing licenses treating them as one thing under three names. The article fixes its vocabulary rather than discovering it: map covers any external artefact standing for spatial relations. Whether human spatial thought resembles a graphic map is a separate empirical question, and the next section asks what neuroscience actually supports.

2. What the brain does with space

Two findings anchor the biological side and neither of them says what the popular phrase cognitive map is usually taken to say.

Hippocampal place cells fire when an animal occupies a particular location. O'Keefe and Nadel proposed that the hippocampus supports an allocentric representation organised relative to the environment (O'Keefe and Nadel 1978). Here, allocentric summarises the work's nonegocentric framework; it is not presented as a literal quotation from pp. 89–90. In rats, certain cells in dorsocaudal medial entorhinal cortex were later found to activate at the vertices of a periodic triangular grid across the experimental environment (Hafting et al. 2005). Their role in a more general spatial system is an interpretation supported by, but not identical to, that firing pattern.

What these findings do not establish is that the brain builds a map in the everyday graphic sense. A classic developmental account distinguishes landmark knowledge, route knowledge organised around sequences, and survey knowledge that integrates spatial relations beyond one route (Siegel and White 1975). The proposed sequence has been influential, but it should not be converted into a universal developmental law. Route knowledge can support successful travel without being a degraded survey map, while its flexibility varies with learning and task.

The relevance to the rest of the article is direct. If human spatial competence rests substantially on sequenced, viewpoint dependent knowledge, then the survey style representation that Europeans came to call a map is not the natural externalisation of how people think about space. It is one option among several, and the others are well attested.

Where That Leaves Us

Place and grid cells are robust findings about neural activity, and they do not show that spatial thought takes graphic form. The landmark, route and survey distinction stays influential without being a universal developmental law. If sequenced, viewpoint dependent knowledge carries much of human competence, the survey style map is one externalisation among several rather than the natural one.

3. Representations that are not pictures of terrain

3.1 Wave charts

Marshallese stick charts are training and teaching artefacts associated with wave navigation. Their elements can encode relations among islands and swell patterns, including the ways islands disturb waves; different chart types do not all have the same function. They are therefore not equivalent to scaled geographic plans, but it is too broad to say that they never represent island arrangement (Genz et al. 2009; Genz 2016). At sea, navigators detect relevant patterns partly through the motion of the canoe.

The important point is not that the charts lack spatial content, but that their spatial content is inseparable from a model of ocean behaviour and navigational practice. Reading them as ordinary topographic maps loses the function for which they were taught.

3.2 Reference frames that move

Caroline Islands navigation includes a technique known as *etak*, in which progress can be conceptualised by reference islands passing through successive star bearings. The system is described in the ethnographic literature on Puluwat and in comparative work on Pacific navigation (Gladwin 1995; Lewis 1994).

The analytical point is that spatial reasoning can organise motion around different reference frames. The ethnographic description establishes the practice; it does not by itself establish that one frame is computationally optimal or that an entire European tradition uses only its opposite.

3.3 The general point

The three books of volume 2 of *The History of Cartography* examine Islamic and South Asian, East and Southeast Asian, African, American, Arctic, Australian and Pacific traditions in their own historical settings (Harley and Woodward 1992; 1994; Woodward and Lewis 1998). Their organisation supports comparison without treating one tradition as the universal template.

Where That Leaves Us

An artefact can carry spatial content while refusing the form of a scaled plan, and reading it as a topographic map discards the function it was made to serve. A wave chart is inseparable from a model of ocean behaviour, and *etak* can organise motion around a reference frame that moves. Older candidate artefacts raise a harder problem of identification.

4. The earliest candidate maps are disputed

The wall painting at Çatalhöyük is usually introduced as the oldest known map, showing a village plan beneath an erupting volcano. The reading has been challenged in detail, with the argument that the composition is better explained as a leopard skin and a geometric design than as a plan and a mountain (Meece 2006). A later study accepted a representational reading of the upper element, arguing on geochronological grounds that it may depict an eruption, without settling the question of whether the lower element is a plan (Schmitt et al. 2014).

The methodological risk is circularity: resemblance to a modern category motivates an identification, and the identification is then cited as proof that the category is ancient. Independent contextual evidence can strengthen the inference, but resemblance alone cannot settle function.

This is not scepticism about ancient spatial representation, for which there is no shortage of unambiguous later evidence. It is scepticism about the specific move that turns a suggestive image into the first map.

Where That Leaves Us

Identifying a prehistoric image as a map from resemblance alone invites circularity: the modern category motivates the reading, and the reading is then offered as proof that the category is ancient. Independent contextual evidence can strengthen such an inference; resemblance alone cannot fix function. The doubt is narrow and does not reach ancient spatial representation generally. What follows does not depend on history.

5. Why no flat map of a sphere can be faithful

Here the argument stops being historical.

A sphere and a plane have different Gaussian curvature, and Gaussian curvature is preserved by local isometries, the Theorema Egregium (do Carmo 1976, p. 234). This article deduces from it that no open region of a sphere can be mapped isometrically onto a plane: some metric property must be distorted. The source states the sphere-plane impossibility explicitly as exercise 4 on p. 237; the demonstration developed here is the article's own reasoning.

That result converts what looks like an engineering problem into a constraint. Every flat map of a surface with nonzero Gaussian curvature, the sphere included, distorts. A projection can preserve angles, or areas, or distances along particular lines, or a compromise among them, and it cannot preserve everything. The cartographer must therefore choose which distortions to accept, while accuracy remains assessable against the map's intended purpose.

The Mercator projection is the standard illustration. On its spherical model it is conformal, and a line of constant compass bearing is straight; its scale and area inflation increase toward the poles. Monmonier documents both its navigational value and its later use in wall maps and other non-navigational settings, where the trade-off may be poorly matched to purpose (Monmonier 2004).

The general lesson is that a projection encodes a decision about what matters, and that a reader who does not know the decision will misread the map in a direction the map does not disclose.

Where That Leaves Us

Because sphere and plane differ in Gaussian curvature, no open region of one lies isometrically on the other, so every flat map of a sphere distorts something. That consequence is this article's deduction from the Theorema Egregium, not a result quoted from the cited book. What looks like an engineering limit is a constraint: projections distribute unavoidable distortion differently, and accuracy must be assessed against purpose.

6. Maps as arguments

Once distortion is unavoidable and selection is necessary, the claim that a map merely shows what is there becomes untenable. The critical cartography literature makes this its central point: maps are rhetorical artefacts whose apparent neutrality is itself an effect they produce (Harley 1989), and it has been argued that cartographic form can naturalise the interests a map serves rather than stating them (Wood 1992; Wood, Fels and Krygier 2010).

Four recurring editorial decisions help expose this selectivity.

Selection. What is on the map. Every map omits almost everything, and the omissions are not random.

Scale. What is too small to appear. Scale is a claim about what counts as a feature.

Orientation. North at the top is a convention with a history, not a property of the Earth. Maps can adopt other orientations according to purpose and tradition.

Classification. What counts as the same kind of thing, which imports a taxonomy that the map then displays as if it were terrain.

This does not make maps arbitrary or unreliable, and the critical literature is often misread as saying so. A map can be checked, corrected and used, and the geometry of section 5 is not a matter of opinion. What follows is narrower: the neutrality a map projects is a property of the genre, and evaluating a map requires knowing its purpose.

Where That Leaves Us

If distortion is unavoidable and selection necessary, a map cannot merely show what is there. Selection, scale, orientation and classification are recurring decisions through which that selectivity works, and the neutrality a map projects is an effect of the genre rather than a property of terrain. This does not make maps arbitrary, and the geometry stands apart from the critique.

7. The digital case

Online interactive mapping has made Web Mercator the dominant display projection. Battersby and colleagues document that dominance and analyse its technical advantages, cartographic limitations and cognitive implications; they do not establish one exclusive cause for its adoption (Battersby et al. 2014). Web Mercator uses spherical Mercator equations, is well suited to tiled display, and retains the severe high-latitude area inflation associated with Mercator-type world views.

The case shows that the geometric constraint in section 5 does not disappear when the medium becomes interactive. Technical architecture, interoperability, inherited standards and design choices can all shape which projection users encounter. Assigning their relative causal weight requires historical evidence beyond the projection's mathematical properties.

Where That Leaves Us

Web Mercator's dominance online shows the geometric constraint surviving the move to interactive media, carrying the same severe area inflation at high latitudes. Technical architecture, interoperability, inherited standards and design can each shape which projection a user meets, and the cited analysis isolates no exclusive cause. Assigning causal weight would need historical evidence the mathematics cannot supply.

8. What the evidence cannot settle

Whether a given artefact is a map. Without independent evidence of use, identifying a prehistoric image as a map is an inference from resemblance, and resemblance to a modern category is weak evidence about an ancient one.

What cognitive maps are. Place cells and grid cells are robust findings about neural activity. The step from that to a claim about the format of human spatial thought is an interpretation, and the route versus survey distinction suggests the format varies by task and by person.

Whether non cartographic systems are maps. Whether a wave chart or a sung route is a map depends on the definition adopted, and this article does not pretend that its own stipulation settles a question that is partly terminological.

Intent behind selection. That maps can encode interests is argued in the critical literature. That a particular map encodes a particular intent requires evidence about its makers, and the critical literature is stronger on the general claim than on individual attributions.

Where That Leaves Us

Four questions resist the evidence assembled here: whether a given artefact was used as a map, what format spatial thought takes, whether non cartographic systems fall under the term, and what intent a particular map encodes. One is partly terminological rather than empirical, and this article's stipulated vocabulary does not settle it. The closing synthesis states what does follow.

9. Conclusion

Space can be known without a graphic artefact, represented without a scaled picture of terrain, and projected from a sphere onto a plane only by accepting geometric trade-offs. A cartographic map is one powerful family of solutions, not the universal form of spatial knowledge.

The recurring error is to treat one solution as the general case and everything else as an approximation to it. A wave model, a star sequence, a route description and a projected grid answer different questions about space. A conformal world map accepts area distortion in order to preserve local angles; that trade-off is precise enough without calling the map a lie.

Where That Leaves Us

Knowing space needs no graphic artefact and representing it needs no scaled picture, but a sphere maps onto a plane only with geometric trade offs. A cartographic map is therefore one powerful family of solutions, not the general form of spatial knowledge. The recurring error takes that family as the standard and reads everything else as an approximation to it, when the alternatives answer different questions.

Boundary note

With KNO-005, Oral Language and Transmission. Spatial knowledge held and transmitted orally, including route instruction and sung routes, is treated in KNO-005 as an instance of constrained transmission. This article takes up the representational question that KNO-005 explicitly deferred: what an external artefact does when it stands for spatial relations. Transmission mechanisms, fidelity and memory are not re examined here.

With KNO-007, Calendars and Time Measurement. The representation of time, the division of cycles and the instruments of timekeeping belong to KNO-007. Where this article touches time, in navigation by star sequence and in dead reckoning, it does so only as an element of spatial method. The general question of how a periodic process becomes a measure is KNO-007's.

With KNO-004 and KNO-013. Notation and the graphic encoding of language belong to KNO-004. The wider question of what a model is and how models relate to reality is KNO-013's, and this article does not anticipate it: it treats projection as a geometric constraint, not as a case study in modelling.

What We Know, and How

- 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 We Still Don't Know

Whether a given prehistoric artefact was made or used as a map remains open, because resemblance to a modern category is weak evidence about an ancient one. What format human spatial thought takes is not settled by the neural findings, and the route versus survey distinction suggests it varies by task and person. Whether a wave chart or a sung route counts as a map is partly terminological. And attributing a particular intent to a particular map needs evidence about its makers that the general argument does not supply.

What Could Change Our Minds

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.

If You Remember Only One Thing

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.

Related Articles

- KNO-004 · From Symbols to Writing
- KNO-005 · Oral Language and Transmission
- KNO-007 · Calendars and Time Measurement
- KNO-013 · Models, Maps and Reality

References

Battersby, Sarah E., Michael P. Finn, E. Lynn Usery, and Kristina H. Yamamoto. 2014. "Implications of Web Mercator and Its Use in Online Mapping." *Cartographica* 49 (2): 85–101. DOI 10.3138/carto.49.2.2313.

do Carmo, Manfredo P. 1976. *Differential Geometry of Curves and Surfaces*. Englewood Cliffs: Prentice-Hall. ISBN 978-0-13-212589-5.

Genz, Joseph H. 2016. "Resolving Ambivalence in Marshallese Navigation: Relearning, Reinterpreting, and Reviving the 'Stick Chart' Wave Models." *Structure and Dynamics* 9 (1). DOI 10.5070/SD991031888.

Genz, Joseph, Jerome Aucan, Mark Merrifield, Ben Finney, Korent Joel, and Alson Kelen. 2009. "Wave Navigation in the Marshall Islands: Comparing Indigenous and Western Scientific Knowledge of the Ocean." *Oceanography* 22 (2): 234–245. DOI 10.5670/oceanog.2009.52.

Gladwin, Thomas. 1995. *East Is a Big Bird: Navigation and Logic on Puluwat Atoll*. Cambridge, MA: Harvard University Press. ISBN 978-0-674-22426-1.

Hafting, Torkel, Marianne Fyhn, Sturla Molden, May-Britt Moser, and Edvard I. Moser. 2005. "Microstructure of a Spatial Map in the Entorhinal Cortex." *Nature* 436 (7052): 801–806. DOI 10.1038/nature03721.

Harley, J. B. 1989. "Deconstructing the Map." *Cartographica* 26 (2): 1–20. DOI 10.3138/E635-7827-1757-9T53.

Harley, J. B., and David Woodward, eds. 1992. *The History of Cartography, Volume 2, Book 1: Cartography in the Traditional Islamic and South Asian Societies*. Chicago: University of Chicago Press. ISBN 0-226-31635-1.

Harley, J. B., and David Woodward, eds. 1994. *The History of Cartography, Volume 2, Book 2: Cartography in the Traditional East and Southeast Asian Societies*. Chicago: University of Chicago Press. ISBN 0-226-31637-8.

Lewis, David. 1994. *We, the Navigators: The Ancient Art of Landfinding in the Pacific*. 2nd ed. Edited by Derek Oulton. Honolulu: University of Hawai'i Press. ISBN 978-0-8248-1582-0.

Meece, Stephanie. 2006. "A Bird's Eye View – of a Leopard's Spots: The Çatalhöyük 'Map' and the Development of Cartographic Representation in Prehistory." *Anatolian Studies* 56: 1–16. DOI 10.1017/S0066154600000727.

Monmonier, Mark. 2004. *Rhumb Lines and Map Wars: A Social History of the Mercator Projection*. Chicago: University of Chicago Press. ISBN 978-0-226-53431-2.

O'Keefe, John, and Lynn Nadel. 1978. *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press. ISBN 978-0-19-857206-0.

Schmitt, Axel K., Martin Danišik, Erkan Aydar, Erdal Şen, İnan Ulusoy, and Oscar M. Lovera. 2014. "Identifying the Volcanic Eruption Depicted in a Neolithic Painting at Çatalhöyük, Central Anatolia, Turkey." *PLOS ONE* 9 (1): e84711. DOI 10.1371/journal.pone.0084711.

Siegel, Alexander W., and S. H. White. 1975. "The Development of Spatial Representations of Large-Scale Environments." *Advances in Child Development and Behavior* 10: 9–55. DOI 10.1016/S0065-2407(08)60007-5.

Wood, Denis. 1992. *The Power of Maps*. New York: Guilford Press. ISBN 978-0-89862-493-9.

Wood, Denis, with John Fels and John Krygier. 2010. *Rethinking the Power of Maps*. New York: Guilford Press. ISBN 978-1-59385-366-2.

Woodward, David, and G. Malcolm Lewis, eds. 1998. *The History of Cartography, Volume 2, Book 3: Cartography in the Traditional African, American, Arctic, Australian, and Pacific Societies*. Chicago: University of Chicago Press. ISBN 0-226-90728-7.

KNO-007 · Calendars and Time Measurement

KNO · Ways of Knowing

Juan Segurola

v0.2 · EN

All Rights Reserved

Article Abstract

Most time reckoning leans on three observable cycles that do not divide into one another. For epoch 2000 the mean tropical year is 365.24219 days and the mean synodic month 29.53059 days, both approximate means that change slowly, so a calendar seeking lunar or seasonal alignment must use rules, observation, intercalation or accepted drift. KNO-007 asks what follows from that. It compares lunar, solar and lunisolar strategies without ranking them as stages of refinement, performs its own arithmetic and marks it as its own, and shows that a calendrical cycle need not track an astronomical period at all. It then separates the SI second, defined by a caesium transition since 1967, from civil time scales still held by convention near a turning Earth. It closes on disputed prehistoric candidates, on what a calendar coordinates, and on the limits of the evidence.

In 60 Seconds

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 This Question 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.

1. Three periods that do not fit

Many systems of time reckoning draw on three observable cycles: day and night, lunar phases and seasonal recurrence.

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, where each day is 86,400 SI seconds (Richards 2013, in Urban and Seidelmann, eds.). The same source stresses that these are approximate mean periods and change slowly with time.

From those two figures three consequences follow by arithmetic alone. **The article performs this arithmetic itself; the products below are deductions from the cited values, not quotations.**

- Twelve synodic months come to 354.36708 days, which is **10.87511 days short of a tropical year**.
- The mean synodic month is not a whole number of 86,400-second days, so fixed-length calendar months cannot remain exactly aligned with every lunation.
- The mean tropical year is not a whole number of mean synodic months, so a fixed sequence of lunar months cannot remain exactly aligned with the seasons without adjustment.

This is not a defect of a particular calendar. It is a consequence of the approximate and changing relations among astronomical periods. Calendars that seek lunar or seasonal alignment must use rules, observations, intercalations or accepted drift. Calendars need not seek either alignment, as section 4 shows.

There is a useful but limited parallel with KNO-006. A projected map faces an exact geometric theorem; a calendar faces empirically measured periods that vary with time. Both make trade-offs visible, but the mathematical status of the two constraints is not the same.

Where That Leaves Us

Three observable cycles carry time reckoning, and the two measured periods are approximate mean periods that change slowly. From those two figures alone, months of fixed length cannot remain aligned with every lunation, and a fixed sequence of them cannot remain aligned with the seasons without adjustment. That arithmetic is performed here, not quoted, and what a calendar does about an empirical constraint is a choice.

2. Three widespread strategies

2.1 Keep the month, lose the season

A calendar of twelve mean synodic months runs about eleven days short of a tropical year, so its dates move through the seasons. By the figures above, the accumulated displacement is approximately one tropical year after 33.6 lunar years.

The Islamic calendar is a lunar calendar: it keeps months in relation to lunations and does not keep its year in synchrony with the Sun (Richards 2013). Its seasonal drift is therefore a consequence of the chosen structure, not evidence that the calendar has failed to be solar.

2.2 Keep the season, lose the moon

The opposite choice abandons the lunar month as an astronomical unit and keeps only its name. The Egyptian civil year is the clearest early case: twelve months of thirty days plus five epagomenal days, three hundred and sixty five in total, described in the standard study as running without an intercalated day (Parker 1950).

A year of exactly 365 days is about a quarter day short, so it too drifts against the seasons, only far more slowly. The Egyptian system tolerated this and tracked the heliacal rising of Sirius separately; Parker records that rising as falling on 17 to 19 July in the Julian calendar throughout dynastic history.

2.3 Keep both, pay in intercalation

The third family keeps true lunar months and forces them back into agreement with the seasons by inserting an extra month at intervals. The arithmetic that makes this workable is a near coincidence:

- Nineteen tropical years come to 6939.60161 days.
- Two hundred and thirty five synodic months come to 6939.68865 days.
- **They differ by 0.08704 days, about 2.09 hours.**

Nineteen years therefore contain almost exactly 235 lunar months, which is twelve ordinary years of twelve months plus seven years with a thirteenth. The nineteen year cycle was in use in the Babylonian calendar from the early fifth century BCE, and the question of whether Greek knowledge of it came from Babylon or was found independently remains open, with an independent Chinese use of the same cycle cited as a reason not to rule out independent discovery (Steele 2020).

The residual is small but not zero. At about two hours per cycle, it accumulates to about 0.44 days, or 10.4 hours, after five cycles (ninety five years). Actual calendrical practice also has to accommodate the fact that the astronomical means themselves vary.

Where That Leaves Us

Keep the month and the year walks through the seasons, which the Islamic calendar does by design; keep the year and the Egyptian civil calendar drifts a quarter day at a time; keep both and the price is a thirteenth month, because nineteen years and 235 months differ by 0.08704 days. Whether Greek use came from Babylon is open, and the nineteen year fit is close without being exact.

3. The Julian error and what the Gregorian reform actually decided

Compared with the epoch-2000 mean used here, a year of exactly 365.25 days exceeds the tropical year by 0.00781 days. **As an illustrative calculation, that is one day every 128 years**, and 9.82 days across 1257 such years. It is not a reconstruction using a constant historical tropical year.

Richards reports that by the thirteenth century the displacement of the vernal equinox from 21 March was recognised, and that debate and attempted remedies preceded the 1582 reform by centuries (Richards 2013). The reform changed both leap-day rules and the method for calculating Easter.

The Gregorian year of 365.2425 days, obtained by dropping three leap days every four centuries, exceeds the tropical year by 0.00031 days, **which is one day in about 3226 years by the same arithmetic.**

The reform did not make the calendar exact; it reduced the long-term solar residual. It also revised the ecclesiastical lunar tables used to calculate Easter, so it should not be described simply as choosing the season and ignoring the Moon. The epoch-2000 comparison above is a transparent approximation, not a permanent error rate.

Where That Leaves Us

Against the epoch 2000 mean a Julian year exceeds the tropical year by 0.00781 days and a Gregorian one by 0.00031: a day every 128 years against one in about 3226. Both are illustrative: the periods change slowly. The 1582 reform reduced the residual without removing it and revised the ecclesiastical lunar tables, so it did not simply choose the Sun. Not every cycle aims at a period.

4. Calendrical cycles need not be astronomical periods

The assumption that a calendar exists only to track an agricultural or astronomical year is too narrow. Richards notes that Central American thirteen-day and twenty-day cycles have no direct astronomical basis (Richards 2013). Their combination yields a 260-day count, a calendrical period that is neither a synodic month nor a tropical year.

Šprajc, Inomata and Aveni infer from orientations of Formative ceremonial complexes that the 260-day calendar was in use between 1100 and 750 BCE, centuries before its previously known textual record (Šprajc, Inomata and Aveni 2023). That is peer-reviewed evidence for an early use, but it remains an archaeological inference from orientations rather than a surviving written calendar.

The case establishes the narrower point needed here: calendrical organisation can combine astronomical observations with conventional numerical cycles. The origin and full range of functions of the 260-day count remain matters for specialist evidence; this article does not assign it one exclusive cause.

Where That Leaves Us

Thirteen and twenty day cycles have no direct astronomical basis, and their combination gives a 260 day count that is neither a month nor a year. Orientations of Formative complexes are read as evidence of its use between 1100 and 750 BCE, an archaeological inference rather than a surviving written calendar. Calendrical order can be conventional as well as observed, and the unit underneath has its own history.

5. From the turning Earth to the caesium atom

Before 1960 the second was defined as $1/86,400$ of the mean solar day. In 1960 it was redefined through a fraction of the tropical year 1900, and in 1967 the CGPM adopted a caesium-based atomic definition. The 2018 wording, effective with the revised SI in 2019, defines the second by fixing the numerical value of the unperturbed caesium-133 ground-state hyperfine transition frequency at 9,192,631,770 Hz (BIPM, “Historical perspective: unit of time, second”; BIPM 2019).

The SI unit is therefore no longer defined by Earth's rotation. That does not mean all timekeeping is independent of Earth: UTC has the rate of atomic time but has been kept close to rotational time, UT1, through leap seconds.

In 2022 the General Conference on Weights and Measures adopted Resolution 4 of its twenty seventh meeting, on the use and future development of UTC. **The resolution's own operative language is that the maximum value for the difference UT1 minus UTC will be increased in, or before, 2035,** with the problem stated as the risk that leap second discontinuities cause serious malfunctions in critical digital infrastructure, and with a direction to propose a new tolerance ensuring continuity of UTC for at least a century.

The wording matters. The resolution decides that the permitted divergence will be increased; it does not itself specify the new maximum or literally say that leap seconds are abolished.

This is related to, but not identical with, section 1. Since 1967 the SI unit has been atomic, while civil time has retained a conventional relation to Earth's rotation. The 2022 decision changes the permitted tolerance between those time scales; it does not change the astronomical periods discussed in section 1.

Where That Leaves Us

The second was defined by the mean solar day, then by a fraction of the 1900 tropical year, and since 1967 by a caesium transition fixed at 9,192,631,770 Hz. The unit left Earth's rotation; UTC did not, kept close to UT1. In 2022 the CGPM decided that difference will be increased, without saying by how much. How far back deliberate reckoning goes is another question.

6. The earliest candidates are disputed

A pit alignment at Warren Field, Crathes, in Aberdeenshire, was interpreted by its investigators as a possible lunisolar time reckoner of the eighth millennium BCE and proposed as very early evidence of deliberate time reckoning (Gaffney et al. 2013). Their article repeatedly uses qualified language such as “appears” and “may”.

Earlier still, incised marks on Upper Palaeolithic bone and antler were argued to constitute deliberate sequential lunar notation (Marshack 1972), an interpretation challenged in a paper whose title puts the objection as a question about wishful thinking (d'Errico 1989).

The methodological warning from KNO-006 applies, but not without alteration. At Warren Field the case combines pit sequence, lunar interpretation and a midwinter alignment; for incised bones the inference depends more heavily on the organisation of marks. In both cases function is inferred rather than written on the artefact. A sequence of notches may be compatible with lunar counting, other tallying or non-notational patterning, so compatibility alone does not establish use.

A specific limitation of this round should be stated. No published scholarly critique specifically addressing the Warren Field interpretation was located during preparation. That is a statement about this search and not evidence that the interpretation is uncontested.

Where That Leaves Us

Warren Field was interpreted by its investigators as a possible lunisolar time reckoner, in language they kept qualified, and no published critique of that interpretation was located in this round, which is a statement about the search. Incised Palaeolithic marks were argued to be deliberate lunar notation, and the argument was challenged. Function is inferred in both cases, and what it was for is harder still.

7. What a calendar is actually for

Section 1 framed calendars as solutions to an astronomical constraint, which is true and incomplete. A calendar is also a coordination device, which helps explain the need for predictable rules, shared procedures and communicable dates.

Calendars can coordinate hunting, agriculture, ritual, contracts and social life (Richards 2013). Different designs trade predictability against observation: arithmetic calendars can be computed by fixed rules, whereas observational calendars may wait for a crescent, solstice or other event and cannot always be constructed fully in advance. Coordination requires shared procedures and a way to communicate dates, not necessarily that every user personally perform the calculation.

This helps frame why reform is not only an astronomical calculation. Changing a calendar also changes a coordination standard, religious observances and administrative practice. The claim that these social costs explain a particular historical delay is an interpretation and requires case-specific evidence.

Where That Leaves Us

The framing in section 1 is true and incomplete. A calendar coordinates hunting, agriculture, ritual, contracts and social life, so its design has practical coordination consequences as well as astronomical ones. Arithmetic designs can be computed in advance while observational ones wait for a crescent or a solstice, and coordination needs shared procedure rather than universal calculation. Explaining a particular delay by its social cost would need case evidence.

8. What the evidence cannot settle

Whether an early artefact is a calendar. Without independent evidence of use, matching an artefact to a modern calendrical function is inference from resemblance.

Intention behind an observed period. That a structure aligns with a solstice does not establish that alignment was intended, and the number of possible alignments in any structure is large.

Whether early systems were understood as approximations. Later users of a nineteen year cycle knew it was inexact. Whether its earliest users did is not recoverable from the cycle's use alone.

Limits of the sources. Four descriptive claims in this article rely on sources whose relevant text was not available in full. Those claims are therefore stated conservatively rather than extended beyond what the accessible evidence supports.

Where That Leaves Us

The evidence stops in four places. Calling an early artefact a calendar is inference from resemblance, with no independent evidence of use. An alignment that exists was not necessarily intended, since the possible alignments in any structure are many. Whether the first users of an inexact cycle knew it was inexact is not recoverable from the cycle. And four descriptive claims rest on sources not read directly here.

9. Conclusion

Mean days, lunar months and tropical years do not occur in integral ratios, and calendars that seek to track them respond with drift, intercalation, observation or arithmetic rules. Other calendrical cycles need not represent an astronomical period at all.

Reading the history as simple progress towards the Gregorian calendar mistakes different purposes for stages on one ladder. A lunar calendar is not a failed solar one, and a 260-day count is not an inaccurate tropical year. The SI second has been atomic since 1967, while UTC still mediates between atomic rate and terrestrial rotation. In 2022 the CGPM decided that the allowed UT1-UTC difference would be enlarged by or before 2035; the new maximum was not fixed by that resolution.

Where That Leaves Us

The ladder is the mistake: it turns different purposes into stages and a lunar calendar into a failed solar one. Mean days, lunar months and tropical years are not in integral ratios, which leaves calendars a choice, and other cycles represent no astronomical period. Since 1967 the second has been atomic, and UTC goes on mediating between that rate and a rotating Earth. What is established, and how, follows.

What We Know, and How

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 We Still Don't Know

Whether Warren Field worked as a lunisolar reckoner is an interpretation its own investigators hedged, and no published critique addressing it specifically was located in this round, which is a statement about the search and not about the literature. Whether Upper Palaeolithic incised marks record lunations is open, because compatibility with a lunar count does not establish one. Whether the earliest users of a nineteen year cycle understood it as inexact is not recoverable from the use of the cycle. The origin and full range of functions of the 260 day count need specialist evidence this article does not supply. And four descriptive claims rest on sources whose relevant text was not read directly in this round.

What Could Change Our Minds

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.

If You Remember Only One Thing

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.

Related Articles

- KNO-006 · *Maps and Spatial Representation*
- KNO-008 · *Instruments That Extend the Senses*

References

BIPM. 2019. *SI Brochure, 9th Edition, Appendix 2: Mise en pratique for the Definition of the Second in the SI*. Bureau International des Poids et Mesures, 20 de mayo de 2019. URL estable: <https://www.bipm.org/documents/20126/41489667/SI-App2-second.pdf> **C-D**

BIPM. «Historical Perspective: Unit of Time, Second». Bureau International des Poids et Mesures. URL estable: <https://www.bipm.org/en/history-si/second>

CGPM. 2022. Resolución 4 de la 27.^a CGPM, «On the Use and Future Development of UTC». Conférence Générale des Poids et Mesures, 15 a 18 de noviembre de 2022. DOI 10.59161/CGPM2022RES4E. URL estable: <https://www.bipm.org/en/-/resolution-cgpm-27-4> **C-D**

d'Errico. 1989. «Palaeolithic Lunar Calendars: A Case of Wishful Thinking?». *Current Anthropology* 30 (1). DOI 10.1086/203721.

Gaffney, Vincent, Simon Fitch, Eleanor Ramsey, Ron Yorston, Eugene Ch'ng, Eamonn Baldwin, Richard Bates, Christopher Gaffney, Clive Ruggles, Tom Sparrow, Anneley McMillan, Dave Cowley, Shannon Fraser, Charles Murray, Hilary Murray, Emma Hopla, y Andy Howard. 2013. «Time and a Place: A Luni-Solar "Time-Reckoner" from 8th Millennium BC Scotland». *Internet Archaeology* 34. DOI 10.11141/ia.34.1. **C-D**

Marshack, Alexander. 1972. *The Roots of Civilization: The Cognitive Beginnings of Man's First Art, Symbol and Notation*. New York: McGraw-Hill. ISBN 978-0-07-040535-6.

Parker, Richard A. 1950. *The Calendars of Ancient Egypt*. Studies in Ancient Oriental Civilization 26. Chicago: Oriental Institute of the University of Chicago. URL estable: <https://isac.uchicago.edu/sites/default/files/uploads/shared/docs/saoc26.pdf>

Richards, E. G. 2013. «Calendars». Capítulo 15 de *Explanatory Supplement to the Astronomical Almanac*, 3.^a ed., editado por S. E. Urban y P. K. Seidelmann. Mill Valley, CA: University Science Books. URL estable: https://aa.usno.navy.mil/downloads/c15_usb_online.pdf

Šprajc, Ivan, Takeshi Inomata, y Anthony F. Aveni. 2023. «Origins of Mesoamerican Astronomy and Calendar: Evidence from the Olmec and Maya Regions». *Science Advances* 9 (1): eabq7675. DOI 10.1126/sciadv.abq7675.

Steele, J. M. 2020. «Geminis and Babylonian Astronomy». En *Instruments, Observations, Theories: Studies in the History of Astronomy in Honor of James Evans*, editado por Alexander Jones y Cristián Carman, 205–212. DOI 10.5281/zenodo.3975741.

KNO-008 · Instruments That Extend the Senses

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

Article Abstract

The phrase in this article's title is a useful shorthand and a poor description. A thermometer does not sharpen anyone's sense of heat; it converts temperature into a length the eye can read. What an instrument builds is a chain of conversions whose last link happens to be perceivable, and the human sensorium is not extended anywhere along it. This article follows the consequences: that trusting a reading means trusting every link, that calibrating an instrument is circular before it is routine, and that the interesting instruments are the ones whose output corresponds to no sense at all. It discharges the longitude problem two earlier articles deferred, and finds the familiar account of how it was solved does not survive contact with the scholarship.

In 60 Seconds

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 This Question 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.

1. What an instrument actually does

Consider the difference between three devices that are usually filed together.

Spectacles correct a defect in an existing sense and change nothing about what kind of thing can be perceived. Light still arrives at the eye and the eye still sees it.

A **telescope** operates in the same modality and pushes past a natural limit. It is still light, still the eye, still seeing, and the objects it reveals are of a kind the unaided eye already knows.

A **thermometer** does something categorically different. Temperature is not a thing eyes perceive at all. The instrument couples a thermometric substance to the quantity of interest and converts it into a length, and the length is what the observer actually sees. The observer does not perceive temperature more acutely. The observer perceives a column of liquid and infers.

Of the three, the spectacles **aid** a sense and the telescope **extends** one beyond its natural range, which are the two literal cases and the least interesting ones; the table below reserves each word for one of them. The third is what many scientific instruments do, and the pattern recurs widely: the device converts an inaccessible quantity into an accessible one, and every step of that conversion is a place where the chain can be wrong.

This suggests a working vocabulary, offered as a stipulation and not a discovery.

	---	---		Aid Restores or improves a sense within its own modality.	Extension Same modality, beyond the natural range.	Transducer Converts a quantity the senses cannot register into one they can.	Inferential instrument Produces no perceivable analogue at all; the output is a number or a statistical statement, and the phenomenon is reached only through a model.
--	-----	-----	--	--	---	---	---

The four are not stages. They coexist, and a single apparatus can contain all four in series.

Where That Leaves Us

Three devices filed together do three different things, and only two involve a sense. Restoring sight and reaching past its natural range are the literal cases and the least interesting. The third turns what no eye registers into what an eye can, making the observer a reader of columns rather than a perceiver of heat, and every conversion between a place the result can fail.

2. The chain has to be trusted link by link

If an instrument is a chain of conversions, then the question of whether to believe it is not a question about the senses. It is a question about the chain, and there is no point along it where perception settles the matter.

The telescope is an early example of instrument-mediated observation. Galileo's observations of 1609 and 1610 were not accepted on sight; the instrument and its reports spread across Europe and were taken up by multiple observers. The precise sequence of that contested reception is not reconstructed here, because the accessible sources do not support a more detailed chronology.

The structure is worth naming because it recurs. A new instrument reports something no one has seen. The report cannot be checked by looking, because looking is what is in question. One check that works is agreement among independent instances of the same chain, built by people with no stake in one another's results. **It is not sufficient on its own.** Copies of an instrument can share a design, a calibration route or a background model, and instruments that share any of those can agree for the same wrong reason. What raises the agreement above that are checks of a different kind: feeding the

chain a **known input** and seeing whether it returns it, measuring the same quantity with an **instrument of another design, intervening** on the phenomenon and seeing the reading follow, and holding the result against an **independent constraint** that does not come through the same chain.

Where That Leaves Us

If belief in a reading is belief in a chain, no amount of looking settles it, and the founding case bears that out: the instrument was trusted once people with no stake in each other built their own and got the same thing. Agreement alone is not enough, since copies can share a design and be wrong together. Known inputs, other designs, intervention and outside constraints are what raise it.

3. Calibration is circular before it is routine

The deepest problem in instrumentation is not building the device. It is establishing what its readings mean, and that problem is circular at the outset.

To justify a thermometer one needs the law relating the behaviour of the thermometric substance to temperature. To establish that law one needs to measure temperature. The circularity is real and has a name: the problem of nomic measurement, set out in a study of eighteenth and nineteenth century thermometry which argues that the circle was broken not by decree but by iterative, self correcting refinement across competing thermometric traditions over a long period (Chang 2004).

Two consequences follow and both are easy to miss.

First, **an instrument does not simply measure a quantity that was already well defined**. Before workable thermometry there was no operational way to say that two bodies were at the same temperature, and fixing the measurement procedure was part of fixing what the quantity was.

Second, **a circularity of this kind can arise where a novel phenomenon is at stake**, and it is a risk to be looked for rather than the structure of every new instrument. Where it does arise, whether an apparatus is working correctly is judged by whether it produces the expected result, and what result to expect is what the apparatus was built to find out. **The nomic-measurement case of the previous paragraph is narrower still**: it belongs to the foundational cases where the quantity itself is being fixed along with its measurement, and a thermometer checked against a maintained standard is not in that position. In the sociology of science this is called the experimenter's regress, associated with Collins (1992). The University of Chicago Press 1992 edition lists *Detecting Gravitational Radiation: The Experimenter's Regress* as chapter 4. That bibliographic fact supports the attribution, but not every broader interpretation of the chapter's argument.

There is a published critique of the regress and it belongs here rather than in a footnote. Franklin argues, from the gravitational wave episode, that the regress was broken by reasoned argument (Franklin 1994). Read it as narrowly as it is written: it denies that the regress is inescapable, not that it never arises. This section already treats the circularity as a risk to look for rather than as the structure of every new instrument, and the critique is what makes that treatment a position in a debate rather than a hedge.

Where That Leaves Us

Establishing what a reading means is harder than building the device, and at first the two requirements chase each other. History shows the circle worn away by rival traditions correcting one another rather than cut by decree. Fixing how to measure was part of fixing what was measured, and the circularity is a risk worth hunting, not the shape of every device. A published critique argues it can be broken.

4. Longitude, and a debt discharged

KNO-006 deferred the longitude problem as belonging to time. KNO-007 examined it and forwarded it here, on the ground that it is not a question about what time is but about transporting a time standard. This article discharges the debt, and the scholarship overturns the familiar story.

The popular account is that John Harrison's marine timekeepers solved the problem. The scholarly account is that longitude finding in practice required hardware, procedures and tables, and skilled people exercising judgment, and that no single artefact solved anything on its own (Miller 2015).

More striking is the fate of the rival method. Lunar distances, made practical by Tobias Mayer's tables in the 1750s and by Maskelyne's *Nautical Almanac* from 1767, were treated by the Board of Longitude in 1767 as co equal with chronometers. The *Endeavour* voyage of 1768 to 1771 used lunar distances, at a time when Harrison's H4 was the only chronometer accurate enough and there was exactly one of it. **This article does not say that they were the primary method, there or on Cook's other voyages**, which is a claim the accessible pages do not settle. Lunar tables remained in nautical almanacs well into the following century, and the method declined as chronometers became cheap rather than because it failed (de Grijns 2020).

What that source also describes is the estimate the method rested on. On the *Endeavour* voyage the lunar work was combined with the tried and tested dead reckoning approach, which remained a fundamental yet supplementary navigation tool when weather interrupted the lunar observations; ship speed was estimated from a chip float on a knotted log line timed by a glass, and the procedure could not correct for drift or currents (de Grijns 2020). **No longitude figure is given here**, because what the passage supplies is a procedure and its limits and not a number.

The general point is about instruments rather than about ships. **An instrumental solution and a computational solution can be equally correct**, and which one wins can be decided by manufacturing cost rather than by accuracy. Reading the outcome backwards as a demonstration that the clock was the right answer imports a conclusion the history does not supply.

Where That Leaves Us

A debt two earlier articles left open closes here, and the familiar answer turns out to compress a longer story. Finding position at sea took hardware, tables, procedures and judgment together, no artefact deciding anything alone. The rival method was treated as an equal, leaned on an estimate blind to currents, and faded as its competitor grew cheap. Cost decided as much as accuracy did.

5. Instruments with no perceptual analogue

Some instruments detect things for which no sense could exist, and these show the transduction structure at its clearest.

Radio astronomy began by accident and not by design. Karl Jansky was employed to trace static interfering with transatlantic telephony (NRAO/AUI Archives, "Papers of Karl G. Jansky"), using a rotating antenna (Jansky 1932) operating at 21 MHz (Kellermann et al. 2023). He first suspected the Sun. By early 1933 the direction of the hiss no longer tracked the Sun, and in February he concluded that it came from a fixed direction in space (Kellermann et al. 2023), in the direction of the centre of the Milky Way (NRAO, Jansky Lectureship). The primary report records that the source of this disturbance had been thought somehow associated with the sun (Jansky 1933), which is the hypothesis the later analysis displaced, and that same institutional history identifies the result as radio waves from the central region of our Milky Way Galaxy. The instrument was not built to observe anything. It was built to reduce noise, and the noise was the observation.

Gravitational wave detection has no perceivable stage at all. The first reported direct observation gives a peak strain of 1.0×10^{-21} (Abbott et al. 2016). Nothing in that chain terminates in a sensation. Moreover the detection itself is a statistical judgement. **One of the methods used for it** is matched filtering: the data are correlated against a bank of theoretical template waveforms, and for Gaussian noise this reduces to a maximised data to template correlation that optimises signal to noise, which is a form of maximum likelihood detection (Jaranowski and Królak 2012). **That method is not the whole of how such a signal is found**, and this article does not present it as the definition of the detection.

The initial detection did not come from template matching at all. The collaboration's own summary records that the initial detection was made by a so called burst search, and that two separate matched filtering pipelines, called PyCBC and GstLAL, recovered the event afterwards (LIGO Scientific Collaboration, science summary). So the templates were a cross check on a detection made without them, and an account that opens with matched filtering has the order the wrong way round.

That is worth stating plainly. **The object of the claim is not seen, heard or registered by any sense. It is inferred from the degree to which a record matches a model of what such an event would produce.** The instrument does not extend a sense. It replaces perception with inference, and where the inference runs through templates it is only as good as the model that supplies them.

Where That Leaves Us

Some instruments reach objects no sense could register, and show the structure plainly. One began as an attempt to clean up interference and ended by reporting where a hiss came from. Another ends in no sensation at all, reaching its object through a model of what would produce it. Even there the initial signal was found without template matching, and the matched filter pipelines recovered it afterwards.

6. What makes two readings comparable

An instrument that only its owner can trust is not much use. What makes a reading in one laboratory comparable with a reading in another is metrological traceability, the chain of calibration back to a common reference, a concept incorporated into the third edition of the international vocabulary of metrology (JCGM 200:2012).

A citation caution belongs here rather than in a footnote. **A widely circulated copy of that vocabulary carries the date 2006 and is a pre publication committee draft, not the published edition.** Anyone citing it from that copy would be citing a draft. The current reference edition is JCGM 200:2012.

The vocabulary defines it as a property of a result, and says what it does not guarantee.

Metrological traceability is a property of a measurement result by which the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement

uncertainty; and the same entry warns that traceability does not ensure that the measurement uncertainty is adequate for a given purpose, and does not ensure absence of mistakes (JCGM 200:2012, VIM 2.41, Note 5).

The conceptual point is that comparability is not a property of an instrument. It is a property of an institutional arrangement that links instruments to one another, and it has to be maintained. An uncalibrated instrument is not a slightly worse instrument; it is an instrument whose readings do not connect to anyone else's.

Where That Leaves Us

An instrument nobody else can check is of little use, and what connects one laboratory to another is a maintained chain rather than a property of any device. The vocabulary that defines that chain is careful about what it does not promise: not that the uncertainty suits the purpose, and not that no mistake was made. An uncalibrated instrument is not slightly worse; its readings simply do not connect.

7. What the evidence cannot settle

Whether an instrument is working, in a novel case. The circularity of section 3 is not a puzzle awaiting a solution. It is a structural feature, managed by iteration and by independent replication rather than removed.

What an instrument would have shown had it been built differently. Every instrument selects. What it does not register is not thereby absent, and the absence of a signal is evidence only when the instrument's response to that signal is known.

Priority in accidental discovery. When an instrument built for one purpose registers something else, deciding who discovered what depends on what counts as recognising a phenomenon, and that is not settled by the record alone.

Three cited works could be verified bibliographically but not read in full for the relevant passages; the claims that rely on them are therefore limited to what could be established.

Where That Leaves Us

Four things stay open here, each for a different reason. Whether a device works in a genuinely new case is structural rather than unsolved. The counterfactual about a differently built instrument has no answer, since selection is what instruments do. Credit for the unintended turns on what counts as recognising. And three sources were verified as publications without their relevant text being read.

8. Conclusion

Many instruments do not extend the senses at all. They enrol them at the end of chains that begin somewhere the senses cannot reach, and in those cases the epistemic weight sits on the chain rather than on the moment of perception at its end. **Some instruments do extend a sense**, and section 1 separates them out; the claim of this article is about the many, not about all.

That is why the history of instruments is not a history of sharper eyes. It is a history of calibration, of replication among people who do not trust one another, of standards maintained by institutions, and of models that turn a correlation into a claim about a black hole merger. When an instrument is believed, what is being believed is all of that, and the reading is only the last and least of it.

Where That Leaves Us

The argument comes back to where the weight sits. For many instruments the epistemic load is carried by the chain and not by the moment of perception at its end, and section 1 marks out the ones where that is not so. Read this way, the history stops being about better eyes and becomes about calibration, mutual distrust, maintained standards and models that make a correlation mean something.

What We Know, and How

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 We Still Don't Know

Several details remain outside the evidential scope used here: the full argument in Collins on the experimenter's regress; the precise disputed reception of Galileo's telescope; the detailed conclusions of the historical thermometry study; whether lunar distances were the primary method on Cook's voyages; the wording of the published 2012 vocabulary preface relative to the earlier committee draft; and whether Sullivan's chapter independently supports the Jansky narrative. None of those points is used here as a foundation for a stronger claim.

What Could Change Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-006 · *Maps and Spatial Representation*
- KNO-007 · *Calendars and Time Measurement*

References

This article cites in the body itself, by author and date or by document designation, and carries no bracketed claim numbers. Each citation resolves to a row of KNO-008_CLAIMS_AND_SOURCES_RECONSTRUCTION.md, which names one academic or official source per claim and, from 11 September 2026, the location in the record where it was read. Uncertain claims, declared scope limits and excluded sources are in KNO-008_UNCERTAINTIES_RECONSTRUCTION.md.

Abbott, B. P., et al. (LIGO Scientific Collaboration and Virgo Collaboration). 2016. «Observation of Gravitational Waves from a Binary Black Hole Merger». *Physical Review Letters* 116 (6): 061102. DOI 10.1103/PhysRevLett.116.061102.

Bucciantini, Massimo, Michele Camerota, y Franco Giudice. 2015. *Galileo's Telescope: A European Story*. Trad. Catherine Bolton. Cambridge, MA: Harvard University Press. ISBN 978-0-674-73691-7.

Chang, Hasok. 2004. *Inventing Temperature: Measurement and Scientific Progress*. Oxford Studies in Philosophy of Science. Oxford: Oxford University Press. ISBN 978-0-19-517127-3.

Collins, Harry. 1992. *Changing Order: Replication and Induction in Scientific Practice*. Chicago: University of Chicago Press. ISBN 978-0-226-11376-0.

de Grijs, Richard. 2020. «A (not so) brief history of lunar distances: Lunar longitude determination at sea before the chronometer». *Journal of Astronomical History and Heritage* 23 (3): 495–522.

Franklin, Allan. 1994. «How to Avoid the Experimenters' Regress». *Studies in History and Philosophy of Science Part A* 25 (3): 463–491. DOI 10.1016/0039-3681(94)90062-0

Jansky, Karl G. 1932. «Directional Studies of Atmospherics at High Frequencies». *Proceedings of the Institute of Radio Engineers* 20 (12): 1920–1932.

Jansky, Karl G. 1933. «Radio Waves from Outside the Solar System». *Nature* 132: 66. Publicado el 8 de julio de 1933. DOI 10.1038/132066a0

Jaranowski, Piotr, y Andrzej Królak. 2012. «Gravitational-Wave Data Analysis. Formalism and Sample Applications: The Gaussian Case». *Living Reviews in Relativity* 15 (1): art. 4. DOI 10.12942/lrr-2012-4.

JCGM. 2012. *International Vocabulary of Metrology: Basic and General Concepts and Associated Terms (VIM)*. 3.^a edición, JCGM 200:2012, versión corregida. Bureau International des Poids et Mesures. URL estable: <https://www.bipm.org/en/committees/jc/jcgm/wg/jcgm-wg2-vim>

Kellermann, Ken, Ellen Bouton, Heather Cole y Jeff Hellerman. 2023. «Jansky @90: The Origins of a New Window on the Universe». National Radio Astronomy Observatory, 27 de abril de 2023. URL estable: <https://public.nrao.edu/news/jansky-90-the-origins-of-a-new-window-on-the-universe/>

LIGO Scientific Collaboration. Resumen científico «How we searched for merging black holes and found GW150914». La página estable consultada no muestra fecha de publicación; no se asigna una fecha no documentada. URL estable: <https://ligo.org/science-summaries/gw150914cbc/>

Miller, David Philip. 2015. «Longitude Networks on Land and Sea: The East India Company and Longitude Measurement "in the Wild", 1770–1840». En *Navigational Enterprises in Europe and its Empires, 1730–1850*, ed. Richard Dunn y Rebekah Higgitt, 223–247. Londres: Palgrave Macmillan. ISBN 978-1-137-52063-0. DOI 10.1057/9781137520647_12.

National Radio Astronomy Observatory. Página «Jansky Lectureship». La página estable consultada no muestra fecha de publicación; no se asigna una fecha no documentada. URL estable: <https://science.nrao.edu/science/jansky-lecture/>

National Radio Astronomy Observatory, NRAO/AUI Archives. «Papers of Karl G. Jansky, 1923–1998». Inventario del fondo. URL estable: <https://www.nrao.edu/archives/jansky-finding-aid>

Sullivan, W. T., III. 1984. «Karl Jansky and the discovery of extraterrestrial radio waves». En *The Early Years of Radio Astronomy: Reflections Fifty Years after Jansky's Discovery*, 3–42. Cambridge: Cambridge University Press. DOI 10.1017/CBO9780511564956.003.

Boundary note

With KNO-007, Calendars and Time Measurement. KNO-007 treats what is agreed to be measured; this article treats the devices and the chains that realise it. **The longitude problem, deferred by KNO-006 and forwarded by KNO-007, is discharged here in section 4 and is not open anywhere else in the series.**

With KNO-011, Measuring Before Theorising. This is the closest boundary in the series and needs care. **KNO-008 covers the circularity of validating a particular instrument**, because that is a problem about the device. **KNO-011 covers measurement without theory as an epistemic strategy**, including whether a research programme can proceed by measuring first, and that broader question is not anticipated here.

With KNO-013, Models, Maps and Reality. Section 5 states that gravitational wave detection reaches its object through a model. **The general question of what a model is and how model dependence bears on realism belongs to KNO-013** and is not developed here.

With KNO-006. The structural resemblance between an instrument's selectivity and a map's is real but limited, and this article does not press it. Map projection is constrained by a geometric theorem; instrument selectivity is a design decision. **They are analogous in consequence and different in kind.**

KNO-009 · Archives, Libraries and Collective Memory

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

Article Abstract

An archive is not a container of the past. It is the output of a long series of decisions about what to keep, and those decisions do not survive in the material they produce, so an archive is evidence about the keeping as much as about the kept, and its surviving holdings cannot by themselves reveal the full shape of what is missing. KNO-009 separates records, archives and libraries by the selection logic each runs on, takes the professional definition of appraisal literally as a decision about what to destroy, and uses a four part account of where silences enter historical production. It examines the archive as an instrument of governance, Alexandria as a case where the dramatic explanation is the weaker one, and why digital material needs continuous action to persist.

In 60 Seconds

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 This Question 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.

1. Three institutions that are not one thing

Records, archives and libraries are routinely discussed together and they operate on incompatible principles.

A record is created to conduct or document an activity; whether it is preserved afterwards is a separate question. A receipt exists to settle an account; a minute exists to bind a decision. Its evidential value comes from the activity for which it was created, while its later survival introduces a separate preservation filter, which is what makes it good evidence and unrepresentative evidence at once.

An archive keeps the records of an activity, and its unit is the body of material that the creating body generated. That creator is often an institution and is often a person or a family: personal papers and acquired collections are archives on the same principle, and using archive as a synonym for institutional records leaves them out. Its organising principle is provenance, which is a principle of context and of keeping material of distinct origins separate rather than a rule for choosing what to take in. What it holds is still determined by whose activity was documented and whose records were judged worth accessioning, and that judgement is appraisal, which section 2 treats on its own.

A library collects works that were made to be circulated. Its selection principle is subject and use. Much of what it holds was published, which means that part already passed through another filter before the library ever saw it, and libraries also hold unpublished material, so library and published are not the same category.

The three therefore have different silences. An archive is silent about people who generated no records or whose records nobody kept. A library is silent about what never entered the circuits it collects from. A record is silent about whatever its purpose did not require it to state. **Treating the three as one repository of the past merges three different distortions into an apparently single view.**

Where That Leaves Us

Three institutions are filed together and run on incompatible principles. A record is created to conduct or document an activity, and whether it is preserved afterwards is a separate question; an archive keeps what one activity generated and organises it by provenance; a library gathers what was made to circulate. Each is therefore blind in a different place, and merging them merges three distortions into one apparent view.

2. Appraisal is a decision about what to destroy

The word the profession uses for deciding what enters an archive is appraisal, defined as the process of identifying materials offered to an archives that have sufficient value to be accessioned, with the underlying professional debate framed as being about what to save and what to destroy (SAA Dictionary of Archives Terminology).

That framing deserves to be taken literally. Everything cannot be kept, so a professional judgement determines what will not remain available through that archival series, and the judgement is made under the constraints and assumptions of its own moment. Archival theory has recognised the archivist as an active appraiser rather than a passive curator, and has traced a sequence of shifting paradigms in how that role has been understood (Cook 2013).

Two consequences follow.

The decision is irreversible, and the surviving series does not show it. What was destroyed leaves no gap there; the series simply reads as complete. That is a property of the series and not of the archive as a whole: appraisal decisions can be documented separately, in disposal schedules and appraisal reports, and where they are, what is unrecoverable is the destroyed content rather than the fact that a decision was taken.

The archive's composition is evidence about the appraiser. What a period thought worth keeping is a fact about that period, and it is legible in the archive whether or not anyone intended it to be.

Where That Leaves Us

Appraisal is the professional word for deciding what enters an archive, and the debate behind it is framed as what to save and what to destroy. The decision is irreversible and the surviving series reads as complete, though disposal schedules and appraisal reports can document that a decision was taken. What a period kept is evidence about that period.

3. Where silences enter

The most useful account of how absences are produced does not treat silence as a single event. It locates four distinct moments:

Trouillot places them at fact creation, which is the making of sources; fact assembly, which is the making of archives; fact retrieval, which is the making of narratives; and retrospective significance, which is the making of history in the final instance (Trouillot 1995, 26).

The value of the four part division is that it separates things usually blurred together. A source may never have been made. A source that exists may never have entered an archive. A record in an archive may never be found or used. And material that is found may be judged unimportant. **These are four different failures with four different remedies, and no amount of better searching fixes the first two.**

It also explains why an archive cannot audit itself from its own holdings alone. Silence can enter at each moment, and surviving holdings alone do not reveal which moment produced a particular absence.

Where That Leaves Us

Trouillot locates silence at four moments: fact creation, fact assembly, fact retrieval and retrospective significance. The value of the division is that these are four different failures with four different remedies, and better searching does not touch the first two. The same division explains why surviving holdings cannot audit themselves: silence can enter at each moment, and the holdings alone do not reveal where a particular absence arose.

4. The archive as an instrument of governance

If archives are produced rather than found, then their production is a subject in its own right. This is the move usually called the archival turn: reading archives as sites of knowledge production and as monuments of states rather than as neutral repositories to be mined for facts (Stoler 2002).

The practical point does not depend on the theory. **An archive answers the questions its creators needed answered.** A colonial administration's files are excellent evidence about what that administration worried about, and poor evidence about the lives of those it governed, and the second failing is not a gap to be filled by looking harder in the same files.

Where That Leaves Us

Once archives are produced rather than found, their production becomes a subject. The practical point survives without the theory: files answer the questions their creators needed answered, so a colonial administration's papers are strong evidence about official anxiety and weak evidence about the governed, and the second weakness is not repaired by searching the same files harder.

5. Alexandria, where the dramatic explanation is the weaker one

The destruction of the Library of Alexandria is the standard emblem of catastrophic knowledge loss, and it is usually attributed to a fire.

One scholarly argument points elsewhere. On Bagnall's account, ordinary decay of papyrus, compounded by the loss of institutional will and funding to keep copying texts, is sufficient to account for the disappearance without any hostile act, and he puts it that even without hostile action the Library, or Libraries, of Alexandria would not have survived antiquity (Bagnall 2002). **That is his argument and it is given here as his**, not as a settled finding of the field.

This matters beyond one library. A collection on a perishable medium is not a thing that persists unless destroyed; it is a process that stops unless maintained. Copying is the mechanism of survival, copying costs money and institutions, and the failure mode is not an event but a withdrawal of attention. **A catastrophe is easier to narrate than a budget, which is one reason the catastrophic account persists.**

Where That Leaves Us

On Bagnall's account ordinary papyrus decay, with institutional will and funding to keep copying withdrawn, suffices to explain the disappearance without a hostile act. That is his argument and it is given as his. The wider point is that a collection on a perishable medium is a process that stops rather than a thing that persists unless destroyed.

6. When what cannot be found effectively disappears

An archive that holds material nobody can locate is functionally equivalent to one that does not hold it. Access is governed by description, and description follows standards; the international general standard for archival description was issued in a second edition by the International Council on Archives (ISAD(G) 2000).

The consequence is that **description governs what can be retrieved**, which is not the same as a second appraisal. What is described at item level can be found; what is described only at the level of a large aggregate can be found only by someone who already suspects it is there. The effect on a researcher can resemble destruction, and the two differ in exactly the way that matters: an absence in a finding aid does not demonstrate an absence in the holdings, and description, unlike appraisal, can be redone.

A boundary belongs here. How categories are built, what they do to the things they sort, and whether a classification can be neutral are the subject of KNO-010 and are not developed in this article. KNO-009 treats description only as a constraint on retrieval.

Where That Leaves Us

Description decides retrievability, which is a constraint and not a second appraisal. What is catalogued in detail can be reached; what is catalogued only as a large aggregate can be reached by someone who already suspects it. To a researcher that can feel like destruction, and it differs where it counts: a finding aid is not the holdings, and cataloguing can be done again.

7. The digital paradox

Digital material is commonly assumed to be permanent because it copies perfectly. The assumption is wrong in its general form, and so is its mirror image. Parchment does not survive neglect indefinitely, and not every digital object requires continuous migration. What holds is narrower: a digital object depends on its carrier and its format remaining readable, and where either stops being so it needs active intervention and an institution that keeps providing it.

Two measured symptoms.

References to web material decay. A study of over one million references from more than three and a half million articles published between 1997 and 2012 reports that one out of five articles in science, technology and medicine suffers from reference rot, meaning it is impossible to revisit the web context that surrounds them some time after publication (Klein et al. 2014).

The institutional response is a reference model for open archival information systems, currently in force as ISO 14721:2025, third edition, jointly maintained with the space data systems committee as CCSDS 650.0-M-3.

The structural point is the same as section 5 in modern form. Survival is not the default state of a collection. It is the result of continuous expenditure, and the digital case makes the expenditure continuous rather than periodic.

Where That Leaves Us

A digital object depends on its carrier and its format staying readable, and where either stops it needs intervention and an institution to keep providing it. A study of over one million references reports that one in five articles in science, technology and medicine suffers reference rot. Survival is continuous expenditure rather than a default state.

8. What the evidence cannot settle

The size and shape of what is missing. An archive cannot measure its own omissions, because the evidence that would measure them is what is absent. Estimates of loss come from outside the archive or not at all.

Whether an omission was deliberate. Destruction, neglect, cost and censorship can leave the same trace, which is none. Distinguishing them needs external evidence about the institution.

Whether the surviving proportion is representative. There is no general warrant for assuming that it is, and in the cases discussed here it is not; by how much and in which direction cannot be determined from the survivors, and neither can the universal claim that survival is never representative.

Sources with unresolved bibliographic identity were excluded rather than retained with a caveat. The remaining argument uses only sources whose identity and role could be stated explicitly.

Where That Leaves Us

The evidence stops at four points. Omissions cannot be sized from the surviving holdings, since the evidence that would size them is the missing evidence. Deliberate removal, neglect, cost and censorship can be indistinguishable from those holdings alone. Representativeness cannot be settled from survivors, and the sweeping denial of it cannot either. Three entries were withdrawn rather than kept with a caveat.

9. Conclusion

Collective memory is not a store that fills. It is a set of institutions that continuously decide what to carry, and the decisions become invisible in the carrying. An archive is therefore two documents at once: the record of what it holds, and the unwritten record of what its keepers thought worth holding.

Reading it well means holding both in view, and accepting that the second cannot be read directly. That is not a defect to be repaired by a better archive. It is what an archive is, and the practical response is not to trust the survivors less but to know what kind of thing they are: the output of a filter that does not describe itself.

Where That Leaves Us

Collective memory is a set of institutions deciding what to carry, and the deciding disappears into the carrying. What a repository holds and what its keepers thought worth holding are two different records, and only the first is directly legible. Treating that as a defect to repair mistakes what the institution is. What is established, and how, follows.

What We Know, and How

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 We Still Don't Know

The size and shape of what is missing, because the evidence that would measure it is what is absent. Whether an omission was deliberate, since destruction, neglect, cost and censorship can leave the same trace and external evidence about the institution is what separates them. Whether the surviving proportion is representative, which cannot be settled from the survivors, and neither can the universal claim that survival never is. Whether Bagnall's account of Alexandria is the field's, rather than his: this article gives it as his argument. The excluded operations are not used to support any claim in this article.

What Could Change Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-008 · *Instruments That Extend the Senses*
- KNO-010 · *Classification as a Knowledge Tool*

References

Bagnall, Roger S. 2002. «Alexandria: Library of Dreams». *Proceedings of the American Philosophical Society* 146 (4): 348–362. URL estable: <https://archive.nyu.edu/bitstream/2451/28263/2/D172-Alexandria%20Library%20of%20Dreams.pdf> CCSDS. 2024. *Reference Model for an Open Archival Information System (OAIS)*. CCSDS 650.0-M-3, Issue 3. Consultative Committee for Space Data Systems. URL estable: <https://ccsds.org/Pubs/650x0m3.pdf> Cook, Terry. 2013. «Evidence, memory, identity, and community: four shifting archival paradigms». *Archival Science* 13 (2): 95–120. DOI 10.1007/s10502-012-9180-7. ICA. 2000. *ISAD(G): General International Standard Archival Description*. 2.^a edición. Ottawa: International Council on Archives, Committee on Descriptive Standards. ISBN 0-9696035-5-X. URL estable: https://www.ica.org/app/uploads/2024/01/CBPS_2000_Guidelines_ISADG_Second-edition_EN.pdf ISO. 2025. *Space Data System Practices: Reference Model for an Open Archival Information System (OAIS)*. ISO 14721:2025, 3.^a edición. Ginebra: International Organization for Standardization. URL estable: <https://www.iso.org/standard/87471.html> Klein, Martin, Herbert Van de Sompel, Robert Sanderson, Harihar Shankar, Lyudmila Balakireva, Ke Zhou, y Richard Tobin. 2014. «Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot». *PLOS ONE* 9 (12): e115253. DOI 10.1371/journal.pone.0115253. SAA. *Dictionary of Archives Terminology*, entrada «Appraisal». Society of American Archivists. URL estable: <https://dictionary.archivists.org/entry/appraisal.html> Stoler, Ann Laura. 2002. «Colonial Archives and the

Arts of Governance». *Archival Science* 2 (1): 87–109. DOI 10.1007/BF02435632. Trouillot, Michel-Rolph. 1995. *Silencing the Past: Power and the Production of History*. Boston: Beacon Press. ISBN 0-8070-4310-9.

KNO-010 · Classification as a Knowledge Tool

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

Article Abstract

A classification is a working instrument, not a picture of the world. It is built so that something further can be done: a specimen found again, a death counted, a book shelved, a population governed. This article reads classification as a tool with declared purposes and measurable costs. No working system Bowker and Star examined meets the three properties an ideal one would have, the ICD shows how one scheme handles what its categories miss, and grouping, naming and identifying are separate operations a single scheme quietly fuses. The species case shows a category that is indispensable and unsettled at once; rank shows a convention that is not comparable across groups. Folk classification sits beside scientific taxonomy rather than below it. Applied to people, the same properties acquire consequences a specimen does not suffer. Revision is paid for twice, and part of the bill falls on everything already counted.

In 60 Seconds

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 This Question 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.

1. What a classification is asked to do

A classification is not a description. It is a device for putting things into groups so that further work becomes possible, and its groups are made by the work they will have to support. Bowker and Star define a classification as a segmentation of the world and a classification system as "a set of boxes (metaphorical or literal) into which things can be put to then do some kind of work" [1]. The phrase to note is the last one. A scheme that supports no further operation is not a poor classification. It is not one.

This immediately makes classification an engineering problem with an epistemic surface. The engineer asks whether the scheme lets the operation run. The reader of the scheme asks what the operation required the world to look like.

Bowker and Star set out three properties an ideal classification would have: it operates by consistent and unique classificatory principles, its categories are mutually exclusive, and the system is complete [2]. Each is reasonable on its own. Bowker and Star report that none of the working systems they examined satisfies all three, and they extend that beyond their sample as an argued doubt rather than a proof: "No real-world working classification system that we have looked at meets these 'simple' requirements and we doubt that any ever could" [3].

This is the first thing to understand about classification and it is not a complaint. The three requirements are a standard against which real schemes are read, not a specification the systems examined were found to meet. A scheme that fails them is normal. A scheme presented as meeting them is being misdescribed.

Where That Leaves Us

A classification is a device built so that a further operation can run, so its groups were shaped by that operation rather than found in the world. Three properties would make one ideal: consistent principles, mutually exclusive categories, completeness. None of the systems Bowker and Star examined satisfies all three, and they doubt any could. Failing that standard is ordinary; claiming to meet it is the thing to distrust.

2. The residue

Incompleteness is the requirement whose failure is easiest to see, because a scheme that is not complete has to do something with whatever its categories did not anticipate. The ICD shows one explicit way a working classification handles cases its main categories do not capture.

In the International Classification of Diseases the residue is explicit and it is distributed. Bowker and Star observe that the phrase "not elsewhere classified" appears throughout the entire ICD but nowhere as a top-level category [4]. That distribution is informative. A single top-level box for the unclassified would gather the unnamed into one place. Distributing the residue through the branches instead keeps the effects of uncertainty local to the branch in which they arise. That is what the observation establishes; what a reader concludes from the arrangement about the scheme as a whole is not established by it.

A residual category is not a gap in a classification. It is a working part of it, and reading a scheme well means reading where its residues sit and how much traffic they carry.

Where That Leaves Us

Incompleteness is the requirement whose failure shows most plainly, because whatever the categories did not anticipate still has to be placed. In the ICD the residual phrase runs through the branches and never sits at the top, which keeps uncertainty local to wherever it arose. That distribution is what the observation establishes. Reading a scheme well means noticing where its residues sit and how much traffic passes through them.

3. Three operations that a single scheme hides

Classification is regularly used as one word for three different operations, and separating them removes a large amount of confusion.

Grouping decides which things belong together. It is the operation the previous sections describe.

Naming attaches a label to a group. It is governed by conventions of its own, and those conventions can be changed without changing a single grouping. Nothing about the world decides whether a group is called by one name or another; a code of nomenclature does.

Identifying decides which group a particular thing belongs to. It runs on evidence available at the moment of the decision, which is often much less than the evidence that established the group.

The three come apart in practice. Two workers can agree on the groups, disagree on the names, and both be right about the code they follow. Two workers can agree on the names and place the same specimen differently because the identification key resolves what they can see rather than what the group is defined by. **When a scheme is criticised, it is worth asking which of the three operations is actually being criticised**, because the remedies are unrelated.

Where That Leaves Us

One word covers three operations. Grouping decides what belongs together, naming attaches a label under a code that can change without moving anything, and identifying places a particular thing using whatever evidence is at hand. They come apart in practice: agreement about groups survives disagreement about names. Sorting a criticism into the three is worth the trouble, because their remedies have nothing to do with each other.

4. The species problem

Biology gives the clearest case of a classification whose most used category has no agreed definition. Wells and colleagues state that species "remain controversial in terms of definition, delimitation and reality" [5]. This is not a minority position and not a recent one. Species are counted, protected by law, listed as extinct and used as the unit of a very large part of biological research, and the unit itself is contested.

Their proposal is worth stating precisely because it is not a retreat. They argue that "treating species as a heuristic can be consistent with a theoretical definition of what species are" [6], the theoretical content being that species are clusters of closely related individuals responding in a similar manner to comparable sets of evolutionary and ecological forces [6]. A heuristic in this sense is not an admission of failure. It is a statement about what kind of instrument the category is: one that does reliable work across a wide range of cases without being reducible to a single decision rule.

The species case shows something general. A category can be indispensable and unsettled at the same time, and the indispensability is not evidence that the definition has been found.

Where That Leaves Us

The species is the unit of much of biology and it is contested, while that same unit is counted, protected by law and listed as extinct. Wells and colleagues call it controversial in definition, delimitation and reality, and argue that treating the category as a heuristic can sit alongside a theoretical account of what species are. Needing a category constantly is no evidence that its definition has been found.

5. Rank is a convention

Above the species the difficulty changes shape. The question is no longer what the category is but what its level means.

Lee and Skinner state that "there are no generally accepted criteria for specifying the rank to which a taxon should be referred" and that difficulties attach to the criteria that have been proposed; they add that "taxa assigned to a particular rank are rarely equivalent entities" across different groups [7]. A family of beetles and a family of primates are not comparable quantities. They share a word.

Phylogenetic nomenclature responded by proposing to name clades without ranks. Benton reports that most proponents hold that ranks should be abandoned, while noting that the PhyloCode itself does not demand this, and he objects that ranks carry usable information about the relative inclusiveness of clades and serve as practical proxies in macroevolutionary and biodiversity work: "why abandon ranks when nothing other than semantic purity is gained by doing so?" [8].

Neither position disputes that ranks are not comparable across groups. What Lee and Skinner report and what Benton argues differ over whether a convention known not to be comparable is still worth keeping for the work it does. They are two stated positions among others, not the two halves of the field. That is a question about the tool, not about nature, and it is the right shape of question to ask of any classification.

Where That Leaves Us

Above the species the question changes from what the category is to what its level means. No criterion for assigning rank is generally accepted, Lee and Skinner report, and taxa at one rank are rarely equivalent across groups. Benton does not dispute that and argues the convention still earns its keep. Their disagreement is about the tool, which is the right kind of question to put to any scheme.

6. Folk classification, placed beside and not below

Scientific taxonomy is one classification of living things among many, and the others are not earlier drafts of it.

Ethnobiological work has found recurrent structure. Atran reports that Berlin established a standard terminology of folk-biological ranks: the folk kingdom rank, the life form rank, the generic or generic species rank, the folk specific rank and the folk varietal rank [9]. The recurrence is the finding. It says that human groups working independently arrive at hierarchies of broadly similar depth, and that one level in particular carries most of the naming and learning.

The fit with scientific taxa is partial and it is uneven in a specific way. Atran states that generic species "often correspond to scientific genera or species, at least for those organisms that humans most readily perceive, such as large vertebrates and flowering plants", and that on occasion they correspond instead to local fragments of biological families, orders, or with invertebrates even higher taxa [10]. Atran reads that unevenness as tracking perceptual salience rather than phylogeny, and the reading is his contribution to a debate rather than a settled result. It is offered here as the pattern he reports across the cases he examined, not as a rule that fixes in advance where any two schemes will converge.

Neither the recurrence nor the mismatch supports a developmental ladder. A classification that groups by what an organism does, where it is found, or what may be done with it is answering a different question from one that groups by descent, and answering a different question is not answering the same question badly.

Where That Leaves Us

Independent human groups converge on hierarchies of comparable depth, and Atran reports the standard terminology Berlin established for their ranks. The correspondence with scientific taxa is partial and uneven, and Atran's reading of the unevenness is his contribution to a debate rather than a settled result. A scheme that groups by use or habitat answers a different question from one that groups by descent.

7. When the thing classified is a person

Everything above applies to specimens. When the objects are people the same structural features acquire consequences that a specimen does not suffer.

Bowker and Star's term for the central mechanism is torque. Where an individual life and an institutional classification do not align, the authors describe a pressure that runs in both directions: "We call this agglomeration torque, a twisting of time lines that pull at each other, and bend or twist both patient biography and the process of metrication" [11]. Both the biography and the process of measurement are twisted by the encounter, on the authors' own wording, and what is recorded afterwards carries that.

Two documented cases show the range.

Library classification. Thornton traces how ideas about human classification from antiquity informed eighteenth and nineteenth century ideas of race and ethnicity, how those ideas became embedded in the Dewey Decimal Classification, and how the scheme's treatment of these subjects changed across four editions over 130 years, reflecting the presentation of changing dominant societal views [12]. The scheme did not merely contain the assumptions of its period. It carried them forward, edition by edition, into the arrangement of collections built long after the assumptions were abandoned.

State classification of population. Posel's study of twentieth century South Africa describes classification criteria that combined appearance and general acceptance with "habits, education and speech and deportment", and finds that officials "had considerable latitude to interpret the bureaucratic criteria for racial classification in ways that allowed them to draw widely and idiosyncratically on popular litanies of biological stereotypes" [13]. The point for this article is structural and it is severe. A classification that cannot be applied consistently does not therefore fail to operate. It operates through the discretion of whoever applies it, and the resulting record shows a category where in fact there was a decision.

Neither is a misuse of a neutral instrument, and the two are not the same case. A library scheme carried the assumptions of its period forward into the arrangement of collections; a state scheme decided what a person was for administrative purposes. Both show the ordinary properties described in sections 1 to 3 operating on objects that can be harmed by being sorted, and they differ in what the sorting did.

Where That Leaves Us

With people the ordinary properties acquire consequences a specimen never suffers. Torque is Bowker and Star's name for a pressure running both ways, bending the biography and the process of measurement together. A library scheme carried its period's assumptions forward into collections built long afterwards; a state scheme decided what a person was for administrative purposes. Neither shows a neutral tool being abused, and what the sorting did differs.

8. Revision, and who pays for it

A classification that is used for long enough must be revised. Revision is paid for twice: in the present, through implementation and the retraining of everyone who codes, and retrospectively, through the data already collected.

The International Classification of Diseases is the clearest instrument for seeing this because it is revised deliberately, at long intervals, by an international body. Its eleventh revision was adopted by the Seventy-second World Health Assembly on 28 May 2019, in a resolution that "ADOPTS the eleventh revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-11), to come into effect on 1 January 2022, subject to transitional arrangements" [14].

The cost of such a change is measurable. Janssen and Kunst examined cause-specific mortality in six European countries from 1950 to 1999 and found that "despite using a concordance table based on three-digit ICD codes, in 10.8% of the revisions to codes for cause-specific mortality, significant discontinuities, regarded as being due to ICD revisions, remained", concluding that discontinuities arising from coding changes "can lead to substantial changes in long-term trends in cause-specific mortality" [15].

Read carefully, that is a statement about the limits of translation between schemes. Across the revisions those six countries passed through between 1950 and 1999, a concordance table was applied as the best available repair and roughly one revision in nine still left a discontinuity that the authors attributed to the revision itself rather than to any change in health. The figure belongs to those transitions and is not a rate for revisions in general. **A long time series built across revisions of a classification therefore contains two superimposed signals, one about the world and one about the scheme**, and only the first is usually what the series is being read for.

Where That Leaves Us

Revision is paid for twice: now, through implementation and retraining, and backwards, through everything already recorded. Janssen and Kunst measured the second bill in six countries across five decades, where a concordance table repaired what it could and about one revision in nine still carried a discontinuity the authors put down to the coding change rather than to health. The figure is theirs and not a general rate.

9. Reading a classification well

What the preceding sections support is not scepticism about classification. It is a short set of questions to put to any scheme in use.

What operation was it built to support? The categories were shaped by that operation and will fit others less well.

Where is the residue and how much passes through it? A large or growing residual category is the scheme reporting its own strain.

Which of grouping, naming and identifying is at issue? The three fail differently and are repaired differently.

Is the category settled? Indispensable and defined are different properties, as the species case shows.

Do the levels mean anything across branches? Frequently they do not, and the shared word conceals it.

Who or what is being sorted, and what happens to something that does not fit? With specimens the answer is a residual category. With people it can be torque.

What did the last revision cost, and to whom? The bill is paid by everything already recorded.

A classification is a good tool in the ordinary sense: it does specific work well, it deforms what it is not built for, and it has to be maintained. Treating it as a report on how the world is divided attributes to the world a structure that came from the work.

Where That Leaves Us

The article ends where a reader can use it: a short list of questions to put to any scheme in use, about the operation it was built for, its residue, which of the three operations is at issue, whether the category is settled, whether levels mean anything across branches, what happens to what does not fit, and what the last revision cost. What is established, and how, follows.

What We Know, and How

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 We Still Don't Know

What a species is, since the category is used everywhere and its definition is contested, and the heuristic proposal is a way of working rather than a settlement. Whether ranks should be kept, which Lee and Skinner and Benton answer differently while agreeing that ranks are not comparable across groups. How far folk generic species and scientific taxa correspond, which Atran reads as perceptual salience and offers as his contribution to a debate. And how much of any long mortality series is the scheme rather than the world, since the 10.8 per cent belongs to the transitions studied and is not a rate for revisions in general.

What Could Change Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-009 · *Archives, Libraries and Collective Memory*

References

The bracketed numbers in the body are claim numbers, not reference numbers. Each resolves to a row of the claims ledger, which names the source below that supports it.

Atran, Scott. 1998. «Folk biology and the anthropology of science: Cognitive universals and cultural particulars». *Behavioral and Brain Sciences* 21 (4): 547–569. URL estable: <https://philpapers.org/rec/ATRFAT>

Benton, Michael J. 2007. «The Phylocode: Beating a dead horse?». *Acta Palaeontologica Polonica* 52 (3): 651–655. URL estable: <https://www.app.pan.pl/archive/published/app52/app52-651.pdf>

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting things out: classification and its consequences*. Cambridge, Mass.: MIT Press. ISBN 0-262-02461-6.

Janssen, Fanny, and Anton E. Kunst. 2004. «ICD coding changes and discontinuities in trends in cause-specific mortality in six European countries, 1950-99». *Bulletin of the World Health Organization* 82 (12): 904–913. **Sin DOI localizado**. URL estable: <https://scielosp.org/article/bwwho/2004.v82n12/904-913/>

Lee, Michael S.Y., and Adam Skinner. 2007. «Stability, ranks, and the PhyloCode». *Acta Palaeontologica Polonica* 52 (3): 643–650. URL estable: <https://www.app.pan.pl/archive/published/app52/app52-643.pdf>

Posel, Deborah. 2001. «Race as Common Sense: Racial Classification in Twentieth-Century South Africa». *African Studies Review* 44 (2): 87–113. DOI 10.2307/525576.

Thornton, Lisa. 2024. «How Did We Get Here? Race and Ethnicity in Dewey Decimal Classification». *Knowledge Organization* 51 (6): 414–424. DOI 10.5771/0943-7444-2024-6-414.

Wells, Tom, Tom Carruthers, Pablo Muñoz-Rodríguez, Alex Sumadijaya, John R I Wood, and Robert W Scotland. 2022. «Species as a Heuristic: Reconciling Theory and Practice». *Systematic Biology* 71 (5): 1233–1243. DOI 10.1093/sysbio/syab087.

World Health Organization. 2019. *Eleventh revision of the International Classification of Diseases*. Resolución WHA72.15, punto 12.7 del orden del día, 72.^a Asamblea Mundial de la Salud, 28 de mayo de 2019. URL estable: https://cdn.who.int/media/docs/default-source/classification/icd/icd11/a72_29-en_icd-11-adoption.pdf

KNO-011 · Measuring Before Theorising

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

Article Abstract

Theory is often said to tell scientists what to measure, and often it does. This article follows the other direction, where numbers arrive first and stand until an explanation reaches them, and treats neither direction as the rule. It opens on the question KNO-010 handed forward, whether sorting into categories already counts as measuring, and names both answers without adopting either. It then takes three cases where the number came first: a thermometer that had to be trusted before anything could check it, long wavelength measurements a famous formula was written to fit, and an orbital residual no available theory accounted for. It closes on the reverse move of 2019, when a constant measured for a century became a definition instead. Experimental design, uncertainty analysis and the structure of data records belong to later articles.

In 60 Seconds

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 This Question 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.

1. Two orders, and why neither is the rule

There is a tidy account in which theory comes first. A theory says that a quantity exists, says roughly how large it should be, and an experiment is then designed to find it. Much of modern physics works this way, and the account is not wrong.

There is an equally tidy account running the other way. People measure something because they can, the numbers accumulate, a pattern appears in them, and a theory is eventually built that has to reproduce the pattern or be discarded. Much of the history of physics and chemistry works this way too, and that account is not wrong either.

What is wrong is treating either as the order of science. The two directions coexist, alternate within a single research programme, and sometimes both operate on the same quantity at different moments. This article argues the second direction because it is the one usually left out, not because it is the true one, and section 6 shows the first direction winning decisively in a case where the second had held for a century.

Where That Leaves Us

Two tidy accounts run in opposite directions and both describe real practice, which is why neither is the order this field follows. Theory naming a quantity and an experiment going after it happens; numbers accumulating until a theory has to reproduce them happens too, sometimes on one quantity at different moments. The second is developed here because it is the one usually left out.

2. What counts as measuring, and the question KNO-010 left open

KNO-010 stopped at the category and handed forward a question it did not argue: whether assigning a thing to a category is already a measurement. It has to be answered before anything else here, because the answer decides what counts as a case.

The permissive answer is the one most often quoted. Stevens defines measurement, in the broadest sense, as "the assignment of numerals to objects or events according to rules" [1]. On that definition his four scale types follow from what the assignment allows one to do. His table separates them by their basic empirical operation: the nominal scale by determination of equality, the ordinal by determination of greater or less, the interval by determination of equality of intervals or differences, and the ratio by determination of equality of ratios [2].

The restrictive answer objects that this makes the question unaskable. Michell observes that under Stevens's view "any rule for making numerical assignments always defines at least a nominal scale" [3]. If any rule at all produces a measurement, then whether the attribute being studied is quantitative in the first place is never put to test, and Michell's charge against a century of quantitative psychology is exactly that: the scientific question of the additivity of the attributes involved was hardly ever addressed [4].

Both positions are recorded here and neither is adopted. The consequence for this article is only procedural, and it is enough. **Under the permissive definition, classifying is measuring on the weakest scale; under the restrictive one, it is not measuring at all until the attribute is shown to be quantitative.** Everything that follows concerns quantities that satisfy the restrictive definition too, so the cases below stand under either reading. Where an article needs the distinction to do real work, it should say which definition it is using, because the two disagree about what has been achieved when a number is produced.

Where That Leaves Us

Whether classifying already counts as measuring turns on the definition in force. The permissive one takes any rule-governed assignment of numerals and sorts the results by what the assignment supports; the restrictive one objects that this never puts to test whether the attribute is quantitative at all. Both are recorded and neither adopted, and everything below satisfies the stricter reading.

3. Validating an instrument with no standard behind it

The first difficulty of measuring before theorising is not the theory. It is that the instrument has to be trusted before anything exists that could check it.

A thermometer is the standing example. To know whether a thermometer reads correctly one needs a known temperature, and to know a temperature one needs a thermometer. The eighteenth and nineteenth century work on fixed points was an attempt to break that circle from the inside, and it is the subject of Chang's study, whose publisher describes it as examining "how they managed to assess the reliability and accuracy of these instruments without a circular reliance on the instruments themselves" [5].

The circle is not a scandal, and in the cases this article discusses it was not avoided. Several measurement standards were at some point established by instruments that the standard was later used to judge. Whether that holds of every standard is not established here and is not claimed. What can be asked of such a process is not that it start from certainty but that each round leave the next one better placed, and that the improvements be visible rather than assumed. This article does not develop the general epistemology of that process, which is the subject of Chang's book and beyond the scope of the series. **What is used from that book here is limited to what was checked directly on its official abstract and table of contents.** Its body was not read, so nothing in this section attributes an argument to it beyond the description its publisher gives.

One consequence matters here. A quantity can be measured well, consistently and usefully for a long time before there is any theory of what is being measured. Temperature was measured for a long period before thermodynamics gave an account of it, and this article does not fix the length of that period, which depends on what is counted as a measurement and on which account of thermodynamics is taken as the arrival. The measurements were not waiting for the theory; the theory had to accommodate them.

Where That Leaves Us

Nothing exists that could check the first instrument, which is the difficulty here, and the thermometer is the case always cited. In the episodes discussed the circle was not avoided; how far that goes is not settled here. What such a process owes is not certainty at the start but visible improvement each turn. A quantity can be measured well long before anything explains it.

4. A formula built to fit the measurements

The clearest case of a theory constructed to reproduce a number is also one of the most consequential, and its usual telling reverses the order.

By 1900 several competing radiation laws existed for the emission of a black body. González de Arrieta records that they "were all consistent with the expected properties of blackbody radiation, but, crucially, each predicted different behaviors at large values of the product λT " [6]. The disagreement was therefore located precisely: the laws were indistinguishable in the accessible range and separated only at long wavelengths, which is a description of an experiment waiting to be done.

Rubens and Kurlbaum did it. They set out to improve an existing residual ray experiment, "using four blackbody cavities spanning a temperature range from -180 to $1600\text{ }^{\circ}\text{C}$ " [7]. The result was decisive in the negative direction first: "none of these expressions reproduced precisely the observed data for all temperatures" [8].

What happened next is the point of this section. On 7 October 1900 Rubens communicated to Planck that the limit of the radiation law at large λT was a linear function of temperature, and Planck reworked the thermodynamic properties of his empirically based blackbody system and produced an equation that predicted the observed phenomena [9]. **The sequence is what the source records, and this article reads it as a sequence:** the communication came first and the equation followed, which is not by itself an account of what caused what. Rubens confirmed the fit, and the two results were presented together: "Rubens replied that the proposed equation fitted the results perfectly, and both Planck and Rubens and Kurlbaum presented their results at the same meeting of the Physical Society on October 19" [10].

The formula was answerable to the measurements from the moment it was written. Its physical interpretation, the one that made it a founding result of quantum theory, came afterwards and had to be reconciled with a formula that already existed and already worked. That is measuring before theorising in its strongest form, and it is worth noticing that the article does not claim the measurement caused the theory. It claims something weaker and more defensible: **the theory had a target it had to hit, and the target was set by measurement.**

Where That Leaves Us

Competing radiation laws agreed everywhere the instruments could reach and parted only at long wavelengths, which describes an experiment waiting to be done. It was done, none of the expressions survived it, and the equation that did come after a communication about how the limit behaved. That order of events is what the source records. The formula answered to the numbers from the moment it existed.

5. A residual that measurement established and theory did not

A measurement can also produce a number that no theory explains, and hold it in place for decades.

In 1859 Le Verrier addressed a letter to Faye on the theory of Mercury and the motion of its perihelion, published in the Comptes rendus of the Académie des sciences [11]. His finding was quantitative and narrow. As the passage is printed, it sufficed to add thirty eight arcseconds to the secular motion of the perihelion in order to represent all the observed transits to within one second [12].

Read carefully, that sentence has an unusual structure for a discovery. It does not report a new object or a new effect. **It reports the size of a correction that has to be applied to an existing theory for that theory to match the observations**, which is to say it reports the exact magnitude of the theory's failure. The correction was small, it was stable, and it was not going away.

A residual of this kind is a peculiarly durable scientific object, and the rule for reading one is worth stating transferably. A residual is an estimate that depends on the observations behind it, on the parameters used and on the model it is a residual of. It therefore constrains only those successor theories that claim to cover the same observational domain, and only to the precision that is relevant there. It is not an anomaly in the loose sense of something odd; it is a number, established by measurement, that a successor theory covering the same ground has to reproduce within that precision. It survived every attempt to explain it away by adjusting the inventory of masses in the solar system, and it was still there when a theory that predicted it arrived more than half a century later. **It was not there unchanged.** Laskar records Le Verrier's figure as roughly thirty eight arcseconds per century, obtained from twenty one transit observations between 1697 and 1848, against the roughly forty three arcseconds per century the later theory gives [12]. The two are close and they are not the same number, and the difference is the ordinary difference between an estimate and a prediction rather than a defect in either. This article deliberately stops at the residual and does not treat the later explanation, which belongs to the history of gravitation rather than to the epistemology of measurement.

Where That Leaves Us

Decades can pass with a figure in place that nothing accounts for. Neither a new object nor a new effect was reported, but the size of a correction an existing theory needed, which is the magnitude of its failure. Such a figure is an estimate, so it binds only successors covering the same ground at the precision that matters there, and the later value is close without being the same.

6. When a measured constant becomes a defined one

The three preceding sections all run from measurement towards theory. The most recent large change in metrology runs the other way, and it is worth setting beside them so that the article is not a one directional story.

Until 2019 the kilogram was an object. As the SI Brochure's own appendix on the kilogram states, "the previous definition of the kilogram was based on the mass of the international prototype of the kilogram (IPK) immediately after the prescribed cleaning procedure" [13]. Everything expressed in kilograms was, ultimately, a comparison with one artefact.

Resolution 1 of the 26th General Conference on Weights and Measures, adopted in 2018 and effective from 20 May 2019, replaced that with a constant. The kilogram "is defined by taking the fixed numerical value of the Planck constant h to be $6.626\,070\,15 \times 10^{-34}$ when expressed in the unit $J\,s$, which is equal to $kg\,m^2\,s^{-1}$ " [14].

The inversion is complete and it is worth stating plainly. The Planck constant was measured, with improving precision, over a long period before its value was fixed; this article does not put a figure on that period. It is not measured any longer, because it no longer has a value to be found: its numerical value is now fixed by definition, and the official statement is unambiguous that "the numerical values of the seven defining constants have no uncertainty" [15]. The uncertainty did not disappear. It moved. Where there was once an uncertain constant and an exact unit, there is now an exact constant and a unit whose realisation carries the uncertainty.

The resolution itself records what happened to the prototype, and it is worth reading exactly. Resolution 1's own Appendix 2, under *Status of constants previously used in the former definitions*, states that from 20 May 2019 the mass of the international prototype of the kilogram is 1 kg within a

relative standard uncertainty equal to that of the recommended value of the Planck constant when the resolution was adopted, 1.0×10^{-8} , and that its value is determined experimentally thereafter [16].

This is not a conserved quantity changing hands. The defining constant has no uncertainty at all, which is the separate statement above with its own source; what this appendix adds is that the artefact now has an experimentally determined mass carrying an inherited uncertainty, so *moved*, in the paragraph above, describes where the work is done and not a quantity passed along.

This is a metrological convention, adopted deliberately, and not a case of theory telling measurement what to do. A committee decided to fix a number that had until then been sought, and it was only possible because measurement had made the constant known well enough that fixing it changed nothing practical. The two directions are not rivals here. The second was earned by the first.

Where That Leaves Us

The most recent large change ran the other way. A constant measured for a century was fixed by decision, and the defining value has no uncertainty because there is no longer anything to find. The prototype keeps a mass determined experimentally and carrying an inherited relative uncertainty. Nothing was passed along: what changed is where the work of determining is done.

7. Reading a measurement well

What the preceding sections support is a short set of questions to ask of any number offered as a measurement.

Which definition of measurement is in play? On the permissive one, a classification with numerals attached qualifies. On the restrictive one, nothing qualifies until the attribute is shown to be quantitative. The two disagree about what has been established.

What validated the instrument, and what validated that? Every chain ends somewhere, and the interesting question is whether the last link was examined or assumed.

Was the quantity theorised before it was measured? If not, the measurements are not confirmations of anything and should not be read as such.

Is this number a value or a residual? A residual reports the size of a theory's failure, and it constrains successors in a way that an ordinary measurement does not.

Is the constant still being measured? Some are not, by decision. A defined constant has no uncertainty because uncertainty has been moved elsewhere, not because it has been eliminated.

What would it take for this number to change? A better instrument, a redefinition, or a reinterpretation are different answers with different consequences, and a number that cannot change for any of the three is probably not a measurement.

Measuring before theorising is neither a lesser activity nor a preliminary one. It produces the constraints that theories are answerable to, and it occasionally produces a number that stands unexplained for fifty years and is right the whole time.

Where That Leaves Us

The article ends in questions to put to any number offered as a measurement. Which definition is in force, what validated the instrument and what validated that, whether the quantity was theorised first, whether the figure is a value or a residual, whether it is still being measured, and what would have to happen for it to change. What the sections establish, and how, comes next.

What We Know, and How

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 We Still Don't Know

Whether the circularity described holds of measurement standards in general, which section 3 declines to establish. How long the Planck constant was measured before its value was fixed, and how long temperature was measured before thermodynamics accounted for it, neither of which this article puts a figure on. What Chang's book argues, as opposed to what its publisher says it examines. And what caused what in October 1900: the order of events is recorded and the causal question is left open.

What Could Change Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-010 · *Classification as a Knowledge Tool*

References

The bracketed numbers in the body are claim numbers, not reference numbers. Each resolves to a row of KNO-011_CLAIMS_AND_SOURCES_RECONSTRUCTION.md, which gives one academic or official source for each and names the entry below that supports it. Uncertain claims, declared scope limits and excluded sources are listed in KNO-011_UNCERTAINTIES_RECONSTRUCTION.md.

BIPM. 2018. *Resolution 1 of the 26th CGPM (2018): On the revision of the International System of Units (SI)*. Conférence Générale des Poids et Mesures. URL estable: <https://www.bipm.org/en/committees/cg/cgpm/26-2018/resolution-1> DOI 10.59161/CGPM2018RES1E

BIPM. 2021. *SI Brochure – 9th edition (2019) – Appendix 2*, 7 de julio de 2021. Bureau International des Poids et Mesures. URL estable: <https://www.bipm.org/documents/20126/41489673/SI-App2-kilogram.pdf>

BIPM. Página «SI defining constants». Bureau International des Poids et Mesures. **Sin fecha mostrada en la página.** URL estable: <https://www.bipm.org/en/measurement-units/si-defining-constants>

Chang, Hasok. 2004. *Inventing Temperature: Measurement and Scientific Progress*. Oxford Studies in Philosophy of Science. Oxford University Press. ISBN impreso 9780195171273. URL estable: <https://global.oup.com/academic/product/inventing-temperature-9780195171273> **Manifestación en línea, Oxford Academic:** ISBN en línea 9780199835003, DOI 10.1093/0195171276.001.0001, <https://academic.oup.com/book/5530>

González de Arrieta, Iñigo. 2022. «Beyond the infrared: a centenary of Heinrich Rubens's death». *The European Physical Journal H* 47: artículo 11. DOI 10.1140/epjh/s13129-022-00044-x

Laskar, Jacques. 2017. «Des premiers travaux de Le Verrier à la découverte de Neptune». *Comptes Rendus Physique* 18: 504–519. DOI 10.1016/j.crhy.2017.10.011

Michell, Joel. 1997. «Quantitative science and the definition of measurement in psychology». *British Journal of Psychology* 88: 355–383. DOI 10.1111/j.2044-8295.1997.tb02641.x URL estable: <https://errorstatistics.com/wp-content/uploads/2015/02/michell-1997-1.pdf>

Stevens, S. S. 1946. «On the Theory of Scales of Measurement». *Science* 103 (2684): 677–680. **Sin DOI localizado sobre página consultada.** URL estable: <https://dl.icdst.org/pdfs/files3/38396a3ee53de9478a57af4605f0e1c1.pdf>

KNO-012 · Experiments and Variable Control

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

Article Abstract

In the assigned comparative designs treated here, an experiment is a comparison that someone arranged, which is its advantage over watching and the source of the difficulties that follow. Many experiments assign nothing to anybody, and none of what follows defines experiment in general. This article follows the arrangement rather than the result. It separates the two jobs the word randomisation is asked to do, generating an unpredictable sequence and concealing that sequence from whoever enrolls participants, and records the measured price of failing at the second. It distinguishes protections that end at assignment from those that must continue afterwards, which is why blinding is assessed separately. It turns to the cases where intervention is impossible and the comparison has to be found rather than made. It closes where the result itself is disputed and judging the competence of the work requires knowing what the answer should have been.

In 60 Seconds

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 This Question 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.

1. The thing an experiment adds

Observation records what happened. An experiment arranges the situation first and then records what happened, and the arrangement is the whole of the difference.

That difference buys one thing. **In an experiment the investigator decides who receives what**, so the question of why one group got the treatment and another did not has a known answer instead of an unknown one. Everything else in experimental method exists to protect that single property, and every case in the following sections is a case of protecting it, losing it, or working without it.

What this article covers, and what it does not. Many experiments assign nothing to anybody: they vary a voltage, a temperature, a reagent or a delay, with no participants, no groups and no randomisation. What follows is about **comparative designs in which units are assigned to conditions**, because that is the family in which the protections discussed here do their work. None of it should be read as a definition of experiment in general.

The word control does two jobs in this field and they are worth separating at the start. **Holding a variable fixed** means arranging that it does not vary across the comparison. **A control group** is the comparison group against which the intervention is judged: it may receive a placebo, usual care, a different intervention or nothing at all, and what makes it a control is the role it plays in the comparison rather than the absence of treatment in it. The first is a property of the setup; the second is a part of the design. A study can have a control group and hold nothing fixed, and it can hold a great deal fixed and have nothing to compare.

No setup holds everything fixed. Complete equality of conditions across a comparison is not attainable, which is why the case for a comparison is always about which differences were controlled and which were left to chance, and never about their absence. **This is a framing statement, declared as such in section 4 of KNO-012_UNCERTAINTIES_RECONSTRUCTION.md.**

Where That Leaves Us

Observation records what happened; an experiment arranges the situation first, and the arrangement is the whole difference. In the assigned comparative designs treated here it buys one property, that the investigator decides who receives what, and the rest of the method protects it. Control has two senses: holding a variable fixed belongs to the setup, a control group to the design. A study can have one without the other.

2. Why allocation is the load bearing step

If the investigator decides who receives what, the decision itself becomes the vulnerable point.

Fisher gave the classical answer, and gave it a heading that states the whole argument: section 9 of *The Design of Experiments* is titled "Randomisation; the Physical Basis of the Validity of the Test" [1]. His claim there is that no further apparatus is required, because "the simple precaution of randomisation will suffice to guarantee the validity of the test of significance" [2].

Two features of that sentence are worth marking, and a third belongs to its context. It is a claim about the **validity of a test**, not about the balance of any particular experiment, and this article does not extend it beyond what it says. It is also conditional rather than unqualified: the guarantee is offered against a background in which other avoidable errors have been avoided, not as a warrant that survives them. **The statement is treated here as conditional on the surrounding experimental context; it is**

not interpreted as a claim that randomisation removes every other avoidable source of error.

And it describes randomisation as a **precaution**, which is the right category: it is a procedure adopted in advance, whose value does not depend on how the resulting groups happen to look afterwards.

Where That Leaves Us

The vulnerable point is the decision itself, once someone is making it. Fisher's answer concerns the validity of a test and not the balance of any one experiment. The article treats the statement as conditional on a background where other avoidable errors were avoided, rather than as a guarantee that randomisation removes them. He calls randomisation a precaution.

3. Two jobs, one word

In practice randomisation is asked to do two different things, and only one of them is arithmetic.

The first is **generating** an allocation sequence that nobody can predict. The second is **concealing** that sequence from whoever is enrolling participants, so that knowing which arm the next person will fall into cannot influence whether that person is enrolled at all. Reporting standards treat these as separate obligations: the CONSORT 2010 checklist carries sequence generation, allocation concealment mechanism and blinding as three separately numbered items, 8a, 9 and 11 [3].

The separation is not bureaucratic tidiness, and there is a number attached to it. Schulz and colleagues compared trials by the adequacy of their concealment and found that "odds ratios were exaggerated by 41% for inadequately concealed trials and by 30% for unclearly concealed trials (adjusted for other aspects of quality)" [4].

Read that result carefully, because it is easy to overstate. It concerns the set of trials examined in that study, it is an association between a reported methodological feature and the size of reported effects, and the exaggeration is measured against trials judged adequately concealed rather than against a known truth. What it establishes is narrow and serious: **a randomisation that is generated properly and concealed poorly is not protected by having been generated properly.** What it does not establish is anything about the other half of the pair. The comparison sorts trials by the adequacy of their concealment, so it does not show that sequences were correctly generated in the trials whose concealment was judged inadequate. Only one of the two properties is what this study sorts on.

Where That Leaves Us

One word covers two jobs. Generating a sequence nobody can predict is arithmetic; concealing it from whoever enrolls participants is not, and reporting standards number them apart. What failing at the second costs was measured in one set of trials, against those judged adequately concealed and not against any known truth. It says nothing about the other half of the pair.

4. Before assignment and after

Allocation protects one moment. Most of an experiment happens after that moment, and different instruments are needed for it.

The Cochrane Handbook draws the line explicitly. Allocation sequence concealment "seeks to prevent bias in intervention assignment by preventing trial personnel and participants from knowing the allocation sequence before and until assignment" [5], and, crucially, "it can always be successfully implemented, regardless of the study design or clinical area" [6]. Blinding is the opposite kind of instrument: it "seeks to prevent bias after assignment and cannot always be implemented" [7].

Those two sentences together are the most useful thing in this article. One protection is available across the designs and clinical areas the handbook is describing, which are randomised trials, and it is time limited. Read outside that setting, the word always would claim more than the source does. The other is not always available at all, because some interventions cannot be disguised, and where it is available it need not run the whole length of the trial: it has to hold for the people, and at the stages, at which knowing the assignment could introduce bias. A trial can therefore be impeccable at the point of allocation and exposed for the rest of its duration, and the two states have to be assessed separately rather than summarised as a single quality score.

The same handbook goes further and separates the assessor from the participant: "blinding of outcome assessors, to avoid bias in *measuring* the outcome, is considered separately, in the 'Bias in measurement of outcomes' domain" [8]. **Who was blinded matters more than whether the word blinded appears.** A trial in which participants could not be blinded but outcome assessors were, and a trial in which the reverse holds, are exposed to different failures and should not be described by the same phrase.

Where That Leaves Us

Allocation protects one moment and most of a trial happens after it. Concealment prevents bias before and until assignment and, within the randomised trials the handbook describes, can always be implemented. Blinding works after assignment and cannot always be implemented at all, and where it can it need not run the whole length. Who was blinded therefore matters more than whether the word appears.

5. When the comparison has to be found rather than made

For many questions the arrangement in section 1 is simply unavailable. The Nobel Committee's background document states the constraint plainly: "due to financial, ethical, or practical constraints, many important questions — particularly in the social sciences — cannot be studied using a controlled randomized experiment" [9].

The response is to look for situations that did the arranging for you. The same document defines the object: "a natural experiment is an event or a situation, that is not under the control of the subjects under study, which generates variation in the variable of interest that is as if it had been randomly assigned" [10].

The load in that definition is carried by two phrases and both are conditions, not descriptions. The event must be outside the control of the subjects, which closes one route by which the people studied could select themselves into the group that suits them. It does not on its own close the others. Selection by some party other than the subjects, anticipation of the event, and systematic differences that the event tracks rather than creates all survive that condition and have to be argued away case by case. Whether the resulting variation is comparable to random assignment is therefore a claim requiring

justification, and the machinery for assessing such justifications is causal inference, which this series takes up in KNO-014 rather than here. And the variation must be *as if* randomly assigned, which is an assumption about the world rather than a fact about the design, because nobody randomised anything.

This is the difference from section 2 and it should not be blurred. In a randomised experiment the reason for the assignment is known because someone produced it. In a natural experiment the reason is argued for. That argument can be strong, and the design can be the only available route to a question that matters, but it is a different kind of warrant and this article does not treat the two as interchangeable.

Where That Leaves Us

Where the arrangement is unavailable the comparison has to be found instead. The definition carries two conditions rather than two descriptions: the event lies outside the subjects' control, which closes one route to self-selection and not the others, and the variation is *as if* randomly assigned, which is an assumption about the world. In a randomised design the reason is known; here it is argued.

6. When the result itself is what is in dispute

The hardest case is not a badly run experiment. It is a well run experiment whose result is contested, because then the question of whether it was well run cannot be settled independently of the result.

Collins put the circularity in the form that gave it its name, in a passage Franklin quotes from him: "we will have no idea whether we can do it until we try to see if we obtain the correct outcome. But what is the correct outcome? What the correct outcome is depends upon whether or not there are gravity waves hitting the Earth in detectable fluxes. To find this out we must build a good gravity wave detector and have a look. But we won't know if we have built a good detector until we have tried it and obtained the correct outcome!" [11].

That is Collins's diagnosis, and it is recorded here as his rather than adopted. On his account competence is judged by the outcome while the outcome is trusted according to the competence. In a settled field this is invisible, because everyone already agrees what a working instrument produces. It becomes visible precisely when a field is not settled, which is when it matters.

It does not follow that the circle cannot be broken, and this article does not claim that it cannot. Franklin, who reports the passage, argues against exactly that conclusion and holds that "the regress can be broken by reasoned argument" [12]. The two positions are recorded here and neither is adopted, and the reader should note that they are not symmetrical: one says a difficulty exists, the other says it is surmountable, and both can be true.

The scope limit fixed earlier in this series is preserved here. The regress is a real problem in some contested cases, not a general law that every calibration is circular. A thermometer checked against a national standard is not in a regress, and describing it as though it were would empty the concept of the content that makes it useful.

Where That Leaves Us

A well run experiment whose result is contested is the hard case, because competence cannot then be judged independently of the result. Collins states that circle and Franklin, who quotes him, holds it can be broken by reasoned argument. Both are recorded and neither adopted, and they are not symmetrical. The regress belongs to contested cases and not to calibration in general.

7. Reading an experiment well

Was anything assigned, or only observed? If nobody assigned, the reason people are in the groups they are in has to be argued rather than known.

Who knew what, and when? Sequence generation and allocation concealment are different failures. So are concealment and blinding, and so are blinding of participants and blinding of outcome assessors.

Which protections ended at assignment? Concealment is available throughout the randomised trials the handbook describes, and stops at assignment. Blinding is not always available, and where it is, it has to hold for the people and stages at which knowing the assignment could bias what happens or what is recorded.

What is the comparison group, and what makes it comparable? In a randomised design the answer is procedural. In a natural experiment it is an argument about the world, and the argument should be visible.

Is the result contested? If it is, expect judgements about competence and judgements about the answer to move together, and look for the reasons offered rather than for a decisive experiment.

What was held fixed, and was that the same thing as having a control group? They are separate, and a study can have one without the other.

An experiment is not a device that produces truth. It is an arrangement that replaces one unknown, why these people and not those, with a procedure, and then spends the rest of its length defending that substitution.

Where That Leaves Us

The article ends in questions a reader can carry to an experiment: whether anything was assigned, who knew what and when, which protections stopped at assignment, what makes the comparison group comparable, whether the result is contested, and whether what was held fixed is the same thing as a control group. An arrangement, not a device producing truth. What is established, and how, follows.

What We Know, and How

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 We Still Don't Know

Whether sequences were correctly generated in the trials whose concealment was judged inadequate, since the Schulz comparison sorts on concealment and not on generation. How far the exaggeration figures travel beyond the set of trials examined, being an association between a reported feature and reported effect sizes rather than a distance from a known truth. Whether a particular natural experiment's variation really is as if randomly assigned, which is an assumption about the world and not a fact about the design. And whether the regress can be broken in a given contested case, on which Collins and Franklin are recorded here and neither is adopted.

What Could Change Our Minds

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.

If You Remember Only One Thing

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.

Related Articles

- KNO-014 · *Causality and Correlation*

References

The bracketed numbers in the body are claim numbers, not reference numbers. Each resolves to a row of the claims ledger, which names the source below that supports it.

Comité del Premio en Ciencias Económicas. 2021. *Answering Causal Questions Using Observational Data. Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021*. The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel, The Royal Swedish Academy of Sciences, 11 de octubre de 2021. URL estable: <https://www.nobelprize.org/uploads/2021/10/advanced-economicsciencesprize2021.pdf>

Fisher, R. A. 1935. *The Design of Experiments*. Edimburgo y Londres: Oliver and Boyd. **Sin ISBN: la obra es anterior al sistema ISBN**. URL estable del ejemplar digitalizado consultado: <https://archive.org/details/in.ernet.dli.2015.502684>

Franklin, Allan. 1994. «How to Avoid the Experimenters' Regress». *Studies in History and Philosophy of Science* 25 (3): 463–491. **Sin DOI mostrado sobre página consultada; identificador impreso 0039-3681/94.** URL estable: https://physics.yale.edu/sites/default/files/files/3_1%20franklin.pdf

Higgins, Julian PT, Jelena Savović, Matthew J Page, Roy G Elbers, y Jonathan AC Sterne. 2024. «Chapter 8: Assessing risk of bias in a randomized trial». En *Cochrane Handbook for Systematic Reviews of Interventions*, versión 6.5. Cochrane. URL estable: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-08>

Schulz, Kenneth F., Douglas G. Altman, y David Moher, for the CONSORT Group. 2010. «CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials». *PLoS Medicine* 7 (3): e1000251. DOI 10.1371/journal.pmed.1000251

Schulz, Kenneth F., Iain Chalmers, Richard J. Hayes, y Douglas G. Altman. 1995. «Empirical Evidence of Bias: Dimensions of Methodological Quality Associated With Estimates of Treatment Effects in Controlled Trials». *JAMA* 273 (5): 408–412. DOI 10.1001/jama.1995.03520290060030 URL estable: https://www.jameslindlibrary.org/wp-data/uploads/2012/02/Schulz_Chalmers_Hayes_Altman_1995.pdf

KNO-013 · Models, Maps and Reality

KNO · Ways of Knowing

Juan Segurola

v0.1 · EN

All Rights Reserved

Article Abstract

A scientific model is not a replica. It stands in for a selected part or aspect of the world, and the selecting is the work. This article treats omission and deliberate distortion as working parts rather than defects, and separates an error, which a modeller would remove, from an idealisation, which a modeller would defend. It records one published position holding that the laws which unify best describe least, without adopting it. It states the constraint a population biologist formulated in 1966 and the objection that denies the constraint is necessary. It shows that the most quoted sentence about models is not in the paper usually cited. It closes on the realism question two earlier articles handed forward, states both standards and adopts neither.

In 60 Seconds

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 This Question 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.

1. A model is not a small copy

The word model suggests a scale replica, and the suggestion is misleading in a specific way. A replica aims to preserve everything at reduced size. A scientific model aims at something narrower: it stands in for a chosen part of the world, and the choosing is the work.

The standard formulation puts the target first. Many scientific models are representational models, and they "represent a selected part or aspect of the world, which is the model's target system" [1]. **The phrase to hold on to is selected part or aspect.** A model does not represent the world; it represents a piece of it, picked out in advance, and the question of whether a model is good is always the question of whether it is good for that piece and that purpose.

A warning about the map analogy, which this article uses sparingly and on purpose. Maps are the standing illustration of selective representation, and an earlier article in this series treats them directly. But the constraint that binds maps most tightly is a geometric theorem about projecting a sphere onto a plane, and **no corresponding theorem about models in general is established here or transported from KNO-006.** Borrowing the map's force without borrowing its mathematics would produce an impressive sentence and an unsupported one. This article therefore uses the map as an illustration of selection and never as evidence about what models can or cannot do.

Where That Leaves Us

The word model suggests a replica, and that is the misleading part. A scientific model stands in for a chosen piece of the world, and choosing it is the work, so whether it is good is whether it is good for that piece and that purpose. Maps illustrate the choosing and are never evidence here, because their own constraint is a theorem this article does not carry over.

2. Leaving things out is not the same as getting them wrong

Models are not one kind of thing, and the distinctions matter for what follows. A **representational** model stands in for a target system. An **abstraction** leaves features out without asserting that they are absent. An **idealisation** deliberately simplifies or distorts what it keeps. Only the third asserts something false on purpose, and if a model only omitted, it would be an abridgement rather than a distortion.

The distinction that matters is between an error and an idealisation. **An error is something the modeller would remove on discovering it. An idealisation is something the modeller put there and would defend.** The reference formulation is explicit about the intent: idealised models "involve a deliberate simplification or distortion of something complicated with the objective of making it more tractable or understandable" [2].

That definition contains a criterion that can be applied. **An idealisation is answerable to a purpose.** A frictionless plane, a well mixed population, a perfectly rigid body: each is false, each is chosen, and each can be asked whether it earns its falsehood in the use to which the model is put. A distortion that no longer earns it has become an error without anyone changing the equations.

Where That Leaves Us

Three things separate here. A representational model stands in for a target; an abstraction omits without claiming absence; an idealisation distorts knowingly, and only it asserts a falsehood on purpose. The test is about intent rather than accuracy: what a modeller would strike out on noticing is an error, and what they would defend has to earn its falsehood against the use the model is put to.

3. A published argument that goes further

There is a stronger position in the literature, and it should be stated as one position rather than as the consensus.

Cartwright argues that fundamental explanatory laws, the deepest and most admired successes of modern physics, "do not in fact describe regularities that exist in nature" [3]. The argument is not that physics fails. It is that explanatory power and descriptive accuracy are obtained by different means and are traded against each other, so the laws that unify most are the ones that describe least.

The same argument carries a distinction that keeps it from collapsing into general scepticism: on this account "theoretical entities, and the complex and localized laws that describe them, can be interpreted realistically, but the simple unifying laws of basic theory cannot" [4]. **The scepticism is aimed at the top of the hierarchy, not at the bottom.**

This article does not adopt that position. It records it because it is a developed argument in the literature and because it sharpens the question section 6 has to answer. A reader should note that it is contested, and that the sentences quoted here come from the publisher's description of the book rather than from a reading of its argument, a limitation stated in the uncertainty record.

Where That Leaves Us

A stronger position exists in the literature and this article records rather than adopts it. On that account the deepest explanatory laws do not describe regularities, because reach and accuracy are bought by different means, and its own distinction keeps it from general scepticism: localised laws and the entities they describe may be read realistically, the unifying ones may not. The quoted sentences come from a publisher's description.

4. The trade you cannot design around

One of the most useful results about modelling is a constraint, and it was stated plainly in 1966. **It is a position in a live debate rather than a settled theorem**, and the objection to it is stated below rather than left implicit.

Levins wrote that "it is of course desirable to work with manageable models which maximize generality, realism, and precision toward the overlapping but not identical goals of understanding, predicting, and modifying nature. But this cannot be done" [5].

Read the sentence carefully, because its structure is often lost. **It names three properties, attaches them to three different goals, and says the three cannot be maximised together in a manageable model.** It does not say that any one of them is unattainable, and it does not rank them. **The word realism here is not the realism of section 6:** in Levins's triad it means how much of a system's detail a model keeps, while in section 6 it is a position about whether the unobservable content of a theory should be believed. The two are unrelated and this article does not let one carry the other. What it says is that within a manageable model the three cannot be maximised together, so that where one is pushed the others are under pressure, and that the trade is a decision whether or not anyone notices making it. **It is Levins's claim about model building as he framed it, and this article does not extend it into a law that every model must obey.**

The objection is direct, and it is not a quibble about wording. Orzack and Sober reject the claim that generality, realism and precision must necessarily trade off against one another, so that the conflict Levins describes is not a universal constraint on model building [12]. Read it as narrowly as it is written: it denies that the trade is necessary, not that a trade ever occurs. A modeller who finds one of the three under pressure has not been refuted, and one who finds all three improving together has not broken a law. **What the objection removes is the entitlement to treat the constraint as settled**, which is what this section was already refusing to claim.

This is the practical content of everything above. Selection, idealisation and the gap between explaining and describing all show up here as a single budget. A model that is very general is unlikely to be very precise about a particular case. A model tuned to one case in detail is unlikely to travel. **Asking which of the three a model gave up is usually more informative than asking whether the model is right.**

Where That Leaves Us

The constraint is that generality, retained detail and precision cannot be maximised together in a manageable model, and detail here is not the doctrine section 6 discusses. The objection recorded beside it denies that the trade is necessary, not that trades happen, so neither pressure on one property nor improvement in all three settles anything. What survives is which was spent, and on what.

5. Wrong, and useful, in two different documents

One sentence about models is quoted far more than the rest of the literature on it, and it is often attributed to the wrong document. **How often is not established here:** no bibliometric study was consulted, and the point of the section does not need one.

The 1976 paper is the one usually cited. What it actually contains is this: "Since all models are wrong the scientist cannot obtain a 'correct' one by excessive elaboration. On the contrary following William of Occam he should seek an economical description of natural phenomena" [6]. That is a complete argument on its own, and it is not the famous sentence. **A direct check of that paper confirms that the phrase but some are useful does not appear anywhere in it** [7].

The famous wording is in a technical report from 1979, where the literal string "ALL MODELS ARE WRONG BUT SOME ARE USEFUL" appears [8], accompanied by the observation that "it would be very remarkable if any system existing in the real world could be exactly represented by any simple model" [9].

The two texts say compatible things and they are not the same text. The 1976 passage is an argument against over elaboration: since no model is correct, adding parameters cannot make one correct, so economy is the right response. The 1979 wording is an argument about the question to ask: not whether the model is true, but whether it is useful. Citing the second sentence to the first paper is a small error with a general moral, and this article states it because the moral is the article's own: **a formulation that circulates on its own is the kind least likely to be checked against its source.**

Where That Leaves Us

One sentence circulates more widely than the literature around it and is commonly credited to the wrong document. The 1976 paper argues that elaboration cannot rescue a model, so economy is the answer, and a direct check finds the famous phrase nowhere in it; the wording belongs to a report from 1979. The moral is this article's own: what travels alone is what nobody checks.

6. The question two earlier articles handed forward

Two earlier articles in this series stopped at the same doorway. One asked whether a classification scheme corresponds to a real division of the world and declined to answer. The other asked whether a measured number corresponds to something real and declined likewise. Both handed the question here, and it is the same question in different clothes.

The two positions should be attributed before any claim of neutrality is made.

Scientific realism, as the reference formulation states it, "is a positive epistemic attitude toward the content of our best theories and models, recommending belief in both observable and unobservable aspects of the world described by the sciences" [10]. On this view the unobservable content of a successful theory is a candidate for belief, and a model that works is evidence that its posits are approximately right.

The rival standard is empirical adequacy, associated with constructive empiricism and given in the same reference through van Fraassen's formulation: "a theory is empirically adequate exactly if what it says about the observable things and events in the world, is true" [11]. On this view a theory that saves the observable phenomena has done all that is required, and belief in its unobservable machinery is optional. **The criterion is not the same as fitting a finite body of data.** As formulated it ranges over what a theory says about the observable, including the observable it has not been tested against, so a good fit to the data in hand is evidence bearing on empirical adequacy rather than the thing itself.

Having stated both, this article adopts neither, and the reason is not diplomatic. The disagreement is about what success licenses, and nothing in the practice of modelling settles it: the same model, performing identically, is a reason for belief on one account and a well functioning instrument on the other. **What the disagreement does settle is a procedural point that both sides accept.** Whichever standard is in force, that a model fits the observable record is itself a finding that has to be established, and it depends on the data used, on how those data were treated, on the rules of comparison, on the assumptions carried into the fitting and on the uncertainty attached to the result. **Fit does not accredit itself.** Any further claim about what the model's parts correspond to is then a separate step again, and has to be argued separately. **The articles that handed this question forward can therefore proceed, and what they get is narrower than an answer.** The realism question is not resolved here and this article does not resolve it. What is established is that a classification and a measurement can each be assessed **operationally**, by what it does and what it was built for, without that assessment waiting on the realism question. The question itself stays open and stays handed forward.

Where That Leaves Us

Two standards are set out and neither is adopted, because the disagreement is about what success licenses and no amount of modelling practice decides it. What both sides grant is procedural. That a model matches the observable record is itself a result to be established, resting on the data, their handling, the comparison rules, the assumptions and the uncertainty. Correspondence is a further step, argued separately.

7. Reading a model well

What is the target? A model represents a selected part or aspect, and criticising it for the parts it never claimed is a category error.

Which distortions are deliberate? Ask which falsehoods the modeller would defend and which they would fix. The second kind are errors; the first kind have to earn their place.

What was traded away? On Levins's account the three cannot be maximised together in a manageable model. Identify which was spent and on what, and treat the constraint as the position it is.

Is the claim about explaining or about describing? At least one developed position holds that the laws that unify best are not the ones that describe best. Whether or not that is accepted, the two virtues are worth distinguishing.

Does the argument need the unobservables to be real? Sometimes it does and sometimes it does not, and an argument that works either way is stronger than one that needs the question settled.

Is a memorable formulation doing load bearing work? If so, check it in the source rather than in the citation, because the most repeated sentences are the least examined.

A model is a deliberate misdescription with a job. Judging it means asking what job, what was left out, what was distorted on purpose, and what was given up to make it usable.

Where That Leaves Us

The article ends in questions to carry to a model: what it targets, which distortions are deliberate, what was given up, whether the claim concerns explaining or describing, whether the argument needs its unobservables to be real, and whether a memorable formulation is bearing weight. A deliberate misdescription with a job to do. What is established, and how, follows.

What We Know, and How

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 We Still Don't Know

Whether the unobservable parts of a working model correspond to anything real, which this article states both standards for and settles neither. Whether Cartwright's position survives a reading of the argument itself, since what is reported here comes from a publisher's description. How often the famous sentence is misattributed, which no bibliometric study was consulted to establish. And whether, in any particular case, generality, retained detail and precision do trade against one another, which the objection leaves open by denying only that they must.

What Could Change Our Minds

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.

If You Remember Only One Thing

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.

Related Articles

- KNO-006 · *Maps and Spatial Representation*

References

The bracketed numbers in the body are claim numbers, not reference numbers. Each resolves to a row of the claims ledger, which names the source below that supports it.

Box, G. E. P. 1979. *Robustness in the Strategy of Scientific Model Building*. MRC Technical Summary Report #1954. Mathematics Research Center, University of Wisconsin–Madison, mayo de 1979. **Sin DOI ni ISBN: informe técnico.** URL estable del ejemplar digitalizado consultado: https://archive.org/details/DTIC_ADA070213

Box, George E. P. 1976. «Science and Statistics». *Journal of the American Statistical Association* 71 (356): 791–799. **Sin DOI localizado sobre página consultada.** URL estable del ejemplar consultado: <https://www-sop.inria.fr/members/Ian.Jermyn/philosophy/writings/Boxonmaths.pdf>

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford University Press. **Manifestación en línea**, ISBN 9780191597152. DOI 10.1093/0198247044.001.0001

Chakravartty, Anjan. «Scientific Realism». *Stanford Encyclopedia of Philosophy*. Publicada por primera vez el 27 de abril de 2011; revisión sustantiva del 12 de junio de 2017. URL estable de la versión archivada consultada: <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>

Frigg, Roman, y Stephan Hartmann. «Models in Science». *Stanford Encyclopedia of Philosophy*. Publicada por primera vez el 27 de febrero de 2006; revisión sustantiva del 2 de abril de 2025. URL estable de la versión archivada consultada: <https://plato.stanford.edu/archives/sum2025/entries/models-science/>

Levins, Richard. 1966. «The Strategy of Model Building in Population Biology». *American Scientist* 54 (4): 421–431. **Sin DOI: la obra es anterior al sistema DOI y ninguna página consultada muestra uno asignado con posterioridad.** URL estable indicada por el propio ejemplar consultado: <https://www.jstor.org/stable/27836590>

Orzack, Steven Hecht, y Elliott Sober. 1993. «A Critical Assessment of Levins's "The Strategy of Model Building in Population Biology" (1966)». *The Quarterly Review of Biology* 68 (4): 533–546. DOI 10.1086/418301

KNO-014 · Causality and Correlation

KNO · Ways of Knowing

Juan Segurola

v0.7 · EN

All Rights Reserved

Ways of Knowing

Epistemic profile: Current scientific interpretation **Licence at publication:** All Rights Reserved

Article Abstract

When two variables move together, what additional evidence is needed before the pattern can support a claim that changing one would change the other? This article examines that question through temporal order, confounding, reverse causation, randomisation, causal diagrams, adjustment, propensity scores, instrumental variables, negative controls and target-trial emulation. It asks how each method changes the inferential problem, which assumptions remain hidden when an association is treated as self-explanatory, and why different designs identify different causal contrasts. It also asks where observational evidence can be genuinely causal without pretending to reproduce a randomised experiment, and where formal methods merely make assumptions more explicit. The goal is to separate the observable pattern from the intervention question and to show what a defensible causal claim must make inspectable.

In 60 Seconds

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 This Question 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.

1. Association Is a Pattern; Causation Is a Counterfactual Question

A statistical association describes how variables co-vary in observed data. A causal effect compares outcomes under different possible exposure or treatment conditions. Hernán's formal epidemiologic presentation makes the distinction explicit: causal effects are defined through contrasts between potential outcomes, while observed associations can differ from those causal contrasts because the groups receiving different exposures may already differ in relevant ways [2]. Holland's account of the Rubin causal model makes the same separation central to statistical reasoning about causation [3].

The counterfactual formulation also exposes a fundamental obstacle. For the same unit at the same time, we do not simultaneously observe the outcome both under treatment and under no treatment. The missing counterfactual is not a technical nuisance that a larger dataset automatically fills. Causal inference therefore depends on comparisons across units, times or regimes together with assumptions that make those comparisons informative.

Claim KNO014-C01. An observed association and a causal effect are different quantities; causal effects require a contrast between potential outcomes or intervention regimes, not merely a covariance or regression coefficient. [2,3]

Claim KNO014-C02. The fundamental causal problem is that, for one unit, only one of the relevant potential outcomes is observed. [3]

Where That Leaves Us

A statistical association describes how variables co-vary in observed data. Hernán's formal epidemiologic presentation makes the distinction explicit: causal effects are defined through contrasts between potential outcomes, while observed associations can differ from those causal contrasts because the groups receiving different exposures may already differ in relevant ways [2]. The next section asks why temporal order helps but does not settle the question.

2. Why Temporal Order Helps but Does Not Settle the Question

A cause must precede its effect in the causal relation under study. This makes temporal ordering an essential constraint. If an alleged cause occurs only after the outcome, that claim fails for that temporal relation.

But precedence is not enough. An exposure may precede an outcome while both are driven by an earlier common cause. A biomarker may change before diagnosis while already reflecting an occult disease process. A policy may be adopted before an economic change because governments anticipate that change. Temporal order removes some causal stories, not all competing stories.

Bradford Hill treated temporality as one consideration in evaluating whether an observed association might be causal, while also warning against treating his considerations as rigid criteria that mechanically prove causation [1]. His framework remains historically important precisely because it addressed inference from imperfect observational evidence, not because it supplied a universal checklist.

Claim KNO014-C03. Temporal precedence is necessary for the cause to precede the effect in the causal ordering, but temporal order alone is not sufficient to identify a causal effect. [2]

Where That Leaves Us

A cause must precede its effect in the causal relation under study. If an alleged cause occurs only after the outcome, that claim fails for that temporal relation. But precedence is not enough. An exposure may precede an outcome while both are driven by an earlier common cause. The next section examines confounding.

3. Confounding: When a Common Cause Produces the Association

Suppose smoking is associated with an outcome, but age influences both smoking history and the outcome. If age is not appropriately handled, some of the observed association between smoking and outcome may reflect age rather than the causal effect of smoking.

This is confounding in its simplest form: a common cause of exposure and outcome opens a non-causal path between them. Modern causal diagrams make this structure explicit. Greenland, Pearl and Robins showed how directed acyclic graphs can encode assumed causal relations and help identify variables that must be controlled to estimate an unconfounded effect [4].

The important word is assumed. A graph is not discovered merely by fitting the data. Its arrows represent causal commitments supplied by scientific knowledge, design information and temporal constraints. Data can contradict some consequences of a graph, but a compatible graph is not thereby proven true.

Claim KNO014-C04. Confounding can produce an exposure-outcome association even when part or all of that association is not the causal effect of the exposure. [4]

Claim KNO014-C05. Causal diagrams can make adjustment assumptions explicit, but the causal structure encoded by their arrows requires substantive assumptions beyond the observed joint distribution. [4,5]

Where That Leaves Us

Suppose smoking is associated with an outcome, but age influences both smoking history and the outcome. If age is not appropriately handled, some of the observed association between smoking and outcome may reflect age rather than the causal effect of smoking. Data can contradict some consequences of a graph, but a compatible graph is not thereby proven true. The next section examines reverse causation.

4. Reverse Causation: When the Arrow Points the Other Way

An association between X and Y does not tell us whether X causes Y, Y causes X, both occur, or a third structure explains the dependence. Reverse causation is especially important when disease processes begin before diagnosis. Weight loss may predict later illness because preclinical illness caused the weight loss, not because weight loss caused the illness.

Longitudinal data can help because they establish sequence, but they do not automatically eliminate reverse causation. The relevant causal process may begin before the first recorded measurement. The correct response is to define time zero carefully, use knowledge of latency and disease history, and test whether plausible lag structures alter the inference.

Claim KNO014-C06. Longitudinal ordering can constrain reverse-causation explanations but cannot by itself exclude causal processes that began before measurement. [2,7]

Where That Leaves Us

An association between X and Y does not tell us whether X causes Y, Y causes X, both occur, or a third structure explains the dependence. Weight loss may predict later illness because preclinical illness caused the weight loss, not because weight loss caused the illness. The next section examines randomisation.

5. Randomisation: Why an Experiment Can Change the Evidential Problem

In an ideal randomised experiment, treatment assignment is generated independently of baseline causes of the outcome. With adequate sample size, concealment, adherence to the protocol and valid outcome measurement, randomisation can make treatment groups comparable in expectation and support estimation of causal effects without the same confounding assumptions required in ordinary observational comparisons [2].

This does not mean every randomised trial is unbiased. Loss to follow-up, non-adherence, outcome misclassification, post-randomisation selection and analysis choices can still create problems. Nor does randomisation answer every causal question: some interventions are unethical, impossible, too rare or too long-term to randomise.

The key epistemic point is narrower. Randomisation changes the assignment mechanism. It does not make causality visible; it creates a design in which a specific causal contrast can be identified under a smaller and more transparent set of assumptions.

Claim KNO014-C07. Proper randomisation can identify causal effects by breaking systematic baseline causes of treatment assignment, but trial validity still depends on implementation, follow-up, measurement and analysis. [2]

Where That Leaves Us

In an ideal randomised experiment, treatment assignment is generated independently of baseline causes of the outcome. This does not mean every randomised trial is unbiased. Loss to follow-up, non-adherence, outcome misclassification, post-randomisation selection and analysis choices can still create problems. The next section examines the proposition that adjustment is not “control for everything”.

6. Adjustment Is Not “Control for Everything”

A common intuition says that observational studies become more causal if they adjust for more variables. Causal diagrams show why this is false.

If a variable is a confounder, adjustment may block a non-causal backdoor path. If a variable lies on the causal pathway, adjusting for it may remove part of the total effect one intended to estimate. If a variable is a collider, meaning it is caused by two variables on a path, conditioning on it can open an association that was absent before conditioning. In such cases, “controlling for” a variable manufactures bias.

Greenland, Pearl and Robins formalised how graph structure guides adjustment [4]. Later methodological work has emphasised that collider control, mediator control and indiscriminate pre-treatment adjustment can each distort causal effect estimates in the wrong structure [6].

Claim KNO014-C08. More statistical adjustment is not monotonically better: adjusting for a confounder may reduce bias, while adjusting for a mediator or collider can alter or introduce bias depending on the estimand and causal structure. [4,6]

Where That Leaves Us

A common intuition says that observational studies become more causal if they adjust for more variables. If a variable is a confounder, adjustment may block a non-causal backdoor path. If a variable lies on the causal pathway, adjusting for it may remove part of the total effect one intended to estimate. The next section examines the proposition that potential outcomes and causal diagrams are complementary languages.

7. Potential Outcomes and Causal Diagrams Are Complementary Languages

Two major modern traditions describe causal inference using different but overlapping languages. Potential-outcome frameworks define effects by comparing outcomes under alternative treatment states. Structural causal models and DAGs represent causal relations and intervention structure graphically and algebraically.

These are not simply rival vocabularies where one must be chosen and the other rejected. In many practical analyses they illuminate different parts of the same problem. Potential outcomes force precision about the causal contrast and target population. DAGs force precision about assumed pathways, confounding, mediation and selection. Both make assumptions more visible than a regression coefficient presented without a causal model.

Pearl emphasised that causal claims require assumptions not contained in ordinary probability distributions and that causal languages are needed to express those assumptions [5]. Hernán and Robins similarly organise causal inference around explicitly specified treatment strategies and counterfactual outcomes [8].

Claim KNO014-C09. Potential-outcome and graphical approaches provide complementary ways to state causal estimands and assumptions; neither reduces causal inference to ordinary association alone. [5,8]

Where That Leaves Us

Two major modern traditions describe causal inference using different but overlapping languages. These are not simply rival vocabularies where one must be chosen and the other rejected. Structural causal models and DAGs represent causal relations and intervention structure graphically and algebraically. Hernán and Robins similarly organise causal inference around explicitly specified treatment strategies and counterfactual outcomes [8]. The next section examines propensity scores.

8. Propensity Scores: Balance Measured Covariates, Not Unknown Reality

Rosenbaum and Rubin introduced the propensity score as the probability of treatment assignment conditional on observed covariates [9]. Under their assumptions, balancing on the propensity score can balance those observed covariates between treatment groups and support causal effect estimation.

But a propensity score does not automatically solve unmeasured confounding. If an important common cause is absent from the measured covariates, balancing the measured variables does not balance the missing cause. The method is therefore powerful within its identification conditions and powerless against violations it cannot see.

Claim KNO014-C10. Propensity-score methods can balance observed covariates under specified assumptions but do not by themselves eliminate bias from unmeasured confounding. [9]

Where That Leaves Us

Rosenbaum and Rubin introduced the propensity score as the probability of treatment assignment conditional on observed covariates [9]. But a propensity score does not automatically solve unmeasured confounding. If an important common cause is absent from the measured covariates, balancing the measured variables does not balance the missing cause. The next section examines instrumental variables.

9. Instrumental Variables: A Different Route with Different Assumptions

Sometimes researchers use an instrumental variable: a variable that shifts treatment or exposure but, under the identification model, affects the outcome only through that treatment and satisfies additional independence and relevance conditions.

Angrist, Imbens and Rubin showed how instrumental-variable reasoning can identify a local average treatment effect for compliers under explicit assumptions, including monotonicity in their framework [10]. This result illustrates a recurring lesson: methods do not produce causal effects in the abstract. They identify particular estimands for particular populations under particular assumptions.

Claim KNO014-C11. Instrumental-variable methods can identify causal effects without ordinary no-unmeasured-confounding assumptions, but only by substituting other strong identification assumptions and often a more specific estimand. [10]

Where That Leaves Us

Sometimes researchers use an instrumental variable: a variable that shifts treatment or exposure but, under the identification model, affects the outcome only through that treatment and satisfies additional independence and relevance conditions. This result illustrates a recurring lesson: methods do not produce causal effects in the abstract. The next section examines negative controls.

10. Negative Controls: Look for Bias Where No Effect Should Exist

A negative-control exposure or outcome is chosen so that, under the causal hypothesis being tested, the target causal effect should be absent. If an association nevertheless appears, it can reveal residual confounding, selection bias or other systematic error.

Lipsitch, Tchetgen Tchetgen and Cohen proposed negative controls as a routine diagnostic for observational causal studies [11]. A negative control is not a universal falsifier. Its interpretation depends on why the control is expected to have no causal effect and on whether it shares the bias mechanism one is trying to detect. Used carefully, however, it turns background assumptions into an empirical stress test.

Claim KNO014-C12. Negative controls can detect some forms of residual confounding or bias when the control is causally null for the target relation but shares relevant bias pathways. [11]

Where That Leaves Us

A negative-control exposure or outcome is chosen so that, under the causal hypothesis being tested, the target causal effect should be absent. A negative control is not a universal falsifier. Lipsitch, Tchetgen Tchetgen and Cohen proposed negative controls as a routine diagnostic for observational causal studies [11]. The next section examines target trials.

11. Target Trials: Design the Observational Question Before Analysing the Data

Many observational studies are implicitly trying to answer the question a randomised trial would have answered if that trial were feasible. Hernán and Robins proposed making that hypothetical trial explicit: eligibility criteria, treatment strategies, assignment procedure, time zero, follow-up, outcome, causal contrast and analysis plan are specified first, and the observational analysis is then designed to emulate those elements as closely as possible [7].

This approach can reveal avoidable design defects. If eligibility is assessed at one time, treatment later, and follow-up begins at a third time, the analysis can introduce immortal-time or selection biases. Target-trial thinking therefore shifts causal inference upstream: design precedes modelling.

Claim KNO014-C13. Explicit target-trial emulation can expose timing and design mismatches in observational analyses before statistical modelling begins. [7,12]

Where That Leaves Us

Many observational studies are implicitly trying to answer the question a randomised trial would have answered if that trial were feasible. Target-trial thinking therefore shifts causal inference upstream: design precedes modelling. If eligibility is assessed at one time, treatment later, and follow-up begins at a third time, the analysis can introduce immortal-time or selection biases. The next section revisits Bradford Hill's evidence-integration framework.

12. Bradford Hill Revisited: Evidence Integration, Not a Causal Scorecard

Hill's 1965 discussion remains influential because it addressed how to reason when controlled experiments are unavailable. Strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment and analogy were offered as viewpoints to consider, not as a validated additive score and not as necessary-and-sufficient conditions [1].

Modern causal inference adds formal tools that Hill did not have: counterfactual estimands, DAGs, g-methods, propensity scores, instrumental-variable identification, negative controls and target-trial emulation. Yet Hill's broader lesson remains compatible with them: causal judgment integrates multiple kinds of evidence while resisting automatic inference from association alone.

The important update is methodological. A strong association can still be confounded. A weak association can still represent an important causal effect. Biological plausibility evolves with background knowledge. Consistency across studies can reflect shared bias. The evidential value of each feature depends on the causal question and the design that generated the evidence.

Claim KNO014-C14. Bradford Hill's considerations are aids to causal reasoning, not a mechanical checklist that proves or disproves causation. [1]

Where That Leaves Us

Hill's 1965 discussion remains influential because it addressed how to reason when controlled experiments are unavailable. Strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment and analogy were offered as viewpoints to consider, not as a validated additive score and not as necessary-and-sufficient conditions [1]. The next section asks what data alone cannot tell you.

13. What Data Alone Cannot Tell You

A sufficiently rich dataset can estimate many conditional associations with great precision. Precision does not identify the causal structure that generated them. Distinct causal models may imply the same observed distribution. The difference between those models appears only when we ask intervention or counterfactual questions.

This is why causal inference cannot be reduced to "use a more powerful model". Machine learning may estimate prediction functions exceptionally well while leaving the causal estimand unidentified. Conversely, a simple analysis can support a strong causal claim when design and assumptions make the contrast identifiable.

Claim KNO014-C15. Predictive accuracy and causal identification are distinct goals; a model can predict well without establishing the effect of an intervention. [5]

Where That Leaves Us

A sufficiently rich dataset can estimate many conditional associations with great precision. Precision does not identify the causal structure that generated them. Distinct causal models may imply the same observed distribution. The difference between those models appears only when we ask intervention or counterfactual questions. The next section turns these points into a practical sequence for analysis.

14. A Practical Causal-Inference Sequence

A disciplined analysis can be organised as a sequence rather than a single statistical test:

1. State the causal question and target population.
2. Define the exposure or intervention and the comparison regime.
3. Define the outcome and time horizon.
4. Specify time zero and the ordering of relevant events.
5. Draw or otherwise state the assumed causal structure.
6. Decide which variables are confounders, mediators, colliders or consequences of selection for the estimand at hand.
7. Choose a design and estimator whose identification assumptions match that structure.
8. Check balance, overlap, measurement quality and missingness.
9. Use negative controls, alternative specifications or sensitivity analyses where appropriate.
10. Interpret the estimate as causal only to the extent that the identification and measurement assumptions remain defensible.

This sequence is not a guarantee. It is a way of making the hidden premises of causal claims inspectable.

Claim KNO014-C16. Causal analysis is best treated as a design-and-assumption problem before it is treated as an estimation problem. [4,7,8]

Where That Leaves Us

Step by step, a disciplined analysis defines the question, contrast, outcome and time zero; states the assumed causal structure and each variable's role; fits design and estimator to that structure; checks balance, overlap and measurement; probes the estimate; and reads it as causal only insofar as its assumptions stay defensible. The analysis guarantees nothing but makes hidden premises inspectable. The next section defines the boundary of KNO-014.

15. The Boundary of KNO-014

KNO-014 is about the **methodological problem of distinguishing causal effects from non-causal associations**. It does not attempt to settle the metaphysics of causation. That conceptual question belongs to IDE-004 · Causality.

It also does not replace KNO-012 · Experiments and Variable Control. KNO-012 asks how experiments create controlled comparisons; KNO-014 asks how causal claims are identified across experimental and observational settings.

KNO-015 · Uncertainty and Probability will take up a different problem: how uncertainty is represented and propagated once the target quantity and inferential structure have been specified. Sampling variability as a competing explanation for an observed association is also handed to KNO-015 and is not developed here.

Where That Leaves Us

KNO-014 is about the methodological problem of distinguishing causal effects from non-causal associations. It does not attempt to settle the metaphysics of causation. That conceptual question belongs to IDE-004 · Causality. This defines the article's boundary and prepares the final synthesis without extending the claim beyond what the preceding sections establish.

What We Know, and How

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 Still 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 Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-012 · Experiments and Variable Control
- KNO-013 · Models, Maps and Reality
- KNO-015 · Uncertainty and Probability
- KNO-016 · How Evidence Is Built: Putting the Tools Together

References

- [1] Hill, Austin Bradford. "The Environment and Disease: Association or Causation?" *Proceedings of the Royal Society of Medicine* 58(5), 1965, 295–300. DOI: 10.1177/003591576505800503. PMID: 14283879.
- [2] Hernán, M. A. "A Definition of Causal Effect for Epidemiological Research." *Journal of Epidemiology and Community Health* 58(4), 2004, 265–271. DOI: 10.1136/jech.2002.006361. PMID: 15026432.
- [3] Holland, Paul W. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81(396), 1986, 945–960. DOI: 10.1080/01621459.1986.10478354.
- [4] Greenland, Sander; Pearl, Judea; Robins, James M. "Causal Diagrams for Epidemiologic Research." *Epidemiology* 10(1), 1999, 37–48. DOI: 10.1097/00001648-199901000-00008. PMID: 9888278.
- [5] Pearl, Judea. "An Introduction to Causal Inference." *The International Journal of Biostatistics* 6(2), 2010, Article 7. DOI: 10.2202/1557-4679.1203. PMID: 20305706.
- [6] Tennant, Peter W. G.; Murray, Eleanor J.; Arnold, Kellyn F.; Berrie, Laurie; Fox, Matthew P.; Gadd, Sarah C.; Harrison, Wendy J.; Keeble, Claire; Ranker, Lynsie R.; Textor, Johannes; Tomova, Georgia D.; Gilthorpe, Mark S.; Ellison, George T. H. "Use of Directed Acyclic Graphs (DAGs) to Identify Confounders in Applied Health Research: Review and Recommendations." *International Journal of Epidemiology* 50(2), 2021, 620–632. DOI: 10.1093/ije/dyaa213.
- [7] Hernán, Miguel A.; Robins, James M. "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available." *American Journal of Epidemiology* 183(8), 2016, 758–764. DOI: 10.1093/aje/kwv254. PMID: 26994063.
- [8] Hernán, Miguel A.; Robins, James M. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020. Suggested citation of the authors; online text consulted separately as a revisable electronic manifestation.
- [9] Rosenbaum, Paul R.; Rubin, Donald B. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1), 1983, 41–55. DOI: 10.1093/biomet/70.1.41.
- [10] Angrist, Joshua D.; Imbens, Guido W.; Rubin, Donald B. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91(434), 1996, 444–455. DOI: 10.1080/01621459.1996.10476902.
- [11] Lipsitch, Marc; Tchetgen Tchetgen, Eric; Cohen, Ted. "Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies." *Epidemiology* 21(3), 2010, 383–388. DOI: 10.1097/EDE.0b013e3181d61eeb. PMID: 20335814.
- [12] Hernán, Miguel A.; Sauer, Brian C.; Hernández-Díaz, Sonia; Platt, Robert; Shrier, Ian. "Specifying a Target Trial Prevents Immortal Time Bias and Other Self-Inflicted Injuries in Observational Analyses." *Journal of Clinical Epidemiology* 79, 2016, 70–75. DOI: 10.1016/j.jclinepi.2016.04.014. PMID: 27237061.

KNO-015 · Uncertainty and Probability

KNO · Ways of Knowing

Juan Segurola

v0.7 · EN

All Rights Reserved

Ways of Knowing

Epistemic profile: Current scientific interpretation **Licence at publication:** All Rights Reserved

Article Abstract

Scientific results often come with intervals, probabilities, error bars or sensitivity analyses, but what kinds of uncertainty do those devices actually represent? This article asks how measurement uncertainty, sampling variation, parameter uncertainty, model uncertainty, structural uncertainty and expert judgement differ, and what happens when they are collapsed into a single number. It examines confidence intervals, credible intervals, p-values, propagation, prediction and sensitivity analysis while asking which statements depend on a chosen statistical framework and which limits remain outside it. It also asks how to communicate uncertainty without converting ignorance into false precision or treating every unknown as quantifiable. The central problem is methodological: what representation of uncertainty matches the claim being made, and what remains unrepresented even after the calculation is complete?

In 60 Seconds

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 This Question 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.

1. Uncertainty is not one substance

The word covers several mechanisms that should not be collapsed into one residual category.

[C01] Measurement uncertainty concerns incomplete knowledge of the value of a measurand given the measurement procedure, inputs, corrections and model. The GUM framework establishes general rules for evaluating and expressing such uncertainty, while later JCGM documents extend the framework to distribution propagation and measurement models. [S01–S03]

[C02] Sampling uncertainty arises because a finite sample is only one realization from a broader data-generating process. Its statistical treatment depends on the sampling or assignment mechanism and on the estimator being studied. It is conceptually distinct from instrument calibration uncertainty or structural model uncertainty. [S01, S08]

[C03] Model uncertainty includes uncertainty about functional form, omitted structure, approximations and alternative plausible specifications. Structural uncertainty cannot always be reduced to a single standard error estimated inside one chosen model. [S07, S09]

These categories can interact. A calibration constant may enter a statistical model; a model may transform measured inputs; a sampling design may determine which parameters are estimable. The methodological requirement is not to force each source into an isolated box, but to prevent one quantified component from being mistaken for all uncertainty.

Where That Leaves Us

The word covers several mechanisms that should not be collapsed into one residual category. Measurement, sampling and model uncertainty are conceptually distinct, yet they can interact, as when a calibration constant enters a statistical model. What matters is that no quantified component is mistaken for all uncertainty. The next section examines the proposition that a measurement result is not just its displayed digits.

2. A measurement result is not just its displayed digits

A digital instrument can display 12.345678 without establishing knowledge to six decimal places.

[C04] In the GUM/NIST approach, individual standard uncertainty components are combined, including covariance where relevant, to obtain a combined standard uncertainty. Expanded uncertainty is then obtained as a coverage factor times the combined standard uncertainty when that reporting form is appropriate. [S01, S04]

[C05] A coverage factor such as $k = 2$ does not mechanically guarantee 95% coverage in every case. The probability interpretation depends on distributional and other assumptions, and small effective degrees of freedom can require different factors. [S04, S05]

This is a general lesson. Precision printed by software is formatting. Evidential precision comes from the uncertainty model and its justified assumptions.

Where That Leaves Us

A digital instrument can display 12.345678 without establishing knowledge to six decimal places. This is a general lesson. Precision printed by software is formatting. Evidential precision comes from the uncertainty model and its justified assumptions. The next section examines the proposition that probability is a representation tool, not a solvent for ignorance.

3. Probability is a representation tool, not a solvent for ignorance

Probability enters science in more than one role. In repeated-sampling models it characterizes the behaviour of random variables under repetitions. In Bayesian analysis it is also used to represent uncertainty about unknown quantities conditional on a model, prior information and observed data. In metrology, probability distributions are used to represent values that could reasonably be attributed to uncertain quantities and to propagate those distributions through measurement models. [S02, S03]

[C06] The same numerical probability can therefore arise from different inferential constructions. A scientifically adequate report must state what random object or uncertain quantity the probability refers to, and under which assumptions. [S02, S06]

KNO-015 does not decide the philosophical meaning of probability. That belongs to IDE-008. Its concern is operational: do not move between interpretations without saying so.

Where That Leaves Us

Probability enters science in more than one role. [S02, S03] KNO-015 does not decide the philosophical meaning of probability. In Bayesian analysis it is also used to represent uncertainty about unknown quantities conditional on a model, prior information and observed data. The next section examines the proposition that confidence intervals and credible intervals answer different questions.

4. Confidence intervals and credible intervals answer different questions

The phrase “95% interval” is incomplete.

[C07] A frequentist confidence procedure is calibrated by its repeated-sampling properties under the assumed model. The usual 95% confidence interpretation concerns the long-run coverage of the procedure, not a posterior probability that the fixed parameter lies in the particular observed interval. [S10]

[C08] A Bayesian credible interval is a probability statement about a parameter or quantity conditional on the specified Bayesian model, including the prior and likelihood. Its interpretation is therefore not interchangeable with that of a frequentist confidence interval, even when the numerical endpoints are similar. [S10]

Neither label makes assumptions disappear. The relevant question is what data, model and inferential construction justify the interval.

Where That Leaves Us

The phrase “95% interval” is incomplete. Neither label makes assumptions disappear. The relevant question is what data, model and inferential construction justify the interval. What matters here is the stated relation between evidence, method and inference, not a generic promise of certainty. The next section examines the proposition that a p-value is not the probability that a hypothesis is true.

5. A p-value is not the probability that a hypothesis is true

This error is so common that it deserves a separate boundary.

[C09] The American Statistical Association states that p-values can indicate incompatibility between data and a specified statistical model, but do not measure the probability that the studied hypothesis is true, do not measure effect size or scientific importance, and should not by themselves determine scientific or policy conclusions. [S06]

A p-value is calculated under a specified model, usually including a null hypothesis and other assumptions. It is not a posterior probability of that hypothesis unless an additional Bayesian construction has been supplied, in which case the resulting posterior probability is a different object.

Where That Leaves Us

This error is so common that it deserves a separate boundary. It is not a posterior probability of that hypothesis unless an additional Bayesian construction has been supplied, in which case the resulting posterior probability is a different object. The next section examines the proposition that uncertainty propagates through models.

6. Uncertainty propagates through models

If an output depends on uncertain inputs, uncertainty in those inputs propagates to the output.

[C10] The GUM provides analytical propagation rules under its conditions of applicability, while JCGM 101 specifies propagation of probability distributions through a measurement model using Monte Carlo methods. [S01, S02]

Monte Carlo propagation is not magic. It inherits the input distributions, dependence structure and measurement model supplied to it. If those are poor, a large simulation can produce a very precise answer to the wrong uncertainty model.

Where That Leaves Us

If an output depends on uncertain inputs, uncertainty in those inputs propagates to the output. Monte Carlo propagation is not magic. It inherits the input distributions, dependence structure and measurement model supplied to it. If those are poor, a large simulation can produce a very precise answer to the wrong uncertainty model. The next section examines the proposition that structural uncertainty deserves its own line of sight.

7. Structural uncertainty deserves its own line of sight

Some uncertainties live outside the parameter covariance matrix of the selected model: alternative mechanisms, boundary conditions, unrepresented feedbacks, unknown dependence, or unresolved expert disagreement.

[C11] The IPCC uncertainty guidance explicitly separates assessments of evidence and agreement, confidence, likelihood or probability, and asks authors to state assumptions and assess structural uncertainties where relevant. [S07]

This separation prevents a category mistake. A probability distribution can summarize uncertainty conditional on a model while leaving uncertainty about the model itself only partially quantified.

Where That Leaves Us

Some uncertainties live outside the parameter covariance matrix of the selected model: alternative mechanisms, boundary conditions, unrepresented feedbacks, unknown dependence, or unresolved expert disagreement. A probability distribution can summarize uncertainty conditional on a model while leaving uncertainty about the model itself only partially quantified. The next section examines the proposition that expert judgment can be disciplined without pretending it is data.

8. Expert judgment can be disciplined without pretending it is data

In some domains direct frequencies are sparse, models disagree, or future conditions are not repetitions of a stable historical process. Expert judgment may then contribute to uncertainty assessment.

[C12] IPCC guidance uses calibrated language and traceable evaluation of evidence and agreement to discipline such judgments, rather than presenting every confidence statement as a frequency observed in data. [S07]

Expert judgment is not automatically weaker or stronger than statistical output. Its evidential role depends on transparency, calibration, independence, elicitation design and the evidence on which experts rely. The crucial rule is provenance: readers should know when a number came from observations, a model ensemble, formal elicitation or a mixture.

Where That Leaves Us

In some domains direct frequencies are sparse, models disagree, or future conditions are not repetitions of a stable historical process. Expert judgment may then contribute to uncertainty assessment. Expert judgment is not automatically weaker or stronger than statistical output. The next section examines the proposition that unknown unknowns are not inside every interval.

9. Unknown unknowns are not inside every interval

An uncertainty budget can only include sources represented in the analysis.

[C13] Scientific reports should therefore distinguish quantified uncertainty from recognized but unquantified limitations and from genuinely unrecognized error. The National Academies recommends reporting recognized uncertainties and potential sources across stochastic, measurement, computational, knowledge, modelling and analysis dimensions rather than treating a single statistical interval as exhaustive. [S08]

This does not make quantified intervals useless. It defines their jurisdiction.

Where That Leaves Us

An uncertainty budget can only include sources represented in the analysis. This does not make quantified intervals useless. Its value lies in making this particular inferential step inspectable rather than treating the method as self-validating. The next section examines the proposition that sensitivity analysis tests the dependence of conclusions on assumptions.

10. Sensitivity analysis tests the dependence of conclusions on assumptions

If a conclusion changes sharply under plausible alternative specifications, the uncertainty is not captured by quoting only the standard error under one specification.

Sensitivity analysis can vary priors, functional forms, nuisance parameters, missing-data assumptions, calibration inputs or plausible structural choices. It is not a universal cure: the analyst must still decide which alternatives are plausible. But it exposes where conclusions depend on choices that would otherwise remain hidden.

Where That Leaves Us

If a conclusion changes sharply under plausible alternative specifications, the uncertainty is not captured by quoting only the standard error under one specification. It is not a universal cure: the analyst must still decide which alternatives are plausible. The next section examines the proposition that prediction uncertainty is not parameter uncertainty.

11. Prediction uncertainty is not parameter uncertainty

Estimating a mean response and predicting a new observation are different tasks. A model parameter can be well estimated while individual future outcomes remain highly variable.

[C14] A complete uncertainty statement must identify the target quantity: a parameter, a latent state, a measurand, a future observation, a population average, or a decision-relevant outcome. Different targets generally require different uncertainty constructions. [S01, S03, S10]

This is why “the uncertainty is $\pm x$ ” is often underspecified.

Where That Leaves Us

Estimating a mean response and predicting a new observation are different tasks. A model parameter can be well estimated while individual future outcomes remain highly variable. This is why “the uncertainty is $\pm x$ ” is often underspecified. The conclusion is therefore useful only in the role defined by this section, with its stated evidential conditions visible. The next section examines a practical uncertainty workflow.

12. A practical uncertainty workflow

For scientific work, the following sequence is more robust than attaching an interval at the end:

1. Define the target quantity and estimand or measurand.
2. Identify the data-generating or measurement process.
3. List material uncertainty sources and dependencies.
4. Separate stochastic variability from incomplete knowledge where the distinction matters.
5. Specify the model and inferential framework.
6. Quantify components that can be justified quantitatively.
7. Propagate them through the model, analytically or computationally as appropriate.
8. Test sensitivity to plausible alternative assumptions.
9. Record important limitations that remain unquantified.
10. Report the meaning of every interval, probability and coverage statement.

The workflow does not guarantee truth. It makes the path from evidence to uncertainty claim auditable.

Where That Leaves Us

Rather than attaching an interval at the end, the workflow defines the target quantity and its data or measurement process, lists uncertainty sources, separates variability from incomplete knowledge, specifies model and framework, quantifies and propagates justified components, tests sensitivity, records what stays unquantified and reports what each interval means. This makes the path auditable without guaranteeing truth. The next section asks what uncertainty does not license.

13. What uncertainty does not license

Uncertainty is not permission to say that every claim is equally plausible. Nor does a narrow interval prove a model correct. A result may be statistically precise and systematically wrong. A wide interval may faithfully report weak information rather than methodological failure.

The strongest uncertainty statement is therefore not always the smallest number. It is the one whose scope, assumptions and construction survive inspection.

Where That Leaves Us

Uncertainty is not permission to say that every claim is equally plausible. A result may be statistically precise and systematically wrong. A wide interval may faithfully report weak information rather than methodological failure. This defines the article's boundary and prepares the final synthesis without extending the claim beyond what the preceding sections establish.

Boundary note

KNO-015 is about **the methods by which scientific work identifies, represents, propagates and communicates uncertainty**. It does not attempt to settle what probability fundamentally is; that conceptual question belongs to **IDE-008 · Probability**. It does not treat uncertainty as a general

ontological or epistemological concept; that belongs to **IDE-017 · Uncertainty**. It builds directly on **KNO-011 · Measuring Before Theorising**, **KNO-012 · Experiments and Variable Control**, **KNO-013 · Models, Maps and Reality**, and **KNO-014 · Causality and Correlation**.

Declared framing statements not represented as numbered claims

This list is a control aid, **not a guarantee of semantic exhaustiveness**. It records substantive framing sentences intentionally left outside the numbered claim ledger and the reason for doing so.

- “Uncertainty is not permission to say that every claim is equally plausible.” Normative framing of calibrated uncertainty, not a distinct empirical claim.
- “A result may be statistically precise and systematically wrong.” Compressed methodological warning already bounded by the sections on systematic/model uncertainty.
- “A wide interval may faithfully report weak information rather than methodological failure.” Interpretive framing of interval width, not an independent quantitative rule.
- “The strongest uncertainty statement is therefore not always the smallest number.” Synthesis of the preceding uncertainty distinctions, not a new source-level proposition.

What We Know, and How

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 Still 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 Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-011 · Measuring Before Theorising
- KNO-014 · Causality and Correlation

- KNO-016 · How Evidence Is Built: Putting the Tools Together
- KNO-021 · Open Science and Reproducibility

References

- [S01] Joint Committee for Guides in Metrology (JCGM). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*. JCGM 100:2008(E). DOI: 10.59161/JCGM100-2008E.
- [S02] Joint Committee for Guides in Metrology (JCGM). *Evaluation of measurement data — Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method*. JCGM 101:2008. DOI: 10.59161/JCGM101-2008.
- [S03] Joint Committee for Guides in Metrology (JCGM). *Guide to the expression of uncertainty in measurement — Part 6: Developing and using measurement models*. JCGM GUM-6:2020. DOI: 10.59161/JCGMGUM-6-2020.
- [S04] Taylor, B. N.; Kuyatt, C. E. *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*. NIST Technical Note 1297. NIST web manifestation consulted 5 September 2026; page updated 6 May 2026. <https://www.nist.gov/pml/nist-technical-note-1297>
- [S05] National Institute of Standards and Technology. *NIST TN 1297: Appendix B. Coverage Factors*. NIST web page, updated 11 September 2025; consulted 5 September 2026. <https://www.nist.gov/pml/nist-technical-note-1297/nist-tn-1297-appendix-b-coverage-factors>
- [S06] Wasserstein, Ronald L.; Lazar, Nicole A. “The ASA’s Statement on p-Values: Context, Process, and Purpose.” *The American Statistician* 70(2), 2016, 129–133. DOI: 10.1080/00031305.2016.1154108.
- [S07] Mastrandrea, Michael D.; Field, Christopher B.; Stocker, Thomas F.; Edenhofer, Ottmar; Ebi, Kristie L.; Frame, David J.; Held, Hermann; Kriegler, Elmar; Mach, Katharine J.; Matschoss, Patrick R.; Plattner, Gian-Kasper; Yohe, Gary W.; Zwiers, Francis W. *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. IPCC, 2010.
- [S08] National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.
- [S09] Draper, David. “Assessment and Propagation of Model Uncertainty.” *Journal of the Royal Statistical Society: Series B* 57(1), 1995, 45–70. DOI: 10.1111/j.2517-6161.1995.tb02015.x.
- [S10] Possolo, Antonio; Gelabert, Maria Victoria; Hibbert, David; Stohner, Jürgen; Bodnar, Olha; Meija, Juris. *Guía Breve sobre la Incertidumbre de Medición*. NIST Technical Note 2330, 2025. DOI: 10.6028/NIST.TN.2330.
- [S11] Joint Committee for Guides in Metrology (JCGM). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement. Amendment 1: Nonlinearity in measurement models*. JCGM 100:2008/Amd.1:2026. DOI: 10.59161/PPDI3267.

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

KNO · Ways of Knowing

Juan Segurola

v0.6 · EN

All Rights Reserved

Ways of Knowing

Epistemic profile: Current scientific interpretation **Licence at publication:** All Rights Reserved

Article Abstract

Evidence is not a substance that appears automatically when data accumulate. What has to be connected for an observation to bear on a scientific claim? This article follows that question from claim formulation and measurement through study design, modelling, causal identification, uncertainty, replication, reproducibility, triangulation, directness, transparency and collective scrutiny. It asks which links are empirical, which depend on assumptions, which can be tested independently and which remain vulnerable to shared error. It also asks why more studies do not necessarily mean stronger evidence if they repeat the same weaknesses, and why evidence can become more reliable without becoming certain. The aim is to make the architecture of evidential support explicit enough that another reader can see not only what conclusion is proposed, but how each step is supposed to support it.

In 60 Seconds

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 This Question 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.

1. Evidence begins with a claim precise enough to fail

Before asking whether evidence is strong, one must know what claim is being supported.

[C01] A scientific claim should identify its target sufficiently clearly that observations or analyses could count against it. Broad narratives that change meaning after each result are difficult to test because the evidential target moves. [S01, S02]

Operational definitions are not mere paperwork. They connect theoretical language to observable procedures. The same word can refer to different estimands, measurements or classifications. If the target shifts between design, analysis and interpretation, the apparent evidence may answer a different question from the one originally posed.

Where That Leaves Us

Before asking whether evidence is strong, one must know what claim is being supported. Operational definitions are not mere paperwork. They connect theoretical language to observable procedures. The same word can refer to different estimands, measurements or classifications. The next section examines the proposition that observations are produced, not simply found.

2. Observations are produced, not simply found

KNO-008 showed that instruments extend the senses; KNO-011 showed that numbers depend on measurement conventions. Here those lessons become part of the evidence chain.

[C02] The evidential value of a measurement depends on how the measurand or variable was defined, how the instrument or procedure was calibrated, what corrections were applied and what uncertainty remains. A numerical output is not self-validating. [S03]

A photograph, spectrum, assay result or survey response is already the result of choices about sampling, resolution, processing and classification. Evidence begins before statistical analysis.

Where That Leaves Us

KNO-008 showed that instruments extend the senses; KNO-011 showed that numbers depend on measurement conventions. Here those lessons become part of the evidence chain. A photograph, spectrum, assay result or survey response is already the result of choices about sampling, resolution, processing and classification. The next section examines the proposition that design determines which alternatives can be excluded.

3. Design determines which alternatives can be excluded

Different questions require different designs.

[C03] Experimental controls, randomization, blinding, repeated measurement and observational comparison structures address different potential biases. None is a universal marker of quality independent of the scientific question. [S04]

For causal questions, design determines which counterfactual contrasts are plausibly identified. For descriptive questions, representative sampling may matter more than random treatment allocation. For historical sciences, convergent traces and mechanistic consistency can carry the evidential load when intervention is impossible.

Where That Leaves Us

Different questions require different designs. For descriptive questions, representative sampling may matter more than random treatment allocation. For historical sciences, convergent traces and mechanistic consistency can carry the evidential load when intervention is impossible. Its value lies in making this particular inferential step inspectable rather than treating the method as self-validating. The next section examines the proposition that models are bridges between observations and claims.

4. Models are bridges between observations and claims

Data rarely speak directly to the proposition of interest. A model translates observations into quantities, mechanisms or predictions.

[C04] Every model-based evidential claim is conditional on modelling choices to some degree. Good practice therefore distinguishes what is directly observed from what is inferred through the model and tests sensitivity to plausible alternatives. [S06]

This is not a defect unique to complex simulations. Even a straight-line fit, a calibration curve or a diagnostic threshold is a model.

Where That Leaves Us

Data rarely speak directly to the proposition of interest. This is not a defect unique to complex simulations. Even a straight-line fit, a calibration curve or a diagnostic threshold is a model. What matters here is the stated relation between evidence, method and inference, not a generic promise of certainty. The next section examines the proposition that causal evidence requires more than association.

5. Causal evidence requires more than association

KNO-014 established the central causal boundary.

[C05] An observed association can support a causal claim only through assumptions and design features that address confounding, reverse causation, selection and other competing explanations. The strength of causal evidence therefore depends on the identification strategy, not simply on the size or statistical significance of the association. [S04, S07]

Different causal methods can be combined, but multiple analyses that share the same hidden bias are not independent confirmation.

Where That Leaves Us

KNO-014 established the central causal boundary. Different causal methods can be combined, but multiple analyses that share the same hidden bias are not independent confirmation. The conclusion is therefore useful only in the role defined by this section, with its stated evidential conditions visible. The next section examines the proposition that uncertainty belongs inside the claim.

6. Uncertainty belongs inside the claim

KNO-015 treated uncertainty as a methodological object rather than a decorative interval.

[C06] Evidence assessment should distinguish uncertainty arising from measurement, sampling, modelling and other recognized sources. A narrow within-model interval does not establish that unmodelled systematic error is small. [S03, S08]

A strong claim therefore states not only an estimate but also the conditions under which the estimate is expected to be reliable.

Where That Leaves Us

KNO-015 treated uncertainty as a methodological object rather than a decorative interval. A strong claim therefore states not only an estimate but also the conditions under which the estimate is expected to be reliable. Its value lies in making this particular inferential step inspectable rather than treating the method as self-validating. The next section examines the proposition that replication and reproducibility test different links.

7. Replication and reproducibility test different links

The terms are often blurred.

[C07] The National Academies distinguishes computational reproducibility, obtaining consistent computational results using the same data, code and methods, from replicability, obtaining consistent results across studies addressing the same scientific question with new data. [S01]

[C08] A successful replication increases confidence but does not prove an original claim universally correct; a failed replication does not by itself conclusively refute it because differences can arise from genuine heterogeneity, uncontrolled variables, methodological defects or other causes. [S01]

This asymmetry matters. Repetition is evidence about robustness, not a logical proof operator.

Where That Leaves Us

The terms are often blurred. Repetition is evidence about robustness, not a logical proof operator. What matters here is the stated relation between evidence, method and inference, not a generic promise of certainty. [S01] A successful replication increases confidence but does not prove an original claim universally correct. The next section examines the proposition that triangulation is strongest when errors differ.

8. Triangulation is strongest when errors differ

Independent evidence does not mean simply “more studies.”

[C09] Triangulation gains force when different methods address the same question while having different key sources of bias. Convergence is more informative when the biases would not be expected to drive all approaches toward the same result. [S07]

Ten studies using the same instrument, same selection mechanism and same model may reproduce a shared bias. A smaller set using genuinely different designs can be more diagnostically valuable.

Where That Leaves Us

Independent evidence does not mean simply “more studies.” Ten studies using the same instrument, same selection mechanism and same model may reproduce a shared. A smaller set using genuinely different designs can be more diagnostically valuable. The next section examines the proposition that evidence is cumulative, but accumulation needs structure.

9. Evidence is cumulative, but accumulation needs structure

Meta-analysis, systematic review and evidence frameworks can combine studies, yet combination is only meaningful when the underlying questions are sufficiently compatible.

[C10] Evidence frameworks such as GRADE explicitly examine risk of bias, inconsistency, indirectness, imprecision and dissemination/publication bias rather than treating study count alone as certainty. [S09]

The point is not that GRADE governs every scientific field. It demonstrates a general architecture: bodies of evidence must be evaluated for systematic error, coherence, relevance and precision, not merely aggregated.

Where That Leaves Us

Meta-analysis, systematic review and evidence frameworks can combine studies, yet combination is only meaningful when the underlying questions are sufficiently compatible. The point is not that GRADE governs every scientific field. It demonstrates a general architecture: bodies of evidence must be evaluated for systematic error, coherence, relevance and precision, not merely aggregated. The next section examines the proposition that directness matters.

10. Directness matters

Evidence can be accurate yet answer the wrong question.

A laboratory surrogate may not predict a clinical outcome. A result in one population may not transport to another. A proxy variable may measure only one component of the intended construct.

[C11] The concept of indirectness formalizes mismatches between the question of interest and the populations, interventions or exposures, comparators and outcomes actually studied. [S09, S10]

External validity and transportability therefore belong to evidence assessment, not to an afterthought after internal validity has been established.

Where That Leaves Us

Evidence can be accurate yet answer the wrong question. A laboratory surrogate may not predict a clinical outcome. A result in one population may not transport to another. A proxy variable may measure only one component of the intended construct. The next section examines the proposition that transparency changes what others can test.

11. Transparency changes what others can test

Evidence is stronger when the path from raw material to conclusion can be inspected.

[C12] Clear reporting of methods, data-processing steps, analytic choices and uncertainty supports computational checking and helps other investigators identify errors or test robustness. Reproducibility is facilitated by access to the relevant digital artifacts and sufficiently complete methods. [S01, S02]

Transparency does not make a flawed analysis correct. It makes flaws more discoverable.

Where That Leaves Us

Evidence is stronger when the path from raw material to conclusion can be inspected. Transparency does not make a flawed analysis correct. It makes flaws more discoverable. The conclusion is therefore useful only in the role defined by this section, with its stated evidential conditions visible. The next section examines the proposition that peer review and consensus are filters, not substitutes for evidence.

12. Peer review and consensus are filters, not substitutes for evidence

Peer review can detect errors, request clarification and improve reporting, but it does not independently reproduce the underlying experiment. Consensus can summarize a mature evidential landscape, but shared assumptions can also propagate through a field.

These institutional processes belong fully to KNO-020. Here the boundary is simpler: social validation can strengthen quality control, but the evidential chain still has to connect observations to claims.

Where That Leaves Us

Peer review can detect errors, request clarification and improve reporting, but it does not independently reproduce the underlying experiment. Consensus can summarize a mature evidential landscape, but shared assumptions can also propagate through a field. These institutional processes belong fully to KNO-020. The next section examines a practical evidence audit.

13. A practical evidence audit

Before accepting a strong scientific conclusion, ask:

1. **What exactly is the claim?** Descriptive, predictive, causal, mechanistic or historical?
2. **What observations anchor it?** What was directly measured or recorded?
3. **How were variables constructed?** What classifications, thresholds or transformations intervened?
4. **What design addresses alternatives?** Controls, randomization, matching, temporal ordering, natural variation, comparative cases?
5. **What model connects data to claim?** Which assumptions are essential?
6. **What uncertainties are quantified?** Which remain recognized but unquantified?
7. **What biases could move the answer?** In what direction?
8. **Has the computation been reproduced?** Can the analytic path be rerun?
9. **Has the finding been replicated or triangulated?** Are the new failure modes genuinely different?
10. **How direct is the evidence for this target population, quantity or mechanism?**

|

11. **Is the result consistent with a broader body of evidence?** If not, is the disagreement informative?

12. **Which conclusion is actually warranted?** Avoid upgrading “compatible with” into “proved.”

Where That Leaves Us

Before accepting a strong scientific conclusion, ask: What uncertainties are quantified? Which remain recognized but unquantified? What observations anchor it? What was directly measured or recorded? How were variables constructed? What classifications, thresholds or transformations intervened? The next section examines the proposition that evidence can become stronger without becoming certain.

14. Evidence can become stronger without becoming certain

Science accumulates constraint. Measurements narrow possible values; experiments eliminate alternatives; models organize implications; independent methods test different vulnerabilities; replication probes robustness; synthesis shows where results cohere and where they do not.

The endpoint is rarely certainty. It is a claim whose remaining alternatives have become progressively harder to sustain.

Where That Leaves Us

Measurements narrow possible values; experiments eliminate alternatives; models organize implications; independent methods test different vulnerabilities; replication probes robustness; synthesis shows where results cohere and where they do not. The endpoint is rarely certainty. This defines the article's boundary and prepares the final synthesis without extending the claim beyond what the preceding sections establish.

Boundary note

KNO-016 is deliberately a **synthesis article**. It does not replace the detailed methods in KNO-008 through KNO-015, and it does not pre-empt KNO-020 on peer review and consensus or KNO-021 on open science. Its role is to show how measurement, experimental design, modelling, causality, uncertainty, reproducibility, replication and triangulation combine into an auditable evidential argument.

What We Know, and How

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 Still 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 Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-012 · Experiments and Variable Control
- KNO-013 · Models, Maps and Reality
- KNO-014 · Causality and Correlation
- KNO-015 · Uncertainty and Probability
- KNO-020 · Review, Consensus and Disagreement

References

- [S01] National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.
- [S02] National Academies of Sciences, Engineering, and Medicine. “Understanding Reproducibility and Replicability,” chapter 3 in *Reproducibility and Replicability in Science*, pp. 39–54. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.
- [S03] Joint Committee for Guides in Metrology (JCGM). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*. JCGM 100:2008(E). DOI: 10.59161/JCGM100-2008E.
- [S04] Hernán, Miguel A. “A Definition of Causal Effect for Epidemiological Research.” *Journal of Epidemiology and Community Health* 58(4), 2004, 265–271. DOI: 10.1136/jech.2002.006361.
- [S05] Schulz, Kenneth F.; Altman, Douglas G.; Moher, David; CONSORT Group. “CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials.” *BMJ* 340, 2010, c332. DOI: 10.1136/bmj.c332.
- [S06] Frigg, Roman; Hartmann, Stephan. “Models in Science.” *The Stanford Encyclopedia of Philosophy*, Summer 2025 Edition; substantive revision 2 April 2025. <https://plato.stanford.edu/archives/sum2025/entries/models-science/>
- [S07] Lawlor, Debbie A.; Tilling, Kate; Davey Smith, George. “Triangulation in aetiological epidemiology.” *International Journal of Epidemiology* 45(6), 2016, 1866–1886. DOI: 10.1093/ije/dyw314.

[S08] Mastrandrea, Michael D.; Field, Christopher B.; Stocker, Thomas F.; Edenhofer, Ottmar; Ebi, Kristie L.; Frame, David J.; Held, Hermann; Kriegler, Elmar; Mach, Katharine J.; Matschoss, Patrick R.; Plattner, Gian-Kasper; Yohe, Gary W.; Zwiers, Francis W. *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*. IPCC, 2010. <https://www.ipcc.ch/site/assets/uploads/2018/05/uncertainty-guidance-note.pdf>

[S09] Neumann, Ignacio; Brennan, Sue; Meerpohl, Joerg; Davoli, Marina; Alonso Coello, Pablo; Akl, Elie; Dahm, Philipp; Skoetz, Nicole; Xia, Jun; Brozek, Jan; Schünemann, Holger J. "Overview of the GRADE approach." *GRADE Book*. Last modified 12 May 2026; consulted 5 September 2026. <https://book.gradepro.org/guideline/overview-of-the-grade-approach>

[S10] "Indirectness." *GRADE Book*. Last modified 12 May 2026; consulted 5 September 2026. <https://book.gradepro.org/guideline/indirectness>. Author display currently contains an apparent duplicated Holger J. Schünemann entry; author chain therefore remains P pending human confirmation and is not normalised in this bibliography.

KNO-017 · Is Mathematics a Language?

KNO · Ways of Knowing

Juan Segurola

v0.6 · EN

All Rights Reserved

Ways of Knowing

Epistemic profile: Current scientific interpretation **Licence at publication:** All Rights Reserved

Article Abstract

Mathematics is often called a language, but what exactly is being compared when we say that? This article asks whether mathematical notation, formal systems, proof practices, abstraction and applied modelling satisfy the same functions as natural language, or whether the analogy works only for selected parts of mathematics. It examines syntax, semantics, interpretation, proof, calculation and external cognitive representation while separating mathematical expressions from the wider practices in which they are created and justified. It also asks why formal symbols can support unusually precise manipulation without carrying meaning by themselves, and why the effectiveness of mathematics in science does not settle what mathematics ultimately is. The question is therefore not whether mathematics uses languages, but how far the language analogy can be pushed before it hides important differences.

In 60 Seconds

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 This Question 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.

1. There Is More Than One Sense of “Language”

Natural languages combine syntax, semantics, pragmatics and social convention. Formal languages deliberately isolate some of these features. In classical logic, a formal language can be specified as a recursively defined set of strings over an alphabet, while deductive rules and semantics provide inferential or interpretive structure [S01].

Mathematics uses many such formal or semi-formal systems, but it also uses natural language and diagrams. The phrase “mathematics is a language” is therefore not a single technical thesis. It can refer to notation, to formal syntax, to proof, to communication, or to representation.

[C01] Parts of mathematics can be represented as formal languages with explicit syntax and inferential rules, but this does not by itself establish that all mathematical practice is one formal language. [S01,S02]

Where That Leaves Us

Natural languages combine syntax, semantics, pragmatics and social convention. Mathematics uses many such formal or semi-formal systems, but it also uses natural language and diagrams. In classical logic, a formal language can be specified as a recursively defined set of strings over an alphabet, while deductive rules and semantics provide inferential or interpretive structure [S01]. The next section examines the proposition that syntax is not yet meaning.

2. Syntax Is Not Yet Meaning

A string can be well formed without yet specifying what its symbols denote. Logic textbooks distinguish syntax from semantics precisely because a formal expression and an interpretation of that expression are different objects [S01].

This matters scientifically. The equation $y = ax + b$ can describe many systems, or none, depending on how its symbols are mapped to quantities and conditions. Formal validity therefore does not certify empirical interpretation.

[C02] Well-formed mathematical syntax does not by itself determine the empirical meaning of the symbols being manipulated. [S01]

Where That Leaves Us

A string can be well formed without yet specifying what its symbols denote. Formal validity therefore does not certify empirical interpretation. The equation $y = ax + b$ can describe many systems, or none, depending on how its symbols are mapped to quantities and conditions. The next section examines the proposition that notation changes what can be seen and manipulated.

3. Notation Changes What Can Be Seen and Manipulated

Notation is not a transparent wrapper around pre-existing thought. A good notation makes some structures easier to compare, transform and reuse. Historical work on mathematical notation shows how symbolic compression can expose patterns that are cumbersome in prose [S03].

This does not mean notation creates mathematical truth. It means representations alter the cognitive and practical accessibility of relationships. Algebraic notation, coordinate systems and operator notation make different operations cheap or expensive to perform.

[C03] Mathematical notation can reorganise the accessibility of relationships and operations even when the underlying mathematical content is unchanged. [S03]

Where That Leaves Us

Notation is not a transparent wrapper around pre-existing thought. This does not mean notation creates mathematical truth. Historical work on mathematical notation shows how symbolic compression can expose patterns that are cumbersome in prose [S03]. It means representations alter the cognitive and practical accessibility of relationships. The next section examines the proposition that formalisation makes rules inspectable.

4. Formalisation Makes Rules Inspectable

Formalisation replaces many tacit conventions with explicit grammars, definitions and inference rules. Mathematical logic made it possible to study mathematical theories themselves as formal objects [S02,S04]. Modern proof assistants extend that idea computationally: definitions and proofs can be encoded in a formal language that a machine parses and checks against a foundational system [S05].

The gain is auditability, not omniscience. Formal verification checks a derivation relative to the encoded definitions, axioms, libraries and software environment. It does not decide whether a scientific formalisation chose the right phenomenon to represent.

[C04] Formalisation can make syntactic and inferential commitments mechanically inspectable, while leaving the choice and interpretation of the formalisation as a separate problem. [S04,S05]

Where That Leaves Us

Formalisation replaces many tacit conventions with explicit grammars, definitions and inference rules. The gain is auditability, not omniscience. Modern proof assistants extend that idea computationally: definitions and proofs can be encoded in a formal language that a machine parses and checks against a foundational system [S05]. The next section examines the proposition that mathematical proof is more than printed symbol strings.

5. Mathematical Proof Is More Than Printed Symbol Strings

Mathematical practice usually communicates proofs in a compressed social form. Published arguments omit elementary transformations, rely on shared definitions and often mix prose, equations and diagrams. Contemporary philosophy of mathematical practice therefore distinguishes formal derivations from the informal or semi-formal proofs mathematicians actually write and evaluate [S06].

A proof can consequently have two related standards: whether its conclusion follows under the accepted mathematical rules, and whether the presentation makes the relevant reasoning intelligible and checkable to its audience.

[C05] Ordinary mathematical proofs are commonly presented in natural language interspersed with formal symbols rather than as fully expanded formal derivations. [S06]

Where That Leaves Us

Mathematical practice usually communicates proofs in a compressed social form. Published arguments omit elementary transformations, rely on shared definitions and often mix prose, equations and diagrams. Contemporary philosophy of mathematical practice therefore distinguishes formal derivations from the informal or semi-formal proofs mathematicians actually write and evaluate [S06]. The next section examines the proposition that formal languages can have many interpretations.

6. Formal Languages Can Have Many Interpretations

One reason formal systems are reusable is that the same syntax can admit different models or interpretations. This is a strength, but it blocks a simple identification between an equation and one physical reality.

In applied science, a mathematical structure becomes a model of a target system only after variables, parameters, boundary conditions and correspondence relations are specified [S01,S07]. KNO-013 treats model adequacy; here the narrower point is representational: mathematical form underdetermines application.

[C06] Reuse of the same mathematical structure across domains depends on interpretation; the formal structure alone does not select its scientific target. [S01,S07]

Where That Leaves Us

One reason formal systems are reusable is that the same syntax can admit different models or interpretations. This is a strength, but it blocks a simple identification between an equation and one physical reality. KNO-013 treats model adequacy; here the narrower point is representational: mathematical form underdetermines application. The next section examines the proposition that abstraction is a feature, not a defect.

7. Abstraction Is a Feature, Not a Defect

Mathematics often suppresses material detail in order to preserve structure. The same group, graph, differential equation or probability space can be instantiated by systems with very different physical composition. This abstraction supports transfer of methods between domains.

But abstraction also creates a gate: the omitted features must genuinely be irrelevant to the question being asked. The mathematical derivation may be exact inside the abstraction while the abstraction is scientifically inadequate.

[C07] Mathematical abstraction supports transfer by preserving selected structure across different domains, but scientific adequacy depends on whether the omitted features are irrelevant to the target question. [S02,S07]

Where That Leaves Us

Mathematics often suppresses material detail in order to preserve structure. But abstraction also creates a gate: the omitted features must genuinely be irrelevant to the question being asked. This abstraction supports transfer of methods between domains. The next section examines the proposition that mathematics can express relations more precisely than prose.

8. Mathematics Can Express Relations More Precisely Than Prose

Natural language can state that one quantity rises with another; an equation can specify the functional relation, parameter dependence and limiting cases. Units, domains and assumptions can be made explicit. This precision is one reason mathematics is indispensable in mature quantitative sciences.

Yet precision of expression must not be confused with precision of knowledge. A model can contain many significant digits while its parameters are poorly measured or its structure is wrong. KNO-015 supplies the uncertainty boundary.

[C08] Mathematical representation can increase precision about relations and assumptions without guaranteeing that the represented quantities or model are empirically accurate. [S07]

Where That Leaves Us

Natural language can state that one quantity rises with another; an equation can specify the functional relation, parameter dependence and limiting cases. Yet precision of expression must not be confused with precision of knowledge. This precision is one reason mathematics is indispensable in mature quantitative sciences. The next section examines the proposition that calculation and explanation are not identical.

9. Calculation and Explanation Are Not Identical

A derivation can produce a correct numerical result without explaining why the phenomenon occurs. Conversely, an explanatory argument may depend on qualitative structure rather than a long calculation. Philosophers of mathematics and science therefore distinguish proof, calculation, prediction and explanation rather than treating them as synonyms [S06,S07].

The distinction matters for communication. “The equation gives the right number” is evidence of empirical adequacy in a tested domain, not automatically an explanation of mechanism.

[C09] Correct calculation, valid proof, successful prediction and scientific explanation are related but non-identical epistemic achievements. [S06,S07]

Where That Leaves Us

A derivation can produce a correct numerical result without explaining why the phenomenon occurs. Conversely, an explanatory argument may depend on qualitative structure rather than a long calculation. Philosophers of mathematics and science therefore distinguish proof, calculation, prediction and explanation rather than treating them as synonyms [S06,S07]. The next section asks why mathematics works so well in science.

10. Why Does Mathematics Work So Well in Science?

Wigner famously framed the effectiveness of mathematics in the natural sciences as something that itself demands explanation [S08]. His essay is historically important because it identifies a genuine puzzle: mathematical structures developed with substantial autonomy often later find striking scientific applications.

But the observation does not settle what mathematics is. It does not prove Platonism, formalism, structuralism or the claim that the universe literally “is mathematics”. Those are philosophical positions that require additional arguments.

[C10] The successful application of mathematics to natural science is an empirical and philosophical fact to be explained, not by itself a proof of any single ontology of mathematics. [S02,S08]

Where That Leaves Us

Wigner famously framed the effectiveness of mathematics in the natural sciences as something that itself demands explanation [S08]. But the observation does not settle what mathematics is. It does not prove Platonism, formalism, structuralism or the claim that the universe literally “is mathematics”. The next section examines mathematics as external cognitive infrastructure.

11. Mathematics as External Cognitive Infrastructure

Written mathematics lets reasoning persist outside a single working memory. Expressions can be copied, transformed, checked and shared. A notation can preserve intermediate states so that a long inference does not have to be mentally reconstructed from scratch [S03,S05].

This connects KNO-017 to KNO-022, which will treat computation as external memory more generally. Here the narrower claim is that symbolic mathematics is an external representational technology as well as an abstract discipline.

[C11] Mathematical inscriptions and formal representations can function as external, revisable states in extended chains of reasoning. [S03,S05]

Where That Leaves Us

Written mathematics lets reasoning persist outside a single working memory. A notation can preserve intermediate states so that a long inference does not have to be mentally reconstructed from scratch [S03,S05]. Here the narrower claim is that symbolic mathematics is an external representational technology as well as an abstract discipline. The next section synthesizes the answer to the article's central question.

12. The Answer: Mathematics Contains Languages, but Is Not Exhausted by Them

The strongest defensible answer is therefore conditional. Mathematics contains formal languages and depends heavily on learned symbolic conventions. These languages provide compact representation and explicit inferential structure. But mathematics as practiced also includes interpretation, informal reasoning, diagrams, examples, proof strategies and judgments about relevance.

Calling mathematics “a language” is useful when it highlights representation and rule-governed inference. It is too strong when it collapses all mathematics into syntax or treats symbols as self-interpreting.

[C12] “Mathematics is a language” is defensible as a methodological metaphor for its representational and rule-governed systems, but not as an exhaustive description of mathematical practice.
[S01,S02,S06]

Where That Leaves Us

The strongest defensible answer is therefore conditional. Mathematics contains formal languages and depends heavily on learned symbolic conventions. These languages provide compact representation and explicit inferential structure. But mathematics as practiced also includes interpretation, informal reasoning, diagrams, examples, proof strategies and judgments about relevance. Calling mathematics “a language” is useful when it highlights representation and rule-governed inference. The next section defines the boundary of KNO-017.

13. Boundary of KNO-017

KNO-017 does **not** decide whether mathematical objects exist independently of minds, whether mathematical truth is discovered or invented, or which philosophy of mathematics is correct. Those questions belong to the conceptual IDE series.

It also does not treat diagrams and visual encodings in depth; KNO-018 does that. Nor does it treat computation as a general external-memory technology; KNO-022 does that.

The methodological result is narrower: mathematical notation, formal languages and proof are tools that make relationships representable, transformable, communicable and inspectable, but their application to the world requires interpretation and evidence beyond syntax.

Where That Leaves Us

KNO-017 does not decide whether mathematical objects exist independently of minds, whether mathematical truth is discovered or invented, or which philosophy of mathematics is correct. It also does not treat diagrams and visual encodings in depth; KNO-018 does that. This defines the article's boundary and prepares the final synthesis without extending the claim beyond what the preceding sections establish.

What We Know, and How

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 Still 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 Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-003 · From Marks to Symbols
- KNO-013 · Models, Maps and Reality
- KNO-016 · How Evidence Is Built: Putting the Tools Together
- KNO-018 · Images, Diagrams and Visualisation

References

[S01] Shapiro, Stewart; Kouri Kissel, Teresa. "Classical Logic." *The Stanford Encyclopedia of Philosophy*, Fall 2022 archive, substantive revision 29 June 2022. <https://plato.stanford.edu/archives/fall2022/entries/logic-classical/>

[S02] Horsten, Leon. "Philosophy of Mathematics." *The Stanford Encyclopedia of Philosophy*, Spring 2022 Edition, substantive revision 25 January 2022. <https://plato.stanford.edu/archives/spr2022/entries/philosophy-mathematics/>

[S03] Mazur, Joseph. *Enlightening Symbols: A Short History of Mathematical Notation and Its Hidden Powers*. Princeton University Press, 2014/2016 online edition. DOI: 10.23943/princeton/9780691173375.001.0001.

[S04] Weir, Alan. "Formalism in the Philosophy of Mathematics." *The Stanford Encyclopedia of Philosophy*, Fall 2024 Edition, substantive revision 20 February 2024. <https://plato.stanford.edu/archives/fall2024/entries/formalism-mathematics/>

[S05] Avigad, Jeremy; Commelin, Johan; Macbeth, Heather; Topaz, Adam. "Anatomy of a Formal Proof." arXiv:2411.11885, 2024. <https://arxiv.org/abs/2411.11885>

[S06] Tanswell, Fenner Stanley. *Mathematical Rigour and Informal Proof*. Cambridge University Press, 2024. Cambridge Elements in the Philosophy of Mathematics.

[S07] Frigg, Roman; Hartmann, Stephan. "Models in Science." *The Stanford Encyclopedia of Philosophy*, Summer 2025 Edition, substantive revision 2 April 2025. <https://plato.stanford.edu/archives/sum2025/entries/models-science/>

[S08] Wigner, Eugene P. "The Unreasonable Effectiveness of Mathematics in the Natural Sciences." *Communications on Pure and Applied Mathematics* 13(1), 1960, 1–14. DOI: 10.1002/cpa.3160130102.

KNO-018 · Images, Diagrams and Visualisation

KNO · Ways of Knowing

Juan Segurola

v0.6 · EN

All Rights Reserved

Ways of Knowing

Epistemic profile: Current scientific interpretation **Licence at publication:** All Rights Reserved

Article Abstract

Scientific images, diagrams and graphs can make patterns visible that are difficult to express in prose, but what exactly happens when data are turned into a visual display? This article examines diagrams, graphical encodings, scientific images and visualisation pipelines as related but distinct tools. It asks how spatial arrangement changes cognitive work, how perceptual channels affect quantitative judgement, how scale, framing and colour influence interpretation, and how acquisition and processing stand between a physical system and the pixels a reader sees. It also asks when alternative views reveal dependence on arbitrary choices, when processing clarifies rather than distorts, and what provenance is needed to make a figure auditable. The focus is not on whether visualisation is useful, but on which transformations a reader must be able to reconstruct before a visual pattern can function as evidence.

In 60 Seconds

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 This Question 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.

1. A Diagram Can Change the Work Required

Larkin and Simon showed that informationally equivalent representations can differ in computational efficiency: a spatial diagram may make relations explicit that a sentential description leaves to be derived [1]. This does not mean diagrams are universally superior. Their advantage depends on the task and on which relations the representation makes easy to inspect.

Where That Leaves Us

Larkin and Simon showed that informationally equivalent representations can differ in computational efficiency: a spatial diagram may make relations explicit that a sentential description leaves to be derived [1]. This does not mean diagrams are universally superior. Their advantage depends on the task and on which relations the representation makes easy to inspect. The next section examines the proposition that seeing is also decoding.

2. Seeing Is Also Decoding

Graphs encode data into visual channels. Cleveland and McGill experimentally studied how accurately viewers decode elementary graphical encodings, establishing that the choice of channel affects quantitative judgement [2]. A visualisation therefore has a perceptual error structure as well as a statistical one.

Where That Leaves Us

Graphs encode data into visual channels. Cleveland and McGill experimentally studied how accurately viewers decode elementary graphical encodings, establishing that the choice of channel affects quantitative judgement [2]. A visualisation therefore has a perceptual error structure as well as a statistical one. The next section examines the proposition that summary numbers can hide structure.

3. Summary Numbers Can Hide Structure

Anscombe constructed datasets with nearly identical conventional summary statistics but strikingly different scatterplots [3]. The lesson is not that plots replace statistics. It is that different representations expose different properties, so numerical and visual summaries should be treated as complementary checks.

Where That Leaves Us

Anscombe constructed datasets with nearly identical conventional summary statistics but strikingly different scatterplots [3]. The lesson is not that plots replace statistics. It is that different representations expose different properties, so numerical and visual summaries should be treated as complementary checks. The next section examines the proposition that every view has a task.

4. Every View Has a Task

Munzner's framework separates data abstraction, task abstraction, visual encoding and validation [4]. Ware similarly links display design to properties of human perception [5]. A figure is not simply 'good' in the abstract: it is good or bad for a specified question, audience, data type and decision.

Where That Leaves Us

Munzner's framework separates data abstraction, task abstraction, visual encoding and validation [4]. A figure is not simply 'good' in the abstract: it is good or bad for a specified question, audience, data type and decision. The next section examines the proposition that scale and framing are part of the claim.

5. Scale and Framing Are Part of the Claim

Axes, limits, bin widths, aggregation and normalisation determine what a viewer can see. Changing them may be legitimate, but it changes the representation. The epistemic requirement is not to freeze one universal view; it is to disclose transformations that materially affect interpretation and to inspect plausible alternatives when conclusions depend on them.

Where That Leaves Us

Axes, limits, bin widths, aggregation and normalisation determine what a viewer can see. Changing them may be legitimate, but it changes the representation. The epistemic requirement is not to freeze one universal view; it is to disclose transformations that materially affect interpretation and to inspect plausible alternatives when conclusions depend on them. The next section examines the proposition that colour is not a neutral label.

6. Colour Is Not a Neutral Label

Colour can separate classes and add dimensions, but hue has no universal quantitative ordering. Wong warned specifically that rainbow schemes are problematic for ordered values [6]. Crameri, Shephard and Heron showed that uneven colour gradients can distort apparent structure and can be inaccessible to readers with colour-vision deficiencies [7]. Modern perceptually designed scales improve this, but no palette removes the need to match encoding to task.

Where That Leaves Us

Colour can separate classes and add dimensions, but hue has no universal quantitative ordering. Modern perceptually designed scales improve this, but no palette removes the need to match encoding to task. Crameri, Shephard and Heron showed that uneven colour gradients can distort apparent structure and can be inaccessible to readers with colour-vision deficiencies [7]. The next section examines the proposition that an image is data plus a rendering pipeline.

7. An Image Is Data Plus a Rendering Pipeline

A microscopy image is not identical to the specimen. Acquisition settings, reconstruction, intensity scaling, filtering and display conversion stand between the physical system and the published pixels. Rossner and Yamada emphasised that digital manipulation can cross from legitimate adjustment into inappropriate alteration of the original data [8]. The key distinction is not 'processed versus unprocessed' but whether processing is scientifically justified, consistently applied where required, preserved, and disclosed.

Where That Leaves Us

A microscopy image is not identical to the specimen. The key distinction is not 'processed versus unprocessed' but whether processing is scientifically justified, consistently applied where required, preserved, and disclosed. Rossner and Yamada emphasised that digital manipulation can cross from legitimate adjustment into inappropriate alteration of the original data [8]. The next section examines the proposition that processing can reveal and conceal.

8. Processing Can Reveal and Conceal

The community-developed microscopy checklists, published online in 2023 and issued in Nature Methods in 2024, stress that image-processing and image-analysis workflows mathematically transform image data and should be transparently documented [9]. Adjustments to intensity range, non-linear contrast, filtering, restoration, segmentation or machine-learning steps can improve interpretability, but they can also remove information or generate misleading appearances. Raw or minimally processed data and workflow provenance therefore matter.

Where That Leaves Us

The community-developed microscopy checklists, published online in 2023 and issued in Nature Methods in 2024, stress that image-processing and image-analysis workflows mathematically transform image data and should be transparently documented [9]. Raw or minimally processed data and workflow provenance therefore matter. The next section examines the proposition that a visual pattern is not yet an explanation.

9. A Visual Pattern Is Not Yet an Explanation

A cluster, gradient or apparent boundary in a plot can be a clue. It does not by itself identify mechanism or cause. The display may reflect the underlying system, preprocessing, sampling, projection, smoothing or the encoding itself. Visualisation generates and tests questions; causal interpretation still needs the evidential machinery developed in KNO-014 and KNO-016.

Where That Leaves Us

A cluster, gradient or apparent boundary in a plot can be a clue. It does not by itself identify mechanism or cause. The display may reflect the underlying system, preprocessing, sampling, projection, smoothing or the encoding itself. The next section examines the proposition that multiple views are a form of stress test.

10. Multiple Views Are a Form of Stress Test

Because every encoding privileges some relations, comparing views can reveal dependence on arbitrary choices. A scatterplot, residual plot, histogram and tabular summary can interrogate different aspects of the same dataset [3,4]. Multiple views do not create independent evidence if they derive from the same observations; that final statement is a declared synthesis of this article rather than a claim directly established by those sources.

Where That Leaves Us

Because every encoding privileges some relations, comparing views can reveal dependence on arbitrary choices. Multiple views do not create independent evidence if they derive from the same observations; that final statement is a declared synthesis of this article rather than a claim directly established by those sources. The next section examines the proposition that clarity is not the same as truth.

11. Clarity Is Not the Same as Truth

A clear figure can present a wrong model beautifully. A cluttered figure can contain valid evidence badly. Perceptual effectiveness, statistical validity and scientific interpretation are distinct dimensions. Good visualisation aligns them without treating any one as a substitute for the others.

Where That Leaves Us

A clear figure can present a wrong model beautifully. A cluttered figure can contain valid evidence badly. Perceptual effectiveness, statistical validity and scientific interpretation are distinct dimensions. Good visualisation aligns them without treating any one as a substitute for the others. The next section asks what makes a figure auditable rather than merely persuasive.

12. The Auditable Figure

An auditable scientific visualisation lets a reader identify the source data, transformations, visual encodings and important omissions. For scientific images, that includes acquisition and processing provenance; for graphs, it includes scales, aggregation and uncertainty where relevant [8,9]. Reproducibility of the rendering pipeline is not proof that the scientific claim is true, but opacity about the pipeline makes the claim harder to evaluate.

Where That Leaves Us

An auditable scientific visualisation lets a reader identify the source data, transformations, visual encodings and important omissions. Reproducibility of the rendering pipeline is not proof that the scientific claim is true, but opacity about the pipeline makes the claim harder to evaluate. This defines the article's boundary and prepares the final synthesis without extending the claim beyond what the preceding sections establish.

Claims

C01. Diagrammatic and sentential representations can be informationally equivalent yet differ in computational efficiency for specific tasks. [1]

C02. Graphical encodings differ in perceptual decoding accuracy. [2]

C03. Similar summary statistics can coexist with substantially different data structures visible in plots. [3]

C04. Visualisation design should be evaluated relative to data, task, encoding and validation level. [4]

C05. Perceptual properties constrain effective display design. [5]

C06. Hue is a problematic sole channel for ordered quantitative magnitude, and rainbow maps can mislead. [6,7]

C07. Colour-map accessibility and perceptual uniformity are scientific communication issues, not

merely aesthetic ones. [7]

C08. Digital scientific-image manipulation can alter evidential meaning and must be controlled. [8]

C09. Image-processing and image-analysis workflows transform numerical image data and require transparent reporting. [9]

C10. Legitimate image processing does not make the displayed image identical to the unprocessed acquisition. [8,9]

C11. Multiple visual views can expose structure or assumptions hidden by a single representation, without becoming independent evidence. **Declared synthesis; sources 3 and 4 provide context, not direct support for the independence clause.**

C12. Reproducible rendering improves auditability but does not itself establish the truth of the scientific interpretation. **Declared synthesis; sources 4, 8 and 9 provide context on validation, integrity and reporting, not direct support for the truth-status clause.**

Boundary Note

KNO-018 concerns visual representation as a method of knowing. It does not replace KNO-006 on cartography, KNO-013 on models, KNO-014 on causal identification, KNO-015 on probability/uncertainty, or KNO-019 on data tables and record structures. It also does not claim that one visual grammar is universally optimal.

What We Know, and How

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 Still 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 Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-006 · Maps and Spatial Representation
- KNO-013 · Models, Maps and Reality

- KNO-014 · Causality and Correlation
- KNO-019 · Data, Tables and Records

References

- [1] Larkin, Jill H.; Simon, Herbert A. (1987). "Why a Diagram is (Sometimes) Worth Ten Thousand Words." *Cognitive Science* 11(1):65–100. DOI: 10.1111/j.1551-6708.1987.tb00863.x.
- [2] Cleveland, William S.; McGill, Robert (1984). "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79(387):531–554. DOI: 10.1080/01621459.1984.10478080.
- [3] Anscombe, F. J. (1973). "Graphs in Statistical Analysis." *The American Statistician* 27(1):17–21. DOI: 10.1080/00031305.1973.10478966.
- [4] Munzner, Tamara (2014). *Visualization Analysis and Design*. CRC Press. ISBN: 978-1-4665-0893-4.
- [5] Ware, Colin (2019). *Information Visualization: Perception for Design*. 4th ed. Morgan Kaufmann, an imprint of Elsevier. ISBN: 978-0-12-812875-6.
- [6] Wong, Bang (2010). "Color coding." *Nature Methods* 7:573. DOI: 10.1038/nmeth0810-573.
- [7] Crameri, Fabio; Shephard, Grace E.; Heron, Philip J. (2020). "The misuse of colour in science communication." *Nature Communications* 11:5444. DOI: 10.1038/s41467-020-19160-7.
- [8] Rossner, Mike; Yamada, Kenneth M. (2004). "What's in a picture? The temptation of image manipulation." *Journal of Cell Biology* 166(1):11–15. DOI: 10.1083/jcb.200406019.
- [9] Schmied, Christopher; Nelson, Michael S.; Avilov, Sergiy; et al. (2024; published online 2023). "Community-developed checklists for publishing images and image analyses." *Nature Methods* 21(2):170–181. DOI: 10.1038/s41592-023-01987-9.

KNO-019 · Data, Tables and Records

KNO · Ways of Knowing

Juan Segurola

v0.7 · EN

All Rights Reserved

Article Abstract

Scientific work leaves behind tables, spreadsheets, datasets and archives, but the survival of a file does not by itself tell us what later investigators can learn from it. What must a record preserve if someone years later is to interpret an observation responsibly and independently, distinguish absence from non-detection, identify the version that was analysed and reconstruct transformations between acquisition and publication? This article examines those questions across tabular structure, spreadsheet practice, metadata, provenance, versioning, persistent identifiers and the ambiguous category of “raw data.” It asks where simple storage ends and an auditable scientific record begins, how much context must travel with numerical values, and which parts of a data-producing history need to remain visible for reanalysis, correction and long-term reuse.

In 60 Seconds

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 This Question 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.

1. A record is not the event

A record is a representation left by an event, observation, measurement or decision. It may preserve numbers, text, images, timestamps or categories, but it is not identical to what produced them. **C01.**

Records extend memory by making observations persistent, but persistence does not guarantee interpretability. Declared synthesis: [3] and [5] support provenance, metadata and reconstruction; they do not directly establish the interpretability clause.

This distinction matters because later users usually encounter the record after the original context has disappeared. The epistemic task is therefore not only to store content, but to retain enough information about its production that another person can understand what the content means.

Where That Leaves Us

A record is a representation of an event or observation, not the event itself, so preservation always depends on what was encoded and what was omitted. That claim is conceptual rather than a denial that records can be reliable evidence. The next question is therefore structural: what organization must data have before relationships among observations can be inspected consistently?

2. Data need structure

A collection of values becomes easier to inspect and analyse when its structure is explicit. Wickham's tidy-data formulation is one influential convention: variables occupy columns, observations occupy rows, and different kinds of observational units occupy different tables. [1]

C02. A consistent tabular structure can reduce ambiguity and make transformation, modelling and visualisation easier, but it is a convention for organizing data rather than a universal ontology of the world. [1]

A row is not naturally an “observation” until someone defines the observational unit. The same event can be represented at the level of a person, visit, instrument reading, image, trial or transaction. The table therefore embodies a modelling decision.

Where That Leaves Us

Structured data reduce ambiguity by making variables, observations and values explicit, but no table design is neutral with respect to later operations. The epistemic gain is consistency and inspectability, not automatic correctness. Once structure is explicit, the next question is what tables reveal well and what they can obscure when relationships are compressed into rows and columns.

3. Tables expose relationships and hide others

Tables are powerful because they align comparable values. They allow sorting, filtering, joining, grouping and calculation. Yet a table can conceal the history that produced its cells. A final column may contain a value generated through several transformations whose intermediate steps are absent.

C03. Tabular organization supports comparison and computation, but the table alone need not preserve the provenance of each value. Declared synthesis: [2] supports disciplined tabular organization and [3] supports provenance; neither source directly states the non-preservation clause.

For this reason, a clean table and an auditable record are not the same thing. Good organization helps analysis; provenance helps reconstruction.

Where That Leaves Us

Tables expose comparisons, joins and patterns efficiently while suppressing other relationships that may require different representations. Their usefulness depends on matching structure to the question being asked. That limitation leads to a practical danger: when table-like data are handled in flexible spreadsheets, what conveniences can quietly weaken the integrity and auditability of the record?

4. Spreadsheet convenience can create epistemic fragility

Spreadsheets are useful because they combine data entry, calculation and presentation. The same flexibility can also mix data with formatting, comments, multiple variables in one cell, embedded calculations or inconsistent missing-value codes. Broman and Woo recommend simple rectangular organization, one piece of information per cell, consistent missing-value conventions and separation of raw data from calculations. [2]

C04. Spreadsheet practices that mix raw observations, formatting and derived calculations can make later interpretation and analysis more error-prone. [2]

This is not an argument against spreadsheets. It is an argument for treating the record structure as part of the method.

Where That Leaves Us

Spreadsheets are powerful working environments, but mixed types, hidden formulas, manual formatting and overloaded cells can make data fragile or ambiguous. The risk is not that spreadsheets are inherently invalid; it is that convenience can hide structure. The next question is what information must accompany the values themselves so that later users can still interpret the record correctly.

5. Metadata preserve meaning around values

Metadata describe the data. They may identify variables, units, instruments, dates, locations, creators, methods, software, access conditions or links to related records. NIST's Research Data Framework treats metadata as essential to effective use, reuse and preservation over time. [5]

C05. Values become more reusable when metadata preserve the definitions and conditions needed to interpret them. [5]

A column labelled temperature is incomplete if the scale, sensor, location, timing and processing are material to the question. Metadata do not magically remove uncertainty; they preserve context needed to reason about it.

Where That Leaves Us

Metadata preserve the context around values: variable definitions, units, methods, identifiers and other information needed to interpret what a record means. Complete metadata do not guarantee valid science, but missing context can make valid observations unusable. The next question is historical rather than descriptive: how can a record preserve not only what values are, but how they came to exist?

6. Provenance records how data came to be

The W3C PROV family defines provenance in terms of entities, activities and agents involved in producing a piece of data or other object. It is intended to support interoperable descriptions of origins and transformations. [3]

C06. Provenance records can document where data came from, who or what acted on them, and which processes contributed to the resulting object. [3]

Provenance is therefore different from a descriptive label. A dataset may have excellent variable names yet still fail to show which source file, instrument run, script or transformation produced the version being analysed.

Where That Leaves Us

Provenance records the lineage of data through acquisition, transformation and use, allowing later users to inspect how an object was produced. Provenance improves reconstructability without proving that every transformation was appropriate. Once lineage matters, a further problem appears: if the dataset changes over time, which exact state of that evolving object does a later citation identify?

7. Versioning changes what a citation points to

Data can change after initial creation: errors are corrected, records are added, classifications are revised and processing pipelines are updated. If a claim depends on a particular state of a dataset, citing only the general dataset name can be insufficient.

The Joint Declaration of Data Citation Principles emphasizes specificity and verifiability, including information sufficient to identify a particular version, timeslice or granular portion of data. [4]

C07. When data change over time, version information can be necessary to determine which exact data supported a claim. [4][5]

The knowledge claim is tied not only to a dataset family but to an identifiable manifestation of it.

Where That Leaves Us

Versioning distinguishes successive states of a data object, so a reference can point to what was actually analyzed rather than to an unspecified moving target. A version label helps only if the version is preserved and identifiable. The next question is how persistent identifiers contribute to this problem and, equally important, which documentation problems they cannot solve on their own.

8. Persistent identifiers solve one problem, not every problem

Persistent identifiers help keep references stable when objects move. Data-citation principles recommend globally unique, persistent, machine-actionable identification. [4]

C08. A persistent identifier improves referential stability, but it does not by itself establish data quality, correct interpretation or evidential relevance. [4]

A DOI can tell us which object was cited. It cannot tell us whether the measurements were valid or whether the analysis answered the intended question. Identification and evaluation are separate epistemic tasks.

Where That Leaves Us

Persistent identifiers make references more stable and machine-resolvable, but they do not supply missing metadata, provenance or version semantics by themselves. They solve identification, not the whole problem of interpretability. The next question moves inside the table again: when a cell has no ordinary value, how should the record distinguish missingness from zero, non-detection or inapplicability?

9. Missing values are records too

Absence in a table can mean many things: not measured, not applicable, below detection, lost, refused, impossible, corrupted or truly zero. Treating these as interchangeable can change analysis.

C09. Missingness requires explicit representation because an empty field, a zero and an unobserved value can have different meanings. [2]

The exact coding scheme is domain dependent. The general methodological point is that a record should distinguish different states when those distinctions matter to interpretation.

Where That Leaves Us

Missing values are themselves states that require explicit representation, because different reasons for absence can imply different analyses and interpretations. Encoding missingness carefully does not recover information that was never observed, but it prevents several states from being collapsed into one. The next question is whether the familiar label “raw data” reliably identifies an untouched starting point.

10. Raw data are not always untouched data

The phrase “raw data” often suggests a pristine layer before interpretation. In practice, instruments may already apply calibration, filtering, compression, thresholding or proprietary conversion before a researcher receives a file. NIST's RDaF explicitly treats instruments, processing operations, parameters, workflows and provenance as relevant parts of data management. [5]

C10. The earliest available digital record may already contain processing decisions, so “raw” should not be assumed to mean untransformed. Declared synthesis: [5] documents instruments, processing operations, parameters, workflows and provenance; it does not directly state that the earliest available file is necessarily transformed.

What matters is to document the earliest retained state and the transformations between states.

Where That Leaves Us

So-called raw data may already include calibration, filtering, digitization or other transformations performed by instruments and acquisition systems. “Raw” is therefore a workflow-relative label rather than a guarantee of untouched reality. That makes provenance more important, not less. The next question is what happens when files are preserved successfully but the contextual knowledge required to interpret them disappears.

11. Data can be preserved yet become unusable

A file may remain byte-for-byte intact while the software, schema, codebook, units or provenance needed to interpret it disappear. NIST notes that poor-quality metadata can render important datasets unusable when their creators are no longer available. [5]

C11. Preservation of bytes is not equivalent to preservation of meaning or reusability. [5]

Long-term records therefore depend on documentation, accessible formats, identifiers and contextual metadata, not storage alone.

Where That Leaves Us

Long-term storage can preserve bytes while losing the knowledge needed to interpret variables, processing choices or instrument context. Preservation is therefore necessary but not sufficient for reuse. The epistemic limit is semantic: survival of a file does not ensure survival of meaning. The next question is how much of the observation-to-analysis chain must remain reconstructable for records to support reproducibility.

12. Records support reproducibility only when the chain is reconstructable

Reproducibility depends on more than retaining final numbers. The National Academies' treatment of reproducibility emphasizes access to data and computational information needed to recompute results, while NIST connects reproducibility with documented metadata, code, methods and instruments. [5][6]

C12. Records contribute to reproducibility when they preserve enough of the data, metadata, code, methods and provenance for the relevant result to be reconstructed. [5][6]

This is not a promise that every study can be exactly recreated. Some observations are unique, destructive or historically unrepeatable. The record's role is to preserve the reconstructable path from observation to claim as far as the research context permits.

Where That Leaves Us

Records support reproducibility when another researcher can reconstruct the relevant chain of data, code, transformations, versions and context well enough to repeat the computational path.

Reconstructability still does not guarantee identical scientific interpretation. The next question is practical: what minimum audit should a researcher perform before treating a dataset or table as a durable evidential record?

13. A practical record audit

For an important dataset, ask:

1. What does one row or record represent?
2. What does each variable mean, and what are its units?
3. How are missing, censored and not-applicable values distinguished?
4. Which record is original, which is cleaned, and which is derived?
5. What transformations connect them?
6. Which instrument, method, software and parameters produced the values?
7. Is the exact version used for a claim identifiable?
8. Can another person interpret the data without relying on the creator's memory?

This checklist is an editorial synthesis for this article, not a universal data-management standard.

Where That Leaves Us

A practical record audit checks structure, units, missingness, metadata, provenance, versions, identifiers and the links between transformed outputs and their inputs. Passing such an audit improves inspectability but does not prove the underlying measurements are correct. The next question is one of boundary: which neighboring KNO topics own visualization, evidence synthesis and reproducibility beyond the narrower problem of records?

14. Boundary of this article

KNO-019 concerns the epistemic role of records and structured data. It does not attempt to prescribe a universal database schema, repository policy or statistical workflow. KNO-018 addressed visual representations; KNO-022 will address computation as external memory; KNO-024 will address loss and preservation bias. Here the narrower question is: **what must survive in a record for later reasoning to remain possible?**

Where That Leaves Us

KNO-019 owns the epistemic problem of structured records, metadata, provenance and versioned data objects; it does not replace KNO-018 on visualization or KNO-016 on evidence synthesis, and it only prepares the ground for KNO-021 on open reproducibility. With that boundary fixed, the article can now summarize what records can preserve, what they cannot, and what would change our confidence in them.

What We Know, and How

We know that explicit tabular organization can make variables and observations easier to analyse [1], that spreadsheet conventions can reduce avoidable ambiguity and mixing of raw and derived content [2], that provenance can represent entities, activities and agents involved in producing data [3], and that data citation requires specificity sufficient to identify the object or version being cited [4]. NIST RDaF Version 2.0 explicitly treats metadata, provenance, versioning, instruments, processing operations and reproducibility as parts of research-data management [5]. The National Academies likewise connect reproducibility with access to the data and computational information needed to recompute results [6].

What We Still Don't Know

No universal record structure guarantees future interpretability. Different disciplines require different units, metadata, provenance granularity, retention periods and representations of missingness. Nor can documentation alone establish that measurements were valid or that an analysis answered the right scientific question.

What Could Change Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-016 · *How Evidence Is Built: Putting the Tools Together*
- KNO-018 · *Images, Diagrams and Visualisation*
- KNO-022 · *Computation as External Memory*
- KNO-024 · *Lost Knowledge and Preservation Bias*

References

- [1] Hadley Wickham. "Tidy Data." *Journal of Statistical Software* 59(10), 2014. DOI: 10.18637/jss.v059.i10.
- [2] Karl W. Broman; Kara H. Woo. "Data Organization in Spreadsheets." *The American Statistician* 72(1):2–10, 2018. DOI: 10.1080/00031305.2017.1375989.

[3] World Wide Web Consortium (W3C). *PROV-Overview: An Overview of the PROV Family of Documents*. W3C Working Group Note, 30 April 2013. Stable URI for the W3C Note: <https://www.w3.org/TR/2013/NOTE-prov-overview-20130430/>

[4] Data Citation Synthesis Group. *Joint Declaration of Data Citation Principles*. FORCE11, 2014. Editor: Maryann E. Martone. DOI: 10.25490/a97f-egy.

[5] Robert Hanisch; Debra L. Kaiser; Alda Yuan; Andrea Medina-Smith; Bonnie C Carroll; Eva Campo. *NIST Research Data Framework (RDaF): Version 2.0*. NIST Special Publication 1500-18r2, 7 February 2024. DOI: 10.6028/NIST.SP.1500-18r2.

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

KNO-020 · Review, Consensus and Disagreement

KNO · Ways of Knowing

Juan Segurola

v0.6 · EN

All Rights Reserved

Article Abstract

Science uses several social procedures to expose claims to criticism, but those procedures do not all do the same job. A manuscript may be reviewed before publication, a published result may face replication or correction, and a community may later describe a position as a consensus while substantial questions remain open. Disagreement can arise from evidence, methods, models, assumptions, uncertainty or values, and those sources need not have the same remedy. This article asks how review, consensus and disagreement should be interpreted as evidence about scientific evaluation. It separates first-order evidence from judgments about that evidence, tracks what publication changes and what it does not, and asks what makes agreement or disagreement informative rather than merely visible across scientific communities and changing evidence.

In 60 Seconds

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 This Question 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.

1. Peer review is a filter, not a truth machine

A journal review asks qualified readers to examine a manuscript before an editorial decision. The process can identify unclear reasoning, missing controls, unsupported claims, methodological problems or relevant literature that authors have overlooked. COPE describes reviewers as playing an important role in the integrity of the scholarly record and requires reviews to be objective, constructive and appropriately matched to expertise. [1]

That role is substantial but limited. A reviewer usually sees a manuscript, its reported methods and the evidence supplied by the authors. Reviewers do not rerun every experiment, independently regenerate every dataset or guarantee that unreported assumptions are harmless. Empirical research on peer review has also documented variability in reviewer judgments and concerns about reliability, bias and predictive validity. [2][3]

Claim C01 - Declared synthesis. Peer review is best understood as a structured opportunity for expert criticism before or around publication, not as a procedure that establishes truth. The first part is directly supported by [1][2]; the “not truth” contrast is an interpretive synthesis from the documented scope and limitations of the process.

Where That Leaves Us

Peer review can expose weaknesses before publication, but its reach is bounded by what reviewers can inspect and by the information authors provide. The epistemic status is therefore provisional: review adds structured criticism without certifying truth. Once that limit is clear, the next question is what to make of disagreement when competent reviewers reach different judgments.

2. Disagreement among reviewers is not a malfunction by itself

Two competent reviewers can read the same paper differently. They may specialize in different methods, notice different failure modes or judge the importance of limitations differently. That disagreement can be useful because it exposes places where a manuscript depends on assumptions that were not obvious to all readers.

Editorial practice does not reduce this to vote counting. Nature Genetics explicitly tells reviewers that other reviewers may have different expertise or views, and states that editorial decisions are not made by counting votes or numerical ranks. Editors weigh the arguments raised by reviewers and authors. [4]

Claim C02. Reviewer disagreement can reveal differences in expertise, assumptions and evidential weighting rather than simply indicating that one reviewer is incompetent. [2][4]

Where That Leaves Us

Reviewer disagreement can be informative because different expertise and assumptions expose different vulnerabilities, and editorial decisions need not reduce those judgments to a vote. The evidence supports disagreement as a diagnostic signal, not as proof that a manuscript is sound or unsound. The next question is what changes when a claim passes from review into publication.

3. Publication is a stage in scrutiny, not its endpoint

Once a paper is published, scrutiny broadens. Other researchers may try to reproduce calculations, repeat experiments, test different populations, challenge measurements or compare alternative explanations. The National Academies distinguishes reproducibility using the same data and computational steps from replicability across studies with new data, and emphasizes that failed replication can sometimes expose limitations or open new scientific questions. [5]

Publication therefore changes a claim's social status without freezing its epistemic status. A paper can later be corrected, qualified, contradicted or retracted. Ioannidis' influential 2005 essay formalized how bias, low power, flexibility and prior odds can make published claims unreliable in some settings; the paper itself later received a correction to one table, a useful reminder that publication and subsequent correction belong to the same record. [6][7]

Claim C03. Publication expands the community able to test a claim; it does not end the process by which that claim is evaluated. [5][6]

Where That Leaves Us

Publication makes a claim more widely inspectable and gives it a durable scholarly status, but it does not freeze its epistemic status. Replication, reanalysis, correction and contradiction can still change the balance of evidence. That leaves a further question: if publication is not finality, what exactly should scientific consensus mean when communities later report broad agreement?

4. Consensus is not the same thing as unanimity

Scientific consensus can mean several things, so the term should be used carefully. In formal consensus-study processes, the National Academies describes consensus reports as evidence-based positions produced by expert committees from gathered information and deliberation, followed by independent peer review. [8]

That is different from asking a room to raise hands. The epistemic value comes from the evidence and the procedures used to challenge interpretations, not from unanimity as such. A mature consensus may still contain uncertainty, unresolved subquestions and legitimate minority interpretations.

Claim C04 - Declared synthesis. In scientific assessment, consensus is most informative when it records convergent expert judgment grounded in an explicit body of evidence and critical deliberation. [8] directly supports the institutional model; the generalization beyond that model is a synthesis.

Where That Leaves Us

Consensus is most informative when it records evidence-based expert convergence under procedures that expose reasoning to challenge; it is not equivalent to unanimity or a headcount. That conclusion is well supported institutionally but remains generalizable only with care. The next question is why agreement reached through partly independent evidential routes should carry more weight than repeated agreement built on shared dependencies.

5. Agreement becomes more informative when routes to it are partly independent

Ten papers that reuse the same dataset, assumption or calibration are not equivalent to ten independent lines of inquiry. Convergence is more informative when different measurements, designs, models or populations fail in different ways. This is the same reason replication and methodological diversity matter: shared weaknesses can create shared conclusions.

The National Academies notes that consistency across independent studies can strengthen confidence while also warning that non-replication has multiple possible causes and must be interpreted in context. [5]

Claim C05. The evidential weight of agreement depends partly on how independent the contributing evidence and methods are. This is a declared synthesis grounded in the logic of replication described in [5].

Where That Leaves Us

Agreement gains evidential force when contributing studies, methods or measurements do not all share the same failure modes. Independence is rarely absolute, so the claim is comparative rather than binary. This leaves a practical difficulty: even a strong majority can be wrong, so how should a minority position be treated without either dismissing it or granting it automatic parity?

6. A minority view is not automatically wrong

History contains cases in which minority positions became mainstream after new evidence, better instruments or improved theory. That possibility is important because peer review and expert judgment can be conservative. Benda and Engels found evidence that peer review can perform reasonably on average while also identifying tension between review and innovative ideas, with bias and preconceptions posing risks. [3]

But the symmetrical mistake is just as serious: the existence of a minority view does not show that the majority is poorly supported. A disagreement has evidential force only to the extent that the competing position supplies methods, observations or arguments capable of changing the balance of evidence.

Claim C06 - Declared synthesis. Minority status should neither disqualify a claim nor grant it evidential parity. [3] supports the risk of conservatism; the parity rule is an explicit methodological synthesis.

Where That Leaves Us

Minority status is not an evidential defect by itself, and review systems can be conservative; yet dissent also does not create equal support where evidence is unequal. The warranted position is therefore conditional on methods, observations and arguments rather than social rank. The next question is how disagreement can be localized so that competing interpretations become testable instead of merely opposed.

7. Disagreement is most useful when it can be localized

“Scientists disagree” is often too coarse to be informative. The disagreement may concern the raw observations, how variables were measured, which cases were excluded, which model is appropriate, how uncertainty was propagated, whether a causal interpretation is identified, or how much weight to assign to competing evidence.

Localizing disagreement matters because different disagreements require different tests. A calibration dispute may be settled by better measurement. A model dispute may require a discriminating prediction. A causal dispute may require a different design. A disagreement about values or acceptable risk cannot be settled by pretending it is merely a measurement problem.

Claim C07 - Declared synthesis. The most productive scientific disagreements specify the level at which competing interpretations diverge and what observation or analysis could discriminate between them. This is a methodological synthesis of the review, replication and model-testing logic developed across KNO-013, KNO-016 and the present article.

Where That Leaves Us

A disagreement becomes scientifically useful when we can specify whether it concerns observations, measurement, model choice, causality, uncertainty or values, because different sources of disagreement require different tests. This is a methodological synthesis rather than a universal taxonomy. Once disagreement is localized, the next question is whether a stable consensus should be treated as settled or merely well supported for now.

8. Consensus can be stable without being final

A consensus may remain stable for decades because successive tests continue to support it. It can also change quickly when a new result removes a central assumption or when a new instrument opens a previously inaccessible regime. Stability therefore tells us something about the history of successful challenge, but it does not create logical immunity from revision.

The National Academies' treatment of reproducibility and replicability explicitly allows non-replication to contribute to discovery rather than treating inconsistency as automatically pathological. [5]

Claim C08. A scientific consensus can be well supported and still remain revisable in principle. [5][8]

Where That Leaves Us

A stable consensus can reflect repeated survival under criticism without becoming logically immune to revision. Its stability is evidence about the history of successful challenge, not a guarantee concerning every future observation. That distinction leads naturally to the scholarly record itself: if claims remain revisable after publication, what should we learn when an explicit correction appears?

9. Correction is evidence that scrutiny continued

Corrections and retractions are sometimes presented as evidence that science “failed.” They can instead show that the record remained open to repair. This does not mean every correction is trivial or that institutional self-correction is always fast. It means the relevant question is whether errors can be detected, documented and propagated through later work.

The 2022 correction to Ioannidis' 2005 paper is a simple example: the journal preserved the original article, linked the correction and specified the exact table error. [7]

Claim C09. A visible correction can be evidence of continuing quality control, although the existence of a correction does not by itself measure the reliability of the rest of the record. [7]

Where That Leaves Us

A visible correction can show that scrutiny continued and that a specific error was documented in the public record, although the existence of a correction does not certify the reliability of everything else. The epistemic gain is auditability, not absolution. The next question is therefore what information a claim of review or consensus must expose if readers are to audit it rather than trust a label.

10. Consensus claims need an audit trail

A statement such as “experts agree” is weaker than a statement that identifies who was assessed, what question they considered, what evidence base they reviewed, how disagreements were handled and when the assessment was made. Formal consensus-study processes are valuable partly because they make at least some of that path inspectable. [8]

The same principle applies to review. COPE emphasizes declared conflicts, relevant expertise, confidentiality and constructive reasoning because a review gains legitimacy from how it was performed, not merely from the label “peer reviewed.” [1]

Claim C10. Claims about review or consensus are more auditable when the process, participants, scope and evidence base are explicit. [1][8]

Where That Leaves Us

Claims about review or consensus are more auditable when they identify participants, scope, procedure and evidential basis. That does not guarantee a good outcome, but it makes the route to the judgment inspectable. Once this audit trail is visible, a deeper limit appears: can agreement or disagreement itself substitute for the first-order evidence under discussion?

11. What disagreement cannot do

Disagreement alone cannot make two claims equally probable. It cannot substitute for missing evidence. Nor can consensus alone turn weak measurements into strong ones. Both agreement and disagreement are second-order information: they tell us something about how a community has processed evidence, but they do not replace the first-order evidence itself.

|

Claim C11 - Declared synthesis. Expert agreement and disagreement are evidence about the state of evaluation, not substitutes for the observations, measurements and arguments being evaluated.

Where That Leaves Us

Expert agreement and disagreement are second-order information about evaluation, not replacements for observations, measurements and arguments. This is a declared synthesis and should not be mistaken for a theorem derived from one source. The next question is practical: given that social signals are informative but limited, how should a reader combine them with first-order evidence when encountering a contested claim?

12. A practical reading rule

When you encounter a contested claim, ask four questions:

1. **What is the first-order evidence?** What was measured, observed or derived?
2. **What was the review process?** Who examined the claim, with what expertise and what opportunity to challenge it?
3. **What exactly is agreed or disputed?** Data, method, model, causal interpretation, uncertainty, or value judgment?
4. **What would change the balance?** A replication, a new measurement, a discriminating experiment, a corrected analysis, or a new theoretical constraint?

That procedure does not guarantee the right answer. It prevents the words *peer reviewed*, *consensus* and *controversy* from doing more epistemic work than they can support.

Claim C12 - Declared synthesis. The strength of a review or consensus claim should be evaluated together with the underlying evidence, process and possible discriminating tests.

Where That Leaves Us

A disciplined reader asks what was measured, how the claim was reviewed, where disagreement lies and what evidence could discriminate between alternatives. The rule does not guarantee the right answer; it prevents institutional labels from doing more epistemic work than they can support. The final question is one of scope: which neighboring KNO articles own the evidence-building and reproducibility issues that this article deliberately leaves outside?

13. Boundary and inherited cases

KNO-020 owns the **social evaluation layer**: how review, consensus and disagreement organize criticism after evidence has been produced. It does not replace KNO-016, which asks how evidence itself is assembled, or KNO-021, which will treat openness, reproducibility infrastructure and practices that make checking possible.

Five earlier KNO articles delegated specific disagreement cases here. Their disposition is deliberately limited. KNO-010's disputes over species and rank remain examples of classification disagreement; KNO-011's permissive and restrictive definitions of measurement remain examples of definitional disagreement; KNO-012's "experimenter returns" problem remains a case of disagreement about

controls and intervention; KNO-013's realism versus empirical adequacy remains a model-interpretation dispute; and KNO-016's statement that peer review and consensus belong here is fulfilled by the present article's treatment of social scrutiny. None of those cases is re-litigated here. They remain illustrative cases in their own articles, while KNO-020 supplies the general framework for interpreting the disagreement.

Where That Leaves Us

KNO-020 now closes the delegated social-evaluation cases without duplicating their scientific content: earlier articles retain their specific disputes, while this article supplies the general framework for interpreting review, consensus and disagreement. KNO-016 owns evidence construction and KNO-021 owns open-science and reproducibility infrastructure. With those boundaries fixed, the article can now state what is known, uncertain and revisable about social evaluation.

What We Know, and How

- Structured peer review can expose problems before publication, but reviewer judgments vary and are not certificates of truth. [1][2][3]
- Editorial decisions can incorporate conflicting reviewer advice rather than simply count votes. [4]
- Reproducibility and replicability extend scrutiny beyond publication and can strengthen, qualify or challenge claims. [5]
- Formal scientific consensus processes can combine evidence review, expert deliberation and independent review. [8]
- Corrections can preserve an auditable record of errors found after publication. [7]

What We Still Don't Know

There is no universal numerical conversion from “degree of expert agreement” to probability that a claim is true. Different fields also use different review models, evidential standards and mechanisms for building consensus. The reliability of those processes depends on the question, the evidence, the incentives and the independence of the people and methods involved.

What Could Change Our Minds

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.

If You Remember Only One Thing

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

Related Articles

- KNO-012 · Experiments and Variable Control
- KNO-013 · Models, Maps and Reality
- KNO-014 · Causality and Correlation
- KNO-015 · Uncertainty and Probability
- KNO-016 · How Evidence Is Built: Putting the Tools Together
- KNO-021 · Open Science and Reproducibility

References

1. COPE Council. *COPE Ethical Guidelines for Peer Reviewers*. Version 2, September 2017. DOI: 10.24318/cope.2019.1.9.
2. Bornmann, Lutz. "Scientific peer review." *Annual Review of Information Science and Technology* 45(1): 197-245 (2011). DOI: 10.1002/aris.2011.1440450112.
3. Benda, Wim G.G.; Engels, Tim C.E. "The predictive validity of peer review: A selective review of the judgmental forecasting qualities of peers, and implications for innovation in science." *International Journal of Forecasting* 27(1): 166-182 (2011). DOI: 10.1016/j.ijforecast.2010.03.003.
4. *Nature Genetics*. "Peer Review." Editorial policies, current web policy. No DOI; live publisher policy page.
5. National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press, 2019. DOI: 10.17226/25303.
6. Ioannidis, John P. A. "Why Most Published Research Findings Are False." *PLOS Medicine* 2(8): e124 (2005). DOI: 10.1371/journal.pmed.0020124.
7. Ioannidis, John P. A. "Correction: Why Most Published Research Findings Are False." *PLOS Medicine* 19(8): e1004085 (2022). DOI: 10.1371/journal.pmed.1004085.
8. National Academies of Sciences, Engineering, and Medicine. "Consensus Study Reports" institutional description, National Academies Press. Live institutional page; no DOI for the description itself.

KNO-021 · Open Science and Reproducibility

KNO · Ways of Knowing

Juan Segurola

v0.3 · EN

All Rights Reserved

Article Abstract

Scientific claims become easier to inspect when the materials needed to understand, repeat or challenge them are accessible. Open science is the family of practices that tries to make that inspection possible: sharing data and code where appropriate, describing methods precisely, distinguishing exploratory from confirmatory analysis, preserving versions and making corrections visible. Reproducibility and replicability are related but not identical. A result can be computationally reproducible and still be wrong, while an independent study can fail to replicate for reasons that do not reduce to misconduct or incompetence. This article asks which parts of a scientific result must remain inspectable for later researchers to reconstruct the analysis, test the claim under new data and discover where disagreement enters. It treats openness as an infrastructure for criticism, not as a certificate of truth.

In 60 Seconds

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.[1,2] Preregistration can separate predictions from analyses chosen after seeing the data, but it cannot make a weak design strong.[3] Registered reports move part of evaluation before results are known, while open data and code let others inspect transformations, exclusions and implementation.[4] 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 This Question 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.

1. Reproducibility Is Not Replication

The Question If two researchers obtain the same numerical result, what exactly has been reproduced: the computation, the observation, or the claim? The answer depends on which inputs, procedures and conditions remain fixed, and which are allowed to change during the new test.

The National Academies distinguishes reproducibility from replicability for practical clarity. Reproducibility concerns obtaining consistent results with the same input data, computational steps, methods and conditions of analysis; replicability concerns new studies addressing the same question with newly obtained data.[1] Goodman, Fanelli and Ioannidis likewise show that the vocabulary has been used inconsistently across fields and warn against treating “reproducible” as a synonym for “true.”[2]

This distinction separates two different failure modes. If the same files and code do not regenerate a published table, the problem may be missing dependencies, undocumented preprocessing, version drift or implementation error. If a new study disagrees, the explanation may instead involve sampling, measurement, context, effect heterogeneity or a false original claim. The diagnostic work begins by asking which kind of sameness was expected.

Claim C01. Reproducibility and replicability test different forms of consistency. [1,2]

Claim C10. Reproducibility should not be used as a synonym for truth. [2]

Where That Leaves Us

Reproducibility and replication test different links in the evidential chain. Matching an old computation checks whether the reported analysis can be reconstructed; collecting new data tests whether the claim survives a fresh encounter with the world. Neither result is a truth certificate. The next question is therefore what must be preserved so that reconstruction is possible at all.

2. A Paper Is Not the Whole Analysis

A paper compresses a much larger process. Raw observations are selected, calibrated, cleaned, transformed and summarized before they become a figure or coefficient. Software libraries contribute defaults; exclusion rules remove cases; intermediate files carry state. When those layers disappear, the publication can remain readable while the analysis becomes impossible to reconstruct.

Reproducible work therefore needs more than prose. Depending on the field, it may require data or a controlled-access substitute, code, environment information, parameter choices, calibration records and a mapping from outputs back to inputs. The point is not to preserve every keystroke. It is to preserve enough of the dependency graph that another competent researcher can identify how the reported result was produced.[1,4]

Claim C02. Reproducibility can fail because data, code, dependencies or analysis conditions are not reconstructible. [1]

Where That Leaves Us

The scientific object under review is often larger than the article that describes it. Data, code, parameters and provenance can carry causal responsibility for a result even when they occupy no printed page. Preserving them makes hidden choices inspectable. That leads directly to the next problem: some choices should be recorded before the result is known, not reconstructed afterward.

3. Preregistration Separates Prediction from Retrospection

Data can suggest hypotheses after they have been observed. That is scientifically useful, but it is different from testing a prediction that existed before the observation. Nosek, Ebersole, DeHaven and Mellor describe preregistration as a way to make that temporal distinction visible by recording hypotheses and analysis plans before outcomes are inspected.[3]

Preregistration does not eliminate exploration, and it does not make an analysis correct. Plans may be underspecified, measurements may fail, and justified deviations may be necessary. Its value is that a later reader can see which decisions were prospective and which were responses to the data. That changes the interpretation of evidential strength without forbidding discovery.

Claim C03. Preregistration can make the timing of hypotheses and analysis choices visible. [3]

Claim C04 - Declared synthesis. Preregistration does not by itself establish design quality or truth.

Declared synthesis: [3] supports preregistration as a procedural commitment made before outcomes are known; the further clause that preregistration does not itself establish design quality or truth is the synthesis.

Where That Leaves Us

Preregistration is a timestamped distinction, not a quality seal. It preserves when a hypothesis or analysis choice entered the workflow, so readers can separate prediction from data-dependent adaptation. That can reduce ambiguity about evidential status while leaving exploration legitimate. The next step moves the same logic into publication: can evaluation begin before the result is known?

4. Registered Reports Change the Order of Evaluation

A conventional article is usually reviewed after the main results exist. Registered-report formats move an important part of review earlier: the question, design and planned analysis can be assessed before the outcome is known. Munafò and colleagues include this among reforms intended to reduce opportunities for selective reporting and hindsight-driven storytelling.[4]

Changing the order does not guarantee a useful question or flawless execution. It changes incentives and the evidential record. A null result from a design accepted in principle is less vulnerable to disappearing because it is unexciting, while deviations from the accepted plan can be documented rather than hidden. Publication becomes more closely tied to the quality of the question and method than to the direction of the outcome.

Claim C05. Registered-report style reforms shift evaluation of questions and methods earlier in the research process. [4]

Where That Leaves Us

Registered reports show that openness is partly about chronology: preserving what was decided before outcomes could influence those decisions. They reduce one route by which the published record can become selectively filtered, but they do not remove all filters. The next question is what happens after publication, when another researcher needs access to the material that made the result possible.

5. Open Data and Code Make Transformations Inspectable

Sharing data can allow others to check coding, exclusions, transformations and alternative analyses. Sharing code can expose how a published number was actually computed rather than how the method was described in prose. Together they reduce the distance between the claim and the operations that produced it.[1,4]

But “open” is not enough. A file without variable definitions, provenance or version information may be technically downloadable and scientifically unusable. Code that depends on undocumented software versions may fail years later. Openness therefore needs curation: identifiers, metadata, licences, dependency information and preservation strategies. KNO-019 established why structured records need context; open science adds the requirement that the context remain accessible to criticism.

Claim C06. Open data and code can expose transformations and implementation choices to scrutiny. [1,4]

Claim C07 - Declared synthesis. Open access without metadata, provenance and version context can be scientifically insufficient. **Declared synthesis:** [1] supports the importance of metadata, provenance and version context for reuse; the judgment that access alone can therefore be scientifically insufficient is the synthesis.

Where That Leaves Us

Open materials extend the inspectable object from the paper to the workflow, but access without context is weak evidence infrastructure. Data and code need metadata, versions and provenance if they are to remain interpretable. This turns openness into a stewardship problem as well as a disclosure problem. It also raises a limit: some scientifically important materials cannot responsibly be made public without restriction.

6. Openness Has Legitimate Limits

Human-subject data can contain information that participants did not consent to make public. Ecological locations can expose endangered species. Security-sensitive methods and contractual obligations can restrict dissemination. Open science therefore cannot mean universal public release of every underlying object.

The stronger principle is inspectability under appropriate governance. De-identification, secure enclaves, data-use agreements, synthetic data, controlled repositories and independent access committees can preserve some auditability without pretending that risk is zero. The design problem is to make restrictions explicit and proportionate so that “cannot be public” does not silently become “cannot be checked.”

Claim C08. Legitimate privacy and security obligations can justify controlled rather than unrestricted access. [1]

Where That Leaves Us

Open science is compatible with controlled access when unrestricted disclosure would violate legitimate obligations. The epistemic requirement is not maximal publicity but a documented path by which qualified scrutiny remains possible where feasible. That makes governance part of evidence infrastructure. The next issue is correction: once errors are found, does the record preserve the fact that the scientific object changed?

7. Versions and Corrections Are Part of the Evidence

A corrected dataset, revised script or retracted conclusion is not the same scientific object as the version that existed before the change. Persistent identifiers, versioned repositories and correction notices make that distinction recoverable. Without them, later users may unknowingly combine claims from one version with data or code from another.

Good correction practice therefore does more than replace a file. It preserves enough history to show what changed, why it changed and which conclusions are affected. This is a form of institutional memory. It allows science to change its mind without erasing the path by which the change occurred. KNO-024 will return to the preservation problem from the opposite direction: what disappears when such records do not survive?

Claim C09 - Declared synthesis. Versioning and correction records distinguish changed scientific objects from earlier versions. **Declared synthesis:** [1] supports explicit version and correction context; treating those records as what distinguishes a changed scientific object from its earlier manifestation is the synthesis.

Where That Leaves Us

Versioning makes correction visible as a change in the evidential object rather than a silent replacement. That matters because later researchers must know which data, code and interpretation a claim refers to. A transparent correction history is evidence about the history of evidence. The final question is how all these practices should be judged without turning openness itself into a badge of correctness.

8. Open Science Is an Error-Correction Infrastructure

Open practices do not guarantee that a dataset is unbiased, a model is well specified or a conclusion is true. They change which mistakes can be found and by whom. Munafò and colleagues frame reproducible practice as a system of interacting reforms rather than a single cure.[4] The National Academies similarly treats reproducibility and replicability as properties to investigate, not as simple verdicts on scientific worth.[1]

The strongest case for openness is therefore procedural. A claim accompanied by inspectable materials exposes more possible points of challenge. An unreplicated result can be diagnosed more precisely. A correction can propagate to the objects that depend on it. The gain is not certainty; it is a shorter path from disagreement to diagnosis and from diagnosis to repair.

Claim C11 - Declared synthesis. Open-science practices can shorten the path from disagreement to diagnosis by exposing more of the evidential workflow. **Declared synthesis:** [1,4] support opening records and workflows to inspection and reuse; the claim that this can shorten the path from disagreement to diagnosis is the synthesis.

Claim C12 - Declared synthesis. Open science is best treated as error-correction infrastructure rather than a truth certificate. **Declared synthesis:** [1,2,4] support transparency, reproducibility practices and correction mechanisms; describing open science overall as error-correction infrastructure rather than a truth certificate is the synthesis.

Where That Leaves Us

Open science is best understood as infrastructure for organized criticism. It preserves more of the route from observation to claim, distinguishes some decisions by when they were made, and makes corrections easier to trace. These properties can improve reliability without proving truth. The boundary is now clear: KNO-020 owns review and consensus; KNO-021 owns inspectability, reproducibility and correction practices.

What We Know, and How

- Reproducibility and replicability test different kinds of consistency and should not be treated as synonyms.[1,2]
- Preregistration can make the timing of hypotheses and analysis choices visible.[3]
- Open data, code and provenance can make more of an analysis inspectable, but access without context is insufficient.[1,4]
- Versioning and visible corrections preserve which scientific object a claim refers to.

What We Still 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 Our Minds

- 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.

If You Remember Only One Thing

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

Related Articles

- KNO-019 · Data, Tables and Records
- KNO-020 · Review, Consensus and Disagreement
- KNO-024 · Lost Knowledge and Preservation Bias

References

[1] National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science. National Academies Press, Washington, DC (2019). DOI 10.17226/25303. <https://doi.org/10.17226/25303> [2] Steven N. Goodman; Daniele Fanelli; John P. A. Ioannidis. What does research reproducibility mean?. Science Translational Medicine 8(341), 341ps12 (2016). DOI 10.1126/scitranslmed.aaf5027. <https://doi.org/10.1126/scitranslmed.aaf5027> [3] Brian A. Nosek; Charles R. Ebersole; Alexander C. DeHaven; David T. Mellor. The preregistration revolution. Proceedings of the National Academy of Sciences 115(11), 2600–2606 (2018). DOI 10.1073/pnas.1708274114. <https://doi.org/10.1073/pnas.1708274114> [4] Marcus R. Munafò; Brian A. Nosek; Dorothy V. M. Bishop; Katherine S. Button; Christopher D. Chambers; Nathalie Percie du Sert; Uri Simonsohn; Eric-Jan Wagenmakers; Jennifer J. Ware; John P. A. Ioannidis. A manifesto for reproducible science. Nature Human Behaviour 1, 0021 (2017). DOI 10.1038/s41562-016-0021. <https://doi.org/10.1038/s41562-016-0021>

KNO-022 · Computation as External Memory

KNO · Ways of Knowing

Juan Segurola

v0.3 · EN

All Rights Reserved

Article Abstract

A computer does more than calculate quickly. It can store intermediate states, apply explicit procedures repeatedly, preserve symbolic structures and make operations inspectable outside the limits of unaided human memory. That makes computation a form of external cognitive infrastructure. Yet outsourcing memory and procedure changes what people remember, what they notice and where errors can hide. A saved file can preserve a result while losing the environment that produced it; an automated pipeline can increase consistency while repeating the same mistake thousands of times. This article asks what computation contributes when it is used as an external memory for science and reasoning: persistence, repeatability, scale, shareable representations and executable procedures. It also asks what computation cannot supply by itself: meaningful variables, justified assumptions, trustworthy inputs or an interpretation of why a result matters.

In 60 Seconds

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.[1] 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.[2] Search engines and networked information can shift memory toward knowing where information can be found.[3] In science, executable code can preserve a procedure more precisely than prose, especially when documentation and computation are kept together.[4] 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 This Question 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.

1. Memory Can Be Moved into the Environment

The Question What changes when information needed for a task is stored outside the person performing it? External memory can reduce internal load, but it can also reorganize reasoning, create new dependencies and change what a user must know in order to retrieve, execute and interpret the stored information.

Cognitive offloading uses physical action or external storage to reduce internal cognitive demand. Risko and Gilbert review evidence that people restructure tasks by writing, arranging or storing information when relying on memory alone would be costly.[1] The gain is persistence and shareability, not immunity to error: an external record can be lost, mislabelled or stripped of context.

Claim C01. People use external actions and representations to reduce internal cognitive demand. [1,2]

Where That Leaves Us

External memory changes who can inspect information, how long it can persist and which errors become possible. Writing and storing reduce some cognitive burdens while creating dependence on labels, location and context. Computation extends this tradeoff by making external states executable. The next question is how representations themselves reshape the work of reasoning rather than merely holding facts.

2. External Representations Change the Cost of Thinking

Kirsh argues that external representations do more than save memory: they can change the cost structure of inference, create persistent referents and provide shareable objects of thought.[2] A diagram, table or symbolic layout can make relations simultaneously available that would be cumbersome to maintain internally. Computation adds the ability to transform one external representation into another by explicit rules.

Claim C02. External representations can alter the cost structure of reasoning, not only memory load. [2]

Where That Leaves Us

External representations reorganize operations, not just storage. A table, plot or symbolic structure changes what can be compared at once and which relations become easy to inspect. Computation adds rule-governed transformation between representations. That leads to a stronger form of external memory: a stored state can be paired with an executable procedure that changes it predictably.

3. A Program Stores a Procedure, Not Only a Result

A saved answer records an endpoint; a program can preserve a sequence of operations capable of regenerating that endpoint from specified inputs. Knuth's literate-programming proposal made the link between executable procedure and human explanation explicit.[4] Code can therefore externalize procedural memory with more operational detail than prose alone, although documentation is still needed to connect machine operations to scientific meaning.

Claim C03. Code can preserve an executable procedure in addition to a stored result. [4]

Claim C10. Literate programming links executable procedure with human-readable explanation. [4]

Where That Leaves Us

Code can externalize procedural memory: not only what answer was reached, but how a system can reach it again from inputs. That increases precision and repeatability for complex workflows, provided the code remains interpretable. Execution alone does not reveal why a variable represents what it claims to represent. The next question is what happens when remembered procedure depends on a changing software environment.

4. Exact Files Can Stop Being Executable

A script may remain byte-for-byte unchanged while libraries alter defaults, operating systems remove interfaces, services disappear and formats age. The text of a program therefore does not fully specify its future behaviour unless relevant dependencies and environment are also known. Containers and environment manifests can reduce ambiguity, but long-lived computational memory still requires migration and testing, not just copying.

Claim C04 - Declared synthesis. Executable preservation depends on software and environment context as well as file identity. **Declared synthesis:** [4] supports preserving computational environment information; the claim that executable preservation therefore depends on that context in addition to file identity is the synthesis.

Where That Leaves Us

Digital persistence has layers. Byte identity protects against one kind of change, while executability depends on software and hardware context, and scientific meaning depends on metadata and interpretation. A preserved script can therefore be intact yet unusable. This exposes the next limit of automation: when a procedure does run reliably, it may repeat a mistaken assumption with extraordinary consistency.

5. Automation Multiplies Consistency and Error

Automation applies the same explicit procedure repeatedly without fatigue, making it possible to process volumes no person could handle manually. That consistency is conditional on the procedure being appropriate. A unit-conversion error inside a loop is not diluted by repetition; it is replicated. Computational scale therefore increases the value of tests, validation data, invariants and independent checks because systematic error can be amplified with perfect regularity.

Claim C05 - Declared synthesis. Automation can reduce variability of execution while amplifying systematic procedural error. **Declared synthesis:** no source is assigned as direct support; both the reduction of run-to-run variability and the possibility of consistently repeating the same procedural error are methodological synthesis here.

Where That Leaves Us

Automation changes the distribution of error. It can suppress inconsistency caused by repeated manual execution, yet magnify any mistake embedded in the procedure. The more widely a pipeline is reused, the more consequential its assumptions become. That makes computational memory inseparable from verification: stored procedures need tests that can fail. The next question is what users remember once machines reliably remember for them.

6. Knowing Where Can Replace Remembering What

Sparrow, Liu and Wegner reported experiments in which expectations about later computer access affected what participants remembered, including memory for where information could be found.[3] The result does not show that the internet simply destroys memory. It illustrates a shift in allocation: when external retrieval is expected, internal memory can prioritize access paths over content, creating dependence on the stability of the external system.

Claim C06. Expectations of later computer access can change what people remember about information and its location. [3]

Claim C09 - Declared synthesis. External computational memory creates dependency on identifiers, interfaces and institutions. **Declared synthesis:** [1,3] support distributed cognition and persistent computational resources; describing their use as a dependency on identifiers, interfaces and institutions is the synthesis.

Where That Leaves Us

External access can change what internal memory is used for. People may retain routes to information rather than the information itself when reliable retrieval is expected. This is efficient while the external system persists, but it creates dependency on identifiers, interfaces and institutions. The next question is how those dependencies are coordinated when computation is distributed across people, artifacts and procedures.

7. Computation Can Be Distributed Across People and Artifacts

Hutchins' study of ship navigation treats cognitive work as distributed across people, instruments, charts, procedures and communication rather than located wholly in one navigator.[5] Modern scientific workflows have similar architecture: sensors produce records, scripts transform them, databases preserve state and specialists validate stages. Reliability can emerge from coordinated division of labour, while responsibility and meaning can also become diffuse at handoffs.

Claim C07. Distributed cognitive systems can span people, artifacts and procedures. [5]

Claim C08. No single participant need represent every transformation in a distributed cognitive workflow. [5]

Where That Leaves Us

Computational knowledge can be distributed across humans and artifacts. No individual needs to remember every intermediate state if the system preserves them and coordinates transformations reliably. This expands collective capability while making provenance more important: responsibility and meaning can be lost at handoffs. KNO-023 will examine the human-group side of this architecture; first, KNO-022 must set its boundary with understanding itself.

8. External Memory Does Not Supply Meaning

Licklider imagined a partnership in which machines perform routinizable work while humans formulate goals, hypotheses, criteria and evaluations.[6] The language is dated, but the methodological warning remains useful. A computer can preserve and execute formal operations without deciding whether the inputs answer the right question. Meaning still enters through modelling choices, measurement conventions and interpretation.

Claim C11 - Declared synthesis. Computation can preserve operations without establishing the scientific meaning of their inputs. **Declared synthesis:** [4,6] support preserving workflows and software-dependent operations; the contrast that preserved operations do not by themselves establish the scientific meaning of inputs is the synthesis.

Claim C12 - Declared synthesis. Computation is epistemically strongest when stored states and procedures retain provenance and interpretation. **Declared synthesis:** [2,4,5] support provenance, reproducible procedures and retained computational context; ranking such systems as epistemically strongest when those elements remain interpretable is the synthesis.

Where That Leaves Us

Computation is external memory plus executable transformation, not an autonomous source of meaning. It can preserve states, procedures and coordination across time, but the scientific significance of those objects still depends on justified representations and questions. This sets the series boundary: KNO-022 owns computational offloading and executable memory; KNO-023 turns to how human groups distribute and combine knowledge.

What We Know, and How

- People use external representations and actions to reduce internal cognitive demand.[1,2]
- External representations can change the structure and cost of reasoning, not only storage requirements.[2]
- Expectations of external access can affect what people remember about information and its location.[3]
- Executable code can preserve procedures with greater operational specificity than prose alone.[4]

What We Still 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 Our Minds

- 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.

If You Remember Only One Thing

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

Related Articles

- KNO-008 · Instruments That Extend the Senses
- KNO-019 · Data, Tables and Records
- KNO-021 · Open Science and Reproducibility

References

[1] Evan F. Risko; Sam J. Gilbert. Cognitive Offloading. *Trends in Cognitive Sciences* 20(9), 676–688 (2016). DOI 10.1016/j.tics.2016.07.002. <https://doi.org/10.1016/j.tics.2016.07.002> [2] David Kirsh. Thinking with external representations. *AI & Society* 25, 441–454 (2010). DOI 10.1007/s00146-010-0272-8. <https://doi.org/10.1007/s00146-010-0272-8> [3] Betsy Sparrow; Jenny Liu; Daniel M. Wegner. Google effects on memory: cognitive consequences of having information at our fingertips. *Science* 333(6043), 776–778 (2011). DOI 10.1126/science.1207745. <https://doi.org/10.1126/science.1207745> [4] D. E. Knuth. Literate Programming. *The Computer Journal* 27(2), 97–111 (1984). DOI 10.1093/comjnl/27.2.97. <https://doi.org/10.1093/comjnl/27.2.97> [5] Edwin Hutchins. *Cognition in the Wild*. The MIT Press (1995). DOI 10.7551/mitpress/1881.001.0001. <https://doi.org/10.7551/mitpress/1881.001.0001> [6] J. C. R. Licklider. Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics HFE-1*, 4–11 (1960).

KNO-023 · How Groups Know More Than Individuals

KNO · Ways of Knowing

Juan Segurola

v0.3 · EN

All Rights Reserved

Article Abstract

Groups sometimes solve problems that defeat any one member, but adding people does not automatically add knowledge. Collective performance depends on how information is distributed, whether errors are independent, how disagreement is expressed and how contributions are combined. Diverse problem-solving perspectives can outperform groups selected only for individual ability under particular model conditions, while social influence can reduce the very diversity that makes aggregation useful.[1,2] Laboratory groups can also outperform their best members on some intellectual tasks.[3] This article asks when a group becomes an epistemic system rather than a crowd: when members hold complementary information, when the structure allows that information to surface, when reasons can be compared and when the final decision preserves rather than erases informative disagreement. It treats collective intelligence as conditional organization, not as a property that appears merely because many people are present.

In 60 Seconds

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.[3] Models by Hong and Page show how diversity of problem-solving approaches can matter under specified assumptions.[1] Yet aggregation can fail. Lorenz and colleagues found that social influence can reduce diversity of estimates without necessarily improving collective accuracy.[2] 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 This Question 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.

1. More People Do Not Automatically Mean More Knowledge

The Question What has to happen between possessing different information and producing a better collective judgment? More people can add evidence, search paths and criticism, but interaction can also erase independence, amplify status and collapse diversity before useful alternatives have been tested.

A group contains more total experiences and observations than a typical member, but those resources matter only if they enter the decision. Shared bias can make ten people repeat the same mistake, while unequal speaking time can hide unique information. The useful unit is therefore not headcount but information structure: who knows what, how independent those pieces are and what mechanism combines them.

Claim C01 - Declared synthesis. Group performance depends on information structure, not headcount alone. **Declared synthesis:** [1-4] support effects of diversity, social influence, group size and interaction process; combining them into the general rule that information structure matters more than headcount alone is the synthesis.

Where That Leaves Us

Collective knowledge begins with distribution but depends on integration. More people can add observations, skills and perspectives, yet those additions have no epistemic effect if they are redundant or suppressed. The first design problem is therefore independence: a group benefits from multiple estimates only when members contribute information that is not merely copied from one another. That is where aggregation can either help or fail.

2. Independence Makes Aggregation Informative

When individual errors have different signs, aggregation can cancel some of them. The benefit weakens when people share sources, cues or incentives because errors become correlated. Lorenz and colleagues experimentally showed that social influence could reduce diversity of estimates while increasing individuals' confidence.[2] Convergence can therefore look like improvement even when it removes the variation that aggregation needed.

Claim C02. Social influence can reduce diversity of estimates and undermine some crowd-aggregation benefits. [2]

Where That Leaves Us

Aggregation is powerful when contributions contain partly independent error. Social influence can reduce that independence before estimates are combined, creating confidence without equivalent accuracy. A collective method should therefore distinguish information gathering from social convergence when possible. The next question is broader than averaging numbers: groups also differ in the problem-solving perspectives they bring to a task.

3. Diversity Can Add Search Paths

Hong and Page modelled agents using different problem-solving heuristics and identified conditions under which a diverse group of capable solvers can outperform a group composed only of the individually highest performers.[1] The result is a theorem within a specified model, not a universal law that every form of diversity improves every task. Its value is to show how heterogeneity can create additional search paths.

Claim C03. Diverse problem-solving heuristics can outperform groups of high-ability problem solvers under specified model conditions. [1]

Claim C04. The Hong-Page result is conditional on a formal model and is not a universal diversity theorem. [1]

Where That Leaves Us

Diversity can contribute by creating distinct search paths, but its value is conditional on the task and on members being able to contribute their differences. The Hong-Page result is strongest as a mechanism demonstration, not a blanket policy conclusion. A group must still combine perspectives effectively. Laboratory problem-solving experiments provide one way to test whether groups actually exceed their strongest members under controlled conditions.

4. Some Groups Beat Their Best Member

Laughlin, Hatch, Silver and Boh compared individuals and small groups on structured letters-to-numbers problems. Groups of three, four and five solved the problems in fewer trials than the best individuals drawn from equivalent numbers of participants, whereas groups of two did not outperform the best of two individuals.[3] The task had demonstrably correct answers and allowed exchange of hypotheses and feedback, so the finding shows a group advantage under defined conditions rather than a universal superiority of groups.

Claim C05. Groups of three or more outperformed the best individuals on the letters-to-numbers tasks studied by Laughlin et al. [3]

Claim C06 - Declared synthesis. That group advantage is task-specific and does not establish universal superiority of groups. **Declared synthesis:** [3] directly supports the advantage on the studied letters-to-numbers tasks; the clause denying universal superiority beyond those task conditions is the synthesis.

Where That Leaves Us

Groups can outperform their best members when the task has correctable structure and communication lets partial progress accumulate. That establishes a real collective gain without implying that all tasks behave the same way. The mechanism points toward interaction quality: how members take turns, respond to proposals and keep useful information in play. This is where research on collective intelligence shifts from composition to process.

5. Interaction Quality Is Part of Collective Intelligence

Woolley and colleagues reported a general collective-intelligence factor across a battery of group tasks in their experiments. Performance was associated with factors including social sensitivity and more equal conversational turn-taking, rather than being explained simply by the maximum or average individual intelligence of members.[4] The study shows that patterns of interaction can predict performance in its experimental setting; it does not establish a fixed trait possessed by every group.

Claim C07. Laboratory group performance has been associated with social sensitivity and more equal conversational turn-taking. [4]

Where That Leaves Us

Collective performance depends on process as well as member ability. Experimental work links better group performance to interaction patterns that allow information to circulate rather than remain concentrated in one voice. This makes communication itself an epistemic mechanism. Yet communication can also distort. The next problem is how groups preserve dissent long enough to distinguish useful minority information from noise.

6. Dissent Can Carry Information

A dissenting member may be wrong, but disagreement can also show that the group has not integrated all available information. If consensus is rewarded too early, members can suppress uncertain observations or adapt estimates to the visible majority. Anonymous estimates, independent preliminary judgments and explicit minority reports are attempts to preserve information until it can be evaluated rather than erased by social convergence.[2]

Claim C08 - Declared synthesis. Dissent can preserve information that premature convergence would otherwise erase. **Declared synthesis:** [2] supports the loss of diversity under social influence; treating dissent as a mechanism that can preserve otherwise-erased information is the synthesis.

Where That Leaves Us

Dissent has epistemic value when it preserves information that consensus mechanisms would otherwise erase. Institutions can protect that value by eliciting judgments independently before discussion or by recording minority reasons explicitly. The goal is not permanent disagreement but informative convergence. That raises the next question: collective knowledge often depends on specialization, where no member can independently verify the whole chain.

7. Specialization Creates Distributed Knowledge

Hutchins' analysis of navigation describes cognitive work distributed across specialists, procedures and artifacts.[5] Scientific communities work similarly: instrument specialists, statisticians, field experts and theorists can contribute different parts of one evidential chain. Specialization increases capacity but creates trust dependencies, so standards, documentation and cross-checks at interfaces become part of the evidence infrastructure.

Claim C09. Specialized collective cognition can be distributed across people, artifacts and procedures. [5]

Claim C10 - Declared synthesis. Collective knowledge creates trust dependencies at interfaces between specialized roles. **Declared synthesis:** [5] supports cognition distributed across specialists, procedures and artifacts; the claim that such distribution creates trust dependencies at role interfaces is the synthesis.

Where That Leaves Us

Specialization lets groups know things no member could maintain alone, but it turns interfaces into epistemic infrastructure. Trust must be attached to roles, records and checks rather than to a fiction that every participant verifies everything. This distributed architecture connects KNO-023 to KNO-021 and KNO-022. The final section asks how a group should combine expertise without converting prestige into evidence.

8. Expertise Must Be Weighted Without Becoming Authority

Equal weighting is not always appropriate because a specialist may hold measurements or technical knowledge that others lack. Yet weighting by expertise is different from accepting status as proof. A defensible collective process asks what competence is relevant to the subproblem and combines domain-specific authority with explicit reasons, documented dissent, independent estimates where useful and opportunities for revision.

Claim C11 - Declared synthesis. Expertise should be weighted by relevance to the subproblem rather than treated as status-based proof. **Declared synthesis:** no source is assigned as direct support; both the relevance-based weighting rule and the rejection of status as proof are explicit methodological synthesis.

Claim C12 - Declared synthesis. Collective intelligence is conditional organization that preserves and integrates distributed information. **Declared synthesis:** [1-5] support distinct mechanisms of diversity, independence, interaction and distributed cognition; defining collective intelligence as their conditional organization is the synthesis.

Where That Leaves Us

Groups can outperform individuals by combining independent estimates, diverse search strategies, complementary expertise and error-correcting discussion. The same social mechanisms can also create correlated error, suppression and premature certainty. Collective intelligence is therefore conditional organization, not crowd size. KNO-020 owns how consensus is assessed; KNO-023 owns how human groups distribute, preserve and combine information before consensus is reached.

What We Know, and How

- Groups can outperform their best members on some structured intellectual tasks.[3]
- Diverse problem-solving heuristics can improve group search under specified model assumptions.[1]
- Social influence can reduce diversity of estimates and undermine some wisdom-of-crowds benefits.[2]
- Interaction patterns such as conversational turn-taking are associated with collective performance in laboratory groups.[4]

What We Still 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 Our Minds

- 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.

If You Remember Only One Thing

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

Related Articles

- KNO-009 · Archives, Libraries and Collective Memory
- KNO-020 · Review, Consensus and Disagreement
- KNO-021 · Open Science and Reproducibility

References

[1] Lu Hong; Scott E. Page. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences* 101(46), 16385–16389 (2004). DOI 10.1073/pnas.0403723101. <https://doi.org/10.1073/pnas.0403723101> [2] Jan Lorenz; Heiko Rauhut; Frank Schweitzer; Dirk Helbing. How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences* 108(22), 9020–9025 (2011). DOI 10.1073/pnas.1008636108. <https://doi.org/10.1073/pnas.1008636108> [3] Patrick R. Laughlin; Erin C. Hatch; Jonathan S. Silver; Lee Boh. Groups perform better than the best individuals on letters-to-numbers problems: effects of group size. *Journal of Personality and Social Psychology* 90(4), 644–651 (2006). DOI 10.1037/0022-3514.90.4.644. <https://doi.org/10.1037/0022-3514.90.4.644> [4] Anita Williams Woolley; Christopher F. Chabris; Alex Pentland; Nada Hashmi; Thomas W. Malone. Evidence for a collective intelligence factor in the performance of human groups. *Science* 330(6004), 686–688 (2010).

DOI 10.1126/science.1193147. <https://doi.org/10.1126/science.1193147> [5] Edwin Hutchins. Cognition in the Wild. The MIT Press (1995). DOI 10.7551/mitpress/1881.001.0001. <https://doi.org/10.7551/mitpress/1881.001.0001>

KNO-024 · Lost Knowledge and Preservation Bias

KNO · Ways of Knowing

Juan Segurola

v0.3 · EN

All Rights Reserved

Article Abstract

What survives from the past is not a random sample of what once existed. Bones weather, manuscripts burn, institutions discard records, file formats become unreadable and datasets vanish when storage devices or contact addresses fail. The surviving archive is therefore shaped by processes of preservation, selection and destruction before any historian or scientist begins to interpret it. Archaeology formalized this problem through formation processes that separate past activity from the material record recovered later.[1] Paleontology likewise studies how preservation filters organisms and environments.[2,3] Archives add human decisions about what is created, kept, described and made accessible.[4] Digital information removes some physical constraints but introduces dependence on formats, software, migration and active stewardship.[5,6] This article asks how knowledge claims should change when absence may reflect non-preservation rather than non-existence, and how preservation systems can make their own selection effects more visible.

In 60 Seconds

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. [1] In paleontology, weathering, transport, burial and mineralization make some organisms and environments much more likely to enter the fossil record than others.[2,3] Archives introduce another layer: people and institutions decide what to document, retain, classify and permit others to access.[4] 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.[5] Digital preservation therefore requires migration, integrity checking, metadata and institutional continuity rather than mere copying.[6] 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 This Question 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.

1. The Record Is Produced by Survival

The Question If most traces disappear before anyone can inspect them, what does the surviving record actually sample? The answer depends not only on what happened in the past, but also on which traces were created, preserved, collected, described and made discoverable to later observers.

A past event and a surviving trace are not the same object. Between them lie processes that create, transform, move, preserve and destroy records. Schiffer's distinction between systemic and archaeological context made formation processes central to archaeological inference: objects recovered later have post-use histories that affect their location and association.[1] Before treating a trace as a sample of past reality, one must ask what made that trace likely to survive.

Claim C01. Archaeological evidence is shaped by formation processes between active use and later recovery. [1]

Where That Leaves Us

The surviving record is generated twice: first by the events that create traces, then by the processes that preserve some traces and remove others. Inference must model both. Preservation is therefore not a nuisance added after observation; it is part of the observation process itself. The next section shows how strongly this matters in the physical record of organisms and environments.

2. Taphonomy Filters the Biological Past

Behrensmeyer calibrated stages of bone weathering with known-age carcasses and showed that weathering depends on time and local conditions.[2] Kidwell and Flessa review variation in completeness, temporal resolution and ecological fidelity of fossil records.[3] Fossil abundance is therefore not identical to past abundance: hard parts, rapid burial and depositional setting can alter preservation probability. Taphonomy treats that filter as evidence rather than noise.

Claim C02. Bone weathering varies with time and environmental conditions and can inform taphonomic interpretation. [2]

Claim C03. The fossil record differs in completeness and temporal/ecological resolution across contexts. [3]

Where That Leaves Us

Fossils are filtered observations. Weathering, transport, burial and mineralization change which organisms, body parts and time intervals become visible. Taphonomy does not make the fossil record useless; it makes the preservation process part of the evidence. The same principle applies to human records, but with an additional complication: people and institutions intentionally decide what to document and keep.

3. Archives Are Built, Not Merely Inherited

Schwartz and Cook argue that archives and records participate in the construction of social memory rather than functioning as neutral containers of facts.[4] Decisions about creation, appraisal, description, access and custody shape what later users can encounter. A missing record can therefore mean non-occurrence, but it can also mean non-creation, destruction, exclusion from acquisition, poor description or restricted access.

Claim C04. Archives and records are shaped by human decisions about creation, appraisal, description and access. [4]

Where That Leaves Us

Human archives add intentional selection to physical survival. Records can disappear because no one created them, because institutions did not retain them, because description made them hard to find or because access remained restricted. Archival provenance therefore includes custody and policy. The next issue is asymmetry: official systems often preserve some voices and activities more systematically than informal ones.

4. Missingness Is Often Structured

Preservation bias is most dangerous when missingness correlates with the question being asked. Durable stone and metal survive differently from textiles, wood or speech; formal institutions document differently from informal communities; crises may leave denser paperwork than routine success. The key questions are which categories had opportunities to create traces, which traces were likely to survive and which were likely to be collected and discovered.

Claim C05 - Declared synthesis. Missingness can be structured when survival probability depends on the property being studied. **Declared synthesis:** [1-4] support selective formation, survival and archival processes; describing missingness as structured when survival depends on the studied property is the synthesis.

Claim C06 - Declared synthesis. Absence in an archive is not automatically evidence that an event did not occur. **Declared synthesis:** [1,4] support multiple pathways by which records can fail to form or survive; the inference that archival absence is therefore not automatically event-absence is the synthesis.

Where That Leaves Us

Preservation bias is strongest when the mechanism of loss is related to the phenomenon under study. Then the surviving record can exaggerate durable, official, exceptional or institutionally valued material. Correct interpretation requires a model of who could create records and which records could survive. Digital media seem at first to remove many physical filters, but they introduce their own dependencies and failure modes.

5. Digital Does Not Mean Permanent

Vines and colleagues requested datasets from 516 biological papers spanning two to twenty-two years and found that, among cases where authors reported data status, the odds of a dataset being extant fell by 17 percent per year in their sample.[5] Broken contact addresses and obsolete storage were major obstacles. The estimate is sample-specific, but it demonstrates that digital data can disappear rapidly without institutional preservation.

Claim C07. In the Vines et al. sample, data availability declined strongly with article age. [5]

Claim C08. Digital preservation requires active stewardship beyond simple copying. [5,6]

Where That Leaves Us

Digital information can be copied cheaply, yet persistence still requires active stewardship. Data may vanish through obsolete devices, broken links, missing credentials, format drift or loss of context. The Vines study shows one empirical form of this decay in research data. Preservation therefore shifts from protecting one physical object to maintaining an interpretable chain of files, software, metadata and institutions.

6. Digitization Is Not the Same as Digital Preservation

Conway distinguishes digitization for preservation from digital preservation as a sociotechnical activity. [6] Scanning can reduce handling of an original and widen access, but the scan itself becomes a digital object that requires management. Resolution, colour, compression, metadata and file format determine what the surrogate captures, while migration and validation determine whether that surrogate remains usable.

Claim C09. Digitization and digital preservation are distinct activities. [6]

Claim C10 - Declared synthesis. A digital surrogate preserves only what its capture and metadata choices encode. **Declared synthesis:** [6] supports distinguishing digitization choices from digital preservation; the clause that a surrogate preserves only what capture and metadata encode is the synthesis.

Where That Leaves Us

Digitization creates a new preservation object rather than ending preservation work. What is captured depends on conversion choices, and the resulting files require ongoing management. A digital surrogate can improve access while omitting features of the source or becoming obsolete itself. The next problem follows: preservation systems must make their own selections and transformations visible enough for later users to judge what survived.

7. Preservation Needs Provenance of Loss and Change

A useful archive can record not only what it holds but, where possible, what was transformed, migrated, deaccessioned or known to be missing. Checksums document byte integrity; migration logs record format changes; collection policies explain selection boundaries. These records cannot reconstruct everything that was lost, but they make some preservation processes auditable and help distinguish original evidence from later transformation.

Claim C11 - Declared synthesis. Preservation systems can improve auditability by recording migration, integrity and selection history. **Declared synthesis:** [6] supports the need to manage digital objects across preservation actions; the claim that recording migration, integrity and selection history improves auditability is the synthesis.

Where That Leaves Us

Preservation systems can reduce bias not only by saving more material, but by documenting the filters they apply. Integrity records, migration histories and collection policies expose parts of the survival process to later scrutiny. They cannot recover what was never recorded, yet they can prevent silent transformation from masquerading as original evidence. The final section asks how claims should be phrased when the preservation filter remains only partly known.

8. Absence Requires a Model of Preservation

An absence is strongest evidence when the system was likely to create, preserve and expose the relevant record had the event occurred. It is weaker when any of those probabilities are low or unknown. Archaeologists model formation processes, paleontologists study taphonomy, archivists examine appraisal and custody, and digital preservation studies formats and stewardship. Different domains express the same discipline: interpret the record through the process that made it observable.

Claim C12 - Declared synthesis. Absence is strongest evidence when creation, preservation and discoverability would all have been likely. **Declared synthesis:** [1-6] support distinct creation, survival, archival and digital-preservation filters; combining them into this criterion for when absence is strongest evidence is the synthesis.

Where That Leaves Us

Lost knowledge is not only missing content; it is uncertainty about the filtering process that produced what remains. Absence becomes informative when creation, preservation and discoverability would all have been likely. Where those conditions are weak, confidence should fall accordingly. KNO-024 closes the Ways of Knowing sequence by making preservation itself part of evidence: what survives constrains knowledge only through a model of how survival happened.

What We Know, and How

- Archaeological records are shaped by formation processes between use and recovery.[1]
- Bone weathering and other taphonomic processes affect which biological traces survive and how they are interpreted.[2,3]
- Archives embody decisions about creation, appraisal, description and access rather than functioning as neutral containers.[4]
- Research datasets can become unavailable over time when preservation depends on individual custodians.[5]

What We Still 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 Our Minds

- 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.

If You Remember Only One Thing

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.

Related Articles

- KNO-009 · Archives, Libraries and Collective Memory
- KNO-019 · Data, Tables and Records
- KNO-021 · Open Science and Reproducibility

References

- [1] Michael B. Schiffer. Archaeological Context and Systemic Context. *American Antiquity* 37(2), 156–165 (1972). DOI 10.2307/278203. <https://doi.org/10.2307/278203> [2] Anna K. Behrensmeyer. Taphonomic and ecologic information from bone weathering. *Paleobiology* 4(2), 150–162 (1978). DOI 10.1017/S0094837300005820. <https://doi.org/10.1017/S0094837300005820> [3] Susan M. Kidwell; Karl W. Flessa. The quality of the fossil record: populations, species, and communities. *Annual Review of Earth and Planetary Sciences* 24, 433–464 (1996). DOI 10.1146/annurev.earth.24.1.433. <https://doi.org/10.1146/annurev.earth.24.1.433> [4] Joan M. Schwartz; Terry Cook. Archives, Records, and Power: The Making of Modern Memory. *Archival Science* 2(1–2), 1–19 (2002). DOI 10.1007/BF02435628. <https://doi.org/10.1007/BF02435628> [5] Timothy H. Vines; Arianne Y.K. Albert; Rose L. Andrew; Florence Débarre; Dan G. Bock; Michelle T. Franklin; Kimberly J. Gilbert; Jean-Sébastien Moore; Sébastien Renaut; Diana J. Rennison. The availability of research data declines rapidly with article age. *Current Biology* 24(1), 94–97 (2014). DOI 10.1016/j.cub.2013.11.014. <https://doi.org/10.1016/j.cub.2013.11.014> [6] Paul Conway. Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas. *The Library Quarterly* 80(1), 61–79 (2010). DOI 10.1086/648463. <https://doi.org/10.1086/648463>