CHAPTER 1

Bad Measures, Ugly Measures, and Good Measures

Bad measures waste your time. Ugly measures destroy it. Good measures transform you.

Walk into almost any company today and you'll find a wall of dashboards. Real-time charts, color-coded scorecards, weekly metric reviews, quarterly business reviews, KPI trees cascading from the CEO down to individual contributors. Companies have never measured more, more often, with more sophistication. And yet a strange thing keeps happening: the numbers look fine, and the business doesn't. Or the numbers look great for a while, and then something quietly breaks. Or two teams stare at the same dashboard and reach opposite conclusions about what it's saying. If you've sat through a meeting where everyone has data and nobody has clarity, you already know the feeling.

This chapter is about why that happens — and it almost never traces back to the data itself. It traces back, again and again, to the **measures** the company chose, how they're defined, what they hide, what they incentivize, and whether they still mean what they used to. Most of the analytics mistakes covered in later chapters are downstream of this one. Get your measures wrong and no amount of clever analysis will save you; get them right and even simple analysis becomes powerful. In the sections that follow, we'll walk through the most common ways measures go wrong — from picking the wrong metric, to gaming, to hidden definitions, to composite scores that aren't what they seem, to the hardest one of all: good metrics that quietly stopped being good while nobody was watching. The goal isn't to make you an expert in metric design. It's to make you the person in the room who knows when to trust the number on the dashboard, and when to push back.

Part I — The Strategic Foundation

Measures Are a Strategic Choice

What you decide to track is not a reporting decision. It is a declaration of your business model.

Mr. Watanabe arrives at Shin-Osaka station fifteen minutes before departure. He already knows the train will leave at exactly 6:03 p.m. So he takes his time — stops for a bento, a cold Kirin beer, a sweet for his son back in Tokyo. No rushing, no uncertainty. When the train departs on time, he settles in, opens his book, and lets the journey begin.

Mr. Watanabe could plan his evening around a six-minute window because JR West had made punctuality — not speed — the metric that ran the entire organisation. That is not an operational detail. It is a strategic choice.

The metric that ran a railroad

In the early 2000s, the president of JR West Japan gave a seminar in which he spent considerable time on a single number: the company's on-time departure rate, which had improved from 95% to 96%. His tolerance band for "on time" was plus or minus fifteen seconds.

To many in the room, this seemed like an oddly narrow obsession. China was building faster trains. Wasn't speed the real competition?

His response reframed the problem entirely.

Shinkansen, he explained, does not compete with Chinese high-speed rail. It competes with Japanese domestic airlines. And when you understand that, the entire logic of what to measure changes.

Airline passengers travel airport to airport. Shinkansen passengers travel door to door — and stations sit in city centres, airports do not. So the relevant measure of time is not train speed alone, but total journey time from origin to destination. On that measure, Shinkansen is highly competitive for routes like Tokyo to Osaka, regardless of top speed.

Then he explained something more subtle. Shinkansen operates on a single-track network with frequent stops. One delay cascades across the entire system. The trains

run so frequently — at peak, just minutes apart — that even small variations in departure times force wider safety intervals between trains, which directly reduces the number of services that can run per day, which reduces total passenger capacity, which affects revenue. The mathematics were precise: tighten on-time performance and you can effectively increase capacity without laying a single metre of new track.

Speed, he noted, is one variable in this equation — but not the controlling one. Push speed too hard and safety requirements force longer gaps between trains, which can actually reduce total capacity. The system optimises for something more complex than any single dimension.

That is why the strategic measure was on-time performance, not maximum speed. It was not a reporting choice. It was the business model expressed as a number.

The same industry, a completely different question

In his TEDx talk Perspective is Everything, Rory Sutherland raises a question about Eurostar that is worth sitting with. Six million pounds, he points out, was spent reducing the London to Paris journey time by around forty minutes. His provocation: for a fraction of that investment, the same train could have had Wi-Fi on every carriage — which would not have shortened the journey, but would have made it significantly more useful and enjoyable. Invest a little more in genuine comfort and service quality, Sutherland argues, and passengers might stop measuring the journey in minutes altogether.

His deeper point is not really about trains. It is about how the way we define a problem determines what we measure, and what we measure determines what solutions we are even willing to consider. If the goal is defined as "reduce travel time," you build faster infrastructure. If the goal is defined as "make the journey feel worthwhile," you invest in experience. Both are coherent strategies. Both are defensible business models. But they lead to entirely different operating priorities, entirely different capital decisions, and entirely different organisations.

Neither JR West nor Rory's Eurostar example is presented here as right or wrong. The point is that they represent genuine strategic choices — and those choices are visible, first and foremost, in what each organisation decides to measure.

The question leaders rarely ask

Ask most senior leaders which measures matter most to their business, and the answers come quickly: sales, profit, market share. These are not wrong answers. But they are incomplete ones. Every competitor in your industry is also watching sales,

profit, and market share. They do not tell you what is distinctive about how your business creates value — only whether it is, in aggregate, succeeding or failing.

The more revealing question is this: what are the measures that reflect how your business model specifically competes? When leaders sit with that question, the answers slow down — and become far more interesting.

That is precisely the shift the Shinkansen example illustrates. On-time performance is not a generic measure of success. It is specific to a business model that competes on total journey reliability, operates on a networked infrastructure where one variance affects everything, and serves passengers whose decision to choose train over plane depends on predictability as much as speed. Remove any one of those conditions and a different measure might become the right one.

What your measures actually say about you

Most organisations inherit their metrics. A dashboard was built before the current leadership team arrived. The executive scorecard reflects priorities that made sense in a previous strategy cycle. The KPIs in the quarterly review have been there so long that no one questions whether they still reflect the business model.

But measures do not sit passively on a slide. They shape behaviour at every level of the organisation. What gets measured gets managed — and more importantly, what gets measured signals to every team, analyst, and frontline manager what the organisation actually values, as opposed to what it says it values. The gap between those two things is where misalignment quietly accumulates.

Consider a sales team measured purely on new revenue. Without being told to, they will prioritise winning new accounts over serving existing ones. No one instructed them to do this. The metric did. Now consider the same team with a measure that equally weights customer retention. Their behaviour shifts — not because of a new strategy document, but because the scorecard changed. This is the quiet power of measures. They do not just report on your organisation. They run it.

Questions for leaders to consider

The following questions are not a checklist. They are a prompt for the kind of deliberate thinking that metric selection deserves — the same quality of thinking applied to strategy, capital allocation, or organisational design.

- When you look at your current KPIs, can you articulate precisely why each one reflects your competitive advantage — not in general terms, but specifically, for

your market and your business model?

• If a new executive joined your organisation today and could only see your performance dashboard, would they be able to accurately infer your strategy from it?

• Which of your measures were designed for a previous version of your business? Are they still fit for purpose, or have they become inherited furniture?

• Where in your organisation do you see capable people working hard on the wrong things? Is there a misaligned measure somewhere upstream that is quietly producing that behaviour?

• Who in your organisation is currently responsible for deciding what gets measured? Is that the right conversation to be delegating?

• What would you measure if you were building your performance framework from a blank page today, knowing what you now know about your customers, your competitive position, and your operating model?

> *The measures you choose are not a summary of your strategy. They are its operating system — the layer that turns intent into daily action across every function in the business.*

Chosen with care, they align organisations around what genuinely matters. Inherited without examination, they quietly run an older strategy that may no longer be yours.

That is why the decision deserves the attention of leadership — not as an afterthought once strategy is set, but as an integral part of setting it.

Part I — The Strategic Foundation

When the Scorecard's Green but the Business Is Bleeding

Most data problems are not data problems. They are wrong-measure problems.

You have seen this dashboard. Twenty-something metrics, mostly green, a few amber, one persistent red. Teams meet weekly to defend their numbers. Quarterly reviews celebrate the greens and investigate the reds. Everyone is occupied. The business is still losing ground.

When this happens, the instinct is to question the data. Is it accurate? Is it complete? Do we need better reporting tools? These are understandable questions, but they are almost never the right ones. Better data will not fix this. More data will not fix this.

The right question is: are we measuring the things that actually predict the outcome we want — or are we measuring things that only tell us, after the fact, how we have already done?

The difference between reporting and predicting

There is a useful distinction that sits at the heart of this problem. Lagging indicators measure outcomes that have already occurred — revenue, profit, market share, customer churn. They confirm whether a strategy worked. Leading indicators, by contrast, measure the conditions or behaviours that drive those outcomes before they materialise. They tell you where you are heading, not just where you have been.

Most dashboards are dominated by lagging indicators. They are easier to define, easier to collect, and easier to defend in a review meeting. But a scorecard built entirely on lagging measures is, in practice, a rearview mirror. You can see exactly what happened. You cannot do anything about it.

This is why organisations find themselves in the position of a green dashboard and a struggling business at the same time. The measures are accurate. They are just not the right measures. They confirm what has already occurred rather than signalling what is coming or guiding what to do next.

The more productive question to ask of any metric is not "does this reflect our performance?" but "does this predict our outcome, early enough for us to act?"

Moneyball: when an entire industry tracked the wrong thing

For most of baseball's history, hitters were evaluated by batting average and runs batted in (RBIs). Pitchers were judged by wins and earned run average. These were the accepted measures. Every scout used them. Every general manager signed contracts based on them.

In the late 1990s, Oakland Athletics general manager Billy Beane faced a constraint that forced him to question this consensus. His budget was among the smallest in Major League Baseball — he could not compete for the players who looked best by traditional metrics. So he and his analyst Paul DePodesta, drawing on the statistical work of Bill James, asked a different question. Not "which players have the best numbers?" but "what actually causes a team to win games?"

Their analysis pointed to two metrics the market was significantly undervaluing: on-base percentage (OBP), which measures how often a batter reaches base by any means, and slugging percentage (SLG), which measures the power and run-generating potential of each at-bat. Together, these two metrics predicted runs scored. Runs scored predicted wins. Batting average and RBIs — the measures everyone had been paying premium prices for — turned out to be far weaker predictors of the outcome that actually mattered.

Beane built a competitive team by acquiring players who excelled at OBP and SLG but were overlooked because their traditional numbers appeared unremarkable. In the 2002 season, Oakland won 103 games, set a record with a 20-game winning streak, and did it with one of the lowest payrolls in the league.

The lesson is not about baseball. It is about how an entire industry can converge on measures that feel right, look familiar, and are universally used — while those measures fail to predict the outcome everyone is actually trying to achieve. The consensus view of what to track had gone unquestioned for so long that it had become invisible. It took someone with both a budget constraint and a willingness to ask a foundational question to find the gap.

> *Familiarity is not the same as predictive power. Industry-standard KPIs get adopted not because they have been proven to predict outcomes in your specific business, but because they are familiar, benchmarkable, and easy to justify.*

How to find the right measures

The method is straightforward in principle, though it demands genuine discipline in practice. It begins by reversing the direction most organisations use when building a

dashboard.

Rather than starting with what is easy to collect, start with the outcome you actually need. Then ask what causes that outcome. Then ask what predicts the cause. The measure belongs at that last step — close to the drivers of performance, not at the level of the result.

To make this concrete: consider a subscription business whose goal is long-term revenue growth. The lagging indicator is obvious — annual recurring revenue. But what drives it? Retention. What drives retention? Customer engagement with the core product. What predicts engagement before it becomes visible in retention data? Frequency of use within the first thirty days. That early-use behaviour turns out to be the leading indicator with the most predictive power — and it is also the measure that product teams and customer success teams can actually act on, before churn has occurred and the revenue impact has already been recorded.

The chain looks like this:

- Outcome (lagging): Annual recurring revenue growth
- Driver: Customer retention rate
- Predictor (leading): Product engagement in the first 30 days
- Measure: Weekly active usage rate, tracked by cohort from onboarding

Each step in the chain must be validated — not assumed. The connection between the leading measure and the lagging outcome should be tested against real data, and revisited as the business evolves. A measure that predicted retention well two years ago may no longer do so if the product or customer base has changed.

It is also worth keeping the final set small. Most organisations need a handful of well-chosen leading indicators, not a comprehensive library of them. The value of a leading measure comes from its predictive power and its ability to direct attention and action. A dashboard of thirty leading indicators produces the same paralysis as a dashboard of thirty lagging ones.

Questions for leaders to consider

As with any strategic decision, the quality of the outcome depends on the quality of the thinking that precedes it. These questions are offered as a starting point for that conversation.

- For each outcome metric on your current scorecard, can you name the two or three leading indicators that most reliably predict it? If not, is the connection

between your daily operations and your results being left to assumption?

• Which of your current measures are genuinely leading indicators, and which are lagging ones dressed up as forward-looking? Revenue targets broken into monthly milestones are still lagging measures — they report on what has already been sold.

• When you look at where your teams are working hardest, are their daily activities connected to measures that predict the outcomes you care about — or are those activities invisible on your current dashboard?

• When did you last trace the causal chain from your leading indicators to your outcome metrics, and verify that the relationship still holds? Markets shift, customer behaviour changes, and a measure that once predicted well may no longer do so.

• If you were building your measurement framework today, using only what you now know about how your customers behave and how your business actually creates value — what would you measure that you are not currently measuring?

> *The most valuable measures in any organisation are rarely the most obvious ones. They sit one or two steps upstream of the outcomes everyone is watching — close enough to the work that teams can act on them, predictive enough that they give leadership genuine early sight of where the business is heading.*

Finding them requires the same rigour and seniority that any other strategic decision deserves. It is not an analytics project. It is a leadership conversation — one that starts with a clear-eyed question about what actually causes the results you are trying to achieve.

Part I — The Strategic Foundation

"Better Than the Competition" Is Not a Strategy

Winning the wrong race is still losing.

Walk into most strategy meetings and you will find a benchmarking deck. Our metric against their metric, dimension by dimension, with helpful colour coding. Red means we are behind. Yellow means we are close. Green means we are ahead. The implicit goal is to turn the reds and yellows into green — to match or exceed the competitor on every dimension and declare victory.

It feels rigorous. It is often the wrong exercise entirely.

The problem is not with knowing what competitors are doing. That is useful context. The problem is with letting competitors determine what you measure and therefore what you optimise for. When your targets are defined as "better than them on dimension X," you have outsourced a fundamental strategic decision — what game we are playing and how we win — to the organisations you are trying to beat. You are letting them set the terms of competition, and then trying to beat them on their own terms.

There is a more useful question. And it starts not with the competitor, but with the customer.

The feature race nobody won

In the late 2000s and early 2010s, smartphone manufacturers engaged in an increasingly intense benchmarking exercise. Each round of devices tried to out-specify the previous one. More megapixels. Faster processors. Larger screens. Longer battery life. Stylus support. Removable storage. The benchmark sheet grew longer and the specifications more impressive with every product cycle.

The first iPhone, by the logic of that benchmark sheet, should not have succeeded. Its camera was inferior. It had no removable battery, no expandable storage, no stylus. On the dimensions everyone in the industry was measuring, it was behind.

It did not matter, because the dimensions everyone in the industry was measuring were not the dimensions that determined customer choice. The question customers

were actually asking — does this device feel intuitive, fast, and genuinely good to use? — was not on the benchmark sheet. It had not been put there because it was harder to quantify, and because no established competitor had made it the basis of their strategy. The benchmark sheet was a map of the existing race. Apple chose not to run it.

The manufacturers who were winning the feature benchmarking contest found themselves with products that were measurably superior on conventional dimensions and commercially struggling against a competitor that had declined to be measured the same way. The metrics were internally coherent and externally irrelevant.

Blockbuster: winning a race that no longer existed

At its peak in the early 2000s, Blockbuster was the undisputed leader of the home video rental industry. It operated over 9,000 stores worldwide, employed more than 60,000 people, and dominated every metric that defined competitive success in its category: store count, geographic coverage, inventory breadth, brand recognition, and revenue. Against any conventional benchmark comparison with its direct competitors it was winning comprehensively.

It was also measuring the wrong things entirely.

Blockbuster's financial model had become structurally dependent on late fees — the charges applied when customers returned DVDs after the rental deadline. At their peak, late fees generated approximately $800 million annually, representing more than 15% of total revenue. This was not a minor revenue stream. It was a load-bearing pillar of the business model.

The critical detail is what that metric represented. Every dollar of late fee revenue was a direct measure of customer frustration. A customer paying a late fee was a customer who had been penalised for an imperfect interaction with the service. Blockbuster's most important revenue metric was, in effect, a measure of how often it was making its customers feel bad. The metric said the business was healthy. The customer relationship said something different.

When Netflix launched its DVD-by-mail subscription model with no late fees, no due dates, and no penalty structure, Blockbuster's internal logic could not process the strategic threat clearly. Eliminating late fees meant destroying more than 15% of revenue overnight. When Blockbuster eventually attempted it in 2005 — five years after Netflix had made the model viable — the move cost the company $200 million in lost revenue and was reversed within a year under shareholder pressure.

But there is a deeper problem than the late fee dependency. Blockbuster was benchmarking against other video rental stores, and by those measures it was the industry leader. What its benchmarks could not show was that the category itself was being made irrelevant. Netflix was not a better video rental store. It was a different answer to a different question — one Blockbuster had never asked.

> *In 2000, Blockbuster's executives were offered the opportunity to acquire Netflix for $50 million. They declined. By 2010, Blockbuster had filed for bankruptcy with over a billion dollars in debt. Netflix's market capitalisation today exceeds $250 billion.*

The question the benchmark never asks

To understand why Blockbuster's benchmarks failed so completely, it helps to understand what they were measuring — and what they were structurally incapable of measuring.

Blockbuster measured category-level competition: how it performed against other companies doing the same thing in the same way. This is what conventional benchmarking does. It tells you where you stand within an existing competitive frame. What it cannot tell you is whether that frame is the right one, whether it is becoming obsolete, or whether the job your customers are hiring you to do is being done better by someone operating entirely outside your category.

This is the insight at the centre of Clayton Christensen's Jobs to Be Done framework, developed through his research at Harvard Business School and published most fully in *Competing Against Luck* (2016). Christensen's central argument is that customers do not choose between competing products in a category. They hire a solution to accomplish a job — a specific kind of progress they are trying to make in a particular circumstance. The competitive set for any product is therefore not other products in the same category. It is every solution a customer might hire to do the same job.

Christensen illustrated this with a study of milkshake sales at a fast food chain. The chain wanted to increase milkshake revenue and initially surveyed customers about taste, thickness, and flavour — the conventional competitive dimensions. The improvements this suggested did not move sales. Then a researcher changed the question. Instead of asking what customers wanted from a milkshake, he observed when and why customers bought one, and asked what they were trying to accomplish.

The answer was unexpected. Most milkshakes were bought in the morning by commuters facing a long, solitary drive. They wanted something that would keep them occupied, stave off hunger, and could be consumed with one hand. Its competitors

were not other milkshakes. They were bananas, bagels, doughnuts, and coffee — all candidates for the same commuter job, each with different strengths and weaknesses. Understanding the job changed the product direction entirely. None of that would have emerged from benchmarking against competing milkshake brands, because the benchmark would have been looking in the wrong direction.

Christensen made the competitive implication explicit: "The competitive field is likely completely different from what you might have imagined." He noted that Facebook competes with cigarettes — because both are hired to accomplish the job of relaxing and taking a short mental break. That is not a category-based comparison. It is a job-based one. And it reveals competitive threats and opportunities that category benchmarking is structurally blind to.

Applied to Blockbuster, the Jobs to Be Done lens asks: what job were customers hiring Blockbuster to do? The answer is not "rent physical DVDs." It is something closer to: convenient, guilt-free, friction-free access to entertainment at home, on their own schedule. Late fees were a direct attack on that job. Mail delivery was a better solution to it. Streaming was a far better solution still. Blockbuster's benchmarks measured how well it was doing the thing it was doing. They said nothing about whether the thing it was doing was still the thing customers most wanted done.

What this means for how you set targets

Targets anchored to beating category competitors optimise for the current race. They do not help you see whether the race is being run on the right track, by the right rules, with the right competitors in the field.

Targets anchored to the job your customers are hiring you to do are different in character. They ask not "are we better than them?" but "are customers choosing us, using us more, recommending us, and staying with us — and do we understand why?" Those are absolute measures of whether you are doing the job well. They remain relevant regardless of which competitors enter your space, because they are grounded in the customer's need rather than the competitor's behaviour.

The most important question to ask of any benchmarking exercise is one that rarely gets asked: is the dimension we are competing on actually what determines our customers' choice? If the answer is yes, benchmark rigorously and set ambitious targets. If the answer is uncertain, the benchmarking exercise may be producing precise measurements of something that does not determine whether you win or lose.

Questions for leaders to consider

• What job are your customers actually hiring your product or service to do? Not the category you operate in — the specific progress your customers are trying to make when they choose you. Can everyone in your leadership team answer that question consistently?

• Who are your real competitors, defined by the job rather than the category? What other solutions might your customer hire to do the same job — including solutions from entirely different industries or categories?

• For each metric on your benchmarking dashboard, ask: does this dimension actually determine whether a customer chooses us? If the answer is uncertain, who decided it was worth benchmarking — and when was that decision last revisited?

• Where in your current business model are you dependent on a metric that is also a measure of customer friction or dissatisfaction? A revenue stream built on customer pain is a structural vulnerability, not a strength.

• If a competitor entered your market by solving your customers' job in a fundamentally different way — not better on your current dimensions, but differently configured around a different understanding of the job — would your current metrics show the threat clearly?

• When your targets were last set, were they derived from what customers need you to do better, or from what competitors are currently doing well? Those two starting points lead to different targets, different investments, and ultimately different businesses.

Blockbuster was not complacent. It was measuring carefully and managing what it measured. The problem was that the things it measured most carefully — store count, inventory coverage, late fee revenue — were a precise map of a competitive landscape that was being made irrelevant. Netflix was not playing a better version of Blockbuster's game. It was asking a different question about what customers needed, and building a business around the answer.

Benchmarking tells you where you stand within an existing frame. The harder and more important question is whether that frame is the right one. That question cannot be answered by looking at competitors. It can only be answered by looking, with genuine rigour, at the job your customers are hiring you to do — and whether you are the best available solution for it.

Part II — Setting Targets

The Trap of "More Is Better": Why Setting Targets Is a Strategic Act

The hardest part of setting targets is not choosing what to maximise. It is choosing what to accept being worse at.

Ask most leadership teams what their targets should be and the answers arrive quickly and confidently: more revenue, higher quality, lower cost, faster delivery, better customer experience, greater market share. The list grows. The ambition sounds right. And somewhere in the process, a critical strategic decision gets quietly avoided.

Because in most businesses, the metrics that matter are not independent. They push against each other. Push speed and you trade off quality. Push acquisition and you trade off retention. Push cost reduction and you trade off capability. Push short-term margin and you trade off long-term brand. Everybody in the room knows this. And almost everybody in the room sets targets as if it were not true — aiming for more across the board, because picking a hierarchy is uncomfortable, and "aim higher on everything" sounds like ambition rather than evasion.

The actual skill — the one that separates competent operators from genuine strategists — is not finding the right number to put against each metric. It is deciding which metrics your business will be excellent at, which it will be adequate at, and which it will deliberately and unapologetically fall behind on. That hierarchy is not a compromise. It is the strategy, made operational.

Why metrics conflict by design

Every business operates with finite resources — capital, time, talent, organisational attention. Allocating more of any of these toward one outcome means allocating less toward another. In engineering, this is treated as a fundamental constraint: you cannot simultaneously optimise a system for speed, cost, and reliability without accepting degradation in at least one. In business strategy, the same constraint applies — but it is far more frequently ignored, because the consequences are slower to appear and easier to attribute to other causes.

The more important point is that trade-offs are not just unavoidable — they are where strategy lives. A business that tries to be excellent at everything ends up with no

coherent reason for a customer to choose it over a competitor that is genuinely excellent at one thing. The choice of which trade-offs to accept, and where to channel the resources freed by accepting them, is the clearest possible expression of what a business actually believes about its customers and its competitive position.

BYD: deliberate trade-off within a product

In 2025, China's BYD overtook Tesla to become the world's largest seller of pure battery-electric vehicles — selling 2.26 million units compared to Tesla's 1.64 million. It is a milestone that would have seemed improbable a decade ago, and it was built substantially on a deliberate decision to be worse at certain things than the competition.

Battery design presents every electric vehicle manufacturer with three competing priorities: cost, charging speed, and energy density. Improving one typically requires accepting degradation in at least one of the others. For its mass-market vehicles, BYD made an explicit hierarchy of choices. Cost and safety were non-negotiable: the company adopted lithium iron phosphate chemistry — cheaper and more thermally stable than the nickel-based battery chemistries used in Tesla's vehicles, but with lower energy density. Charging speed was engineered not to its maximum, but to its most commercially relevant point: BYD's research showed that most drivers care about charge speed up to around 80% of battery capacity, not the full 100%. Energy density was treated as adequate for urban and mid-range driving, compensated for by the innovative blade pack design.

The result was a vehicle that was deliberately inferior to premium competitors on certain measurable dimensions, and superior on the dimensions that determined purchase decisions for the largest segment of the market. For BYD's premium models — the Yangwang U8 and others — the hierarchy shifts entirely: 800-volt architecture, faster charging, higher energy density, higher cost. Different customer, different trade-off, different target hierarchy.

The point is not that BYD found better battery technology. It is that BYD was more rigorous than most competitors about deciding which metrics to excel at for which customers — and was willing to accept being measurably worse on the others.

Ryanair: deliberate trade-off as a business model

Ryanair's story makes the same point at the level of an entire business model — and does so with a directness that is unusual in corporate strategy.

Most airlines compete by trying to be broadly good across a range of dimensions: reasonable fares, reasonable comfort, reasonable punctuality, a loyalty programme, a hub network, some form of in-flight service. Ryanair looked at that set of trade-offs and made a different choice. Its target customer was the price-sensitive traveller who wanted to fly, period — and for whom a meaningfully cheaper ticket would justify accepting degradation across every other dimension of the experience.

Once that customer was defined, the trade-off hierarchy became clear and was pursued without apology. Ryanair flies to secondary and regional airports — which reduces landing fees dramatically and enables rapid 25-30 minute aircraft turnarounds that maximise daily utilisation. The entire fleet is a single aircraft type, the Boeing 737, eliminating the cost and complexity of multiple maintenance and training programmes. Seating density is high. There are no complimentary meals, no assigned seats at the base fare. Ancillary services are unbundled and priced separately, growing from roughly 15% of total revenue in 2010 to over 30% by 2019.

The result is an airline that scores poorly on conventional comfort and convenience metrics, and has never tried to hide this. Michael O'Leary has made the trade-off explicit in terms most executives would never use publicly: the airline exists to get passengers from one place to another at the lowest possible fare, and every operational decision is evaluated against that single priority.

> *Ryanair is now Europe's largest airline by passenger numbers and consistently among the most profitable in the industry — achieved not by being broadly good at everything, but by being unbeatable at one thing and deliberately, systematically mediocre at several others.*

Neither BYD nor Ryanair discovered a way to escape the fundamental trade-offs in their industry. They decided which side of those trade-offs to be on, for which customers — and held that decision consistently across every target they set, every investment they made, and every operational choice that followed.

The trade-off that is hardest to see

The trade-offs between cost and quality, speed and capacity, breadth and focus are at least visible. They show up on dashboards and in strategic reviews, even if leaders avoid making the call explicitly.

The trade-off that causes the most damage is rarely visible at all: the trade-off between now and later.

Almost every metric on a typical corporate dashboard is short-term. This quarter's revenue, this month's customer acquisition cost, this week's pipeline. Long-term metrics — customer lifetime value, brand health, employee capability, technical debt — are harder to measure, slower to move, and consistently deprioritised in favour of the metrics that have a deadline attached. The short-term metric wins not because it is more important, but because it is more urgent, more visible, and more directly connected to someone's performance review.

Consider a software company whose product requires customers to set it up and learn it before they see any value. The sales team, under pressure to close the quarter strongly, cuts the onboarding support process from two weeks to two days to push more deals through faster. Acquisition numbers improve. The quarter looks strong. Six months later, churn rises. Customers who were not properly onboarded never understood the product well enough to find value in it. They cancel. The customer acquisition cost saving was real. The customer lifetime value destruction was larger. But the lifetime value loss appeared two reporting cycles later, attributed to product issues or market conditions, and nobody traced it back to the onboarding decision that caused it.

The practical response is not to ignore short-term metrics. It is to pair each short-term measure explicitly with its long-term counterpart and review both together. Customer acquisition cost alongside customer lifetime value. Quarterly revenue alongside retention cohort data. Cost-per-hire alongside twelve-month performance ratings of those hires. When the pairing is made visible, the trade-off becomes a decision. When only the fast-moving number has a dashboard, the trade-off happens anyway — silently, incrementally, and without anyone making the choice.

Why "more is better" persists

The first reason is that it avoids a hard internal decision. Saying "we will be excellent at these three things and deliberately mediocre at these five" requires picking — and picking creates internal losers. The teams whose metrics get deprioritised will notice, and they will object. Saying "aim higher on everything" skips that conversation. It sounds like ambition and functions as avoidance.

The second reason is that deliberate trade-offs feel like failure in cultures that celebrate all-around excellence. Accepting that the business will not be great at something sounds like giving up, rather than like the focused resource allocation it actually represents.

The third reason is that targets are often set without a genuine operational conversation about what achieving them would require. When numbers come down

from leadership without a real discussion about what would need to be sacrificed to hit them, the result is a long list of stretch goals that generate neither prioritisation nor accountability — only the quiet cynicism that accumulates when ambitious targets are set and then quietly not met.

Questions for leaders to consider

- For each of your top performance metrics, which other metric does it push against? If you cannot name the tension, the trade-off is still happening — you are just not making it consciously.
- If you were forced to rank your metrics in order of non-negotiable importance — first, second, third — what would that ranking be? Does your current resource allocation reflect that ranking, or does it reflect a "more of everything" default that avoids the choice?
- Where in your organisation are teams being held accountable for metrics that are structurally in conflict with each other, without any explicit guidance on which takes priority when they collide?
- For each of your short-term performance metrics, what is its long-term counterpart — the metric that captures what you are trading against it? Are both visible on the same dashboard, reviewed at the same time, by the same people?
- Where has your organisation accepted being deliberately behind on a metric — not by accident or neglect, but by design, because being behind there is what funds being excellent somewhere else? If you cannot name any such trade-off, it is worth asking whether the strategy is as explicit as it should be.
- When your targets were last set, was there a genuine conversation about what achieving them would require giving up? Or were the targets set, and the trade-offs left to surface on their own — later, elsewhere, at greater cost?

BYD did not overtake Tesla by being better at everything. Ryanair did not become Europe's largest airline by being broadly good across all dimensions. Both built dominant positions by being exceptionally clear about which metrics defined their competitive advantage, which metrics they would meet adequately, and which they would accept losing — and by holding that hierarchy consistently across every operational and investment decision they made.

Setting targets is not a numerical exercise. It is a strategic one. The number is the easy part. The hard part — and the part that determines whether the targets actually produce the business you are trying to build — is deciding, explicitly and honestly, what you are willing to be worse at in order to be excellent at what matters most.

Part II — Setting Targets

Metrics Have a Shelf Life

The metric that built your company can be the same metric that ends it.

In 1996, Kodak generated approximately $16 billion in revenue and employed 140,000 people. It dominated the global photography market. Its leadership team tracked the metric that had defined the company for nearly a century: film sales volume and market share in film.

By every analytical standard, it was a good metric. It tied directly to the business model. It was precise, consistently measurable, and every part of the organisation — manufacturing, retail relationships, R&D; investment — was aligned around it. It satisfied the tests that good metrics are supposed to satisfy: it reflected the competitive strategy, it drove operational behaviour, and it had decades of evidence behind it.

It also blinded them.

Kodak invented the digital camera in 1975. An engineer named Steven Sasson built the first working prototype inside Kodak's own laboratories. Leadership saw it, understood what it was, and shelved it — because by their core metric, digital photography was a threat to film sales. Why would you invest in a technology that cannibalised your own scoreboard?

By 2012, Kodak had filed for bankruptcy. The metric they had ridden to dominance was the same metric that prevented them from acting on the disruption that would end them.

This is the part of the metrics conversation that almost nobody talks about. Even a perfectly designed metric has a shelf life. Customer behaviour changes. Technology shifts. Competitors enter from unexpected directions. The metric that captured your strategy three years ago may capture nothing useful today. And unless someone asks the question deliberately, the dashboard will keep loading, the trend lines will keep rendering, and the weekly review will keep proceeding — while the business quietly drifts away from what the numbers are measuring.

Three ways metrics expire

The business moves to a new stage. A startup, a growth-stage company, and a mature incumbent need fundamentally different things from their metrics — even in the same industry, serving the same customers. Early-stage companies that focus too early on profit margin are a classic example. The metric is universally respected and eventually essential. At the earliest stage, however, it is the wrong measure entirely. A company that does not yet know whether it has genuine product-market fit is not in a position to optimise for margin — it is in a position to experiment, test, and learn. Tracking profit margin at that stage causes companies to cut the spending on experimentation they most urgently need, in order to protect a number that does not yet have strategic meaning.

Amazon's approach is instructive. For nearly two decades, Bezos declined to manage the business around the profit metric that Wall Street expected. Instead he pointed to free cash flow per share and customer cohort behaviour — measures that reflected the stage Amazon was actually in: building durable customer relationships and infrastructure that would take years to pay off. The metric that would have satisfied quarterly expectations would have prevented the long-term investment the business required.

The reverse is equally costly. Mature companies that continue tracking only growth metrics — customer acquisition, new market entry, revenue expansion — long after their core market has saturated, tend to over-invest in acquisition that no longer generates sustainable returns. The metric was correct and productive at an earlier stage. Nobody updated it when the stage changed.

Customer behaviour or technology changes underneath you. Blockbuster's late fee revenue and store-level rental counts were exactly the right metrics for a brick-and-mortar video rental business in a world where that was the only available model. Then Netflix arrived with a flat-rate subscription model that eliminated late fees by design, and DVD-by-mail delivery that eliminated the need for a physical store entirely. Blockbuster's metrics were not wrong. They were measuring exactly what they were designed to measure. They were measuring the previous business — and by the time leadership added subscription and digital metrics to the scorecard, Netflix had already established the habits of a generation of customers.

Netflix itself is the more instructive counterpart. The company has evolved its measurement framework several times — from DVD-by-mail metrics focused on subscriber acquisition and disc inventory efficiency, to streaming metrics focused on content hours watched, to engagement and retention metrics, to household penetration and advertising tier performance. Each transition was uncomfortable. Each reflected a genuine shift in what the business was actually doing. The company that

insisted on measuring disc shipment efficiency in 2015 would have been optimising for something the business had already left behind.

The metric definition silently drifts. The dashboard has not changed. The number is still being reported on schedule. But somewhere in the background, the underlying reality the metric was designed to measure has shifted. Customer composition has changed. A platform algorithm has been updated. The definition of a key event was quietly revised in the data pipeline. Year-over-year comparisons are no longer measuring the same thing, and nobody has updated the interpretation.

Social media engagement rate illustrates how this happens at scale. In the early years of social platforms, engagement rate was a meaningful signal of genuine audience interest. Over the following decade, platform algorithms began optimising for engagement signals in ways that inflated certain types of interaction. Content designed to provoke outrage routinely outperformed content designed to inform, because the metric could not distinguish between the two. By the mid-2010s, many organisations were optimising budgets around an engagement rate that no longer reliably predicted the commercial outcomes it was originally used to represent. The metric was still being calculated correctly. Its strategic meaning had silently expired.

Why even good leaders miss this

The first reason is that nothing breaks. There is no error message when a metric becomes obsolete. The dashboard loads cleanly, the numbers populate, the review proceeds. Metric irrelevance is silent — it announces itself only in the gap between what the numbers say and what the business is actually experiencing, and that gap can take years to become undeniable.

The second reason is that the leaders most likely to miss metric expiration are often the most experienced ones — those whose current metrics helped build their success. Kodak's executives were not strategically naive. They were experienced, and their experience pointed in exactly the wrong direction when the environment changed. The very confidence that comes from having built something with a metric makes it harder to let that metric go.

The third reason is structural: the cost of recognising metric expiration is paid immediately, while the benefit is realised later. Replacing a well-established metric means short-term confusion, organisational resistance, and an implicit acknowledgement that the previous story was incomplete. Urgent questions consistently win over important ones, and this is among the most important questions a leadership team rarely gets around to asking.

The discipline: hold the intent, question the metric

The practical response to metric expiration is not to change measures frequently for its own sake. Instability in measurement creates its own problems — organisations cannot learn or improve what they cannot observe consistently over time. The discipline is more specific than that.

The distinction that matters is between strategic intent and the metric used to track it. Strategic intent — the underlying goal the organisation is pursuing — tends to be durable. The specific measure used to track progress toward that intent should be treated as provisional, subject to revision as the world and the business change around it.

Google's advertising business illustrates what this looks like in practice. The intent has been consistent for over two decades: connect users to relevant commercial information at the moment they are looking for it. The metric used to track progress toward that intent has evolved roughly every five to seven years — from click-through rate, to quality score, to conversion rate, to return on ad spend, and more recently toward privacy-preserving aggregated attribution models. Each transition caused internal disruption. Each was necessary because the metric had drifted from the intent as the technology and the market changed around it. The intent never moved. The metric kept moving with the world.

> *Write down the intent behind each metric, separately from the metric itself. The intent rarely needs updating. The metric often does — and the habit of asking the question is what separates organisations that catch the drift from those that discover it too late.*

Questions for leaders to consider

- For each of your most important metrics, can you write down the strategic intent it was designed to serve — in one sentence, separately from the metric itself? When you read that intent alongside the current metric, does the measure still capture what you actually care about?
- Which of your current metrics have been in use the longest? The metrics with the most history in your organisation deserve more scrutiny, not less — because they are most deeply embedded in how decisions get made and hardest to question when the world around them changes.
- Is there a gap between what your performance dashboard says and what you observe in customer behaviour, market position, or competitive pressure? If the dashboard says green and the business feels otherwise, which one do you believe

— and what would it take to investigate the gap rather than trust the number?

• At what stage of its development is your business right now — and were your current metrics designed for that stage, or for a previous one? What would you measure differently if you were designing your measurement framework for the business as it exists today?

• Who in your organisation is responsible for questioning whether your metrics are still the right metrics — as distinct from whether you are hitting them? If the answer is unclear, the question is probably not being asked.

• If the world changes significantly in the next three years — in customer behaviour, technology, regulation, or competitive structure — which of your current metrics would be the first to stop reflecting reality?

Kodak did not fail because it measured poorly. It failed because it measured well — and then kept measuring the same thing long after the world had moved on. The film metric was accurate, operationally aligned, and strategically coherent for the business Kodak had built. What it could not do was show the business Kodak needed to become. No metric can do that indefinitely. The discipline is not to find a measure so good it never needs revisiting. It is to hold the intent behind your measures firmly enough that you are willing to let go of the measure itself when the world requires it.

Part III — How Measures Go Wrong

The Cobra Effect: When Your Metric Becomes the Problem

Good intentions, the wrong measure, and the surprisingly predictable ways things go wrong.

There is a parable that has become a staple of economics classrooms, and it earns its place there.

During British colonial rule in India, the government faced a serious problem: too many venomous cobras in Delhi. The solution seemed straightforward — offer a bounty for every dead cobra brought in. The incentive worked immediately. Large numbers of snakes were killed and presented for payment. The streets appeared safer. The programme looked like a success.

Then officials noticed something unusual. The number of cobras being presented for the bounty was not declining — it was rising. When they investigated, they found that enterprising locals had begun breeding cobras specifically to kill them and collect the reward. The metric was being satisfied with perfect efficiency. The actual problem was getting worse.

When the government shut the programme down, the breeders did what rational people do with a now-worthless inventory: they released the snakes. The cobra population ended up higher than when the programme began.

The German economist Horst Siebert gave this pattern a name in his 2001 book: The Cobra Effect. Whether the original Delhi story is documented history or a parable that hardened into accepted fact matters less than what it illustrates so vividly. A measure introduced to fix a problem had become the problem — not through negligence or bad faith, but through the entirely predictable behaviour of people responding rationally to the incentive in front of them.

Goodhart's Law: the principle underneath

To fully understand why this happens — and why it happens so consistently — it helps to start with a more formally documented observation.

In 1975, the British economist Charles Goodhart was studying monetary policy when he noticed something that applied far beyond central banking. His insight has since become known as Goodhart's Law: when a measure becomes a target, it ceases to be a good measure.

The mechanism is straightforward. When a measure is used purely for observation — to understand what is happening — it reflects reality reasonably well. The moment it becomes a target tied to consequences, people begin optimising for the measure itself rather than the underlying reality it was designed to capture. The measure and the thing it was meant to represent gradually pull apart. The number improves. The reality may not.

Three examples make this concrete. In education, schools evaluated primarily on standardised exam pass rates tend, over time, to shift their teaching toward exam technique rather than deep subject understanding. Pass rates rise. Whether students genuinely understand the material becomes a secondary question, because it is not what is being measured with consequences attached.

In customer service, call centre agents measured on average handle time — how quickly they close each call — learn to end calls efficiently rather than resolve problems thoroughly. Calls get shorter. Customers whose problems were not properly resolved call back, and total call volume increases. The metric moves in the intended direction. The actual outcome moves in the opposite one.

In healthcare, hospitals in the United Kingdom measured against a target requiring patients to be seen within four hours of arriving at emergency departments responded, in some cases, by holding patients in ambulances outside the entrance. The clock did not start until they crossed the threshold. The target was met. The patient's wait was unchanged. The measure had been satisfied without the underlying objective being served at all.

In each case, the people involved were not behaving dishonestly in any straightforward sense. They were responding to the rules of the system they operated within, finding the most efficient path to satisfying the number they were being held accountable for. Goodhart's Law is not a theory about human corruption. It is a theory about how measurement and incentives interact — and how reliably that interaction produces distortion when the two are combined without care.

From distortion to reversal: the Cobra Effect

Goodhart's Law describes a measure losing its reliability as a signal. The Cobra Effect describes something more severe: a measure that does not merely stop reflecting the

goal but actively works against it.

The distinction is worth holding clearly. A school that drills exam technique has drifted from its educational purpose — that is Goodhart's Law at work. But the school has not made its students less educated than they were before the target was introduced. The metric is distorted; the underlying situation is stagnant at worst.

The Cobra Effect crosses a further line. The Delhi bounty programme did not merely fail to reduce cobras — it created more of them. Wells Fargo's account metric did not merely fail to build loyal customer relationships — it systematically destroyed them while appearing, on the dashboard, to do the opposite.

> *The Cobra Effect is what Goodhart's Law looks like when it reaches its most damaging conclusion: not just a measure that has drifted from reality, but one that has inverted the relationship between effort and outcome.*

Wells Fargo: a modern cobra farm

In the mid-2000s, Wells Fargo's leadership concluded that the path to growth ran through cross-selling — getting existing customers to hold more products with the bank. The target became the number of new accounts opened per customer. Bonuses were tied to it. Performance reviews were built around it. Branch managers faced sustained pressure to deliver it.

Employees responded the way people respond to pressure on a number: they hit the number. Over a period stretching from roughly 2002 to 2016, more than 3.5 million accounts were opened that customers had never requested and in many cases never knew existed. Fees were charged on accounts people did not know they had. Credit scores were affected by products people had not applied for.

When the scandal surfaced publicly in 2016, the initial fine from federal regulators was $185 million — at the time, the largest the Consumer Financial Protection Bureau had ever imposed. Total settlements eventually reached $3 billion. CEO John Stumpf resigned, was personally fined $17.5 million, and received a lifetime ban from the banking industry. The bank cycled through two further CEOs before stabilising. Its reputation, once considered among the strongest of the major American banks, has not fully recovered.

The story contains both phenomena described above. In Goodhart's terms, the account metric stopped reflecting what it was meant to measure. But Wells Fargo also crossed into Cobra Effect territory. The bank did not merely fail to grow loyal customer relationships. It actively damaged them — destroying the trust of millions of existing

customers while the dashboard reported that cross-selling was thriving. The metric showed a programme succeeding. The underlying business was being hollowed out.

Why this keeps happening

Both Goodhart's Law and the Cobra Effect are well documented. Most educated leaders have encountered them in some form. And yet the pattern recurs with notable consistency across industries and institutions. Three reasons make it harder to prevent than it appears.

The first is the gap between the metric and the goal going unexamined at the point of design. The British programme did not distinguish between wild cobras killed and farmed cobras killed. Wells Fargo's account metric did not distinguish between accounts customers wanted and accounts they did not. In both cases, the gap between what the metric captured and what the goal required was present from the beginning — but nobody worked through its implications before the incentive was attached.

The second is a persistent underestimation of adaptive behaviour. Wherever a metric controls money, status, or job security, people will find its boundaries faster than its designers expect. This is not cynicism about human nature. It is a straightforward observation about how people engage with the rules of any system they operate within.

The third, and most fundamental, is that the problem definition was imprecise before the metric was ever chosen. The British did not actually want fewer dead cobras. They wanted fewer cobra-related injuries and deaths. Wells Fargo did not actually want more accounts. It wanted more profitable, lasting customer relationships. When the metric is a loose proxy for the real goal, it is the proxy that gets delivered — with precision and at scale — while the real goal recedes.

Designing your way out: the paired metrics principle

Understanding the problem is necessary but not sufficient. The more useful question is: what does a measurement system that resists these dynamics actually look like?

The most robust answer is not to find a perfect single metric — one does not exist. It is to design metrics in pairs that pull against each other, so that satisfying one at the expense of the other becomes immediately visible.

The logic is straightforward. A single metric creates a single target, and a single target creates a single direction in which behaviour can be gamed. Two well-chosen metrics

that point in complementary directions create a corridor: to stay within it, behaviour must satisfy both simultaneously. Gaming one metric at the expense of the other moves you outside the corridor, which shows up plainly in the data.

Consider how this principle applies across different contexts. In manufacturing, Toyota does not measure production output alone. Its system rests on two explicit pillars: Just-in-Time production — produce only what is needed, when needed — and Jidoka — stop and fix problems at source before they travel downstream. Neither is optional; neither can be sacrificed for the other. A line running fast but producing defects fails the quality pillar. A line with perfect quality but negligible output fails the efficiency pillar. Both measures must be satisfied together, which makes gaming either one structurally difficult.

Wells Fargo's failure is instructive in precisely the same terms. The bank measured account openings but not account activation, usage, or customer satisfaction. Had a genuine engagement metric sat alongside the volume metric — one that would have collapsed the moment customers discovered unauthorised accounts — the gap between the metric and the goal would have surfaced far earlier. The second metric would have made the gaming of the first one visible.

The same principle applies broadly. Call handling speed paired with first-call resolution rate. Patent counts paired with commercial value realised. Student exam scores paired with independently assessed comprehension. In each case, the second metric does not replace the first — it constrains it, ensuring that the path to satisfying one cannot run through the destruction of the other.

> *The measure you deploy is not a passive instrument. It is an active instruction — one that will be read carefully, tested for its edges, and optimised for, often in directions you did not plan.*

Questions for leaders to consider

- For each target metric in your organisation, can you state in plain language the real goal it is meant to represent? And can you describe the gap between what it directly measures and what you actually care about?
- What is the most efficient path to satisfying this metric — not the path you intend, but the one requiring the least genuine effort or the most creative interpretation of the rules? Is that path visible anywhere on your current dashboard?
- For each of your high-stakes metrics, what is its paired constraint — the second measure that would make gaming the first one immediately visible? If you do not have one, what would it be?

- Where in your organisation do you see metrics being consistently met while the underlying outcome they were designed to drive remains flat or is deteriorating?
- Do the people in your organisation feel safe surfacing problems with current metrics — or does the pressure to defend green numbers outweigh the willingness to question whether those numbers reflect reality?
- When did you last revisit whether the metrics you are holding people accountable for still accurately represent the goals you are pursuing?
- Returning to Wells Fargo: if the real goal was profitable, lasting customer relationships — what would you have measured instead of, or alongside, number of new accounts opened?

The Cobra Effect and Goodhart's Law are not warnings about bad intentions. The British colonial administration was not careless. Wells Fargo's leadership was not uniquely reckless. The call centre managers who introduced handle-time targets were trying to improve efficiency. Each made a version of the same mistake: they defined the goal imprecisely, chose a metric that captured something adjacent to it, attached high-stakes incentives, and trusted that people would pursue the intended outcome rather than the efficient one.

Designing that instruction well — with a clear definition of the real goal, a paired constraint that makes gaming visible, and an organisation willing to hear uncomfortable truths from its own data — is not a technical task. It is a leadership one.

Part III — How Measures Go Wrong

The Metrics Cave: When Your Numbers Show Shadows of Reality

A perfectly correct metric can still hide more than it reveals.

Plato told a story about prisoners chained in a cave their entire lives, facing a wall, watching shadows cast by objects passing behind them. The shadows were their entire reality. They had named each one. They had theories about how shadows behaved and predicted what would come next. When one prisoner escaped, saw the actual world outside, returned, and told the others that the shadows were not real — they refused to believe him. Some wanted to kill him.

Most metrics are shadows.

Not wrong, exactly. Just incomplete projections of something more complicated. And if you stare at one shadow long enough, you start mistaking it for the whole picture. The danger is not that the number lies. The danger is that it tells a partial truth confidently enough that no one thinks to ask what it is leaving out.

Two kinds of shadow

Before examining how metrics hide reality, it is worth naming two distinct ways they do it — because they require different responses.

The first is the **ambiguous shadow**: a metric that different people define differently, without realising it. Everyone in the room is using the same word and looking at different things. The disagreement that follows is not about reality — it is about definitions that were never made explicit.

The second is the **incomplete shadow**: a metric that is precisely defined and correctly measured, but captures only one dimension of a more complex reality. Add a second measure and the story shifts. Add a third and it shifts again. The first metric was not wrong. It was just not enough.

Both types are common. Both are costly. And both are preventable.

The ambiguous shadow: when the same word means different things

Try this experiment in your own organisation. Find three teams that all report on "active users" — product, marketing, customer success, finance, whoever happens to use the term. Ask each of them, separately, to define exactly what an active user means in their numbers.

You will get different answers. Almost guaranteed.

One team counts anyone who logged in this week. Another counts anyone who completed a core product action in the past month. A third counts anyone with a live subscription, regardless of whether they have opened the product at all. Same label on the dashboard. Three completely different populations. The numbers do not reconcile, and no one is wrong — each team is answering a slightly different question while using identical language.

What this looks like in practice is a leadership meeting where two executives spend an hour debating whether engagement is up or down. Both have data. Both teams trust their data. The disagreement is not about what is happening with customers — it is about a definition nobody made explicit. Senior leadership time is consumed by a problem that does not exist, while the actual question goes unanswered.

The fix is unglamorous but effective. Write down the definition. Where does this number come from? What is included, what is excluded, what is the time window, who is the population being measured? One clear paragraph per metric, accessible to everyone who uses it. Most organisations do not do this, and pay for it repeatedly in conference room hours and misaligned decisions.

The incomplete shadow: when one metric hides the rest of the story

The second and subtler problem is a metric that is precisely defined, correctly measured, and genuinely useful — but incomplete. It shows you one dimension of a more complicated reality, and because it looks authoritative, the other dimensions go unexamined.

Consider a platform or e-commerce business reporting strong growth in Gross Merchandise Volume — the total value of all transactions processed through the platform in a given period. It is a natural headline metric: large, visibly growing, and compelling at scale. But watch what happens to that story as additional metrics are layered in.

Layer 1 — GMV alone. Year 1: $100M. Year 2: $150M. Year 3: $210M. The story: strong, consistent growth. A business scaling well.

Layer 2 — Add Take Rate (the percentage of each transaction the platform retains as revenue). Year 1: 18%, giving $18M revenue. Year 2: 15%, giving $22.5M. Year 3: 12%, giving $25.2M. The story shifts. Revenue is still growing, but the rate of growth is slowing sharply — and the take rate is falling. The platform may be discounting to drive volume, or losing pricing power to competitors. The GMV number alone did not raise the question.

Layer 3 — Add Contribution Margin (revenue minus the variable costs directly associated with each transaction). Year 1: 40%, giving $7.2M contribution profit. Year 2: 28%, giving $6.3M. Year 3: 15%, giving $3.8M. The story shifts again — and now it is concerning. The business is generating more transactions, more GMV, and more revenue. But the profit generated by each unit of revenue is collapsing. The business is, in the most meaningful operational sense, becoming less efficient with every period of apparent growth.

Layer 4 — Add Repeat Purchase Rate (the proportion of customers who made more than one purchase, indicating whether the platform is building genuine loyalty or relying on continuous acquisition). Year 1: 45%. Year 2: 38%. Year 3: 29%. Now the full picture is visible. GMV is growing. Take rate is declining. Margin is deteriorating. And fewer customers are coming back. This is not a growth story. It is a business that is spending its way to volume — acquiring customers at increasing cost, retaining fewer of them, and generating less profit from each transaction. The original headline told none of this.

> *Four metrics. Four different shadows on the wall. None of them false. Each one revealing something the previous one concealed.*

How shadows become beliefs

A single incomplete metric is a manageable problem. The more serious risk is what happens over time when an organisation repeatedly analyses the same incomplete measure and draws consistent conclusions from it.

Consistency feels like truth. If the same analysis produces the same finding quarter after quarter, the finding starts to seem less like a product of the metric and more like a fact about the world. Teams build strategies around it. New hires are onboarded into it. Leaders who question it are seen as contrarian rather than rigorous.

This is the dynamic Plato's returning prisoner encountered. The other prisoners had not simply seen one shadow — they had built an entire system of knowledge around their shadows. Challenging the shadow was not experienced as useful information. It was experienced as a threat to the system.

In organisations, this manifests as strategic decisions made with confidence on the basis of a GMV number, while the contribution margin question was never asked. Or customer satisfaction scores that consistently look acceptable, while the underlying driver of dissatisfaction — never measured — compounds quietly beneath them. The incomplete metric does not just hide one fact. Over time, it shapes what questions get asked, what analyses get commissioned, and what the organisation believes it understands about itself.

Designing for better visibility

The solution is not to add metrics until the dashboard becomes unmanageable. It is to be deliberate about what each metric reveals and what it conceals, and to pair measures that catch what the primary metric misses.

Start with the primary metric and ask one question: what could improve on this measure while the underlying business reality gets worse? That gap — between the metric moving in the right direction and the business moving in the wrong one — is precisely where the complementary metric needs to sit. GMV can rise while margins fall: add contribution margin. Customer count can grow while loyalty declines: add repeat purchase rate. Revenue can increase while customer satisfaction erodes: add a retention or churn measure.

Definitions deserve the same attention as the metrics themselves. Before any number is reported, the definition should be written down in plain language and agreed across every team that uses it. Not stored in a SQL query that only the data team can access — written in a sentence that a new executive could read and understand. This is not a technical task. It is a communication task, and it belongs to leadership.

Finally, metrics should be revisited as the business evolves. A measure that captured the right dimension of the business at an earlier stage may no longer do so. The shadow on the wall shifts when the object casting it changes shape. The discipline is not just to choose measures well at the outset, but to continue questioning them as the business they are meant to represent continues to change.

Questions for leaders to consider

• What would have to be true for your headline metric to look strong while your business was quietly deteriorating? Whatever your answer is — that is the complementary metric you are currently missing.

• Do the teams in your organisation that all use the same metric share a single written definition of it? If not, the next disagreement in a leadership meeting about whether a number is up or down is a definition problem, not a data problem.

• Which of your current metrics have been in use long enough that their conclusions feel like organisational facts rather than outputs of a measurement choice? When did someone last question the measure itself rather than the number it produced?

• Where in your business are decisions being made primarily on a single metric? What are the two or three dimensions of that decision that the single metric does not capture?

• How accessible are your metric definitions to the people who use them? Could a new leader joining your organisation find a clear explanation of what each dashboard number includes, excludes, and was designed to measure?

• When your data tells a consistently comfortable story quarter after quarter, is that because the business is genuinely performing well — or because the metric was designed in a way that makes discomfort invisible?

Plato's prisoner who escaped the cave and returned with news of the real world was not wrong. He was inconvenient. What he had seen could not be reconciled with the system of knowledge the others had built around their shadows — and so he was disbelieved, and the system was preserved.

The metrics on your dashboard are shadows too. Useful, necessary, and incomplete. The discipline is not to distrust them, but to know precisely what they reveal and what they do not — and to build the habit of occasionally asking whether the shadow on the wall still bears a reasonable resemblance to the thing casting it.

Part III — How Measures Go Wrong

The One-Number Trap: When a Composite Score Becomes a Substitute for Thinking

Every clean score is hiding a messy story. The question is whether you know what is inside it.

There is something deeply appealing about a single authoritative number. It fits on a slide. It travels through an organisation without losing anything in translation. It makes comparison easy, progress trackable, and accountability clear. Composite scores — single numbers built from multiple underlying inputs — have become the standard currency of measurement in business, finance, and policy precisely because they promise to compress complexity into something manageable.

They are everywhere. GDP reduces the entire economic output of a nation into a single growth rate that governments live and die by. University rankings collapse the quality of an institution's research, teaching, graduate outcomes, and student experience into a number that determines where families send their children. Employee wellness scores aggregate sleep quality, stress levels, physical activity, and dietary habits into a single index. Brand health scores combine awareness, consideration, preference, and advocacy into one dashboard metric. Social engagement scores fold together likes, shares, comments, reach, and sentiment into a composite that agencies present as evidence of campaign success.

In each case, the appeal is identical: take something genuinely complex, compress it into something quotable, and make decisions accordingly. The compression is useful. The danger is in what gets lost in it — and in the degree to which the assumptions embedded in the construction become invisible once the number acquires authority.

The problem is not that composite scores are wrong. The problem is that they are trusted in a way that their construction does not warrant. Behind every composite score sits a set of choices: what to include, what to exclude, how to measure each component, and how much weight to assign each one. Those choices are rarely visible in the final number. And when a composite score is used to make high-stakes decisions — capital allocation, executive compensation, regulatory compliance, investment mandates — the assumptions buried inside it are doing far more work than most of the people relying on it realise.

ESG ratings offer the most thoroughly researched and consequential illustration of this problem available today. And the story they tell applies to every composite score on every dashboard in every organisation.

What ESG ratings are — and what they claim to measure

Environmental, Social, and Governance ratings are composite scores designed to evaluate a company's performance across three broad dimensions: its environmental impact, its treatment of employees and stakeholders, and the quality of its internal governance structures. They were developed to give investors a single, comparable measure of corporate sustainability — one number that could sit alongside financial metrics and inform capital allocation decisions.

Today they influence decisions across the investment chain. ESG scores are used in index construction, determining which companies appear in funds held by millions of ordinary savers. They are incorporated into executive compensation packages at major corporations. They inform regulatory compliance frameworks in the European Union. A number that once lived at the fringes of responsible investment has become load-bearing infrastructure in global capital markets.

Three agencies, three different questions

The major ESG rating providers — including MSCI, Sustainalytics, and S&P; Global — each approach the problem from a different philosophical starting point.

MSCI uses an industry-relative, risk-management approach. It identifies which ESG issues are most material for a specific industry, then assesses how well a company manages those risks relative to its industry peers. A company is not penalised for operating in a high-carbon industry — it is assessed on how effectively it manages the ESG risks inherent to that sector.

Sustainalytics takes a risk-exposure approach. It measures both a company's exposure to ESG risks and the degree to which those risks remain unmanaged. Notably, Sustainalytics treats a company's failure to disclose data on a material issue as a potential signal of weak management, rather than a neutral data gap.

S&P; Global takes a performance-and-peer-comparison approach, measuring how a company performs on ESG factors relative to peers within the same industry, integrated with traditional financial analysis.

These are not minor methodological variations. They reflect three genuinely different answers to the question: what does it mean to be a good ESG performer? A company

can answer all three questions very differently — and receive three very different scores — without anything being technically wrong.

The numbers that reveal the problem

In 2022, Florian Berg, Julian Kölbel, and Roberto Rigobon at MIT Sloan published a landmark paper titled *Aggregate Confusion: The Divergence of ESG Ratings* in the *Review of Finance*. Analysing data from six prominent ESG rating agencies, the research team found that the average pairwise correlation between agencies' ESG ratings ranged from 0.38 to 0.71, with an average of 0.61.

To put that in context: the correlation between Moody's and S&P; credit ratings for the same set of firms stands at 0.99. Credit ratings achieve near-perfect agreement because they are measuring something with a relatively objective definition — the probability of default on a financial obligation. ESG ratings are measuring something for which no equivalent objective standard exists. The 0.61 average correlation is not evidence that agencies are doing their work carelessly. It is evidence that they are answering fundamentally different questions while using the same label.

The research team decomposed the divergence into three sources. Measurement divergence — where agencies assess the same concept using different indicators — accounts for 56% of the total disagreement. Scope divergence — where agencies assess different things entirely — accounts for 38%. Weight divergence — where agencies agree on what to measure and how, but assign different levels of importance — accounts for just 6%.

The paper also detected what the researchers called a "rater effect": a rater's overall impression of a company influences how they score specific individual categories. The composite is not even fully systematic within a single agency. The general view shapes the specific measurement, which then feeds back into the general score.

Tesla and Chevron: the same company, opposite conclusions

The research findings are made vivid by two well-documented real-world cases.

Tesla is one of the world's most recognised electric vehicle companies — a business whose core product exists specifically to reduce carbon emissions. MSCI assigned Tesla a rating of AA, near the top of its scale. S&P; Global dropped Tesla from its ESG index entirely in 2022, citing concerns about workplace conditions, racial discrimination claims, and governance structure. The same company, at the same point in time, was simultaneously near the top of one agency's scale and excluded from another's index.

The explanation reveals the methodology at work. MSCI's industry-relative approach weighted Tesla's environmental contribution heavily. S&P;'s approach weighted social and governance factors alongside environmental ones, and found Tesla's performance on those dimensions significantly weaker. Both agencies were measuring real things about the same company. They were measuring different things and weighting them differently, and arrived at opposite conclusions.

The oil company Chevron provides an equally stark example. At the same point in time, Sustainalytics rated Chevron as high ESG risk — placing it in its second-worst category out of five. MSCI rated Chevron A, placing it in its third-best category out of seven. The same company, the same period, opposite ends of the quality spectrum from two agencies both claiming to measure ESG performance.

> *These are not edge cases or anomalies. They are illustrations of what a 0.61 average correlation looks like when applied to specific companies. The number on the dashboard has the appearance of a fact. The research shows it is closer to an opinion.*

Three layers of the composite problem

The ESG case illustrates what every composite score does, in three layers, to the information it claims to represent.

The first layer is the **construction layer**. Every composite score is built from a set of choices about what to include, how to define each component, and how to combine them. Those choices reflect the builder's view of what matters — which is a legitimate and often thoughtful view, but a view nonetheless. When MSCI and Sustainalytics reach opposite conclusions about Tesla, neither is wrong in its own terms. They built different things and called them by the same name.

The second layer is the **assumption layer**. Inside the construction are assumptions — about which proxies adequately represent underlying concepts, about how to handle missing data, about which industries face which risks, about how much weight each factor deserves. These assumptions are not visible in the final score. Most of the people using the score have never examined them, do not know where to find them, and would struggle to evaluate them if they did.

The third layer is the **authority layer**. A composite score, once published by a credible agency and adopted into an investment process or a governance framework, acquires institutional legitimacy. It stops being treated as a structured opinion and starts being treated as an objective measurement. Capital moves on the basis of it. Executive pay is linked to it. Regulatory frameworks are built around it. The authority of the number

outruns the evidence for its construction — and the gap between the two is where the risk lives.

What this means beyond ESG

The composite score problem is not unique to ESG ratings. It is the standard form of any attempt to compress multi-dimensional reality into a single number — customer health scores in SaaS businesses, employee engagement indices, brand equity scores, credit ratings, innovation indices, productivity scores.

What the ESG case provides is an unusually well-researched, unusually consequential, and unusually honest window into what every composite score is actually doing. Most organisations using composite metrics in daily decision-making have neither the underlying methodologies nor the analytical tools to decompose the divergence. They have the number, and the authority it has accumulated, and the assumption that it is measuring something real.

The discipline required is not to abandon composite metrics. They serve a genuine purpose — they roll up complexity into something actionable, which leadership dashboards genuinely need. The discipline is to treat them as structured opinions rather than objective facts, to know what is inside them, and to ask regularly whether the assumptions embedded in their construction still reflect the reality you are trying to manage.

Questions for leaders to consider

- For each composite metric your organisation relies on, can you name the underlying inputs it is built from, how those inputs are weighted, and what gets excluded? If you cannot, the methodology is invisible — and invisible methodologies carry invisible risks.

- When a composite score moves, do you know which underlying input drove the change? A customer health score dropping from 75 to 68 could mean very different things depending on whether the movement came from usage data, support ticket volume, or a change in contract value. The composite conceals the cause.

- If two colleagues used two different but reputable sources to assess the same thing and reached substantially different conclusions, would your organisation have a way to understand why? Or would the disagreement simply produce confusion?

• Are the composite scores in your organisation ever revisited for methodology drift? Inputs change, proxies get redefined, data sources shift, and survey questions get reworded. A score that was validly constructed three years ago may not be measuring the same thing today, even if it carries the same name.

• Where in your organisation are high-stakes decisions — compensation, investment, supplier selection, strategic prioritisation — being made on the basis of a composite score whose construction has never been examined by the people making those decisions?

• What would the decision look like if the composite score were decomposed back into its underlying components? Would the same conclusion follow? If the answer is uncertain, the composite is doing work it should not be trusted to do alone.

A composite score is a convenience and a compression. It makes the complex manageable, and that is genuinely useful. What it is not is a substitute for understanding what you are actually measuring. The ESG case shows, with unusual precision and at unusual scale, what happens when a composite score is treated as ground truth: agencies measuring the same company reach opposite conclusions, investors make decisions based on numbers that agree with each other only 61% of the time, and the authority of the score outruns the evidence for its construction.

The same dynamic operates, at smaller scale and with less visibility, in every organisation that trusts a composite number without examining what is inside it. The number is not the reality. It is a structured opinion about the reality — and the structure deserves as much scrutiny as the opinion.

Part III — How Measures Go Wrong

The Method Is the Message: How the Way You Measure Shapes What You Find

The tool you use to collect a metric is not neutral. It is already making choices about what you will and will not be able to see.

There is a category of measurement failure that is harder to spot than a wrong number or a poorly chosen metric. It is the failure that lives not in the data itself, but in the assumptions built into the method used to collect it. The data is accurate. The analysis is technically correct. And the conclusion is still wrong — because the instrument had a blind spot that nobody examined.

This kind of failure is particularly dangerous precisely because it does not look like a failure. The numbers add up. The dashboard is populated. The reports go out on time. What nobody questions is whether the method of measurement was capturing the right thing, from the right population, in the right unit, at the right moment.

Four cases illustrate how differently this problem can manifest — and how consistently it goes unnoticed until the consequences have already arrived.

When two systems speak different languages

In 1999, NASA lost the Mars Climate Orbiter — a $327 million spacecraft — nine months into its journey to Mars. The investigation found no software malfunction, no hardware failure, and no navigational error in the conventional sense. The cause was simpler and more embarrassing: the engineering team at Lockheed Martin had expressed thruster force in pound-seconds, the imperial unit. NASA's navigation system read that data assuming it was in newton-seconds, the metric equivalent. One pound of force equals 4.45 newtons. Every trajectory correction over nine months was therefore off by a factor of 4.45. When the spacecraft reached Mars, it entered the atmosphere at the wrong altitude and disintegrated.

Two teams. Both measuring correctly. Both recording data accurately. The catastrophic failure came entirely from an assumption — that the two systems were using the same unit — that nobody had verified.

This is the most fundamental form of methodological failure: not a flaw in the measurement itself, but an unexamined inconsistency in the language the measurement was expressed in. It requires no negligence, no incompetence, and no bad data to produce a catastrophic outcome. It only requires two parties to assume they are speaking the same language when they are not.

When the sample excludes the people who matter most

In 1948, three major polling organisations — Gallup, Roper, and Crossley — predicted that Thomas Dewey would defeat incumbent Harry Truman in the US presidential election. The polls were large, the methodology appeared rigorous, and the prediction was confident enough that the Chicago Tribune printed "Dewey Defeats Truman" on its front page before the results came in.

Truman won the Electoral College 303 to 189.

The error was methodological, not mathematical. Telephone surveys in 1948 reached households that owned telephones — which in that era meant predominantly wealthier, urban, Republican-leaning households. Working-class voters, who formed Truman's base, were systematically absent from the sample. The instrument worked perfectly. The population it was applied to was not the population that decided the election.

The lesson generalises immediately to business. A customer satisfaction survey sent only to users who completed a purchase excludes everyone who abandoned. An employee engagement survey excludes the people who already left. A product review dataset excludes the customers who never bothered to leave one. In each case the data accurately reflects a real population. The question is whether that population is the one you actually needed to understand.

When the data you cannot see matters more than the data you can

During the Second World War, military analysts mapped bullet hole patterns on returning bombers and recommended reinforcing the most damaged areas. The reasoning seemed sound: put armour where the planes are getting hit.

Statistician Abraham Wald saw the flaw immediately. The planes they were studying were the ones that came back. The areas with the most bullet holes were precisely the areas a plane could sustain damage and still survive. The areas showing no damage on returning planes were where the planes that never returned had been hit. The correct recommendation was the opposite of the intuitive one: reinforce where there

are no bullet holes.

The entire analysis was based on a dataset that was structurally incomplete in a way that was invisible from inside the data. The missing planes were not in the room. Their absence looked like nothing — like empty space — rather than like information. But the empty space was where the most important signal was hiding.

In business this pattern appears wherever data is collected only from the visible population. Strategies built by studying only successful product launches, without examining the ones that failed. Churn models built only on data from retained customers, without understanding those who left. Market analysis based on companies currently operating, without accounting for those that no longer exist. In each case the methodology produces a dataset that is internally consistent and externally misleading.

When the unit of measurement is not what you think it is

Consider a consumer brand sold across two major e-commerce platforms. Each platform is managed by a separate team, and each team uses the data provided by their platform to track new buyers of the brand. Leadership sees new buyer growth as a critical indicator of brand health, so both teams report their numbers monthly, the figures are combined, and the total is presented as the brand's new buyer count.

For several consecutive periods, new buyer numbers grow consistently. The business celebrates. The metric is moving in the right direction.

Then someone notices a problem. Despite steady new buyer growth, total brand revenue is growing far more slowly than the new buyer figures would suggest. The investigation reveals the issue. Platform A counts any customer who makes their first purchase on Platform A as a new buyer. Platform B counts any customer who makes their first purchase on Platform B as a new buyer. Neither platform has visibility into the other. A consumer who buys on Platform A in January and Platform B in March is counted as a new buyer twice — once by each team. The unit being measured was a platform transaction, not a person. And because the brand's goal was to grow its customer base — people, not transactions — the metric was answering a different question from the one leadership thought they were asking.

There are two further layers worth noting. First, the overlap itself is strategically interesting information that the current method makes invisible. Customers who buy across both platforms are likely the brand's most loyal and engaged buyers — exactly the population leadership should most want to understand. The methodology was not only producing a wrong total number. It was actively hiding the most valuable segment within that number.

Second, this failure required no negligence from either team. Both teams were measuring accurately and reporting honestly within the scope of their own data. The problem was architectural — built into the measurement design before either team collected a single data point.

The assumption that goes unasked

All four cases share the same structural root. In each one, the measurement method carried an assumption that was never examined: that two teams were using the same unit; that the surveyed population represented the deciding population; that the visible data represented the complete data; that the unit being counted corresponded to the concept being managed.

These are not exotic failure modes. They are the ordinary assumptions that measurement systems run on — invisible precisely because they seem too basic to question. The NASA engineers did not think to ask whether both systems were using the same units because, in a project of that technical sophistication, the answer seemed obvious. The polling organisations did not think to ask who owned telephones because telephone surveys were the standard method. The e-commerce teams did not think to ask whether a platform account and a person were the same unit because the platform data presented them as equivalent.

> *The discipline required is not technical. It is a habit of interrogating the method itself, not just the numbers it produces. The question worth asking is not only "what does this number say?" but "what did the method assume in order to produce it — and have we verified that assumption?"*

Questions for leaders to consider

- For each of your important metrics, can you describe precisely what unit is being counted? Is that unit the same as the concept you are actually trying to manage — or is it a proxy that seemed equivalent when the measurement was designed?
- Where in your organisation do multiple teams report on the same underlying concept using different data sources? Has anyone verified that those sources are measuring the same population in the same unit, or is the consistency assumed?
- Which populations are systematically absent from your measurement processes? Customers who churned before completing a survey. Users who never activated. Markets where you have no presence yet. The absence of data from these groups is itself a signal worth examining.

- When a metric grows but the underlying business outcome it is meant to represent does not keep pace, what is your organisation's first instinct — to question the business, or to question the method?
- When did you last ask not "what do our numbers show?" but "what did our measurement method assume?" — and verified that those assumptions still hold for the business you are running today?

The four cases in this article involved no dishonesty, no negligence, and no analytical incompetence. They involved smart, capable people working diligently with data that was, in every conventional sense, correct. What each case lacked was a habit — the habit of asking not just what the measurement shows, but what the method of measurement was built to see, and what it was therefore structurally incapable of showing.

Data does not arrive neutral. It arrives shaped by the choices made before it was collected — about who to ask, what unit to count, which population to observe, and what to assume about the systems involved. Those choices are invisible in the final number. Making them visible is not a technical task. It is a leadership one.

What This Chapter Asks You to Hold

Measurement is not a technical problem. It is a leadership one — and it does not end when the dashboard is built, the targets are set, or the quarterly review is complete. It is an ongoing act of judgment about what matters, how to track it honestly, and when to let go of a measure that once served you but no longer does.

The sections in this chapter have approached that judgment from different angles. Some are about choosing the right measure in the first place. Some are about the ways even well-chosen measures quietly break down — through gaming, through incompleteness, through methodological blindness, through the slow drift of a world that moves while the dashboard stays still. Taken together, they point toward the same discipline: treat your measures not as furniture, but as active strategic decisions that deserve the same scrutiny you give to any other choice about how to run the business.

Before moving on, three questions worth sitting with:

What is your most important metric doing that you did not intend?

Every measure shapes behaviour beyond its stated purpose. The gap between what you intended and what the metric actually produces in your organisation is where the most important work often lives.

What does your current dashboard make invisible?

Not what it shows — what it structurally cannot show, because of how it was built, who was included in the data, or what questions nobody thought to ask when the measures were first chosen.

If the world changes significantly in the next three years, which of your current measures will be the first to stop being true?

That metric deserves your attention now, not when it has already misled you.