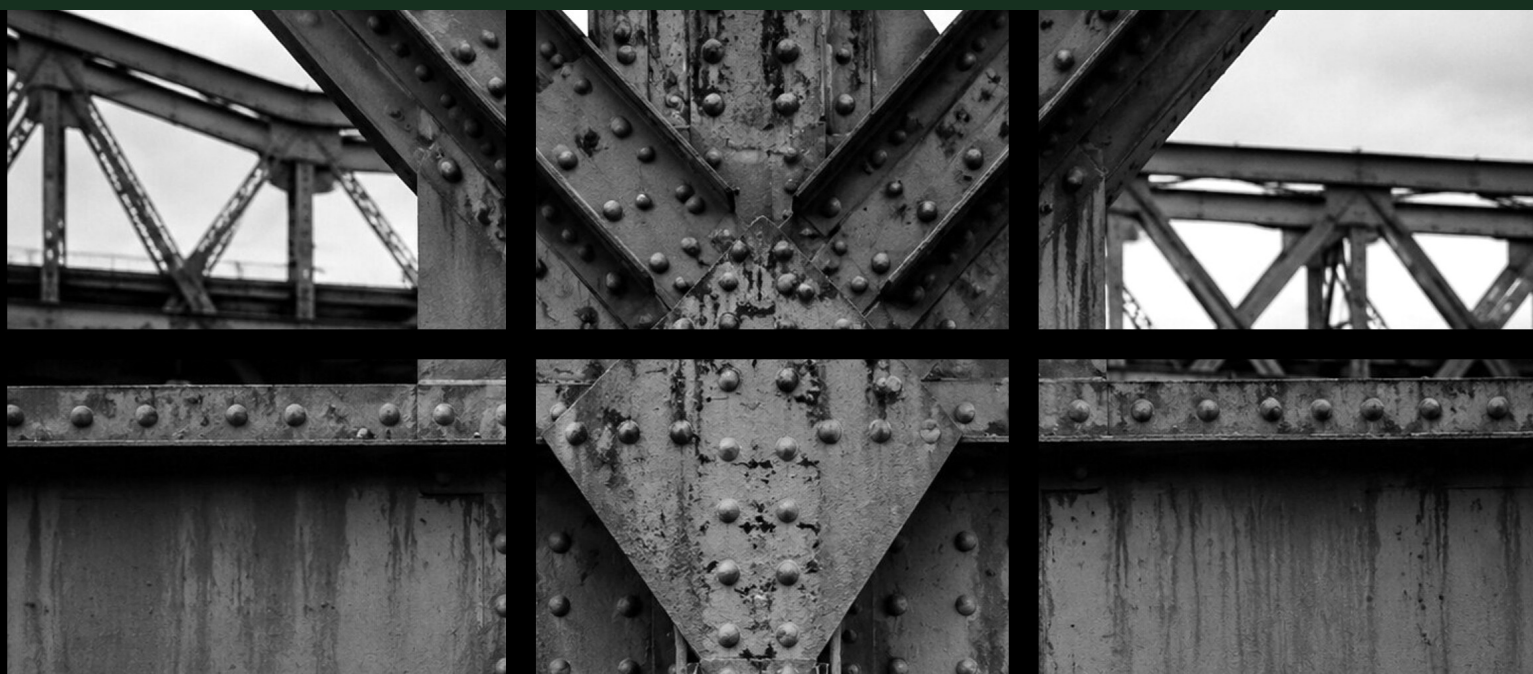


BEYOND THE GREEN REPORT

When the Measure Becomes the Target: Goodhart's Law and the Slow Corruption of Programme Metrics

How the numbers that govern large programmes come, over time, to describe their own management rather than the work — and what to do about it



Giovanni Leonardi

July 2007

Contents

1. Executive Summary	3
2. The Report That Reads Green	3
3. A Law, Not a Failing	4
4. Why Programmes Are Especially Exposed	5
5. The Anatomy of a Gamed Metric	5
6. The Seductive Wrong Answer	6
7. What the Corruption Is Really Telling You	7
8. Living With the Law	7
9. Coda	8

When the Measure Becomes the Target: Goodhart's Law and the Slow Corruption of Programme Metrics

Programme · No. 98

Originally written July 2007 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

*“The report lies in aggregate while being true in every particular,
and that is the most interesting failure mode in the whole of
programme management.”*

1. Executive Summary

Every large programme eventually produces a document that lies without anyone having decided to lie. The monthly report is green. Milestone completion stands at ninety-four per cent. The budget variance is within tolerance. And yet anyone who walks the floor can feel that the thing is stuck — that the work described on the page and the work happening in the building have quietly come apart. No one falsified anything. Each number was assembled honestly by someone doing their job. The report lies in aggregate while being true in every particular, and that is the most interesting failure mode in the whole of programme management.

This essay is about why that happens, and it argues that the cause is not weak governance, dishonest managers, or the wrong choice of metrics. The cause is a law — as close to a law as the management of human systems affords — first stated by an economist studying monetary policy and since rediscovered in every field that tries to steer complex work from a distance. When a measure becomes a target, it ceases to be a good measure. Goodhart's Law is not a warning about bad metrics. It is a statement about what happens to any metric, however well chosen, the moment people's fortunes come to depend on it.

The argument proceeds in four moves. First, that programmes are unusually exposed to Goodhart's Law because they can only ever manage proxies — you cannot measure "transformation," so you measure milestones, spend, and defect counts, and the gap between proxy and reality is precisely where the corruption grows. Second, that the corruption is not fraud but adaptation: intelligent people responding rationally to the signal the organisation actually sends. Third, that the instinctive cure — more measurement, better measurement, harder audit — makes the disease worse, because measurement cannot police measurement indefinitely. And fourth, that the only durable responses are structural: never let one number serve both learning and accountability, hold measures loosely, protect the honesty of the reporting chain, and keep someone who understands the work close enough to the work to notice when the numbers have started to describe themselves.

2. The Report That Reads Green

Begin with the divergence, because it is the symptom everyone recognises and almost no one diagnoses correctly. A transformation programme of some ambition — say, the replacement of a core operational platform across a dozen business units — has been running for fourteen months. The steering committee receives, each month, a report of

real craftsmanship. Red-amber-green indicators across nine workstreams, mostly green. A milestone burn-up showing ninety-four per cent of planned milestones achieved to date. Earned value metrics — a schedule performance index of 0.97, a cost performance index of 1.01 — that a textbook would call the very picture of a well-run programme.

Now walk the floor. The integration team has not had a stable environment to test against in six weeks. Three of the twelve business units have privately concluded the new platform will not do what they need and have begun, without telling anyone, to plan how they will keep the old one running. The "achieved" milestones include four that were achieved by the expedient of narrowing what they meant. The programme is, in every sense that matters, in serious trouble. And the report is green.

Here is the point that unsettles people when they first sit with it: no one in this story has behaved dishonestly. The workstream lead who reported green did so because, against the criteria he was given, his workstream was green. The person who narrowed the milestone did it in a planning session that everyone attended, for reasons that seemed sensible at the time. The earned value figures are arithmetically correct. The lie is emergent. It is a property of the system, not of any person in it, and that is exactly why tightening the honesty of individuals will not fix it.

3. A Law, Not a Failing

The mechanism has a name, and naming it correctly is the beginning of taking it seriously. In its original form it was an observation about central banking: any observed statistical regularity will tend to collapse once pressure is placed upon it for the purpose of control. A monetary aggregate that reliably tracked the economy stopped tracking it the moment policymakers began targeting it, because the act of targeting changed the behaviour that had produced the correlation. The same insight has been stated, independently, by a social scientist studying educational and social reform — the more any quantitative indicator is used for decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort the processes it was meant to monitor — and given its crispest modern phrasing by an anthropologist: when a measure becomes a target, it ceases to be a good measure.

The reason this deserves the word "law" is that it does not depend on anyone behaving badly. It depends only on two things being true, both of which are almost always true in a programme: that the number is a proxy for something we actually care about but cannot directly see, and that people's standing depends on the number. Given those two conditions, human ingenuity does the rest. Effort flows, entirely rationally, towards the thing that is rewarded — the number — and away from the thing the number was standing in for. The proxy and the reality, once loosely coupled, begin to drift apart, and they drift fastest precisely where the pressure to perform is highest.

A metric is trustworthy in almost exact proportion to how little depends on it. The instant a measure carries real consequences, it recruits the intelligence of everyone it governs — not to improve the underlying reality, but to satisfy the measure. This is not cynicism. It is what it feels like, from the inside, to be held to account by a number.

This is why the phenomenon is so often misdiagnosed. Leaders see the gap between the green report and the stalled programme and conclude they have a people problem — someone was not straight with them — or a tooling problem — the metrics were the wrong

ones. They almost never conclude they have a law problem, because the law is counter-intuitive: it says the failure would have happened with honest people and well-chosen metrics too. The better the metric, in fact, the more tempting a target it makes, and the more completely gaming it can substitute for doing the work.

4. Why Programmes Are Especially Exposed

The public debate about targets in the delivery of public services over the past few years has made the general pattern familiar — waiting times that improve on paper while the experience of waiting does not, examination results that rise while the capability they supposedly certify does not. Programmes are exposed to the same forces, and for structural reasons they may be exposed more acutely.

- Everything of value in a programme is unmeasurable directly. You cannot put a meter on "the organisation has successfully adopted the new way of working." You can only count proxies — milestones passed, modules deployed, users trained, tickets closed. Every one of these is a shadow of the real thing, and every shadow can be produced without the real thing casting it.
- The measured control the measurement. In most programmes the people who report progress are the same people whose progress is being judged. This is not corruption; it is org design. But it means the reporting chain has, at every link, a mild and understandable interest in the amber not becoming red, and mild interests compounded over many links produce a strong systematic bias towards green.
- The consequence lags the gaming by months. When a milestone is quietly hollowed out, nothing bad happens immediately; the bad thing happens later, at integration, at go-live, at the first month-end on the new platform. By the time reality arrives, the person who shaved the scope has been promoted, the reporting period has closed, and the causal thread connecting the early game to the late failure has been cut. The feedback loop that would teach the system not to game is too slow to teach it anything.

Put these together and a programme starts to look like an almost perfectly designed apparatus for Goodhart's Law: irreducibly reliant on proxies, reported on by the people it judges, and shielded from the consequences of distortion by long time lags. The wonder is not that programme metrics corrupt. It is that anyone expects them not to.

5. The Anatomy of a Gamed Metric

None of this stays abstract for long once you know what to look for. The corruption of a programme metric follows recognisable patterns, and it is worth setting them out plainly, because naming the move is most of the defence against it.

Metric under pressure	How it is satisfied without doing the work	What it stops measuring
Milestone completion	The milestone's definition of done is narrowed in a planning session until it fits what is actually finished	Whether meaningful progress occurred

Metric under pressure	How it is satisfied without doing the work	What it stops measuring
Schedule performance	Future work is rebaselined so that today's position sits on the new curve	Whether the programme is on track to its original promise
Defect count	Severities are reclassified downward; defects become "known limitations" or "change requests"	The true quality of what is being built
Adoption or training numbers	People are trained and logged; whether they then use the system is not measured	Whether behaviour has actually changed
Budget variance	Cost is moved between capital and operating lines, or parked in a later phase	Whether the programme will cost what was promised

Consider the milestone line, worked through with numbers, because it is the most common and the most invisible. A workstream has a milestone: "Payments module integration complete," due at month twelve. At month eleven it is clear integration will not be complete on any honest reading. Rather than report red, the team proposes — reasonably, in a room full of people who want the programme to succeed — to split the milestone: "Payments module integration complete (core flows)" now, "Payments module integration complete (exception handling)" deferred to month fifteen. The month-twelve milestone is achieved. Completion stays at ninety-four per cent. And the exception handling — which is where, as every practitioner knows, the genuine difficulty of payments always lives — has been moved out of sight and out of the metric. Nothing false was reported. The measure has simply been taught to look past the hard part.

Multiply that single move across nine workstreams and fourteen months and you have the green report over the stalled programme, fully explained, with no villain anywhere in it.

6. The Seductive Wrong Answer

Faced with all this, the instinct of a competent, conscientious leadership is to reach for more of what it already trusts: better measurement. If the milestone metric was gamed, define done more rigorously. If schedule performance was manipulated by rebaselining, lock the baseline and audit every change. Triangulate — add leading indicators, add independent quality gates, add a second reporting line that does not report to the programme. Measure the measuring.

This is the most serious objection to the argument of this essay, and it should be met head-on rather than waved away, because it is half right. Rigorous definitions of done are better than loose ones. Independent assurance does catch some gaming. A programme with no measurement at all is not honest; it is merely blind. The case for more and better measurement is not stupid, and a certain amount of it is simply good practice.

But it fails as a general strategy, for a reason that follows directly from the law itself. Measurement cannot police measurement past a point, because every new metric you add is a new target, and every new target recruits the same ingenuity to satisfy it. Harden the definition of done and effort flows to satisfying the letter of the new definition. Add an independent quality gate and the work reorganises to pass the gate. Introduce a second

reporting line and, within two cycles, the programme has learned what the second line looks at and manages that too. You have not escaped Goodhart's Law; you have paid more to run into it again, and you have added cost, delay, and a thicket of indicators that obscure the few signals that still mean something. The distortion is not a bug in this or that metric that a better metric would remove. It is a structural consequence of governing through metrics under pressure, and no quantity of metrics is its cure.

“The programme that responds to a gamed number by adding three more numbers has not become harder to fool. It has merely become more expensive to fool, and more confident that it cannot be.”

7. What the Corruption Is Really Telling You

Step back and the phenomenon reframes itself. Metrics exist because we govern at a distance. A steering committee cannot watch the work; it watches numbers about the work, because attention is scarce and the programme is large and no one at the top has time to walk every floor. The metric is a substitute for presence — a way of knowing without looking. Goodhart's Law, seen this way, is simply the tax levied on governing without looking; the further the decision-maker sits from the work, the more the reported reality and the real reality are free to diverge, and the more the system optimises for the report.

This is the same gap, wearing programme clothes, that haunts every discussion of why transformation intent so rarely becomes transformation reality. The intent lives in the numbers — coherent, green, on track. The reality lives on the floor — partial, contested, behind. The measurement system does not merely fail to close that gap; left to itself it slowly widens it, because it applies pressure to the one side of the gap it can see, and the work quietly rearranges itself to relieve the pressure rather than to close the distance. A programme that is managed purely by its metrics is, over time, being optimised to look transformed. Whether it is transformed becomes, alarmingly, a separate question that the reporting system is structurally unequipped to answer.

8. Living With the Law

You cannot repeal Goodhart's Law, and the counsel of despair — abandon metrics, trust everyone, look at everything yourself — is not available to anyone running real work at real scale. What is available is a set of structural habits that keep the law's damage within tolerable bounds. They are not techniques so much as dispositions, and they cost something, which is why they are rare.

1. Never let one number serve both learning and accountability. This is the single most important discipline. The moment a measure that people were using to understand the work becomes the measure they are judged by, it stops telling the truth about the work. Keep the two functions on separate numbers, and keep the learning numbers safe — explicitly, visibly safe — from being used to reward or punish. A metric that can

end a career will only ever tell you what someone wants their career to look like.

2. Hold measures loosely and rotate them. A metric watched forever is a metric gamed forever. Metrics that change — that are retired once understood and replaced with different views of the same reality — give the gaming less time to calcify, because the target keeps moving before the workarounds mature. Stability of measurement feels like rigour; it is often just a longer runway for corruption.
3. Protect the honesty of the reporting chain by rewarding early red. In most programmes, red is punished and green is rewarded, which teaches everyone to defer red as long as possible — the exact opposite of what a leader needs. The counter-move is deliberate and cultural: make the person who raises the early, honest red visibly better off than the person who held green until it collapsed. A programme learns what it rewards, and if it rewards reassurance it will be reassured all the way into the wall.
4. Keep someone who understands the work close to the work. Every metric is an abstraction, and abstractions can only be sanity-checked by someone who can still see the thing abstracted. The oldest instinct in operational management — go and look at the actual work, in the actual place — is not nostalgia; it is the only reliable defence against a number that has begun to describe itself. One person who understands payments, standing in the integration team's space for a day, will learn more about the true state of the "ninety-four per cent" than a quarter of steering reports.

9. Coda

The green report over the stalled programme is not a scandal and not an accident. It is Goodhart's Law doing exactly what it always does, in a setting almost perfectly built for it. The mature response is not outrage at the people who produced the report, nor faith that a better dashboard would have told the truth. It is the recognition that every number used to govern complex work is, from the day it acquires consequences, in a slow race between its usefulness and its corruption — and that the job of anyone who governs through metrics is to keep changing the terms of that race, never to imagine it can be won once and for all. The measures will always, in the end, become targets. The only choice is whether you notice in time.