CONTENTS

FOR DTC FOUNDERS AND OPERATORS · A FIELD GUIDE

THE KEPT PROMISE

Delivery you can count on,
contacts you never get,
and what to do when it breaks anyway.

ANDREW LAUCHNER Author of The Second Order and The Whole Machine

September 2026 · 15 chapters · About 70 minutes

A note before you start

Every order is a promise: this thing, in this condition, by this date. Keep it and the customer barely notices. Break it and they notice twice, once when it happens and again when they decide whether to order from you next time.

5–10%	70%
less spent afterward by Uber riders whose trips ran furthest past the promised arrival time (Halperin and colleagues, 2022, a study of 1.5 million riders)	cut in contacts per order at Amazon during the three years Bill Price ran its customer service (Price, 2005)

Those two numbers are the guide in miniature. The first says a broken promise costs repeat revenue even when the product itself was fine: the car came, the rider got there, just later than the app had said. The second says the contacts a broken promise creates are not a fact of life. Most of them come from things other teams did: a date the site shouldn't have shown, a package that left the warehouse late, a tracking page that said nothing. Fix the cause and the contact never happens.

So this guide treats delivery, support contacts and recovery as one system: count every failure in one weighted number, promise dates you can keep, tell customers bad news first, send each contact reason back to the team that causes it, and make things right without believing that a good recovery beats no failure at all.

> **The promise kept is the retention metric. Everything else is a report on it.**

It builds on The Second Order, which covers the second purchase and cohorts, and The Whole Machine, which covers contribution margin and the core flows. Returns have their own guide in this series, at /returns; they appear here only where a wrong or damaged order creates one.

How to read it

Start with The Promise Audit. Your lowest checks name the chapters to read first. Or follow a path:

- **Founder:** What a Broken Promise Costs, One Number for Every Failure, Paying for the Promise.
- **Operations lead:** One Number for Every Failure, An Accurate Date Beats a Fast One, Bad News First.
- **Support lead:** Contacts per Order, Every Reason Has an Owner, The Recovery Grid, How to Apologize.
- **Analyst:** What a Broken Promise Costs, The Promise Scorecard, For Your Analyst.

Four tools run in the page. Nothing you type leaves your browser.

What's proven and what isn't

- **Published:** peer-reviewed research, books and essays by the researchers themselves, and official law and regulator guidance, with the date it was true.
- **Reported:** what a named company, author or outlet published, including executives describing their own companies.
- **Derived:** my arithmetic, with the inputs shown.

Examples that open with Say or Picture use made-up round numbers. Every source is listed in Appendix C. The section on the FTC's shipping rule is an operator's summary as of September 2026, not legal advice.

Andrew Lauchner

Scottsdale, Arizona

FRONT

TEN POSITIONS

What this guide argues, and what would prove each claim wrong.

A position says what would prove it wrong. Test each on your own store.

01 **A broken delivery promise costs repeat revenue, not just a support ticket.**

Wrong if customers whose first order was late reorder as often as matched customers whose first order was on time.

What a Broken Promise Costs

02 **Count every kind of failure in one weighted number, not on-time rate alone.**

Wrong if your on-time rate and your contacts per order have always moved together, week for week.

One Number for Every Failure

03 **An accurate delivery date beats a fast one you miss.**

Wrong if a later, safer checkout date cost more in conversion than it earned in repeat orders.

An Accurate Date Beats a Fast One

04 **Tell customers about a delay before they find it.**

Wrong if a random group that got no delay notice contacted you no more, and reordered no less, than the group that did.

Bad News First

05 **Showing the work makes waiting easier to accept.**

Wrong if customers who can see where their order is contact you as often as customers who can't.

Show the Work

06 **Most support contacts are broken promises made elsewhere in the business.**

Wrong if fewer than half your contacts, by reason code, trace to a decision made outside support.

Contacts per Order

07 **Every contact reason needs an owner outside support, with a target.**

Wrong if contacts fall as fast when support owns every reason as when each causing team owns its own.

Every Reason Has an Owner

---

08 **A recovered customer is worth less than a customer you never failed.**

Wrong if customers who had a failure and a good recovery reorder more than matched customers who had no failure.

The Recovery Paradox, Checked

---

09 **An apology works when something comes with it, and stops working when you repeat it.**

Wrong if an apology with no credit, or a third apology to the same customer, raises their later spending against a holdout.

How to Apologize

---

10 **Pay the whole team on the promise kept, not only the people who answer the phone.**

Wrong if a shared bonus tied to one delivery number doesn't move that number within two quarters.

Paying for the Promise

FRONT

THE PROMISE ON ONE PAGE

Six moments between the checkout and the next order. Each has a promise, a number that says whether it was kept, and a team that owns it.

Most brands measure the first moment and the fifth. The four in between are where promises break.

MOMENT	THE PROMISE	THE NUMBER	OWNER
1. The checkout	A delivery date, stated or implied	Share of orders delivered by the date shown	Ecommerce, with operations
2. The warehouse	The right items, packed to survive, out on time	Ship-on-time rate; wrong-item and short-ship rate	Fulfillment or the 3PL
3. The carrier	Transit in the days the date assumed	Transit time at the median and the 90th percentile, by lane	Operations
4. The doorstep	Arrives intact, all of it, once	Damaged, lost and split-shipment rates	Operations and packaging
5. The contact	An answer without chasing	Contacts per order, by reason	The team that caused the reason
6. The recovery	Made right, fast, once	Repurchase after each failure type	Support, with a budget

The moments feed each other. An overconfident date at moment 1 becomes a late delivery at moment 3, a contact at moment 5 and a credit at moment 6.

> **The contact queue is where the business's broken promises arrive. It's rarely where they're made.**

Do this

- **Fill in the number column** for your store from whatever you have today. Leave blanks blank and write a name next to each. Most brands can fill in two.

START HERE · CHAPTER 1

THE PROMISE AUDIT

Twelve checks on whether your store keeps its delivery promises, knows when it doesn't, and makes it right. About forty minutes with your order data, your help desk and a phone.

The audit isn't about how fast you ship. It's about whether the checkout promise matches what arrives, whether you hear about the gap before the customer does, and whether anyone outside support feels it when it breaks.

Open your order export, shipping platform, help desk and your checkout on a phone. Score each check 0 to 2: 0 if it failed or nobody can answer it, 1 if partly true, 2 if clean.

> **A promise nobody measures is a guess you made on the customer's behalf.**

The twelve checks

01

You know your on-time rate against the date you showed · 4 min

Look at: Last month's delivered orders, with the delivery date the checkout showed and the date the carrier scanned delivery.
Good: Someone can give the share delivered by the promised date, split by carrier and shipping method, today.
Cost if wrong: You track speed and miss the broken promises customers remember.
Read next: An Accurate Date Beats a Fast One

02

One weighted failure number · 4 min

Look at: Your weekly operations report.
Good: Late, very late, wrong, damaged, missing and unannounced split orders are counted in one index, weighted by how much each hurts, and reported per 1,000 orders.
Cost if wrong: The team improves on-time rate while damage and wrong items grow unseen.
Read next: One Number for Every Failure

03

Repurchase is measured after each failure type · 5 min

Look at: Whether anyone has compared repeat rates of customers whose first order failed and didn't.
Good: A table exists, by failure type, with a matched comparison group, refreshed each quarter.
Cost if wrong: Unpriced failures lose every budget argument to acquisition.
Read next: What a Broken Promise Costs

04

The checkout date comes from data, not a template · 3 min

Look at: Your product page and checkout on a phone, for an address near your warehouse and one far away.
Good: The date changes with address and cutoff time, set from your own transit data.
Cost if wrong: A fixed "3 to 5 business days" is broken for every far-away customer, every week.
Read next: An Accurate Date Beats a Fast One

05

Delay notices go out before the customer asks · 3 min

Look at: The last time a shipment stalled or an item went out of stock after purchase.
Good: An automatic message went out the day the delay was known, with a new date and a choice.
Cost if wrong: The customer finds out by waiting, then contacts you, angrier.
Read next: Bad News First

06

Your delay process meets the FTC shipping rule · 4 min

Look at: Your stated ship times, your backorder and preorder flow, and your refund timing.
Good: You ship within the time stated or 30 days; when you can't, the customer gets a revised date and the option to cancel for a prompt refund.
Cost if wrong: Civil penalties per violation, and a pattern regulators find easy to prove.
Read next: Bad News First

---

07 **The tracking page says something useful · 3 min**

Look at: The tracking link from your last shipping email, opened on a phone.
Good: It shows your brand, the expected date, each step so far in plain words, and how to get help.
Cost if wrong: "Label created" for three days reads as nothing happening, and becomes a contact.
Read next: Show the Work

---

08 **Contacts per order, by reason, every week · 4 min**

Look at: Your help desk's reporting.
Good: Every ticket, chat and call has one primary reason code, and contacts per order is reported weekly for each reason.
Cost if wrong: You staff up for volume instead of removing its causes.
Read next: Contacts per Order

---

09 **Each reason has an owner outside support · 3 min**

Look at: Your top five contact reasons.
Good: Each has a named owner in the team that causes it, with a target and a review date.
Cost if wrong: Support absorbs the cost of other teams' decisions forever.
Read next: Every Reason Has an Owner

---

10 **Agents have a recovery grid and a budget · 3 min**

Look at: What an agent can offer without asking a manager.
Good: A written grid says what each failure gets, and each agent has a monthly budget to spend beyond it.
Cost if wrong: Recovery depends on who picks up and how hard the customer pushes.
Read next: The Recovery Grid

---

11 **Apologies carry a remedy, and repeat failures are flagged · 3 min**

Look at: Your apology templates and your customer records.
Good: Every apology comes with a fix or a credit, and an agent can see when this customer was last failed.

Cost if wrong: Words alone do little, and the same apology twice does harm.
Read next: How to Apologize

12 **Someone outside support is paid on the promise · 2 min**

Look at: Bonus and review criteria for operations, fulfillment and ecommerce.
Good: At least one delivery or failure number sits in their goals, and ideally a shared one for everybody.
Cost if wrong: The people who can prevent failures are paid for speed and cost alone.
Read next: Paying for the Promise

Score as you go; your band appears when all twelve are in.

RUN YOUR NUMBERS

Score the twelve checks

0: failed, or nobody can answer it. 1: partly true. 2: clean. Scores stay in this browser.

01 You know your on-time rate against the date you showed 4 min	0	1	2
02 One weighted failure number 4 min	0	1	2
03 Repurchase is measured after each failure type 5 min	0	1	2
04 The checkout date comes from data, not a template 3 min	0	1	2
05 Delay notices go out before the customer asks 3 min	0	1	2
06 Your delay process meets the FTC shipping rule 4 min	0	1	2
07 The tracking page says something useful 3 min	0	1	2
08 Contacts per order, by reason, every week 4 min	0	1	2

09 **Each reason has an owner outside support** 3 min	0	1	2
10 **Agents have a recovery grid and a budget** 3 min	0	1	2
11 **Apologies carry a remedy, and repeat failures are flagged** 3 min	0	1	2
12 **Someone outside support is paid on the promise** 2 min	0	1	2

0 of 24 points	0 of 12 checks scored

Score all twelve to read your band. 12 to go.

RESET SCORES

Read your score

SCORE	WHAT IT MEANS	READ NEXT
20–24	You keep your promises and know when you don't. Your job now is pricing each failure and paying the whole team on the index.	What a Broken Promise Costs, then Paying for the Promise
14–19	The basics are in place, but some failures reach customers before they reach you. Fix the zeros first.	The chapter linked from your lowest check, then One Number for Every Failure
8–13	Support is absorbing failures made elsewhere, and nobody can say what they cost. Start counting.	Part one, starting at What a Broken Promise Costs
0–7	You don't yet know which promises you're breaking. Measure on-time against the date you showed, and	An Accurate Date Beats a Fast One, then The First Thirty Days

SCORE	WHAT IT MEANS	READ NEXT
	send delay notices, this month.	

If you use a 3PL, checks 1, 2 and 4 depend on data they hold. Put it in the contract: order-level ship scans, delivery scans and error codes, daily.

PART ONE · COUNT WHAT BREAKS · CHAPTER 2

WHAT A BROKEN PROMISE COSTS

The refund and the reship are the small part. The large part is the next order that never comes, and you can measure it on your own store.

Ask a team what a late order costs and they'll add up the credit, the agent's time and maybe a reship. The cost they can't see is the customer who quietly orders less often afterward. It shows up months later, in a cohort report nobody connects to the day the package was late.

What the research finds

- **Printed goods, online.** Shashank Rao, Stanley Griffis and Thomas Goldsby followed previously loyal customers of an online retailer after an order was delayed, against

comparable customers whose orders weren't. Afterward, they ordered less often and placed smaller orders PUBLISHED . These were the loyal customers a brand assumes will forgive it.

- **Ride-hailing.** In a study of 1.5 million Uber riders by Basil Halperin, Benjamin Ho, John List and Ian Muir, riders whose trips ran furthest past the arrival time Uber had shown went on to spend 5% to 10% less on the platform than comparable riders PUBLISHED . Nothing was wrong with the product: the car came, the rider got there. The only failure was the promise.
- **Food delivery.** Alice Harter, Lucas Stich and Martin Spann looked at about two million orders from a European quick-delivery service. Orders that arrived later than communicated lengthened the time to the customer's next order, and lateness hurt more than earliness helped PUBLISHED .
- **Restaurants, over years.** Tim Norvell, Piyush Kumar and Mayukh Dass tracked diners after service failures. Customers whose problem was fixed well came back about as often as unaffected customers for a while, then drifted down toward the customers whose problems were never fixed. Over the long run, the customers who were never failed visited most PUBLISHED .

Two details matter later. Harter's team, and a study of ratings at a South American online retailer by Serkan Akturk and colleagues, both found that the damage from lateness grows more slowly as the delay gets longer PUBLISHED . The first day late does a large share of the harm, which makes the promised date in chapter 4 the cheapest lever in this guide. And Norvell's result previews chapter 10: recovery softens a failure but doesn't erase it.

> **A late order doesn't cost you a refund. It costs you part of the next order, and the one after that.**

Measure it on your own store

Studies give the direction. Your own data gives the size, and the size wins budget. The method is a failure cohort:

---

01 **Take first-time customers from a full year ago**

So everyone has had twelve months to come back.

02 **Label what happened to that first order**

On time, late 1 to 2 days, late 3 or more, wrong or missing, damaged, lost, split without warning. One label per order, the worst.

03 **Compare repeat rates within matched groups**

Late orders cluster in far regions and peak weeks. Compare within the same month, region and first product.

04 **Read the gap, with its uncertainty**

Pool quarters until each group has a few hundred customers.

Matching narrows the bias without removing it. The cleanest evidence comes from failures you didn't choose, like a carrier's regional outage: customers caught in it were failed more or less at random. Keep a dated list of such events. The query is in Appendix A.

A worked example

Say a brand ships 10,000 orders a month and 6% of them fail in some way: 600 orders. Customers with no failure come back within a year at 30%. The failure cohort shows that a failure cuts that rate by a fifth, to 24%. So of the 600 failed customers, 36 who would have come back don't. Each returning customer places 2.5 more orders in the year, at $25 of contribution each.

That's 36 × 2.5 × $25 = $2,250 of contribution lost each month, from one month's failures, before any refund or reship. At $12 of direct cost per failure, the direct cost is $7,200 a month. The hidden cost is almost a quarter of the total, and it's the quarter nobody reports DERIVED .

RUN YOUR NUMBERS

What does a broken promise cost?

Example numbers. Replace with yours. A failure is any order that was late, wrong, damaged, lost or split without warning.

Orders shipped per month

Share of orders that fail

10000	6 %
12-month repeat rate, customers with no failure	How much a failure lowers that rate (relative)
30 %	20 %
Orders a returning customer places in 12 months	Contribution per order
2.5	$ 25
Direct cost per failure: credit, reship, contact	
$ 12	

36 repeat customers lost per month of failures	$15.75 total cost per failure, across 600 failures a month	24% of that cost is lost repeat contribution	$113,400 a year, at this failure rate
$18,900 a year, for each point of failure rate you remove			

> **About 24% of each failure's cost is lost repeat contribution that no report shows. Cutting the failure rate by one point is worth about $18,900 a year: use that as the budget ceiling for the fix.**

Lost repeat contribution per failure = repeat rate × relative drop × orders per returning customer × contribution per order. It counts only the first year after the failure and ignores word of mouth, so it's a floor, not an estimate of the whole cost.

With the defaults, each failure costs $15.75, 24% of it in lost repeat contribution, and one point off the failure rate is worth about $18,900 a year. The 20% drop is a placeholder; replace it from your failure cohort, since it decides the answer. For scale, the Uber study's 5% to 10% fall in spending followed a car ride that arrived late; a damaged or lost order is a bigger failure than that.

Do this

- **Build the failure cohort this month** for first orders placed twelve to eighteen months ago. Put the repeat rate for each failure type beside the no-failure rate, on one page.
- **Run the tool with your own drop** and put the per-point value at the top of the operations roadmap, so a packaging fix can compete with an ad budget.

PART ONE · COUNT WHAT BREAKS · CHAPTER 3

ONE NUMBER FOR EVERY FAILURE

FedEx counted every kind of failure, weighted by how much it hurt, in one daily number. The idea fits a DTC brand almost unchanged.

An on-time rate of 96% sounds good and tells you almost nothing. It hides the 4% that were late and treats a day-late package the same as one that never arrived. It says nothing about damage, wrong items or the customer who had to write in twice. A team told to raise it will raise it, sometimes by shipping faster and packing worse.

How FedEx did it

According to a Federal Highway Administration review of its methods, until 1989 FedEx assumed that on-time delivery was what its customers valued most, and customer research

showed they expected much more REPORTED . So it built the Service Quality Indicator: twelve kinds of failure, each counted every day and multiplied by a weight that reflected how much it hurt customer satisfaction. When FedEx won the Malcolm Baldrige National Quality Award in 1990, the award profile noted that SQI reports went daily to workers at every site, management met daily to discuss the previous day's performance, and executives were evaluated on the SQI PUBLISHED .

FAILURE	WEIGHT
Right day, late	1
Wrong day, late	5
Traces not answered	1
Complaints reopened by customers	5
Missing proofs of delivery	1
Invoice adjustments requested	1
Missed pickups	**10**
Lost packages	**10**
Damaged packages	**10**
Aircraft delay, in minutes	5
Overgoods (packages that lost their labels)	5
Abandoned calls	1

REPORTED The weights as given in case materials on Federal Express. A version cited by the Federal Highway Administration, from a 2003 article in California CPA, lists an international indicator in place of aircraft delay. The structure, twelve items weighted by their effect on satisfaction, is from the 1990 Baldrige award profile.

Two things are worth copying. The weights are blunt, 1, 5 and 10, not decimals from a regression. And two items are about the service around the delivery: a complaint the

customer had to reopen, and an abandoned call. FedEx counted the customer having to chase as a failure in its own right.

> **A raw count treats a lost package like a late one. A weighted count tells the team which failure to fix first.**

The DTC version

Here's the index I'd start on, with FedEx's 1, 5, 10 scale until your failure cohort from chapter 2 lets you set each weight in proportion to what that failure costs.

FAILURE	COUNTED WHEN	STARTING WEIGHT
Late, 1 to 2 days	Up to two days past the date shown	1
Late, 3 days or more	Three or more days past it, or past a date that mattered	5
Split without warning	Arrived in unannounced pieces	1
Wrong or missing item	Any line wrong or short	5
Damaged	Arrived unusable	10
Lost	Never received	10
Contact reopened	The customer had to come back about it	5

Report it weekly as points per 1,000 orders, beside the perfect-order rate: the share of orders with no failure at all. Show the board the perfect-order rate. Run the operation on the index, because it says where the points come from.

RUN YOUR NUMBERS

Your weighted failure index

Example numbers for one month. Replace with yours. Enter a count and a weight for each failure; zero is fine for a count.

Orders delivered in the period

10000

Late 1 to 2 days (count)	350	Weight	1
Late 3 days or more (count)	90	Weight	5
Split without warning (count)	60	Weight	1
Wrong or missing item (count)	50	Weight	5
Damaged (count)	40	Weight	10
Lost (count)	15	Weight	10
Contacts reopened (count)	45	Weight	5

1,885	188.5	94.0%	Orders late by 3 days or more
index points in the period	points per 1,000 orders	perfect orders, at least	biggest source of points: 24% of the index, 14% of the failures

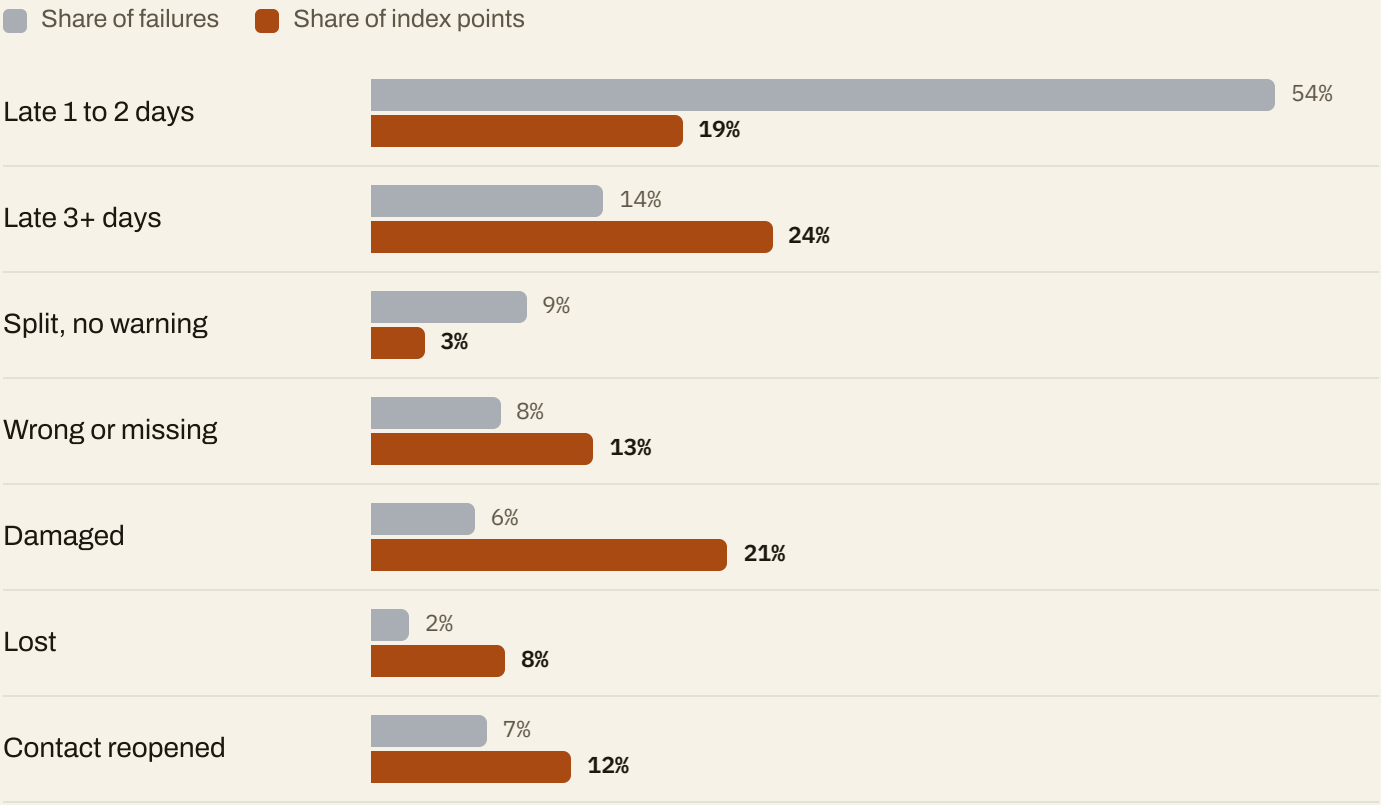
Start with orders late by 3 days or more: 24% of your index from 14% of the failure count. Its weight makes it count for more than its share of failures, so a raw count would understate it.

Perfect orders assumes no order had two failures; if some did, the true share is a little higher. Reopened contacts count in the index but not against perfect orders, since they're about the service after delivery.

With the defaults, the month scores 1,885 points, or 188.5 per 1,000 orders, with a perfect-order rate of at least 94.0%. The biggest source is orders late by three days or more: 24% of the index from 14% of the failures. The chart shows why the weights matter.

The most common failure isn't the most expensive

Share of failures and share of weighted index points, by type. The example month from the tool



DERIVED 650 failures and 1,885 points from the tool's example counts and starting weights. Bars are scaled to 60%.

By raw count, the team would spend the quarter on short delays. By points, it looks at long delays and damage first, and the damage fix is often a packaging change costing cents per order.

Keep it honest

- **Count from systems, not complaints.** Damage and wrong items reported through contacts undercount; the delivery check question in chapter 6 finds the silent ones.

- **Change weights once a year,** from the failure cohort, and restate last year on the new weights.
- **Report the worst lane,** the carrier, region or product that fails most, not only the average.

Do this

- **Put last month's counts into the tool** and circulate the result with the chart. Ask each team which of the top two sources it can move.
- **Add the index and perfect-order rate** to the weekly report, with the worst lane.

PART TWO · PROMISE WHAT YOU CAN KEEP · CHAPTER 4

AN ACCURATE DATE BEATS A FAST ONE

Speed sells. A missed date costs more than speed earns. Set the checkout date from your own delivery data, at the share of orders you're willing to be wrong about.

You can improve a delivery promise by making delivery faster, which costs warehouses and carrier upgrades, or by making the promise truer, which costs a little conversion. Most brands spend on the first. The research says the second is where the retention is.

Speed does sell

When a US apparel retailer opened a new distribution center that cut delivery times to its western customers, Marshall Fisher, Santiago Gallino and Joseph Xu found online sales rose about 1.45% for each business day saved, from a starting point of seven business days, with a spillover to the retailer's stores PUBLISHED . On Alibaba's Tmall, Vinayak Deshpande and Pradeep Pendem estimated that cutting three-day deliveries to two days would lift average daily sales for third-party sellers by 13.3% PUBLISHED . In a business-to-business online store, Shin Oblander and Kinshuk Jerath found each day off the promised time lifted demand 1.82%, like a 2.21% discount, though buyers barely reacted to promises under a week PUBLISHED .

The promise is a separate lever, and it cuts both ways

One study separates the promise from the delivery. Ruomeng Cui, Zhikun Lu, Tianshu Sun and Joseph Golden worked with Collage.com, which sells custom photo products, to change the delivery estimates shown on the site while the actual delivery speed stayed the same. A faster promise raised sales and profits. It also raised returns and reduced customer retention PUBLISHED . The faster date bought orders today with broken promises that cost orders later.

Recall from chapter 2 that lateness hurts repurchase more than earliness helps. And when Nooshin Salari, Sheng Liu and Zuo-Jun Max Shen built a model on JD.com data that forecasts the whole spread of possible delivery times and sets each promise with the cost of being late in mind, their simulations put the sales gain over JD.com's existing policy at 6.1% PUBLISHED . A promise set from data, not a template, is worth money in both directions.

> **A fast promise wins the order. A kept promise wins the next one.**

Set the date from the tail, not the middle

Most checkout promises are set from the typical delivery: "usually 3 to 5 business days." But the slow tail, not the typical order, generates the contacts. If half your orders arrive in four days and one in ten takes seven or more, a five-day promise breaks far more often than the team thinks. Pick the share you're willing to deliver late, then promise the date that share implies. I'd start at 90% to 95% on time against the date shown, by region.

RUN YOUR NUMBERS

What date can you promise?

Example numbers. Replace with yours, for one region or shipping method. Days are order to door, measured in fractions of a day if you can.

Median days, order to door

4 days

90th percentile days, order to door

7 days

Days you promise now

5 days

Share you want on time

95 %

Orders per month on this lane

10000

69.5%	9 days	3,047 → 317	4.8 days
on time at your current 5-day promise	to promise for 95% on time (96.8% at 9 days)	late orders a month: now, then at that promise	90th percentile you'd need to keep today's 5-day promise at 95%

> **Your 5-day promise is kept 69.5% of the time: about 3,047 broken promises a month. Either promise 9 days, or cut the slow tail until one order in ten takes no more than 4.8 days. Most brands do some of each, by region.**

Fits a lognormal curve to your median and 90th percentile, a common shape for delivery times, which have a long slow tail. It's an approximation: check it against the share of last month's orders that actually arrived by the date shown.

With the defaults, a five-day promise on a lane with a four-day median and a seven-day 90th percentile is kept about 70% of the time: roughly 3,000 broken promises a month on 10,000 orders. Hitting 95% would take a nine-day promise, or a tail short enough that one order in ten takes no more than about 4.8 days. Neither is comfortable. The template "3 to 5 business days" hides that choice; the tool puts it on the table.

What to do with the answer

- **Show a date, not a range.** “Arrives by Thursday, October 8” is a promise you can measure. It should move with the address and the cutoff time.
- **Promise by region.** One national promise is too slow for near customers and broken for far ones.
- **Attack the tail, not the median.** The misses sit in the warehouse over a weekend, miss the pickup or cross the country by ground. Cutoffs, weekend picking and a faster service for far zones move the tail.
- **Test the promise the way Collage.com did.** Change the date in some regions and not others, and read conversion, lateness, contacts and 90-day repeat orders together.

Shipping price and free-shipping thresholds are covered in The First Offer. This chapter is only about the date.

Do this

- **Run the tool for your largest region** from last month’s order-to-door times, and check it against the actual on-time share.
- **Change the checkout promise** to what your data supports, region by region, and log the date.

PART TWO · PROMISE WHAT YOU CAN KEEP · CHAPTER 5

BAD NEWS FIRST

The worst way for a customer to learn an order is late is by waiting for it. Tell them the day you know, with a new date and a choice. The law already requires most of this.

Every delay gets discovered. The question is who discovers it first. If it's you, you explain it and offer a choice. If it's the customer, they find it on a tracking page that hasn't changed in three days, and the first thing you hear is a complaint.

Why the notice matters more than the delay

David Maister's 1985 essay on waiting set out propositions that have held up: uncertain waits feel longer than known ones, unexplained waits longer than explained ones, and anxiety makes waits seem longer PUBLISHED . Shirley Taylor's 1994 study of delayed airline passengers found delays hurt evaluations mainly through anger and uncertainty PUBLISHED . A delay notice attacks exactly those: it makes the wait known, explains it, and shows someone is in control.

Order matters too. Angela Legg and Kate Sweeny found that people receiving mixed news mostly wanted the bad news first, while the people delivering it had to be nudged to give it that way; hearing it first left recipients less worried PUBLISHED . The delay goes in the subject line and the first sentence.

> **Put the bad news in the subject line. Everything after it is the part they'll actually read.**

A caution: I haven't found a published field experiment isolating the effect of a proactive delay email on ecommerce repurchase. The case rests on the waiting research and the studies in chapter 2. So hold back a random group, as below, and measure it yourself.

When to send

Send the notice the day the delay becomes likely, not the day it's certain. Five triggers catch most delays:

- **Not shipped by the promised ship date.**

- **No carrier scan within a day of the label,** which usually means a missed pickup.
- **No new scan for two days in transit,** or a carrier exception.
- **The carrier's estimate passes the date you showed.** Compare them daily.
- **Out of stock after purchase:** an oversell, a split, a backorder.

What it says

01 **The news, first**

"Your order is running late."

02 **The new date**

Set from your data, not a hope. If you can't give one, say when you'll write next.

03 **The reason, in one sentence**

"The carrier missed its pickup on Tuesday," not "unforeseen circumstances."

04 **The choice**

Wait, cancel for a full refund in one tap, or swap to an item in stock.

05 **The remedy, if the grid calls for one**

From chapter 11.

Templates are in Appendix B. This is a service message, not marketing, so it goes to every buyer. Keep promotions out of it, and send by SMS only to customers who gave you a number for order updates.

What the law requires

In the US, the FTC's Mail, Internet, or Telephone Order Merchandise Rule sets a floor under all of this. It was last amended in 2014 PUBLISHED . In summary:

- **You need a reasonable basis to ship on time:** within the time you state, or 30 days if you state none (50 if the order depends on a credit application). Shipping means into the carrier's hands, not delivery.
- **If you can't, you must ask.** Before the promised time passes, give a revised ship date (or say you don't know one) and the choice to consent or cancel for a prompt refund.
- **Silence counts as yes only for a first delay of 30 days or less.** For a longer or indefinite delay, or a second one, you need express consent; without it, you cancel and refund.
- **A prompt refund** means within seven working days, card payments included. One billing cycle applies only where you extend the credit yourself.
- **Email can carry the notice.** An order-status page alone fails if the buyer doesn't see it in time.

Violations can carry civil penalties per violation; the inflation-adjusted maximum is $53,088, as set in January 2025 PUBLISHED . A delivery date shown at checkout is also a claim under the FTC Act's ban on deception. States and other countries add their own rules. This is an operator's summary as of September 2026, not legal advice; have counsel review your preorder, backorder and delay flows.

PUBLISHED 16 CFR Part 435, as amended at 79 FR 55619 (September 17, 2014); FTC, Business Guide to the FTC's Mail, Internet, or Telephone Order Merchandise Rule. Full references in Appendix C.

Prove it works

Hold back 10% of delayed orders, at random, from the early notice for one month. They still get everything the law requires. Compare contacts per delayed order, cancellations and 90-day repeat rate. If the notice doesn't cut contacts, rewrite it before you blame the idea.

Do this

- **Build the five triggers** in your shipping platform or help desk, each firing the same notice with a new date and a choice.
- **Send your backorder and preorder flows to counsel** with the rule's summary above, and fix anything that ships late without asking.

PART TWO · PROMISE WHAT YOU CAN KEEP · CHAPTER 6

SHOW THE WORK

Customers value work they can see. A tracking page that shows real progress buys patience; a page that says “label created” for three days spends it.

Between the order confirmation and the doorstep, most brands go dark. The customer’s only window is a carrier page written in scan codes. Ryan Buell’s research at Harvard Business School explains why that silence is expensive.

The labor illusion

In experiments on travel and dating websites, Buell and Michael Norton found people valued a service more when it showed its work, even if that meant waiting longer for the same result. They called it the labor illusion PUBLISHED .

In food-service field experiments, Buell, Tami Kim and Chia-Jung Tsay let customers and cooks see each other. Customer-reported quality rose 22.2% and throughput times fell 19.2% PUBLISHED . And in Boston, with Ethan Porter, Buell found that residents who reported a problem through the city’s app and got a photograph of the fix submitted 60% more requests afterward; people shown the government’s work were 14% more trusting and 12% more supportive of it PUBLISHED . It worked because the city was actually responding. Transparency shows the work; it can’t replace it.

> **Transparency multiplies whatever it shows. Show real work and it earns patience. Show a stalled order and it earns a contact.**

What that means for an order

- **Own the tracking page.** On your domain, in your voice: the promised date in large type, each step in plain words, one clear way to get help.
- **Fill the dark days.** Tell customers when the order is picked, packed and handed over. “Packed at our Reno warehouse this morning” is the labor illusion working for you.
- **Translate the scan codes** into a sentence about what each status means for the date.
- **Show the problem when there is one.** If the date moves, the page should say so, with the same new date and choice as the notice in chapter 5.
- **Ask after delivery.** One question in the delivery email, “Did everything arrive as expected?” The no answers feed the index in chapter 3 and find damage customers would never have reported.

Transparency is not a stream of notifications. Send a message only when it changes what the customer expects or answers a question they’d ask. Four emails for an on-time order train customers to ignore the one about the delay. How the flows are built is in The Whole Machine.

Do this

- **Follow your last tracking link on a phone** and list the questions a worried customer would still have. Each is a line the page should answer.
- **Add the delivery check question** to the delivery email this week, and route every “no” to a same-day reply.

CONTACTS PER ORDER

Support teams are usually measured on how fast they answer. The number that matters is how often customers need to ask, and why.

Response times and handle times measure how well support copes with contacts. They say nothing about why the contacts exist. The best service, as Bill Price and David Jaffe titled their book, is no service: the contact that never has to happen.

The Amazon number

Price was Amazon's first global vice president of customer service. The metric his team put in place was contacts per order: every call, email and chat, divided by orders. By his own account, it fell 70% during his almost three years at the company REPORTED . Amazon kept watching it after he left. Jeff Bezos's letter to shareholders for 2002 called it "our most sensitive measure of customer satisfaction" and reported a 13% improvement that year REPORTED .

It's sensitive because almost nobody contacts a store for fun. A contact means something didn't work. Count contacts per order, split them by reason, and you have a map of every broken promise in the business, drawn by the customers who hit them.

> **Every contact is a customer telling you, at your expense, where the business broke a promise.**

The seven principles

Price and Jaffe's seven principles, in their order: eliminate dumb contacts; create engaging self-service; be proactive; make it easy to contact the company; own the actions across the company; listen and act; deliver great service experiences REPORTED . The order matters. Most brands start at the second, with a chatbot, which moves contacts to software without asking why they exist. Start at the first. Chapter 5 applies the third to delivery, and chapter 8 the fifth.

Reason codes that work

Most help desks have forty tags, applied inconsistently, several per ticket. What works:

- **One primary reason per contact,** chosen at close from twelve to fifteen, named by cause: "late against date shown," not "email inquiry."
- **Every contact tied to its order,** so contacts join to the failure index and carrier data.
- **Audited monthly.** Re-code twenty random contacts blind. If you disagree with agents on more than one in five, the list is unclear.

A starting list is in Appendix B.

RUN YOUR NUMBERS

What would fewer contacts save?

Example numbers. Replace with yours. Contacts means every inbound email, chat, call and message, counted once per conversation.

Orders per month: 10000

Contacts per month: 1800

Share about order status or lateness: 40 %

Share of those that notices and tracking remove: 50 %

Share of all other contacts removed by fixing causes: 15 %

Agent minutes per contact, including wrap-up: 8 min

Loaded cost per agent hour: $ 30

0.180 → 0.128	522	70	$25,056
contacts per order, now then after	contacts avoided per month	agent hours freed per month	saved a year in agent time

> **Removing 522 contacts a month takes contacts per order from 0.180 to 0.128 and frees about 70 agent hours a month, $25,056 a year. The larger prize is the customers who never had to ask: put the failures behind these contacts into the cost tool in chapter 2.**

Counts agent time only. It leaves out help-desk seat costs, the retention value of customers who never had to ask, and the contacts that reopen when the first answer didn't resolve the problem, so the real saving is larger.

Say a brand ships 10,000 orders a month and receives 1,800 contacts, four in ten about where an order is. With those defaults, contacts per order falls from 0.180 to 0.128 and about 70 agent hours a month are freed, worth about $25,000 a year in agent time alone DERIVED . That's modest next to the cost of the failures behind the contacts, which is the point: the support saving is the smaller half. Replace the made-up shares with your own reason counts before quoting it.

Deflection is not elimination

A help-center article that answers "where is my order?" deflects the contact. The customer still had the worry and still found your promise broken. Report contacts per order and self-service sessions per order side by side. If the second rises while the first falls, you've moved the problem, not solved it.

Do this

- **Replace your tag list** with twelve to fifteen primary reasons named by cause, starting from the list in Appendix B.
- **Report contacts per order by reason, weekly,** starting with the last twelve weeks.

PART THREE · CONTACTS ARE A SYMPTOM · CHAPTER 8

EVERY REASON HAS AN OWNER

Support answers the contact. The team that caused it has to remove it. Put a name, a target and a date next to every reason code.

Most contact reasons are decisions made somewhere else: the date by ecommerce, the packaging by operations, the discount rule by marketing. Support meets the consequences and has no power to change them. So each reason gets an owner in the team that causes it, measured on that reason's contacts per order.

The routing table

REASON	USUAL CAUSE	OWNER	FIRST FIX
Where is my order	Silence after checkout; a vague date	Ecommerce, with operations	A dated promise, an owned tracking page
Late against the date shown	A promise the lane can't keep	Operations	Promise by region; delay notice
Damaged	Packaging and handling	Operations and packaging	Drop-test the top five products

REASON	USUAL CAUSE	OWNER	FIRST FIX
Wrong or missing item	Picking errors	Fulfillment or the 3PL	Scan-to-verify at pack
Change or cancel an order	No self-serve edit window	Ecommerce	An edit window on the order page
Discount or code didn’t work	Unclear rules	Marketing	One rule per offer, printed with the code
Product question before buying	The page doesn’t answer it	Merchandising	Add the top three questions to the page
How to use the product	Missing instructions	Product	An insert and a first-use email
Unexpected charge	Descriptor, split charges, renewals	Finance or retention	A clear descriptor; renewal reminders
Return or exchange	Fit, expectations	Merchandising and ops	See the returns guide

Subscription contacts belong to whoever runs the program (The Standing Order); product-page questions overlap with The Honest Test; first-use questions with the first-use guide.

> **Support should own the answer. The team that caused the question should own the number.**

The monthly contact review

Once a month, for forty-five minutes, each owner brings one slide: their reasons’ contacts per order for three months, one fix shipped with its date and expected drop, and one planned. Support reads ten verbatim contacts from the top reason aloud. The next review checks whether each fix moved its number. A fix that didn’t isn’t a fix.

Give it a price

Owners move faster when a contact has a cost in their own budget. Put a notional price on each contact reason: the agent time from the tool in chapter 7, plus, for failures, the repeat contribution from chapter 2. Say the damaged reason runs 150 contacts a month; at $4 of agent time and $3.75 of lost repeat contribution each, that's about $1,160 a month charged to packaging. A sturdier box at $0.15 more on 10,000 orders costs $1,500. Now the owner has a real decision to make, with both sides priced.

Do this

- **Fill in the owner column** for your top ten reasons, with a name, not a team.
- **Book the first monthly review** and ask each owner for one fix, with its date and expected drop.

PART THREE · CONTACTS ARE A SYMPTOM · CHAPTER 9

EFFORT, MEASURED CAREFULLY

Making things easy for customers is the right goal. The famous single-question survey built to measure it predicts retention poorly. Measure effort by what customers had to do.

In 2010 Matthew Dixon, Karen Freeman and Nick Toman published “Stop Trying to Delight Your Customers” in Harvard Business Review. Its argument has held up. Its metric hasn’t.

The argument

From a study of more than 75,000 people who had used contact centers or self-service, the authors concluded that going above and beyond made little difference to loyalty; customers wanted their problem solved simply. Among customers who reported low effort, 94% said they intended to buy again, 88% that they’d spend more, and only 1% that they’d speak badly of the company REPORTED . Their prescriptions: head off the next likely problem, stop making customers switch channels, attend to how the customer feels, learn from those who struggled, and put solving ahead of speed. A customer who writes twice or repeats their order number to a second agent has done work that was yours to do.

The metric, checked

The article proposed a Customer Effort Score, one survey question on how much effort the customer spent, reported to predict loyalty better than satisfaction or the Net Promoter Score. That claim failed an independent test. Evert de Haan, Peter Verhoef and Thorsten Wiesel compared the three metrics on customers of 93 firms in 18 industries and tracked who stayed. The effort score “in itself has little to no predictive power and performs the worst of all” the metrics studied, even among customers who had actually contacted the company. Top-two-box satisfaction, the share who gave one of the two highest scores, predicted retention best PUBLISHED .

> **Reduce effort as a design rule. Don’t adopt a one-question effort score as your loyalty metric.**

The authors’ caution is about scope: the effort question asks about one past interaction, while retention depends on the whole relationship, including the delivery that caused the contact. The contact is a symptom; its effort is a symptom of the symptom.

Measure effort by behavior

Your help desk already records effort:

MEASURE	DEFINED AS	WHAT IT CATCHES
Repeat contact rate	Contacts about the same order within 7 days of a previous one, over all contacts	First answers that didn't resolve it
Contacts per resolution	Contacts, over issues marked solved	Back-and-forth to get one thing done
Channel switches	Issues touched in more than one channel, over all issues	Chat that sends people to email, email that sends them to phone
Transfers	Contacts handed to a second person, over all contacts	Agents without the authority to finish

Repeat contacts are FedEx's "complaints reopened," weighted 5 in its index and in the one in chapter 3. If you survey after contacts, report the top-two-box share, which predicted retention best in de Haan's study.

Do this

- **Add repeat contact rate and transfers** to the weekly support report, beside contacts per order.
- **Read twenty repeat contacts** to find their top source. Usually it's a missing permission, like an agent who can't issue a reship, or missing data, like the carrier scan.

THE RECOVERY PARADOX, CHECKED

A great recovery can leave a customer more satisfied than no failure at all. It doesn't leave them more likely to buy again. Plan on recovery softening the loss, never reversing it.

Every support team has a story about the customer whose order went wrong, got an extraordinary fix, and became the brand's biggest fan. The research calls it the service recovery paradox, and it's used to justify treating failures as opportunities. The evidence says treat them as costs.

What the meta-analysis says

In 2007, Celso de Matos, Jorge Henrique and Carlos Rossi pooled the studies that had tested the paradox. The effect was real for satisfaction: on average, customers who had a failure and a good recovery rated their satisfaction higher than customers who had no failure. For repurchase intentions, word of mouth and the company's image, the pooled effect was not significant: no evidence of a paradox at all PUBLISHED . Even the satisfaction effect varied with study design, such as whether subjects were students.

The field studies since then point the same way. Stefan Michel and Matthew Meuter tested the paradox with more than 11,000 interviews with bank customers about real service encounters. It was a rare event, and where it appeared the differences were small PUBLISHED . And the restaurant study in chapter 2, which followed behavior over years, found recovered customers drifting down toward those whose problems were never fixed PUBLISHED .

> **Recovery raises how customers feel about the fix. It doesn't raise how often they come back above what they'd have done if nothing broke.**

Why it matters for the budget

If you believed the paradox, you'd spend freely on recovery and worry less about prevention. Since it doesn't hold for repurchase, the math runs the other way: recovery wins back part of the repeat contribution a failure puts at risk, and prevention protects all of it. Set recovery spending against the lost repeat contribution per failure from the cost tool, and prevention spending against the whole cost. Recovery isn't optional; an unrecovered failure is worse in every study above. It's something to do well, at a known cost, not something to celebrate.

Do this

- **Add a recovery column to your failure cohort:** repeat rates with and without a remedy, beside the no-failure rate. Expect recovered customers to sit between.
- **Stop reporting recovery stories as wins** in the weekly meeting. Report them as costs, next to the failure that caused them.

PART FOUR · MAKING IT RIGHT · CHAPTER 11

THE RECOVERY GRID

What each failure gets, written down, so recovery doesn't depend on who picks up or how loudly the customer

complains. Plus a budget each agent can spend without asking.

Without a grid, recovery is set by things that shouldn't matter: which agent answers and how hard the customer pushes. The polite customer with a damaged order gets less than the furious one with a late order. The grid fixes that; the budget handles what it can't foresee.

Four rules for the grid

01 **Fix first, then remedy**

A reship or replacement goes out the same day, before any credit.

02 **Scale with severity and history, not volume**

A lost order gets more than a late one; a second failure more than a first. Shouting gets nothing extra.

03 **Don't make cheap failures prove themselves**

If a replacement costs less than the agent time spent asking for photos, skip the photos. Watch for abuse in the data.

04 **Record every remedy against the customer**

So the next failure is treated as a repeat. Chapter 12 explains why.

A starting grid

FAILURE	FIX	REMEDY, FIRST FAILURE	REMEDY, REPEAT WITHIN 12 MONTHS
Late 1 to 2 days, notified first	New date	None beyond the notice	Shipping refunded
Late 1 to 2 days, not notified	New date, apology	Shipping refunded	Credit of about 10% of the order
Late 3+ days, or past a date that mattered	Upgrade or reship if faster	Credit of about 10% of the order	Credit of about 20%, from a person
Split without warning	Say what's coming, when	None unless late	Shipping refunded
Wrong or missing item	Ship the right item now	Credit of about 10%	Credit of about 20%
Damaged	Replace now	Credit of about 10%	Refund the item too
Lost	Replace now, before the carrier claim	Credit of about 15%	Refund the order too

These are starting points, not findings. Set them so the expected remedy per failure sits below the lost repeat contribution per failure from chapter 2. A credit toward the next order gives a reason to come back, and the Uber study found apologies worked best with a future-trip credit attached. Keep credits tied to failures, or customers learn to complain for them; The First Offer covers how discounts train customers.

The agent budget

A grid can't foresee the gift that arrived after the birthday or the third failure in a month. For those, give each agent a monthly budget to spend without asking anyone.

Size it from the value at stake. Say an agent handles 150 failure contacts a month, each putting $3.75 of repeat contribution at risk, as in the example in chapter 2: about $560 a month. A budget of a fifth to a third of that, $110 to $190, covers the cases that matter most DERIVED .

- **A per-case limit,** say twice the grid remedy, so one case can't empty the month.
- **Review for patterns, not policing.** Recurring uses become grid rows.
- **No rollover.** It's not a savings target; a team that underspends is probably under-recovering.

> **A written grid makes recovery fair. A budget makes it human. You need both.**

Do this

- **Pull last month's remedies** by failure type and agent. The spread between agents is the argument for the grid.
- **Publish the grid and the budget** to the team in one page, with the date you'll review both.

PART FOUR · MAKING IT RIGHT · CHAPTER 12

HOW TO APOLOGIZE

A field experiment with 1.5 million Uber riders tested apologies for late trips. Words alone did little. A credit helped. Apologizing again and again made things worse.

Most apology templates are written by instinct: say sorry, say it warmly, promise it won't happen again. A very large field experiment suggests the

instinct is wrong about the last part.

The Uber experiment

Basil Halperin, Benjamin Ho, John List and Ian Muir worked with Uber on riders whose trips arrived later than the app had estimated. Riders were randomly assigned to messages like these, some carrying a $5 credit toward a future ride, or to no message at all PUBLISHED :

TYPE	THE MESSAGE
Basic apology	“Your trip took longer than we estimated, and we know that’s not ok.”
Status apology	“We underestimated how long your trip would take, and that’s our fault.”
Commitment apology	“We’re working hard to give you arrival times that you can count on.”
No apology	“You have places to go and people to see. Enjoy $5 off your next ride.”

PUBLISHED Excerpts from each message, as reported in the paper. Outcome: riders’ net spending over the next twelve weeks.

Three findings matter. First, an apology in words alone had little effect, and was sometimes counterproductive. Second, “money speaks louder than words”: the messages that came with a credit did best, and the credit alone raised riders’ net spending by about 1.5% over the first week and about 0.8% over twelve weeks. Third, apologizing repeatedly to the same rider after repeated bad trips reduced their future spending. By the third late trip, an apology with a credit had a significantly negative effect PUBLISHED . The authors’ advice: use apologies sparingly, and ideally only after outcomes that were unexpectedly bad.

The commitment apology deserves its own warning. The paper names apologies that promise to do better, with repeated ones, as cases where apologizing can be worse than sending nothing PUBLISHED . A promise to improve is another promise, and a rider who’s late again has now seen two broken.

> **An apology is a promise about the future.**
> **Don’t make one you haven’t already kept.**

Seven rules for an apology

01 **Lead with the failure, in plain words**

"Your order arrived damaged," not "sorry for any inconvenience."

02 **Take responsibility without excuses**

Name the cause in one sentence, if you know it.

03 **Say what's already been done**

"A replacement shipped this morning." The fix before the feelings.

04 **Attach the remedy from the grid**

Words alone did little at Uber.

05 **Promise only what you've already fixed**

"We've changed how we pack glass" is fine if you have. "It won't happen again" isn't.

06 **Check the history before you send**

If this customer got an apology in the last few months, escalate to a person instead of repeating the template.

07 **Keep it short**

Four or five sentences. A long apology reads as a performance.

Rule 6 needs the failure history that the grid in chapter 11 records. Templates are in Appendix B.

Do this

- **Rewrite your apology templates** against the seven rules. Delete every "won't happen again" you can't back up.
- **Add a failure count to the customer record** in your help desk, visible on every ticket, and a rule that a second failure in 90 days goes to a senior agent.

PART FIVE · RUNNING IT · CHAPTER 13

PAYING FOR THE PROMISE

In 1995 Continental Airlines promised a monthly bonus to every hourly employee, all 35,000 of them, if the company hit one shared goal. Theory said a bonus shared that widely shouldn't work. It did.

When Gordon Bethune became chief executive of Continental in October 1994, the airline ranked last in most of the industry's performance measures REPORTED . He told the story of the turnaround in his book From Worst to First. The best evidence on why one part of it worked comes from two economists who studied it afterward.

What Continental did

In February 1995, Continental introduced an incentive scheme that promised monthly bonuses to all 35,000 of its hourly employees in any month the company achieved a firm-wide performance goal PUBLISHED . It was one goal for the whole company, not a goal for the gate agents or the mechanics: one number for everybody, paid to everybody.

On paper, that shouldn't work. With 35,000 people sharing a goal, one employee's effort barely affects whether the bonus is paid, so theory predicts people will coast: the free-

rider problem.

Why it worked

Marc Knez and Duncan Simester studied the scheme and found it did raise employee performance. Their explanation: employees worked in autonomous groups, like the crew at one airport, where people could see each other's work and pressed each other to do their part. They call it mutual monitoring. The firm-wide bonus gave every group the same reason to care, and the small group gave each person someone watching PUBLISHED . Continental went on from last in most performance categories to winning more J.D. Power customer satisfaction awards than any other airline REPORTED .

> **Pay everyone on the same kept promise, in teams small enough to see each other work.**

What a DTC brand can take from it

- **One shared number.** The perfect-order rate or the index from chapter 3, not a goal per department. A warehouse paid on speed and support paid on response time each hit their number while the order goes wrong between them.
- **Monthly, paid close to the month.** A bonus paid next February doesn't change a Tuesday in April.
- **Small teams that see each other,** like a packing shift or a support pod, so mutual monitoring can work.
- **A number that's hard to game.** On-time alone can be bought with worse packing; the weighted index can't.
- **Include the 3PL,** with the index in the contract and a monthly credit or bonus attached.

Does it pay?

Say 25 people each get $75 in any month the perfect-order rate beats target: $22,500 a year if it's hit every month. At the cost tool's default, where each point of failure rate costs about $18,900 a year, the bonus pays for itself if it takes about 1.2 points off the failure rate, say

from 6% to 4.8% DERIVED . Publish the number daily where the team can see it, the way FedEx sent its SQI to every site.

Do this

- **Pick the one number** and publish it daily for a month before attaching money, so the team trusts the count.
- **Price the bonus against the cost tool** and propose it at the next budget meeting, with the failure-rate reduction it needs to pay for itself.

PART FIVE · RUNNING IT · CHAPTER 14

THE PROMISE SCORECARD

One page, every week. Each number with its denominator, split by the lane where it's worst, and one quarterly line that says what failures cost in repeat orders.

Delivery usually shows up in the weekly meeting as shipping cost per order, and support as response time. Neither says whether the business kept its promises. This page does.

NUMBER	DEFINED AS	WHAT IT CATCHES
Perfect-order rate	**Orders with no failure, over orders delivered**	**The headline**
Failure index	**Points per 1,000 orders, top two sources, worst lane**	**Which failure to fix first**
On time against the date shown	Delivered by the checkout date, by region	Promises the lane can’t keep
Ship on time	With the carrier by the promised ship date	Warehouse delays, early
Order-to-door days	Median and 90th percentile, by region	A lengthening tail
Told first	Delayed orders notified before any contact	Delay triggers that stopped firing
Contacts per order	Total, and for each of the top five reasons	A new reason appearing, a fix that didn’t hold
Repeat contact rate	Same-order contacts within 7 days	Effort: answers that didn’t resolve
Delivery check “no” rate	“No” answers to the delivery check	Damage and errors customers didn’t report
Recovery spend	Remedies and budget use per failure	Drift in what failures cost to fix
Repurchase after failure (quarterly)	12-month repeat rate by failure type, against no failure	What each failure costs in the orders that follow

> **A number without its denominator is a mood, and an average without its worst lane is an alibi.**

Two rules for the page

First, every rate shows its worst lane beside it: the carrier, region or product where it's worst. Customers on that lane don't experience the average. Second, watch the "told first" line. A delay trigger that fires zero times in a week isn't a good week; it's usually a broken integration.

Reading it

If the perfect-order rate improves while the index worsens, fewer orders fail but those that do fail worse. Contacts lag failures by about a week. When you change something, write the date on the page.

Do this

- **Build the page this week** with whatever you can pull, even if half the cells are blank. Put a name next to each blank. The page gets complete by being used.

PART FIVE · RUNNING IT · CHAPTER 15

THE FIRST THIRTY DAYS

Measure, then promise, then route the contacts, then write down how to make it right. Four weeks, in that order.

The order of work is the same whoever you are. Find out which promises you're breaking. Stop making the ones you can't keep. Send each contact back to its cause. Then make recovery consistent.

01

Week one: measure

Pull last month's orders with the date shown at checkout and the carrier delivery scan, and compute on time against the date shown, by region (chapter 4). Count last month's failures by type and run the index (chapter 3). Replace the help desk's tags with twelve to fifteen primary reasons (chapter 7).

02

Week two: promise

Run the date tool for each region and change the checkout promise to what the data supports. Build the five delay triggers and the notice (chapter 5). Send your backorder, preorder and delay flows to counsel with the FTC rule summary. Add the delivery check question and fix the tracking page's worst gaps (chapter 6).

03

Week three: route

Name an owner for each of the top ten contact reasons (chapter 8). Hold the first monthly contact review. Add repeat contact rate and transfers to the support report (chapter 9). Start the 10% holdout on delay notices.

04

Week four: make it right

Publish the recovery grid and the agent budget (chapter 11). Rewrite apology templates and add a failure count to the customer record (chapter 12). Run the failure cohort for customers from a year ago (chapter 2) and put the scorecard in front of the team (chapter 14).

At day thirty you won't yet know what the changes did to repeat orders; that takes a quarter. You'll have a store that knows which promises it breaks, tells customers first, and knows who owns each reason they write in.

> **Measure before you promise. Promise before you apologize.**

Do this

- **Put the four weeks on the calendar today,** with an owner for each line. If the same person owns the checkout date, the warehouse and the help desk, give them this guide.

CLOSE

DAY ONE

Six things whoever owns the promise needs on the first day.

Whoever owns delivery reliability and support, a new operations lead or you on the Monday you decide the ticket queue is telling you something, needs six things on day one.

01 **Orders with the promise attached**

Every order with the checkout date shown, the ship scan and the delivery scan.

02 **Carrier and 3PL access**

The shipping platform, the 3PL portal and the contracts.

03 **A help desk export**

Six months of contacts, with reason, order, remedy and reopens.

04 **Contribution per order**

From finance, so a failure and a remedy can be priced.

05 **The policies, with their history**

Shipping, delay, backorder and recovery rules, with the dates each changed.

06 **Authority to change the promise**

The right to change the checkout date and switch on delay notices without waiting for a meeting.

Do this

- **Ask for all six in one message** today. Whatever comes back slowest is where the business knows least about the promises it makes.

CLOSE

THE SHELF

The books and papers worth reading next, and what to take from each.

The Best Service Is No Service

Bill Price and David Jaffe, 2008

The case for contacts per order, from the man who ran Amazon's early customer service. Read it for the seven principles and the idea that every contact reason has an owner.

From Worst to First

Gordon Bethune, 1998

The Continental turnaround in the chief executive's own words. Read it for how one shared, visible number changed an organization, and read Knez and Simester for why.

The Effortless Experience

Matthew Dixon, Nick Toman and Rick DeLisi, 2013

The book-length version of the effort argument. Take the design rules; be skeptical of the single-question score, for the reasons in chapter 9.

“The Psychology of Waiting Lines”

David Maister, 1985

A short essay that explains why an unexplained, uncertain wait feels longer than a known one. The theory behind every delay notice.

“Toward an Understanding of the Economics of Apologies”

Basil Halperin, Benjamin Ho, John List and Ian Muir, 2022

The Uber field experiment. Read the message wording and the results on repeated apologies before you write another template.

“Service Recovery Paradox: A Meta-Analysis”

Celso de Matos, Jorge Henrique and Carlos Rossi, 2007

The paper to hand anyone who says a failure is an opportunity. Satisfaction, yes. Repurchase, no.

“Operational Transparency”

Ryan Buell, Harvard Business Review, 2019

The practitioner’s summary of the research behind chapter 6, on why showing the work makes customers value it more.

The research behind each chapter is listed in Appendix C.

CLOSE

ABOUT THE AUTHOR

Andrew Lauchner is a growth and retention operator for consumer brands and the author of sixteen free books on the method. He fixes the second-purchase problem through his practice, Growth Legend, as one engagement: The Second-Purchase Sprint. He is the author of The Second Order, on turning first-time buyers into second-time buyers, and The Whole Machine, on the fundamentals of DTC growth, along with a series of field guides for DTC operators at andrewlauchner.com.

As Senior Director of Growth and Retention Marketing at Gallery Furniture, he rebuilt the customer journey and the sales playbooks together. He has worked on growth and retention at Binance and 3Commas, and has been Head of Growth and Retention at Greatness Wins and at Nexus Agriscience.

WHAT COLLEAGUES SAY

“Andrew led retention, lifecycle, and email/SMS, but what separates him from most in this space is how deeply he understands the role retention plays in the overall growth engine.”

Akram Khan, Head of Marketing at Gallery Furniture, senior to Andrew but didn’t manage Andrew directly

Andrew answers every note from operators working on this. Write to andrew@growthlegend.com or message him on LinkedIn.

APPENDIX A

FOR YOUR ANALYST

The formulas behind the four calculators, and four queries every store should be able to run.

The formulas

FOR	FORMULA	NOTES
Cost per failure	$r \times d \times n \times c + x$	r: repeat rate, no failure. d: relative drop. n: orders per returner. c: contribution. x: direct cost.
Failure index	$\Sigma\ (count_i \times weight_i)\ /\ orders \times 1{,}000$	Perfect-order rate = 1 – failed orders / orders.
On time at a promise of D days	$\Phi((\ln D - \ln m)\ /\ \sigma), \sigma = (\ln p90 - \ln m)\ /\ 1.2816$	m: median days. p90: 90th percentile. Lognormal fit; Φ is the standard normal CDF.
Days to promise for target t	$\lceil\ m \times e^{z_t\ \sigma}\ \rceil$	z_t: normal quantile for t (1.645 for 95%).
Contacts avoided	$C \times w \times r + C \times (1 - w) \times x$	C: contacts. w: status share. r: share removed. x: other share removed.

Check the lognormal fit against last month’s actual on-time share.

On time against the date shown

```
-- share of delivered orders that arrived by the date shown at checkout
-- orders.promised_delivery_date must be stored when the order is placed
SELECT o.shipping_region,
       f.carrier_service,
       COUNT(*)                                                    AS delivered,
       AVG(CASE WHEN f.delivered_at::date <= o.promised_delivery_date
                THEN 1.0 ELSE 0 END)                               AS on_time_share,
```

```
       PERCENTILE_CONT(0.5) WITHIN GROUP
         (ORDER BY EXTRACT(EPOCH FROM f.delivered_at - o.created_at) / 86400) AS median_da
       PERCENTILE_CONT(0.9) WITHIN GROUP
         (ORDER BY EXTRACT(EPOCH FROM f.delivered_at - o.created_at) / 86400) AS p90_days
FROM orders o
JOIN fulfillments f ON f.order_id = o.id
WHERE f.delivered_at >= CURRENT_DATE - INTERVAL '30 days'
GROUP BY o.shipping_region, f.carrier_service
ORDER BY on_time_share;
```

The syntax is Postgres. If you don't store the promised date, start today; until then, reconstruct it from the shipping method and order time, and call the result an estimate. The median and 90th-percentile columns feed the date tool in chapter 4.

The failure index

```
-- one row per order with its worst failure, for the index and perfect-order rate
WITH flags AS (
  SELECT o.id AS order_id,
         CASE WHEN f.delivered_at IS NULL AND f.lost_at IS NOT NULL THEN 'lost'
              WHEN EXISTS (SELECT 1 FROM tickets t WHERE t.order_id = o.id
                           AND t.reason = 'damaged')                  THEN 'damaged'
              WHEN EXISTS (SELECT 1 FROM tickets t WHERE t.order_id = o.id
                           AND t.reason = 'wrong_or_missing')         THEN 'wrong_or_miss
              WHEN f.delivered_at::date >= o.promised_delivery_date + 3   THEN 'late_3plus
              WHEN f.delivered_at::date >  o.promised_delivery_date       THEN 'late_1_2'
              WHEN o.split_unannounced                                    THEN 'split'
              ELSE 'none' END AS failure
  FROM orders o JOIN fulfillments f ON f.order_id = o.id
  WHERE o.created_at >= DATE_TRUNC('month', CURRENT_DATE) - INTERVAL '1 month'
    AND o.created_at <  DATE_TRUNC('month', CURRENT_DATE)
)
SELECT failure, COUNT(*) AS orders
FROM flags GROUP BY failure;
```

Count reopened contacts separately and add them in the tool. Each order counts once, at its worst failure, which keeps the perfect-order rate honest.

The failure cohort

```
-- 12-month repeat rate of first-time customers, by what happened to the first order
WITH firsts AS (
  SELECT DISTINCT ON (customer_id) customer_id, id AS order_id, created_at,
         shipping_region, first_product_id
  FROM orders ORDER BY customer_id, created_at
)
SELECT DATE_TRUNC('month', fo.created_at) AS cohort_month,
       fo.shipping_region,
       fl.failure,
       COUNT(*) AS customers,
       AVG(CASE WHEN EXISTS (
             SELECT 1 FROM orders o2
             WHERE o2.customer_id = fo.customer_id
               AND o2.created_at > fo.created_at
               AND o2.created_at <= fo.created_at + INTERVAL '12 months')
           THEN 1.0 ELSE 0 END) AS repeat_12m
FROM firsts fo
JOIN flags fl ON fl.order_id = fo.order_id   -- the flags logic above, run for these orde
WHERE fo.created_at < CURRENT_DATE - INTERVAL '12 months'
GROUP BY 1, 2, 3
ORDER BY 1, 2, 3;
```

Compare each failure group with the ‘none’ group within the same month and region, then average the gaps, weighted by group size. Add whether a remedy was given to get the comparison in chapter 10. Pool quarters until groups reach a few hundred.

Contacts per order by reason

```
-- weekly contacts per order, by primary reason
SELECT DATE_TRUNC('week', t.created_at) AS week,
       t.reason,
       COUNT(*)::numeric / NULLIF(w.orders, 0) AS contacts_per_order
FROM tickets t
JOIN (SELECT DATE_TRUNC('week', created_at) AS week, COUNT(*) AS orders
      FROM orders GROUP BY 1) w
  ON w.week = DATE_TRUNC('week', t.created_at)
WHERE t.created_at >= CURRENT_DATE - INTERVAL '12 weeks'
  AND t.direction = 'inbound'
GROUP BY 1, 2, w.orders
ORDER BY 1, 3 DESC;
```

Table and column names are generic; rename them to match your exports. Count conversations, not messages. Repeat contact rate follows the same pattern: contacts with an earlier contact on the same order in the previous 7 days, over all contacts.

APPENDIX B

TEMPLATES

Delay notices, apologies, reason codes and the delivery check. Copy them into whatever your team already uses.

These messages are transactional: about an order the customer placed, sent to every buyer, with no promotion. Keep your standard footer with your address and a preferences link. Have counsel review the option notice.

Delay notice, email

```
SUBJECT    Your Order Is Running Late
PREVIEW    new date inside, and your options

Hi [first name],

Your order [#1234] is running late. It should now arrive
by [Thursday, October 8], instead of [Monday, October 5].

What happened: [the carrier missed its pickup from our
warehouse on Tuesday].
```

```
What you can do:
  Wait for it          Nothing to do. We'll write again if
                       the date moves.
  Cancel for a refund  [Cancel order] (one tap, full refund
                       within 7 working days)
  Swap it              [Choose something in stock]

[Only if the recovery grid calls for it:]
We've added a [$X] credit to your account for your next order.

[Name], [Brand] customer care
Reply to this email to reach a person.

[Brand, postal address] | Email preferences
```

Delay notice, SMS

```
[Brand]: Your order #1234 is running late. New date: Thu Oct 8.
Wait, cancel for a refund, or swap: [short link]. We're sorry
for the delay. Reply STOP to opt out.
```

Send by SMS only to customers who opted in to order updates by text.

Can’t ship on time: the option notice

```
SUBJECT    We Can't Ship Your Order On Time
PREVIEW    please choose: wait or cancel

Hi [first name],

We can't ship [item] from order [#1234] by [the date we
promised]. We now expect to ship it by [revised date].
[Or: We don't know yet when we can ship it.]

You can:
  Wait     [I'll wait]
  Cancel   [Cancel for a full refund]

If you cancel, we'll refund [$X] within [7 working days].

[If the delay is 30 days or less and this is the first delay:]
If we don't hear from you, we'll ship it when it's ready.
```

```
[If longer, indefinite, or a second delay:]
If we don't hear from you by [date], we'll cancel this item
and refund you in full.

[Brand, postal address] | Email preferences
```

Apology with remedy, first failure

```
SUBJECT    Your Order Arrived Damaged. A Replacement Is On Its Way.
PREVIEW    shipped this morning, arriving by [date]

Hi [first name],

Your [item] arrived damaged. That's on us.

A replacement shipped this morning and should arrive by
[date]. You don't need to send anything back.

We've also added a [$X] credit to your account for your
next order.

[What we've changed, only if it's true: We've moved this
item to a sturdier box.]

[Name], [Brand] customer care

[Brand, postal address] | Email preferences
```

Primary contact reasons

```
WHERE IS MY ORDER          no update, customer asking
LATE AGAINST DATE SHOWN    past the checkout date
DAMAGED                    arrived unusable
WRONG OR MISSING ITEM      any line wrong, short or absent
LOST                       never arrived, or marked delivered
CHANGE OR CANCEL ORDER     after checkout, before shipping
DISCOUNT OR CODE           didn't apply, or unclear
PRODUCT QUESTION, BEFORE   pre-purchase question
PRODUCT USE, AFTER         how to use, results, fit
BILLING                    unexpected or unclear charge
SUBSCRIPTION               skip, change, cancel
RETURN OR EXCHANGE         start, status, refund
```

```
ACCOUNT                    login, details
OTHER                      review monthly; split anything over 5%
```

The delivery check

```
IN THE DELIVERY CONFIRMATION EMAIL
  "Did everything arrive as expected?"
  [Yes]  [No, something's wrong]

"No" opens a short form: damaged / wrong or missing /
not received / other, with an optional photo.
Every "no" gets a same-day reply and a grid remedy.
```

APPENDIX C

SOURCES

Every external source, by chapter. Web sources were read in September 2026.

What failures cost (chapters 2 and 10)

- Shashank Rao, Stanley E. Griffis and Thomas J. Goldsby. “Failure to deliver? Linking online order fulfillment glitches with future purchase behavior.” Journal of Operations Management 29(7–8), 2011, 692–703.
- Basil Halperin, Benjamin Ho, John A. List and Ian Muir. “Toward an Understanding of the Economics of Apologies: Evidence from a Large-Scale Natural Field Experiment.” Economic Journal 132(641), 2022, 273–298. NBER working paper 25676.

- Alice Harter, Lucas Stich and Martin Spann. “The Effect of Delivery Time on Repurchase Behavior in Quick Commerce.” Journal of Service Research 28(2), 2025.
- Tim Norvell, Piyush Kumar and Mayukh Dass. “The Long-Term Impact of Service Failure and Recovery.” Cornell Hospitality Quarterly 59(4), 2018, 376–389.
- M. Serkan Akturk, Rakesh R. Mallipeddi and Xingzhi Jia. “Estimating impacts of logistics processes on online customer ratings: Consequences of providing technology-enabled order tracking data to customers.” Journal of Operations Management 68(6–7), 2022, 775–811.
- Celso Augusto de Matos, Jorge Luiz Henrique and Carlos Alberto Vargas Rossi. “Service Recovery Paradox: A Meta-Analysis.” Journal of Service Research 10(1), 2007, 60–77.
- Stefan Michel and Matthew L. Meuter. “The service recovery paradox: true but overrated?” International Journal of Service Industry Management 19(4), 2008, 441–457.

The failure index (chapter 3)

- National Institute of Standards and Technology. Malcolm Baldrige National Quality Award 1990 recipient profile: Federal Express Corporation.
- Federal Express case study, Exhibit 4, “Service Quality Indicators” (copy read; publisher not stated in the copy).
- Federal Highway Administration. “The Use of Operations Objectives and Performance Measures in Private and Public Organizations,” FHWA-HOP-10-029, citing D. D. Morris and R. J. Baker, “Measuring What Matters,” California CPA, August 2003.
- Quirk’s. “Federal Express uses an ongoing study to track customer satisfaction,” November 1990.

Speed and the promise (chapter 4)

- Marshall L. Fisher, Santiago Gallino and Joseph Jiaqi Xu. “The Value of Rapid Delivery in Omnichannel Retailing.” Journal of Marketing Research 56(5), 2019, 732–748.
- Vinayak Deshpande and Pradeep K. Pendem. “Logistics Performance, Ratings, and Its Impact on Customer Purchasing Behavior and Sales in E-Commerce Platforms.” Manufacturing & Service Operations Management 25(3), 2023, 827–845.
- Shin Oblander and Kinshuk Jerath. “The Nonlinear Impact of Promised Delivery Time on Online Purchasing.” Manufacturing & Service Operations Management 28(4), 1051–1067; published online September 2025.
- Ruomeng Cui, Zhikun Lu, Tianshu Sun and Joseph M. Golden. “Sooner or Later? Promising Delivery Speed in Online Retail.” Manufacturing & Service Operations Management 26(1), 2024, 233–251. Emory Goizueta summary.
- Nooshin Salari, Sheng Liu and Zuo-Jun Max Shen. “Real-Time Delivery Time Forecasting and Promising in Online Retailing: When Will Your Package Arrive?” Manufacturing & Service Operations Management 24(3), 2022, 1421–1436.

Delay notices and the law (chapter 5)

- David H. Maister. “The Psychology of Waiting Lines.” In The Service Encounter, 1985.
- Shirley Taylor. “Waiting for Service: The Relationship between Delays and Evaluations of Service.” Journal of Marketing 58(2), 1994, 56–69.
- Angela M. Legg and Kate Sweeny. “Do You Want the Good News or the Bad News First? The Nature and Consequences of News Order Preferences.” Personality and Social Psychology Bulletin 40(3), 2014, 279–288.
- Federal Trade Commission. Mail, Internet, or Telephone Order Merchandise Rule, 16 CFR Part 435, as amended at 79 FR 55619, September 17, 2014; Business Guide to the FTC’s Mail, Internet, or Telephone Order Merchandise Rule, including the 2025 civil penalty amount.

Transparency (chapter 6)

- Ryan W. Buell and Michael I. Norton. “The Labor Illusion: How Operational Transparency Increases Perceived Value.” Management Science 57(9), 2011, 1564–1579.
- Ryan W. Buell, Tami Kim and Chia-Jung Tsay. “Creating Reciprocal Value Through Operational Transparency.” Management Science 63(6), 2017, 1673–1695.
- Ryan W. Buell, Ethan Porter and Michael I. Norton. “Surfacing the Submerged State: Operational Transparency Increases Trust in and Engagement with Government.” Manufacturing & Service Operations Management, 2021. INFORMS release, November 2021.
- Ryan W. Buell. “Operational Transparency.” Harvard Business Review, March–April 2019.

Contacts and effort (chapters 7 to 9)

- Bill Price and David Jaffe. The Best Service Is No Service. Jossey-Bass, 2008. Interview with Bill Price on the seven principles, CustomerThink.
- Bill Price. “For a Driving-Force Metric, Consider CPX.” CustomerThink, August 2005.
- Amazon.com. 2002 letter to shareholders.
- Matthew Dixon, Karen Freeman and Nick Toman. “Stop Trying to Delight Your Customers.” Harvard Business Review, July–August 2010. Figures as quoted by MeasuringU.
- Evert de Haan, Peter C. Verhoef and Thorsten Wiesel. “The predictive ability of different customer feedback metrics for retention.” International Journal of Research in Marketing 32(2), 2015, 195–206. Full text.

Continental (chapter 13)

- Marc Knez and Duncan Simester. “Firm-Wide Incentives and Mutual Monitoring at Continental Airlines.” Journal of Labor Economics 19(4), 2001, 743–772.
- Gordon Bethune. From Worst to First: Behind the Scenes of Continental’s Remarkable Comeback. Wiley, 1998.

- Wikipedia. Continental Airlines, on Bethune’s appointment in October 1994 and the airline’s rankings and J.D. Power awards.

THE KEPT PROMISE

Andrew Lauchner · September 2026 · The Whole Machine · The Second Order · Close the Loop · The First Offer · The Standing Order · The Honest Test · Free tools · andrewlauchner.com

Published figures come from peer-reviewed research, books and essays by the researchers themselves, and official law and regulator guidance, as of the dates given. Reported figures come from named companies, authors and outlets, including executives describing their own companies. Derived figures are my arithmetic, with inputs shown. Examples that open with Say or Picture use made-up round numbers. Shipping, consumer-protection and messaging rules vary by state and country and change often; have counsel review your delay, backorder and refund flows. None of this is legal or financial advice.

Privacy