

tl;dr-

- In 2023, Optimism ran two megarounds. In 2024, Optimism ran one round per domain per year. We've learned that mega rounds devolve into popularity contests and annual feedback loops are too slow. In 2025, Optimism should focus on fewer domains, iterate more rapidly, and refine what works.
- In 2023, Optimism had everyone vote on everything. In 2024, Optimism ran experiments around expertise-based voting and metrics-based voting. We've learned what humans are good at - and what data is good at. In 2025, Optimism should take the best of both (and set aside the rest).
- In both 2023 and 2024, we struggled at measuring the success of Retro Funding itself. The time for not knowing how well this works is over.

Full post...

This was supposed to be simple

As Vitalik [wrote](#) when initially describing the mechanism: "The core principle behind the concept of retroactive public goods funding is simple: it's easier to agree on what was useful than what will be useful."

Now, nearly two years into this experiment, we should be able to look back at the mechanism of Retro Funding itself and apply this same analysis.

Specifically, we want to understand:

- Can people actually agree on what was useful?
- Under what circumstances does retroactive funding produce superior outcomes?

Simple, right? Not entirely.

My team at OSO has been trying to help Optimism answer some of these questions since [Round 2](#). In this post, we'll look at a very high level at what's happened across five rounds of Retro Funding (Rounds 2-5) and share some of our observations.

The good news is there are some clear learnings about what people are good at agreeing on and how to create the right conditions for consensus. These learnings are the result of deliberate experiments that were undertaken in 2024, for instance, experimenting with metrics-based voting in Round 4 and expertise-based voting in Round 5.

The not-so-good news is that we don't (yet) have the data to show that retroactive funding produces superior outcomes.

In our view, solving the measurement problem is the most important thing to get right in 2025. We need to build an engine for measuring the impact of each allocation cycle—one that gives us more than a vague sense that it's working and one that gives the collective more than a single round (per domain) per year to see what's working.

Our recommendations include:

- Continuing to move away from project-based voting
- Finding the right balance of metrics and experts' subjective assessments
- Focusing on a relatively narrow set of domains but with more rapid feedback cycles

Basically, if 2024 was all about experimenting on expertise-based voting and metrics-based voting, then 2025 should be about combining the best of both. Once we discover the optimal combinations, we can expand the scope and complexity of rounds.

What humans are good at (and what data is good at)

Optimism has now completed five rounds that use badgeholder-based voting to allocate tokens.

Each round had different design parameters and these have allowed us to learn different things about what humans and data are good at.

Here's a quick summary:

Round

Design Parameters

Key Lessons Learned

Round 2

Projects had to be nominated by a badgeholder in order to make the round. Badgeholders had to determine outright token allocations for each of the 200 projects with minimal tooling.

Project nomination process was awkward and time-consuming for badgeholders. Badgeholders needed more structure in the voting process; everyone came up with their own scattershot method (eg, sharing spreadsheets to rate their favorite projects).

Round 3

Any project could sign-up for one or more domains and self-report their metrics. Badgeholders would do a light review to filter out spam, but left everything else intact. Projects had to get at least 17 votes in order to receive rewards. Over 600 projects ended up getting approved. Badgeholders still had to determine outright token allocations for each of the 600+ projects. Badgeholders could also create “lists” recommending projects and outright allocations.

Voters felt overwhelmed. It was very difficult to differentiate between weak projects and good ones with little reputation; borderline projects led campaigns to reach the quorum line. The domain categories were not strictly enforced and thus mostly useless to badgeholders. Impact metrics were not comparable. Every list was different and there was no quality control. Onchain projects received a smaller share of the token pool than most badgeholders felt they deserved.

Round 4

Focused only on onchain builders with strict eligibility requirements (based on onchain activity). Badgeholders were provided metrics instead of projects to vote on. Over 200 projects were approved out of 400+ applicants. Projects could receive a multiplier if they were open source.

Badgeholders found voting much easier. Results had a steep power law distribution, but voters generally felt the top projects were rewarded fairly. Metrics alone couldn't capture quality, momentum, and other nuances, highlighting the need for more complex evaluation signals. Metrics did not work as well for certain “cutting-edge” sub-domains (eg, 4337 related apps). The open source multiplier was too complex to enforce consistently.

Round 5

Focused only on OP Stack contributions with strict eligibility requirements (enforced by a small review team). Returned to project-based voting, grouping voters by expertise and assigning each to a single category of 20-30 projects, rather than hundreds.

Results were very flat, and voters felt the top projects were not rewarded sufficiently. Perverse incentive to divide work across multiple smaller contributions / teams in order to receive more tokens. Grouping by expertise revealed significant differences between experts and non-experts both in project selection and allocation strategy. After seeing both allocations, voters preferred experts' selections.

In Round 6, which is currently underway, Optimism is experimenting with impact attestations and a more aggressive set of distribution options for badgeholders.

Clearly, the optimal Retro Funding design hasn't been found yet. But we do see some recurring themes:

- Humans are good at relative comparisons (what's more valuable)
- Humans are bad at outright comparisons (how much more valuable)
- Data is good at providing comprehensive coverage of things that are countable
- Data is bad at dealing with nuances and qualitative concepts that experts intuitively understand

We've also learned that people only reveal their true opinions after seeing the result. This follows basic product theory: you need to show people something and iterate based on their reactions in order to build something they actually want.

Metrics-driven, experts in the loop

These hard-earned lessons inform our recommendation for how Optimism approaches future round designs. The goal should be to combine the best parts of what humans and data are good at.

Here is the basic framework we propose:

- We use metrics-based evaluations to propose initial token allocations within a domain.
- We let subject matter experts review, fine-tune the metrics, and choose the best allocations.
- Over time, we identify which metrics best align with experts' qualitative assessments, refining models through consistent backtesting.

This approach leverages data's systematic reach with human intuition's nuanced adjustments. Metrics establish a quantitative foundation, ensuring that projects are assessed objectively and fairly, while expert review adds layers of qualitative nuance, including quality, innovation, and momentum. An iterative feedback loop lets us adjust metrics based on expert insights—particularly valuable when experts consistently revise scores or highlight lower-scoring projects. This process is similar to RLHF (Reinforcement Learning with Human Feedback) in machine learning, but with an emphasis on retaining clear, interpretable inputs for expert adjustments.

Practically, we implement this by establishing metrics within a domain, generating proposals for initial allocations (by weighting metrics into an evaluation algorithm), and refining the allocations with expert input. Such an approach should work best in domains with lots of verifiable data, e.g., onchain builders and software dependencies.

This framework should also perform well in fast-evolving domains. Experts can adjust allocations to reflect Optimism's shifting priorities (e.g., prioritizing interoperability transactions over standard transactions), address data blind spots (e.g., fine-tuning metrics for 4337-related projects), and reward innovation (e.g., favoring fast-growing, high-potential projects over more established but static ones).

One essential element is running funding rounds even more frequently and continuously. Doing so lets us pinpoint metrics most correlated with desirable outcomes and backtest these metrics and evaluation algorithms against historical data. Each round remains an experiment, but the cumulative impact across rounds should reveal a clear, positive trend over time.

Hard Choices → Easy Life

The end goal remains super ambitious: to develop predictive models that guide economic policy for the collective. For instance, in 2025, we may want to discover that incentivizing certain behaviors predictably leads to increased interoper transaction volume.

Reaching this goal won't be easy. It requires focus, as the outcomes are highly path-dependent.

Currently, hard choices need to be made around domain scopes. Any changes may be unpopular, especially among community members accustomed to existing Retro Funding patterns. However, narrowing the scope and committing to continuous improvement within scopes are essential steps in reaching the top of the mountain.

[

image

800×661 27 KB

](<https://europe1.discourse-cdn.com/bc41dd/original/2X/e/e8c2684c34954a269f489a8549d29f4e6abe8a64.png>)

Looking back, we'll likely see some initial assumptions were overly optimistic or naive. But we can't improve by continuing to "spray and pray." Governance is ultimately about making hard choices with limited resources.

Optimism has spent two important years learning, but it's time to double-down on what works in 2025.