The Ancestry of Belief

Epistemic lineage as a first-class primitive for bounded intelligences, human and machine

Tad MacPherson
Researcher, Metavolve Labs
ORCID: 0009-0002-8659-7479

Working draft v1.3 (review), June 2026

A position/theory paper. Claims are conditional and labelled; the civilization-scale framing is the horizon, the load-bearing claims are deliberately tractable. We use Arweave/AO as a single existence proof, but the argument is about substrate properties*, not any product. It should survive if any particular implementation disappeared.*

Abstract

Almost everything any intelligence knows, it did not observe. Every generation inherits vastly more knowledge than it directly observes. The question is not whether knowledge is inherited, but whether the ancestry of that inheritance survives. A human inherits the shape of history, science, and geography from testimony they cannot personally check; a language model inherits its world from a corpus it cannot re-derive; a robot inherits a map it did not survey. Every bounded mind, biological or artificial, runs overwhelmingly on *inherited belief*. The classical theories of knowledge, including truth-maintenance systems and the AGM tradition of belief revision, were built for the opposite case: a single, private, fully-controlled belief base. They ask *what should I believe and when should I revise?* That question assumes a thinker who owns their beliefs. It is the wrong question for a world in which most belief is inherited from others, across time, and cannot be independently verified.

We propose that for bounded intelligences operating at scale in fluid, adversarial information environments, a different primitive is operationally fundamental: **epistemic lineage**, the durable, attributable, externally-inspectable ancestry of how a belief came to be held and how it changed. The central move can be stated in one line: *classical revision theory treats revision history as disposable internal bookkeeping; epistemic lineage treats revision history as a durable object of reasoning shared between minds.* Truth does not stop mattering; it remains the target. But where truth cannot be directly established, which is the default condition for any mind acting on inherited belief, lineage becomes the operational mechanism for allocating trust. This is not a novel need. Civilization has hand-built partial lineage systems for millennia: citation, peer review, archives, precedent, the footnote. Those were *institution-dependent*, surviving only as long as a keeper behaved. A *neutral, append-only, permanent, attributable* substrate is the first to make lineage *substrate-dependent*, guaranteed by properties that hold rather than institutions that behave. From the primitive two theorems follow, which we develop as the paper's signatures. **Coordination through traceable disagreement** is a third way to hold a society of minds coherent, beside consensus and authority. **Accountable compression** is the requirement that lineage be not merely preserved but traversable. The same primitive operates at three scales: it lets an *individual* inspect inherited belief, lets *institutions* preserve reasoning across turnover, and lets *societies of minds* coordinate without consensus. The scale changes; the primitive does not. We are explicit about the bound. Lineage is an **accountability primitive, not a truth**

oracle: a permanent substrate preserves a lie as faithfully as a fact. That is the precise and defensible scope. The paper does not claim to solve truth. It claims to make *the path by which a claim entered the world* durable, attributable, and open to inspection.

1 The question a bounded mind can actually answer

Consider any intelligence that cannot personally observe everything it must act on. It is the situation of a person, an institution, an agent, a robot, a civilization. Beyond the narrow horizon of direct experience, every such mind must *inherit* belief. It accepts claims asserted by others, conditioned on evidence it did not gather, revised by hands it cannot see.

For this mind, the classical epistemic question, *is this belief true?*, is frequently unanswerable in the moment of action. The robot cannot verify the map before it must move; the model cannot re-derive the training example before it must answer; the analyst cannot reproduce the experiment before the decision is due. Truth remains the *target* of inquiry, but it is rarely available as an *input* to action. What is available, what can always in principle be inspected, is the **ancestry** of the belief: who asserted it, when, on what evidence, and how it has been revised since.

This suggests a reframing we state carefully, because the careless version invites attack. We do **not** claim that ancestry replaces truth, or that truth ceases to matter:

> *For bounded intelligences operating at scale, lineage is the primary operational mechanism for managing uncertainty when truth cannot be directly established.*

Phrased this way the claim is almost obviously true, which is the point. A mind acting on inherited belief is *already* trusting a chain. The only question is whether that chain is **inspectable**. When it is not, as through nearly all of human history and in most machine pipelines today, the mind must trust blindly, by reputation, brand, or force. When the chain *is* inspectable, trust can be *reasoned about*. The shift is not from truth to ancestry. It is from *blind inheritance* to *inspectable inheritance*.

The human stakes are not a metaphor. For humans the un-inspectability of lineage has never been neutral. Knowledge has been, across recorded history, among the most reliably gated of assets and a durable lever of control. What makes it gateable is precisely that the chain can be hidden: the source made unverifiable, the citation enclosed, the provenance stripped, the inconvenient past quietly revised by whoever holds the pen. We are careful not to overstate. Power is resourceful and always finds new levers, and no substrate abolishes control. The precise claim is narrower, and stronger for it: *the substrate weakens one specific and historically powerful lever, the ability to invisibly alter or obscure the lineage of knowledge.* The human reader is inside this argument's scope, not adjacent to it.

2 Civilization has always run on lineage

The primitive we will define is not an invention. It is the formalization of something humanity has approximated, imperfectly, for as long as it has kept records. The clearest proof is the object you are reading.

A scientific paper *is* a manually maintained lineage graph. Its citations, footnotes, prior-work section, and dependencies exist to let a stranger reconstruct not whether each claim is true but *how it came to be held*: on whose authority, by what evidence, building on which earlier argument. We are, in writing this, doing the very thing we describe. The paper is only the sharpest instance of a pattern that is everywhere once seen. Science is lineage: citation networks, replication, the methods section. Law is lineage: precedent, the chain from constitution to

statute to ruling, the preserved dissent. Scholarship is lineage: the primary source, the critical edition, the archive. Journalism is lineage: attribution, the named source, the correction notice. Civilization spent thousands of years building these because it learned, at great cost, that the ancestry of a claim is what lets a stranger trust it. **The claim of this paper is therefore not that lineage is new.** It is that lineage has always been valuable enough that civilization *repeatedly reinvented partial versions of it*, and that we may now, for the first time, have a substrate able to represent it directly.

Two features of these historical systems matter for what follows. The first is that every one of them is, on inspection, also a **compression** scheme: a citation compresses an argument to a pointer, a textbook compresses a field, a headnote compresses a precedent. Nobody reads the whole lineage; we read summaries that *stand in* for it. We return to this in Section 6, because it is the framework's central practical challenge as much as its oldest solution.

The second is that every one of these systems is **institution-dependent**. Its integrity rested on a keeper behaving well: the academy, the publisher, the court, the state archive, the library. The keeper has always been simultaneously the system's *guarantor* and its *gate*. The Library of Alexandria is the emblem of the failure mode. When lineage lives in an institution, it shares that institution's mortality and its politics; the same body that preserves the record can enclose it, capture it, defund it, or lose it. (Tellingly, the compression schemes are where capture is easiest: control the textbook and you control a field's view of its own ancestry without touching a single primary source.)

This locates precisely what the substrate of Section 3 adds, and it is one specific thing, stated with care. The substrate does **not** make lineage *custodian-free* in the sense of no one being responsible. It changes *what kind of guarantee you are relying on.* Libraries survive because librarians behave: a **social** guarantee, contingent on conduct, solvency, and politics. A neutral permanent substrate survives because **properties hold**: an **architectural** guarantee, contingent on the substrate's design rather than any party's good faith. The move is from *institution-dependent lineage* to *substrate-dependent lineage*, and the two are fundamentally different kinds of assurance, one behavioural, one structural. We resist the grandiose reading and state the tractable one. We do not propose to replace science, law, or the archive, nor imagine those institutions dissolving. We propose that the substrate supplies, for the first time, a property every prior lineage system lacked, *integrity that does not depend on a keeper who is also a gate*, and we ask what follows when an ancient need finally has a planetary-scale, property-guaranteed implementation.

3 Defining the primitive

3.1 What lineage is, and what changes when it leaves the mind

The single most important claim of this paper can be stated before any definition:

> *Classical revision theory treats revision history as disposable internal bookkeeping; epistemic lineage treats revision history as a durable object of reasoning shared between minds.*

Everything else follows from that sentence. The definition makes it precise.

By **epistemic lineage** we mean the durable, attributable, externally-inspectable ancestry of a belief: not merely *where a datum came from* (provenance in the narrow sense), but the full structure by which a belief came to be held and changed. It comprises four interlocking parts:

1. **Provenance**: the origin of the asserted claim and the evidence offered for it.
2. **Justification structure**: the dependency graph linking a belief to the beliefs and evidence

that support it. This is the contribution of truth-maintenance systems (Doyle, 1979) and the AGM/iterated-revision tradition, adopted here as foundations, not adversaries: they formalize *reasoning over why a belief is held.*

3. **Revision history**: the temporal record of how the belief changed, what superseded what, and when.

4. **Attribution**: who asserted, who revised, who corroborated; the signatures on each link.

The novelty is not in these parts but in what happens to the *revision operation* when they are externalized. In classical truth maintenance and belief revision, revision is an **internal consistency-maintenance operation inside a single mind's private, mutable store**. The agent updates its beliefs to restore coherence, and the *history* of that update is an implementation detail the agent is free to discard once consistency is restored. Externalize the belief base onto a neutral, append-only, permanent, attributable substrate, and the operation changes character. Revision becomes a **public, attributed, queryable amendment whose own history is now a first-class object**. It is not discarded once coherence is restored, but retained as part of what *other* minds, and future versions of the same mind, reason over. The agent no longer merely *has* a revised belief; it leaves a durable, inspectable record of *having revised it, and why.* Revision history is promoted from a forgettable internal detail to a shared input that downstream minds use to calibrate trust (Section 4) and to locate where their conclusions diverge from another's (Section 5). The move reduces to a single sentence, and the whole of this paper is its elaboration: *revision history, long treated as disposable internal bookkeeping, becomes a first-class object of reasoning.*

3.2 The four substrate properties

For lineage to be a *first-class, shared, substrate-guaranteed* primitive, rather than a private byproduct each mind keeps or discards, or an institution-dependent record a keeper can gate, the substrate must be:

- **Neutral**: no single party can rewrite, withhold, or disappear the record. Neutrality is what converts the guarantee from social to architectural. Without it, the holder of the record becomes the new gate.

- **Append-only**: the past is never overwritten, only extended.

- **Permanent**: the record persists beyond the lifespan, solvency, or interest of any individual mind or institution. Lineage that dies with its keeper is not lineage; it is memory.

- **Attributable**: every assertion and revision carries a verifiable signature, so the chain is not merely preserved but *accountable.*

We make no claim that these require a blockchain in principle. We claim only that they must hold *together*, and that systems satisfying all four at scale are historically rare. Civilization's lineage machinery reliably satisfied append-only-ish and attributable; it never satisfied neutral-and-permanent together. Section 9 discusses the first system we know of that satisfies all four jointly. The theory is stated over the properties and would stand if that implementation vanished. Permanent storage here is the **transistor**, not the point; epistemic lineage is the computation it enables.

3.3 The mechanism: correction and erasure become different operations

The objection to permanence is always the same: *a permanent record preserves errors and lies forever.* The objection is correct, and it is the engine of the proposal rather than a flaw, once one sees that permanence does not make change difficult. It makes **change and erasure different operations**.

In any mutable store, "correct the record" and "hide that the record changed" are the *same* act. You overwrite, and the overwrite destroys the evidence of its own occurrence. That fusion is what makes adversarial rewriting work and knowledge gateable. A neutral, append-only, permanent, attributable substrate **decouples** them. Honest correction becomes *cheaper and easier than ever*: a new signed version, attributed, instant, superseding the old as the authoritative pointer while the prior state remains visible beneath it. But covert revision becomes **impossible**. The prior state cannot be deleted, and every change is signed and timestamped. Honest actors gain; their correction history becomes the very lineage we wish to preserve. Adversaries lose the one thing they require: revising the past without anyone able to tell. *Because erasure is impossible, the lineage is complete; because honest correction is cheap and attributed, the lineage stays current.* Completeness and currency, usually in tension, are reconciled by separating the two operations a mutable store welds together. This dissolves the objection for *erroneous* content, which becomes attributed correction rather than silent overwrite. It does not, by itself, address the harder case of content that is harmful precisely *because* it is true and accurate, which cannot be "corrected" away. We take that case up in Section 8.

3.4 The bound: accountability, not truth

Lineage is an accountability primitive, not a truth oracle. The substrate preserves the ancestry of a belief faithfully; it cannot make the *genesis assertion* honest. The first link in any chain may be a lie. What changes is that the lie is now durable, attributable, timestamped, and open to inspection. It is traceable to its author and weighable against its corroboration, even when it cannot be refuted outright.

This is not a weakness smuggled in at the end. It is the precise scope and, we argue, the more fundamental property. A mind cannot always determine whether an inherited belief is *true*. It can, given inspectable lineage, determine how *far to trust* it, and act safely within that bound without resolving the underlying truth. **Trust becomes a computable function over ancestry** rather than a verdict asserted by authority.

4 Trust as a function of ancestry

If lineage is inspectable, trust stops being a black box. Today, an intelligence deciding how far to rely on an inherited belief has mostly crude instruments: the reputation of the source, the brand of the institution, the bare assertion of an authority. None is a function of the belief's *actual history*. They are proxies, gameable precisely because the history they stand in for is hidden.

With first-class lineage, trust can be computed *over the ancestry graph itself*, as a function of inspectable features: the **depth** of the chain (steps back to a primary source or direct observation), the **diversity** of its sources (genuinely independent provenance versus a single origin laundered through many voices), the **corroboration** among branches (do independent ancestries converge), and the **revision history** (stable, or repeatedly superseded, and were the supersessions attributed and evidenced). We do not propose a specific scoring function. We claim that lineage makes such functions *possible at all*, by supplying the inspectable substrate they range over.

This makes trust a different kind of object than the two it replaces. Reputation loses most of its

force the moment it is moved out of the context that earned it; a belief instead carries its ancestry *with* it, so trust becomes **portable without laundering the basis for it away**. Authority collapses the moment the authority is distrusted; an ancestry can instead be re-inspected by anyone, so trust becomes **legible**, a thing whose work can be shown rather than a verdict that must be accepted. This is the natural generalization of provenance-based trust scoring, which today operates on single artifacts (a file, a credential, a wallet), lifted to operate on *belief itself*.

The bound of Section 3.4 holds here and sharpens the claim. A function over ancestry does not compute *truth*; it computes *warranted reliance*. A deep, diverse, corroborated, stably-revised ancestry can still terminate in a false genesis assertion. What the function buys is calibrated exposure: an intelligence can bound how much it stakes on a belief by the shape of that belief's history, which is exactly the decision a bounded mind actually has to make. But computing over an ancestry presumes the ancestry can be *read*, and a complete record may be too large to traverse. That is not a footnote. It is the second theorem, and we turn to it after the first.

5 First theorem: coordination through traceable disagreement

If epistemic lineage is the primitive, its first major theorem is a new way for a society of minds to remain coherent. We give it the room because primitives are remembered for the *consequences* that fall out of them.

There have classically been two mechanisms for holding a group of minds coherent:

- **Consensus**: *we agree.* The group converges on a shared belief. Powerful when achievable, but brittle. It fails under genuine, persistent disagreement, and manufacturing it requires persuasion that may not arrive or coercion that corrupts.
- **Authority**: *we obey.* A trusted party fixes the belief and the rest defer. Scalable, but it concentrates power and inherits every failure of the authority: capture, error, mortality, bad faith.

Inspectable lineage enables a third, which we state as the theorem:

> *Coordination through traceable disagreement. Minds that do not agree on what is true can nonetheless coordinate, provided each can inspect why the other believes what it believes and trace the divergence to its root. They do not converge (it is not consensus) and no one dictates (it is not authority); they locate exactly where their ancestries diverge and continue cooperating across that fault line with eyes open.*

The proposition earns its weight by naming the real failure mode of institutions under disagreement. **Institutions rarely fail because disagreement exists; they fail because disagreement becomes *opaque*.** When two parties can no longer reconstruct *where* they began to diverge, disagreement curdles into faction, mistrust, and stalemate. Each side sees only the other's conclusion, not the path to it, and an unexplained conclusion is indistinguishable from bad faith. Lineage makes divergence inspectable. It converts "you are wrong" into "here is the exact assertion, the exact evidence, the exact revision where our chains parted." That is a tractable object: examinable, arguable, sometimes resolvable, and, crucially, *coordinated around even when not resolved*.

We are not claiming traceable disagreement is new. A skeptical reviewer will rightly observe that science already coordinates through documented disagreement, that law preserves dissent beside the majority, that version control records every fork. Just so. These are the partial reinventions of Section 2. The claim is sharper: *existing systems achieve traceable disagreement only within bounded institutions and through expensive, manual, institution-dependent processes;*

substrate-dependent lineage makes it a native property of the information layer itself, available across institutional boundaries, between mutually-distrusting parties, and without a custodian who is also a gate. The mode is old. Making it native and scalable is the contribution.

Three concrete instances show its reach, each stated at the tractable altitude:

- **A human institution under narrative attack.** When the information environment is adversarially flooded with cheap synthetic claims, manufactured corroboration, and quietly edited histories, the defense is not a better arbiter of truth (which the attacker contests) but inspectable ancestry: the ability to show *where* a contested claim entered, *who* asserted it, and *whether* its apparent corroboration traces to independent sources or to one source laundered. The institution survives by making the divergence between its record and the injected one *traceable*, not by winning a contest over truth it has been engineered to lose.
- **A multi-agent economy under partial distrust.** Autonomous agents operated by rival principals must cooperate without a shared orchestrator any of them trusts. Inspectable lineage lets each agent reason about *the ancestry of another's claim*, its sources, corroboration, and revision record, and transact across disagreement it can see into, rather than refusing all interaction with parties it cannot verify. Coordination scales precisely because it does not require trust in the counterparty, only inspectability of the counterparty's chain.
- **A robot fleet sharing a world model.** As units' sensors drift and individual robots are lost and replaced, the fleet will hold *divergent* beliefs about the same environment: a doorway one maps as open, another as closed. Consensus discards real signal; a central authority creates a single point of failure. With inspectable lineage, a unit encountering a conflicting belief can trace it to *which robot observed what, when, with which sensor, superseded by which later observation*, and bound its trust accordingly without halting. The fleet stays coherent not by agreeing but by being able to inspect *why* it disagrees. (Tractable claim: lineage makes the disagreement *locatable*; it does not make either sensor correct.)

Consensus does not scale to mutually-distrusting parties; authority does not survive their distrust; traceable disagreement asks neither for agreement nor for a sovereign, only for inspectable lineage. That is why we hold it the most important thing the primitive yields, provided the divergence can actually be *found*, which is the burden of the second theorem.

6 Second theorem: accountable compression

A lineage that cannot be traversed is no better than one that was never kept. A complete ancestry, faithfully preserved, can be too large or too tangled to read; an inspectable record that no one can afford to inspect calibrates no trust (Section 4) and locates no divergence (Section 5). The first theorem assumes the second. Put plainly: *without accountable compression, lineage is an archive; with it, lineage is a system.*

> *First-generation lineage systems solve preservation. Second-generation lineage systems must solve traversability. The way to solve it without re-importing a gatekeeper is accountable compression: lossy summaries that themselves carry a verifiable pointer back to the full ancestry they compress.*

History already discovered that lineage must be compressed: the citation, the textbook, the headnote (Section 2). But historical compression was **trusted** compression. A keeper decided what to summarize away, and the summary became the new gate, re-importing the institution-

dependence the substrate was meant to escape. Control the textbook and you control a field's view of its own past without touching a primary source. The danger is not compression; it is *unaccountable* compression.

The framework's requirement is therefore precise: a summary must be *lossy yet inspectable*. It compresses, because no one can read everything, but its compression can be reopened and checked against the underlying lineage by anyone who doubts it. A trust function over ancestry (Section 4) is exactly such a compression: it reduces a vast graph to a calibrated reliance-bound whose derivation descends, on demand, into the chain it summarizes. The difference from a textbook is not that the substrate version is uncompressed, since it must compress, but that its compression is **inspectable rather than authoritative**. You read the summary, and when the stakes warrant, you descend into the lineage it stands for.

This is why we state it as a theorem rather than a caveat: it defines a research agenda with natural descendants. *Search* over lineage (find the relevant ancestry), *summarization* of lineage (compress it faithfully), *explainability* (a model's output as a traversable ancestry rather than an opaque verdict), and *agent memory* (a working summary that can be re-grounded in full history) are all instances of accountable compression over an inspectable substrate. Whether such compression can be made both faithful and cheap at scale is, we believe, the central open engineering problem the framework opens (Section 10), and the right successor question to "is the record permanent?" The record being permanent is necessary; the record being *traversable* is what makes permanence useful.

7 Lineage as institutional memory: a theory of anti-forgetting

There is a frame underneath this paper larger than AI, truth, or storage: a crisis of **institutional memory**. Civilizations forget. So do companies, governments, and scientific communities, and so will societies of agents. The mechanism is monotonously similar each time. People leave, records are deleted or quietly migrated into oblivion, the context that made a decision intelligible vanishes, revisions overwrite the history that produced them, and what remains decays into folklore: a conclusion whose ancestry no one can reconstruct.

Lineage is, in this light, almost a *theory of anti-forgetting*, and the distinction in what it preserves is the subtle, important point. It does not primarily preserve *facts*; archives already do that, badly, when the custodian survives. It preserves **how facts arrived**: the justification structure and revision history that let a later mind reconstruct not just *what* was concluded but *why*, *by whom*, *against what alternatives*, and *through which corrections*.

> *A fact without its ancestry is folklore; a fact with its ancestry is knowledge.*

Institutions forget in the second sense long before they lose the first: the conclusion survives in a slide deck while the reasoning that warranted it, and the dissent that qualified it, evaporates with the people who held it. This reframes the substrate's value for any long-lived organization, human or machine. The asset is not the immutable file. It is the **inheritable justification**: the ability of a future member to inspect the ancestry of an inherited belief instead of receiving it as opaque inheritance. That member might be a new employee, a successor government, a scientist a century on, or a freshly instantiated agent. Substrate-dependent lineage is the first form of institutional memory that does not decay when the institution's *people* turn over, because it does not depend on those people to carry it. The discipline of Section 1 still applies: we claim not that this makes organizations *wise*, but that it makes their reasoning *recoverable*, a precondition for wisdom, not a substitute for it.

8 Dual-use: the shadow of an inerasable record

A primitive this strong is dual-use, and a theory that frames only the upside is advocacy. The same properties that make lineage valuable make it dangerous along three axes, each stated as a live problem rather than resolved.

Inerasable lineage is also inerasable surveillance. A record no party can disappear protects the truth-teller from the censor, and exposes the dissident to the regime, permanently. The decoupling of Section 3.3 cuts both ways: a society that cannot make the past disappear also struggles to grant a person a fresh start, a sealed record, or a genuine right to be forgotten. Part of the boundary is firm. Content whose *bare existence* is the violation, material unlawful to possess at all, must simply never reach the substrate; permanence is the wrong tool, and there is no clever way around it. But for the larger class of content that is injurious through *amplification* rather than existence, accurate yet harmful, or subject to a right to be forgotten, the framework opens a structural direction we name but deliberately do not develop. A served *index* is separable from the permanent *substrate*, and, by Section 6, is an accountable compression of it. Such content might therefore be forgotten *operationally*, delisted, no longer surfaced or amplified, without being forgotten *evidentially*: that it was said, and by whom, remains, and the delisting itself is recorded rather than silent. Designing and, above all, *governing* such a mechanism (who delists, by what criteria, under what authority) is its own paper, and plausibly the first governance framework this primitive opens. Here we only mark that the shadow has a possible structure, and leave it to future work.

Inerasable error is un-revisable in the wrong design. Amendable immutability (Section 3.3) answers this for *correctable* belief: supersede the pointer, leave the error visible as attributed history. It does not answer for an agent whose committed, on-substrate policy is itself harmful. A persistent autonomous system that outlives its creator and cannot be stopped is, to a safety reviewer, an uncontrollable agent, and permanence is what makes it uncontrollable. The honest position pairs the substrate's persistence with **verifiable controllability on the same substrate**: corrigibility constraints, escalation triggers, and multi-party governance encoded in the permanent record and auditable by anyone. *Persistent but governable* is a design requirement, not a property the substrate confers for free.

Substrate-dependence relocates trust; it does not abolish it. The "architectural rather than social guarantee" of Section 2 is the paper's strongest move and its most exposed: it holds only insofar as the substrate's properties *actually* hold. Neutrality that degrades toward capture, permanence that rests on contingent economics, availability that depends on choke-points an adversary can pressure: each would convert the architectural guarantee back into a social one wearing architectural clothing.

9 An existence proof, and its limits

The theory is stated over properties; this section asks whether anything instantiates them, and answers honestly. The Arweave permanent-storage network and the AO compute protocol are, to our knowledge, the first large-scale system to satisfy neutrality, append-only persistence, and attributability *jointly*: content no single operator can rewrite, beneath an actor-model compute layer whose ordered message history is itself written permanently, so that state is recomputable from an inspectable log rather than asserted by a node. As an existence proof that the four properties can co-exist at scale, it suffices, and that is all this section claims of it.

As a *guarantee* it is bounded, and we say so. Permanence rests on an economic assumption that may not hold indefinitely; the project's own materials concede the network is not expected to produce blocks "in true perpetuity." Data permanence is not data *availability*: retrieval is

mediated by gateways that sit outside the core protocol and can apply policy. And the compute layer's stronger verifiability claims rest, at present, substantially on project-affiliated sources rather than independent evaluation. That is itself a finding: the properties this theory most needs are precisely the ones still *under-verified by neutral parties*, exactly the verification a theory like this should invite rather than assume. None of these undoes the existence proof. All bound how far one should lean on this particular instantiation, and all are reasons the theory is stated over properties rather than over a product. If a better substrate supersedes it tomorrow, the argument re-homes without loss, which is the test a theory should want to pass.

10 What would falsify the framework, and what remains open

A theory worth keeping must say how it could be wrong.

The framework weakens if the architectural guarantee proves illusory in practice, if every real instantiation of "neutral and permanent" drifts back toward an institution-dependent gate (capture, economic failure, choke-points), such that substrate-dependence is social-dependence in disguise. It weakens if inspectable lineage cannot be made **traversable**: if accountable compression (Section 6) cannot be both faithful and cheap, the lineage is complete but unread, and an unread record calibrates no trust and locates no divergence. This is, we believe, the single most important open problem, the successor to "is it permanent?" It weakens if trust-as-a-function-of-ancestry cannot be made robust against **ancestry forgery**: adversarial construction of deep, diverse-*looking*, mutually-corroborating chains that are in fact one source laundered, the lineage analogue of a Sybil attack. Inspectability makes such forgery *detectable in principle* (independence can be examined), but detecting it cheaply at scale is unsolved. And it weakens if coordination through traceable disagreement (Section 5), real in small documented settings, fails to scale past the point where inspecting the divergence costs more than the coordination is worth.

It is strengthened by the opposite: reasoning that becomes recoverable across institutional turnover that previously lost it; agent or robotic collectives that hold coherence through inspected rather than enforced agreement; trust and compression functions over ancestry that measurably resist laundering. These are experiments, not assertions, and we would rather discover the framework's limits than defend its reach.

11 Conclusion

This is, in the end, a theory of **knowledge lineage**: of what becomes possible when the ancestry of a belief, rather than the belief alone, is made durable, attributable, and open to inspection. Its starting observation is humbling and old. A bounded mind, human or machine, knows almost nothing of its own observation and must inherit the rest. Its starting move is small and, we think, exact: take the one thing classical theories of belief discard, the history of how a belief was revised, and make it a first-class, shared object rather than private bookkeeping. Everything else in the paper is consequence.

Two consequences we have argued are theorems. *Coordination through traceable disagreement* is the social one: a third way to hold minds together, neither forcing agreement nor installing a sovereign, that becomes native to the information layer once lineage is substrate-dependent rather than institution-dependent. *Accountable compression* is the practical one: the discovery that a lineage no one can traverse is no lineage at all, and that the way to make it traversable without restoring a gatekeeper is summarization whose own derivation can be inspected. The first tells us what the primitive is *for*; the second tells us what it will take to *use*.

We have been deliberate about the bound throughout, because the bound is what makes the

claim survivable. Stated cleanly, it is this. We do not claim that lineage produces truth. We claim that it preserves the ancestry from which truth-claims emerge. And the contribution of a neutral permanent substrate is not that it resolves disagreement, but that it makes disagreement *inspectable*, across time, institutions, and minds. That is less than a theory of truth and, for intelligences acting on inherited belief, more useful than one.

There is a word for what these theorems share, and we have kept it for the end because it is better discovered than declared. Every condition this paper addresses is a form of *discontinuity*. People leave, institutions forget, agents restart, sensors and maps diverge, records are rewritten, civilizations lose the thread. Lineage matters not merely because it records ancestry but because it carries reasoning *across* those breaks. Epistemic lineage is, in the end, the preservation of the continuity of reasoning across discontinuities of time, identity, institution, and belief. The fear it answers is older and more human than the loss of data. It is the fear of forgetting: of a conclusion severed from the path that gave it meaning.

If the argument is right, the deepest stakes are not technological. Humanity has wanted inspectable ancestry for as long as it has kept records, and built fragile, custodial approximations of it: the citation, the archive, the dissent, each surviving only as long as its keeper. That they were fragile was not a failure of will but the absence of a primitive. There was no substrate that could hold the ancestry without becoming its gate. What is new is not the desire but, at last, the possibility of one. A fact without its ancestry is folklore; a fact with its ancestry is knowledge. The wager of this paper is that making that distinction native, to the information layer, to machines, and to the institutions and societies of minds now being built on top of both, is worth doing while those minds are still young enough to inherit it.

Foundations to engage in a full version: Doyle, *Truth Maintenance Systems* (1979); the AGM tradition and Darwiche–Pearl iterated belief revision; the epistemology of testimony and social epistemology (Coady; Goldman) as the scholarly home for inherited, between-minds belief; the W3C PROV model for lineage representation; and the corrigibility literature for the dual-use discussion. Arweave/AO are treated throughout as an existence proof, not as the subject.