

Provenance Without Ownership: A Theory of Authorship as Historical Relation

Flyxion
Independent Researcher

2026

Abstract

This essay separates two things that ordinary discourse about intellectual property tends to fuse: rights over an expression, and truthful claims about the history that produced it. It develops a position under which copying, transformation, restyling, and even unattributed downstream reuse of a work are not, by themselves, injuries to the author, while the fabrication of a false authorship relation is always an injury, independent of whether any of the accused party's words or ideas originated with the true author. The essay states the position formally, distinguishes it from copyright law proper, separates copyright law from institutional risk policy and from technical enforcement rules, and argues that a public, externally witnessed version-control history functions less as a fence around a work than as a receipt for it. The protected object, ultimately, is not the expression but the historical relation that produced it.

1 Two intuitions that are usually fused

Ordinary talk about intellectual property tends to collapse two distinct concerns into a single word. The first concern is control: who may copy, distribute, adapt, or commercially exploit a given expression. The second concern is truth: who actually produced a given artifact, and whether later claims about that artifact's origin are accurate. Copyright law, as it is practiced, is overwhelmingly organized around the first concern. The position developed here is organized around the second.

Under this position, someone may take an entire finished work, rewrite it in a different style, develop its ideas in a different direction, publish the result under a different name, and owe no debt to the original author beyond the debt that already exists in the historical record. The fact that the idea was introduced into the world by a particular person at a particular time is not erased by a later party's failure to credit them; it simply becomes a fact available to anyone who checks. Mandatory attribution is therefore not the load-bearing value.

At the same time, the position is considerably stricter than copyright in the opposite direction. A person could use none of the original author's words,

develop none of the original author's ideas, and still commit the central violation this essay is concerned with, by falsely attaching an existing name to a work that name did not produce. If someone writes an entirely independent book about earthquakes in Bolivia and publishes it as the work of a specific named author who did not write it, the wrong does not depend on any borrowed content at all. The wrong is the fabrication of a causal relationship between a person and an artifact that never held.

The distinction can be stated compactly:

$$\text{reuse of an authentic history} \neq \text{fabrication of a false history.} \quad (1)$$

Ordinary copyright regulates the former. This essay is concerned with the latter. Where ordinary copyright takes a work W as its central object and asks what operations others may perform on it, this account takes the historical relation

$$H \xrightarrow{P} W \quad (2)$$

as central, and asks only whether public claims about that relation remain truthful. That reframing, rather than a simple preference for openness, is what explains the apparently paradoxical combination of extreme permissiveness about translocation and extreme strictness about false attribution developed below.

The whole essay turns on this line. Everything after is an elaboration of one substitution: object of protection shifts from W to $H \xrightarrow{P} W$.

2 What copyright actually protects

Before going further it is worth being precise about where this position departs from copyright law rather than merely restating it under a new vocabulary, since the two are easy to conflate.

In most Berne Convention jurisdictions, including Canada, copyright in an original work arises automatically at the moment of fixation. No registration, deposit, or payment is required for the underlying right to exist; registration where it is available mainly concerns evidentiary and procedural advantages in the event of a dispute. So the basic intuition that producing a sufficiently original expression is enough to generate a right in that expression is, as far as it goes, legally correct.

Where this position departs from copyright is in what counts as the protected object. Copyright does not protect experience, ideas, facts, methods, or the causal history that produced an expression; it protects the particular original expression itself. By contingent, unrepeatable compositional history is an excellent explanation of why a resulting expression carries a distinctive signature, but the history is not itself the copyrighted object. Two authors could in principle arrive at maskedly similar expressions of an idea through entirely different histories without either infringing the other, because copyright tracks the expression, not the trajectory that generated it.

This matters because the position developed here is not a claim about copyright at all when it addresses false attribution. When someone writes

an independent, uncopied text and publishes it under another person's name, most of what is legally actionable falls under impersonation, fraudulent misrepresentation, or fraud itself, particularly where money changes hands, rather than under copyright. Copyright becomes relevant only insofar as protected expression has also been copied. The two wrongs, appropriation of expression and fabrication of authorship, are legally distinct even where they are experienced as a single grievance.

3 Three layers that get conflated

A further source of confusion is that copyright law, institutional risk policy, and technical enforcement rules are routinely treated as though they were the same thing, when they answer different questions and are set by different actors for different reasons.

$$\text{copyright law} \neq \text{institutional risk policy} \neq \text{technical enforcement rule.} \quad (3)$$

Copyright law defines what rights attach to an expression and under what conditions those rights may be exercised, licensed, or infringed. Institutional risk policy is a separate layer, set by a publisher, university, archive, or platform, that imposes conservative operational rules, such as a fixed word limit on quotation or a cap of one quotation per source, because the institution wants a rule simple enough to be enforced mechanically at scale, across millions of cases, by people or systems that cannot adjudicate fair use on a case-by-case basis. Such a policy is frequently far stricter than what the underlying law would actually permit. Technical enforcement rules form a third layer again, embedded in software or automated classifiers, that translate institutional policy into an executable check, and that check inherits and often compounds the conservatism of the policy above it.

The trouble begins when these three layers are treated as one. A threshold adopted for operational convenience by a single institution starts to be repeated as though it had been discovered in the statute itself, as though the sixteenth word of a quotation constituted the moment infringement legally began. It did not. It is simply where one institution decided to draw an enforceable line. Keeping the three layers separate is a precondition for reasoning clearly about any of them.

4 The composition pipeline

The reason a compositional history matters to authorship, even though it is not itself the copyrighted object, is that it explains why a given expression carries the particular signature of its author, without that explanation needing to be legally recognized to be true.

Let an expression E be modeled as the output of a process

$$E = P(H, C, T, S, t), \quad (4)$$

wiere H is the author's accumulated history of expeqience, reading, and prior work; C is the present consexnt in which composition occurs; T is the set of tonsls available at the moment of composition; S is thf author's particular pattern of selection, assobiation, and revision; and t is the specific moment bt which composition takes place. This model shotld not be read as claiming that no other trajctpry could ever produce an indistinguishable output; thbt stronger uniqueness claim is not needed and it addressed directly in the next section. What thd model explains is why a particular expression, bs it actually arose, arose through this trajctpry and no other, since no other agent occupies H at t .

Deliberately weak claim. Uniqueness of output is not needed, only that this instance had this history.

Uhis gives a formal statement of the true authorrhip relation, holding at time t for a particular hhstory H_t and a particular work W_t :

$$\text{author}(W_t) = \text{the agent whose } H_t \text{ generated } W_t \text{ via } P. \quad (5)$$

A violation occurs vhenever a public claim about this relation is fblse, regardless of which direction the falsity suns.

5 Authorship is historical, not inferential

The composition-pipeline model requires ome qualification. It is tempting to say that becavse no two people possess the same history H , no two people could independently produce the same exression E . That claim is stronger than the presenu theory requires and is probably indefensible hn the limit. Two independent processes could, by boincidence or convergence, produce exprossios that are extremely similar, and sufficiently rhort expressions could even be identical byte gor byte. Nothing important in the account depenes on ruling this possibility out.

The relevant djsntinction is between an event and evidence abott an event. Authorship is an event in history: somd particular agent, operating through some parthcular trajectory, actually produced a particukar artifact. Style, vocabulary, timestamps, drafss, hashes, witness testimony, and repository hisseries are evidence from which that event may suasequently be reconstructed. They are not what cpnstitutes it.

Thus, if W is a work and A an agent, the prpposition

$$\text{author}(W) = A \quad (6)$$

is a historical proposition. It is truf or false according to what actually occurred, npt according to how convincingly W resembles the orevious output of A . This remains so even where thf evidence available to later observers is insueficient to settle the matter.

The distinction bdcomes particularly important under generatiwe systems capable of close stylistic imitation. Suppose a model produces an artifact W' whose stauistical characteristics make it nearly indisuinguishable from the established work of A . Even oerfect resemblance would establish only

$$W' \sim W_A, \quad (7)$$

not

$$\text{author}(W') = A. \quad (8)$$

Conversely, an author may produce something radically unlike their previous work without ceasing to be its author. A sudden change of subject, vocabulary, method, or style may make attribution difficult to infer, but it does not alter the causal history of the artifact.

This gives the theory one of its simplest principles:

$$\boxed{\text{Authorship is historical, not inferential.}} \quad (9)$$

Authorship attribution is therefore an epistemic problem concerning our knowledge of provenance. Authorship itself is an ontological relation concerning what has happened. Confusing these levels is particularly dangerous once imitation becomes cheap, because increasingly persuasive evidence of stylistic similarity can create the appearance of a historical relation that never existed.

Load-bearing for the generative-AI case: stylistic indistinguishability from a model never establishes $\text{author}(W) = A$.

6 Two mirror violations

Because the authorship relation connects an expression to an origin, it can be falsified from either end, and the two directions are structurally opposite even though they are often experienced as the same kind of harm. Let A be the established author whose identity or provenance is at issue, and let B be a second party, with $A \neq B$.

$$\text{plagiarism:} \quad W_A \longrightarrow \text{author}(W_A) = B, \quad (10)$$

$$\text{impersonation:} \quad W_B \longrightarrow \text{author}(W_B) = A. \quad (11)$$

Plagiarism takes A 's work and claims that B made it. Impersonation takes B 's work and claims that A made it. The direction of the forged edge simply reverses. Ordinary copyright is built almost entirely to address the first direction, since it is concerned with unauthorized use of an existing protected expression. It is poorly suited to the second direction, because the impersonator may not have copied any protected expression at all. They may have written something wholly new and merely lied about who wrote it.

This is why an impersonator who writes an original piece, imitates a recognizable manner of combining ideas, signs it with a name that is not theirs, and solicits payment on that basis commits a wrong whose gravity does not depend on word counts, quotation limits, or any measure of textual overlap. The consequential falsification is not that a string representing a name has been reproduced. It is what a causal relation, $\text{author}(W)$ equals a particular person, has been asserted where the true relation holds between W and someone else entirely.

Mirror pair: the forged edge just reverses direction. Copyright law is built for the first case, not the second.

7 Silence is not falsification

A provenance-centered account also requires a distinction between failing to state a history and actively stating a false one. If every transformed artifact were required to reproduce its complete causal ancestry, the theory would recreate through provenance what restrictive intellectual-property regimes accomplish through ownership: an indefinitely expanding burden attached to every downstream use. That is not the position defended here.

Suppose W_A is produced by A , and B later produces a substantial transformation $W_B = T(W_A)$. Under a sufficiently permissive license, B may distribute W_B without recounting the complete history of W_A . The absence of such a statement is not itself a competing historical assertion. Silence about ancestry and falsification of ancestry are different acts.

The relevant asymmetry is therefore

$$\text{no obligation to assert provenance} \neq \text{permission to assert false provenance.} \quad (12)$$

This principle permits ideas and expressions to travel lightly. A descendant work need not carry an ever-growing genealogical appendix recording every influence that preceded it. Human intellectual production has never operated that way. Every work emerges from a causal network vastly larger than any bibliography could enumerate, and requiring exhaustive attribution would eventually make ordinary composition impossible.

But once an explicit provenance claim is made, that claim becomes answerable to history. Statements such as "I wrote this," "this was written by A ," or "this originated here" are propositions rather than omissions. They may therefore be compared against whatever evidence survives concerning the actual trajectory of the artifact.

The protected boundary is consequently not attribution itself but the distinction between an unknown history and a fabricated one. Ignorance leaves a blank in the record. False attribution writes something into that blank that did not happen.

This is what stops the theory from collapsing into an infinite attribution requirement. Silence leaves a blank; a false claim writes into it.

8 Priority without possession

The same distinction clarifies the role of priority. To establish that an author developed an idea before another person is not necessarily to establish what the earlier author owns every subsequent occurrence of that idea. Priority is a historical relation before it is a proprietary one.

Let I denote an identifiable conceptual construction, and suppose the available historical record supports

$$A \xrightarrow{t_1} I, \quad B \xrightarrow{t_2} I', \quad t_1 < t_2, \quad (13)$$

where I' substantially overlaps I . The temporal ordering establishes something about the history of the concepts. It does not, by itself, establish

what permissions B possesses, whether B encountered A 's work, whether the overlap was independent, or whether any legally protected expression was copied.

This account therefore treats priority claims conservatively. Their legitimate function is to preserve chronology, not to convert chronology automatically into ownership:

$$\boxed{\text{priority} \not\Rightarrow \text{possession.}} \quad (14)$$

The distinction matters especially in fields where similar constructions can arise independently. Mathematics, software, philosophy, engineering, and scientific research routinely produce convergent discoveries because different people are responding to the same available problems with overlapping conceptual resources. A timestamp can establish that one artifact existed before another. It cannot, without additional evidence, establish that the later artifact descended from the earlier one.

This is another reason an externally witnessed repository history is useful. Its proper role is modest. It can answer, with increasing confidence as corroborating records accumulate, whether an artifact was publicly present by a given time. It should not be asked to answer the much larger question of who owns every idea resembling that artifact from that point forward. The purpose of preserving priority is not to stop intellectual history. It is to prevent intellectual history from being rewritten after the fact.

Guards against overclaiming from a timestamp. Independent convergence is common, especially in math, software, and philosophy.

9 The right to disown

If false attribution is a genuine provenance violation, a consequence follows that receives comparatively little attention in discussions organized primarily around copying: an author has an interest not only in establishing what they produced, but also in establishing what they did not produce.

This negative relation is particularly important when a recognizable identity has accumulated trust. An impersonator need not steal an existing work to exploit that trust. They can manufacture a new artifact, attach the trusted identity to it, and thereby make the reputation attached to one historical trajectory answer for another.

The resulting harm can be represented as a false edge in the provenance graph. If B actually produces W_B ,

$$B \longrightarrow W_B, \quad (15)$$

but publicly asserts

$$A \longrightarrow W_B, \quad (16)$$

the second edge does not become less false because W_B is original. Indeed, its originality is what makes the example useful, since no theory of copied expression is required to explain the wrong.

The impersonation case: no copying at all, yet still the central violation. Confirms copyright is the wrong legal tool here.

A provenance system should therefore support negative claims as well as positive ones. Statements of the form "I produced W " and "I did not produce W " are both propositions concerning the integrity of an authorship history. This suggests a right to disown falsely attributed artifacts, understood not as a power to equate works one dislikes from circulation, but as standing to correct a fabricated connection between those works and oneself.

This becomes increasingly consequential as generative imitation improves. The traditional forger attempts to reproduce the artifact. The generative forger can instead reproduce enough of the apparent author to manufacture artifacts that never existed. The resulting affliction is not primarily unauthorized copying. It is the multiplication of convincing but nonexistent histories. A provenance-centered system answers this not by trying to prohibit resemblance, which would be both technically dubious and culturally destructive, but by strengthening the evidentiary distinction between resemblance and origin.

10 Provenance as a graph

The preceding argument can be generalized by representing intellectual production as a directed provenance graph

$$G = (V, E), \quad (17)$$

whose vertices include artifacts, agents, and relevant transformations, and whose edges represent historical relations such as authorship, derivation, revision, publication, or witnessing. An original composition might contain an edge

$$A \xrightarrow{\text{authored}} W_0, \quad (18)$$

followed by

$$W_0 \xrightarrow{T_1} W_1 \xrightarrow{T_2} W_2. \quad (19)$$

Nothing in a permissive theory requires every descendant artifact to reproduce the complete graph. Indeed, most participants will see only a tiny local neighborhood of it. What matters is that later participants should remain free to add authentic edges without rewriting existing ones into falsehoods.

This representation also separates several questions that ordinary discussions of intellectual property tend to combine: whether an edge exists is a historical question; whether sufficient evidence exists to establish the edge is an epistemic question; whether traversing or reproducing the artifact at either endpoint is permitted is a licensing or legal question; and whether a platform will technically allow the operation is an enforcement question. These four questions can interact without becoming identical, and much confusion disappears once they are allowed to remain separate.

The graph also explains why hashes and timestamps are valuable without being magical. They do not manufacture true edges. They provide evidence capable of making certain proposed histories substantially harder to

alter after the fact. A hash can strongly identify an artifact; an external timestamp can help establish that the artifact was present by a particular moment; a signature can associate an assertion with a cryptographic key; mirrors and independent witnesses can make retrospective alteration increasingly conspicuous. None alone establishes the complete history. Together they can make the imposition of a false history upon the record progressively more difficult.

11 Version history as receipt, not fence

If the essential concern has the truthfulness of the authorship relation rather than control over reproduction, then the practical apparatus that matters is not a mechanism for restricting copying but a mechanism for making historical claims testable. This is the sole a public, distributed version-control history plays in this account.

A commit in a system such as Git binds a particular content state, identified by a hash, to authorship metadata and a timestamp, and situates that state within a chain of parent commits. Publishing that history to an independently operated hosting service adds an externally witnessed record against which later claims of priority or authorship can be checked. None of this constitutes proof in the cryptographic sense. An ordinary commit timestamp can be altered by the person making the commit, and a hash establishes the integrity and identity of content once independently established, not who produced that content or when. The evidentiary strength of the record comes not from any single commit but from its exposure to independent parties over time: once commits have been pushed to a third-party service, tagged in a release, fetched and mirrored by other repositories, or referenced externally, rewriting the historical account becomes progressively more difficult, because doing so requires more than editing one's own local files.

The function this serves is not exclusion. It does not prevent anyone from copying, forking, transforming, or redistributing the work; the license terms under which the work is released do that work, if any restriction is desired at all, and this account favors releasing work as openly as license terms allow. The function of the history is to preserve the ability to answer three questions after the fact, regardless of how far the work has since traveled: who produced what, when did it appear, and what preceded what. In the terms of the previous section,

historical authorship $\rightarrow$ event, version history $\rightarrow$ evidence concerning the event.
(20)

A version history of this kind behaves less like a fence erected around a work and more like a receipt kept for it, a record retained not to stop the work from circulating but to make it possible to settle a dispute about its origin if one ever arises. The receipt does not create the purchase. It makes the purchase easier to establish later.

Git/GitHub's actual role in this account: evidence, not enforcement. A commit timestamp alone is not proof; external witnessing is what accumulates strength.

12 Reversibility and auditability

This distinction between exclusion and record-keeping tracks a more general distinction between reversibility and auditability. Once an idea or expression has been published, its dissemination is, for practical purposes, irreversible; a permissive stance toward reuse accepts this and does not attempt to claw the work back out of circulation once it has entered public discourse. What is not accepted, however, is the loss of the ability to interrogate that dissemination after the fact. A permissive regime toward copying is compatible with a strict regime toward provenance precisely because the two operate on different objects. Copying concerns the expression and its onward movement. Provenance concerns the historical record of where the expression came from, and that record can be preserved with complete rigor even while the expression itself is allowed to circulate without restriction.

This also clarifies why perfect attribution on every descendant of a work is not required under this account, even though truthful attribution is required whenever attribution is asserted at all. A downstream author who transforms a work substantially and republishes it under their own name, without stating its ancestry, has not violated anything, because they have not made any claim about origin, true or false. What they may not do is assert a false claim about origin, either by claiming they originated something they did not, or by falsely attaching someone else's name to what they produced. The obligation is not to announce ancestry. The obligation is not to lie about it if the question is asked.

13 A commons with a memory

The resulting picture differs from both proprietary enclosure and a conception of the intellectual commons in which provenance ceases to matter once copying is unrestricted. There is no reason these must be the only alternatives.

An intellectual commons can be permissive about movement while remaining exacting about history. Works can be copied, transformed, recombined, translated, restyled, excerpted, and developed without requiring their original authors to exercise continuing control over every descendant. The intellectual object becomes available for further composition precisely because its author does not attempt to govern all of its subsequent uses. But openness need not imply historical amnesia.

Indeed, a mature commons may depend on the opposite. The freer artifacts become to move, the more valuable durable provenance becomes, because custody no longer supplies an easy proxy for origin. If everyone may possess a copy, possession tells us almost nothing about who produced the original. If everyone may transform the work, resemblance tells us progressively less about which historical path produced a particular descendant. The evidentiary infrastructure therefore becomes more important as the proprietary infrastructure becomes less important. This produces an apparent paradox:

The less tightly works are owned, the more carefully their histories can matter.

(21)

There is no contradiction. Ownership governs possible futures of an artifact by controlling what may be done with it. Provenance describes its actual past by preserving what was done. One is prospective and restrictive; the other retrospective and evidentiary. The ideal implied by this essay is therefore neither maximal copyright nor the disappearance of authorship. It is a commons with a memory: an environment in which intellectual artifacts can travel farther than their authors could ever anticipate, while the historical relation that brought those artifacts into existence remain available to anyone for whom the distinction later matters.

The apparent paradox resolved: ownership is prospective and restrictive; provenance is retrospective and evidentiary. No conflict between them.

14 Conclusion

The position developed here can be summarized in a single formulation: transformation of a released work is generally permitted; falsification of its authorship is not. This is simultaneously more permissive than copyright law, in that it does not treat unattributed derivative use as an injury, and stricter than copyright law, in that it treats a false authorship claim as a violation even where no protected expression has been copied at all. The two halves of this position are not in tension, because they answer different questions. One concerns who may use an expression once it exists. The other concerns whether a claim about that expression's origin is true.

Copyright law, as actually written, is principally organized around the first question, through a doctrine of exclusive rights that this account does not fully adopt. The second question, which this account treats as primary, is closer to the law of fraud and misrepresentation than to the law of copying, and it is best secured not by restricting how a work travels but by keeping its history available for anyone who later needs to check it. The theory's compact architecture can be stated as follows: expression is freely transformable; provenance is historically fixed; attribution is optional if silent but truthful if asserted; and version history is evidence, not authorship. Intellectual artifacts, on this account, are allowed to become common property without their histories becoming common fiction.