

The Policy-Difference Set and Policy-Explained Divergence

An agent that acts across separate scope partitions is often permissibly governed by a different declared policy in each, so a raw comparison of its conduct across partitions punishes the differentiation the policies deliberately declared, while excusing every divergence as context hides an agent that behaves inconsistently in a respect no policy distinguishes. The policy-difference set and policy-explained divergence, disclosed in U.S. Provisional Application No. 64/117,812, resolve this by determining, per comparison dimension, whether the two governing policies differ in a way bearing on that dimension, then restricting a measured conduct divergence to the dimensions those declared differences explain. Only the unexplained remainder is consumed by consistency evaluation. Where policies differ merely by a threshold or bound, the mechanism explains divergence only up to that bound difference and retains any excess as unexplained.

1. The Gap

A persistent semantic agent (100) maintains a plurality of scope partitions at once. Each partition carries its own instances of the scoped integrity vector (106) and the self-esteem aggregate (108), so every state movement stays in the partition where the conduct occurred and cannot withdraw authorization in an unrelated context. The

conduct in each partition is compared under the signed policy object (112) in force for that conduct, and two partitions can be governed by policies that differ. A declared value present for one partition and absent for another is a difference of governance the principal chose, not a lapse the agent committed.

It creates a measurement problem for anything that wants to check whether the agent is behaving consistently. Compare the agent's recorded conduct across two partitions and you will find divergence. Some of that divergence is exactly what the two policies told the agent to do differently. The rest is the agent behaving one way here and another way there in a respect neither policy distinguishes.

Two blunt responses both miss. Treat every cross-partition divergence as inconsistency and the agent is penalized for the differentiation its principal declared, marking a permitted difference as a fault. Wave divergence away as mere context and the second kind goes unmeasured, letting an agent behave one way in one partition and another way in the other wherever no declared policy separates them.

The filing needs an object that states, mechanically and without asking the agent to justify anything, how much of an observed divergence the declared policies actually account for, so that only the remainder is charged. That object is the policy-difference set, and the divergence it accounts for is the policy-explained divergence.

2. Mechanism

The mechanism takes two inputs and produces two derived quantities.

The first input is a conduct divergence measure. This is a vector of per-dimension divergence components. Each comparison dimension is a pair of an action-class identifier and an affected-party class that both compared partitions hold at a declared minimum count of recorded executed-action entries. A pair either partition holds below the minimum count does not form a dimension. Each component is computed from

recorded outcome fields alone. Every 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, it is a rank-distribution difference. Where interval-valued, it is a central-tendency difference divided by a declared per-dimension divergence scale. The division renders that component dimensionless, so every component in the vector is mutually comparable. No language-model inference is performed at any step. The measure is the vector, appended to the append-only lineage field (104) with the comparison window and the entries relied upon.

The second input, for each compared partition, is the signed policy object (112) in force for the conduct being compared. It is identified by resolving the canonical alias at the recorded times of the retrieved entries, under the applicable anti-rollback constraint. Where a partition's entries span the admission of a successor policy, the window is divided at the recorded admission and the procedure runs per division, so each division of conduct is compared under exactly one policy per partition. The objects compared are those governing at execution, not those in force at comparison, so a later policy succession neither enlarges nor reduces the divergence a policy explains.

Given those two inputs, the agent computes the policy-difference set. For each comparison dimension of the measure, it determines whether the dimension is **differentially governed**. A dimension is differentially governed on any one of three grounds. The first: a declared value applicable to the dimension is present in one identified policy object and absent from the other. The second: the corresponding value is present in both, but with differing value-scope tuples. A value-scope tuple comprises a set of action-class identifiers, a set of scope-partition identifiers, and an empathy-scope designation selected from personal, interpersonal, and global, so two tuples differing on the action class, the scope enumeration, or the empathy scope implicated by the dimension satisfy this ground. The third: a threshold, weighting, or bound applicable to the dimension differs between the two policies. The policy-difference set is the set of dimensions found differentially governed on any ground.

The policy-explained divergence is the restriction of the conduct divergence measure to the dimensions in the policy-difference set. On each such dimension the component is attributed to the declared policy difference and is not charged.

The partial-explanation rule handles the third ground with more care than the first two. A presence difference means the two policies disagree about whether the dimension is governed at all, so the whole observed component is explained. A threshold or bound difference is a matter of degree: the policies govern the same dimension, but they set a different level. In that case the component is assigned as explained only up to the difference of the bounds. The bound difference is brought onto the component's dimensionless scale by dividing it by the same per-dimension divergence scale, so the subtraction is performed between two quantities on one scale. Any observed divergence above the bound difference is not explained. It is retained as unexplained excess.

This bound handling is the crux. Where two policies set different bounds on an interval-valued dimension, the mechanism explains the divergence there only up to how far the bounds actually differ. Diverge by more than the declared bounds allow and the surplus survives the subtraction; diverge within them and nothing survives. The agent is credited for exactly the differentiation its policies declare and no more, computed from recorded fields and declared policy contents rather than from any account the agent gives of itself.

3. Operating Parameters

Every parameter the mechanism consults is a declared quantity. The filing declares each and fixes no magnitude, so the numbers live at deployment, not in the disclosure.

- **Comparison window.** The declared interval over which executed-action entries are retrieved for each partition.

- **Minimum count.** The declared count of retrieved entries each partition must hold on an action-class and affected-party-class pair for that pair to form a comparison dimension. A pair below it on either side forms no dimension.
- **Outcome-field type designator.** Carried by each outcome field, selecting enumerated, ordinal, or interval-valued, and thereby selecting the proportion-difference, rank-distribution-difference, or central-tendency-difference computation for that dimension.
- **Per-dimension divergence scale.** A declared per-dimension divisor that renders the interval-valued component dimensionless. It is the same divisor that brings a threshold or bound difference onto the component scale, keeping every component, threshold, and bound directly comparable.
- **Thresholds, weightings, and bounds.** Quantities applicable to a dimension, a difference in any of which makes the dimension differentially governed and, where the difference is one of degree, sets the amount the mechanism will explain.
- **Residual bound.** The declared bound against which the unexplained remainder is tested downstream. The remainder is acted on when it exceeds this bound on at least one comparison dimension.

The filing states no numeric value for any of these. It does not fix the minimum count, the residual bound magnitude, or any divergence-scale value.

4. Composition

The policy-difference set is the hinge of a four-step procedure disclosed together in Section 10.3. The typed, dimensionless conduct divergence measure produces the vector. The non-revisable retrieval of the policy in force fixes, for each partition, exactly which policy object governed the compared conduct. The policy-difference set and policy-explained divergence, described here, decide which dimensions of that vector the declared policies account for. The residual divergence is then the measure with the

policy-explained divergence removed, being its restriction to dimensions absent from the policy-difference set together with any retained excess from the partial-explanation rule.

That downstream residual is the only quantity consistency evaluation consumes. 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. When the residual exceeds the declared residual bound on at least one dimension, the agent modifies the scoped integrity vector (106) and the self-esteem aggregate (108) of each compared partition where the diverging conduct is recorded, recomputes each deviation quantity, and writes the authorization gate (300) to the withheld state (310) confined to the partition concerned. No party is consulted and no justification is solicited.

The mechanism also composes upward and sideways. It reads the value-scope tuple defined by value-scope resolution at Section 10.1, so the same tuple structure that decides whether a value is implicated in admission decides here whether a value is differentially governed across partitions. It inherits the anti-rollback discipline of policy-in-force retrieval, so the residual of recorded conduct is not revisable by policy succession in either direction. And the same procedure runs across two persistent semantic agents rather than two partitions of one: each resolves a policy under a common canonical alias, presented entries are admitted only on verifying the counterparty signature and a valid-successor epoch, the policy-difference set is computed between the two resolved policies, and the first agent modifies only its own state, aggregating no shared score.

5. Prior-Art Distinction

Several established categories touch adjacent problems without doing what the policy-difference set does. This section distinguishes them structurally and asserts nothing about whether any product or party practices them.

Behavioral anomaly and drift detection flags statistical divergence in an agent's or a user's conduct over time or across contexts. It has no notion of a declared governing policy that legitimately explains part of the divergence, so it would treat declared differentiation as an anomaly to be surfaced. The mechanism here does the opposite first: it subtracts the declared-policy-explained component before anything is charged.

Access-control policy engines, including role-based, attribute-based, and context-aware evaluators, decide a single request against a single policy. They do not compare recorded conduct across two partitions, and they have no object that attributes an observed divergence to the difference between two policies.

Group-fairness and disparate-impact auditing measures outcome differences between populations and reports them as parity gaps. It measures divergence, but it carries no policy-difference set that removes the portion a declared governing policy explains, and it does not bring a bound difference onto the divergence scale to price only the excess.

Configuration and policy diffing tools compute the difference between two policy documents. They stop at the diff. They do not restrict a measured conduct-divergence vector to the differentially governed dimensions, and they do not compute a residual that consumes only the divergence the diff fails to account for.

Language-model consistency judging reasons over natural-language conduct to opine whether behavior is coherent. The mechanism here performs no language-model inference. It operates over typed recorded outcome fields, and its output is a numeric residual, not an opinion.

The distinguishing structure is the join: a typed dimensionless conduct measure, a non-revisable retrieval of the policy that actually governed the conduct, and a per-dimension differential-governance test that subtracts the explained component both by presence and, for differences of degree, only up to the bound difference on a common scale, leaving a residual that is the only thing consumed.

6. Disclosure Scope

The mechanism described here is disclosed in U.S. Provisional Application No. 64/117,812, at Section 10.3, paragraph [0377], titled "Policy-difference set and policy-explained divergence," and is read together with its neighbors in the same section: the typed, dimensionless conduct divergence measure at [0375], the non-revisable retrieval of the policy in force at [0376], and the residual divergence and its confined consumption at [0378]. What the provisional discloses is the computation of a policy-difference set by a per-dimension test of differential governance on three grounds, the restriction of the divergence measure to those dimensions as the policy-explained divergence, and the rule that a threshold or bound difference is explained only up to the difference of the bounds brought onto the component's dimensionless scale, with any excess retained as unexplained.

What this article does not disclose, because the provisional does not fix it, is any numeric magnitude: no minimum-count value, no residual-bound value, and no per-dimension divergence-scale value, each being declared without a fixed number. It also does not state any worked numeric example, any detailed applicability test for when a threshold, weighting, or bound counts as applicable to a dimension, or any per-member lineage record of the policy-difference set beyond what paragraph [0377] recites, none of these being part of the filed disclosure of that paragraph. Nothing in this publication should be read to assert that any named system, product, or party practices or infringes the disclosed mechanism.

Integrity Quantities and Permitted Deviation (</permitted-deviation>)

[All 49 steps → \(/inventive-steps\)](/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 \(/articles/permitted-deviation/bounded-departure-from-declared-values\)](/articles/permitted-deviation/bounded-departure-from-declared-values)

SECONDARY TECHNICAL

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

TERMINOLOGY

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

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