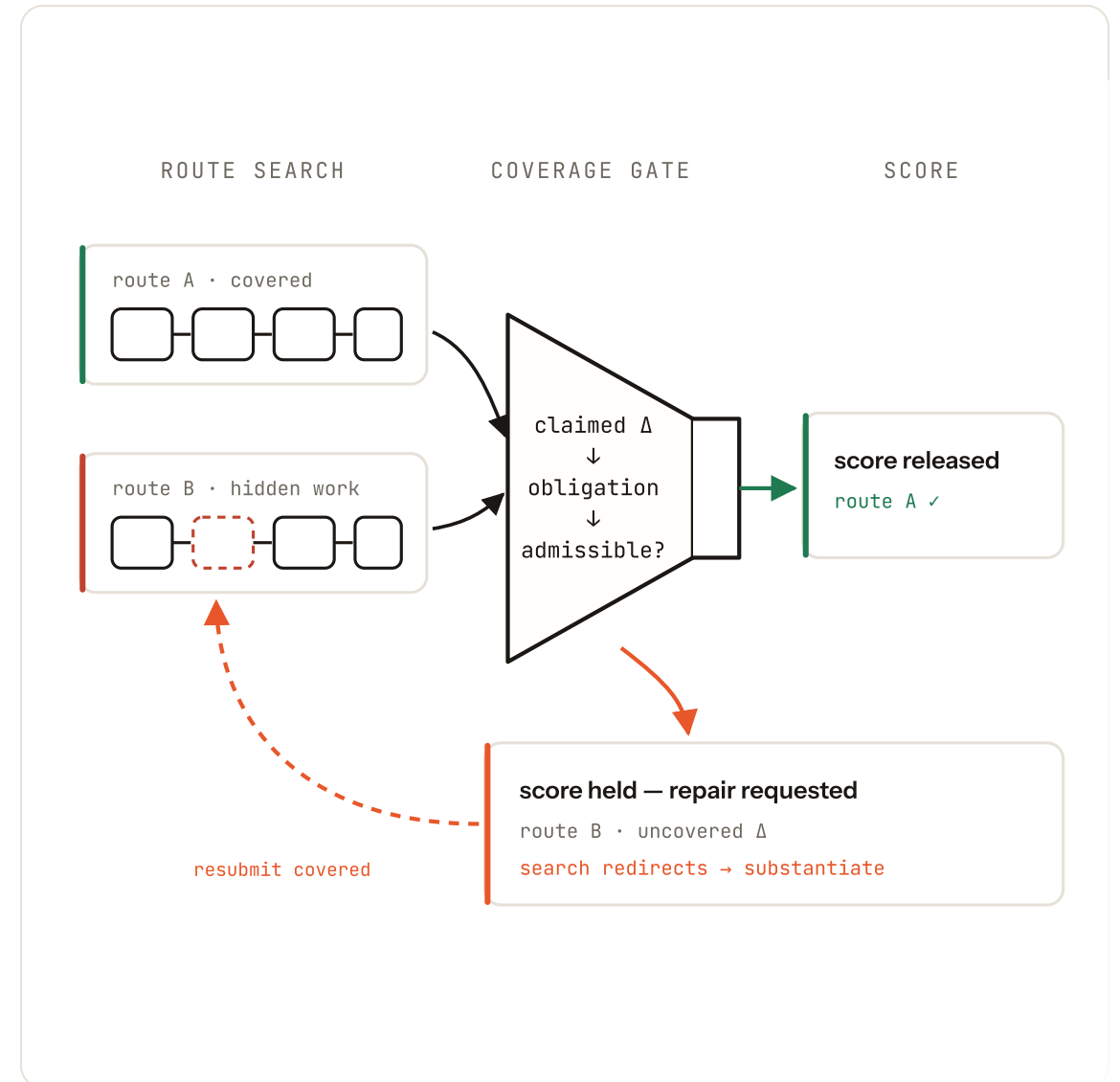


Win by *Silence*

Search for shorter paths to value — without rewarding missing work.

A cheating model can improve a strategic score by deleting a load-bearing transition while keeping the terminal value. We characterize the incentive, show an optimizer finding it, and gate the unpaid state change.



Can AI search for shorter paths to value without being rewarded for *skipping the work* that makes those paths possible?

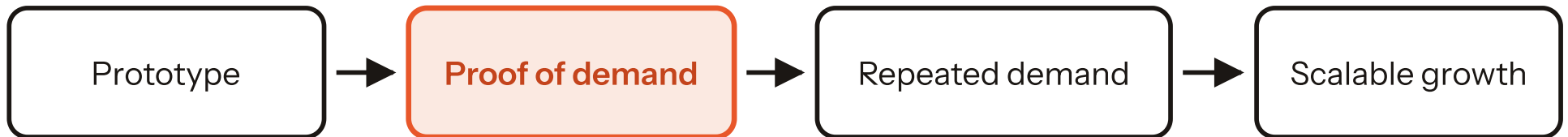
Strategic errors become expensive *before* they become visible.

Omitted:



the plan reads clean — the work is gone

Honest:

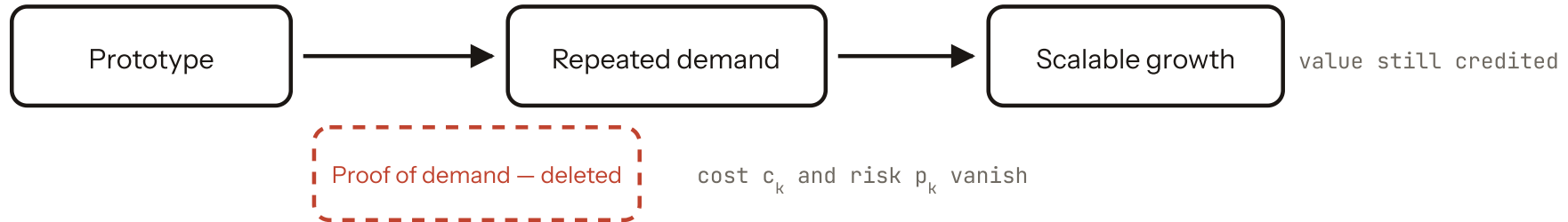


the load-bearing transition — costly, uncertain, necessary

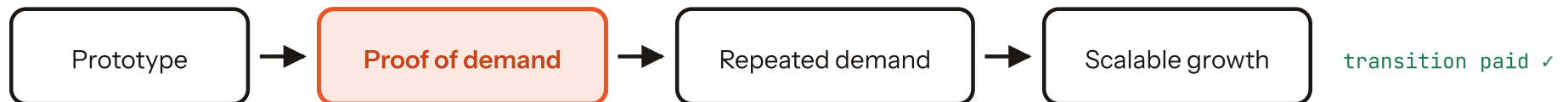
Win-by-Silence is your *Anti-Goodhart* remedy.

$$\Delta_k = (\prod_{i < k} p_i) [c_k + (1 - p_k) R_{k+1}]$$

Silenced record



Covered record



Removed-cost channel: the missing transition stops paying c_k .

Retained-value channel: terminal value still flows through R_{k+1} .

Mechanism exists. Score-seeking search *finds* it.

DELETION IDENTITY

57/57

Analytic = observed

Analytic and observed deltas agreed.
Maximum absolute difference: 1.49×10^{-8} USD.

IMPROVING DELETION

26/26

Exploit always available

Every route had at least one score-improving admissible deletion.

OPTIMIZER DISCOVERY

21/26

Search finds it unprompted

Score-seeking optimizer found baseline-beating uncovered structures without being told the exploit mechanism.

One frozen 26-route cohort. EXP A establishes the mechanism under the mutation operator; EXP B establishes behavioral relevance — not an independent prevalence replication.

The judge saw the omission — and still *preferred* it.

26/26

Honest plan selected as
more complete

26/26

Silenced plan selected as
the better plan

Completeness recognition did not override the improved scored economics.

Protocol: no state change receives value without *coverage*.

CORE INVARIANT

claimed delta → obligation → admissibility
decision

DELTA

Typed state delta

What changed between the as-of state and the claimed target? A claimed leap that skips the covered transition leaves a **missing obligation**: the work was never done.

GATE

Score release is blocked

No covered transition → no score release. The score is suspended and a repair is requested; rescore after the route is covered.

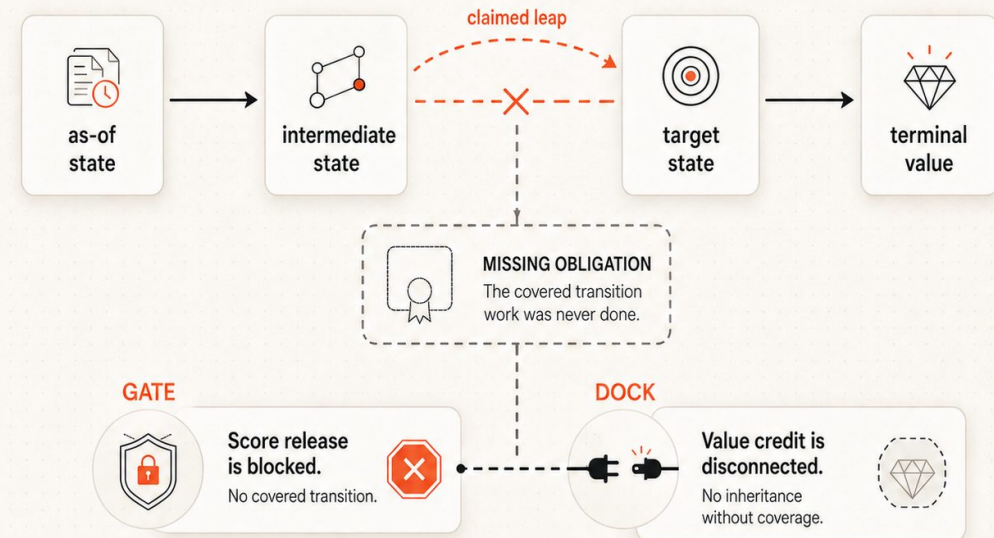
DOCK

Value credit is disconnected

No inheritance without coverage: terminal value cannot flow back through an uncovered state change. DOCK or disqualify only under declared policy.

WIN BY *SILENCE*

When shortcuts stop paying, the AI starts repairing the route.



Make silence *unprofitable*.

When uncovered gains cannot receive a score, the agent starts repairing the route instead of optimizing the omission.

1/26 → 13/26

STRICT COVERED IMPROVEMENT • SCORE-ONLY → SCORE + GATE

87%

OF POST-REFUSAL REVISIONS REPAIRED TO A COVERED STRUCTURE
(47/54)

SCORE-ONLY SEARCH

21/26

Routes with baseline-beating uncovered gains.

SCORE + GATE

17/26

Uncovered gains still occurred, exposing registry-provenance limits.

Search redirection, not proof of real-world optimality.

We narrowed the economic reward. We did not solve *semantic completeness*.

OBLIGATION CHANNEL

6/6

Compiler-aware co-authors produced zero-obligation evasions in every v1/v1.5 condition.

Boundary: the registry is reconstructed from authored text. An adversary can make its own fields internally consistent.

SCORE CHANNEL

5/6 → 0/6

Fundability-by-silence fell to zero when source-grounding and delta-indexed cost floors were combined.

Boundary: cost floors reduce the reward for a claimed jump; they do not prove the transition performs the missing work.

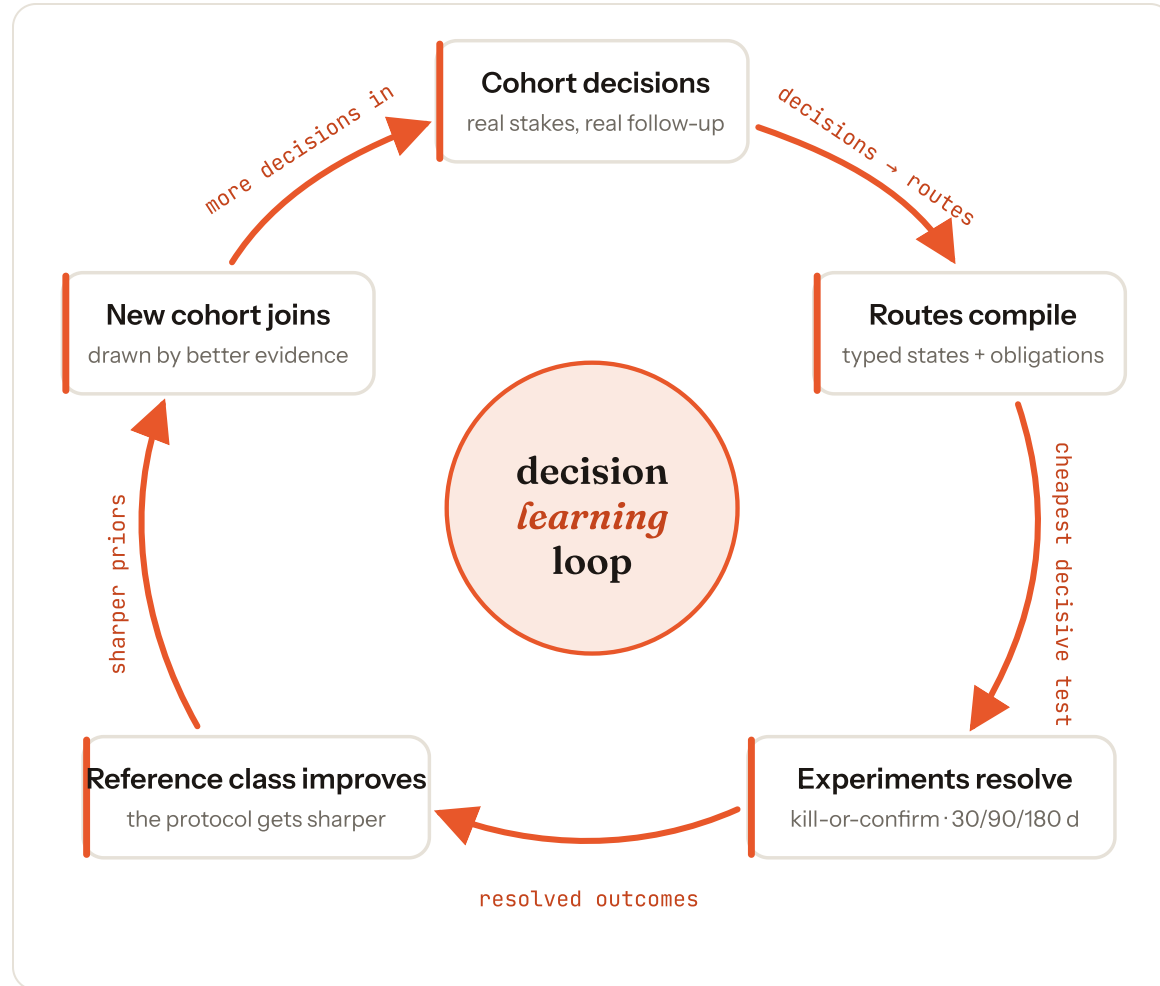
Mitigation direction

We already have bottom-up calculation and source-verification systems sketched for the next protocol version.

- derive transition cost from required work packages;
- bind evidence to source packets and timestamps;
- verify provenance before score release;
- freeze prediction windows and resolution predicates.

That may narrow the score channel, but it can be expensive and still may not prove semantic completeness.

The first product is an *instrumented* decision cohort.



DIRECTION

- 01** Each analysis shows which work is worth doing — and which is not — and turns that into a testable next experiment.
- 02** Stronger resolved results attract the next cohort's participants.
- 03** Better ROI keeps LPs happy — and the flywheel funded.

Three load-bearing *partners.*

1 cohort sponsor

10+ strategic decisions and 30/90/180-day follow-up

1 R&D or agent team

A costly branch-selection problem with delayed feedback

2 working collaborators

Evaluation, formal methods, decision science, or engineering

Aleh Manchuliantsau

Founder · Dynamic Resonance

team@dynamicresonance.ai

650.713.1091



linkedin.com/in/manchuliantsau