

Where Polanyi Meets Galois: Tacit Knowing as an Algebro-Deterministic Operation, Toward Re-rooting AI-Generated Science

Hiroyuki Chuma¹ Kanji Otsuka² Yoichi Sato³ Kodo Yamauchi³

October 12, 2026

Abstract

We argue that tacit knowing is better modelled as an operating regime of a single memory substrate than as a separate kind of content. Our argument draws on VaCoAl, an algebro-deterministic hyperdimensional memory built from XOR-and-shift operations over $\text{GF}(2)$; on CASRAM, its commercial SRAM-CAM realization; and on PyVaCoAl, a DRAM-CAM realization with the extended VaCoAl architecture.

In VaCoAl/PyVaCoAl, the same directed acyclic graph (DAG) of bound records can be read in two ways. With every address collision repaired (Rescue mode), the trace is exactly recoverable and auditable, but no path outranks another. With collisions tolerated (Don't Care mode), a path-integrated confidence, CR_2 , emerges; it prunes circuitous routes and ranks direct ones without ever being stated as a proposition. We call this trade-off, first noted in our companion work on Minsky's Emotion Machine, *audit-valuation complementarity* and propose it as a formal counterpart of Polanyi's claim that tacit and explicit knowledge are not sharply divided.

Five results follow. (1) Polanyi's from-to structure reads as non-commutative binding, his face Identikit as role-filler composition, and recognition as majority-vote coincidence. (2) Explicating the valuation completely destroys it (Lemma 1), an algebraic analogue of Polanyi's pianist; yet the valuation is articulable at the level of law. Tacitness thus comes in two classes: contingent, which a new query brings out, and regime-bound, which can be stated only in a regime where it no longer exists. Read through these classes, Polanyi, Nonaka, Tsoukas and Collins each describe part of one substrate correctly and are narrow in a characteristic way; Nonaka's knowledge conversion, even on the continuum of Nonaka and von Krogh, cannot tell the two classes apart. (3) Collins's relational, somatic and collective tacit knowledge map onto bound structure, substrate valuation and value binding respectively; the last is absent from the substrate. (4) The mentor-student DAG of our earlier computational prosopography is a graph of tacit-knowledge transmission, so its superhighway and its Kuhnian shift are events in how tacit knowledge travels. (5) Terence Tao's observation that AI-solved problems now spawn few follow-up activities is diagnosed as *rootless explicit knowledge*: results that no human holds tacitly. We propose a division of labour that re-roots them—LLMs generate, VaCoAl/PyVaCoAl keeps the auditable trace and its confidence, and the human community values. Virtue epistemology explains why this last role stays human: the intellectual virtues it calls on cannot be offloaded to technology. We close with testable predictions.

¹Professor Emeritus, Hitotsubashi University

²Professor Emeritus, Meisei University

³Shuhari System

Table of Contents

1	Introduction	3
2	Background	4
2.1	Five positions on tacit knowledge	4
2.2	The VaCoAl programme	5
2.3	Tacit knowledge and machines	6
3	Polanyi’s structure of tacit knowing, read algebraically	7
3.1	From-to as directed binding	8
3.2	Subsidiary and focal as two reads of one vector	8
3.3	The Identikit as role-filler composition	8
3.4	Knowing how without telling how	8
3.5	Indwelling and the fixed scaffold	9
4	Audit–valuation complementarity: tacit knowing as an operating regime	9
4.1	A relational definition	9
4.2	The trade-off	10
4.3	The pianist, exactly	12
4.4	Articulation at the level of law	12
4.5	Four positions, one substrate	13
4.6	Externalization with a trace	14
4.7	A conjecture on ineffability	14
5	Collins’s three tacit knowledges as strata of one substrate	14
5.1	Relational: tacit because no one could survey it	14
5.2	Somatic: tacit because of what the memory is	15
5.3	Collective: tacit because it lives in the community	15
5.4	A biological precedent for keeping both readouts	15
6	The mentor–student DAG as a tacit-knowledge transmission graph	16
6.1	What an edge carries	16
6.2	The hourglass as institutionalized socialization	16
6.3	Newton, Thurston, and the follow-up gap	16
6.4	The Monastery Wall as a change of medium	17
6.5	Is a second wall forming?	17
7	Rootless explicit knowledge: a Polanyian diagnosis of the follow-up gap	17
7.1	Definition	17
7.2	A worked case	18
7.3	Why LLM results are rootless by default	19
7.4	A SECI reading of the gap	20
7.5	Scooping, possession and transmission	20
8	Re-rooting: auditable abduction and a three-layer division of labour	20
8.1	Three layers	21
8.2	Why the middle layer has to be exact	21
8.3	Why the third layer is human: a virtue-epistemic reading	22
8.4	Auditable abduction	22
8.5	Re-rooting already happens by hand	22
8.6	A complementary direction	23
9	Predictions, limitations and open problems	23
9.1	Predictions	23
9.2	Limitations	24
9.3	Open problems	24
10	Conclusion	25
	References	25
	Acknowledgments	27

1 Introduction

Polanyi opened his account of “tacit knowing” with a face (Polanyi 1966, 4). We can pick out a familiar face among a thousand, even a million, yet we usually cannot say how we do it. He then described the police Identikit: a witness composes a likeness by choosing a nose, a mouth and a hairline from a stock of drawn features. The kit does communicate a face, but only because the witness can tacitly match each drawn feature against the remembered whole—which is why, in his phrase, “we can know more than we can tell” (Polanyi 1966, 4–5).

The first commercial deployment of the VaCoAl architecture is also a face. Shuhari System’s CASRAM, a content-addressable memory that realizes VaCoAl on SRAM inside an FPGA, is offered in a webcam face-recognition system that identifies up to 30 people at once (Shuhari System n.d.; Shuhari System 2026). In that system a GPU-based feature extractor turns each face into an embedding whose dimensions no one can interpret, and CASRAM matches the embedding by block-wise majority vote, whose agreement can be read off as a confidence. The system knows more than it can tell. Unlike us, it can also show how strongly its parts agreed. This paper takes that coincidence seriously.

The question has become urgent for a reason outside our programme. In July 2026 a counterexample to the Jacobian conjecture in dimension three, found with the Fable AI model, was announced; it can be verified by a brief calculation. Terence Tao observed that, presented this way, the construction looks like a “massive miracle”, and he wrote a “digestion” of it as an exercise to himself (Tao 2026b). In a Mathstodon thread Tao described the broader pattern (Tao 2026a). Responsible AI use could support all the follow-up activities that a solution normally sets off, but often the opposite happens: results come from prompts who cannot answer questions about them or give talks, fewer seminars, workshops and collaborations follow, few newcomers join the field, and open directions are withheld for fear of being scooped. In a guest post on Tao’s blog, Grant Sanderson compressed the point into one line: “every AI-generated proof is born an unsolved exposition problem” (Sanderson 2026).

We read these observations through Polanyi’s dictum that “all knowledge is either tacit or rooted in tacit knowledge” (Polanyi 1969, 144). An AI-produced result that no human holds tacitly is explicit knowledge without a root. It can be checked, but it is hard to teach, extend or value, because those activities draw on tacit knowing. We call such results *rootless explicit knowledge*, and we ask what kind of substrate could help re-root them.

Our answer builds on the VaCoAl programme (Chuma, Otsuka & Sato 2026a–g). VaCoAl is a hyperdimensional memory whose every operation reduces to XOR-and-shift over GF(2). Its binding is exactly reversible, its retrieval is a majority vote with a readable confidence, and its multi-hop traversal of a mentor–student DAG of 25.5 million paths has been demonstrated in the software version of PyVaCoAl. Our companion paper on Minsky’s Emotion Machine observed that audit and valuation are one circuit at opposite settings of one parameter (2026f): in VaCoAl/PyVaCoAl, the same substrate yields either an exactly auditable trace or an emergent valuation of paths, not both at full strength. The step taken here is epistemological: we argue that this trade-off is the right formal model of the tacit–explicit relation.

The paper makes five contributions.

1. **An algebraic reading of Polanyi’s structure of tacit knowing (§3).** The “from–to relation” (attending from individual clues to the whole they form; for example, when we recognize a face from its features) becomes non-commutative binding, “subsidiary awareness” (awareness of clues that we rely on but do not attend to; for example, the feeling in the palm that grips a hammer) becomes bundled-but-unqueried content, “focal awareness” (awareness of the object we attend to; for example, the nail being driven in) becomes unbinding, the Identikit becomes role–filler composition, and recognition becomes majority-vote coincidence (Polanyi 1958, 55–56; 1966, 4–10).
2. **Audit–valuation complementarity (§4).** Repairing every collision makes the trace

exact and erases path valuation; tolerating collisions creates valuation and makes the read approximate. Polanyi’s pianist, who falters when he attends to his fingers (Polanyi 1958, 56), has an exact algebraic counterpart. The trade-off separates contingently tacit structure, which a new query brings out, from regime-bound tacitness, which explication destroys. It also shows why the debate over tacit knowledge has lasted so long: Polanyi, Nonaka and Takeuchi, Tsoukas and Collins each describe part of the substrate correctly and are narrow in a characteristic way; even on the continuum of Nonaka and von Krogh (2009), knowledge conversion cannot tell the two classes apart.

3. **A stratified substrate account of Collins’s three tacit knowledges (§5).** Relational, somatic and collective tacit knowledge map onto bound structure, substrate valuation and value binding, the last absent from the substrate; the two hippocampal pathways give a biological precedent for keeping both readouts.
4. **The mentor–student DAG as a tacit-knowledge transmission graph (§6).** The Leibnizian hourglass, the Kuhnian shift and the Monastery Wall are reread as events in how tacit knowledge travels, and the Newton–Leibniz case as the historical prototype of the follow-up gap.
5. **A diagnosis and a remedy for the follow-up gap (§§7–8).** Rootless explicit knowledge is defined, current evidence on LLM self-explanation is reviewed, and a three-layer division of labour—LLM generation, VaCoAl/PyVaCoAl trace and confidence, community valuation—is proposed as a way to re-root AI results. Virtue epistemology explains why the third layer stays human: the intellectual virtues exercised there cannot be offloaded to technology (Pritchard 2023).

Section 2 reviews the tacit-knowledge debate, the VaCoAl programme and recent claims about tacit knowledge in language models; Section 9 states testable predictions and limitations.

We use “tacit” in the structural sense defined in §4: information that shapes retrieval and inference without being represented as a proposition the system can state. Whether a memory substrate can be a knower in Polanyi’s “personal” sense (Polanyi 1958) lies outside our claims. Our claim is narrower and testable. The structure Polanyi described—an integration that focal scrutiny disrupts, yet which can be articulated at another level—appears in VaCoAl/PyVaCoAl as an exact, parameter-controlled trade-off.

2 Background

Two literatures meet in this paper: a seventy-year debate on whether tacit knowledge can be made explicit, and a seven-paper programme that builds a memory in which that question can be posed algebraically.

2.1 Five positions on tacit knowledge

Polanyi described tacit knowing as a from–to relation: we attend *from* subsidiary particulars *to* a focal whole, and the knower’s participation integrates the two (Polanyi 1958; 1966). He insisted that tacit and explicit knowledge are not sharply divided, because explicit knowledge must itself be tacitly understood and applied (Polanyi 1969). He also held that an art which cannot be specified in detail passes from master to apprentice by example rather than by prescription (Polanyi 1958). Kuhn built on Polanyi when he located much of a scientist’s competence in shared exemplars acquired through practice rather than in explicit rules (Kuhn 1962/1970).

The management literature then split. Nonaka and Takeuchi treated tacit and explicit knowledge as distinct contents that convert into each other in a spiral; in their externalization step, tacit knowledge takes the shape of metaphors, analogies, concepts, hypotheses or

Table 1: Five positions on tacit knowledge.

Position	What makes knowledge tacit	Can it become explicit?
Polanyi (1958, 1966, 1969)	Subsidiary particulars integrated toward a focal whole by the knower	Not sharply divided; explicit knowledge always rests on tacit understanding
Kuhn (1962/1970)	Competence carried by shared exemplars learned in practice	Only partly; exemplars do work that rules cannot
Nonaka & Takeuchi (1995); Nonaka & von Krogh (2009)	Personal, context-bound know-how and mental models; at the far end, embodied knowledge	Yes, by externalization through metaphor, analogy, concept, hypothesis, model; the embodied end cannot be represented (2009)
Tsoukas (2003)	An ineffable dimension of skilled knowing	No; it can be displayed and re-punctuated in dialogue, not converted
Collins (2010)	Three kinds: relational, somatic, collective	Relational: yes. Somatic: by machine reproduction. Collective: not in the foreseeable future

models (Nonaka & Takeuchi 1995). Tsoukas objected that reading tacit knowledge as content awaiting conversion ignores the ineffability Polanyi stressed (Tsoukas 2003). Nonaka and von Krogh (2009) replied that tacit and explicit knowledge lie on one continuum, at whose tacit end embodied knowledge cannot be represented. Collins then showed that the word covers three different things: relational, somatic and collective tacit knowledge (Collins 2010). Relational tacit knowledge is tacit for contingent reasons, so the same item can be tacit at one moment and explicit at another; somatic tacit knowledge can be explicated in the sense of reproduction by a machine; collective tacit knowledge, embedded in society, is the hard case.

Table 1 shows why the debate has stalled. Each position is defensible, and each answers a different question: Polanyi asks about the structure of knowing, Nonaka about organizational practice, Tsoukas about what conversion would destroy, and Collins about where the knowledge resides. What has been missing is a single formal object in which all four questions have answers. We propose that VaCoAl/PyVaCoAl is such an object. In that object, each position turns out to describe one class of tacitness correctly and to be narrow in a characteristic way (§4.5).

2.2 The VaCoAl programme

The programme consists of seven arXiv papers, cited here as 2026a–g in order of submission (Table 2).

Two notes on the list. 2026a has been revised several times; numbers and section references here follow version 5 (21 May 2026). 2026e describes itself as a draft revision of 2026c that develops the argument from Thagard’s viewpoint; we cite both because each contributes a different part of what follows.

VaCoAl (Vague Coincident Algorithm) replaces the random projection of Kanerva’s sparse distributed memory with deterministic diffusion over $\text{GF}(2)$ (2026a). An input polynomial $P(x)$ is partitioned into N blocks; each block b diffuses its segment with its own long-period feedback polynomial $G_b(x)$, implemented as a linear-feedback shift register, and writes an entry address into its own SRAM or DRAM table:

$$\Psi_b(P(x)) = x^m P(x) \bmod G_b(x). \quad (1)$$

Retrieval is a majority vote over the N blocks. A single flipped input bit disturbs only the one or two blocks that contain it, and the avalanche effect scatters their wrong votes across the address space while the clean blocks agree (2026a). The share of agreeing blocks, CR_1 , is a confidence that can be read off every retrieval.

Table 2: The seven papers of the VaCoAl programme.

Ref.	Paper	What this paper takes from it
2026a	Beyond LLMs, Sparse Distributed Memory, and Neuromorphics: VaCoAl (arXiv:2604.11665)	Architecture, Galois-field diffusion, Rescue vs Don't Care, emergent STDP-like selection, mathematical auditability, HDC-AI as a complement to LLMs
2026b	Bridging Silicon and the Hippocampus (arXiv:2605.15652)	Two hippocampal regimes, CR ₂ as a model of multiplicative replay decay, Pearl's rungs 2 and 3
2026c	How to Build Marcus's Algebraic Mind: Substrate over Galois Fields (arXiv:2605.21379)	Operations over variables, recursive structure, individuals vs kinds; reversible binding and non-commutative bundling
2026d	Institutions and the Transmission of Upper-Tail Human Capital (arXiv:2605.31470)	Leibnizian hourglass, 46-fold rise in society membership, possession vs transmission, the Monastery Wall
2026e	How to Build Marcus's Algebraic Mind: Thagard's Brain-Mind Viewpoint (arXiv:2607.16573)	A reversible substrate for deeply nested bindings; demonstrated vs physically realized claims
2026f	How to Build Marcus's Algebraic Mind: Minsky's Emotion-Machine Viewpoint (arXiv:2607.16623)	Introspection as exact recovery of one's own trace; valuation in Don't Care, audit in Rescue
2026g	Computational Prosopography across a Millennium (arXiv:2608.17915)	Tool criticism: every ranking decision is reconstructable, and the macro-structures survive switching the bias off

Binding is XOR-and-shift. With σ a fixed shift, a role R is bound to a filler F and recovered exactly in the noiseless case (2026a; 2026c):

$$\text{Bind}(R, F) = R \oplus \sigma(F), \quad \text{Unbind}(\text{Bind}(R, F), R) = \sigma^{-1}(R \oplus R \oplus \sigma(F)) = F. \quad (2)$$

Because σ acts on the second argument only, $\text{Bind}(R, F)$ differs from $\text{Bind}(F, R)$; the programme builds its non-commutative bundling on this asymmetry, which is how it distinguishes “the dog bites the man” from “the man bites the dog” (2026c).

Multi-hop traversal multiplies per-step confidences into a path integral:

$$\text{CR}_2(n) = \text{CR}_2(n-1) \cdot \text{CR}_1(n-1) \approx \text{CR}_1^n. \quad (3)$$

Two operating modes of VaCoAl/PyVaCoAl matter for everything that follows. In Rescue mode a four-stage circuit repairs every block collision; with full rescue the output is bitwise identical to a Python dictionary, CR₁ is pinned at 1.0, and candidates inside the frontier fall back to lexicographic order. In Don't Care mode collided blocks simply abstain; the small residual variance in CR₁ then accumulates in CR₂ and becomes the only basis for ranking candidates (2026a).

We follow the programme's own division of claims (2026e). Results that depend on the rescue circuit, CR₂ selection or million-dimensional scale are demonstrated in the software version of PyVaCoAl. Results that depend on in-memory speed and power belong to CASRAM, which is physically realized and commercial but not yet benchmarked. Nothing in the present paper depends on the second kind.

2.3 Tacit knowledge and machines

Whether machines can hold tacit knowledge is an older question than this programme. Dreyfus argued that symbolic AI would fail because skilled performance rests on embodied, situated know-how that rules cannot capture (Dreyfus 1972). Collins argued that expert systems cannot

Table 3: Polanyi’s anatomy of tacit knowing and its counterparts in VaCoAl.

Polanyi’s term	What it names	VaCoAl counterpart
From-to relation	We attend from proximal particulars to a distal whole; the relation has a direction	Non-commutative binding $\text{Bind}(R, F) = R \oplus \sigma(F)$; directed DAG edges
Subsidiary awareness	Particulars present and functioning but not attended	Fillers bundled into a composite vector that no query has unbound
Focal awareness	The object of attention	The filler returned by unbinding the composite against a chosen role
Tacit integration	Particulars fuse into a meaningful whole	Bundling and majority vote: the whole is the argmax of many partial agreements
The Identikit	A face communicated by assembling chosen features	Role-filler composition, $\text{Face} = \sum_i \text{Bind}(\text{role}_i, \text{feature}_i)$
Knowing how without telling how	Recognition without stateable rules	Coincidence of votes across blocks: CR_1 reports how much agreement, not why
Indwelling	We dwell in the subsidiaries, as in a probe or a tool	Fixed diffusion polynomials, used on every read and never inspected during use

hold the knowledge that lives in social practice (Collins 1990), and later that deep learning, however good at precedent-based pattern recognition, does not acquire the social understanding on which intelligence depends (Collins 2018).

Language models have reopened the question from both sides. Budding argues that LLMs can acquire tacit knowledge in Martin Davies’s sense—internal structure that admits semantic description, syntactic structure and causal systematicity—although Davies himself denied this of neural networks (Budding 2025). Collins and Thorne reply from the research front. The judgements by which scientists decide which papers to take seriously rest on collective tacit knowledge formed in the spoken discourse of small, closed expert groups, which LLMs cannot currently reach; much of the apparent progress of LLMs, they argue, reflects human discourse being written down and formalized, so that the intelligence lies with the humans (Collins & Thorne 2026).

Our question differs from both. They ask whether a given kind of machine has tacit knowledge, and the machine in question is a learned network whose internals are hard to read. We ask where, in a substrate whose every operation can be inspected, the tacit–explicit boundary falls, and what moves it. The answer developed in §4 agrees with Collins and Thorne that collective tacit knowledge lives in communities (§5.3). It also gives Budding’s Davies-style notion—structure that does causal work without being stated—a counterpart in what §4.1 calls structurally tacit information, and adds an account of what moves the boundary between the two.

3 Polanyi’s structure of tacit knowing, read algebraically

Each element of Polanyi’s anatomy of tacit knowing (Polanyi 1958; 1966) has a counterpart among VaCoAl’s operations, and the correspondence is exact enough to tabulate. Table 3 is a dictionary, not yet an argument; the argument begins in §4, where the dictionary is used to locate the tacit–explicit boundary.

3.1 From-to as directed binding

Polanyi’s relation is asymmetric: the “proximal” term is attended *from*, the “distal” term *to* (Polanyi 1966, 10), and reversing them changes what is known. Binding in VaCoAl has the same asymmetry, because the shift σ marks which argument is the filler; $\text{Bind}(R, F)$ and $\text{Bind}(F, R)$ are different vectors (2026c). The prosopographic graph carries the asymmetry at the level of data. Its edges run from mentor to student, and the 369 bidirectional assertions found among roughly 470,000 records had to be removed before the structure could be traversed as a DAG (2026g). Read through Polanyi, that purification step makes the direction of transmission consistent before anything is inferred from it.

3.2 Subsidiary and focal as two reads of one vector

A composite $H = \sum_i \text{Bind}(R_i, F_i)$ carries every filler at once, and none of them is singled out until a query asks for it. Unbinding H against a role R_k brings F_k forward while the other terms recede into noise near the 0.5 similarity baseline. The programme already uses both kinds of read on the same vectors. The “calculus degree” of a scholar is a holistic read: the Hamming similarity of a whole composite to a bundled reference concept. Extracting a scholar’s FIELD component is an analytic read: one role is unbound and the rest is ignored (2026a; 2026d).

Polanyi’s two kinds of awareness therefore become two kinds of read on one representation. The holistic read answers “how much is this like that?” without naming parts; the analytic read names one part and loses the whole. Polanyi observed that one cannot attend focally to a particular while keeping it in its subsidiary role. In the algebra, the two reads are different operations with different outputs, and the integrated similarity is not recoverable from any single unbound filler.

3.3 The Identikit as role–filler composition

Polanyi used the police Identikit to show that communicating a face still depends on tacit matching: the witness can assemble a likeness from drawn features only because she can match each feature against the remembered face, and she cannot say how she does that (Polanyi 1966). Hyperdimensional computing gives the composition an exact form. A face record is a bundle of bound pairs, $\sum_i \text{Bind}(\text{role}_i, \text{feature}_i)$, and each feature can be recovered by unbinding its role.

The matching, in contrast, is a similarity search in which stored record and probe vote block by block. CASRAM’s face-recognition system divides the work between two machines (Shuhari System 2026). An edge PC with a GPU extracts a feature embedding from each face; an FPGA module carrying CASRAM, connected to it over PCIe, performs the 1:N match; and the interface overlays each identified face with an ID and a confidence. We do not claim that this product binds faces as role–filler pairs.

What the product shows is that the tacit–explicit boundary can be moved. For Polanyi’s witness, the composition could be told and the matching could not. In this system it is the other way round. The matching is done by a memory whose decisions are majority votes with a readable confidence (2026a), while the encoding that produced the embedding is the part no one can state. The boundary has not disappeared; it has moved, and it now runs along a physical link between two pieces of hardware. That is the thesis of §4 in miniature: where the boundary falls depends on the substrate and its operations, not on the knowledge itself.

3.4 Knowing how without telling how

The name VaCoAl, Vague Coincident Algorithm, already states the model of recognition: an answer is what many partial, individually unreliable votes coincide on (2026a). Such a system cannot state a rule for why the winner won. It can, however, state how much its blocks agreed,

which blocks dissented, and how the agreement decayed along a path. Retrieval confidence is the vote share CR_1 , and multi-hop confidence is CR_2 (2026a).

This is a third condition between Polanyi’s poles. The system is articulate about agreement and inarticulate about reasons. People are usually inarticulate about both; we feel more or less sure but cannot count our votes. The partial telling matters for §8, because confidence and trace are exactly what a human needs in order to re-root a result.

3.5 Indwelling and the fixed scaffold

Polanyi described tools and probes as things we come to “dwell in” (Polanyi 1966, 17–18): the blind man feels the end of his stick, not its pressure on his palm. VaCoAl’s per-block feedback polynomials play this role. They are used on every read and write, they shape everything the system can distinguish, and they are never themselves the object of a query. The programme links this design choice to biology: grid-cell codes appear to reach hippocampal CA3 through a fixed projection laid down early in life, onto which later memories are orthogonalized (2026a; 2026d, citing Chandra et al. 2025). On this reading indwelling is not a mystery added to cognition. It is the ordinary condition of any substrate whose orthogonalizer is hardwired and therefore invisible to its own queries.

4 Audit–valuation complementarity: tacit knowing as an operating regime

In VaCoAl/PyVaCoAl the boundary between tacit and explicit is not a property of content. It is set by the operating regime and by the family of read operations available, and the regimes trade exact auditability against emergent valuation.

4.1 A relational definition

Let M be a memory in operating regime ρ (for VaCoAl/PyVaCoAl: the rescue rate RR, memory depth 2^m , block count N and frontier size FS), and let Q be the family of read operations that can be applied to it. We say that information I is:

- **explicit** in (M, ρ, Q) if some query in Q returns I as a stated item;
- **structurally tacit** in (M, ρ, Q) if I changes the output of some query in Q , but no query in Q returns I as an item.

Structurally tacit information then comes in two kinds:

- **contingently tacit** if some enlarged family Q' , run under the same regime ρ , returns I as an item;
- **regime-bound tacit** if every family that returns I as an item requires a regime in which I no longer exists.

In plain words, the definitions rest on three ideas. A *query* is a question put to the memory. An *item* is an answer that states something outright, the way a lookup returns a stored record. The *regime* is the way the memory is being run. In VaCoAl/PyVaCoAl the setting that matters most here is the rescue rate RR, which decides whether address collisions are repaired (Rescue mode) or tolerated (Don’t Care mode). The board games of §7.1 serve as running examples; we preview them here and give the facts and sources there. Read a game program as the memory, the way it plays (for AlphaGo, neural networks guiding a search) as its regime, and what we can ask it as its queries.

Structurally tacit. Information can be at work in a memory without ever being stated by it. We can tell that it is there because the answers come out differently with it than they would without it, yet none of the questions we can ask gets it back as an answer. The knowledge that lets us recognize a face is like this (§1): it decides the answer, “that is my friend”, but it is not part of the answer, and we cannot say what it is. We call such information structurally tacit because its tacitness is a fact about the memory, the way it is run and the questions it can answer, not about the information itself. Take AlphaGo, and suppose the only question we can put to it is “what is your move?” Everything that led it to Move 37, its famous move against Lee Sedol in 2016, is then structurally tacit. The futures it weighed, and the hunches it weighed them with, made it choose that move rather than another, yet the answer itself is a single stone.

Contingently tacit. Some structurally tacit information is unstated only because no one has yet asked the right question. Add a new kind of question, change nothing else about how the memory is run, and the information comes out as an answer. Its tacitness is an accident of which questions happen to be available, which is why we call it contingent; once the new question exists, the information is explicit. AlphaGo’s search kept the futures it weighed in a game tree. Add one new question, “show me your tree”, and those futures come out as stated items, while AlphaGo goes on playing exactly as before. The reasons of IBM’s chess program Deep Blue were of the same kind. They were rules that people had written, so a record of the rules it applied would have stated them.

Regime-bound tacit. Other structurally tacit information exists only while the memory is run in a particular way. As long as the memory runs that way, no new question will bring it out. A question that could state it works only if the memory is run differently, and once the memory is run differently, it no longer contains the information. Asking for it destroys it: it is bound to the regime, which is why we call it regime-bound.

- **The human case** is Polanyi’s pianist (§4.3). His skill is at work while his attention is on the music. If he shifts his attention to his fingers to examine the skill, his playing may falter, and the skill is no longer there to be examined.
- **AlphaGo’s hunches give a board-game picture.** The tree records each hunch only as a number, and the networks say nothing about what lies behind it. Suppose we demanded that what lies behind every judgment be a rule a person can read, as Deep Blue’s rules were. AlphaGo would then judge positions as Deep Blue did, and the hunch would be gone: the sense of who is ahead at a glance that AlphaGo needed in order to win (Graepel 2026). For AlphaGo this is an illustration, not a finding, since whether its hunches could be stated in some other way is an open question.
- **In VaCoAl/PyVaCoAl the corresponding case is exact** (§4.2). The traversal’s preference for direct over circuitous lineages guides the search while collisions are tolerated. It vanishes when every collision is repaired so that each step can be inspected.

The definition makes tacitness relative to a substrate and its operations, as Collins’s relational tacit knowledge is relative to social arrangements. It also turns a recurring question—can a new operation produce explicit knowledge?—into a question about enlarging Q .

Table 4 fills the classes from the programme’s own findings.

4.2 The trade-off

The preference of the traversal for direct over circuitous lineages is the clearest example of regime-bound tacit information. In Don’t Care mode it decides which candidates fill the frontier. Under CR₂ ordering Leibniz ranks second and Euler, J. Bernoulli, Lagrange and Poisson enter the top 20; under the dictionary’s lexicographic ordering their places go to anatomists, surgeons, physicians and humanists, and Leibniz falls to twentieth (2026d). No query returns “prefer

Table 4: The classes of §4.1 filled with findings of the programme.

Class	Example from the programme	What exposes it	Under full rescue
Explicit	A single recorded assertion that one person trained another	Direct lookup	Unchanged
Contingently tacit	The Leibnizian hourglass: 5.3 paths per node upstream, 53.4 downstream	Exhaustive path counting over 25.5 million paths	Survives (2026d, Appendix J)
Contingently tacit	Possession versus transmission: Newton 91st, Leibniz 2nd on the Giant Score	A new composite of unbinding, similarity and path counts	Survives (2026d, Appendix J)
Regime-bound tacit	The preference for direct over circuitous lineages	No query returns it; a closed-form law describes it	Dissolves into lexicographic order
Outside the substrate	That the mainstream of mathematics ought to rank high	Judgement by historians and mathematicians	Not affected

direct routes” as an item. And the regime that would make every step exactly inspectable—full rescue—pins CR_1 at 1.0 and dissolves the preference (2026a; 2026d).

The companion paper on Minsky’s Emotion Machine states the general form: audit and valuation are one circuit at opposite settings of one parameter (2026f). Repair every collision and the trace is exactly auditable but no path outranks another; tolerate collisions and the CR_2 decay that credits direct routes makes the read approximate. We name this *audit–valuation complementarity* and propose it as the formal counterpart of Polanyi’s claim that tacit and explicit knowledge are not sharply divided: one substrate, one dial, two readouts that cannot both be at full strength.

The core of the trade-off can be stated exactly.

Lemma 1 (No valuation without imperfection). *Because $CR_2(n) = CR_2(n-1) \cdot CR_1(n-1)$, with CR_2 starting at 1, a regime that repairs every collision, so that $CR_1 = 1$ at every step, assigns $CR_2 = 1$ to every path, and CR_2 cannot rank candidates. Conversely, two candidates can receive different CR_2 only if $CR_1 < 1$ at some step of at least one of their paths. In the traversal of 2026d, where queries are exact stored vectors, $CR_1 < 1$ arises from collided blocks abstaining, that is, from tolerated collisions.*

Conjecture 1 (Monotone complementarity). *At fixed memory depth, the dispersion of CR_2 among candidates inside the frontier is non-increasing in the rescue rate RR , and it is zero at $RR = 1$ by Lemma 1.*

The lemma is elementary, and that is its point: the valuation exists only because of the imperfection that a full audit removes. The conjecture is what §9 proposes to measure.

One qualification keeps the claim honest. The coarse historical structures—the Leibnizian hourglass, the Islamic convergence, the Monastery Wall—survive in both modes almost bit-identically; the modes differ in the ranking of candidates inside the frontier, which is exactly where CR_2 adds semantic content (2026d, Appendix J). Explicit structure is robust to the dial. Tacit valuation is fragile to explication. Whether valuation contrast falls monotonically as RR rises from 0 to 1 has not been measured, and we list it as a prediction in §9.

One dial sets where the tacit-explicit boundary falls

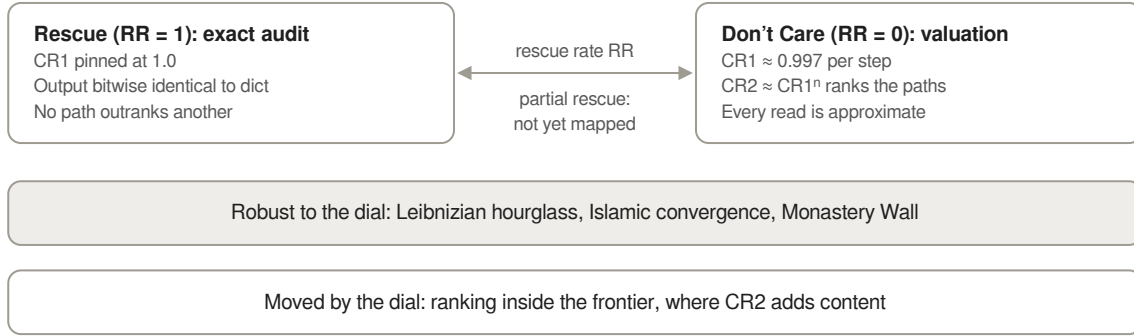


Figure 1: One dial, two readouts. Full rescue yields an exact trace with no path valuation; Don't Care yields valuation with approximate reads. The coarse historical structures do not move with the dial; the ranking inside the frontier does. Intermediate rescue rates have not been mapped.

4.3 The pianist, exactly

Polanyi observed that a pianist who shifts attention from the music to his fingers may falter (Polanyi 1958). Full rescue is the algebraic analogue. The attempt to make every step of the traversal exactly inspectable removes the small per-step imperfection, $CR_1 \approx 0.997$, whose accumulation was the traversal's skill. The change is not metaphorical: the within-frontier ordering switches from a CR_2 ranking with historical content to a lexicographic one without it. In both cases the skill was never stored as an item; it is a property of how the integration runs, and it returns when the integration is allowed to run again—for the pianist when attention goes back to the music, for VaCoAl/PyVaCoAl when RR goes back to 0.

4.4 Articulation at the level of law

The valuation is regime-bound, but it is not ineffable in absolute terms. Its magnitude is predicted in closed form: with $CR_1 \approx 0.997$ at sufficient memory depth, CR_2 at generation 56 should be about $0.997^{56} \approx 0.846$, against a measured value of about 0.905 across 25.5 million paths, with the residual consistent with survivor bias along deep paths (2026d). The system cannot say why it prefers direct routes. Its designers can, as a law about the whole regime rather than as an item retrieved from it. This is the structure of Collins's somatic tacit knowledge: the cyclist cannot state the balancing rule, while physics can (Collins 2010). "We cannot tell" therefore splits into two statements that are both true at once—the system cannot tell at the object level, and an outside theory can tell at the level of law.

Why does the regime produce a valuation at all? Because the search is bounded. On a DAG whose paths grow multiplicatively with depth, an absolute exact match over all paths cannot exist, so VaCoAl/PyVaCoAl builds in a cognitive bound, the frontier size, and ranks the candidates inside it by CR_2 (2026d, Appendix C). Without the bound every path could be kept and none would need to be preferred. This places tacit valuation where Simon placed satisficing: it is what a bounded searcher uses to decide what to keep when it cannot keep everything (Simon 1955). Polanyi's name for what guides a discoverer among too many possibilities was "intellectual passion" (Polanyi 1958, ch. 6); in the substrate, the corresponding guide is a closed-form decay that favours direct routes.

Table 5: The four positions of §2.1, read as descriptions of one substrate. Each is right about the class it describes and narrow in a characteristic way.

Position	Class it describes (§4.1)	What it gets right, in VaCoAl/PyVaCoAl terms	What it leaves out
Polanyi	Every class at once	Tacit and explicit are not separate stores but two readouts of one memory, set by one dial	What sets the boundary in a given substrate, and what moves it
Nonaka & Takeuchi; Nonaka & von Krogh	Contingently tacit; since 2009 also an embodied end that cannot be represented	Contingently tacit structure can be externalized by enlarging the family of operations; metaphor, analogy, hypothesis and model act as new queries	Why something stays unstated: one continuum of content cannot tell structure no query has yet reached from structure the stating regime destroys
Tsoukas	Regime-bound tacit	Regime-bound valuation cannot be converted at the object level; converting it destroys it	Contingently tacit structure, which a new operation does externalize (§4.6)
Collins	Relational: contingently tacit; somatic: regime-bound; collective: outside the substrate	“Tacit” names several things, which sit at different strata of the substrate (§5)	The dial that trades the somatic stratum against exact audit (§4.2)

4.5 Four positions, one substrate

The relational definition of §4.1 also explains why the debate of §2.1 has lasted so long. Each position describes one class of tacitness correctly, and each is narrow in a characteristic way (Table 5).

Nonaka and Takeuchi’s conversion is correct for contingently tacit structure. Their externalization works through metaphor, analogy, hypothesis and model, and each of these acts as a new query put to the same memory run the same way; §4.6 shows the programme doing exactly this. Critics read the theory as treating all tacit knowledge as convertible content (Tsoukas 2003). Nonaka and von Krogh reject that reading: the theory “should not reduce all tacit knowledge to that which can be potentially articulated”, because at one end of a continuum lies embodied knowledge that cannot be represented (Nonaka & von Krogh 2009, 642). Their continuum, however, orders knowledge by how far it can be articulated, and movement along it leaves knowledge “transformed and also enriched” (2009, 642). They describe articulation as “more fluid than a discrete shift from subsidiary to focal awareness” (2009, 641), the shift that, for Polanyi, disrupts action. A continuum of this kind cannot say why something stays unstated. It may be that no query has yet reached it, which is contingent tacitness, or that every regime that could state it is one in which it no longer exists, which is regime-bound tacitness. Only in the second case does explication cost what it explicates.

Tsoukas is right about regime-bound tacitness and narrow in the opposite direction. He denies conversion for all tacit knowledge, yet contingently tacit structure does become explicit when a new operation is added (§4.6).

Collins’s account divides the classes most finely, because his three kinds fall into different classes. Relational knowledge is contingently tacit, somatic knowledge is regime-bound yet articulable at the level of law (§4.4), and collective knowledge lies outside the substrate (§5). What it does not show is the dial that trades the somatic stratum against exact audit. Polanyi’s dictum covers every class at once. What it leaves open is what sets the boundary in a given substrate and what moves it, and that is what the dial of §4.2 supplies.

4.6 Externalization with a trace

The programme has already externalized contingently tacit structure several times by adding operations. The “calculus degree” and the Giant Score are new compositions of binding, bundling, unbinding and Hamming similarity (2026a; 2026d). They turned a feature of the graph into a stated fact: Newton and Leibniz have nearly equal calculus-content scores (0.534 and 0.569) but very different path counts to the 64 Fields Medalists (4 and 47), so Newton ranks 91st on the Giant Score and Leibniz second (2026d). In 2026d we summarized this as the difference between possessing and transmitting upper-tail human capital. Unbinding the FIELD component over 50-year windows recovered a second channel the mentor–student graph cannot show by construction: an astronomy signal of +0.0025 in 1250–1300, aligned with the Toledo and Maragha translation era, and +0.0065 in 1450–1500, while algebra stays near the 0.5 baseline (2026d).

Kanerva’s “What is the dollar of Mexico?” shows the same move in miniature: a mapping vector computed from two country records returns “peso” [without decomposing either record](#) (Kanerva 2010). Nonaka and Takeuchi located externalization in metaphor, analogy, hypothesis and model; hyperdimensional operations are a computational form of exactly those moves. What VaCoAI adds is the trace. Because binding is exactly reversible, every externalized item arrives with its derivation—which bundles, which roles, which paths—so a result that looks like magic can be taken apart afterwards by anyone who holds the substrate.

4.7 A conjecture on ineffability

The semantic pointers that Thagard adopts from Eliasmith bind by circular convolution, whose unbinding is only a quasi-inverse: noise grows with the number and depth of bindings, and a clean-up memory is needed to use the result at all (2026e). In a substrate of that kind a deeply nested integration cannot be taken apart without loss, so it cannot be told. In 2026c we stated the same point in one line: “what cannot be exactly decomposed was never composed, only mixed.” We conjecture that part of what Polanyi called ineffable is a property of a substrate’s read operations rather than of knowledge as such: integration that runs through similarity completion resists articulation, while integration written as exact relational bindings does not. Section 5 gives the hippocampal version of this conjecture, and §9 states how it could fail.

5 Collins’s three tacit knowledges as strata of one substrate

Collins’s three kinds of tacit knowledge sit at three different strata of the substrate, and only two of them lie inside it (Table 6).

5.1 Relational: tacit because no one could survey it

The Leibnizian hourglass was present in Wikidata before anyone measured it. It was tacit for a contingent reason of scale. The graph has 25.5 million paths to the 64 tracer Medalists, a path-to-node ratio of about 68:1, and standard tools fail on it: a Python dictionary exhausted 128 GB of memory at FS = 25,000, the GPU library torchhd requested more than 100 GB of VRAM and stopped, and centrality measures ran but could not tell a mathematician from an anatomist in the same position (2026d, Appendix B). Once an operation existed that could survey the whole graph semantically, the structure became a set of stated facts: 47 of 64 lineages through Leibniz, a 10.1-fold thickness ratio, a 46-fold rise in learned-society membership per scholar across the same window (2026d). This is Collins’s relational tacit knowledge in its purest form: nothing about it resisted telling except the cost of looking.

Table 6: Collins’s three kinds of tacit knowledge as strata of the substrate.

Collins’s kind	Collins’s criterion	Stratum in VaCoAl/PyVaCoAl	What reaches it	Status
Relational	Could be told with effort; tacit for contingent reasons	Bound structure: DAG edges and composite vectors	New read operations (unbinding, bundle matching, multi-hop traversal)	Explicable by enlarging Q (§4.6)
Somatic	The body does it and cannot say how; reproducible by machine	Regime-level valuation: CR_2 decay, Occam-like selection, phase structure across N	A closed-form law about the regime, not a retrieved item	Explicable at the level of law (§4.4)
Collective	Located in society; flexible social performance	Significance for a community: which lineage matters, which result deserves a seminar	Value binding, which the architecture lacks (2026f)	Outside the substrate; held by its users

5.2 Somatic: tacit because of what the memory is

Collins locates somatic tacit knowledge in the physical nature of bodies and brains, and allows that science can explain it and machines can sometimes reproduce it (Collins 2010). VaCoAl/PyVaCoAl’s path valuation is tacit for the analogous reason: it follows from the physical facts of the memory—finite depth per block, residual collisions, the multiplicative structure of CR_2 . The “body” also matters for what the memory comes to prefer. Under a fixed total capacity $N \cdot 2^m = 2^{34}$, a configuration with $N = 128$ settles on a Leibniz-dominant lineage, $N = 256$ on an alternative, Gauss-dominant one, and $N = 512$ purifies the Leibniz–Euler line and alone reaches the full depth of 57 generations (2026d, Appendix F). Different bodies sense the same history differently, and each sensing is historically defensible. Like a cyclist’s balance, none of this is stored as an item, and all of it is explained from outside in closed form.

5.3 Collective: tacit because it lives in the community

The architecture does not know which lineage matters. The Minsky paper is explicit that evaluative self-assessment requires value binding the architecture does not have, and places Minsky’s Self-Reflective layer out of scope (2026f). There is a reflexive point here. Our own claim that the CR_2 ranking is “historically valid”—that the mainstream of mathematics ought to rise to the top 20—is itself a judgement drawn from the collective tacit knowledge of historians and mathematicians. Even the validation of VaCoAl/PyVaCoAl’s somatic-like valuation depends on collective knowledge held outside it. Section 8 builds on this: the community is not an optional add-on to the system but one of its layers.

5.4 A biological precedent for keeping both readouts

The hippocampal paper describes two regimes (2026b). Regime A, the entorhinal–CA3 route, is a scaffolded vector-index substrate in which similarity and attractor completion dominate. Regime B, the trisynaptic route from entorhinal cortex through the dentate gyrus and mossy fibres to CA3, supports sparse, directed relational writes—edge-list or ontology-like commitments—gated by neuromodulation and extended by adult neurogenesis. The brain does not partition cleanly: both regimes contribute to both kinds of query, with task-dependent gain set by acetylcholine and dopamine, and the silicon rescue flag is only a readout-statistics analogue of the biological distinction (2026b). Both pathways have been conserved for more than 520 million years (2026b).

Read through the present paper, the hippocampus keeps both readouts and mixes them by neuromodulatory gain: a similarity-completion regime suited to holistic, hard-to-tell recognition,

and a typed-commitment regime suited to relations that can be stated. This gives the conjecture of §4.7 a biological form. What resists articulation should be what was integrated through Regime-A-like completion; what can be told should be what was written as Regime-B-like typed edges. 2026c adds a caution that matters here: bit-exact reversibility holds in silicon and only approximately under biological noise, so even typed commitments in a brain should leave some residue that cannot be told. We state this as an interpretation, not a finding of 2026b, and return to how it might be tested in §9.

6 The mentor–student DAG as a tacit-knowledge transmission graph

The graph our programme traversed is not a graph of explicit knowledge. Its edges are apprenticeships, the channel Polanyi named for passing on what cannot be specified, so its three macro-structures are events in the history of tacit-knowledge transmission.

6.1 What an edge carries

Each edge asserts that one person trained another; the relation merges Wikidata’s doctoral-advisor, student and student-of properties (2026d, Appendix A). Polanyi held that an art which cannot be specified in detail is transmitted by example from master to apprentice (Polanyi 1958). Collins’s study of laser builders found the same at the bench: groups that had spent time with a working laser’s builders could build one, while groups working only from published accounts could not (Collins 1974). An apprenticeship edge is therefore a socialization channel in Nonaka’s sense, and the topology of the graph records the routes by which tacit knowledge travelled.

The nodes carry something else. Field, language, employer, society membership and dates are explicit attributes, bound into each hub’s composite vector (2026d, Appendix C). The programme’s own results show that the two layers can be read separately. The mentor–student topology cannot see the twelfth-century translation movement, which moved knowledge through books and translators rather than through masters; unbinding the FIELD component of the node vectors recovers it as an astronomy signal (2026d). A DAG memory of this kind thus holds both channels at once: tacit transmission in its edges, codified content in its node vectors. That is the concrete form of a DAG memory that accumulates both kinds of knowledge in one structure.

6.2 The hourglass as institutionalized socialization

Leibniz built an apparatus of transmission: he founded the Berlin Academy of Sciences in 1700, kept one of Europe’s largest scholarly correspondences, and started a lineage running through Johann Bernoulli, Euler, Lagrange, Laplace and Poisson (2026d). In the graph this appears as an hourglass, 5.3 paths per node upstream against 53.4 downstream, and every one of his direct students produced descendants in the top tier of network influence (2026d). Across the same window learned-society membership rose from 6.5 to 82.1 percent of the cohort, from 0.08 to 3.81 memberships per person (2026g; 2026d). We read this as Mokyr’s Republic of Letters, as 2026d does, and add one gloss. Academies, correspondence and lineages are institutions for multiplying socialization. They are, nearly item for item, the follow-up activities Tao lists—talks, meetings with experts, workshops, newcomers joining a field.

6.3 Newton, Thurston, and the follow-up gap

Newton and Leibniz have nearly equal calculus-content scores in the graph, but only four of the 64 lineages pass through Newton, against 47 through Leibniz; Newton, for all his genius, worked largely in isolation and trained few successors (2026d). In 2026d we framed the priority

dispute as possession versus transmission: Newton possessed the calculus, Leibniz possessed and transmitted it. The graph records only one channel, documented mentorship. Newton’s influence travelled abundantly through texts, and nothing here measures it. But the apprenticeship channel is exactly the channel of tacit knowledge, so the finding is about the right thing.

Alongside classical cases like Newton, Thurston recorded a similar human precedent (Thurston 1994). After his classification theorem for foliations, the field began to empty within a couple of years. His papers explained little of the background they presupposed and set down only the key points, so the barrier to entry was high: many graduate students and mathematicians struggled to learn and understand the proofs of key theorems and left the field. He concluded that “more than the knowledge, people want personal understanding.” The follow-up gap is therefore not new; what is new is the rate at which answers can now be produced without anyone acquiring the understanding that used to come with them. The AI-solved results Tao describes are the limiting case: possession without any human possessor, and therefore without the apprenticeship edges along which understanding normally flows (Tao 2026a). Leibniz, Thurston and Tao’s observation are one structure seen at three densities: a result whose transmission channel is thin or absent becomes a node with few outgoing edges, however valuable its content.

6.4 The Monastery Wall as a change of medium

Every traversal stops at the 57th generation, at Lanfranc of Canterbury (1005–1089), who taught at Bec, or at Ivo of Chartres (1040–1116), whose own teachers are unrecorded (2026d). Before them, education was organized in monastic and cathedral schools as a communal affair, and such channels produced no structured mentor–student records; the Wall marks where personal academic mentorship first became record-generating in Europe (2026d; 2026g). In Collins’s terms the Wall is a shift in the medium of tacit transmission, from a communal medium close to collective tacit knowledge to a dyadic medium of apprenticeship. It is not an absence of transmission. It is the point at which transmission changed form and began to leave the kind of trace our instrument can read.

6.5 Is a second wall forming?

If the follow-up gap widens, a future prosopography of twenty-first-century science could meet a boundary of a new kind. Results would be attached to nodes that have few or no human apprentice edges, because what was learned in producing them was learned by no person. We call this hypothetical boundary an *AI wall*. Unlike the Monastery Wall, it would not mark a change from one human medium to another, but a channel in which no human holds the tacit root. Section 9 gives a measurable signature. We stress that this is a prediction, not an observation, and that the re-rooting practices described in §8 are what would prevent it.

7 Rootless explicit knowledge: a Polanyian diagnosis of the follow-up gap

An AI result that is verified but held tacitly by no one is explicit knowledge without a root, and the follow-up gap is what rootlessness looks like from outside.

7.1 Definition

Board games make the definition easy to picture. In March 2016, in the second game of its match against Lee Sedol, one of the greatest Go players of all time, DeepMind’s AlphaGo put a stone on the fifth line of the board. Move 37 looked so strange that some commentators took it for a bug in the program. AlphaGo won the game, and the match 4–1 (Graepel 2026). We

use this move, and IBM’s Deep Blue at chess before it, to illustrate three conditions. When all three hold together, we call a result *R* *rootless*.

1. ***R* is explicit: it is stated openly, and others can check it.** Move 37 was a stone on the board in full view, and the rest of the game put it to the test.
2. **There is no record of the route to *R* that a person can follow step by step.** Here the machines differ.
 - When Deep Blue beat the reigning world champion Garry Kasparov in 1997, it searched six to eight moves deep for each side, scoring 200 million positions a second with rules that people had programmed into it (Graepel 2026). There were far too many steps for anyone to read, but each one applied rules written in human terms.
 - AlphaGo worked differently. Its neural networks supplied hunches about which moves looked promising and which positions looked won, and its search tested those hunches by reading ahead through possible replies. The search kept a record: a game tree with thousands of branches, each a possible future, in which every move and position carried its networks’ judgments (Graepel 2026). The tree can be walked, but the judgments on it come from networks that give no reasons.
 - A large language model typically keeps no such record. The chain of thought it writes out looks like deliberation, but research shows that it is often constructed after the fact, “reaching an answer by one route but reporting another” (Graepel 2026; see §7.3).

Condition 2 is therefore a matter of degree: Deep Blue’s route was long but readable, AlphaGo’s was recorded but rested on its networks’ hunches, and an LLM’s is neither recorded nor reliably reported.

3. **No one yet has the tacit understanding of *R* (the feel for why it works) needed to extend it, teach it, or judge what it is for.** When Move 37 was played, it had almost no roots in human play. AlphaGo’s own policy network, trained to predict the moves of strong human players, gives a measure of this: it estimated the chance that an expert would play the move at roughly one in 10,000 (Graepel 2026).

Polanyi’s dictum implies that a rootless result is not yet complete knowledge. If explicit knowledge must be tacitly understood and applied, a result that nobody understands in that way is true but idle: it can be cited, but not used as knowledge is used. Condition 2 is the one a substrate can change, as the three machines show; AlphaGo’s game tree is the kind of record that the second layer of §8.1 is meant to keep. Conditions 1 and 3 belong to the result and to the community respectively.

Pritchard draws the same line between knowing and understanding: a student can know Pythagoras’ theorem on the testimony of an expert teacher while lacking any grasp of why it holds or how to apply it, and it is understanding, which involves that grasp, that education aims at (Pritchard 2023, 195–197). Ten years after Move 37, Go shows this line at work, with the AI in the teacher’s place. Professional players now train by copying AI moves as closely as they can, “even when the machine’s thinking remains mysterious to them”, and a Go professor observes that “top-tier players haven’t yet been able to deduce the general principles behind AI moves” (Kim 2026). A rootless result is known in the first way: it can be checked or accepted, but no one yet understands it well enough to extend, teach or judge it.

7.2 A worked case

The Jacobian counterexample satisfies all three conditions at the moment of its announcement. Tao noted that it can be verified by a brief calculation, but that the vanishing of every non-constant coefficient of the Jacobian looks like a massive cancellation—1,329 equations against

360 degrees of freedom for a generic degree-seven map—so the example was very unlikely to have been located by brute force (Tao 2026b). His digestion rebuilds the construction from the multiplication of low-degree polynomials and keeps the remaining “miracles” to a minimum, while noting that one of them still lacked a fully satisfactory geometric explanation; he also disclosed that he had used an AI chatbot to discuss the problem and to confirm calculations. Weinreich, who argues for total opposition to AI in mathematics, describes the same example as easy to verify but hard to see the origin of, and reports that tens of sequel papers appeared within a week, many written with AI help (Weinreich 2026).

The policy question that divides Tao, Sanderson and Weinreich is outside the scope of this paper. What matters here is that all three describe the same epistemic object: a result verified as explicit knowledge, whose tacit root has to be grown afterwards by human digestion. Sanderson’s phrase is the precise statement of condition 2: every AI-generated proof begins life as an unsolved exposition problem (Sanderson 2026).

7.3 Why LLM results are rootless by default

Four lines of evidence explain why results produced by large language models tend to arrive rootless.

1. **The derivation exists, but it is not a walkable trace.** A transformer’s forward pass is itself a directed acyclic computation, and interpretability tools can recover parts of it as a computational graph. In Claude 3.5 Haiku, the question about the capital of the state containing Dallas activates an intermediate representation of Texas before the answer Austin; replacing that representation with California changes the answer to Sacramento (Anthropic 2025a). The method captures only a fraction of the computation, and understanding the circuits of a prompt of a few dozen words currently takes hours of human effort (Anthropic 2025a). The DAG is there; the trace is not.
2. **The implicit DAG is directional and patchy.** Models fine-tuned on facts that appear in separate documents fail at two-hop questions without chain-of-thought, at chance level, yet succeed when the facts appear together in training or in the prompt (Balesni, Korbak & Evans 2024). Models trained on “A is B” do not generalize to “B is A”, although they can deduce the reverse when “A is B” is in context (Berglund et al. 2024). Both results describe relations that run one way and compose unreliably inside the weights, in contrast to binding that is exactly invertible (§2.2).
3. **Self-explanation imitates human externalization.** Asked how it added 36 and 59, Claude describes the schoolbook carrying algorithm, while the traced mechanism combines a rough-magnitude path with an exact last-digit path; the researchers suggest this may be because the model learned to explain arithmetic by simulating explanations written by people, while learning to compute by other means (Anthropic 2025a). Introspection is not absent but is weak: under the best protocol, Claude Opus 4.1 detected an injected concept about 20 percent of the time, and when a concept was injected after the fact it accepted a word it had not intended as its own and confabulated a reason (Anthropic 2025b). In Polanyian terms, the model’s explanation is a borrowed externalization, detached from its own integration.
4. **Generation outruns evaluation.** In a blind comparison with more than 100 NLP researchers, LLM-generated research ideas were judged more novel than expert ideas ($p < 0.05$) and slightly less feasible, and the study identified failures of LLM self-evaluation and a lack of diversity in generation as open problems (Si, Yang & Hashimoto 2024).

Taken together, these findings support a precise version of the claim that motivated this paper. LLMs can produce abductive hypotheses—Peirce’s term for the only inference that introduces a new idea (Peirce 1903)—by composing knowledge along implicit paths. But each

hypothesis arrives without a walkable trace, with an explanation that may imitate human explanation rather than report the process, and without reliable self-assessment of its significance. It is rootless by default, and its significance is noticed only when someone outside the model supplies the evaluation.

Rootlessness is a default, not a law. Condition 3 of §7.1 concerns people, not models: a result stops being rootless as soon as someone acquires the understanding needed to extend it, teach it and judge it, as the digestions described in §8.5 show. What makes a result rootless is autonomous generation that no one has yet digested.

7.4 A SECI⁴ reading of the gap

In Nonaka and Takeuchi’s spiral, combination joins explicit knowledge to explicit knowledge; internalization turns explicit knowledge into tacit know-how by doing; socialization passes tacit knowledge between people. Autonomous AI problem-solving performs combination at very high speed. It skips the two moves that need people. The follow-up activities Tao lists—talks, questions after talks, workshops, collaborations, newcomers—are where internalization and socialization happen (Tao 2026a). When they do not happen, the spiral stalls after combination, and explicit results accumulate faster than anyone’s tacit understanding of them. Collins and Thorne locate the same gap at the research front: the judgements that separate serious work from noise rest on collective tacit knowledge formed in the spoken discourse of closed expert groups, which LLMs cannot currently reach (Collins & Thorne 2026). Talks, workshops and collaborations are that discourse.

7.5 Scooping, possession and transmission

Tao’s last symptom is that promising open directions are withheld for fear of being scooped (Tao 2026a). The historical analysis of §6 suggests why this matters. When generation is cheap, the scarce assets are a direction and the tacit sense of why it is worth pursuing, so people guard them. Guarding them, however, starves transmission, which is what 2026d found to separate Leibniz’s legacy from Newton’s in the documented record. A credit system that rewards possession pushes a field toward Newton’s thin thread; one that rewards transmission pushes it toward Leibniz’s hourglass. Sanderson’s proposal to credit motivated explanations, and Tao’s remark, quoted by Weinreich (2026), that prestige should move toward the people who verify and digest AI-generated proofs, can both be read as attempts to reward transmission again (Sanderson 2026).

8 Re-rooting: auditable abduction and a three-layer division of labour

A rootless result can be re-rooted if it passes through three layers in turn—a generator, a walkable trace with its confidence, and a community that values it—and VaCoAI/PyVaCoAI can supply the middle layer, which is the one now missing.

⁴SECI stands for the four modes of knowledge conversion in Nonaka and Takeuchi’s theory of organizational knowledge creation (Nonaka & Takeuchi 1995, 62–73): socialization (tacit to tacit), sharing experience, as apprentices learn a craft by observation, imitation and practice; externalization (tacit to explicit), articulating tacit knowledge as metaphors, analogies, concepts, hypotheses or models; combination (explicit to explicit), systematizing and combining separate bodies of explicit knowledge; and internalization (explicit to tacit), embodying explicit knowledge as tacit know-how through learning by doing. Repeated cycles of the four modes form the knowledge spiral, which carries knowledge from individuals to groups and organizations.

Re-rooting puts an exact trace between generation and valuation

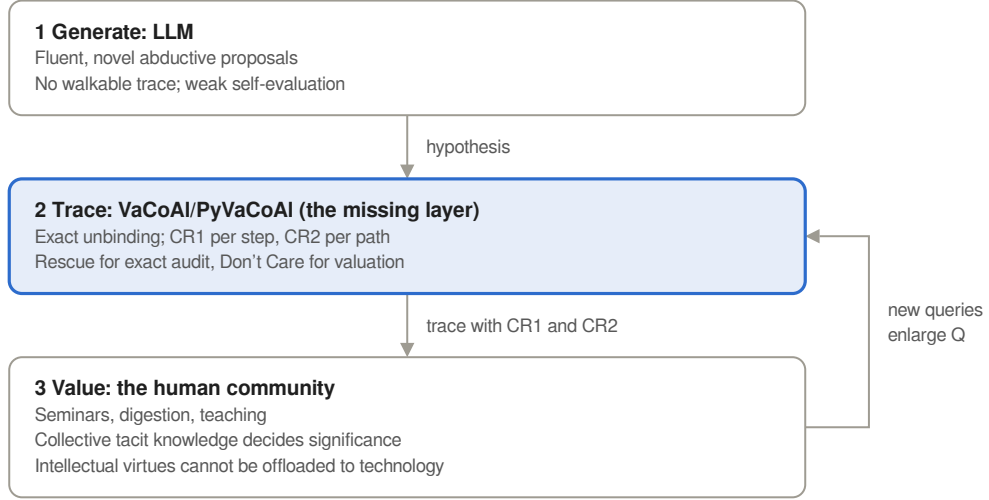


Figure 2: Three layers, one feedback loop. A result moves from generation to valuation through an exact trace; questions raised by the community return to the trace layer as new queries. The third layer is where people exercise the intellectual virtues that technology cannot take over (§8.3).

8.1 Three layers

The first layer generates. LLMs propose hypotheses, constructions and candidate connections fluently and with measurable novelty (Si, Yang & Hashimoto 2024). Their outputs are the abductive raw material.

The second layer keeps the trace. A candidate is re-expressed as bound records and multi-hop paths in a VaCoAI/PyVaCoAI memory, where binding is exactly reversible and every retrieval carries CR_1 and every path CR_2 (2026a). The Minsky paper describes what this buys: a trace readable constituent by constituent, a self-report faithful by construction rather than by verbalization, and credit assignment by counterfactual unbind–rebind (2026f). The prosopography paper shows the same property in use: because the traversal is algebraically reversible, every ranking decision it makes can be reconstructed afterwards (2026g). The dial of §4 is available to the user of this layer, not hidden from them: Rescue mode for exact audit, Don’t Care mode for valuation, and the difference between the two read as information.

The third layer values. Which result matters, which deserves a seminar, which should be taught to the next cohort: these judgements draw on collective tacit knowledge and stay with the community (§5.3). This is also where internalization and socialization happen, so it is where the tacit root actually grows. In the terms of virtue epistemology, it is also where people exercise the intellectual virtues that technology cannot take over (§8.3).

Questions raised in the third layer return to the second as new queries. That loop is how contingently tacit structure gets externalized in practice (§4.6): the community asks, the substrate is given a new operation, and something that was structurally tacit becomes a stated, traceable fact.

8.2 Why the middle layer has to be exact

The two-hop and reversal results point to a precise complementarity. LLMs compose facts reliably when the chain is in front of them and unreliably when it must be assembled inside the weights (Balesni, Korbak & Evans 2024; Berglund et al. 2024). A multi-hop VaCoAI/PyVaCoAI traversal produces exactly what the model lacks: the intermediate nodes as stated items, in order,

with a confidence for each step. This is the concrete content of the division of labour proposed in 2026a, linguistic fluency on one side and multi-hop logical integrity on the other. Exactness matters because the trace is for people. A trace that is itself approximate in an unknown way would reproduce the problem of §7.3 one level down; a trace whose only approximation is a closed-form CR_2 decay can be read with its error bars attached.

8.3 Why the third layer is human: a virtue-epistemic reading

Virtue epistemology explains knowledge as a cognitive success creditable to the knower’s cognitive abilities, whether innate faculties such as perception, memory and reasoning, or acquired intellectual virtues such as conscientiousness, which involve a desire for truth (Pritchard 2023, 54–59). Pritchard’s discussion of technology draws a contrast that fits the first two layers. Using a calculator puts an intellectual distance between a person and the answer: the device acts as an insentient “testifier”, and how the answer was reached may stay hidden, whereas working it out oneself makes the route visible (Pritchard 2023, 187–188). An LLM’s result arrives in the first way. A VaCoAl/PyVaCoAl trace shortens the distance: a reader can follow the route step by step, with CR_1 at each step and CR_2 for the whole path, and so check the derivation rather than take the answer on trust.

The third layer supplies what the trace cannot. Intellectual virtues such as intellectual humility, conscientiousness and intellectual tenacity require motivation, effort and reflection, so they cannot be offloaded onto technology, however much else can be; they also direct how a person’s other cognitive abilities are used (Pritchard 2023, 189–191). They are, moreover, what lead people to seek understanding rather than rest content with knowing (Pritchard 2023, 197–200). Seminars, refereeing, digestion and teaching are where these virtues are exercised, and where a verified result comes to be understood. The substrate can shorten the distance between a person and a result; only the person can close it.

8.4 Auditable abduction

Pearl’s procedure for counterfactuals begins with abduction: infer the exogenous variables consistent with the evidence, then intervene by setting a variable and cutting the arrows into it, then predict the outcome in the modified model (Pearl 2009; Pearl, Glymour & Jewell 2016). The hippocampal paper argues that reversible $GF(2)$ binding supplies the surgical algebra of the do-operator and that the dual-orthogonalizer architecture supplies the parallel substrate counterfactuals need (2026b). Combined with the three layers, this gives a design we call *auditable abduction*. An LLM proposes a hypothesis in Peirce’s sense. The hypothesis is written into the DAG memory as bound records. Its consequences, and the counterfactual “what if this edge were absent?”, are computed by unbind–rebind, and the change shows up as a change in CR along named paths. The human reading the result sees which bindings the hypothesis depends on before deciding what it is worth.

This is a proposal, not a demonstration. The Minsky paper states that the reachability of its Reflective layer is argued rather than shown, and that its sufficient conditions—a meta-control loop and context tagging—remain future work (2026f). Auditable abduction inherits those limits.

8.5 Re-rooting already happens by hand

Two episodes from 2026 show re-rooting done without such a substrate. In April, Liam Price submitted a solution to Erdős Problem 1196 that came from his interaction with GPT-5.4 Pro. Nat Sothanaphan and Jared Lichtman interpreted the AI’s approach and rewrote it in human-readable form, and in May a paper by Alexeev, Barreto, Li, Lichtman, Price, Shah, Tang and Tao expanded the key idea, clarified neighbouring problems and gave a cleaner proof of the

Erdős primitive set conjecture (Sanderson 2026). In July, Tao’s digestion rebuilt the Jacobian counterexample from multiplication of low-degree polynomials (Tao 2026b).

In both episodes people did the work of the second layer by hand: they reconstructed a walkable trace before the third layer could value the result. Our claim is modest. An exact-trace substrate could make part of that reconstruction mechanical and checkable, and so leave more of the scarce human effort for understanding and teaching. It cannot replace the understanding it serves.

8.6 A complementary direction

The three layers keep the LLM’s internal computation opaque and supply a trace beside it. A complementary route reads the LLM itself with hyperdimensional algebra: the Hyperdimensional Probe uses vector-symbolic encodings to project a model’s residual stream onto interpretable concepts and reports that this helps identify model failures (Bronzini et al. 2025). Exact, reversible binding of the VaCoAl kind may sharpen such probes. Both routes share the premise of this paper: the trace is the scarce good, and algebra is a way to make it.

9 Predictions, limitations and open problems

The account makes six predictions that could fail, and it rests on limitations a reader should weigh before accepting it.

9.1 Predictions

1. **Valuation falls as rescue rises.** On the existing 25.5-million-path corpus at fixed memory depth, the dispersion of CR_2 among candidates inside the frontier should fall monotonically as RR rises from 0 to 1, and the CR_2 -ordered top- k should converge on the dictionary’s lexicographic top- k . This is Conjecture 1 of §4.2; a non-monotone curve would show that the dial is not one-dimensional and that the conjecture is false.
2. **A test that sorts discoveries.** Any structure exposed by a new composite operation can be re-run in Rescue mode. Structure that survives is contingently tacit, an externalization in Nonaka’s sense; structure that disappears is regime-bound. The coarse findings of 2026d already pass (Appendix J); the prediction is that the CR_2 -dependent ranking results do not, and that no new discovery will fall between the two classes.
3. **Self-reads fail for structural reasons.** Following 2026f, the error of a reflexive unbind of a deliberative bundle should grow with collision saturation and not with training-data coverage, the opposite signature from an LLM’s self-report. Injecting known deliberation bundles into PyVaCoAl and measuring recovery error against collision rate would test this directly.
4. **Traces re-root results in people.** Give researchers an AI-generated result in one of three forms: the result alone, the result with an LLM’s self-explanation, or the result with a VaCoAl/PyVaCoAl trace and its CR values. After a fixed time, measure what Tao’s follow-up activities require: answering questions on the result, presenting it, and extending it. We predict that the trace condition beats both others on questions and extension, while the self-explanation condition may score high on felt understanding without matching gains in performance.
5. **An AI wall would be measurable.** Using a 21st-century analogue of the prosopographic graph—mentorship, co-authorship and acknowledged collaboration—results whose prove-nance is mainly AI-generated should show thinner downstream traffic than comparable

human results within a fixed horizon, measured by the same paths-per-node statistic that separates 53.4 downstream from 5.3 upstream at Leibniz (2026d). The prediction fails if digestion practices of the kind in §8.5 close the gap, which would itself be evidence that re-rooting works.

6. **Articulability tracks the route of integration.** In the hippocampal reading of §5.4, relations learned under conditions that recruit the typed-commitment route should be easier to state than relations of similar complexity learned by similarity completion. This is the most speculative prediction here, and its experimental design is left open.

9.2 Limitations

- **Scope of the word.** “Tacit” is used in the structural sense of §4.1. Nothing here addresses whether a substrate can be a knower in Polanyi’s personal sense, or the phenomenology of knowing.
- **One corpus.** The empirical anchors come from one relation (documented mentorship), one source (Wikidata, integrating the Mathematics Genealogy Project and MacTutor) and one tracer set (the 64 Fields Medalists). Wikidata over-represents Western, university-affiliated and doctoral-advisor transmission (2026d), and the mentorship channel misses texts, correspondence and translation except where node semantics recover them.
- **Analogies are correspondences.** The pianist, the cyclist and the Identikit are matched to VaCoAl/PyVaCoAl operations by structure, not identity. In particular, the CASRAM face-recognition product matches feature embeddings computed on a GPU; we do not claim that it binds faces as role-filler pairs.
- **Only the poles are mapped.** The audit-valuation trade-off is documented at $RR = 1$ and $RR = 0$. Intermediate rescue rates have not been characterized.
- **Hardware claims are not used.** CASRAM is physically realized and commercial, but its in-memory speed and power remain unbenchmarked, and its SRAM capacity gives a small frontier size compared with the PyVaCoAl DRAM-CAM realization (2026e; 2026f). No argument in this paper depends on CASRAM performance.
- **Proposals are not demonstrations.** Auditable abduction (§8.4) and the Reflective layer it relies on are designs; 2026f states that the Reflective layer’s sufficient conditions are future work and its reachability is argued, not shown.
- **LLM evidence ages.** The findings in §7.3 come from specific models and methods. Introspective ability appears to grow with model capability (Anthropic 2025b), so the default rootlessness of LLM outputs may weaken, and the argument should be re-checked as models change.
- **Reproducibility.** The code for the software version of PyVaCoAl is proprietary and available for commercial and collaborative use (2026a); 2026d states that queries, configurations, the tracer set and a reference implementation are to be deposited with a DOI on acceptance. Independent replication of the claims used here depends on that release.

9.3 Open problems

Three questions are left open. Can audit-valuation complementarity be stated as an inequality between a measure of audit exactness and a measure of valuation contrast, with the closed-form CR_2 law as its equality case? Can significance be represented in a substrate without becoming one more explicit item, or is value binding the formal signature of collective tacit knowledge,

available only through a community? And what interface lets the three layers of §8 work together in practice, so that a researcher can move from an LLM’s proposal to a VaCoAl/PyVaCoAl trace to a seminar without losing the thread?

10 Conclusion

Tacit knowing, on this account, is an operating regime of a memory rather than a kind of content stored in it. VaCoAl/PyVaCoAl makes the claim concrete because one substrate, under one dial, yields either an exactly auditable trace or an emergent valuation of paths, and not both at full strength. Read with a relational definition of tacitness, that single trade-off gives each classical position its place: Polanyi’s non-separation, the pianist who falters under focal scrutiny, Tsoukas’s ineffability at the object level, Nonaka’s externalization by new operations, and Collins’s three kinds as three strata, the last of which lives in the community rather than in the substrate.

The same reading changes what our earlier results are about. The mentor–student DAG is a record of tacit transmission, the Leibnizian hourglass is institutionalized socialization, and the Monastery Wall is a change in the medium through which tacit knowledge travelled. Newton’s thin thread, Thurston’s emptied field and Tao’s AI-solved problems without seminars share one structure: possession without transmission.

AI-generated results are rootless by default. They can be verified, but nobody yet holds them tacitly, and their derivations leave no walkable trace. We have proposed that VaCoAl/PyVaCoAl can supply the missing middle layer between generation and valuation—an exact trace with its confidence—and stated predictions by which that proposal can fail. Understanding a result and judging its worth remain with people: as virtue epistemology explains, the intellectual virtues they call on cannot be offloaded to technology. People will always know more than they can state. A substrate that records how a result was reached gives them something they have rarely had: a way to show the route, which is where tacit knowledge has always begun to travel.

References

The VaCoAl programme

- Chuma, H., Otsuka, K., & Sato, Y. (2026a). Beyond LLMs, sparse distributed memory, and neuromorphics: A hyper-dimensional SRAM-CAM “VaCoAl” for ultra-high speed, ultra-low power, and low cost. arXiv:2604.11665. <https://arxiv.org/abs/2604.11665>
- Chuma, H., Otsuka, K., & Sato, Y. (2026b). Bridging silicon and the hippocampus: Algebro-deterministic memory “VaCoAl” as a substrate for Vector-HaSH and TEM. arXiv:2605.15652. <https://arxiv.org/abs/2605.15652>
- Chuma, H., Otsuka, K., & Sato, Y. (2026c). How to build Marcus’s algebraic mind: Algebro-deterministic substrate over Galois fields. arXiv:2605.21379. <https://arxiv.org/abs/2605.21379>
- Chuma, H., Otsuka, K., & Sato, Y. (2026d). Institutions and the transmission of upper-tail human capital: Scientific lineages across a millennium. arXiv:2605.31470. <https://arxiv.org/html/2605.31470>
- Chuma, H., Otsuka, K., & Sato, Y. (2026e). How to build Marcus’s algebraic mind: From Thagard’s brain–mind viewpoint. arXiv:2607.16573. <https://arxiv.org/pdf/2607.16573>
- Chuma, H., Otsuka, K., & Sato, Y. (2626f). How to build Marcus’s algebraic mind: From Minsky’s emotion-machine viewpoint. arXiv:2607.16623. <https://arxiv.org/abs/2607.16623>
- Chuma, H., Otsuka, K., & Sato, Y. (2026g). Computational prosopography across a millennium: Mathematically oriented lineages traced from the Fields Medalists. arXiv:2608.17915. <https://www.alphaxiv.org/abs/2608.17915>
- Shuhari System (n.d.). CASRAM (SRAM-based CAM). https://shuharisystem.com/?page_id=1336
- Shuhari System (2026). Webcam face-recognition system with CASRAM technology (in Japanese), 29 March 2026. https://shuharisystem.com/wp/wp-content/uploads/2026/04/casram_face_rec

Tacit knowledge, machines and the history of science

- Collins, H. M. (1974). The TEA set: Tacit knowledge and scientific networks. *Science Studies*, 4(2), 165–186.
- Collins, H. M. (1990). *Artificial Experts: Social Knowledge and Intelligent Machines*. MIT Press.
- Collins, H. (2010). *Tacit and Explicit Knowledge*. University of Chicago Press.
- Collins, H. (2018). *Artificial Intelligence: Against Humanity's Surrender to Computers*. Polity Press.
- Dreyfus, H. L. (1972). *What Computers Can't Do: A Critique of Artificial Reason*. Harper & Row.
- Kim, M. (2026). AI is rewiring how the world's best Go players think. *MIT Technology Review*, 27 February 2026. <https://www.technologyreview.com/2026/02/27/1133624/ai-is-rewiring-how-the-worlds-best-go-players-think/>
- Kuhn, T. S. (1970). *The Structure of Scientific Revolutions* (2nd ed.; 1st ed. 1962). University of Chicago Press.
- Mokyr, J. (2016). *A Culture of Growth: The Origins of the Modern Economy*. Princeton University Press.
- Nonaka, I., & Takeuchi, H. (1995). *The Knowledge-Creating Company*. Oxford University Press.
- Nonaka, I., & von Krogh, G. (2009). Tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory. *Organization Science*, 20(3), 635–652. <https://doi.org/10.1287/orsc.1080.0412>
- Peirce, C. S. (1903). Lectures on pragmatism. In *Collected Papers of Charles Sanders Peirce*, Vol. 5 (1934). Harvard University Press.
- Polanyi, M. (1958). *Personal Knowledge: Towards a Post-Critical Philosophy*. University of Chicago Press.
- Polanyi, M. (1966). *The Tacit Dimension*. Doubleday.
- Polanyi, M. (1969). *Knowing and Being* (M. Grene, Ed.). University of Chicago Press.
- Pritchard, D. (2023). *What is this thing called Knowledge?* (5th ed.). Routledge. <https://doi.org/10.4324/9781003356110>
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1), 99–118.
- Tsoukas, H. (2003). Do we really understand tacit knowledge? In M. Easterby-Smith & M. A. Lyles (Eds.), *The Blackwell Handbook of Organizational Learning and Knowledge Management*. Blackwell.

Hyperdimensional computing, memory and causality

- Bronzini, M., Nicolini, C., Lepri, B., Staiano, J., & Passerini, A. (2025). Hyperdimensional probe: Decoding LLM representations via vector symbolic architectures. arXiv:2509.25045. <https://arxiv.org/abs/2509.25045v2>
- Chandra, S., Sharma, S., Chaudhuri, R., & Fiete, I. (2025). Episodic and associative memory from spatial scaffolds in the hippocampus. *Nature*, 638, 739–751.
- Eliasmith, C. (2013). *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford University Press.
- Kanerva, P. (1988). *Sparse Distributed Memory*. MIT Press.
- Kanerva, P. (2010). What we mean when we say “What’s the dollar of Mexico?”: Prototypes and mapping in concept space. AAAI Fall Symposium on Quantum Informatics for Cognitive, Social, and Semantic Processes.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference* (2nd ed.). Cambridge University Press.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. Wiley.
- Thagard, P. (2019). *Brain–Mind: From Neurons to Consciousness and Creativity*. Oxford University Press.

Language models

- Anthropic (2025a). Tracing the thoughts of a large language model, 27 March 2025. <https://www.anthropic.com/research/tracing-thoughts-language-model>
- Anthropic (2025b). Signs of introspection in large language models, 29 October 2025. <https://anthropic.com/research/introspection>
- Balesni, M., Korbak, T., & Evans, O. (2024). The two-hop curse: LLMs trained on $A \rightarrow B$, $B \rightarrow C$ fail to learn $A \rightarrow C$. arXiv:2411.16353. <https://arxiv.org/abs/2411.16353v1>
- Berglund, L., Tong, M., Kaufmann, M., Balesni, M., Stickland, A. C., Korbak, T., & Evans, O. (2024). The reversal curse: LLMs trained on “A is B” fail to learn “B is A”. ICLR 2024. arXiv:2309.12288. <https://arxiv.org/abs/2309.12288>
- Budding, C. (2025). What do large language models know? Tacit knowledge as a potential causal-explanatory structure. *Philosophy of Science*, 92(4). arXiv:2504.12187. <https://www.cambridge.org/core/product/9475F1504081116099098C37D6F57611>
- Collins, H., & Thorne, S. (2026). Large language models and scientific discourse: Where’s the intelligence? *Synthese* (accepted). arXiv:2603.23543. <https://arxiv.org/pdf/2603.23543>
- Graepel, T. (2026). Don’t be fooled—LLMs don’t reason. *MIT Technology Review*, 2 October 2026. <https://www.technologyreview.com/2026/10/02/1145639/dont-be-fooled-llms-dont-reason/>
- Si, C., Yang, D., & Hashimoto, T. (2024). Can LLMs generate novel research ideas? A large-scale human study with 100+ NLP researchers. arXiv:2409.04109. <https://arxiv.org/abs/2409.04109v1>

Mathematics in the age of AI

- Sanderson, G. (2026). If math is more than proof, we need to better celebrate the rest of it. Guest post on *What’s new*, 18 September 2026. <https://terrytao.wordpress.com/2026/09/18/if-math-is-more-than-proof-we-need-to-better-celebrate-the-rest-of-it/>
- Tao, T. (2026a). Thread on AI tools and the follow-up activities of mathematical results. Mathstodon. <https://mathstodon.xyz/@tao>
- Tao, T. (2026b). A digestion of the Jacobian conjecture counterexample. *What’s new*, 21 July 2026. <https://terrytao.wordpress.com/2026/07/21/a-digestion-of-the-jacobian-conjecture-counterexample/>
- Thurston, W. P. (1994). On proof and progress in mathematics. *Bulletin of the AMS*, 30, 161–177. arXiv:math/9404236. <https://arxiv.org/abs/math/9404236>
- Weinreich, M. (2026). The crisis of AI-generated mathematics. arXiv:2608.02859. <https://arxiv.org/pdf/2608.02859>

Acknowledgments

We thank the Wikidata, Mathematics Genealogy Project and MacTutor communities, whose data underlie the results reused here. The hypotheses advanced in this paper mainly originate with the first author (H.C.), grounded in the results of the authors’ seven VaCoAI papers (2026a–g). Claude, an AI assistant developed by Anthropic, was used as a writing aid.