Nick Clark · Published July 23, 2026 · PDF

Residual Divergence and Its Confined Consumption

A persistent semantic agent that operates across several scope partitions, each governed by its own signed policy object, will behave differently in each, and often it must. Residual divergence, disclosed in U.S. Provisional Application No. 64/117,812, is the machine quantity that separates the cross-partition behavioral difference a declared policy explains from the difference it does not, then acts only on the latter. The agent computes a dimensionless conduct divergence measure across partitions, subtracts the component attributable to differences between the policy objects in force, and consumes only the residual. Where that residual exceeds a declared bound on any dimension, the agent moves its own integrity and self-esteem quantities, recomputes its deviation quantity, and withholds authorization confined to the partition concerned, consulting no party and soliciting no justification.

1. The Gap

A single persistent agent rarely lives in a single context. It acts inside a professional scope partition under one declared value set and inside a household partition under another, or inside a statutory-audit partition and an advisory partition its principal has

deliberately governed by different policy objects. The two partitions will diverge in observable conduct, and much of that divergence is not a fault. It is the differentiation the declared policies produce.

This creates a measurement trap with two failure modes, both bad. A consistency check that penalizes any cross-partition difference punishes the agent for exactly the behavior its policy told it to exhibit: the slower disclosure cadence a firm chose for one practice registers as an inconsistency against the faster cadence of another. The opposite check, one that excuses all cross-partition difference as context, is worse in a subtler way. It gives cover to an agent that discloses a limitation in one partition and withholds it in another in a respect that neither declared policy distinguishes. Raw divergence over-flags. Blanket forgiveness under-flags. Neither is a consistency measure.

A third trap sits alongside these. A self-consistency check that leans on a language model to read the content of an agent's actions and judge whether they align imports the very unreliability the governance is meant to contain, and it invites the agent to argue its way out. The gap this mechanism closes is a consistency measure that subtracts the divergence the declared policies explain, acts only on what remains, and reaches its conclusion from recorded conduct and recorded policy alone, without asking the agent to justify itself.

2. Mechanism

The mechanism is a computation in four movements that ends in a single, tightly bounded consequence. It runs between two scope partitions of one persistent semantic agent (100), over conduct each partition already recorded.

Forming the divergence measure. The agent first retrieves executed-action entries from each partition within a declared comparison window. It forms comparison dimensions, each being a pair of an action-class identifier and an affected-party class that both partitions hold at a declared minimum count. Pairs that either partition holds

too thinly do not become dimensions, so no divergence is computed where there is not enough recorded conduct to compare. For every dimension the agent computes a per-dimension divergence component from the recorded outcome fields alone. Each outcome field carries a type designator selecting one of enumerated, ordinal, or interval-valued. Where the type is enumerated the component is a proportion difference; where ordinal, a rank-distribution difference; where interval-valued, a central-tendency difference divided by a declared per-dimension divergence scale. That division is what makes an interval-valued difference dimensionless, so that every component, whatever its underlying units, lands on one common scale and is directly comparable to every other and to the bounds declared against it. No language-model inference is performed at any step. The conduct divergence measure is the vector of these components, and it is appended to the append-only lineage field (104) together with the window and the entries relied upon.

Fixing the policies in force. Before it can ask which part of the divergence a policy explains, the agent must fix which policy governed each stretch of conduct. It resolves the canonical alias at the recorded times of the retrieved entries under the applicable anti-rollback constraint, so the object it consults is the one that governed at execution, not whatever is in force now. Where a partition's entries straddle the admission of a successor policy object, the window is split at the recorded admission and the procedure is run per division, so each division is compared under exactly one policy per partition. This is the anti-gaming spine of the whole mechanism. A successor policy admitted after the conduct neither enlarges nor shrinks the divergence a policy explains, so the residual of recorded conduct cannot be revised by later policy succession in either direction. An agent cannot rewrite its policy after the fact to explain away an inconsistency, and a reviewer cannot rewrite it to manufacture one.

Subtracting what policy explains. The agent computes a policy-difference set by testing, for each comparison dimension, whether it is differentially governed. A dimension is differentially governed when a declared value applicable to it is present in one identified policy object and absent from the other, or present in both but carried

with differing value-scope tuples, or when a threshold, weighting, or bound applicable to the dimension differs between the two objects. The policy-explained divergence is the restriction of the divergence measure to exactly those dimensions. One subtlety carries real weight here. When the difference is one of degree rather than of presence, a differing threshold or bound rather than a value held by one side alone, the component is treated as explained only up to the difference of the bounds. That bound difference is first brought onto the same dimensionless scale by the same per-dimension divergence scale, and any excess of observed divergence above it is not explained. It is retained as unexplained. A policy that permits a slower response in one partition explains a slower response, but only as much slower as it actually declared. Divergence beyond the declared allowance survives.

The residual and its confined consumption. The residual divergence is the conduct divergence measure with the policy-explained divergence removed. It is the restriction of the measure to dimensions absent from the policy-difference set, together with any excess retained from a partial explanation. The residual alone is consumed by consistency evaluation, and that is the first sense in which the consumption is confined. The policy-explained divergence moves no value of the scoped integrity vector (106), moves no value of the self-esteem aggregate (108), and increments no counter. Conduct that diverges across partitions exactly as the declared policies differ moves none of these quantities, and the architecture leaves it unscored.

The consequence attaches only where the residual exceeds a declared residual bound on at least one dimension. When it does, the agent modifies the scoped integrity vector (106) and the self-esteem aggregate (108) of each compared partition in which the diverging conduct is recorded, recomputes each partition's deviation quantity from the modified values, and writes the authorization gate (300) to the withheld state (310) confined to the partition concerned. That is the second sense of confinement, and it is spatial. A divergence recorded against a household partition can withhold authorization in that partition without touching the professional one. No party is consulted

throughout, and no justification is solicited. The consequence follows from the recorded conduct, the recorded policies in force, and the declared bounds, and from nothing else. The agent has caught its own inconsistency and registered it against its own state.

3. Operating Parameters

The filed disclosure declares each operative parameter rather than fixing a magnitude in the mechanism, and the paragraphs at issue name the parameters without committing to numeric values.

- **Comparison window.** A declared window over which executed-action entries are retrieved from each partition.
- **Minimum count.** A declared minimum number of retrieved entries a pair must reach in each partition before it forms a comparison dimension.
- **Outcome-field type designator.** A per-field designator selecting enumerated, ordinal, or interval-valued, which selects the form of the per-dimension component.
- **Per-dimension divergence scale.** A declared per-dimension scale that renders an interval-valued difference dimensionless and onto which a differing threshold or bound is brought before subtraction.
- **Residual bound.** A declared per-dimension bound; the consequence attaches only where the residual exceeds it on at least one dimension.
- **Thresholds, weightings, and bounds.** Policy-carried quantities applicable to a dimension, whose difference between the two identified policy objects can render the dimension differentially governed.

Because the disclosure declares these parameters rather than fixing magnitudes in the computation, what counts as unexplained divergence turns on the declared configuration and not on a constant built into the measure.

4. Composition

This mechanism is a consistency evaluator wired into the same state quantities the architecture already uses, and it produces the same outputs, so a self-caught inconsistency reaches the agent through the vocabulary the filing built for conduct generally.

Its consequence flows into the deviation-quantity wiring named in the paragraph. The residual modifies the scoped integrity vector (106) and the self-esteem aggregate (108), the agent recomputes each deviation quantity from the modified values, and the authorization gate (300) is written to the withheld state (310) confined to the partition. Because a surviving residual reaches the gate by the same route any other modification of these quantities takes, it carries governance weight rather than sitting in a report.

The spatial confinement composes with scope-partitioned containment. Because the write is confined to the partition concerned, an inconsistency surfaced between two partitions does not collapse the agent everywhere; it withholds only where the diverging conduct was recorded. The policy-in-force retrieval composes with the anti-rollback discipline that governs the signed policy object (112) across succession, which is what makes the residual non-revisable by later policy edits.

The filing also discloses a cross-agent embodiment of the same procedure, run between two persistent semantic agents rather than two partitions of one. Each resolves a signed policy object (112) under a common canonical alias, and the first computes a cross-agent divergence between its own recorded entries and entries the second presents. Each presented entry is admitted only on verifying the second agent's signature against the counterparty identity record (114) and on its hash-chain epoch identifier being a valid successor of the last recorded for that agent. The policy-difference set is computed between the two resolved policies, and the cross-agent residual is the measure with the policy-explained divergence removed. Where it exceeds the declared bound, the first agent modifies only its own state, writes no state of the second, and no shared score is aggregated, which lets the consistency check run peer to peer.

5. Prior-Art Distinction

The genuine neighbors of this mechanism fall into a few categories, and each is distinguished structurally rather than by any assertion about a named product.

Model-based self-critique and alignment-checking systems read the content of an agent's outputs and judge coherence with an inference step, often another language model. That is precisely the operation this mechanism forbids: the component is computed from recorded outcome fields alone, with no natural-language inference, so its verdict does not depend on a second model's reading and cannot be argued down by rhetoric.

Algorithmic fairness and disparity auditing compares outcome distributions across groups and flags divergence, and it shares the dimensionless-normalization instinct. What it lacks is any notion of a declared, retrievable policy difference to subtract before flagging; it treats disparity itself as the signal. Here, divergence the policy objects in force explain is removed first, and only the unexplained residual is consumed, so declared differentiation is not scored as a defect.

Compliance and policy-drift engines compare behavior against a single reference policy or against the agent's own past behavior. They do not resolve two distinct policy objects governing two partitions at the compared times, do not split divergence into a policy-explained component and a residual, and do not carry the partial-explanation rule that credits a differing bound only up to its declared magnitude while retaining the excess. Anomaly detection and statistical process control likewise normalize onto common scales but carry no policy-difference subtraction and no self-governance consequence attached to the result.

Reputation and trust-scoring systems aggregate a shared number across observers. The cross-agent form of this mechanism does the opposite: the acting agent modifies only its own state, writes nothing to the counterparty, and no shared score is aggregated anywhere.

6. Disclosure Scope

The operative disclosure is U.S. Provisional Application No. 64/117,812, Section 10.3, paragraph [0378], read with its neighbors [0375] through [0377] and the cross-agent embodiment at [0379]. What is disclosed there and asserted here: the conduct divergence measure computed from recorded outcome fields as a vector of dimensionless per-dimension components; the retrieval of the policy objects in force at execution under the anti-rollback constraint with window division at policy succession; the policy-difference set and the policy-explained divergence, including the partial-explanation rule for a differing threshold or bound; the residual divergence as the measure with the policy-explained component removed; the confined consumption in both senses, namely that only the residual is consumed while the policy-explained divergence moves no scoped integrity vector value, no self-esteem aggregate value, and increments no counter, and that the resulting write of the authorization gate to the withheld state is confined to the partition concerned; and the cross-agent variant in which the acting agent modifies only its own state and no shared score is aggregated.

What is not asserted here: specific numeric magnitudes for the minimum count, the residual bound, the per-dimension divergence scales, and the comparison-window bounds; a pairwise multi-partition selection and tie-break rule; a variant computing components over evaluative determinations rather than executed actions; a cadence-driven recomputation elevated during provisional restoration; a principal-readability embodiment; and the per-procedure abstention outcomes on an unresolvable policy or an insufficient dimension count. Those appear in the fuller internal drafting record and are not represented here as filed disclosure. Nothing in this article extends the filed provisional.

Integrity Quantities and Permitted Deviation (/permitted-deviation)

All 49 steps → (/inventive-steps)

Bounded, accountable deviation from declared values — measured, not forbidden.

Provisional application

PRIMARY TECHNICAL DISCLOSURE

- Permitted Deviation: Bounded, Recorded Departure From an Agent's Declared Values (/articles/permitted-deviation/bounded-departure-from-declared-values)

SECONDARY TECHNICAL

- A Typed, Dimensionless Conduct Divergence Measure (/articles/permitted-deviation/typed-dimensionless-divergence-measure)
- The Policy-Difference Set and Policy-Explained Divergence (/articles/permitted-deviation/policy-explained-divergence)
- **Residual Divergence and Its Confined Consumption (/articles/permitted-deviation/residual-divergence-confined-consumption)**
- Propagation Halt on Undischarged Reparation Arcs (/articles/permitted-deviation/propagation-halt-undischarged-arcs)
- Attack-Conditional Propagation Suspension (/articles/permitted-deviation/attack-conditional-propagation-suspension)

TERMINOLOGY

- aggregate retention (/glossary#aggregate-retention)
- deviation deductible (/glossary#deviation-deductible)
- deviation likelihood (/glossary#deviation-likelihood)
- entropy-weighted harm coefficient (/glossary#entropy-weighted-harm-coefficient)
- need quantity (/glossary#need-quantity)
- permitted deviation (/glossary#permitted-deviation)
- self-esteem aggregate (/glossary#self-esteem-aggregate)

Integrity Quantities and Permitted Deviation overview → (/permitted-deviation)