

Card Network Chargebacks: Counterparty-Directed Remedy in Practice

A purchasing agent departs from its declared policy, records the departure, and then closes its own file. That pattern is common in automated governance, and it is why remedy so often lands in a pool rather than with the party that absorbed the loss. Card network chargeback systems addressed a narrower version of this problem long ago by routing a remedy back to the disputing cardholder across organizational lines. U.S. Provisional Application No. 64/117,812 describes a different arrangement for conduct between autonomous agents: an other-directed reparation arc that the deviating agent cannot discharge by repairing itself, that completes only on a matched pair whose second half the harmed counterparty produces, and whose accumulation, past a declared aggregate retention, forecloses the agent's permission to deviate at all.

When an Agent Harms a Counterparty and Then Grades Its Own Repair

A purchasing agent overrides a declared handling constraint because a shipment is about to miss a cutoff. The override is recorded. A supplier's agent is the party left worse off. In most systems deployed today, the party that caused the harm also writes the entry closing the file.

That arrangement fails in three recurring ways.

Pooled remedy: compensation goes to a fund, a reserve, or a regulator, and the counterparty holding the loss receives nothing traceable to it.

Self-certified repair: an agent free to mark its own remediation complete can produce remediations indefinitely, and a repair produced and evaluated by the deviating party alone demonstrates conscientious behavior, not restoration.

Time as a solvent: where a quiet interval restores capability, the absence of new incidents stands in for the settling of old ones, and a party that waits out its penalty period recovers full authority having settled with no one.

These compound in multi-agent deployments, where there is often no shared operator, no platform adjudicator, and no registry holding a verdict. The open question is not whether harm gets recorded but whether the record can ever be closed, by whom, and at what cost to the party that caused it. Payments infrastructure has long carried an answer to a narrower version of that question.

The Dispute Path a Cardholder Already Has

Card network chargeback systems are a long-established mechanism for directing a remedy at a specific harmed party across organizational boundaries. As publicly described, the flow begins when a cardholder tells the card issuer that a transaction was unauthorized, that goods or services never arrived, or that what arrived did not match what was represented. Where the issuer accepts the claim, it may credit the cardholder and pass the dispute into the network, which routes it to the acquiring bank serving the merchant.

The merchant is then given an opportunity to respond with supporting evidence, a step commonly called representment. Also as publicly described, the networks define the categories of dispute, often called reason codes, the evidence responsive to each, and

the time limits within which each side must act. Where the two sides do not converge, further stages follow, publicly described as including a pre-arbitration exchange and a network-administered arbitration allocating liability between the financial institutions.

Four properties of that design deserve credit, and are easy to overlook once a contrast is drawn.

- **The remedy reaches the party that suffered the loss.** The cardholder is made whole in the cardholder's own account, not through a general fund.
- **It works between parties with no bilateral relationship.** Cardholder and merchant need no contract, no shared platform, no prior enrollment.
- **It is rule-governed and reconstructible.** Each stage produces a record, and the applicable rules are published in advance.
- **It carries cost beyond the disputed amount.** As publicly described, merchants incur fees when disputes are raised, and the networks operate monitoring programs for merchants whose dispute rates run high.

Participants raise limits publicly too: merchants describe the evidentiary burden as heavy and outcomes as hard to predict, and first-party misuse, widely called friendly fraud, is an openly acknowledged problem. The structure presupposes an identified cardholder, issuer, and acquirer. It is a settlement mechanism for payment obligations among financial institutions, and it is very good at that.

Reparation Arcs That Only the Harmed Party Can Close

The architecture disclosed in U.S. Provisional Application No. 64/117,812 begins from a different premise: the deviating agent holds the obligation and cannot discharge it alone.

When a semantic agent admits a departure from the policy in force, it appends a permitted deviation record. Before any reparation structure is created, it performs an affected-party resolution: the affected-party class carried in the conduct descriptor of the admitted action is extracted and tested against the counterparty identity records held in the scope partition where the deviating action is recorded. No separate structure of class definitions is required. A record resolves as a member where the extracted class maps to the interpersonal scope and the identity primitive it carries is recorded, in an entry of the agent's own append-only lineage field, as a party to or a recipient of the deviating action. The outcome is one of exactly two: an identified-counterparty outcome naming each resolving primitive, or a no-identified-counterparty outcome.

Harm up to a declared deviation deductible is borne by the agent and creates no arc; the amount exceeding it creates one. On the identified-counterparty outcome, that arc is designated other-directed and bound to the counterparty's identity primitive, both written at creation and neither removed nor modified afterward, so a later change in the records the agent holds does not redesignate an arc already created. Where several counterparties resolve, the deductible is drawn once against the aggregate harm and the remainder is apportioned in shares summing to it exactly.

Then the governing constraint. An other-directed reparation arc is not dischargeable by a restorative mutation the agent performs on itself. Where the agent performs one anchored to the deviation entry the arc references, the record is appended and retained as a non-discharging restorative mutation record: the arc remains pending, the accumulated amount is not reduced, the self-esteem aggregate receives no positive increment, and the arc keeps contributing unrectified dissonance to the integrity compliance score. It is retained rather than rejected, so the lineage discloses that restoration was attempted and discharged nothing.

Discharge runs instead through a matched pair. The agent emits a reparation acknowledgment artifact naming the permitted deviation record that occasioned the arc, the policy constraint overridden, the amount apportioned to that counterparty, and

the action classes to which it is directed; emission decrements the agent's authorization budget by an amount declared in the signed policy object, without regard to whether it was warranted. The second half is a reparation acceptance determination produced by the bound counterparty within a discharge window declared in that policy object and counted in successor epochs of the agent's own hash chain, so no other party advances it and no clock at a hosting execution node manipulates it. That determination is not produced by the agent, by its principal, or by any adjudicator, arbiter, or other third party appointed to determine the sufficiency of the reparation. Where the pair completes within the window, each party appends a settlement-lineage entry, the arc is discharged, the accumulated amount leaves the retention register, and the agent receives the positive increment it was denied for repairing itself.

Where the class resolves to no counterparty the agent holds a record for, the arc is designated unaddressed and undischageable, accumulating at a declared multiple not less than unity. The specification is careful here: the agent's own records disclose no counterparty of the class, which is not an assertion that none exists. The affected party is designated the structurally silent party.

Foreclosure is conditioned on accumulation, not on any single arc. Where the amount accumulated in the retention register exceeds a declared aggregate retention, the permission condition is foreclosed: a retention foreclosure record is appended, and a deviation likelihood exceeding unity thereafter produces the withholding outcome rather than the admission outcome. Discharge of pending arcs returning the register below the aggregate retention lifts that condition. A successor policy object declaring a larger aggregate retention does not, and is not admitted while the foreclosure record stands unreversed. Where the accumulation consists wholly of unaddressed arcs, the loss is permanent.

Two Different Answers to the Question of Who Decides

The convergence is real: both designs decline to let the party that caused the harm close its own case, both direct the remedy at the party holding the loss, and both attach cost to the process and not only to its outcome. Four differences follow.

Who renders the decision. In card network chargeback systems, as publicly described, the issuer makes the initial call and the network resolves what the parties cannot, up to an arbitration it administers. Authority sits with institutions that are neither the harmed party nor the accused. Under the filed architecture, the acceptance is produced by the bound counterparty and by no one else.

Where the process starts. A chargeback begins with the harmed party raising a claim. A reparation arc begins with the deviating agent, off its own admitted record, and emitting the acknowledgment spends authorization budget without regard to whether it was warranted. Nothing requires the counterparty to notice, complain, or assert.

What silence means. In chargeback flows, as publicly described, a merchant that does not respond within the applicable time limit generally leaves the dispute standing against it. Silence resolves. Under the filed architecture, non-response resolves nothing against either party: the agent appends a reparation non-response outcome, the arc stays pending, no counter of the counterparty is incremented, and no adverse record is appended to the counterparty identity record held for it. The burden stays where the deviation occurred.

What the consequence bears on. A chargeback consequence is monetary, plus the program-level exposure already noted. An arc's amount is denominated in the units of the entropy-weighted harm coefficient, and its consequence bears on the agent's own future permission to depart from policy. Past the declared aggregate retention the agent

does not pay a penalty; it ceases to be permitted to deviate for as long as the accumulation stands above that retention, and the ceiling cannot be raised by a successor policy object while the foreclosure stands.

One case the two designs do not share: chargebacks presuppose a cardholder, an issuer, and an acquirer, whereas an affected-party class resolving to no counterparty in the agent's records still creates an arc, undischageable and accumulating at the declared multiple.

A further treatment has no payments analogue. Where the bound counterparty has itself emitted a non-execution attestation enumerating the action class within which an acceptance would be produced, the agent may not count that withholding, score it, compare it against a threshold, or append anything adverse to the counterparty identity record. The arc's contribution to the accumulation is instead held at the amount then standing, and resumes when the attestation is released. Nothing in that suspension discharges, redesignates, or reduces the arc.

Running Both Layers in One Deployment

Consider a business buying through an autonomous purchasing agent that deals with a supplier's agent and settles on a card. Two layers run at once without overlapping.

The money layer belongs to the card network. If goods never arrive, the buyer disputes the transaction with its issuer and the process runs under the applicable network's published rules. Nothing in the filed architecture displaces authorization, clearing, or those rules.

The conduct layer sits apart from it. Suppose the purchasing agent, on a deviation likelihood exceeding unity computed from its own quantities, departed from a declared constraint governing how it handled the supplier's terms and admitted that departure as a permitted deviation. Its affected-party resolution names the supplier agent from

records it already holds, and an other-directed arc is created for the amount above the deductible. The agent emits its acknowledgment, spending budget to do so. Should the supplier agent produce an acceptance determination within the discharge window, the arc discharges. Otherwise the arc stays pending, the supplier agent suffers nothing for its silence, and the purchasing agent carries the accumulation toward its declared aggregate retention.

The chargeback settles the invoice. The arc governs whether unsettled harm eventually forecloses the agent's permission to make that kind of call again.

Five limits, stated plainly.

- **No money moves.** A reparation acceptance determination is a governed observation, not a payment, and discharge removes an amount from the agent's own retention register.
- **No merits decision is rendered.** The deductible is drawn without any determination of whether the deviation was well founded, and the acknowledgment costs budget either way.
- **Counterparties are not compelled.** One that never answers is never penalized, and its silence never becomes a magnitude.
- **Parties absent from the agent's records stay out of reach.** The no-identified-counterparty outcome describes those records, not the world.
- **Consumer protection, fraud detection, and regulatory compliance sit outside its scope.**

Teams building agent commerce plausibly want both layers. Payments settle who is out of pocket. The conduct layer addresses what payments were not built to reach: whether an agent causing harm it has not settled keeps the discretion that produced it.

Disclosure Scope

The architecture described above is disclosed in U.S. Provisional Application No. 64/117,812, which is pending. Nothing here asserts a patent right, and the scope of any right that may issue is determined by the claims as eventually allowed. Named structures such as the other-directed reparation arc, the reparation acceptance determination, and the retention foreclosure record refer to elements shown in the filed drawings.

References to card network chargeback systems are to public materials and are used for comparison only; no relationship, endorsement, or infringement is asserted.

Descriptions of card network dispute handling here are qualitative, drawn from widely available public sources rather than from any network's current operating rules, which change. Readers evaluating a real deployment should consult the applicable network's published materials.

Counterparty-Directed Reparation ([All 49 steps → \(/inventive-steps\)](#) [counterparty-reparation](#))

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 \(/articles/counterparty-reparation/completion-by-non-worsening-re-encounter\)](/articles/counterparty-reparation/completion-by-non-worsening-re-encounter).
- [Partial Reparation Acceptance and the Residual Arc \(/articles/counterparty-reparation/partial-reparation-and-residual-arc\)](/articles/counterparty-reparation/partial-reparation-and-residual-arc).
- [Dual-Lock Escrow of the Discharge Itself \(/articles/counterparty-reparation/dual-lock-escrow-of-discharge\)](/articles/counterparty-reparation/dual-lock-escrow-of-discharge).
- [Per-Integrity-Scope Aggregate Retention and OR-Firing Foreclosure \(/articles/counterparty-reparation/aggregate-retention-or-firing-foreclosure\)](/articles/counterparty-reparation/aggregate-retention-or-firing-foreclosure)

APPLICATIONS · GENERAL

- [Automated Consumer Redress: Repair Directed at the Party Harmed \(/articles/counterparty-reparation/consumer-redress-automation\)](/articles/counterparty-reparation/consumer-redress-automation)
- [Insurance Claims Networks: Discharging Harm to the Counterparty That Holds It \(/articles/counterparty-reparation/insurance-claims-agent-networks\)](/articles/counterparty-reparation/insurance-claims-agent-networks).

APPLICATIONS · SPECIFIC

- [Card Network Chargebacks: Counterparty-Directed Remedy in Practice \(/articles/counterparty-reparation/payment-chargeback-systems\)](/articles/counterparty-reparation/payment-chargeback-systems).

TERMINOLOGY

- [other-directed reparation arc \(/glossary#other-directed-reparation-arc\)](/glossary#other-directed-reparation-arc)
- [reparation arc \(/glossary#reparation-arc\)](/glossary#reparation-arc)
- [restorative mutation \(/glossary#restorative-mutation\)](/glossary#restorative-mutation).
- [unaddressed reparation arc \(/glossary#unaddressed-reparation-arc\)](/glossary#unaddressed-reparation-arc).

[Counterparty-Directed Reparation overview → \(/counterparty-reparation\)](/counterparty-reparation).