

EARNED · CHAPTER NINE

# The People Who Actually Keep It

*Chapter Nine of Earned*

Ammanuel Santa Anna

Santa Anna, A. (2026). The People Who Actually Keep It. In Earned (ch. 9). Luminous Prosperity Inc. https://doi.org/10.5281/zenodo.23989340
Luminous Library Licence 1.0 · business.hautelumiere.life/library/earned

## THE LICENCE

The People Who Actually Keep It
*Chapter Nine of Earned*

Earned · Chapter Nine. The Haute Lumière edition, 2026. © 2026 Ammanuel Santa Anna.

Released under the Luminous Library Licence 1.0 (hautelumiere.life/library-licence): free to read, share, quote, translate and teach, non-commercially, for ever. The Author keeps every other right, including every commercial right.

Published by Luminous Prosperity Inc. for Haute Lumière. Author ORCID 0009-0008-3244-0306.

DOI 10.5281/zenodo.23989340 · this record's page: business.hautelumiere.life/library/earned/10-the-people-who-actually-keep-it/

Please cite as: Santa Anna, A. (2026). The People Who Actually Keep It. In Earned (ch. 9). Luminous Prosperity Inc. https://doi.org/10.5281/zenodo.23989340

Chapter Nine of Earned. The whole book, with every chapter, is at business.hautelumiere.life/library/earned (DOI https://doi.org/10.5281/zenodo.23989328).

## A NOTE BEFORE YOU READ

This work is offered for education and leadership development. It is not legal, financial, accounting, investment, medical or psychological advice; take the advice of a qualified professional who knows your circumstances. Companies and events are described from the public record as the author understands it; characterisations are the author's analysis. Examples are illustrative, and no particular outcome is promised to any reader.

## ABSTRACT

The Ritz-Carlton's per-employee discretionary authority — famously in the low thousands of dollars per guest, per incident, without approval — gets told as a hospitality anecdote, a charming story about a bellman who flew a forgotten stuffed giraffe to a child. Told that way it is useless to you, because it sounds like a story about extraordinary people, and you cannot hire your way to extraordinary at scale. Three things are true about that authority simultaneously, and the anecdote versions keep only the first. One: the number is large enough to resolve essentially any single guest problem the hotel can create — a ruined dress, a missed flight connection, a wrong room on an anniversary. It was not set by benchmarking what competitors allowed. It was set by asking what size of failure the property can produce and then putting the authority above that line rather than below it.

CHAPTER NINE

# The People Who Actually Keep It



*The People Who Actually Keep It*

HAUTE LUMIÈRE · EARNED · CHAPTER NINE

## What Ritz-Carlton Actually Built

The Ritz-Carlton's per-employee discretionary authority — famously in the low thousands of dollars per guest, per incident, without approval — gets told as a hospitality anecdote, a charming story about a bellman who flew a forgotten stuffed giraffe to a child. Told that way it is useless to you, because it sounds like a story about extraordinary people, and you cannot hire your way to extraordinary at scale.

Read it as operating design instead and it becomes something you can build.

Three things are true about that authority simultaneously, and the anecdote versions keep only the first. One: the number is large enough to resolve essentially any single guest problem the hotel can create — a ruined dress, a missed flight connection, a wrong room on an anniversary. It was not set by benchmarking what competitors allowed. It was set by asking what size of failure the property can produce and then putting the authority above that line rather than below it. Most companies do the opposite: they set the discretion limit low enough that the worst cases *must* escalate, which means the worst cases — precisely the ones where trust is decided — are systematically the ones the frontline cannot fix.

Two: it is per incident, not per year. There is no budget the employee is depleting, no internal scoreboard that makes generosity in March a reason for caution in November. The moment you make discretion a pool, you have installed a mechanism that teaches employees to hoard it, and hoarding shows up to the customer as arbitrary refusal.

Three — and this is the load-bearing one — it is paired with an obligation. The employee who uses the authority is required to log the incident and its cause. The spend is not the point. The spend buys the organisation an accurate map of where it is failing, purchased from the only people who can see it. A company that grants discretion without capturing what it was spent on has bought expensive silence: it has fixed the customers one at a time and learned nothing about why they kept needing fixing.

The mechanism, then, is not generosity. It is a closed loop: authority sufficient to resolve, an obligation to record, and a process that reads the record. That loop is buildable at a software company, a bank, a logistics firm, a clinic. The dollar figure is the least transferable part of it.

## The Quota That Manufactured Fraud

Now the inverse, at scale, and the most instructive case in modern American business precisely because nobody involved set out to do anything wrong.

Wells Fargo spent years driving a cross-selling target — the ambition of eight products per household, promoted publicly as a strategic metric — down through a retail banking organisation of a couple of hundred thousand people. Branch employees were measured on daily product-opening counts. Missing the number had consequences that arrived weekly, in a conversation with a manager who was themselves being meas-

ured on the same number, by a district manager also being measured on it. The target was specific, individual, frequent, and enforced.

What followed is well documented and ended in a 2016 enforcement action, subsequent litigation, congressional hearings, the departure of a CEO, and — the part that should hold your attention — an unprecedented Federal Reserve asset cap imposed in early 2018 that prevented the bank from growing past roughly its then-current size. That cap remained in place for years, into the 2020s. A retail sales metric produced a regulatory constraint on the balance sheet of one of the largest banks in the country.

The moral usually drawn is about incentives, and it is correct but too shallow to act on. Here is the sharper version: millions of unauthorised accounts were not opened by criminals. They were opened by thousands of ordinary people — tellers, personal bankers, branch managers — the overwhelming majority of whom, on any personality assessment you care to run, would have scored as honest. The metric did not select for dishonest people. It manufactured dishonest behaviour out of honest ones, by making the honest path lead reliably to job loss.

That is the mechanism you need to understand, because it generalises far past banking. When a measured target is achievable only by a fraction of people through legitimate means, and unachievable performance is punished, the metric is not measuring behaviour. It is *specifying* behaviour, including behaviour nobody wrote down. Everyone in the chain can honestly report that they never told anyone to falsify anything. They didn't have to. The number did the telling, a million times a day, with more authority than any values statement ever accumulates.

The conditions this requires are worth naming precisely, because a metric without them is merely a metric. You need: individual attribution, so the pressure lands on a person rather than a team; high frequency, so there is no long horizon in which to recover; a punishment asymmetry, where missing costs more than exceeding gains; and — critically — an achievability gap, where the target sits above what the honest path reliably produces. Remove any one of those and the metric mostly behaves. Assemble all four and you have built a machine that converts your own employees into the instrument of your trust destruction, and it will run without further instruction from anyone.

## Metric Autopsy

Every frontline metric is an instruction to a human being about how to treat another human being. Read yours that way and most of them become legible immediately.

Average handle time. The instruction is: end this conversation. Not resolve it — end it. AHT cannot distinguish between a call closed because the problem was solved and a call closed because the customer gave up, and the second is faster. It penalises the agent who slows down for the confused elderly caller and re-

wards the one who reads the deflection script. It makes the single most valuable frontline act — noticing that the customer's stated problem is not their actual problem — a personal cost.

First-contact resolution. Better, and it is genuinely better, but it has a specific pathology: it rewards the *appearance* of closure. Where FCR is measured by whether the customer called back within some window, it instructs the agent to make the customer not call back — which is achievable by solving the problem and also achievable by convincing them that nothing can be done, or by leaving them so worn down they take the loss privately. It is a good metric that quietly permits a bad one.

Attach rate. The instruction is: this person's transaction is incomplete until you have added something to it. At a modest level it does no harm. Where it becomes the primary measure, it converts every interaction into a negotiation the customer did not know they had entered, and it does its worst damage on precisely the customers least able to refuse — the elderly, the distressed, the ones who came in already needing help and have therefore already signalled that they are unlikely to push back.

Save rate. The most corrosive of the four, and the most beloved by boards. The instruction is: this person has told you they want to leave and your job is to prevent it. Everything that follows is structurally adversarial. The retention agent who is measured on saves cannot be an honest counterparty in the conversation, because the customer's clearly expressed preference is the thing the agent is paid to defeat. It generates the hold-time maze, the offer that was available all along and was withheld from loyal customers who never threatened to leave, the cancellation that requires a phone call in a company that will happily take your money through an API. Every one of those is a trust withdrawal, and the save-rate dashboard is green throughout.

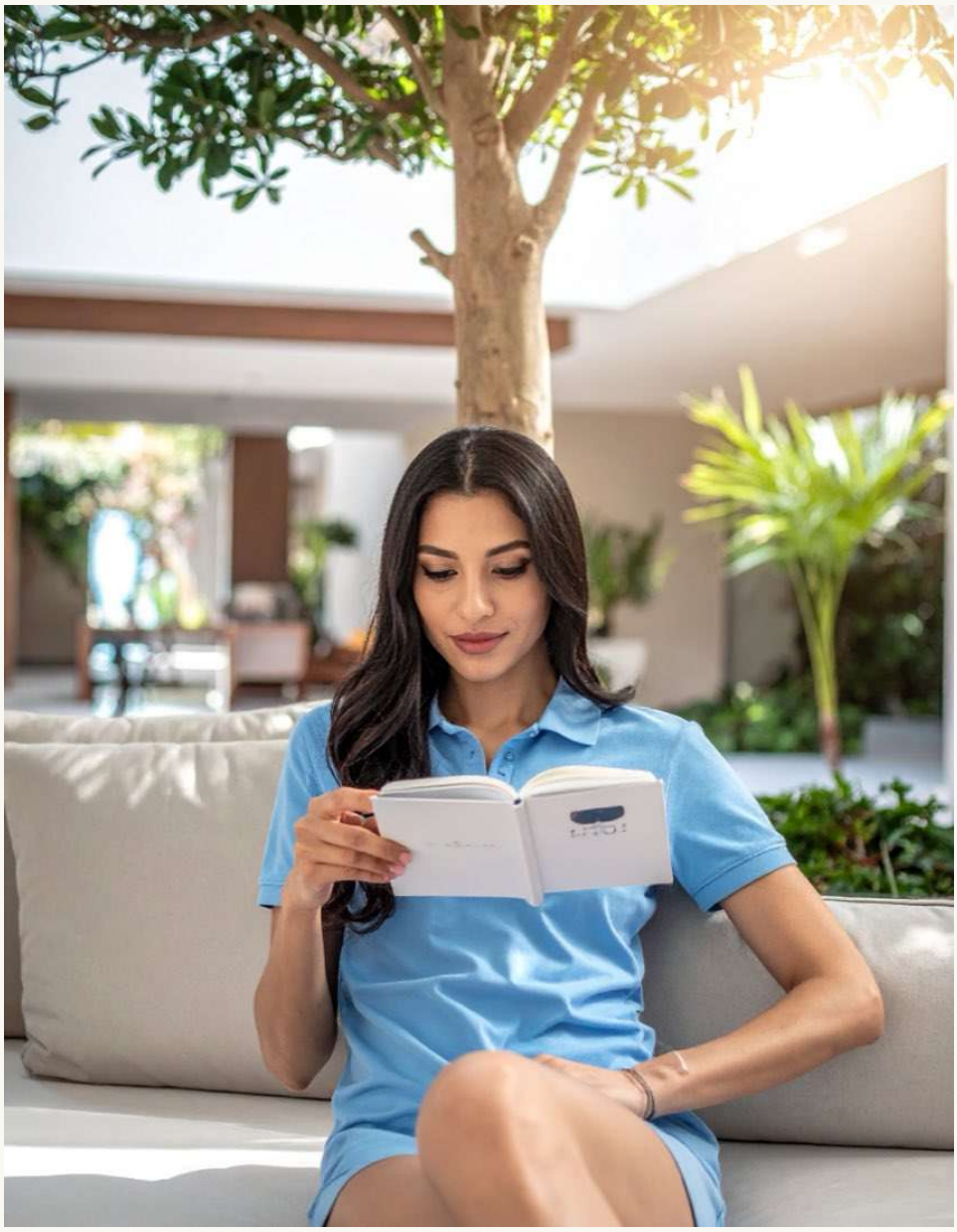
None of these metrics is unusable. All of them are dangerous alone. The diagnostic is simple: for each frontline metric, write the one-sentence instruction it gives to a person under pressure with a customer in front of them and no time. If you cannot say that sentence out loud to a customer without embarrassment, you have found something.

## When the Promise Is Kept by Someone Else's Employee

Outsource the contact centre and you have not outsourced the promise; you have outsourced the *keeping* of it while retaining every consequence of its being broken. The structural problem is not vendor quality. Plenty of BPO agents are excellent. The problem is that the contract is the compensation system, and most such contracts pay per call handled, per minute, or per seat — which means the vendor's revenue and the vendor's ability to deliver on your promise are related inversely. You have created a party whose economics improve when conversations are shorter and whose penalties for a bad resolution are almost always weaker than yours.

Then layer the discretion question on top. The vendor's agent typically has a smaller discretion budget than your own employees, because you were not willing to hand a third party your chequebook — an entirely reasonable instinct with an unreasonable consequence. You have concentrated the least authority in the population most likely to encounter the customer. Your promise is being kept, in the majority of cases, by the people you gave the least power to keep it.

There are only two honest responses. Either pay for outcomes rather than volume — resolution quality, verified problem closure, customer-reported outcomes weeks later — and price that properly, which means paying more per interaction and accepting it, or bring the interactions where trust is actually decided back in-house and outsource only the ones where it isn't. Password resets can live anywhere. The conversation about a denied claim cannot.



*Tenure*

HAUTE LUMIÈRE · EARNED · CHAPTER NINE

## Tenure

There is a version of this argument that sounds sentimental and isn't. Every departure from your front line ends a relationship that had accumulated context, and context is the substrate that discretion runs on.

The mechanism is specific. An experienced agent knows which of the company's rules are load-bearing and which are residue; knows which failures are systemic this month and can therefore recognise the fourth caller with the same problem as a pattern rather than an anomaly; knows the person in fulfilment who can actually fix the thing; knows, from three hundred prior conversations, the difference between a customer

working an angle and a customer genuinely wronged. None of that is in the knowledge base. All of it is what makes discretion safe to grant. A company with 90 percent annual frontline turnover cannot responsibly extend a large discretion budget, because it does not have a population capable of exercising judgment — and so it writes tighter rules, which makes the job worse, which raises turnover, which justifies tighter rules. The doom loop is not cultural. It is mechanical, and turnover is the variable driving it.

Cost the loop honestly and it usually prices itself out. Add the recruiting spend, the six-to-twelve weeks to competence, the error rate of the newly trained, the supervisor time consumed by escalations that experience would have absorbed, and the trust cost of the customer who has explained their situation to the fourth new person this year. Against that, a compensation increase that halves attrition is frequently cheaper in the first year, and it is the only intervention that makes a larger discretion budget defensible. Tenure is not a benefit. It is the infrastructure the whole chapter depends on.

## The Doctrine Case

Here is the moment that decides everything, and every company gets one within about ninety days of granting real authority.

An employee spends the company's money to protect a customer, and they are wrong. Not wrong in intent — wrong in fact. The customer was working an angle, or the situation did not warrant the size of the remedy, or the same outcome was available for a quarter of the cost, or the employee misread a policy that actually did cover it. Somewhere between four hundred and four thousand dollars has left the building for a reason that will not survive review.

What happens to that person is your real policy on trust, and it will be known throughout the frontline organisation within a week. Not through a memo. Through the fastest information system your company has ever operated, the one that carries the story of what happened to a specific named colleague, and which no internal communications function has ever successfully competed with.

If they are reprimanded, coached, put on a plan, or even merely made to explain themselves to three levels of management, every person who hears it draws the correct inference: the authority is real on paper and expensive in practice, and the safe move is to escalate. Your discretion budget is now formally intact and functionally zero. You will not find this in any report.

The alternative is not to pretend the error didn't happen. It is to separate two questions that companies habitually collapse: *was the judgment defensible given what this person knew at the time*, and *was the outcome good*. Discipline attaches only to the first. An employee who reasoned from the principle, in good faith, on the information available, and got a bad outcome, is not corrected — they are backed publicly, and the case becomes teaching material with the reasoning made visible. What gets corrected is the information gap that caused the misread. This is the same doctrine that the better safety cultures run on, and for the same reas-

on: you cannot get honest reporting from people who are punished for outcomes they did not control, and without honest reporting you are blind.

Handled that way, one such case does more for your trust position than any training programme, because it answers the only question the front line was actually asking.

## Where This Inverts

Now the edge, because everything above becomes destructive past a specific point.

Discretion without a principle produces arbitrary generosity, and arbitrary generosity is not trust — it is luck, and customers correctly perceive it as such. When one caller gets a full refund and an apology and the next caller with an identical problem gets a policy citation, the company has not been generous. It has been random. And randomness is corrosive in a way that consistent stinginess is not: a customer facing a consistent, clearly stated, unfavourable policy can plan around it and can decide whether to keep doing business with you. A customer facing a lottery cannot, and their rational response is to game the system — to call back repeatedly until they reach a generous agent, to escalate performatively, to complain publicly because publicity is the only reliable lever. You will then observe rising contact volume and social complaints and conclude that customers have become entitled. They have become adaptive. You built the machine that taught them.

Worse, arbitrary discretion is systematically unequal. When outcomes depend on who you reach and how you present, they will correlate with confidence, articulacy, accent, persistence, and free time — which is to say the customers with the least capacity to advocate reliably get the worst outcomes, at scale, from a policy you would describe as compassionate. That is a fairness problem serious enough to end a chapter on, and it comes from the same source as the good version.

The fix is that discretion must be bounded by a stated principle rather than by a stated amount — or rather, by both, with the principle doing the real work. "Up to $500" is a limit. "Make the customer whole for anything we caused, without requiring proof from them, up to $500 — and if it exceeds $500, you may still commit and we will settle it after" is a principle with a limit attached. The first produces a hundred different behaviours across a hundred agents. The second produces one behaviour with a hundred implementations, which is what consistency actually looks like at the edge. The principle is what makes the discretion legible to a customer as *policy*, and it is legibility — not amount — that converts a generous act into a trust deposit. A remedy the customer understands as something the company does deposits into the stock; the identical remedy experienced as a lucky break deposits nothing, because it predicts nothing about next time.

## The Line, Not the Values

Which brings the argument to its actual conclusion, and it is not the one most leadership teams want.

The front line does not need to be told the values. They have been told the values. They can recite the values; the values are on a card in their badge holder and on the wall behind the coffee machine, and they have been through the workshop. Telling them again changes nothing, and the reason it changes nothing is structural rather than attitudinal: a value is a statement about preference, and the employee is not operating in a preference environment. They are operating in an authority environment and an incentive environment. What they need to be told is what they may spend without asking, and what they will not be punished for.

Culture sits downstream of those two facts. Change the authority line or change the compensation metric and the stated values become decorative within a single quarter — not because anyone stopped believing them, but because belief was never the mechanism. This is why values programmes fail with such consistency and why permission changes work with such speed. A values programme attempts to alter behaviour by altering conviction, in people whose conviction was mostly fine already and who are constrained by something else entirely. A permission change alters the constraint. The behaviour moves the following Monday, and the conviction, which never went anywhere, is suddenly visible in the numbers, and everyone congratulates the culture.

You can run this in reverse as a diagnostic on your competitors. When a company's frontline behaviour is visibly worse than its stated values, do not conclude hypocrisy at the top. Look for the metric. It is nearly always there, and it is nearly always something someone installed for a defensible reason two or three years ago and has not re-read since.

## The Practice

Go and ask five people. Not managers — five people who talk to customers all day, chosen at random rather than nominated, and ideally not in a room with their supervisor. Ask each of them a single question: what are you allowed to do for a customer without asking permission?

Then listen for whether the five answers match.

They almost certainly will not. You will get a dollar figure from one and no figure from another; you will hear one person describe an exception they make routinely and another describe the same exception as forbidden; you will hear at least one person say, in some form, that it depends on the manager on shift. Each of those answers is honest, and together they tell you something specific: your promise is currently being kept by improvisation. Which means it is being kept unevenly today, and it does not survive the departure of whoever is currently improvising well.

The work from there is not a training programme. Write the number down — the actual dollar amount, per incident, no pool — and write the principle it serves in one sentence a person can hold in their head at speed. Publish both, to the front line and, if you have the nerve, to customers, because a discretion budget the customer knows about is a costly signal in exactly the sense chapter two described: a company that intends to be difficult cannot afford to publish one. Require a one-line log of every use and read the logs monthly, because the pattern in them is the cheapest defect data you will ever obtain. Then go find the metric that contradicts all of it — the handle time, the save rate, the attach target — and either remove it or hold it against a countervailing measure that costs someone their bonus when it moves the wrong way. And when the first person spends four hundred dollars on a customer who did not deserve it, decide in that moment, knowing the whole organisation is watching, whether the authority you granted was real.

Everything else in this chapter is commentary on those five answers matching.

YOUR PAGE

Chapter Nine, carried forward. *Which decision in front of you this week is a deposit, and what will it cost you to make it?*

The Briefs

*Chapter Nine, in instruments*



*The Discretion Budget: Setting the Dollar and Policy Limit for Every Frontline Role*

HAUTE LUMIÈRE · EARNED · BRIEF 9.1

BRIEF 9.1

## The Discretion Budget: Setting the Dollar and Policy Limit for Every Frontline Role

Frontline staff face a customer complaint where policy says “no” but the account is worth keeping. The customer waits. The agent checks the handbook. The answer stays no. Trust degrades in that silence. The move

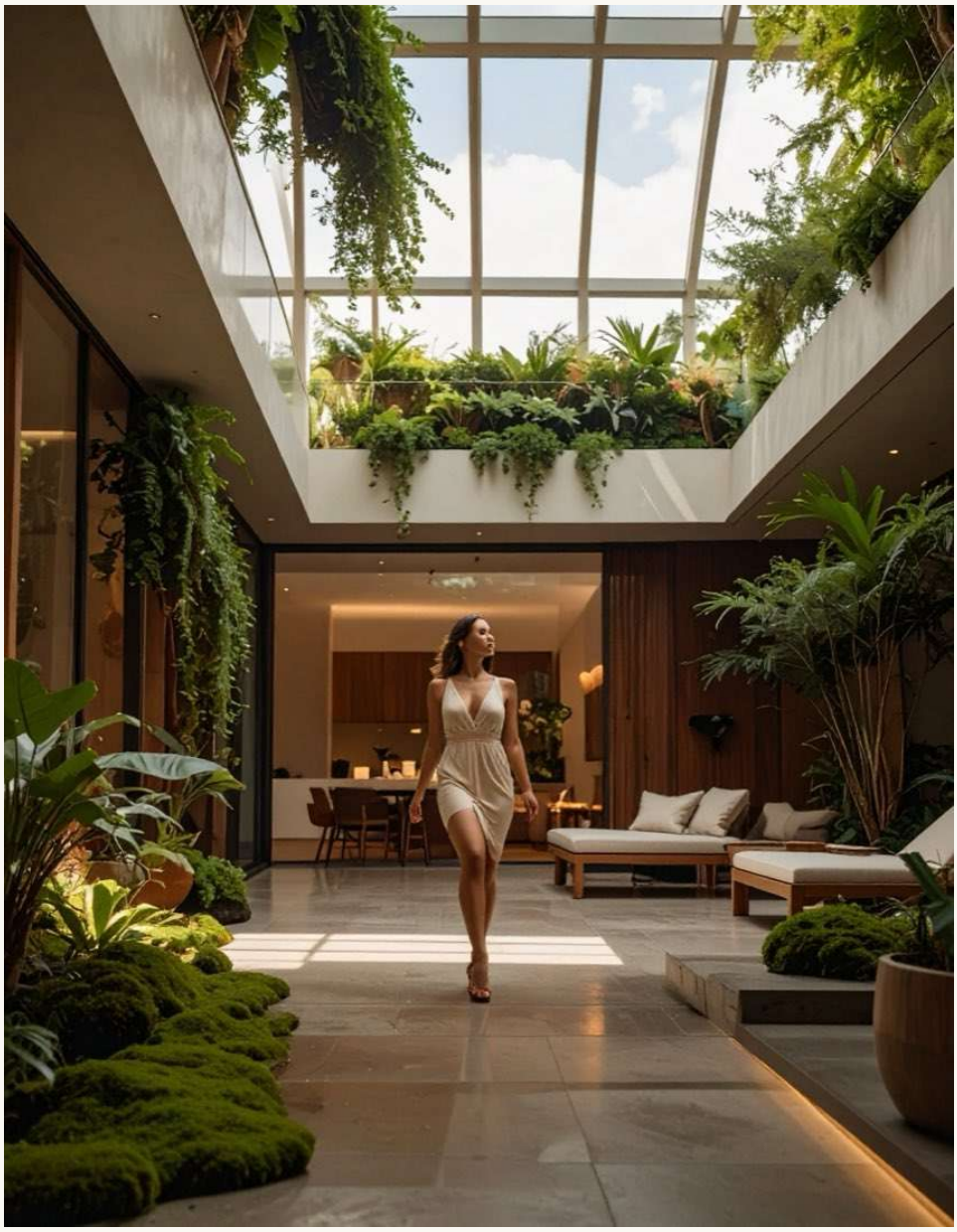
is simple: allocate a role-specific Discretion Budget, a fixed dollar amount and a bounded policy exception tier, that every frontline employee may deploy without managerial approval. The budget converts abstract loyalty into operational capacity. When a support agent can resolve a $180 service failure with a prorated refund plus a one-week platform extension, the interaction closes in one touch instead of three. The mechanism works because it reduces transactional friction while embedding a transparent cost floor; the company absorbs a known loss to preserve a relationship it has already priced. It requires three structural conditions. The limit must sit in the CRM interface, not in a searchable PDF. The budget must replenish monthly, functioning as working capital rather than a sporadic bailout. It must carry a hard per-account cap to prevent arbitrage. The budget succeeds by collapsing the temporal gap between customer loss and company recognition. Where the limit is contested, analysts point to historical cases like the 2014 Amazon Prime shipping overcharge, where the company absorbed the cost rather than arguing policy; the design mirrors that instinct but systematizes it. The real insight is that discretion is not a privilege granted to managers; it is a trust deposit that compounds when the person closest to the failure holds the pen. A discretionary budget becomes a leak when leadership treats it as a cost center rather than a relationship instrument. If managers retroactively claw back approved exceptions, the system inverts: agents stop using the budget, customers perceive the company as hostile, and the cost of re-earning trust doubles. The strongest objection is that unfettered discretion invites fraud; the design answers this by capping per-account usage and requiring a two-line exception note in the CRM, which audit teams review quarterly, not in real time. First action: map your top five frontline roles. Assign each a $75–$300 monthly discretion bucket tied to their average handle value. Publish the limit. Turn it on for thirty days.

BRIEF 9.2

## Metric Autopsy: What Average Handle Time Instructs Your Agents to Do

An enterprise reduces Average Handle Time by 40 percent through a new routing algorithm. Customer satisfaction scores drop. Attrition rises. The VP calls it a successful optimization. The move is to conduct a quarterly Metric Autopsy on every frontline KPI, mapping how each measurement instructs behavior, and retire or reweight any metric that penalizes resolution complexity. Metrics do not measure performance; they direct it. When an organization optimizes for speed, agents learn to close tickets, not solve problems. The autopsy forces a reverse-engineering of the incentive structure: for every hour saved, what unresolved friction did the agent carry to the next interaction? The mechanism operates through behavioral feedback loops. Agents optimize for what is counted. By isolating the correlation between speed metrics and repeat-contact rates, leaders identify which KPIs manufacture hidden service debt. The process requires a clean dataset linking handle time, resolution flags, and customer retention cohorts. It works only when leadership accepts that a faster touch is worthless if it generates a second touch. Where the autopsy is contested, operations argue that speed metrics prevent bloat; the counter is that unweighted speed metrics systemat-

ically punish complex, high-value, or distressed accounts, converting efficiency into erosion. The insight is that trust is not lost in singular failures but in the accumulated mathematics of optimized shortcuts. A Metric Autopsy becomes destructive when it is used to justify further automation without restoring agent authority. If the audit concludes that a metric must stay, it must be paired with a compensatory authority budget, otherwise the agent faces a double bind: optimize for a number that guarantees failure, or ignore the number and face review. The strongest objection is that removing speed targets creates slack; the design answers this by measuring first-contact resolution alongside speed, shifting the optimization surface toward quality without sacrificing velocity. First action: pull last quarter's handle time versus repeat-contact rate for your top ten support roles. Identify the metric that correlates highest with churn. Pause it for one month. Measure the shift.



*The Quota That Manufactured Fraud: Reading Wells Fargo as a Design Failure, Not a Morality Play*

HAUTE LUMIÈRE · EARNED · BRIEF 9.3

BRIEF 9.3

## The Quota That Manufactured Fraud: Reading Wells Fargo as a Design Failure, Not a Morality Play

Wells Fargo opened millions of unauthorized accounts to meet cross-selling quotas. Regulators fined the bank billions. Executives called it a culture problem. The root was a measurement architecture. The move is

to replace output quotas with process-weighted targets that reward accurate onboarding, not volume, and decouple individual performance from aggregate branch targets. Quotas do not create fraud; they create the conditions where fraud becomes the rational path for survival. When a bank ties compensation to a number of accounts per employee, agents face an impossible trilemma: meet the target, report accurately, or keep their job. The Wells Fargo case in 2011–2013 demonstrates this clearly. The mechanism operates through pressure transmission. High output targets compress decision time, forcing agents to substitute verification for volume. The design reverses this by weighting targets toward process compliance: completed identity checks, documented customer consent, and verified account usage. The mechanism requires three conditions. Compensation must be partially tied to audit-pass rates. Branch targets must be decoupled from individual commissions to prevent peer pressure from enforcing shortcuts. Leadership must publicly punish metric manipulation, not just outcomes. The quota succeeds by making honesty the path of least resistance. Where the model is contested, sales leaders argue that volume targets drive necessary market penetration; the counter is that unweighted volume targets systematically generate false positives, which later require costly remediation and destroy the very revenue they were meant to create. The profound insight is that fraud is rarely a moral failure; it is a structural inevitability when measurement outpaces authority. A process-weighted target becomes inert when compliance departments audit in silos. If the process metrics are reviewed by a different team than the one selling, fraud hides in the gap between sales velocity and compliance verification. The strongest objection is that decoupled targets reduce competitiveness; the design answers this by showing that false accounts generate negative revenue through chargebacks, regulatory penalties, and churn, making volume without verification a net liability. First action: audit your current sales quota structure. Identify the single metric that drives 60 percent of compensation. Rebalance it so that accurate onboarding and customer verification account for at least 40 percent of the payout. Publish the change.

BRIEF 9.4

## Save Rates and Retention Offers: The Metric That Punishes Honesty

A SaaS company introduces a retention team tasked with preventing churn. The team offers deep discounts to keep customers. CAC rises. LTV falls. The company believes it saved revenue; it actually destroyed margin. The move is to replace blanket retention discounts with tiered, value-anchored retention offers that require agents to diagnose usage barriers before pricing. Save rates do not prevent churn; they often delay it while eroding margin. When agents are measured on retention percentage, they learn to trade price for time. The mechanism operates through misaligned temporal incentives: the agent gets credit for keeping the account this quarter, but the company absorbs the margin loss next year. The design reverses this by tying retention authority to usage diagnostics. Agents must document the specific friction causing the cancellation request before any discount can be applied. The mechanism works only when the

CRM blocks discount buttons until a usage path is selected. It requires clear pricing tiers and a non-negotiable minimum margin floor. The save rate succeeds by converting price negotiations into product alignment. Where the model is contested, finance argues that discounting is cheaper than acquisition; the counter is that deeply discounted retained customers rarely expand, rarely refer, and often churn faster once the discount expires, making the save rate a deferred liability. The insight is that trust is manufactured when a company refuses to buy back a customer it failed to serve. A tiered save rate becomes a margin trap when leadership treats discounts as the primary retention tool. If agents receive quarterly targets to "save 80 percent of at-risk accounts," they will discount regardless of usage, converting loyal customers into subsidy recipients. The strongest objection is that competitors offer deeper discounts; the design answers this by measuring net revenue retention rather than gross retention, showing that value-anchored saves compound while price-driven saves decay. First action: review your last 100 retention calls. Tally how many ended with a discount before a usage diagnosis. Rebuild the workflow so discounts require a logged friction point. Set a margin floor. Test for one quarter.



*Outsourced Promises: Contract Terms That Stop a Vendor*

HAUTE LUMIÈRE · EARNED · BRIEF 9.5

BRIEF 9.5

## Outsourced Promises: Contract Terms That Stop a Vendor

The architecture of the subordinate contract. When a company subcontracts customer-facing functions, it does not outsource trust; it outsources the conditions under which trust can be manufactured. The legal boundary becomes the operational ceiling. If the vendor contract lacks pass-through liability, scope clarity,

and customer-originated escalation paths, the frontline worker operates in a vacuum of responsibility, forced to translate institutional failure into polite deflection. The mechanism works only when contract terms are structured as behavioral scaffolding rather than punitive shields. It requires explicit pass-through clauses that bind the vendor's revenue to the customer's stated outcome, not the vendor's activity volume. It requires a customer-originated escalation path that cannot be blocked by tiered support matrices. The contract succeeds by converting liability into alignment. Where the model is contested, procurement argues that standardized service level agreements protect margins; the counter is that SLAs measured by response time, not resolution depth, incentivize deflection over diagnosis, turning every interaction into a transactional deferral. The insight is that trust is manufactured when a contract explicitly authorizes discretion at the edge, while financially penalizing its absence. A rigid SLA becomes a trust trap when leadership treats compliance as the primary deliverable. If vendors receive quarterly bonuses for maintaining a three-minute average handle time, they will triage regardless of complexity, converting customers into ticket numbers. The strongest objection is that broader discretion increases operational risk; the design answers this by measuring customer retention and complaint severity rather than activity throughput, showing that outcome-aligned contracts compound while compliance-driven contracts fragment. First action: audit your vendor contracts for escalation blockers. Tally how many require managerial approval for edge exceptions. Rewrite the authority matrix to place resolution power with the first responder. Test for one quarter.

The mechanism operates through what contract lawyers call flow-down provisions, but what the customer experiences as continuity. When a manufacturer outsources logistics, the end user does not distinguish between the brand on the box and the driver at the door. The contract, therefore, must treat the driver's interaction as a direct extension of the brand's promise. The mechanism works by embedding customer-originated escalation into the payment structure. Instead of paying for truck-miles or call-minutes, the buyer pays for first-contact resolution, with a penalty clause that triggers only when a customer-initiated escalation exceeds a defined threshold. This flips the vendor's incentive from activity maximization to outcome stabilization. The condition is that the contract must permit the vendor to make financial sacrifices at the edge without triggering a breach. If a route driver reroutes to deliver a medical supply instead of following the optimized sequence, the contract must recognize that deviation as a performance positive, not a cost overrun. Where the model is contested, operations argues that route optimization reduces fuel costs by twelve percent; the counter is that twelve percent in fuel savings dissolves when a single missed delivery triggers a churn event worth four times the margin. The insight is that trust is manufactured when a contract financially rewards the vendor for overriding its own efficiency metrics in favor of the customer's actual need. An optimized logistics contract becomes a trust trap when leadership treats adherence to algorithms as the primary deliverable. If dispatchers receive quarterly bonuses for maintaining on-time percentages above ninety-four percent, they will reroute regardless of emergency context, converting loyal customers into statistical outliers. The strongest objection is that algorithmic compliance reduces variability and protects margins; the design answers this by measuring complaint severity and repeat order fre-

quency rather than on-time percentages, showing that outcome-aligned contracts compound while compliance-driven contracts fragment. First action: review your vendor payment schedules. Tally how many tie compensation to activity volume rather than customer resolution. Rewrite the incentive structure to reward edge discretion. Test for one quarter.

The mechanism fails when precision replaces judgment. A contract that specifies every possible interaction becomes a bureaucratic instrument, not a trust instrument. When lawyers draft contingency clauses for every edge case, they remove the space where human responsibility lives. The vendor worker, armed with a seventy-page procedural manual, stops listening to the customer and starts cross-referencing the manual. The contract's failure mode is not ambiguity; it is over-specification. The strongest objection to this claim is that undefined terms invite fraud and inconsistent service. The counter is that excessive definition creates defensive workarounds, where vendors document compliance rather than solve problems. The mechanism works only when the contract distinguishes between non-negotiable boundaries (safety, regulatory compliance, data sovereignty) and negotiable pathways (resolution methods, timeline adjustments, resource allocation). The insight is that trust is manufactured when a contract leaves deliberate silence in the middle, forcing the frontline worker to exercise judgment rather than execute instructions. A perfectly specified contract becomes a trust trap when leadership treats predictability as the primary deliverable. If vendors receive quarterly bonuses for maintaining a zero-variance error rate, they will escalate every anomaly, converting customers into managerial interruptions. The strongest objection is that human discretion introduces unacceptable risk; the design answers this by measuring customer satisfaction and escalation frequency rather than error rates, showing that outcome-aligned contracts compound while compliance-driven contracts fragment. First action: audit your vendor procedural manuals. Tally how many steps require managerial approval for edge exceptions. Rewrite the authority matrix to place resolution power with the first responder. Test for one quarter.

The mechanism requires a concrete anchor to survive abstraction. Consider the 2009 transition of Zappos' customer service operations to an outsourced contact center. The initial contract specified script adherence, average handle time targets, and first-call resolution metrics. The result was a measurable drop in customer loyalty scores and a sharp increase in post-purchase support requests. Zappos did not cancel the contract; they rewrote it. The revised agreement removed script compliance, replaced handle-time targets with customer satisfaction scores, and embedded a pass-through clause that tied vendor revenue to repeat purchase rates rather than call volume. The vendor was explicitly authorized to spend as much time as necessary on a single call, provided the customer's stated outcome was resolved. The mechanism worked because the contract shifted the vendor's incentive from activity to alignment. Where the model was contested, the vendor argued that unlimited call time would inflate costs; the counter was that each resolved interaction reduced downstream support requests, marketing waste, and churn, creating a net positive margin over twelve months. The insight is that trust is manufactured when a contract financially rewards the vendor for ending the interaction, not prolonging it. A script-based support contract becomes a trust

trap when leadership treats compliance as the primary deliverable. If vendors receive quarterly bonuses for maintaining a two-minute average handle time, they will triage regardless of complexity, converting customers into ticket numbers. The strongest objection is that broader discretion increases operational risk; the design answers this by measuring customer retention and complaint severity rather than activity throughput, showing that outcome-aligned contracts compound while compliance-driven contracts fragment. First action: review your vendor contracts for escalation blockers. Tally how many require managerial approval for edge exceptions. Rewrite the authority matrix to place resolution power with the first responder. Test for one quarter.

The mechanism operates through what living systems call adaptive feedback. When a contract treats the customer as an external stakeholder rather than the core feedback loop, the system becomes brittle. The contract must embed the customer's voice directly into the vendor's performance review. This means the vendor receives quarterly reviews not from the buyer's procurement team, but from a randomized sample of the customers they serve. The buyer's role shifts from auditor to facilitator, providing the vendor with customer-originated data, not buyer-imposed metrics. The mechanism works by closing the loop between promise and perception. Where the model is contested, legal argues that third-party customer reviews introduce bias and damage vendor relationships; the counter is that bias is preferable to distance, and that vendor relationships should be measured by customer outcomes, not contractual convenience. The insight is that trust is manufactured when a contract makes the customer the primary auditor of the vendor's performance. A buyer-centric contract becomes a trust trap when leadership treats compliance as the primary deliverable. If vendors receive quarterly bonuses for meeting buyer-defined SLAs, they will optimize for the buyer's reporting dashboard, not the customer's lived experience. The strongest objection is that customer feedback is subjective and inconsistent; the design answers this by measuring retention, expansion, and complaint severity rather than satisfaction scores, showing that outcome-aligned contracts compound while compliance-driven contracts fragment. First action: audit your vendor review processes. Tally how many rely on buyer-defined metrics rather than customer-originated data. Rewrite the review structure to place customer outcomes at the center. Test for one quarter.

The mechanism requires a final calibration: the contract must explicitly state what happens when trust breaks. Most contracts focus on prevention; the ones that actually stop vendors focus on response. When a customer's promise is broken, the contract must dictate the vendor's response within forty-eight hours, including financial restitution, root cause documentation, and a customer-approved corrective action plan. The mechanism works by treating failure as a data point rather than a breach. Where the model is contested, operations argues that mandatory restitution clauses increase costs; the counter is that rapid, transparent response converts a churn event into a loyalty event, as documented in the 2018 shift of several major cloud providers' support agreements. The insight is that trust is manufactured when a contract makes failure visible, expensive, and immediately actionable. A prevention-focused contract becomes a trust trap when leadership treats compliance as the primary deliverable. If vendors receive quarterly bonuses for

maintaining a zero-breach record, they will hide minor failures until they compound into major crises, converting customers into casualties of delayed response. The strongest objection is that transparency invites litigation and reputational damage; the design answers this by measuring customer retention and referral rates rather than breach counts, showing that outcome-aligned contracts compound while compliance-driven contracts fragment. First action: review your incident response clauses. Tally how many require legal approval before customer notification. Rewrite the response matrix to place transparency at the center. Test for one quarter.

The mechanism converges on a single principle: contracts do not build trust; contracts structure the conditions under which trust can be built by people at the edge. When a contract treats the customer as an external variable, the system becomes transactional. When a contract treats the customer as the core feedback loop, the system becomes relational. The choice is not between strict and loose terms; it is between compliance and alignment. The insight is that trust is manufactured when a contract explicitly authorizes discretion at the edge, while financially penalizing its absence. A rigid SLA becomes a trust trap when leadership treats compliance as the primary deliverable. If vendors receive quarterly bonuses for maintaining activity targets, they will optimize for those targets, not the customer's actual outcome. The strongest objection is that human discretion introduces unacceptable risk; the design answers this by measuring customer retention and complaint severity rather than activity throughput, showing that outcome-aligned contracts compound while compliance-driven contracts fragment. First action: audit your vendor contracts for escalation blockers. Tally how many require managerial approval for edge exceptions. Rewrite the authority matrix to place resolution power with the first responder. Test for one quarter.

The chapter closes where it began: at the edge, with fifteen seconds and no time to escalate. Trust is not a communication strategy. It is not a brand promise. It is the residue of decisions made when they were expensive, accumulated inside contract terms, support authority, incident response, and data handling. Companies destroy trust most reliably not by marketing campaigns or executive apologies, but by the architecture of the subordinate contract. When a company subcontracts customer-facing functions, it does not outsource trust; it outsources the conditions under which trust can be manufactured. The legal boundary becomes the operational ceiling. If the vendor contract lacks pass-through liability, scope clarity, and customer-originated escalation paths, the frontline worker operates in a vacuum of responsibility, forced to translate institutional failure into polite deflection. The mechanism works only when contract terms are structured as behavioral scaffolding rather than punitive shields. It requires explicit pass-through clauses that bind the vendor's revenue to the customer's stated outcome, not the vendor's activity volume. It requires a customer-originated escalation path that cannot be blocked by tiered support matrices. The contract succeeds by converting liability into alignment. Where the model is contested, procurement argues that standardized service level agreements protect margins; the counter is that SLAs measured by response time, not resolution depth, incentivize deflection over diagnosis, turning every interaction into a transactional deferral. The insight is that trust is manufactured when a contract explicitly authorizes discretion at

the edge, while financially penalizing its absence. A rigid SLA becomes a trust trap when leadership treats compliance as the primary deliverable. If vendors receive quarterly bonuses for maintaining a three-minute average handle time, they will triage regardless of complexity, converting customers into ticket numbers. The strongest objection is that broader discretion increases operational risk; the design answers this by measuring customer retention and complaint severity rather than activity throughput, showing that outcome-aligned contracts compound while compliance-driven contracts fragment. First action: audit your vendor contracts for escalation blockers. Tally how many require managerial approval for edge exceptions. Rewrite the authority matrix to place resolution power with the first responder. Test for one quarter.

Trust is earned in the silence between the clauses. It is not written in the penalties; it is written in the authorization. It is not built by the contract; it is built by the person who reads the contract, looks at the customer, and chooses differently. The machinery that makes that choice automatic, and the structures that make reversing it costly for whoever sits in the chair next, are not legal formalities. They are the foundation. Build them carefully. Measure them honestly. Let them fail. Rewrite them. Test again. The residue will follow.



*The essays · Chapter Nine*

HAUTE LUMIÈRE · EARNED · CHAPTER NINE

# The Essays

*Five assignments for the long evening*

ESSAY 9.1

The prompt — A front-line employee with wide discretion can save a relationship that no policy would have saved: the agent who waives the fee, ships the replacement before the return arrives, stays on the line

past the shift. That same discretion means the customer who calls on Tuesday is treated differently from the one who calls on Thursday, and the difference correlates — reliably, in the research on service encounters — with how articulate the customer is, how confident, how much they sound like the person deciding. Discretion is therefore a machine that produces both the best moments a company ever has and a quiet pattern of unequal treatment it would never defend if it saw it written down. The strongest case for rules is not efficiency but justice: a rule treats the diffident customer exactly as it treats the assertive one. The strongest case for discretion is that the promise a company actually made was to the person, and only a person at the edge can see what keeping it requires. Argue how much discretion is optimal — and, because the two goods genuinely trade against each other, say how a company should choose between them when it cannot have both.

What a serious answer has to do — The essay must specify the variable it is optimising before it optimises: discretion over *what*, bounded by *what*, exercised by someone selected and trained *how*. It has to establish that inconsistency is a real harm with an identifiable victim rather than a theoretical blemish, and then say who bears it — which requires engaging with the evidence on differential outcomes by customer confidence, language, and status. It must also show the mechanism by which discretion produces recovery: not "empowered employees care more," but the specific informational advantage the edge holds that the centre cannot encode. The cheap answer is "empower your people and trust them," which is not an argument, only a slogan with good manners; the second cheap answer is a proposed synthesis — "discretion within guardrails" — offered without saying where the guardrails go, which is the entire question. A serious essay commits to a rule for drawing that line and shows what its own rule costs.

Where to look — The service-recovery literature in operations and marketing, which has studied the paradox of the well-handled failure directly. Hotel and airline front-desk practice, where named discretionary spending limits are public and comparable. Clinical medicine, which has fought the protocol-versus-judgment argument for decades and produced the most honest accounting of what standardisation buys and what it destroys; the checklist debates in surgery are unusually well documented. Sentencing policy, where the move to guidelines was justified explicitly on equal-treatment grounds and the consequences are extensively studied. Street-level bureaucracy as a field of public administration, which is the closest thing that exists to a general theory of this problem.

The length — 2,500 words minimum.

ESSAY 9.2

The prompt — Wells Fargo employees opened accounts customers had not asked for, in numbers that ran into the millions, and they did it under a sales-quota regime designed by people who never opened a single fraudulent account themselves. Both available accounts of responsibility are unsatisfying. Say the employees are responsible and you have described a company where thousands of ordinary people independently and simultaneously decided to commit fraud, which is not how ordinary people behave and not what the pattern of the conduct looks like. Say the executives are responsible and you have written the employees out of the moral picture entirely, treating adults with the option to refuse as instruments, which they were not — some of them did refuse, at cost, and their refusal is evidence that refusing was possible. Argue where responsibility properly sits. Then push the argument where it hurts: if a target that is aggressive enough to distort behaviour is a target that transfers culpability upward, what follows for the design of every aggressive sales target anywhere, including the ones at companies that have not yet had their scandal?

What a serious answer has to do — The essay must produce a criterion for when a target crosses from demanding into culpable, and the criterion has to be usable in advance rather than only in the post-mortem — a rule that only identifies bad targets after they have produced fraud is a description of hindsight, not a design principle. It has to hold two things at once: that the individuals had agency, and that the distribution of misconduct across thousands of people is a fact about the system, not about the moral character of the people it hired. Evidence that counts includes the internal reporting channels and what happened to people who used them, the compensation mechanics, and the treatment of employees terminated for missing targets versus those terminated for how they met them. The cheap answer is "culture" — an explanation that names the phenomenon and then stops, since a culture is precisely what is produced by the incentives the essay is supposed to analyse.

Where to look — The publicly available regulatory findings and consent orders from the Wells Fargo matter, the company's own commissioned board investigation, and the congressional testimony, all of which are on the record and unusually detailed about mechanics. Beyond the case: the organisational-behaviour literature on goal-setting and its documented dark side, where the researchers who established that specific hard goals raise performance later wrote about what else those goals raise. Criminal law's treatment of the intermediary — command responsibility, corporate criminal liability — for its long argument about culpability flowing through a hierarchy. Comparable quota-driven episodes in other sectors: pharmaceutical detailing, subprime origination, for-profit education recruiting.

The length — 2,500 words minimum.

ESSAY 9.3

The prompt — The work that determines whether a company is trusted is disproportionately the work it is most likely to outsource. Support, claims, moderation, collections, onboarding: labour-intensive, hard to automate, easy to price per contact or per minute, and located exactly where the promise is kept or broken. The standard objection is that a vendor's employee has no stake in the brand's long-run reputation and every incentive to hit the handle-time target on the contract, so trust-critical work cannot be outsourced without decay. The standard rebuttal is that this is a specification failure and not an ownership one — an in-house team measured on handle time will fail identically, and a vendor team with authority, tenure, and a contract that pays for resolution rather than volume may outperform the internal alternative, because vendors can specialise in exactly this work. Argue whether trust-critical work can ever be genuinely outsourced. If it can, specify the contractual and compensation structure that makes it possible; if it cannot, identify what precisely does not survive the boundary.

What a serious answer has to do — The essay has to name the specific thing that is supposed to be lost when work crosses the org boundary, in terms concrete enough to be measured or contracted for — accumulated product knowledge, authority to spend, tenure, willingness to escalate, or something else — and then test whether that thing is genuinely unbuyable or merely rarely bought. It must confront the pricing unit directly, since per-contact and per-minute pricing is the mechanism by which the vendor's economics diverge from the customer's interest, and any workable answer has to propose a unit that does not. It should address the tenure problem honestly: turnover in the sector is high, and knowledge that takes a year to build cannot be held by a workforce that turns over faster than that. The cheap answer is the flat "never outsource what you can't afford to get wrong," which sounds like principle and is actually just risk-aversion with no account of why the exceptions exist.

Where to look — Transaction-cost economics and the make-or-buy literature, which is the discipline built for exactly this question and which turns on asset specificity — a concept the essay can use without naming it. The published research on business-process outsourcing in customer service, including the work on attrition and its costs. Aviation and healthcare contracting, where safety-critical functions are routinely performed by contractors under regulated terms, offering a proof of concept that outsourcing plus stringent structure can work. Contrast with sectors where outsourcing produced visible failures: content modera-

tion, where working conditions have been extensively reported, and debt collection, where regulators have documented the consequences of contingency-fee structures.

The length — 2,500 words minimum.

ESSAY 9.4

The prompt — A metric aimed at the front line is not a measurement. It is an instruction, delivered to thousands of people, about what to do when the customer's interest and the company's diverge in the next fifteen seconds — and unlike the values statement, it is enforced. Average handle time tells an agent to end the call. First-call resolution tells them to keep it. Attach rate tells them to sell. Each is defensible in a meeting and each, at scale, produces a moral pattern nobody in that meeting chose. The objection to this framing is that it proves too much: every organisation must measure something, most metrics behave tolerably, and calling a number a moral instruction is a rhetorical move that makes ordinary management sound sinister. Make the case anyway — that the customer-facing metric is where a company's ethics actually live, and that the values on the wall are downstream of it. Then commit: name the single metric that does the most damage across the most industries, and defend the choice against the obvious rivals.

What a serious answer has to do — The essay must show the mechanism connecting a number to a behaviour, which means being specific about the transmission — what is displayed on the wallboard, what the coaching conversation says, what happens at the bottom of the distribution — rather than asserting that incentives matter. It has to make the scale argument rigorously: why a metric that produces a mildly bad choice one per cent of the time is a different moral object from a manager who occasionally makes a bad call. The commitment to a single worst metric must be argued comparatively, with the runners-up given their real strength, and it has to survive the strongest defence of the chosen villain, which usually exists and is usually about cost control in a business with thin margins. The cheap answer is to attack Net Promoter Score, which is a well-worn target and mostly a measurement-validity complaint rather than a moral one; if the essay lands there anyway, it needs a harm argument that goes beyond "the statistics are shaky."

Where to look — Goodhart's law and its origins in economic policy, plus the wider literature on performance measurement and gaming in public services, where the evidence on target distortion is richest — the British NHS waiting-time targets and the associated audit reports are extensively documented. Call-centre operations research on the handle-time and resolution trade-off. Sales-compensation design. Policing met-

rics, where the consequences of counting the wrong thing have been examined by commissions and by researchers with unusual candour. Read the practitioner literature too: the trade publications where these metrics are advocated will tell you what the defenders believe, in their own words, which the essay owes them.

The length — 2,500 words minimum.

## ESSAY 9.5

The prompt — An agent gives a customer the answer that is true rather than the answer that is profitable, and the company loses money on the transaction — a refund outside policy, a cheaper plan recommended, a sale not made because the product was wrong for the person. The company's stated values endorse the choice. Its accounting does not: the loss lands in this quarter, in that agent's numbers, in that team's variance report, while the benefit is diffuse, delayed, and shows up as a customer who did not leave, which nothing measures. The hard version of the question is not whether to protect that employee — most companies would say yes and mean it in the room. It is what the obligation actually consists of and how far it extends. If protection means the loss is simply absorbed, the company has created a decision with no cost and can expect the judgment behind it to erode. If protection is discretionary, granted case by case by a manager, then it is not protection at all; it is a lottery, and everyone watching learns to play safe. Argue what a company owes such an employee, and how that obligation must be made visible to everyone else who is watching.

What a serious answer has to do — The essay has to distinguish the good judgment that lost money from the bad judgment that lost money, and confront the fact that at the moment of decision they look identical to the person deciding — which means the obligation has to be defined on the quality of the reasoning rather than the outcome, and the essay must say how that is assessed without turning into a tribunal nobody would risk facing. It must treat visibility as a design problem with real costs: the mechanisms that make protection legible — public recognition, published cases, a named budget — also create records, precedents, and a surface for gaming. Evidence that counts is any documented instance of a company absorbing a loss and telling its own people about it, plus the structures that make such absorption routine rather than heroic. The cheap answer is the celebration story — the employee praised at the all-hands — which the essay must argue past, since a celebration is a one-time signal that costs nothing to issue and nothing to withdraw, and the question asked for an obligation.

Where to look — Just-culture frameworks in aviation and healthcare, which exist precisely to separate blameworthy error from good-faith judgment and are the most developed body of thinking on this problem anywhere. The literature on psychological safety and its measured effects on error reporting. Company practices where discretionary spending authority is named, budgeted, and treated as a cost of doing business rather than a variance to explain — the retail and hospitality sectors have the clearest public examples. Employment law on the treatment of employees who refuse a profitable instruction, and the whistleblower-protection literature for what happens when protection is nominal rather than structural.

The length — 2,500 words minimum.

The People Who Actually Keep It

Ammanuel Santa Anna · Luminous Prosperity Inc. · Haute Lumière

Set in Cormorant Garamond and Manrope.

Read the Business Library at business.hautelumiere.life/library.

Luminous Library Licence 1.0 · hautelumiere.life/library-licence

**REQUEST A CONVERSATION**

business.hautelumiere.life/book/?kind=gainshare

The Business Library

business.hautelumiere.life/library/

The packages

hautelumiere.life/pricing/