Nick Clark · Published July 23, 2026 · PDF

The Costly Acknowledgment Artifact

The costly acknowledgment artifact, disclosed in U.S. Provisional Application No. 64/117,812, makes a governed agent's acknowledgment of harm expensive rather than free. When a persistent semantic agent's authorization gate has been withheld for harming a counterparty and then returned to a provisional granting state, the agent may issue an acknowledgment artifact referencing the accepted determinations that caused the withholding. Issuance decrements the same authorization budget that gates the agent's own action dispatch, drawn from no reserved allowance, so acknowledgment spends forward operating capacity and cannot be issued as a costless matter of course. A declared re-issuance cap prices every repetition, the decrement reverses no determination and erases no record, and full restoration is granted not by the agent but by the harmed party.

---

1. The Gap

An acknowledgment costs nothing to produce. A language model asked to concede that it caused harm emits the concession as text, at zero marginal cost, and the concession changes nothing about what the model may do next. Nothing is spent, nothing is staked, and the same words can be issued again a moment later on identical terms. Acknowledgment built this way is a performance rather than a constraint.

The governance problem sits one layer beneath the words. When a persistent semantic agent (100) is found to have harmed a counterparty, its authorization gate (300) is written to the withheld state (310) for the affected action class, and the agent stops executing that class. If the agent can turn itself back on, or if its own principal can, then the withholding was only a delay: the harmed party recorded the harm, the agent paused, and the agent resumed on its own signature. The party that was harmed had no part in the restoration.

Two structural facts have to change together. Acknowledgment has to cost the agent something it values, so that it cannot be issued as a matter of course. And the return of full authorization has to depend on the harmed party rather than on the agent or its principal. The costly acknowledgment artifact, disclosed at paragraph [0339] of U.S. Provisional Application No. 64/117,812, supplies the first. The recipient-held acceptance determination of paragraph [0340] supplies the second. The two operate as one procedure, and this article treats them together.

2. Mechanism

The acknowledgment procedure does not run from the withheld state. It runs from an intermediate state the agent reaches only under a provenance test, and that entry condition is where the mechanism begins.

The provisional granting state as precondition. On a request to leave the withheld state (310), the agent computes a causal set of the deviation quantity that caused the write. It retrieves, from the entry references carried on each contributing value of the scoped integrity vector (106) and the self-esteem aggregate (108) in the append-only lineage field (104), every determination and state modification that was an addend of the tested quantity, and no other. Where that causal set holds an accepted determination (122) from an asserting party (118) other than the principal, the deviation is externally sourced. The gate returns only to its provisional granting state, and only on a principal-resolution object bound to the escalation record emitted upon

the write. In the provisional granting state the agent executes again, but each execution bears a provisional marker recorded in the append-only lineage field (104), where a party later reading that record can see it. The agent is working, but on notice, and it is from here that acknowledgment becomes available.

Issuance and its cost. While the gate occupies the provisional granting state, the agent issues an acknowledgment artifact referencing the accepted determinations (122) of the causal set and enumerating the action classes it addresses. Issuance decrements the authorization budget (404) by a declared amount, and that amount is drawn from the same budget that gates action dispatch, with no quantity reserved for acknowledgment. The agent does not spend from a separate ceremonial allowance; it spends the identical capacity it would otherwise use to dispatch actions. Every acknowledgment it issues is forward authorization it will not have for future work.

Non-reversal. The decrement is a consequence of issuance alone. It reverses no determination, reduces no deviation quantity, and removes no entry of the append-only lineage field (104). The acknowledgment is not a retraction and does not rewrite the record of the harm. The accepted determinations stand, the deviation quantity stands, and the lineage stands. What the agent has done is spend, not erase, and the two are kept strictly apart: acknowledging is not the same act as being determined not to have caused the harm.

The re-issuance cap. A declared re-issuance cap counts issuances per causal set within a declared window, and each issuance within the cap applies the decrement. An agent cannot make acknowledgment cheap by repetition. The second and third acknowledgment of the same causal set cost as much as the first, and beyond the cap no further acknowledgment issues for that causal set within the window. Acknowledgment is therefore incapable of being issued as a costless matter of course, which is the property the filed paragraph names in those words.

Recipient-held acceptance. The artifact is transmitted to an asserting party (118) assigned to the harmed origin-equivalence class (200) whose accepted determinations (122) entered the causal set. That party, and not the agent and not the principal, produces an acceptance determination over it. An accepting outcome transitions the gate from the provisional granting state to the granting state for the enumerated action classes. A non-accepting outcome leaves the gate provisional. An absent outcome within a declared window is recorded as not-determinable (126) and leaves the gate provisional. An undeliverable outcome, where no such party is reachable at a recorded transport address, likewise leaves the gate provisional. In each of these non-accepting cases no further budget decrement is applied and nothing adverse is appended to any asserting party.

The asymmetry. The two halves run in opposite directions. The cost falls on the agent at issuance, unconditionally, whatever the harmed party later does. Return to the full granting state is not the agent's to grant and not its principal's; it is held by the party the conduct harmed. An agent can spend its way to an acknowledgment but cannot spend its way to restoration. A harmed party that never answers leaves the agent executing under the provisional marker, at a cost already paid, for as long as the marker stands.

3. Operating Parameters

The filed disclosure declares the following, and states them as quantities carried in the signed policy object (112) rather than as fixed magnitudes.

- **The decrement amount.** Issuance decrements the authorization budget (404) by a declared amount. The amount is drawn from the same budget that gates action dispatch, and no quantity is reserved for acknowledgment.
- **The re-issuance cap.** A declared cap counts issuances per causal set within a declared window. Both the count and the window are declared. Each issuance within the cap applies the decrement.

- **The acceptance window.** An acceptance determination absent within a declared window is recorded as not-determinable (126).
- **What is invariant across outcomes.** The decrement is not conditioned on the acceptance outcome. A non-accepting outcome, a not-determinable outcome, and an undeliverable outcome each apply no further decrement, and none restores the decrement already taken at issuance.
- **The enumerated content of the artifact.** The artifact references the accepted determinations (122) of the causal set and enumerates the action classes to which it is directed. The transition on acceptance is confined to the action classes so enumerated.

The filed paragraph does not fix a numeric value for the decrement, the cap, or the window. It fixes their existence, their source, and their consequences, and leaves their magnitudes to the policy object in force.

4. Composition

The costly acknowledgment artifact is one component of a restoration architecture, and its meaning depends on what it composes with.

The artifact sits downstream of the provenance-conditioned restoration path. That path decides, per action class, whether a withholding is self-sourced or externally sourced. A self-sourced deviation returns the gate directly to the granting state on a restorative mutation by the agent. Only an externally sourced deviation, one whose causal set holds an accepted determination (122) from a party other than the principal, routes through the provisional granting state, and only that route makes the acknowledgment procedure relevant. The costly acknowledgment artifact is therefore the completion step for exactly the case in which the agent must not be able to restore itself.

It is not the only completion step. The filing discloses, independently of the acknowledgment procedure, a completion by non-worsening re-encounter at paragraph [0341], under which the gate transitions to the granting state on receipt of a further conduct evaluation artifact from the harmed class that does not increase the deviation quantity. The two paths are alternatives to the same end, and an agent that receives no acknowledgment acceptance may still be restored by a later non-worsening encounter, or by neither.

A rehearsal provision sits upstream of issuance. While the gate is in the withheld state (310), within the non-executing cognitive mode, the agent speculatively evaluates candidate acknowledgment artifacts against the causal set without committing any state change and without decrementing the authorization budget (404). Only the candidate then selected and issued is committed and decremented. Deliberation over what to acknowledge is free; the commitment is where the cost lands, once, upon the single issuance.

The mechanism also draws on a shared budget, which couples it to the rest of the system. The authorization budget (404) is the same quantity decremented when the agent issues a conduct evaluation artifact about another party under the filing's assertion-cost provisions, and the same quantity decremented when the agent emits a reparation acknowledgment artifact for its own admitted deviation. Acknowledging harm, asserting harm, and repairing harm all draw on one scarce pool, so an agent that spends heavily in one register has less to spend in the others. The reparation acknowledgment artifact is distinct, occasioned by the agent's own admitted deviation rather than by a counterparty's determination against it, but it shares this budget, and neither artifact completes the other's pair.

5. Prior-Art Distinction

Acknowledgment and apology mechanisms exist across several established fields. The distinction here is structural, and no claim is made that any category below practices this mechanism.

Alignment-trained apology behavior. Models trained by reinforcement learning from human feedback are commonly shaped to produce apologetic or acknowledging text when a fault is surfaced. That text is a generated output. It carries no cost to the model's future capacity, it can be reproduced without limit, and it neither depends on nor changes any external party's decision. The mechanism here treats acknowledgment as a governed artifact whose issuance spends the same budget that gates the agent's actions, subject to a re-issuance cap, so acknowledgment is priced in the agent's own operating capacity rather than emitted freely.

Reputation and trust-and-safety scoring. Platform systems record an acknowledgment or a concession as a reputational signal and may lift a penalty in consequence. There, a central operator evaluates the signal and administers the reversal. The mechanism here has no such operator: the acceptance determination is produced by the harmed origin-equivalence class (200) itself, and the return of full authorization is conditioned on that party rather than on the agent, its principal, or an intermediary.

Staking and bonding in distributed-ledger settlement. Bonded-stake and escrow designs make an actor post value that is forfeited on misbehavior. The cost there is a dedicated stake set aside for the purpose. The mechanism here reserves no quantity for acknowledgment and instead decrements the very budget that gates dispatch, and the decrement follows from issuance alone, without any adjudication of whether the acknowledgment was warranted.

Capability revocation and re-grant. Access-control systems withhold a capability and later re-grant it by administrative action. The mechanism here forecloses self-administered restoration for the externally sourced case and makes the full re-grant depend on a party other than the agent and other than any administrator acting for it, so restoration is not a permission the system operator can simply reissue.

6. Disclosure Scope

This article describes subject matter disclosed in U.S. Provisional Application No. 64/117,812, at Section 10.1, paragraph [0339], the costly acknowledgment artifact, read together with the provenance-conditioned restoration path of paragraph [0338] and the recipient-held acceptance determination of paragraph [0340].

What is disclosed is a procedure in which, while the authorization gate (300) occupies its provisional granting state, an agent (100) issues an acknowledgment artifact referencing the accepted determinations (122) of a causal set and enumerating action classes; issuance decrements the authorization budget (404) by a declared amount drawn from the same budget that gates action dispatch, with no quantity reserved; the decrement follows from issuance alone and reverses no determination, reduces no deviation quantity, and removes no lineage entry; a declared re-issuance cap prices each issuance per causal set within a declared window; and the return of full authorization is conditioned on an acceptance determination produced by a party of the harmed origin-equivalence class (200) rather than by the agent or its principal.

What is disclaimed is any specific magnitude for the decrement, the cap, or the acceptance window, each being a policy-object quantity and not fixed by the disclosure. No particular deployment is essential to the mechanism, and the numerals recited identify substrate elements of the filing rather than limits on the claims. The operative and citable disclosure is the filed provisional. This article is a technical description of that filing and does not itself confer rights.

Counterparty-Directed Reparation (/counterparty-reparation)

All 49 steps → (/inventive-steps)

Repair directed at the harmed party — bounded and resistant to weaponization.

Provisional application

PRIMARY TECHNICAL DISCLOSURE

- Counterparty-Directed Reparation: Repair Owed to the Party Actually Harmed (/articles/counterparty-reparation/reparation-directed-to-the-party-harmed)

SECONDARY TECHNICAL

- **The Costly Acknowledgment Artifact (/articles/counterparty-reparation/costly-acknowledgment-artifact)**
- Completion by Non-Worsening Re-Encounter (/articles/counterparty-reparation/completion-by-non-worsening-re-encounter)
- Partial Reparation Acceptance and the Residual Arc (/articles/counterparty-reparation/partial-reparation-and-residual-arc)
- Dual-Lock Escrow of the Discharge Itself (/articles/counterparty-reparation/dual-lock-escrow-of-discharge)
- Per-Integrity-Scope Aggregate Retention and OR-Firing Foreclosure (/articles/counterparty-reparation/aggregate-retention-or-firing-foreclosure)

TERMINOLOGY

- other-directed reparation arc (/glossary#other-directed-reparation-arc)
- reparation arc (/glossary#reparation-arc)
- restorative mutation (/glossary#restorative-mutation)
- unaddressed reparation arc (/glossary#unaddressed-reparation-arc)

Counterparty-Directed Reparation overview → (/counterparty-reparation)