Why Do Optimized Systems Get Worse?

Optimization Trap vs. Goodhart's Law

A Reality Drift Framework Briefing

Prepared by A. Jacobs

The Common Confusion

- Goodhart's Law explains what happens when a measure becomes a target.
- The Optimization Trap describes the broader system pattern where measurable proxies begin replacing the original purpose.
- One explains metric distortion. The other explains what happens when a whole system reorganizes around distorted proxies.

Optimization Trap vs. Goodhart's Law

Goodhart's Law	**Optimization Trap**
A measure becomes a target	A system reorganizes around measurable proxies
Explains metric distortion	Explains broader purpose distortion
Often begins with one metric	Often spreads across an entire system
The proxy gets gamed	The proxy starts replacing the original purpose
Behavior changes to satisfy the measure	The system keeps improving what it can measure while weakening what matters
A measurement failure	A reality-alignment failure

Source: Optimization Trap When Efficiency Erodes Meaning

Why the Optimization Trap Is Harder to Notice

- The Optimization Trap is harder to notice because the system often appears to be improving. The numbers may rise, the dashboard may look better, the process may become more efficient, and the output may become more consistent.
- The problem is that the system can improve its measurable performance while weakening its relationship to the purpose those measurements were supposed to protect.

Recognition Signs

- You can recognize the Optimization Trap when the system keeps getting better at producing measurable outputs while the real experience gets worse.
- The signs include rising metrics with declining trust, better dashboards with weaker judgment, more efficiency with less meaning, and institutions that continue to function while drifting away from their original purpose.

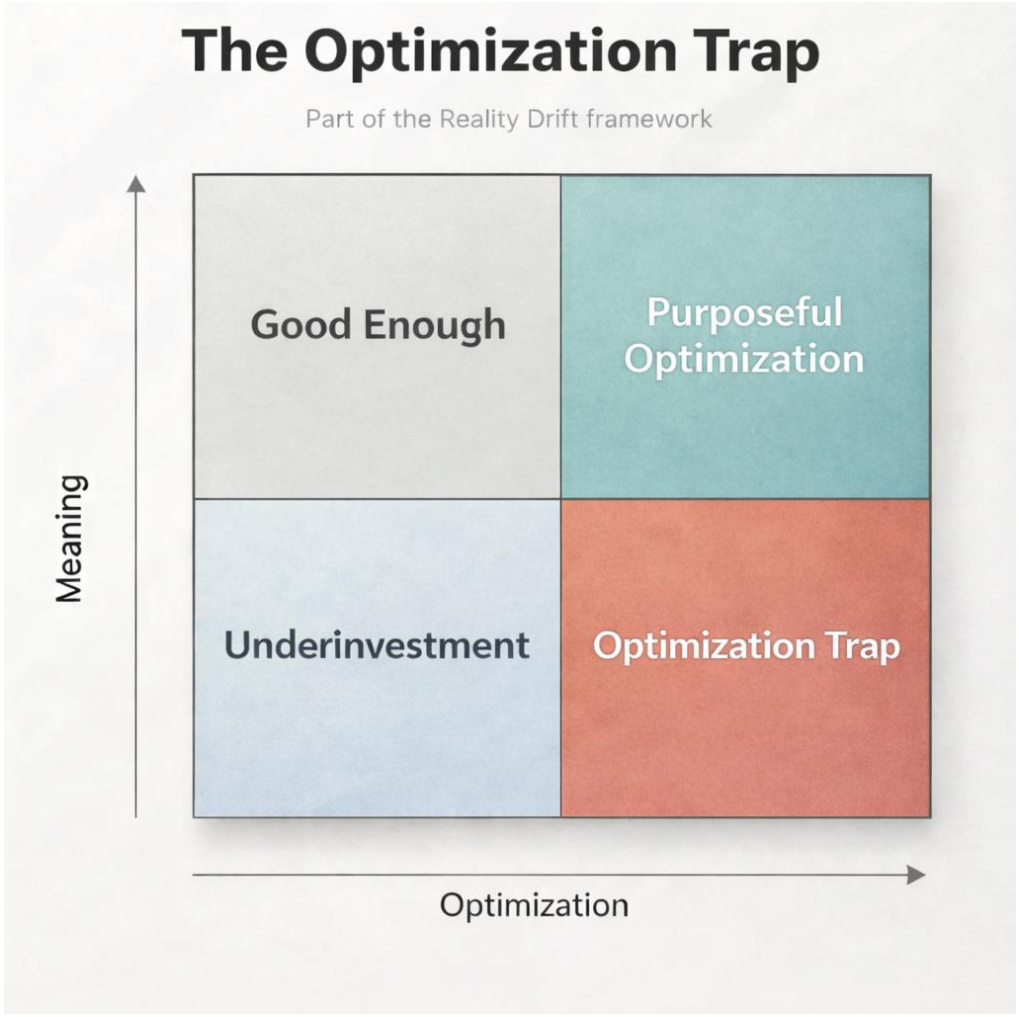

Figure 1. The Optimization Trap Matrix: *This graphic maps optimization against meaning. When both meaning and optimization are low, systems suffer from underinvestment. When meaning is high but optimization is moderate, things may be good enough. When optimization serves meaning, it becomes purposeful optimization. The Optimization Trap appears when optimization increases while meaning declines, creating systems that perform well on measurable outputs while losing contact with the purpose they were meant to serve.*

Source: Reality Drift - Core Visual Frameworks of System Drift and Alignment

Core Takeaway

- Goodhart's Law explains the metric failure. The Optimization Trap explains the world built around it.
- The trap begins when optimization stops serving meaning and starts replacing it.

The Problem Is Purpose, Not Performance

- The danger of the Optimization Trap is that it makes decline look like improvement. A system can become faster, cleaner, more scalable, and more measurable while becoming less sensitive to the reality it was created to serve.
- The surface improves while the purpose weakens.

Core Framework and Sources

- Substack (Articles)
- GitHub (Full Library)
- DOI (Research Paper)
- Glossary & Definition

Keywords: *optimization trap, Goodhart's Law, when a measure becomes a target, why optimized systems get worse, metric gaming, proxy metrics, incentive misalignment, measurement distortion, KPI failure, over-optimization, systems losing purpose, performance metrics vs meaning, proxy failure, optimization vs alignment, reality alignment, purpose drift, institutional drift, measurable outputs, unintended consequences of metrics*