THOUGHT LEADERSHIP

The Most Durable Problem in Management

For two decades, companies have left about 33% of the value their strategies promise on the table. Every method we've thrown at that gap has helped at the edges, but none has closed it. A layer of AI agents may be the first thing that does, and it gives your strategy execution office superpowers.

Jay Goldman July 27, 2026

I spend most of my time with the people who run transformation inside large companies, and lately, the same conversation keeps coming up. They can tell you, in detail, what their strategies are supposed to deliver, but ask them what actually landed, and everyone gets a little uncomfortable and squirmy and has approximate answers that never quite add up.

That distance between what a strategy promises up front and what's delivered in the end is the most enduring problem in management. Certainly not for a lack of trying, but it's outlasted every remedy prescribed for it, from program offices to balanced scorecards to even transformation methodologies and quarterly objectives. Despite all that, companies still capture only two-thirds of the value their strategies set out to deliver, which is roughly the same as they captured when the figure was first measured twenty years ago.

So, a little skepticism is healthy when a new technology shows up claiming to fix what nothing before it could. That is more or less the claim now being made for AI agents, and it's pretty hard to tell the hype apart from the reality, especially when previously unimaginable enterprise values rest in the balance. A narrower version is certainly true, though. For the first time, maybe ever, the execution office can get the always-on intelligence it has always needed, and that changes what those offices can do.

WHY IT NEVER CLOSED

Why the gap survived this long

Let's start with why the gap has been so stubborn.

We're big fans of the Toyota Kaizen Five Whys approach, which is a fast and easy way to get to root cause analysis. Keep asking "Why?" five times, and you'll get through the surface reasons and excuses. In this case, we only need three.

Why has the gap survived? The first layer under it is failures of coordination and information. These manifest as problems that were never prioritized, risks no one raised before business cases were signed, programs slipping quietly for six weeks before SteerCos noticed, and dependencies that fell into the space between two siloed teams.

Why did those failures of coordination and information happen? None of this is intellectually hard. It's even very easy if you just have one project to worry about. The problem is scale. Humans aren't very good at finite attention to detail, magnified across programs that each contain many workstreams. It's restless, continuous, unglamorous work that no leadership team has the hours to keep up, and the program office we invented to supply it is doing that work inside a fixed number of hours, against a portfolio that keeps growing.

The work that would actually close the gap was always possible. There has simply never been enough time and resources, because it's expensive human labor that has to be laser-focused to keep up with all the reporting needs in a quarter. Across the programs we work on, we see teams give 25% or more of their time to status reporting. Think about that as a budget allocation for a moment: one quarter of your headcount cost is spent on reporting on status.

Final why: why does all that status reporting not help? Even if you had enough eyeballs to pay attention to all the right things, and enough brains to do the analysis and write reports, the sheer volume of reporting would drown any SteerCo or board's ability to keep up. That scarcity, not a shortage of methods or understanding, is the root cause of the gap.

Sit through enough of these SteerCo meetings, and you know how it goes: mostly arguing about whose numbers are right. The decisions that would actually move the strategy get whatever time is left over, which is never much. The only real solution is to shift from status reporting to exception handling, so

that all the numbers are correct and the actually green projects take care of themselves while you focus scarce human attention where it can make a real difference.

THE SHIFT

What an agentic transformation management office is

An agentic transformation management office, or ATMO, is what emerges when that always-on intelligence is no longer a constraint. Set aside the science-fiction version in which an autonomous system runs the company while the executives play golf. We're still a long way from that, and if the robots even let us golf, the recent World Humanoid Robot Games suggest we're going to get quickly out-golfed. The useful version is more modest and possible today: a layer of software agents that continuously perform the gathering, reconciling, monitoring, and prompting that the execution office always wanted but has never had the capacity to keep up with.

Matthew Prince, CEO of Cloudflare, wrote a WSJ op-ed on how they decided who to replace with AI when letting go of 20% of their workforce. He categorized their team into three broad categories: builders, sellers, and measurers. He says, "Builders create products. Sellers sell those products. Measurers do everything else: internal audit, revenue recognition, finance, legal, compliance, middle management, operations, and on and on."

Unsurprisingly, the vast majority of the people they laid off were measurers. They cut middle managers because AI enables more managers to measure and mentor larger teams. They consolidated operations teams, removed their marketing team's measurers, and automated many finance functions. These efficiency gains are real and available to your strategy, transformation, and value creation teams today if you put the right capabilities in place.

EXHIBIT 1

The agentic operating loop

Read it this way. The agents read the systems where work actually happens, the financials, the delivery tools, the tickets, the documents, the message threads, reason over what they find, and either surface it to a person or, within limits agreed in advance, act on it. The office keeps the part that was always its own and always crowded out by administration: judgment. What changes is not how smart the people in the room are. It is how much analysis they can bring to bear, which rises from whatever a stretched team could manage to something close to continuous. **Turquoise nodes run continuously. Purple nodes stay human.**

In this version, judgment stays with the humans. What doesn't is the continuous data gathering, analysis, and reporting underneath that judgment. That's the part that used to be too expensive and no longer is.

When we started building this execution velocity into Conductor, I assumed the hard part would be the technology. It was not. The models are ready, and more capable than most people realize. The hard part is everything upstream, which is where the rest of this piece focuses.

THE COST OF THE GAP

What the gap costs

Hidden costs are the worst kind. Being hidden doesn't make them zero, but it makes it significantly harder to find and attribute them correctly. Losing 25% of your team's time to status reporting doesn't typically show up as a spreadsheet line item, but it's costing you very real money on a very ongoing basis.

Think about it this way: the teams we work with are typically running the biggest bets their companies will make. They're in charge of the strategy office that greenlights everything the company does to deliver on its mission and create value. They're leading post-merger integrations and transformations to change how the company operates, its business models, and how it achieves benefits. Our customers and partners deploy the most advanced approaches to doing that, with the latest technologies, but they very rarely have the opportunity to turn that lens inward and evaluate how they get that done.

We've had a seat at those tables across thousands of projects in hundreds of enterprise-scale programs. That gives us an unparalleled perspective into how to run effective programs, and how to connect strategy all the way through execution to maximize returns. Across all those programs, we've identified and documented the four most common ways value leaks, supported by independent research.

EXHIBIT 2

The gap, four ways

63%

of a strategy's value is realized on average. About a third leaks in execution, and the number has barely moved in twenty years.

MANKINS & STEELE, HBR

1/6

major IT programs becomes a "black swan," an overrun averaging around 200 percent, in a database of 16,000+ projects.

FLYVBJERG, OXFORD

72%

of PMOs still spend half a day or more every month just assembling reports by hand.

WELLINGTONE, 2026

~10¢

of every dollar is wasted on poor project delivery, roughly one dollar in ten.

PROJECT MGMT INSTITUTE

None of this is a strategy problem. The strategies are usually sound. It is an execution problem, and underneath that, an intelligence problem. Value leaks at the handoffs, in the space between what leadership committed to and what the organization can actually see, week to week, without someone staying up to watch it.

THE METHOD

Where value leaks: the four gaps

We've run the Five Whys exercise with our customers to understand why those four measures keep recurring. We keep coming across the same four gaps, all of which can be fully or, at least, mostly resolved with always-on intelligence. Agents can help establish clarity on what matters on an ongoing basis throughout long, complex programs. They can keep alignment across many workstreams and projects, and identify resource allocation gaps that don't match the commitments. They can keep the cadence consistent between SteerCo meetings. And, maybe most critically, they can deliver insights from the same perspective across workstreams, ensuring hidden KPI truths are seen and understood the same way.

EXHIBIT 3

The four gaps, worked continuously

01 Clarity

An agent keeps the ranked priorities and their intended benefits legible, and flags the new request that maps to none of them. It can't decide what matters. It can stop the portfolio sprawling back to forty things without anyone noticing.

02 Alignment

It watches whether funding, capacity, and ownership actually sit behind the stated priorities, and surfaces the program everyone calls strategic that is quietly staffed by two contractors. Those discoveries stop happening by accident, and late.

03 Cadence

It assembles what moved, what is blocked, and the decision each owner needs, so the review stops being a status ritual and becomes a forum for decisions. A slipping dependency escalates the day it slips, not at the next steering committee.

04 Insight

It measures benefits against the business case continuously, from the real operational and financial data rather than a self-reported register, and flags the program falling below plan the week it starts to, not at the closure review.

The pattern underneath all four is the same. The work that closes the gap was never intellectually hard, only relentless, the continuous analysis no team ever had the hours to sustain. That is the scarcity agents lift, and continuous measurement is what finally makes continuous reallocation possible.

WHAT IT LOOKS LIKE

A day in the life of an ATMO

Picture the office on an ordinary Tuesday.

Overnight, a benefits agent reconciled the portfolio's claimed benefits against the source data and found one program running eight percent below its business case. It drafted a one-page exception report with the likely cause and three recommendations. A dependency agent noticed that a data-migration workstream is a week from blocking two others, and already routed the question to the person who can mitigate it. A clarity agent flagged a new project request from a business unit head that arrives with a healthy budget and no connection to any of the year's strategic priorities.

The transformation director opens a single view that already knows what changed and why. The next hour is devoted to three pressing and key decisions: whether to intervene in the slipping program, how to clear the dependency, and whether to let the disconnected request through. The always-on intelligence has compiled everything and presented it within minutes, instead of the three days it would once have taken to assemble the same picture. Because the data is correct, the green statuses acknowledged, and the exceptions flagged for attention before anyone sits down, the executive review later that week runs for 40 minutes and produces four actionable and immediate decisions.

THE NUMBER THAT MATTERS

The North Star metric: execution velocity

There is a single measure underneath our thinking. We call it execution velocity: how fast, and how completely, an organization turns strategy into realized value. Importantly, this isn't a vanity metric of activity or milestone throughput. Execution velocity is all about realized value and the time it took to get there. Every one of the four gaps agents helps close is in service of the same thing: improving execution velocity.

THE ONE MEASURE TO OWN

Execution velocity =

$$\frac{\text{priority initiatives} \times \text{realization rate} \times \text{average value per initiative}}{\text{time to benefit}}$$

More of the right work, more of its value actually realized, achieved faster. Almost no one measures it yet.

THE ROAD THERE

Four stages of maturity

Most organizations sit at the first of four stages, and the distance between them is mostly a question of how much the office is willing to trust something other than a person to do the watching and the analysis.

We've seen the same pattern play out in coding. At first, engineers didn't trust AI coding agents and gave them routine, lower-thinking tasks like writing tests. Then they started taking on code writing, but only strictly supervised, with every update reviewed by a human. Now they write all the code, and the humans have moved up the abstraction ladder to supervising teams of agents and providing direction. That's the place we should sit, using our creativity and taste to guide the outcomes, and it's the goal we see leading transformation and value creation teams setting for themselves.

EXHIBIT 4

The road there, in four stages

01 · ASSISTIVE	Agents draft the status and summarize the register. Copilots for the analysts. Useful, and a long way from the prize.
02 · CONTINUOUS	Agents monitor benefits and dependencies unprompted and raise exceptions between reviews. This is the jump that matters.
03 · ORCHESTRATING	Agents run the cadence end to end, composing the review and routing each decision to the person who has to take it.
04 · BOUNDED AUTONOMY	Within limits agreed in advance, agents re-sequence work and reallocate capacity on their own, bringing only the exceptions to a human.

Most offices sit at stage one. The leap worth making is from one to two, from a faster way of producing reports to a system that analyzes the portfolio when no one is looking. Most of the value lives there, and so does most of the difficulty in adopting it, because it is the point at which the office has to trust something other than a person to do the analysis.

WHAT STAYS YOURS

Clarify first, then automate

Garbage in, garbage out remains as true as ever, but the definition of garbage has changed. It used to be that piles of unstructured, messy data could be the death of a transformation team, but now they can point agents at them and get back clean, structured data. The critical part is the prompts to those agents. Bad, unclear prompts with interpretable goals will result in a higher rate of hallucinations and bad outcomes. The critical part is the clarity that human judgment and taste bring.

An ATMO sitting on top of an unclear strategy and messy prioritization won't help. An agent can enforce a clear strategy, but it can't invent one. It will leak value faster, with more conviction, and hand you beautifully evidenced reports all the way down. A human still has to decide what matters, what it's worth, and what to stop. The humans on your team need to clarify first, then the agents can automate at scale.

EXHIBIT 5

Where the line sits

HUMANS KEEP

The consequential, human calls

- Deciding what matters, what it is worth, what to stop
- Killing a program with a powerful champion
- Moving scarce specialists against the politics
- Holding an executive to a benefit they missed
- Governing the agents, and answering for them

AGENTS TAKE

Continuous, and answerable to a human

- Gathering status and reconciling the numbers
- Watching benefits against the business case
- Catching a slipping dependency the day it slips
- Flagging the request that maps to no priority
- Composing the review before anyone sits down

One new job the office never had: governing the agents. What they can do on their own, what they must bring to a person, and who answers when one of them is confidently wrong. If you run on agents, you are accountable for them, and that part never automates.

THE PRACTICAL PART

Three steps to start

If you're at stage one, here's where to begin.

Rank clearly. Focus the portfolio on the few priorities that actually carry the strategy, and stop or park the rest. This is the human act on which the whole system rests. An agent can hold the line once you've drawn it, but it can't draw it for you.

Wire one benefit to real data. Pick a flagship program and connect its business case to the systems where the work and the benefits are tracked, so its health is measured continuously instead of once a quarter.

Run the loop on that one program. Put agents on continuous watch for it, keep judgment with your human team, and what you learn there sets the pattern for the next program, and the one after that.

With that you've already moved from stage one to stage two. It is also the only version of this that compounds: the standards you set on the first program are what make the next one faster, and the one after that.

WHERE I AM GOING WITH THIS

An invitation

This is the problem my team spends its days on. We're getting close to launching the next version of Conductor, and it's built on everything in this post. The closer we get, the clearer one thing becomes: the opinion that matters most to me now doesn't come from yet another analyst or pundit. It's yours, from the people who actually run strategy execution every day.

So, if you're curious to see what we're building, reach out, and I'll walk you through it myself. You understand this work better than anyone else, and I'd like your fingerprints on it.

Last thing, this newsletter is where I share our thinking. Subscribe to stay in the loop. Comment below to share your thoughts and experiences, ask questions, or ask for a tour of the all-new Conductor. I read and respond to every comment.

Jay

Sources

1. Strategy-execution gap, about 63 percent of value realized and roughly a third lost: Michael Mankins & Richard Steele, Turning Great Strategy into Great Performance, Harvard Business Review, 2005.

2. Megaproject overruns and "black swan" IT programs: Bent Flyvbjerg, University of Oxford, research on large-project performance across 16,000+ projects.

3. Reporting overhead in PMOs: Wellingtone, State of Project Management, 2026.

4. Wasted investment on poor project performance: Project Management Institute, Pulse of the Profession.

The 25% status-reporting figure is our own observation across the programs we work on, not published research. Draft for review; the office name is pending Rich's confirmation.

READ ONLINE

Read this article on our website

getconductor.com/resources/the-most-durable-problem-in-management

GET STARTED

Talk to sales

Walk through an example with our team and see Conductor on your own programs.

Talk to Sales →

Scan to book