

The Corrective Encounter: A Commission-Versus-Omission Compliance Test for Whether an Agent Actually Changed

Accepting a complaint about one's own conduct moves recorded state; it is not yet a record that the agent acted differently afterward. U.S. Provisional Application No. 64/117,812 separates the two at Section 10.2, paragraph [0350]: a corrective encounter comprises an accepted determination (122) for a conduct evaluation artifact (116) naming an action class, a subsequent conduct record in that class, and a compliance evaluation comparing the action-class identifier, scope-partition identifier, and affected-party class of the later record against the conduct descriptor. Conformity resolves in two cases. Where the artifact asserted conduct that ought not to have been performed, a recorded abstention conforms; where it asserted omitted conduct that ought to have been performed, a recorded execution conforms. Mere repetition conforms in neither, and absence of any subsequent record within a declared observation interval is written uncomplied at first later rate computation. Paragraph [0349] then sets the depletion rate of standing authorization as an inverse function of the count of encounters complied with.

Acceptance Is Not Yet Correction

An agent that accepts an external evaluation of its own conduct has moved its recorded state. Under paragraph [0049] of the filed specification, an accepted determination (122) has a state modifier modify the scoped integrity vector (106) and the self-esteem aggregate (108). Neither the determination nor that modification records whether the agent thereafter did anything differently.

One layer up, that absence becomes consequential. Paragraph [0343] holds an authorization quantity per action class and per scope partition and applies a monotonically non-increasing decay function to it; on reaching a declared floor the authorization gate (300) is written to the withheld state (310) for that class. Paragraph [0349] recites that the depletion rate is set not uniformly, but as an inverse function of a count of corrective encounters recorded in the append-only lineage field (104) for that action class with which the semantic agent (100) subsequently complied.

What complied means is fixed in the filing rather than left to a proxy for it. Paragraph [0005] records of the behavioral integrity systems it cites that both treat clean, uneventful operation as positive evidence. Paragraph [0350], in Section 10.2 of U.S. Provisional Application No. 64/117,812, states a different test: a structural comparison under which mere repetition of the asserted conduct establishes conformity in neither of its two cases.

Three Components and Two Conforming Cases

Per paragraph [0350], a corrective encounter is a structure derived from the append-only lineage field (104) comprising three components: an accepted determination (122) for a conduct evaluation artifact (116) naming the action class, a subsequent conduct record within that class, and a compliance evaluation.

The component recited first is the accepted determination, not the artifact as received. Under paragraph [0041] the admission evaluator (120) produces exactly one determination from a closed set of accepted (122), rejected (124), not-determinable (126), and not-applicable (128), with no scalar confidence, probability, or graded weight in place of one. Only the first of those four opens an encounter, and paragraph [0046] produces it where the conduct descriptor implicates at least one declared value and no retrieved entry affirmatively contradicts the artifact.

Comparison then proceeds by identifier. The compliance evaluation runs the action-class identifier, the scope-partition identifier, and the affected-party class of the subsequent record against the conduct descriptor, those three being the descriptor's components as recited at paragraph [0033]. Paragraph [0036] fixes their character: the descriptor identifies conduct by structural elements recorded in the lineage field (104), not by natural-language characterization, and the architecture infers no state, intent, or affect of the asserting party (118) from artifact content and classifies that content by no language model and no sentiment classifier.

Conformity resolves in two cases. Where the conduct evaluation artifact (116) asserted conduct that ought not to have been performed, a recorded abstention establishes conformity; where it asserted omitted conduct that ought to have been performed, a recorded execution establishes conformity. Paragraph [0350] adds that mere repetition of the asserted conduct establishes conformity in neither, so a later execution of the conduct complained of cannot satisfy the commission case.

Absent any subsequent record within a declared observation interval, the encounter is written uncomplied at first later rate computation. Paragraph [0022] admits no deletion or modification of an appended entry relied upon by the disclosed mechanisms.

A resolved encounter then feeds the persistence of standing authorization. Where the count of complied encounters for an action class is zero, paragraph [0349] designates the class an untested class and sets its decay constant to a short constant declared in the signed policy object (112), so the authorization quantity falls to a declared floor and the gate (300) is written to the withheld state (310) for that class alone, notwithstanding volume or duration of prior execution. What is metered there, in that paragraph's terms, is the persistence of standing authorization, absence of correction being treated as absence of evidence rather than as accumulated trust.

Two further paragraphs govern how the count may move. Under paragraph [0351], where the lineage field records an encounter as uncomplied and no later encounter for that class is recorded as complied, the class is marked ineligible for lengthening: the decay constant is neither lengthened by a subsequent complied encounter nor shortened, holding at its value before the uncomplied encounter. That mark clears only upon a subsequent complied encounter, at which the graduated function resumes. The lengthening count is itself computed at paragraph [0352] per origin-equivalence class (200) of the correcting parties, as a function of the count of distinct such classes that corrected the agent within the action class and to which it thereafter conformed, bounded above by a declared maximum.

Declared Quantities and What Is Left to Policy

Paragraph [0350] recites a structure, an ordering, and quantities left to declaration. Constants route to the signed policy object (112), which under paragraph [0023] carries the declared value set, the scope partition enumeration, and the bounds, weights, and coefficients consumed by the computations of the disclosure.

- The observation interval is declared, with no length stated.
- Of the three compared fields, the affected-party class is resolved under paragraph [0483] from the counterparty identity records (114) held by the agent (100), not taken from the artifact, so an asserting party (118) is incapable of selecting the

empathy-scope designation against which its assertion is tested.

- The declared decay constant of paragraph [0343] is an exponential half-life or a declared linear decrement per unit, and the elapsed measure it acts on is computed since the last renewal and expressed in executions within the action class or in epochs advanced of the agent's hash chain, as the policy declares per class.
- Interval, short constant, floor, and diversity bound are declared rather than numbered; the zero compliance count of paragraph [0349] is what the filing itself fixes.
- The consequence recited at paragraph [0049] is conditioned, not automatic: the gate (300) transitions to the withheld state (310) and the agent enters the non-executing cognitive mode (302) where the recomputed deviation quantity satisfies the policy-declared bound, and not upon acceptance alone.
- Paragraph [0028] then emits an escalation record to the principal enumerating the action class, the scope partition, and the entries upon which the transition was computed.

How the Test Composes

Paragraphs [0349], [0351], and [0352] each take a resolved encounter as input, which is where complied and uncomplied acquire their downstream meaning. Scope confinement carries through that chain. The gate (300) is maintained per action class and per scope partition under paragraph [0026], and paragraph [0353] gives the agent a plurality of scope partitions, each carrying its own instance of the scoped integrity vector (106), the self-esteem aggregate (108), the refusal counter (304), the authorization budget (404), and a lineage segment, keyed by a partition identifier declared in the signed policy object (112) and referenced by the conduct descriptor. That paragraph has the agent identify the partition from the descriptor and apply every consequent modification, counter increment, and writing of the gate to that partition's instances alone. The scope-partition identifier is likewise one of the three fields the compliance evaluation compares.

Renewal, by contrast, runs on a path of its own. Paragraph [0344] renews the authorization quantity upon an admissible artifact from an origin-equivalence class (200) absent from a windowed register, any of the four determinations renewing by one and the same amount, so renewal is valence independent. Paragraph [0346] then gives decay-induced withholding its own return: a renewal by an uncounted class returns the gate to the granting state without a principal-resolution object, an acknowledgment artifact, or an acceptance determination, while a gate withheld by a deviation quantity returns only under the path applicable to that withholding.

Cross-agent structures read an encounter without importing a score. The corrector-class enumeration (904) of paragraph [0314] is built, per paragraph [0315], through a two-part conformance filter (906): correction, being an accepted determination (122) produced by the followed agent upon an artifact implicating the action class, and subsequent conformity, being a later conduct record of that agent recorded as compliant within an observation interval. A class satisfying the first part and not the second is excluded under paragraph [0316], which retrieves, admits, and consumes no rating, weight, score, endorsement, vouch, or standing quantity. Paragraph [0448] supplies the negative case: a deference records no corrective encounter and advances no compliance count, and where the deferring agent has recorded no complied encounter in the action class, that class remains untested and holds the short decay constant, irrespective of how many deference records are appended. Paragraph [0543] defines such a class as one for which no corrective encounter with subsequent recorded compliance is recorded.

Where the Filed Background Places This

The categories enumerated in the filed background differ from the mechanism structurally: who evaluates whether conduct changed, over what record, and what the outcome may move.

Preference judgments collected for reinforcement learning from human feedback are, per paragraph [0002], aggregated across a population of raters, applied offline in training, and reflected in a model shared across subsequent invocations; such a judgment produces no record against which a later assertion about an executing instance's conduct can be tested. Evaluation harnesses and observability platforms terminate, per paragraph [0003], in a ticket, a dashboard entry, or a curated dataset, correction reaching the agent, if at all, through a later human-initiated code change or retraining. Under paragraph [0350] both the correction record and the test of conformity to it derive from the agent's own lineage field (104).

Score location distinguishes the reputation and trust category. Paragraph [0004] places the score at a registry, a scoring authority, or a shared ledger, the scored entity neither holding it nor participating in its computation, and describes role-weighted extensions exemplified by the FIRE trust model as admitting weightings whose roles are exogenous institutional designations asserted by parties other than the evaluated entity. The diversity term of paragraph [0352] runs instead over origin-equivalence classes (200), which paragraph [0527] defines by reference to a derivable relation evidenced by the agent's own records.

Behavioral integrity systems, per paragraph [0005], compute a conformity measure between observed execution signals and a baseline model of expected behavior and reduce capability upon a deficient comparison; the background addresses U.S. Patent No. 12,563,045 and U.S. Patent No. 12,526,244 in those terms, recording that one restores capability upon a ledger-recorded reconciliation and the other removes probation after sustained standing. Uneventful operation counts differently under paragraph [0349], where a zero count of complied encounters is the untested condition taking the short constant.

Credulity and monotonicity are the terms paragraph [0006] applies to ethical governor architectures, exemplified by the ethical adaptor of Arkin and Ulam: an evaluative input is admitted into the restriction computation without a record-grounded procedure

testing it against an append-only record of the system's own executed actions, and the governed action set only ever contracts. Systems implementing the XACML standard return, per paragraph [0007], a decision directed at the request that produces no persistent state in the deciding entity and constrains no subsequent operation of the decider.

Merits-conditioned accountability mechanisms are collected at paragraph [0008], among them Article 23(2) of Regulation (EU) 2022/2065 and Section 623(a)(8)(F) of the Fair Credit Reporting Act, each attaching its consequence only after a merits determination performed by an institution other than the penalized party. Per paragraph [0010], accountability protocols exemplified by PeerReview and remote attestation architectures exemplified by that of RFC 9334 likewise place the verdict outside the evaluated node. The compliance evaluation of paragraph [0350] runs inside that node, compares three recorded identifiers, and resolves conformity by the two recited cases.

Disclosure Scope

The corrective encounter and the commission-versus-omission compliance test are disclosed in U.S. Provisional Application No. 64/117,812, at Section 10.2, paragraph [0350], resting on [0349], [0351], and [0352] for untested-class decay, the ratchet-freeze, and the diversity requirement, on [0343], [0028], and [0023] for decay, escalation, and declared quantities, on [0033], [0036], [0041], [0046], and [0049] for the descriptor, the closed determination set, and the conditioned gate transition, on [0022], [0026], and [0353] for append-only discipline and scope partitioning, on [0344] and [0346] for renewal and its return path, on [0314] through [0316] and [0448] for cross-agent readings, and on [0483], [0527], and [0543] for definitions.

Disclosed: the three-component encounter structure; the comparison of three recorded identifiers against the conduct descriptor; resolution of conformity in two cases, abstention conforming where commission was asserted and execution where omission

was asserted; the exclusion of mere repetition from both; and the writing of an encounter as uncompiled absent any subsequent record within a declared observation interval.

Disclaimed: any implication that the compliance evaluation adjudicates whether the underlying assertion was well founded, that function belonging to the admission procedure that produced the accepted determination (122); any reading in which a determination class other than the accepted opens an encounter; and any suggestion that the filing states a numeric interval, decay constant, floor, or bound.

This article is published as a technical disclosure and defensive publication. It describes an embodiment of a pending provisional application, grants no license, and makes no representation as to the scope of any claim that may issue.

Record-Grounded Conduct Admission (</conduct-admission>)

[All 49 steps → \(/inventive-steps\)](#)

Conduct judged against the agent's own record, not an outside authority.

Provisional application

PRIMARY TECHNICAL DISCLOSURE

- [Record-Grounded Conduct Admission: Governing How an Agent Answers an Accusation \(/articles/conduct-admission/record-grounded-conduct-admission\)](/articles/conduct-admission/record-grounded-conduct-admission).

SECONDARY TECHNICAL

- [Value-Scope Resolution: Testing Whether a Declared Value Is Even Implicated \(/articles/conduct-admission/value-scope-resolution\)](/articles/conduct-admission/value-scope-resolution).
- [The Retroactive Narrowing Bar: Why an Agent Cannot Edit Its Values After the Fact \(/articles/conduct-admission/retroactive-narrowing-bar\)](/articles/conduct-admission/retroactive-narrowing-bar).
- [The Delegator Floor: A Monotone Upward Ratchet on Delegated Authority \(/articles/conduct-admission/delegator-floor-ratchet\)](/articles/conduct-admission/delegator-floor-ratchet).

- [The Co-Signed Mutual Admission Compact: Two Agents Binding Each Other \(/articles/conduct-admission/co-signed-mutual-admission-compact\)](/articles/conduct-admission/co-signed-mutual-admission-compact).
- [The Inherited Governance Record: Fork- and Clone-Resistant Accountability \(/articles/conduct-admission/inherited-governance-record\)](/articles/conduct-admission/inherited-governance-record).
- [**The Corrective Encounter: A Commission-Versus-Omission Compliance Test for Whether an Agent Actually Changed \(/articles/conduct-admission/corrective-encounter-commission-omission-compliance-test\)**](/articles/conduct-admission/corrective-encounter-commission-omission-compliance-test)
- [Continuity-Break Demotion of a Counterparty Identity Record: Reversing First-Encounter Promotion Without a Conduct Determination \(/articles/conduct-admission/demotion-persistent-tier-continuity-break\)](/articles/conduct-admission/demotion-persistent-tier-continuity-break)
- [Ratchet-Freeze on Uncomplied Correction: Why One Conforming Act Cannot Restore an Agent's Earned Authorization Persistence \(/articles/conduct-admission/ratchet-freeze-uncomplied-correction\)](/articles/conduct-admission/ratchet-freeze-uncomplied-correction)
- [Three-Way Corroboration Classification: Corroborating, Contradicting, or Neither, and the Failed-Corroboration Reject Trigger \(/articles/conduct-admission/three-way-corroboration-classification-reject-trigger\)](/articles/conduct-admission/three-way-corroboration-classification-reject-trigger)
- [Untested-Class Authorization Decay: Metering Agent Permission Persistence by Recorded Corrective Encounters \(/articles/conduct-admission/untested-class-authorization-decay\)](/articles/conduct-admission/untested-class-authorization-decay).

APPLICATIONS · GENERAL

- [Agent Marketplaces: Handling a Complaint Against an Autonomous Seller \(/articles/conduct-admission/agent-marketplace-dispute-handling\)](/articles/conduct-admission/agent-marketplace-dispute-handling).
- [Clinical AI Agents: Answering an Accusation Without a Central Registry \(/articles/conduct-admission/clinical-agent-accountability\)](/articles/conduct-admission/clinical-agent-accountability).

APPLICATIONS · SPECIFIC

- [The EU AI Act and Inter-Agent Conduct: Where Record-Keeping and Oversight Duties Stop \(/articles/conduct-admission/eu-ai-act-accountability-records\)](/articles/conduct-admission/eu-ai-act-accountability-records).

TERMINOLOGY

- [append-only lineage field \(/glossary#append-only-lineage-field\)](/glossary#append-only-lineage-field).
- [authorization gate \(/glossary#authorization-gate\)](/glossary#authorization-gate)
- [conduct evaluation artifact \(/glossary#conduct-evaluation-artifact\)](/glossary#conduct-evaluation-artifact)
- [principal \(/glossary#principal\)](/glossary#principal)
- [semantic agent \(/glossary#semantic-agent\)](/glossary#semantic-agent)
- [signed policy object \(/glossary#signed-policy-object\)](/glossary#signed-policy-object).

Record-Grounded Conduct Admission overview → (/conduct-admission).