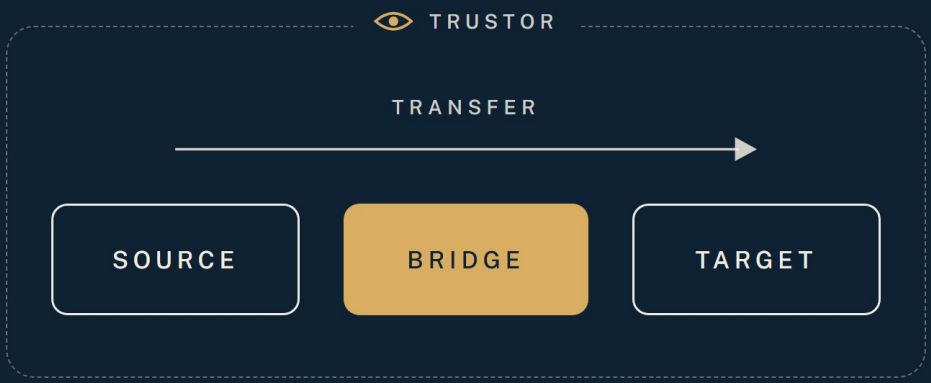
Trust Transfer

A Unified Theory of How Trust Moves Through Networks, Brands, and Institutions



VIVEKA MOHAN DAS

Mindful Machines Press

TRUST TRANSFER

A Unified Theory of How Trust Moves Through Networks, Brands, and Institutions

Viveka Mohan Das

Mindful Machines Press

Copyright © 2026 Viveka Mohan Das

All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.

Published by Mindful Machines Press

First Edition, 2026

ISBN: 978-1-7649186-1-9 (ebook)

An earlier version is available on SSRN: https://doi.org/10.2139/ssrn.7086358

This book is a work of nonfiction. The views, interpretations, and conclusions expressed are those of the author and do not necessarily reflect the views of Mindful Machines Press.

Mindful Machines Press

Australia

I acknowledge the Dharawal people, the Traditional Custodians of the land on which this book was written, and pay my respects to their Elders past, present and emerging.

Contents

Author's Note

The Model in Two Minutes

PART I. The Pattern: How Trust Moves

- Chapter 1: Why We Trust What We Don't Know
- Chapter 2: What Trust Transfer Is
- Chapter 3: One to One
- Chapter 4: The Borrowed Name
- Chapter 5: Many to One
- Chapter 6: The Tempo of Trust

PART II. The Domains: Where Trust Is Borrowed

- Chapter 7: Brands and Influencers
- Chapter 8: Platforms, Search, and Algorithms
- Chapter 9: Machines We Never Inspect: Trust Through the Surroundings
- Chapter 10: Outcomes of Target-Based Evidence

PART III. The Breaks: Failure and Response

- Chapter 11: Borrowed Legitimacy
- Chapter 12: Mistrust Contagion
- Chapter 13: The Contain–Repair–Monitor Model

PART IV. Building Better Trust: Design and the Machine Frontier

- Chapter 14: Designing Trust Signals
- Chapter 15: Verification at Scale
- Chapter 16: A Field Guide for the Trustor
- Chapter 17: The Machine Problem: Can Trust Transfer to AI?

In Closing: What to Carry Forward

Appendix: A Decision Sheet for Organisations

Key Terms and How to Use Them

References

List of Figures and Tables

Acknowledgements

Author's Note

The framework in this book isn't the one I started with. The first version, posted as a working paper, put the Trustor at the end of the chain — the last stop, after trust had already passed through a brand, a platform, or an endorsement. That followed the literature it grew out of: Stewart's 2003 Trust Transfer model, and the trust-formation work behind it from Mayer, Davis, and Schoorman in 1995 and McKnight, Choudhury, and Kacmar in 2002. In the diagram I drew from it, belief travelled down a line and settled on whoever was standing at the end. That placement never sat right with me. It turned the person extending trust into a destination — someone trust simply arrived at, once everyone upstream had finished their work.

That isn't how trust behaves. The person deciding whether to rely on something isn't standing at the end of a line, waiting their turn. They're watching the whole chain as it unfolds, weighing the Bridge, deciding how far to lean before anyone has proven anything. I wanted a model that kept the Trustor in charge of that judgement throughout — not a slot the machinery fills in at the very end.

That's where the Eye of the Trustor came from — the one piece of this model I haven't found in the earlier literature, though I wouldn't claim no one has proposed something similar. In the diagrams I started from, the Trustor sat at the end of the line. Earlier researchers knew that people interpret what they are told; what I wanted was a model that keeps that interpretation in view at every step. Here it runs the whole way through: it decides what counts as a credible Source, how a Bridge gets read, and how much Vulnerability — how much is at stake, of what kind, how reversible — the Trustor accepts before the Target has proven anything at all. Earned Trust and Earned Mistrust are what that exposure can earn once the evidence comes in, and sometimes the honest answer is still "not yet". Trust was never something deposited on the Trustor at the finish line.

Getting there took more drafts than I'd like to admit. I moved the Trustor around the diagram more times than I can count before I saw that the fix wasn't a better position in the chain. It was a different relationship to the chain altogether — inside the model's logic, outside its sequence.

If you come across the original working paper — posted to SSRN as "Trust Transfer: How Credibility Moves Through Networks, Brands, and Machines" (DOI: 10.2139/ssrn.7086358, 2026) — you'll recognise that earlier architecture. This book is where it finally made sense to me. By "unified" I mean one organising synthesis of research that is usually told in separate pieces. It is a way of seeing and checking, not a theory this book has tested.

The Model in Two Minutes

Most of us regularly trust people and things we do not know well. We try a tradesperson because a friend recommends them, choose a product because it appears on a familiar website, or use a new tool because a trusted organisation provides it.

You try a new café because a friend recommends it. Your friend is the Source, the recommendation is the Bridge, and the café is the Target. You are the Trustor: the person deciding whether to rely on that connection. Your friend's experience gives you a reason to try the café; it does not guarantee that you will like it. After a few visits, you have evidence of your own. You might rely on the café more, rely on it less, or still be unsure. This book asks what justifies each step.

Source → Bridge → Target

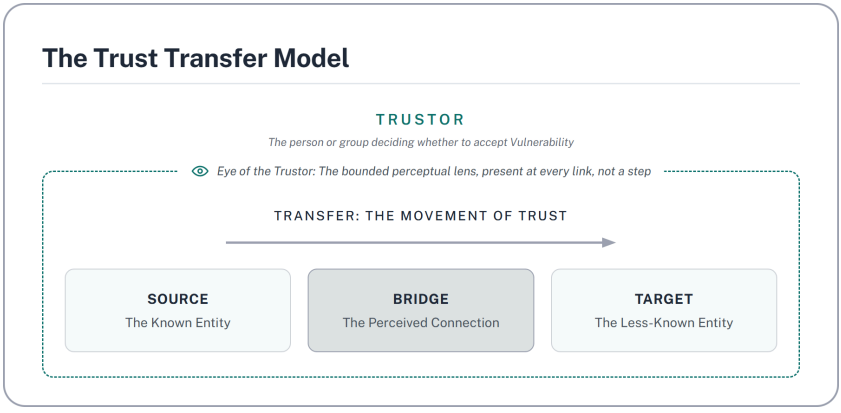


Figure 0.1 The Trust Transfer Model. Trust transfers from a known Source, across a perceived Bridge, to a less-known Target. The Trustor reads all three through a bounded lens the book calls the Eye of the Trustor.

The book calls your point of view at every step the Eye of the Trustor: what you notice depends on what you know and what you stand to lose. At first, your confidence is borrowed. Once you deal with the Target directly, its conduct begins to supply its own evidence. That experience may produce Earned Trust or Earned Mistrust, or leave you unsure for now.

The central question is simple: What does this connection genuinely give me a reason to believe?

The full Key Terms, with a short question for each, are at the back of the book. You do not need them to begin.

Part I

The Pattern: How Trust Moves

Why trust is transferred rather than built, the components of the model, and the mechanics of single links, borrowed names, converging Sources, and tempo.

Chapter 1

Why We Trust What We Don't Know

It is easy to imagine that trust is built slowly: one careful judgement, one experience, one verified fact at a time. In much of life, that is not how it begins.

We make decisions about people, products, platforms, experts, institutions, and systems that we do not directly know. We choose a specialist because a doctor refers us. We try an unfamiliar product because it appears on a marketplace we already use. We accept a recommendation because it comes through a school, a workplace, a friend, a publication, a regulator, or an interface that feels familiar. We do not begin from certainty. We begin from connections that make an unknown thing seem safe enough to approach.

This is a structural feature of ordinary life, not evidence of careless judgement. No one can audit every balance sheet, inspect every supply chain, test every technical system, read every contract, or establish every claim from first principles. Requiring direct verification before every decision would bring ordinary life to a halt.

People will rely on signals from elsewhere because they must. The work is to see what a signal actually establishes, who stands behind it, and where its assurance ends.

When Confidence Vanished

For a time, FTX did not look like the sort of company people should fear. It looked like the sort of company that had already passed someone else's test.

That appearance was powerful because the underlying activity was difficult to inspect. The Securities and Exchange Commission later stated that Bankman-Fried had presented FTX to investors as a safe and responsible crypto-asset trading platform with sophisticated, automated measures to protect customer assets. The SEC alleged that, while raising more than $1.8 billion from investors, he concealed the diversion of customer funds to Alameda Research, special treatment for Alameda on the platform, and risks connected to Alameda's holdings (SEC, 2022).

The collapse exposed the distance between appearance and underlying conduct. In November 2022, Reuters reported, citing two people familiar with the matter, that at least $1 billion of customer funds had vanished from FTX and that internal records showed roughly $10 billion had been moved to Alameda Research; the report also recorded Bankman-Fried's disagreement with its characterisation of that transfer (Berwick/Reuters, 2022). The later criminal

outcome was not a matter of allegation alone. In March 2024, the U.S. Department of Justice reported that Bankman-Fried had been found guilty after trial, sentenced to twenty-five years in prison, and ordered to pay $11 billion in forfeiture for multiple fraudulent schemes involving the misappropriation of billions of dollars of FTX customer funds (U.S. Department of Justice, 2024). FTX is an extreme case used because breakdown makes the movement of trust unusually visible; it is not presented as a typical outcome of Borrowed Confidence.

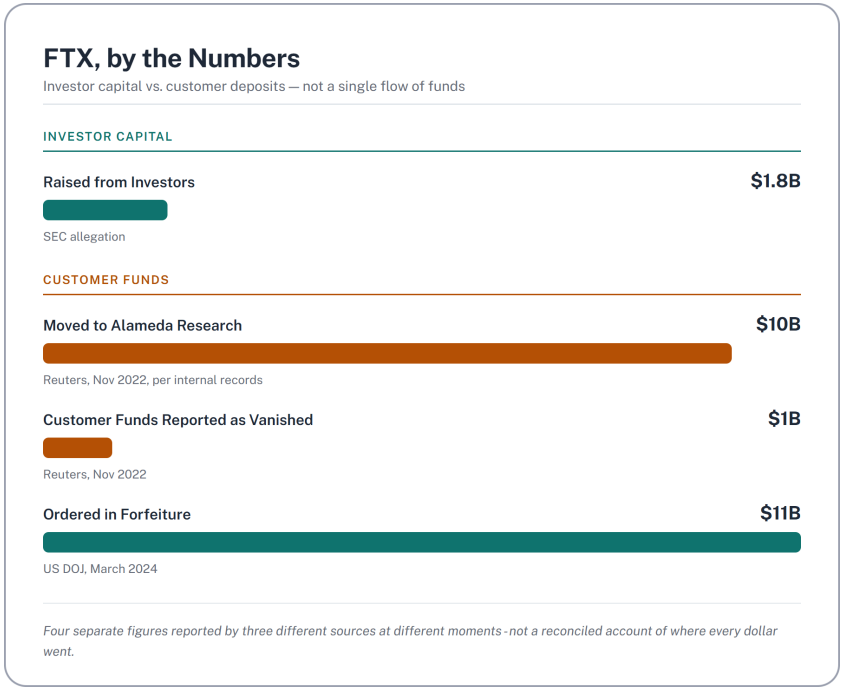


Figure 1.1 FTX, by the Numbers. Four figures from three separate sources, reported at three different moments — a record of scale, not a reconciled account of where the money went.

Most decisions made through Borrowed Confidence do not end in criminal collapse. They work silently: a recommendation saves time, a professional referral helps someone find care, a familiar marketplace makes an unfamiliar seller easier to consider. FTX reveals what can happen when the assurances surrounding an unfamiliar organisation are allowed to do more work than the organisation's own record can bear.

The people who used FTX did not all make the same judgement for the same reason, and the investors, public figures, and institutions associated with the company did not all play the same role. No single relational map can capture that history without concealing its ambiguities. What matters here is simpler:

when a difficult-to-inspect organisation appears alongside credible people, established institutions, visible safeguards, or familiar settings, those associations can lower the threshold for reliance long before Direct Evidence is available.

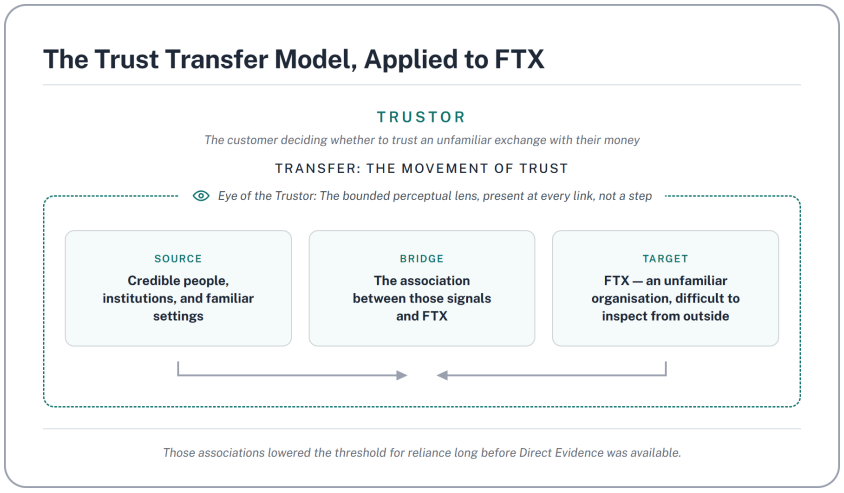


Figure 1.2 The Trust Transfer Model, Applied to FTX. Those associations lowered the threshold for reliance long before Direct Evidence was available.

Why Borrowed Confidence Is Practical

A person who accepts a referral is not usually surrendering judgement. They are using a practical answer to a practical problem. Someone else may have more experience, a better vantage point, a recognised responsibility, or a reputation worth protecting. A familiar setting may make an unfamiliar option easier to assess. A well-designed service may make it clear what has been checked, by whom, and for what purpose.

Full independence is not a realistic alternative. For many ordinary decisions, it would be paralysing. A parent cannot test every educational tool a child might use. A patient cannot independently evaluate every clinical instrument placed inside a hospital system. A buyer cannot inspect the history of every seller on a platform before making an ordinary purchase. People therefore look for cues that tell them where prior knowledge, oversight, or accountability might reside.

Some of those cues are substantive. A professional referral may reflect expertise. A certification may indicate that a recognised body has assessed a defined standard. A trusted institution may be accountable for the systems it places in front of people.

Other cues are more ambiguous. A prominent ranking may reflect popularity rather than quality. A badge may refer to identity, payment, compliance, or something much narrower than the reader assumes. A sleek interface can make a weak claim feel settled. A partnership can mean close scrutiny, a limited commercial arrangement, or little more than public co-presence.

That difference determines whether a reader receives useful orientation or merely a borrowed feeling of safety. When a cue does not disclose what it means, who stands behind it, or where its limits lie, confidence can travel farther than the available evidence warrants.

Why This Matters More Now

Trust Transfer has always existed. Family names mattered before brands did. Seals and uniforms mattered before verified badges and blue checkmarks. Religious, professional, and civic institutions have carried confidence between people for centuries. What has changed is the environment in which that movement now happens.

Scale. Digital systems let one Source project a signal across millions of Targets at once. A platform can rank, label, recommend, or suppress at a reach no single referral ever had. A well-known figure can redirect attention instantly. One interface pattern can shape millions of judgements in a day.

Speed. Confidence can now form before reflection catches up. A person sees the badge, the ranking, or the recommendation, and acts, often within seconds.

Opacity. Many of the relationships that matter most are hard for an outsider to inspect. Why was this result shown first? Why was this seller marked trusted? Was this recommendation earned, purchased, inferred, or simply predicted from past behaviour? Often the person relying on the signal has no way to know, and the signal still works anyway.

Layering. Confidence used to move through one or two visible intermediaries. Now it often travels through a chain: a person trusts a platform because they trust the device it runs on, trusts a creator because the platform recommends them, and trusts a product because the creator endorses it. Each layer adds confidence, even when the relationship underneath any single link is thin.

That is why a modern trust environment can look sturdy right up to the moment it fails.

The Hidden Work of Signals

Modern life is full of small design choices that perform large interpretive work. A recommendation label, a crest, a search position, a familiar logo, a default setting, a verified badge, or a mention in a respected publication can all change the way an unfamiliar option is encountered. Each may suggest that someone else has already done some of the work of checking.

The suggestion may be warranted, partial, or misleading.

Signals compress complex histories into quick impressions. They can make a reader feel that several questions have already been answered: Who is vouching for this? What was checked? What incentives shape the connection? What remains uncertain? What happens if the thing being presented fails? Under time pressure, the impression often arrives before the question.

Compression makes complex societies more usable, but it creates a vulnerability. A signal can be accurate and still be over-read. A Bridge can be real and still be too weak to justify the confidence placed on it. An institution can be respected in one domain and assumed, without evidence, to have guaranteed another. Efficiency then conceals the difference between a meaningful connection and a decorative one.

Placement, certification, and endorsement make different claims. A search ranking may say only that an option was surfaced. A certification may say that a defined standard was met, not that the option is right for every decision. An endorsement tells the reader that someone is lending a name or reputation; it does not reveal the extent of that person's knowledge. Responsible design keeps those distinctions visible.

Reliance should be proportionate to what a signal can actually bear. That is different from refusing every recommendation, distrusting every institution, or starting every decision from scratch.

When Assurance Proves Too Thin

A visible failure can expose more than the failure itself. It can reveal the relationships through which people came to rely on the thing that failed. Questions then spread quickly: Who introduced it? Who appeared to stand behind it? What did the platform, institution, or expert actually know? Which assurance was evidence, and which was merely atmosphere?

This is why the cost of misplaced confidence is rarely confined to one organisation. When a connection is over-read, disappointment can strain the wider relationships around it. A trusted expert may lose credibility after endorsing an unworthy product. A marketplace may become suspect after

repeatedly surfacing unsafe goods. A regulator may be asked to account for the meaning of its approval after harm becomes visible. In larger networks, doubt can reach adjacent organisations and categories that were not directly responsible for the original failure.

The direction and reach of that doubt depend on what failed and how the surrounding relationships were understood. Chapter 12 examines that movement in full. For the present argument, the relevant point is that Borrowed Confidence can enlarge the consequences of a later failure beyond its immediate facts.

What to Notice from Now On

The next time something feels credible, pause before treating that feeling as a conclusion. Where is the confidence actually coming from, and what connects the familiar person, institution, or setting to the unfamiliar thing now being offered? What has actually been checked, who had the responsibility to check it, and what remains unknown?

Is the connection visible and meaningful, or merely implied — and does it reflect expertise, accountability, payment, popularity, placement, convenience, or some mixture of these? What would change if the reassuring connection turned out to be weaker than it first appeared?

These questions do not require cynicism. They make confidence more legible. They help the Trustor distinguish a signal that provides real orientation from one that borrows more authority than it has earned.

Organisations that make introductions, platforms that rank and label, designers who decide what becomes visible, regulators who certify, and institutions that allow their reputation to appear beside someone else's claim all shape the confidence that may move through that connection.

What This Book Will Do

The chapters ahead develop one argument across the four parts of this book.

Part One establishes the pattern. It shows how confidence moves from a known Source, across a Bridge, towards a less-known Target; how the Trustor interprets that relationship; and how the strength, independence, and timing of a Bridge shape what it can reasonably support.

Part Two follows that pattern into commercial, platform, and machine environments, where signals can be repeated and scaled far beyond a single referral. It concludes with the point at which the Target becomes directly observable and Borrowed Confidence gives way to Earned Trust or Earned Mistrust, or remains unresolved.

Part Three turns to failure. It examines how a Target can borrow more legitimacy than its Bridge warrants, how doubt spreads through the relationships that once carried confidence, and what a credible response must stop, correct, and test over time.

Part Four turns the framework into practice. It develops principles for designing honest trust signals, examines how verification operates across whole markets, gives Trustors a practical method for reading the Bridges around them, and tests the framework at the machine frontier.

The promise is simple: to explain why people rely on things they cannot fully verify, how to judge what the connections around them genuinely establish, and what happens when confidence is earned, overextended, or broken.

Chapter Takeaway

Much of modern trust begins before direct knowledge is possible. Familiar people, institutions, and settings make unfamiliar things seem safe enough to use. That Borrowed Confidence is neither naive nor corrupt by definition; it is one way a complex society stays usable.

The danger begins when a signal makes uncertainty feel resolved. The less clear the connection, the more easily confidence can outrun the evidence that should support it.

Chapter 2 now gives this familiar experience a precise vocabulary. It asks what happens when confidence moves from something already known toward something not yet known, and how to inspect the connection before it carries more than it can bear.

Chapter 2

What Trust Transfer Is

The problem is established: people must make decisions before they can know enough directly. They rely on recommendations, institutions, familiar settings, visible affiliations, and signals that suggest someone else has already done part of the work of checking.

This chapter gives that experience a grammar.

Trust Transfer is the movement of confidence from a known Source toward a less-known Target across a perceived Bridge. The movement is interpreted through the Eye of the Trustor: the bounded perceptual lens through which the Trustor (the person or persons whose judgement is at stake) reads this particular trust situation. The lens stays in place after the crossing, reading the Target's own conduct as closely as it read the signal that pointed there.

The model is deliberately spare:

Source → Bridge → Target

Transfer is not a fourth part at the end of the chain. It is the movement that becomes possible when the Bridge makes the Source relevant to the Target. Before that movement, the Source has a reputation and the Target has an uncertain record. After it, the Target may appear more credible, safer, more legitimate, or simply more worth trying than it would have appeared alone.

That appearance may be useful. It is not yet proof.

Direct Evidence and Borrowed Confidence

Confidence can form from Direct Evidence. You return to a café because you have eaten there repeatedly and the experience has met the standard you care about. You rely on a colleague because you have watched their judgement under pressure. In each instance, the Target's own conduct has become available for assessment.

Borrowed Confidence begins elsewhere. You try the café because a friend whose taste you respect recommends it. You give the new colleague responsibility because a manager you trust selected and praised them. You have not yet learned enough about the café or the colleague to rely solely on their own record. Instead, you use evidence about someone or something already known as provisional orientation.

The difference is not whether confidence feels real. Both forms can guide action. The difference is where the confidence is resting. Direct Evidence rests

principally on the Target's observed conduct. Borrowed Confidence rests, at least initially, on the Bridge between a known Source and a less-known Target.

This distinction matters whenever a decision must be made before direct knowledge is available. It lets the reader ask a more precise question than whether they trust something. The better question is: what am I trusting through, and what does that connection actually establish?

The Three-Node Chain

The Source is the known entity from which confidence originates. It may be a person, a brand, an institution, a regulator, a marketplace, a publication, a familiar interface, or a system with an established record. "Known" does not mean universally reliable. It means known enough, in the situation at hand, to carry some existing credibility.

The Target is the less-known entity whose credibility is being assessed. It may be a person, product, organisation, claim, service, tool, or decision. "Less-known" is relational rather than absolute. A well-known company can still be a Target when it enters an unfamiliar domain; a new feature inside a familiar service can still be a Target when its own conduct has not yet been observed.

The Bridge is the perceived connection that makes the Source relevant to the Target. A Source and a Target can exist side by side without transfer occurring. The Bridge is what gives their association meaning: a recommendation, a certification, a ranking, a shared setting, a formal integration, or another connection that a reader takes to signify something.

Consider a reader who sees an unfamiliar book recommended by a critic whose judgement they know. The critic is the Source. The book is the Target. The recommendation is the Bridge. Confidence may move because the reader interprets the recommendation as evidence that the critic has read, judged, and chosen the book. The reader may buy it before reading a page.

The same grammar can describe an apparently more technical situation. A hospital adds a new digital service to its patient portal. The hospital may be the Source, the new service the Target, and the portal integration the Bridge. The patient may reasonably infer that the hospital has made some decision about the service. The critical question is not whether that inference is automatically correct. It is what the integration actually means: clinical validation, procurement approval, a limited administrative arrangement, or merely technical availability.

The chain makes that question visible.

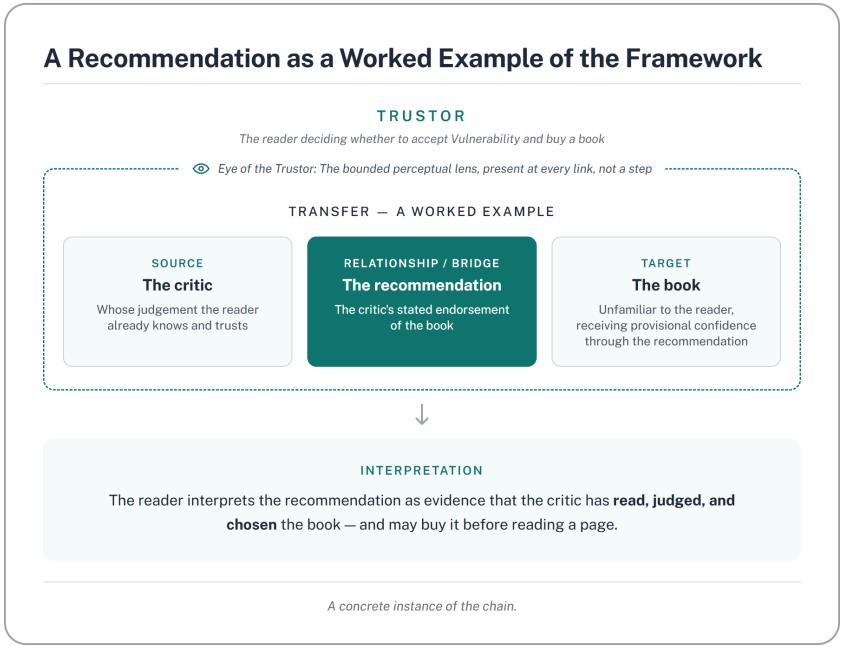


Figure 2.1 A recommendation as a worked example of the framework. A critic's recommendation can carry provisional confidence from a known Source towards an unfamiliar book. The reader still needs Direct Evidence from the Target before confidence becomes earned.

Roles are relational, not permanent labels. The same organisation may be a Source in one instance and a Target in another; a person within an organisation may also create a separate, linked Bridge rather than occupy one fixed place in every account. When a case contains nested Bridges, the model should show more than one chain or state the ambiguity. It should not force a complex situation into a single clean assignment simply because the diagram is simple.

The Eye of the Trustor Is a Bounded Lens

The Trustor is the person or persons interpreting the Source, the Bridge, and the Target, and deciding whether to accept Vulnerability and transfer trust to the Target. What they see through is not neutral. This book calls that bounded perceptual lens the Eye of the Trustor. It is the piece of the model this book adds rather than carries over from existing trust-transfer research, at least among the sources reviewed here.

The lens is bounded because it belongs to a particular trust situation. It is shaped by the decision at hand, the information available, the relationship involved, the alternatives open to the person, and the consequences if the judgement is wrong. It is not a fixed personality type or a total worldview. The

same person may interpret an institutional certification differently when choosing a low-cost household product than when consenting to a medical procedure, because the risks, consequences, and available evidence differ.

The lens is present throughout the chain. It shapes what counts as a credible Source, whether a Bridge appears meaningful, which uncertainties feel tolerable, and what evidence the Target must later provide. This does not make every judgement arbitrary. A certificate either identifies a defined standard or it does not; a ranking either discloses its basis or it does not. The lens identifies the situation in which those facts acquire practical weight, and it is what the Trustor is reading when the moment comes to accept Vulnerability and transfer trust to the Target, or decline to do either.

The Trustor interprets the Bridge. The Bridge can still be more or less legible, proportionate, and accountable. Neither cancels the other: the reading is the Trustor's, and those qualities belong to the Bridge.

Vulnerability is what turns that interpretation into a stake. The stronger a Bridge feels, the more a Trustor may accept: more money committed, more responsibility handed over, less independent checking done before acting. That accepted exposure does not disappear once the crossing is made — it sets the terms for what happens next. When the Target finally becomes observable, Direct Evidence lands against however much Vulnerability was accepted going in. The same conduct from a Target can register as a minor confirmation after a small bet, and as full Earned Trust or Earned Mistrust after a large one. Vulnerability, in other words, is what gives Earned Trust and Earned Mistrust something to be earned against.

Vulnerability varies along three lines. The first is size: how much is on the table. The second is kind: money, health, time, reputation, or the safety of someone else. The third is reversibility: whether a bad outcome can be undone. A friend's recommendation of a wine and the same friend's recommendation of a surgeon use the same Source and the same form of Bridge. The exposure is nowhere near the same. A disappointing bottle costs an evening. A poor operation may cost years, and there is no refund.

The person reading the Bridge is not always the person who carries the risk. A parent choosing a childcare centre reads the accreditation certificate on the wall, but the child spends the day there. A hospital purchasing officer reads a supplier's quality mark, but patients meet the product. When the exposure falls on someone other than the Trustor, checking can get lighter, because the cost of a wrong call is felt second-hand. That is a possibility to watch for, not a rule: plenty of parents check harder for a child than they would for themselves. The model therefore asks two questions: who is doing the accepting, and who will be hurt if the Target fails?

People do not always scale their checking to their exposure. The same row of gold stars can get the same quick glance whether it sits beside a forty-dollar hotel room or an eight-thousand-dollar kitchen renovation. That is where many of the failures in this book begin. The Bridge was about the right size for a small bet and was asked to carry a large one. The section on Bridge-Strength below takes up that match directly.

Bridge Quality: Orientation Is Not a Guarantee

A good Bridge gives a reader usable orientation. It makes the basis of the connection legible: what was checked, by whom, for which purpose, under what standard, and with what accountability. It does not need to resolve every uncertainty. Its responsibility is to avoid making uncertainty look settled when it is not.

A weak Bridge obscures those questions. It can borrow the appearance of scrutiny without disclosing its scope. A badge may confirm identity while being read as a quality endorsement. A partnership may be narrow while being read as a full guarantee. A sponsored position may look like a merit-based ranking. A familiar interface may make a new tool seem institutionally approved when it is simply available for use.

The failure is often an over-reading of the relationship. The Source may have done less than the person interpreting the Bridge assumes. The Target may borrow more legitimacy than the Source intended to lend. The Bridge can be real and still be too weak to support the confidence placed on it.

For that reason, the central diagnostic questions are modest: what is the connection, what has actually been checked, what does the Source stand behind, and what remains for the Target to show?

Those questions do not demand that a reader start every decision from nothing. They make delegated confidence more proportionate to the available evidence.

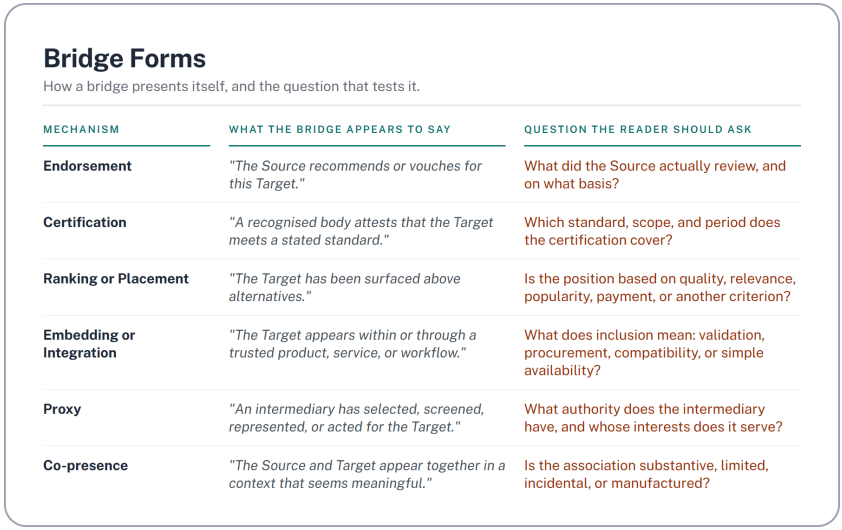
Bridge-Strength

Bridge-Strength is how much weight a Bridge can bear: the degree of reliance its form and basis actually justify. A Certification audited against a published standard carries more than Co-presence, where a Source and a Target simply appear together. Strength comes from the form of the Bridge and the work behind it, not from how prominent the signal looks. It is distinct from Bridge Quality, which concerns whether that basis is made legible in the first place.

What Makes a Bridge Consequential

A Bridge can take many forms, but the book uses six names to keep their differences visible. They are overlapping working categories, not an exhaustive classification. They are types of Bridge, not additional nodes in the chain.

Endorsement: someone lends a judgement. Certification: a body attests to a stated standard. Ranking or placement: an ordering or position is offered as a reason to look. Embedding or integration: a Target sits inside a system the Trustor already uses. Proxy: someone is taken to have checked on the Trustor's behalf. Co-presence: Source and Target merely appear together. These are forms of the Bridge, not extra nodes. A badge, a default, or a logo can wear any of them. Strength comes from the form and the work behind it, not from how prominent the signal looks.



▼ Text version of this table

Mechanism	What the Bridge appears to say	Question the reader should ask
Endorsement	"The Source recommends or vouches for this Target."	What did the Source actually review, and on what basis?
Certification	"A recognised body attests that the Target meets a stated standard."	Which standard, scope, and period does the certification cover?
Ranking or Placement	"The Target has been surfaced above alternatives."	Is the position based on quality, relevance, popularity, payment, or another criterion?
Embedding or Integration	"The Target appears within or through a trusted product, service, or workflow."	What does inclusion mean: validation, procurement, compatibility, or simple availability?
Proxy	"An intermediary has selected, screened, represented, or acted for the Target."	What authority does the intermediary have, and whose interests does it serve?
Co-presence	"The Source and Target appear together in a context that seems meaningful."	Is the association substantive, limited, incidental, or manufactured?

Table 2.1 Bridge Forms. The visible form of a connection does not by itself establish its scope, basis, or accountability.

These forms often overlap. A marketplace may rank a seller, display a verification mark, hold payment in escrow, and place the seller inside a familiar interface. A reader does not experience those functions one by one. They experience an overall impression that someone else has made the seller safer to consider.

The analytical task is to separate the impression into its parts. A ranking is not necessarily an endorsement. Certification is not a blanket guarantee. Integration is not proof that every feature has been independently tested. Co-

presence may be a meaningful association, a limited commercial arrangement, or an attempt to borrow attention without accountability.

The distinction is practical because people often extend more confidence than the Bridge can justify. They see a badge and infer competence. They see a familiar logo and infer responsibility. They see a high placement and infer a considered recommendation. Each inference may be sensible, weak, or wrong depending on what the Bridge actually carries.

A Brief Stress Test

The same grammar becomes especially useful when an unfamiliar service is encountered inside a familiar institutional setting. Consider an employee opening their company's benefits portal during annual enrolment. The portal is where payroll, leave, and superannuation already live, and the employee already trusts it for those. This year, it recommends a single private health insurer as the default option.

The employer is a strong Source. The portal is the Bridge. The recommended insurer is the Target. But what has actually been established? Being the default option can mean the employer negotiated a bulk discount, that a benefits broker earns a commission for placement, that the insurer met a minimum compliance threshold, or that someone in HR genuinely reviewed the market and chose the best plan. Each of those is a different Bridge wearing the same interface.

The employee rarely has the means to tell these apart from inside the portal. The default is doing real work: most people will not compare five insurers once one is already pre-selected. But what that default actually certifies stays invisible from where the employee is standing.

Chapter 9 will take up automated systems directly, where the same problem recurs at greater scale and with even less visibility into the basis of the recommendation. For now, the point is narrower: a Target is rarely encountered in a vacuum. The organisation, workflow, interface, and setting around it can all become Bridges that shape a first judgement before Direct Evidence accumulates.

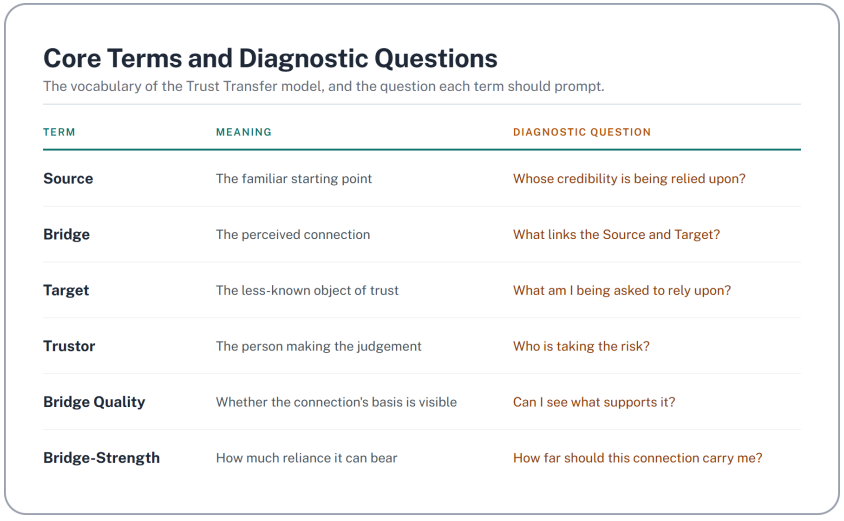
From Borrowed Confidence to Direct Evidence

Borrowed Confidence is not a mistake by definition. It is a way of beginning under uncertainty. Yet the Target's own conduct eventually matters. A recommendation can make a new service worth trying; it cannot permanently replace evidence of how the service performs. A certification can establish compliance with a stated standard; it cannot answer every question a particular

user has. An institutional setting may support provisional reliance; it cannot settle whether the Target will merit continued confidence.

When the Target becomes directly observable, that evidence can support more reliance, justify less, or leave the judgement unresolved for now. Chapter 10 names and develops these outcomes in full.

A different path is also possible. Failure can send doubt away from the failed link through surrounding Bridges and networks. That is Mistrust Contagion. Its direction depends on what failed: Source, Target, Bridge, or network. Chapter 12 develops that typology; Chapter 2 only needs to establish that Borrowed Confidence can create consequences beyond the first encounter.



▼ Text version of this table

Term	Meaning	Diagnostic question
Source	The familiar starting point	Whose credibility is being relied upon?
Bridge	The perceived connection	What links the Source and Target?
Target	The less-known object of trust	What am I being asked to rely upon?
Trustor	The person making the judgement	Who is taking the risk?
Bridge Quality	Whether the connection's basis is visible	Can I see what supports it?
Bridge-Strength	How much reliance it can bear	How far should this connection carry me?

Table 2.2 Core Terms and Diagnostic Questions. The vocabulary of the Trust Transfer model, and the question each term should prompt.

Chapter Takeaway

Trust Transfer has one shape: a known Source, a perceived Bridge, and a less-known Target, all read through the Eye of the Trustor before any Vulnerability is accepted. That shape does not tell a reader whether a given Bridge deserves the confidence placed on it. It only tells them where to look.

The risk this chapter has traced is not that people rely on Bridges. It is that a Bridge can feel persuasive before anyone has checked what it is actually built from, so placement gets mistaken for recommendation and availability for validation.

The framework does not decide whether trust is correct. It shows where confidence came from, what carries it, and what remains for the Target to prove.

Chapter 3 takes the simplest possible Bridge, one person telling another where to look, and asks what a referral can honestly establish, and what it cannot.

Chapter 3

One to One

The smallest consequential form of Trust Transfer is familiar. A colleague says, “Use this accountant.” A neighbour says, “Call this plumber.” A friend says, “You should speak to her before you decide.” The person receiving the recommendation may know very little about the accountant, plumber, or adviser. Yet the recommendation changes what happens next. An unknown option becomes worth considering.

That is not a trivial social courtesy. It is a compact Endorsement Bridge.

A referral does not make the Target known. It changes the conditions under which the Target is first approached. Instead of beginning with a list of unfamiliar alternatives, the recipient begins with one option that has arrived through someone whose judgement, experience, or relationship already carries some weight. The referral supplies orientation while Direct Evidence is still scarce.

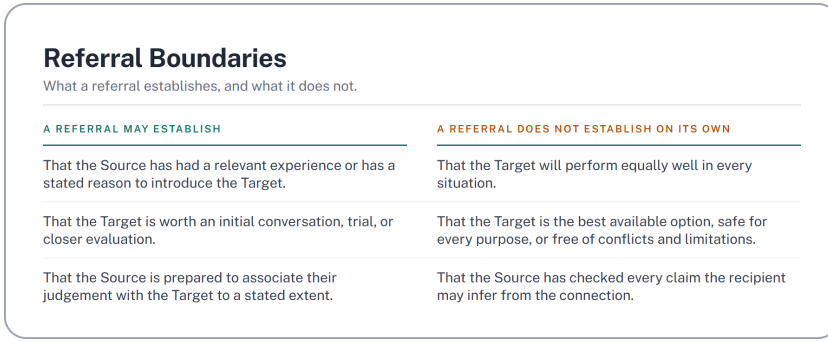
A referral is an Endorsement Bridge: one known Source, one less-known Target, read by the recipient.

The Trustor here is the recipient, reading the referral through everything at stake in this particular decision: what they already know, what they stand to lose, how much the Source’s experience actually applies. The same recommendation can feel throwaway when the decision is low-stakes and decisive when time is short or the alternatives are hard to judge.

What a Referral Can Establish

A referral is often stronger than a loose association, but it is narrower than a general certificate of quality. “I know this person” may establish contact or co-presence. “I have used this person for this work, and this is what I found” gives the recipient more: an identifiable experience, a stated basis for the recommendation, and a starting point for a decision. The difference is not semantic. It determines whether the Bridge carries a judgement or merely an affiliation.

The scope of that judgement should stay visible.



▼ Text version of this table

	A referral may establish	A referral does not establish on its own
1	That the Source has had a relevant experience or has a stated reason to introduce the Target.	That the Target will perform equally well in every situation.
2	That the Target is worth an initial conversation, trial, or closer evaluation.	That the Target is the best available option, safe for every purpose, or free of conflicts and limitations.
3	That the Source is prepared to associate their judgement with the Target to a stated extent.	That the Source has checked every claim the recipient may infer from the connection.

Table 3.1 Referral Boundaries. A referral can provide useful orientation without becoming a guarantee of fit, quality, safety, or future performance.

A referral can therefore be both meaningful and incomplete. It is meaningful when it tells the recipient where some relevant experience begins. It is incomplete because the recipient's situation may differ, the Target's performance may change, and important features of the decision may never have been within the Source's view. This boundary is not a defect in referral. It is the condition that keeps provisional confidence proportionate.

A Recommendation Changes the First Question

Before a referral, the first question is often broad: where should I begin? After one, the question becomes narrower: is this particular option worth my time, attention, or trial?

That narrowing is the practical value of a referral. It can reduce the cost of an initial search without claiming to settle the decision. A friend may have used a restaurant often enough to know its strengths. A colleague may know which adviser understands a specialised problem. A general practitioner may be able to identify a specialist whose practice fits a particular clinical question. In each case, the Source may provide a reason to investigate the Target before the Target has earned confidence independently.

The distinction between orientation and guarantee is essential. A referral can make an initial appointment, conversation, or trial reasonable. It cannot establish the quality of every future interaction, the fit with every need, or the Target's conduct under conditions the Source has not seen. The more the recipient treats a referral as a guarantee, the more confidence outruns the evidence available.

A useful referral therefore directs checking rather than ends it. It gives the recipient a starting point and a set of questions: what did the Source actually experience? How close is that experience to my own situation? What did the Source assess, and what did they merely assume? What remains for the Target to demonstrate?

The Weight of a Single Link

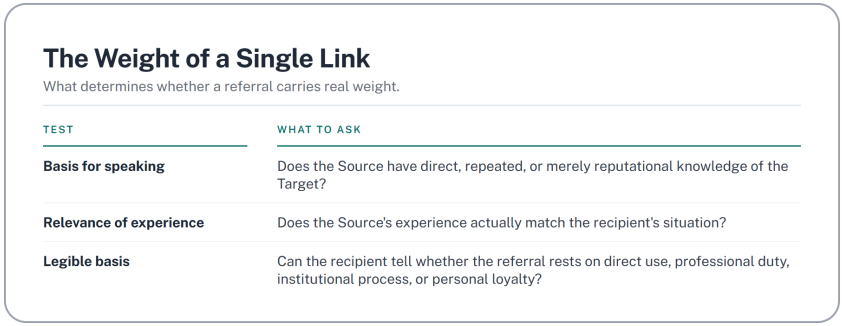
One-to-one referral does not work because all personal recommendations are equally trustworthy. It works, when it works, because the recipient, reading it from their own situation, interprets the link as meaningful.

Three features commonly shape that interpretation.

First, the Source needs a basis for speaking. A recommendation from someone who has used a service repeatedly may carry a different implication from a recommendation based only on acquaintance or reputation. A professional referral may carry a different implication again: the recipient may reasonably ask whether the professional has direct knowledge of the Target's competence, a formal duty to consider suitability, or simply a name within a familiar network.

Second, the Source's experience must be relevant to the decision. A friend can know that a restaurant is enjoyable without knowing whether it can accommodate a severe allergy. A colleague can admire an accountant without knowing whether the accountant is right for a complex cross-border matter. A referral becomes less useful when the Source's experience and the recipient's problem diverge, even if the Source is otherwise credible.

Third, the Bridge needs a legible basis. The recipient should be able to tell whether the recommendation rests on direct use, professional assessment, institutional process, commercial association, personal loyalty, or something else. A recommendation can be genuine and still be narrow. A professional introduction can be responsible and still leave important questions open. A disclosed incentive, connection, or limitation does not automatically invalidate a referral; it identifies part of what the recipient must take into account when interpreting it.



▼ Text version of this table

Test	What to ask
Basis for speaking	Does the Source have direct, repeated, or merely reputational knowledge of the Target?
Relevance of experience	Does the Source's experience actually match the recipient's situation?
Legible basis	Can the recipient tell whether the referral rests on direct use, professional duty, institutional process, or personal loyalty?

Table 3.2 The Weight of a Single Link. A referral's force depends on what the Source actually knows, whether that knowledge fits the decision, and whether the basis for the recommendation is visible.

These are not formal hurdles every ordinary recommendation must meet. They are a way of resisting a familiar error: treating the warmth of a referral as evidence that the Source has checked more than they have. The recipient need not begin from zero, but should know what the referral is actually carrying.

Peer Referral and Authority-Based Referral

A referral from a friend and a referral from an authority figure share the same basic shape, but they place different burdens on the Bridge.

A peer referral often derives its force from proximity to the recipient's everyday concerns. A friend may know the kind of service you value, the constraints you face, or the standard you use when you say something is "good." The Bridge can be persuasive because the friend's experience feels recognisable and their judgement feels relevant to your own decision.

An authority-based referral introduces a different reason for confidence: domain knowledge, professional responsibility, or access to information the recipient does not possess. When a doctor refers a patient to a specialist, the patient may see the referral as carrying more than personal preference. The referral may signal familiarity with the specialist's field, a view of clinical fit, or an institutional pathway for care.

That added authority does not convert the referral into a blank cheque. The recipient may still need to understand the basis of the referral, the range of available options, and the Target's own record. Authority can make a Bridge more consequential while also increasing the cost of a misplaced inference. The recipient may assume that a professional has assessed competence, availability, cost, safety, or suitability when the referral in fact signals only one of those things.

The question changes with the Source: not simply whether the recipient respects this person, but what relevant knowledge, duty, or evidence this person brings to this recommendation.

What the Evidence Can and Cannot Show

The one-to-one mechanism is broader than any single study. Its basic form is a conceptual claim of this book: confidence can begin with a known Source and move provisionally across a perceived Bridge toward a less-known Target.

A study of industrial buyers marks a useful boundary for that claim. In their 1997 Journal of Marketing article, Patricia M. Doney and Joseph P. Cannon developed a framework of five cognitive processes through which industrial buyers may develop trust in a supplier firm and its salesperson. They tested the model with data from more than 200 purchasing managers (Doney & Cannon, 1997).

The study did not examine friendship referrals, consumer reviews, or every form of personal recommendation. It examined industrial buyer–seller relationships. Its value here is more specific. The authors distinguished trust in an individual salesperson from trust in the supplier firm and reported that salesperson trust influenced anticipated future interaction indirectly through supplier-firm trust. In that setting, confidence in a person and confidence in the organisation they represented were related but not interchangeable.

This is a useful warning against a common simplification. A single referral can involve more than one object of confidence. A recipient may respond to the person making the introduction, to the person or service introduced, and to an organisation associated with either of them. Those relationships may reinforce one another, but the model should not collapse them into one undifferentiated feeling of trust.

The limits of the study matter too: the publisher abstract reports that, after accounting for previous experience and supplier performance, neither firm trust nor salesperson trust influenced current supplier selection in the model. That result limits how far the study can be stretched. It does not confirm this book's account of referral, and a null result could not. It is compatible with a

narrower reading: in that setting, once experience and performance were known, trust in the firm or the salesperson added little to the choice itself. That fits the idea that a referral opens a door rather than settling the choice, but it does not test it.

Referral Is a Beginning, Not a Substitute

A referral has done its best work when it brings the recipient to the point where Direct Evidence can begin. The introduced accountant answers questions. The plumber arrives and completes the work. The specialist explains the options. The Target's conduct now becomes available, and the Borrowed Confidence can be tested against what actually happens.

That later evidence may support continued reliance. It may also justify reduction, caution, or exit. In the book's terms, Direct Evidence can lead to Earned Trust or Earned Mistrust, or leave the question open for now. Chapter 10 develops those outcomes in full. The practical point here is earlier and simpler: the Source can make a Target worth approaching, but cannot permanently earn the Target's credibility on its behalf.

This is why a referral is neither naïve nor sufficient. It is a rational response to uncertainty when it shortens the path to a better first question. It becomes risky when it is treated as the last question.

Chapter Takeaway

A referral is a bounded loan of judgement, not a verdict. It is only as good as three things: whether the Source's own experience actually fits the decision at hand, whether the connection between Source and Target is legible, and whether the recipient understands what has, and has not, been checked.

Treating a single referral as more than one route into a decision is the recurring failure this chapter has traced. When the underlying experience does not fit, or the Target simply fails once someone looks directly, a personal introduction buys no immunity from that.

Chapter 4 puts the same referral on a public stage, in front of a name many people already recognise, and asks what changes once everyone nearby is watching.

Chapter 4

The Borrowed Name

On 17 September 1996, Oprah Winfrey's book club announced The Deep End of the Ocean by Jacquelyn Mitchard as its first pick. The announcement did not change the novel's pages, its author, or the experience a later reader would have of the book. It changed the public context in which the book was encountered. Oprah's name now stood beside a work many viewers had not read.

That is the power of a public endorsement. A private referral gives one person a reason to investigate a Target. A recognised name, attached visibly to a Target, can give that reason to thousands or millions at once. The Bridge is no longer confined to a conversation. It becomes part of the Target's public meaning.

A public endorsement is the same Endorsement Bridge with a visible audience.

The Trustor here is whoever is reading the endorsement. What they bring to it, what is at stake, what they know of the Source, and how much Direct Evidence about the Target they already have decide how much weight the name on the cover actually carries. It can feel decisive to one reader and merely interesting to the next.

A Name Is Not a Warranty

Public endorsement is potent partly because it compresses attention. The Source's name tells a reader or buyer where to look first. It may also suggest that someone already known has taken a view worth considering. But a borrowed name is not a general certificate of every quality that an observer may infer from it.

A book-club selection may say that a host wishes to bring a book to an audience. It does not by itself establish the factual status of every claim in that book. An athlete's association with a product may create an expectation of relevance within a domain. It does not, on its own, establish the quality, safety, or suitability of every product carrying that association. A public endorsement can be meaningful and still be narrow.

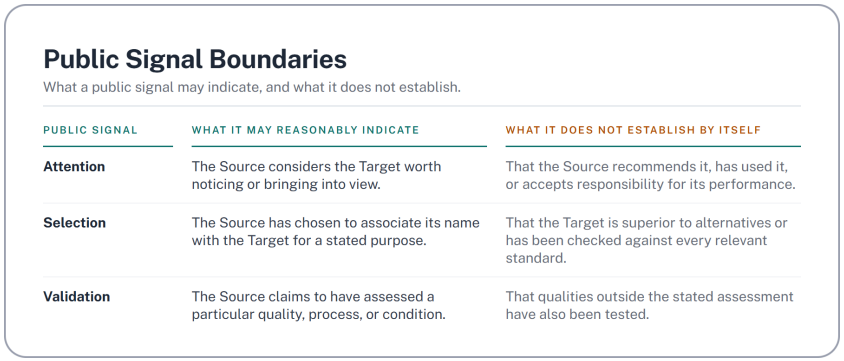
The reader's task is to keep the relationship legible. What has the Source actually endorsed? What relevant knowledge or experience can reasonably be inferred? Is the connection a recommendation based on direct engagement, a commercial partnership, an institutional selection process, or a looser association? What remains for the Target to demonstrate?

These questions do not make public endorsement useless. They prevent its ordinary convenience from becoming an over-read claim. The Bridge is

strongest when the scope of the relationship is visible enough for a person to calibrate the confidence it deserves. It is weak when an endorsement's publicity is mistaken for evidence that the Source has checked everything that matters.

What the Bridge Actually Carries

A public name beside a Target can carry at least three different kinds of signal. They are often blended together in practice, but they should not be treated as interchangeable.



▼ Text version of this table

Public signal	What it may reasonably indicate	What it does not establish by itself
Attention	The Source considers the Target worth noticing or bringing into view.	That the Source recommends it, has used it, or accepts responsibility for its performance.
Selection	The Source has chosen to associate its name with the Target for a stated purpose.	That the Target is superior to alternatives or has been checked against every relevant standard.
Validation	The Source claims to have assessed a particular quality, process, or condition.	That qualities outside the stated assessment have also been tested.

Table 4.1 Public Signal Boundaries. Attention, selection, and validation make different claims; readers should not treat them as interchangeable.

The reader's error is often not believing a false statement. It is moving silently from one level of signal to the next: from attention to selection, or from selection to validation. The Target then receives more confidence than the actual relationship can bear. A clear public endorsement makes its own scope easier to inspect. A vague one asks observers to fill the gaps with the Source's reputation.

This distinction also makes a practical demand on the Source. A public figure or institution need not guarantee a Target to make a responsible endorsement. It

should, however, avoid allowing a narrow connection to stand in for a broader assessment it did not perform. The Target, in turn, cannot rely forever on borrowed name recognition. Its own conduct, evidence, and direct experience must eventually carry the decision.

Visibility Raises the Stakes

A public endorsement differs from a private referral not because it has a different mechanism, but because its relationship is exposed to many interpretations at once. More people can see the Source, the Bridge, and the Target. More people can borrow confidence from the same act. The Source's name becomes part of the Target's public story, and the Target's conduct can in turn reshape what the endorsement means.

That exposure changes the risk. In a private referral, an error may remain within a small circle. In a public endorsement, the relationship itself becomes observable. Supporters, critics, competitors, journalists, and later audiences can all reassess what the Source's name was taken to guarantee. The wider the implied claim, the more the Bridge is asked to carry.

This does not mean every public association should be read as an endorsement. Public proximity can be co-presence rather than support; a brand may appear near a person without borrowing that person's judgement, and a person may be associated with an institution without speaking for it. The analytical question is not simply whether two names appeared together. It is whether there is a relationship that a reasonable observer would take as a basis for extending confidence from Source to Target.

When the Target Fails

The James Frey episode illustrates a Target-side failure. A Million Little Pieces was selected for Oprah's book club. In January 2006, Oprah publicly confronted Frey about fictionalised accounts in the memoir. In the published exchange, she said she felt "really duped" and asked whether a report detailing factual discrepancies was right; Frey replied that most of it was accurate and acknowledged discrepancies in the facts (Winfrey, 2006).

The immediate issue was the Target: the book's presentation as memoir and the accuracy of central claims. But the public relationship made the consequences wider than a dispute between author and reader. Once the Target's factual basis was challenged, people could also ask what Oprah's selection had meant, what had been checked, and whether the Bridge had carried more assurance than its basis could support.

This is not evidence that a Target's failure always destroys confidence in its Source. Oprah's own records do not provide a measurement of audience trust before and after the episode, and the case cannot supply one by itself. It shows a narrower mechanism. A Target-side failure can prompt scrutiny that moves from the Target back toward the named Bridge and the Source whose name made the Target more consequential.

The distinction is practical. The question after a Target-side failure is not only whether the Target was wrong. It is also what the Source's endorsement legitimately represented. If the public inferred comprehensive validation where the Source had only made a selection, the Bridge was over-read. If the Source made a broader claim than its basis could sustain, the Bridge may itself require correction.

When the Source Fails

The Tiger Woods case points in the other direction. Here, Woods functioned as a public Source in endorsement arrangements; sponsor companies were Targets receiving some of the credibility associated with his public persona. When the Source becomes the subject of a reputation event, the question becomes whether the Targets have tied their public meaning too closely to what the Source seemed to stand for.

Christopher Knittel and Victor Stango examined the stock-market effects of the Tiger Woods scandal on sponsors and competitors. Their event study reports that, in the 10–15 trading days after the scandal began, the full portfolio of sponsors lost more than 2% of market value, with losses concentrated among Electronic Arts, Nike, and PepsiCo's Gatorade business (Knittel & Stango, 2014). The study also reports correlations between day-by-day losses and endorsement-related search and news indicators, and interprets the event as a negative marketwide signal about reputation risk associated with celebrity endorsements.

The evidence has a specific scope: an event-study estimate of sponsor-portfolio market value, not a measure of individual consumers' trust or any single buyer's revised judgement. What it does show is that a Source-side reputation event can coincide with a measurable financial consequence for the portfolio of firms publicly connected to that Source.

The case makes the relationship legible in reverse. The companies had not necessarily endorsed every aspect of Woods's private life. Yet the public connection was strong enough for a reputation event involving Woods to become relevant to how investors evaluated the companies associated with him.

A public Bridge can therefore create exposure even when the Target's own product or service has not changed.

Two Failures, Two Questions

The Frey and Woods cases should not be collapsed into one generic story of reputational harm. Their structural difference is the point.

In the Frey case, the Target's factual failure prompted attention to the Source and the Bridge. In the Woods case, a Source-side reputation event placed pressure on Targets attached to that Source. Both involve public endorsement; neither justifies the shortcut that every failure travels in one fixed direction. The direction depends on which link has failed and what relationship the public believes was being claimed.

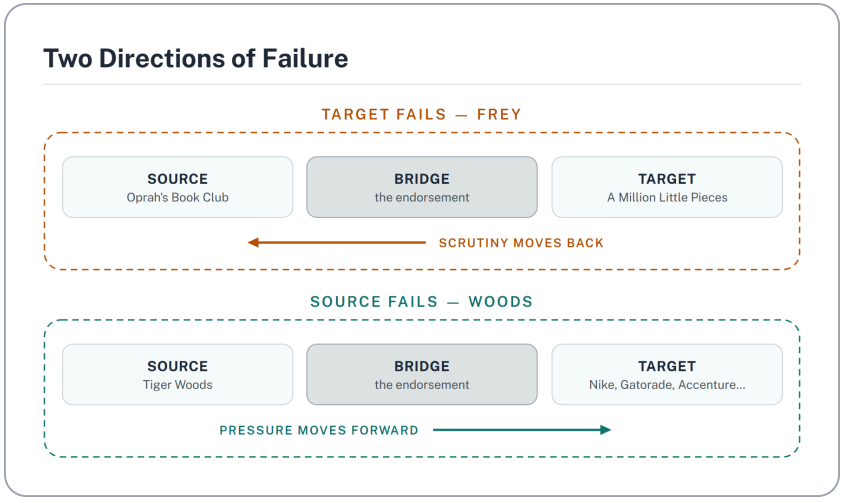


Figure 4.1 Two Directions of Failure. A Target's failure sends scrutiny back toward the Source; a Source's failure sends pressure forward onto every Target still wearing its name.

There can also be role ambiguity. An institution may be both a Target of an endorser's credibility and a Source of the endorsement arrangement that made the association visible. A public figure may be the Source of a recommendation but also part of a larger Target (such as a platform, club, or company) whose conduct shapes the endorsement's meaning. It is better to name that ambiguity than to force each case into a falsely clean assignment.

Chapter 12 develops the full Mistrust Contagion logic. The narrower lesson here is that public visibility increases the cost of ambiguity. The more weight an endorsement asks the Trustor to place on it, the more clearly its basis and limits need to be stated before a failure tests them.

Chapter Takeaway

Public endorsement is the one-to-one Bridge with a bigger audience: a recognised Source attaches its name to a less-known Target and buys attention, even provisional confidence, but not the Target's own evidence. That still has to come later.

The failure this chapter has traced is mistaking a public name for a warranty. When the Target fails, people look back at the Source that vouched for it. When the Source fails, a Target still wearing its name can pay for an association it never fully controlled, whether or not it did anything wrong.

Chapter 5 asks what happens when several apparently independent names point at the same Target, and when that convergence gives confidence a better basis rather than just a more confident-looking one.

Chapter 5

Many to One

One recognised name can direct attention to a Target. Chapter 4 showed why that visible relationship must be read for what it actually carries: attention, selection, or validation. Yet a Target rarely arrives with only one name beside it. It may have several ratings, reviewers, professional affiliations, references, placements, or technical assurances. The pattern can feel stronger than a single endorsement because it appears to say that more than one Source has arrived at the same conclusion.

That impression can be useful. Several Bridges may bring materially different experience, evidence, scrutiny, or accountability to the same Target. They may reduce different uncertainties: whether a product performs, whether a supplier can deliver, whether a claim has been tested, or whether an institution is answerable. But a pattern of agreement is not self-interpreting. Five visible signals do not necessarily mean five independent examinations. They may rest on the same dataset, a shared incentive, a common intermediary, copied language, or a single prior judgement that others have repeated.

The question is therefore not simply how many signals point in this direction. It is what additional basis each signal contributes. In the language established in Chapter 2, the pattern can contain several Sources, each connected by a Bridge, to one Target. Transfer names the movement of provisional confidence across those Bridges; it is not another signal to count. The Trustor, here, is whoever is weighing the pattern against the stakes of the decision, the evidence on hand, and the scope of each relationship — not one more voice inside it.

More Advice Can Add Information

The case against counting signals too quickly is not a case against seeking more information. In a laboratory study of factual estimation, Ilan Yaniv and Maxim Milyavsky asked undergraduate participants to estimate 24 historical dates. The participants then revised their answers after receiving either two, four, or eight estimates, randomly drawn from a pool of earlier undergraduate responses. The three conditions contained 55, 61, and 54 participants respectively. Participants could earn an accuracy-based bonus, and they did not receive the correct answers while revising their estimates (Yaniv & Milyavsky, 2007).

Their average accuracy improved in all three conditions. The authors report mean-absolute-error reductions of roughly 27% after two advisory opinions, 28% after four, and 33% after eight. The gains rose with the number of opinions, but the increments diminished rapidly. Participants also did not make the best

use of the advice available to them: their intuitive revisions were less accurate than some of the median and consensus-trimming rules examined in the study.

The result offers a modest but important point. Additional input can improve a judgement when it brings information that a decision maker did not already have. It does not license a rule that more opinions always deserve more confidence. The advisers in the first experiment were estimates sampled by computer from a peer pool, not named specialists who disclosed their methods, incentives, or accountability. Yaniv and Milyavsky therefore did not test real-world endorsements, source incentives, institutional accountability, or whether each advisor had investigated a Target independently. The experiment establishes a possibility, not a shortcut.

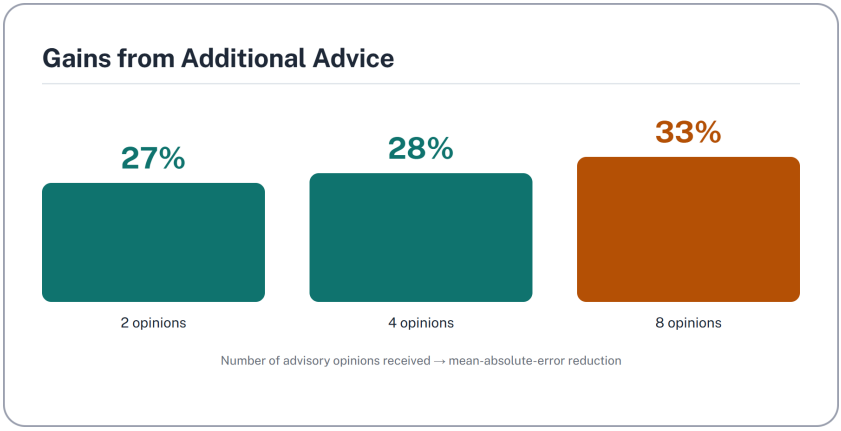


Figure 5.1 Gains from Additional Advice. In this experiment, additional opinions improved average accuracy, but the gains were uneven and should not be treated as a general rule that more advice produces better decisions (Yaniv & Milyavsky, 2007).

When Visibility Produces Convergence

A second experiment marks the other side of the problem. Jan Lorenz and colleagues recruited 144 students into twelve laboratory sessions, each with twelve participants. Working separately, participants made estimates on six factual questions. Across five rounds, they either saw no information about other estimates, saw the group's average estimate, or saw the individual estimates and trajectories of other participants. They were paid for accuracy, and their reported confidence was collected after the first and last estimate (Lorenz et al., 2011).

The researchers found that exposure to the others' estimates reduced the diversity of group responses without significantly reducing collective error. They also reported a confidence effect: participants became more confident as

their estimates converged even though this convergence did not produce a corresponding improvement in collective accuracy. Because the same sessions encountered the information conditions across different questions, in randomised order, the experiment could compare response patterns with and without feedback from others. This was still a deliberately controlled estimation task. The authors note that it abstracts from features of ordinary social life such as authority, competition, and group pressure. It does not test public endorsement, investment, consumer choice, or Trust Transfer.

Taken together, the two experiments draw a boundary that is easy to miss in practice. Multiple inputs can be valuable; visible agreement can also narrow apparent disagreement without adding accuracy. Neither experiment supplies a general law about trust. Their shared lesson is narrower: quantity and convergence are not substitutes for an account of how the information was produced, selected, and connected.

When Ratings Looked Like Corroboration

The structured-finance rating system before the 2007–09 financial crisis provides a high-stakes illustration. A complex mortgage-related security could be accompanied by investment-grade ratings, issuing from organisations familiar to market participants. Those ratings were recognised Certification Bridges: they translated an assessment of creditworthiness into a compact, portable signal. For an observer unable to inspect every loan, model, legal structure, or cash-flow assumption, the signals could make a distant Target appear pre-checked.

The Financial Crisis Inquiry Commission's majority report concluded that major firms and investors had relied on credit-rating agencies as arbiters of risk. Its conclusion described the failures of rating agencies as "essential cogs in the wheel of financial destruction" and "key enablers of the financial meltdown" (FCIC, 2011). That is the Commission majority's account of an interlocking crisis. It does not reduce that crisis to ratings, and it does not say that every investor, security, or rating agency acted alike.

A 2012 report by the staff of the U.S. Securities and Exchange Commission examined the rating process for structured-finance products and the conflicts of interest associated with issuer-pay and subscriber-pay models. It also described the statutory question of whether issuers, sponsors, or underwriters should be prevented from selecting the agencies that would determine and monitor initial ratings (SEC Division of Trading and Markets, 2012). The report was a staff study, and the Commission expressly stated that it had not adopted the report's analysis, findings, or conclusions. Its relevance here is not a finding that any particular rating was purchased or fraudulent. It makes a structural question

visible: when the party seeking a rating participates in selecting and paying the rating agency, what does a cluster of familiar ratings add beyond one another?

The case therefore does not ask the reader to reject ratings as useless. It asks for a more exact reading of what they can bear. A rating can offer a real Certification Bridge. It cannot, merely by appearing alongside other rating signals, establish that several fully independent investigations have taken place. The relevant questions are specific: did each agency use a distinct evidentiary basis; did it face a distinct selection and compensation arrangement; did its assessment reduce a different uncertainty; and could an observer see those differences? The case record does not answer each question for every instrument. It shows why the questions cannot be skipped. The basis of each Bridge remains part of the judgement.

The Independence Test

The Independence Test is a conceptual diagnostic, not a survey-validated scale. It helps the Trustor examine whether apparent convergence adds a new basis for provisional confidence or repeats an existing one in a different form. Table 5.1 sets out its checks.

Independence Checks

Questions to ask before treating multiple Bridges as independent confirmation.

QUESTION	WHAT TO ASK	WHAT THE ANSWER CAN ADD
Observation	Did each Source independently observe, investigate, or experience the relevant feature of the Target?	Distinct first-order information.
Evidence basis	Do the Bridges rest on different data, methods, or records, or are they reading the same material?	A check against repeated evidence.
Incentives	Are the Sources paid, selected, exposed, or otherwise motivated in the same way?	A check against shared interests.
Accountability	Who must answer if a Bridge proves misleading, and is that accountability actually distinct?	A check against concentrated responsibility.
Uncertainty reduction	Which particular uncertainty does this Bridge reduce that the other Bridges have not already reduced?	A test of incremental value.

▼ Text version of this table

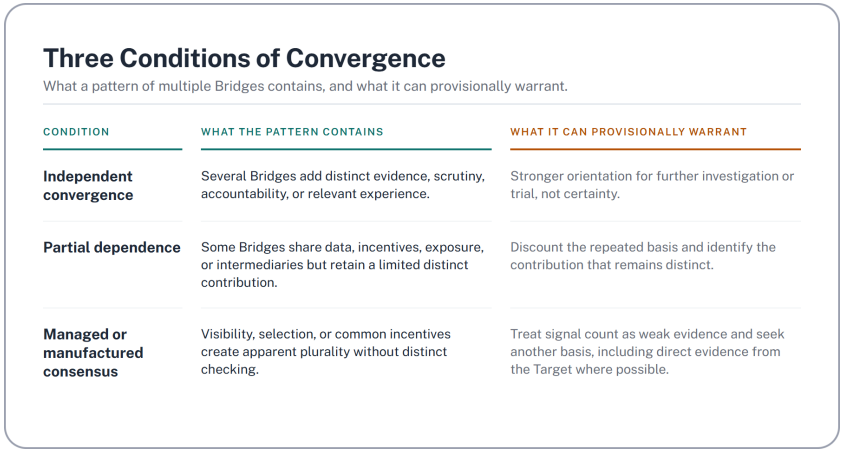
Question	What to ask	What the answer can add
Observation	Did each Source independently observe, investigate, or experience the relevant feature of the Target?	Distinct first-order information.
Evidence basis	Do the Bridges rest on different data, methods, or records, or are they reading the same material?	A check against repeated evidence.
Incentives	Are the Sources paid, selected, exposed, or otherwise motivated in the same way?	A check against shared interests.
Accountability	Who must answer if a Bridge proves misleading, and is that accountability actually distinct?	A check against concentrated responsibility.
Uncertainty reduction	Which particular uncertainty does this Bridge reduce that the other Bridges have not already reduced?	A test of incremental value.

Table 5.1 **Independence Checks.** Multiple signals are more informative when they contribute distinct observation, evidence, incentives, accountability, or uncertainty reduction.

The point is not to demand perfect independence. Most real decisions involve overlap. A regulator may use industry data; two specialists may read the same clinical record; several reviewers may know one another; a ranking may depend on an underlying certification. Independence is therefore a matter of degree and of relevance to the decision, not a ceremonial label. A second assessment that begins with the same record may still add value if it tests a different assumption, uses a different method, or carries a distinct accountability. A different-looking label that merely republishes the first assessment may add little. Overlap does not erase every contribution. It changes the weight a reader can reasonably give to a repeated signal.

The test also prevents an opposite error: treating disagreement as automatic proof of independence. Sources can disagree because they have distinct evidence, but they can also disagree because they are applying different standards, incentives, time horizons, or interpretations. Convergence deserves inspection; divergence does too. What matters is the traceable basis of the Bridge, not whether a pattern looks tidy from a distance.

Three Conditions of Convergence



▼ Text version of this figure

Condition	What the pattern contains	What it can provisionally warrant
Independent convergence	Several Bridges add distinct evidence, scrutiny, accountability, or relevant experience.	Stronger orientation for further investigation or trial, not certainty.
Partial dependence	Some Bridges share data, incentives, exposure, or intermediaries but retain a limited distinct contribution.	Discount the repeated basis and identify the contribution that remains distinct.
Managed or manufactured consensus	Visibility, selection, or common incentives create apparent plurality without distinct checking.	Treat signal count as weak evidence and seek another basis, including Direct Evidence from the Target where possible.

Figure 5.2 Three Conditions of Convergence. What a pattern of multiple Bridges contains, and what it can provisionally warrant.

The distinctions are practical rather than moral. A managed pattern may be entirely legitimate when it is made transparent: a shared data source can be appropriate, and a coordinated certification process can make responsibility easier to locate. Equally, independent convergence does not remove the need to inspect the Target. It can improve the starting point for a judgement, but it does

not convert several partial assessments into a comprehensive one. The risk arises when shared origins are invisible and the audience is invited to infer independent confirmation that the pattern cannot supply.

A reader does not need to resolve every question before acting. Confidence often begins provisionally. The Independence Test clarifies what sort of reliance is sensible at that stage. A cluster of Bridges may justify paying attention, comparing alternatives, or beginning a limited trial. It should not be made to carry assurance that only distinct scrutiny or later Direct Evidence could provide.

Several visible signals are valuable only to the extent that their bases add something new. The failure risk is an echo mistaken for corroboration: common evidence, incentives, selection, or exposure can give Borrowed Confidence a stronger appearance than its underlying basis can justify. The Independence Test keeps the reader from treating a settled-looking pattern as an end point. It identifies what remains to be checked and what form of reliance is proportionate before further evidence arrives.

Yet even actually distinct signals have a history. A certification may be current or stale; an observer may learn of a recommendation after the circumstances that supported it have changed; repeated exposure may make one old judgement feel newly confirmed. Direct experience with the Target can also accumulate while earlier borrowed signals retain their public force.

Chapter Takeaway

Several signals pointing at the same Target can add real information, or they can be one signal wearing several faces. The Independence Test gives questions for assessing which: whether the sources share an origin, an incentive, or an audience that would make them move together regardless of what the Target actually deserves.

A cluster of Bridges is enough to justify attention, a comparison, or a limited trial. It is not enough to substitute for Direct Evidence, no matter how settled the pattern looks from a distance.

Chapter 6 turns to that temporal problem: how the order, recency, repetition, and later experience of Bridges alter the confidence they can carry.

Chapter 6

The Tempo of Trust

A Bridge Has a Time Condition

Chapter 5 asked whether several signals add independent basis or merely repeat one another. That question remains essential. A cluster of familiar signals can look like corroboration when the same organisation, record, or assumption sits beneath each of them. But an assessment of a Bridge is incomplete if it asks only where the signal came from. It must also ask whether the signal is still fitted to the situation now being judged.

A Bridge has a history. It was issued or formed at a particular moment, in response to particular information and under particular conditions. Those conditions may still hold. They may have narrowed, changed, or been overtaken by later facts. The passage of time is not, by itself, a defect. An older recommendation, certification, ranking, or placement can remain highly informative. Nor does a new signal automatically deserve greater weight merely because it is recent. The central question is more exact: does the evidence that made this Bridge useful still apply to this Target and this decision?

This chapter calls that question the tempo of trust. Tempo is not a personality trait, a universal speed of belief, or a promise that confidence always develops slowly and disappears quickly. It is the temporal condition of a perceived Bridge: its sequence, recency, repetition, expiry, and place in the evidence that has accumulated since it first appeared. The book argues that a Bridge may be legitimate and useful at one time yet require a different weight at another.

From First Signal to Later Evidence

At the start of a relationship, direct knowledge of the Target may be thin. A known Source can then make a connection meaningful. A colleague's referral, a certification mark, a familiar host, or a platform's placement can offer a reason to consider a less-known Target. The confidence that crosses the Bridge is provisional. It is not a property transferred once and kept unchanged.

A certification is one type of Bridge. Its value depends not only on who issued it, but also on its stated scope, the information on which it rested, and the conditions under which it remains relevant.

The same institutional signal can mean almost nothing in a low-stakes comparison and a great deal once its scope, currency, and limits actually matter to the decision at hand.

Time changes what that lens has available. A Bridge may be the first reason to look. Later evidence can supplement it, qualify it, or reduce its relevance. That later evidence need not be dramatic. A changed condition, an expired assurance, a revised standard, a new record, or a direct encounter with the Target can all alter the basis on which the original signal should be read. The point is not to distrust every early signal. It is to keep Borrowed Confidence responsive to the evidence that follows it.

Conditional Certification Is Not Timeless Certification

The U.S. Food and Drug Administration's emergency use authorization for Evusheld provides a bounded illustration. In December 2021, the FDA announced an emergency use authorization for Evusheld for a defined pre-exposure purpose and population. Its announcement distinguished that authorisation from FDA approval. It described an EUA as a decision made on the evidence then available, subject to terms and conditions, while safety and effectiveness continued to be evaluated (FDA, 2021).

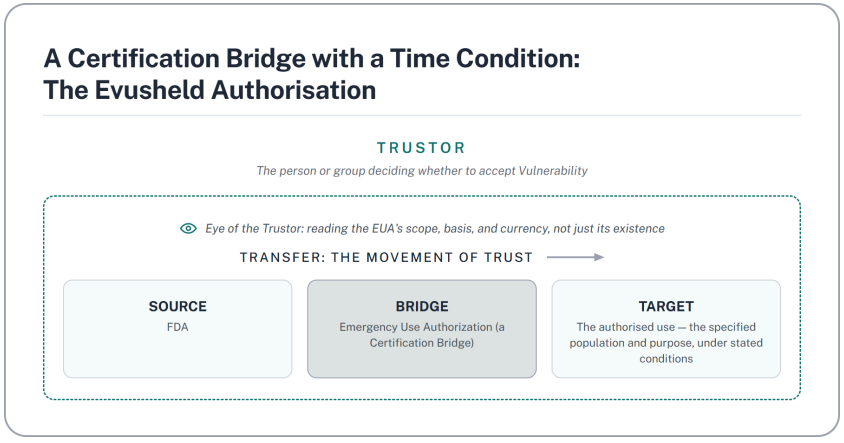


Figure 6.1 A Certification Bridge with a Time Condition: The Evusheld Authorisation. The FDA served as Source, the EUA as a Certification Bridge with a defined scope and conditions, and the authorised use — not Evusheld in general — as the Target, whose boundaries shifted when the authorisation was revised in 2023.

Figure 6.1 maps this case onto the model: the FDA as Source, the EUA as a Certification Bridge with a specified scope and stated conditions, and the authorised use, not Evusheld in general, as the Target. That precision matters. A formal Bridge is not a free-standing symbol of universal approval. It carries the boundaries of the decision that created it.

On 26 January 2023, the FDA reissued the authorisation and revised its scope. The letter limited authorised use to circumstances in which the combined

national frequency of non-susceptible variants was at or below 90%, using available information that included variant susceptibility and national variant frequencies (FDA, 2023b). The accompanying healthcare-provider fact sheet records the original authorisation date as December 2021 and the revised date as January 2023. It also states that the EUA operated for the duration of the underlying emergency declaration unless it was terminated or revoked sooner (FDA, 2023a).

These records do not show that an authorisation was worthless at the outset, that a product failed in every setting, or that anyone made a particular decision wrongly. They establish a narrower point. The Bridge had a time condition. Its warranted meaning depended on stated circumstances, and its scope was revised as those circumstances changed. The authorisation did not become a different kind of evidence merely because it had become familiar. Its present relevance had to be read from its present terms.

That distinction is useful beyond this case. A certification can deserve confidence without being timeless. A recommendation can be well founded without applying to a changed context. A ranking can faithfully report an earlier measurement without answering a later question. The temporal error is not trust in a Bridge as such. It is treating the memory of a Bridge as though it were evidence that its original basis remains unchanged.

Repetition Is Not Renewal

A signal can also appear current because it has been encountered repeatedly. Repetition may be useful when it directs attention to actually updated evidence. Yet repetition alone does not show that the information has been refreshed, independently checked again, or made applicable to changed conditions.

Hassan and Barber examined this distinction in two controlled experiments using trivia statements. Participants encountered statements at different frequencies and later rated their truthfulness. The authors found that repeated statements received higher truth ratings than new statements. The largest increase appeared when a statement was encountered for a second time, with incrementally smaller increases as repetition increased further (Hassan & Barber, 2021). This effect is usually called the illusory truth effect. Their first experiment used a one-week interval between sessions; it was a controlled task involving online adult participants and truth ratings of statements, not a study of regulatory authorisations or decisions in the world.

The boundary is important. The study does not test whether a certification is reliable, whether an official authorisation should be followed, or whether repeated information caused anyone to make a particular choice. It does not test

Trust Transfer. It supplies a narrower caution: in a laboratory truth-rating task, recurrence affected perceived truth. The chapter's conceptual inference is therefore limited. Familiarity may make an unchanged signal feel as though it has acquired new force. But renewed force must come from new or still-applicable evidence, not from the number of times an earlier signal reappears.

The difference can be easy to miss. A repeated certification mark, a recycled recommendation, or a frequently displayed rank may feel validated by visibility. It may instead be the same old basis seen again. Equally, an assurance that appears unchanged may have been revised with relevant new evidence. Tempo cannot be read from frequency alone. It requires attention to what, if anything, has actually changed behind the signal.

The Temporal Check

The book's temporal check is a practical way to keep that distinction visible. It asks four questions of any important Bridge:

1. When was the assurance created?
2. What conditions made it applicable?
3. Has its evidence been renewed or merely repeated?
4. What Direct Evidence has appeared since?

The date does not settle its value, but it locates the evidence in time. A recommendation may describe a Target before a material change; a certification may record compliance with an earlier standard; a ranking may reflect an earlier measurement window. Knowing the date turns vague freshness into a claim that can be assessed.

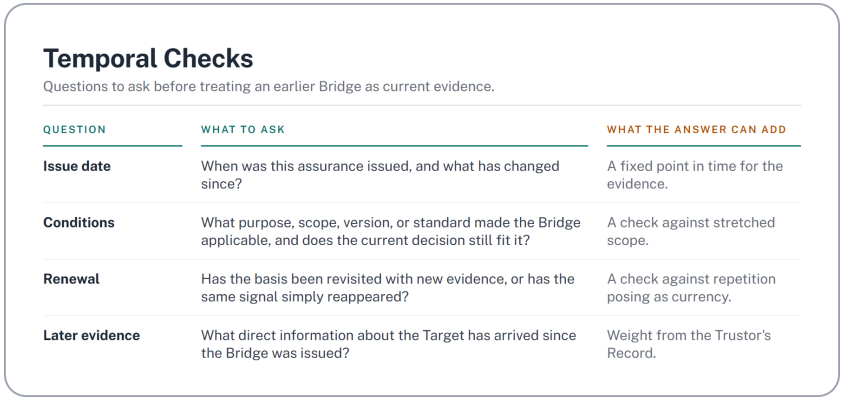
Every Bridge has a context, even when that context is not displayed beside the signal. The FDA case makes the issue explicit: the authorisation was defined by a purpose, population, and conditions that the Source later revised. Other Bridges may depend on a location, a version, a relationship, a standard, or a set of assumptions. The question is not whether a condition makes the Bridge weak. It is whether the current decision still fits the condition.

Renewal means that the underlying basis has been revisited or that new information speaks to the present case. Repetition means only that the existing signal has appeared again. The two can occur together, but they should not be confused. A signal can be widely visible and stale; it can also be modestly visible and current. The book argues that visibility is not a substitute for currency.

Time does not act only on the Bridge. It also creates opportunities for observation. Performance, conduct, experience, track record, and changed

circumstances can add evidence that was unavailable at the moment of introduction. That accumulating history is the Trustor's Record, and it grows independently of whatever the original Bridge said. It does not erase the Source or make the original Bridge irrelevant by default. It changes the evidentiary setting in which the Bridge should be weighed.

The Record ages as well. Entries built up before a material change describe a Target that may no longer exist in the same form. A café that has been reliable for years can change hands, keep its name and signage, and replace its kitchen staff within a month. The regular's Record still says "good", but it is now a Record of the previous owners. The temporal questions that apply to Bridges apply here too: when was the last direct interaction, and has the Target changed since?



▼ Text version of this table

Question	What to ask	What the answer can add
Issue date	When was this assurance issued, and what has changed since?	A fixed point in time for the evidence.
Conditions	What purpose, scope, version, or standard made the Bridge applicable, and does the current decision still fit it?	A check against stretched scope.
Renewal	Has the basis been revisited with new evidence, or has the same signal simply reappeared?	A check against repetition posing as currency.
Later evidence	What direct information about the Target has arrived since the Bridge was issued?	Weight from the Trustor's Record.

Table 6.1 Temporal Checks. An earlier Bridge still counts as evidence only if its date, conditions and basis still fit the decision being made now.

Note: the temporal check is not a validated scale, nor a decision rule for medical, financial, legal, or other high-stakes choices.

Together, these questions protect a basic distinction. A Bridge may still deserve confidence, deserve less confidence, or require a more qualified reading. The

answer depends on the continuing fit between the signal, the Target, and the decision, not on age alone, novelty alone, or frequency alone.

When Signal and Context Fall Out of Time

Temporal mismatch occurs when a signal's apparent force outlasts the condition that made it informative. A formal assurance may remain recognisable after its scope has narrowed. An earlier review may remain publicly visible after the Target has changed. A recommendation may circulate after the relationship or circumstances that produced it no longer exist. In each instance, the risk is not necessarily that the Bridge was deceptive or useless. The risk is that its current evidentiary weight is read as if nothing had changed.

This makes the failure mode subtler than simple error detection. A reader may accurately remember that a Source once supported a Target while still giving that past support more weight than the present conditions justify. The most useful response is not reflexive suspicion. It is a return to the temporal questions: what was known then, what is known now, and what connects the two?

The same discipline prevents a reverse mistake. New information should not automatically cancel an earlier Bridge merely because it is newer. Later information may be irrelevant, weak, or unrelated to the condition the original Bridge addressed. Timing becomes informative when it is connected to substance. A change in date warrants attention only when it marks a change in evidence, scope, context, or direct experience, not because time itself has a fixed moral direction.

From Timing to Signal Environments

Brands and platforms can keep an old signal visible long after its evidentiary basis has changed. Part Two examines how those environments shape what people continue to trust.

Chapter Takeaway

A Bridge has a time condition built into it, even when nothing on the label says so. Well grounded when it was issued, a Bridge can still go stale, get repeated without ever being renewed, or lose relevance once Direct Evidence about the Target starts arriving.

Familiarity is what makes that hard to catch. It keeps an old assurance looking exactly as it did the day it was issued, long after the reasons behind it have moved on, and continuity of appearance gets mistaken for continuity of evidence.

Part Two turns from the pattern itself to the domains where it plays out at scale, starting with the brands and influencers who now decide how often a signal reappears and how long it stays convincing.

Part II

The Domains: Where Trust Is Borrowed

Where Borrowed Confidence does its heaviest work: commercial endorsement, the platforms that rank and display it, and the machines now receiving it. This part closes by asking what happens once the Target becomes directly observable, when Borrowed Confidence gives way to Earned Trust or Earned Mistrust, or remains unresolved.

Chapter 7

Brands and Influencers

A familiar name can make an unfamiliar offer easier to notice. A creator wears a dress, a brand appears beside a service, or a public figure is placed beside an event. The association may make the Target feel less distant. But visibility alone does not tell us what kind of relationship connects Source and Target, what experience the Source brings to it, or what evidence lies behind the association. A commercial signal can be useful without being independent scrutiny.

This distinction is especially important when a relationship looks personal while it is also paid, arranged, or otherwise commercially structured. A photograph may appear to be an individual expression of taste. It may instead sit inside a campaign with specified content, compensation, approval, and disclosure terms. Those possibilities do not make every commercial endorsement weak. They do change the question a reader needs to ask before allowing the association to carry much confidence.

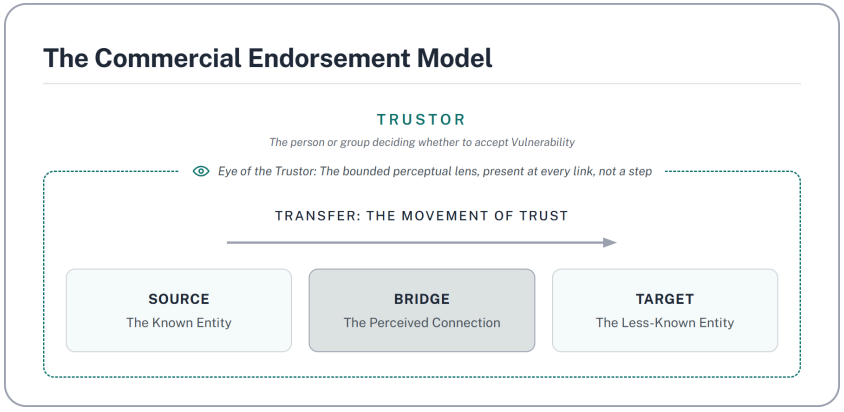


Figure 7.1 The Commercial Endorsement Model. Trust transfers from a brand or creator (Source) across a commercial association (Bridge) to an offer (Target).

Here, as Figure 7.1 shows, the Source is the brand or creator; the Bridge is the commercial association, often an Endorsement; the Target is the offer. The Trustor reads that relationship against the decision at hand and what is at stake.

Commercial relationships have more than one dimension. Payment may reveal an incentive. Pre-approval may reveal who shaped the message. A required label may reveal that the message is commercial. Relevant experience may reveal why a Source's observation deserves attention in one setting. These facts do not carry identical implications, and no single one settles the quality of the

Target. Separating them prevents a familiar face from becoming a compressed substitute for all the questions the Target itself still raises.

For commercial signals, the central question is not whether an endorsement looks persuasive. It is what the relationship gives a person reason to infer. A disclosed paid relationship establishes that a Source is associated with a Target and that the association has a commercial basis. Relevant experience may establish a different kind of connection. Simple co-visibility may establish only that two things have been placed together. None of these, by itself, establishes independent testing, product quality, suitability, or a reason to act.

When an Endorsement Appears Independent

The Federal Trade Commission's 2016 Lord & Taylor matter provides a precise place to begin. The FTC complaint alleged that Lord & Taylor's March 2015 Design Lab campaign gave the Paisley Asymmetrical Dress to 50 selected fashion influencers and paid them between US$1,000 and US$4,000 to post a photograph during a specified weekend. The complaint further alleged that the campaign contracts required brand identifiers and that the company pre-approved proposed posts, but did not require the influencers to disclose compensation (Federal Trade Commission, 2016a).

The case is not useful because it supplies a general story about brands or creators. It is useful because its records make the relationship legible. In the apparent signal, a particular fashion influencer's image appeared beside a defined dress. In the commercial arrangement alleged by the FTC, the Bridge included a gift, payment, specified attribution, and advance review. The Target was the named Design Lab dress, not the general category of fashion, the whole Lord & Taylor business, or the quality of a garment. Once those elements are separated, a visible image cannot simply be read as an unstructured statement of independent experience.

The final consent order is important for a different reason. It states that Lord & Taylor neither admitted nor denied the complaint allegations, other than facts necessary to establish jurisdiction. For the order's purposes, a "material connection" is a relationship that materially affects the weight or credibility of an endorsement and that would not reasonably be expected by consumers. The order prospectively prohibits misrepresenting an endorser as an independent user or ordinary consumer and requires clear and conspicuous disclosure of a material connection where one exists (Federal Trade Commission, 2016b).

Evidence boundary: This case makes the commercial relationship visible. It does not establish product quality, measure individual reliance or prove that the endorsement caused a purchase.[1]

The order makes a narrower point: relationship terms can affect the weight given to an endorsement. The question here is evidentiary rather than moral. When a person sees a familiar figure beside a commercial Target, is the basis of the relationship visible enough to be read accurately, and what does it still leave unestablished?

A disclosed relationship is not the same as a disinterested relationship, and a commercial relationship is not the same as an unreliable one. Disclosure identifies a relevant condition for interpreting a signal; it does not resolve whether the Source has relevant experience, whether the claim is specific, or whether the Target can support the confidence being transferred. The distinction gives the Trustor more information about the Bridge, not a verdict on the Target. The chapter therefore treats legibility as a condition for interpretation rather than a guarantee of good judgement.

A Familiar Signal on a Larger Stage

Fyre Festival makes the same question memorable on a much more familiar public stage. A pitch-deck exhibit filed with the 2017 Reel v. Fyre Festival complaint described a "Fyre Starters" coordinated influencer-marketing campaign for a defined festival event. The exhibit placed named public figures, including Kendall Jenner, alongside the campaign. The complaint separately alleged particular paid promotion and incomplete advertising disclosure (Reel v. Fyre Festival complaint, 2017).

The analytical point is deliberately limited. A recognised creator's image beside a festival can make the event worth noticing. It can direct attention toward a Target that might otherwise remain remote. Yet public association does not itself show that the creator independently assessed the event, possessed relevant information about it, or could answer for what the offer would deliver. The visible relationship may carry some provisional confidence, but the scope of that confidence depends on what relationship is actually present and what information supports it.

Evidence boundary: This case makes the commercial relationship visible. It does not establish creator knowledge, ticket-purchase causation or the later event outcome.[2]

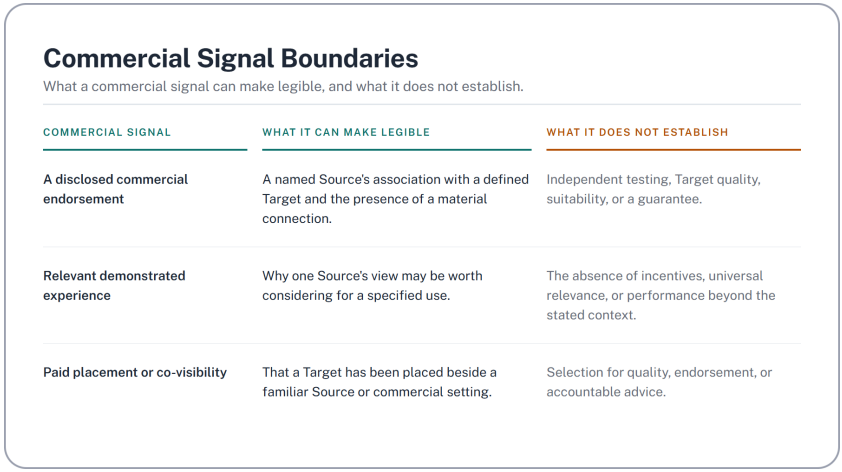
That caveat matters because the drama of the later festival can overwrite the narrower question: what was it reasonable to infer from the association when it was encountered? Not that public figures prove a Target is sound, and not that every visible creator relationship is suspect, but that the commercial Bridge must be named and read before it is allowed to carry more than it can support.

What a Label Changes

A sponsorship label can make part of that Bridge more visible. In Study 1 of an online experiment, Cao and Belo randomly assigned Prolific participants to see otherwise matching simulated Instagram posts with or without a “Paid partnership with” header. The no-header condition contained 290 participants and the header condition 298. Participants in the header condition gave higher mean ratings of advertising awareness and perceived sponsorship transparency. The study reported no statistically significant main difference in stated willingness to engage (Cao & Belo, 2024).

The result is narrow but useful, and its limits travel with it. A prominent label can alter whether people recognise the commercial status of an otherwise matching post, under these online experimental conditions. It does not show that the labelled Target was better or worse. It does not show that the Source had more relevant experience, or that disclosure converted a commercial relationship into an independent assessment. It tells us a label makes a relationship more legible, not that legibility adds evidence about the Target.

A label therefore neither cleans nor condemns an endorsement; it separates one question from another. It may disclose that a message was sponsored. It leaves open whether the Source has relevant experience, and whether independent information about the Target exists. Its absence does not prove that no commercial connection exists. It shows only that the Trustor has less to go on. The task is always the same: distinguish what is known about the Bridge from what is merely assumed from the signal’s appearance.



▼ Text version of this table

Commercial signal	What it can make legible	What it does not establish
A disclosed commercial endorsement	A named Source's association with a defined Target and the presence of a material connection.	Independent testing, Target quality, suitability, or a guarantee.
Relevant demonstrated experience	Why one Source's view may be worth considering for a specified use.	The absence of incentives, universal relevance, or performance beyond the stated context.
Paid placement or co-visibility	That a Target has been placed beside a familiar Source or commercial setting.	Selection for quality, endorsement, or accountable advice.

Table 7.1 Commercial Signal Boundaries. Disclosure can clarify a relationship without converting a commercial association into independent evidence of quality or fit.

The practical discipline is neither cynicism nor automatic belief. It is to ask four connected questions. What is the Source's relationship to this Target? What makes that relationship relevant to the decision being considered? Which commercial connection is visible, and which connection remains unclear? What does the association still fail to establish? The answers will vary with the Target and the decision. A low-stakes aesthetic choice does not call for the same evidence as a consequential purchase or commitment. The point is not to eliminate reliance on commercial Bridges, but to keep their weight proportionate to their basis.

Chapter Takeaway

Much commercial trust runs on borrowed credibility rather than on direct experience with a brand or creator. A label can make a relationship visible; it cannot make a thin relationship strong. What decides whether an endorsement is fair is not whether money changed hands, but whether the Source actually had a basis to speak and said so clearly enough for the Target to eventually be judged on its own evidence.

Chapter 8 turns to the platforms and interfaces that display, rank, and repeat these signals at scale, where the same over-reading happens with far less visibility into who is doing the vouching.

1 The FTC complaint supplies the campaign-specific allegations; the final consent order states that Lord & Taylor neither admitted nor denied those allegations and supplies the order's terms (Federal Trade Commission, 2016a, 2016b).

2 The pitch deck is campaign material filed with the Reel v. Fyre Festival complaint; the payment and disclosure statements are allegations in that complaint, not adjudicated findings (Reel v. Fyre Festival complaint, 2017).

Chapter 8

Platforms, Search, and Algorithms

When people say they trust a platform, they usually mean something narrower and something broader at once. Narrowly, they mean the site or app feels usable, familiar, and safe enough. More broadly, they mean the platform seems to have done some of the checking for them. It has filtered, ranked, verified, recommended, highlighted, suppressed, and organised the chaos. This is precisely what makes platforms so consequential in the modern trust economy. They do not merely host choices. They shape which choices feel credible before users ever inspect them directly.

This is one of the clearest places where Trust Transfer becomes infrastructural. A consumer buying from an unfamiliar seller on Amazon, an employee relying on a result surfaced by internal search, a voter reading a widely circulated post, or a patient clicking the first health result on Google is not evaluating the Target in isolation. The Target appears inside a system already carrying trust. The platform becomes the Source, the result or seller becomes the Target, and ranking, placement, verification, or presentation become the Bridge through which confidence moves.

That transfer is often efficient. No one can personally investigate every seller, article, service, expert, or claim. Platforms reduce the cost of navigating abundance. They offer cues that allow action under uncertainty. But the same efficiency creates a problem. Once people learn to treat platform signals as evidence that checking has happened somewhere upstream, the platform begins lending credibility even when the underlying basis is weak, opaque, or purely commercial.

This is why platform trust is powerful and dangerous. It allows strangers to transact, information to travel, and discovery to scale. It also allows weak Targets to inherit stronger legitimacy than they have earned. Rankings, recommendations, and featured answers do that routing before the Trustor has inspected the Target.

The Platform as Source

A marketplace, search engine, app store, or social platform does not need to say we endorse this for users to feel endorsed into trusting it. Presence itself can act as a Bridge. If something appears on a known platform, users may infer that some threshold has been met. The threshold may be safety, relevance, quality, popularity, compliance, or payment. The user usually does not know which. But the cue still works.

The same mechanism explains why platform trust is so easily borrowed. A new seller on Etsy inherits some reassurance from Etsy. A new app in Apple's App Store inherits some reassurance from Apple. A restaurant appearing high on Google Maps inherits some reassurance from Google. The user is not wrong to make these inferences. In many cases, platforms really have imposed standards. The risk is that users can generalise far beyond the platform's actual promise.

That mismatch between what the platform has checked and what the user assumes has been checked is one of the most important hidden gaps in digital life. A blue badge may mean identity verification, not competence. A high rank may mean engagement, not accuracy. A recommended label may reflect popularity, not suitability. A sponsored placement may visually mimic a vetted one. The mind tends to read these signals as interchangeable because they all occupy the same psychological slot: somebody else probably looked into this.

Platforms therefore function as large-scale trust routers. They direct attention, but they also direct confidence. They do not merely answer the question what is visible. They increasingly answer the more consequential question what feels safe enough to act on.

Ranking as Hidden Endorsement

Search transformed trust because it turned order into meaning. In a physical store, shelf position matters. In digital search, position matters far more because it looks neutral while carrying huge inferential weight. Early placement is easily read as evidence of quality, authority, or relevance. The ranking itself becomes the Bridge.

That Bridge is psychologically efficient. If a result is first, it is easy to assume it has earned the place, and not to ask what earned means. It may mean backlink authority, freshness, popularity, location, personalisation, ad spend, platform priorities, model confidence, or some unstable mixture of them all. Search compresses many variables into one output, and the user reads that output as a simplified trust cue.

Search engines have therefore always been more than discovery tools. They are authority allocators. They do not only find information. They implicitly tell users what deserves attention first. That ordering is not experienced as rhetoric. It is experienced as environment.

PageRank made that Bridge technical. It treated a link as a sign that one page had recognised another, then turned the resulting network into an order. Users did not need the mechanism. They needed the habit that a higher place often felt safer than a lower one.

Algorithmic Curation

Social feeds, recommendation systems, and marketplace sort orders extend this logic further. Search at least begins with an explicit query. Recommendation systems often begin before the user asks. They decide what to surface, what to suppress, what to keep adjacent, and what to make frictionless. They do not merely reflect trust. They build the sequence through which trust becomes possible.

This is hidden trust routing. A platform recommends a creator, then the creator recommends a product, then the platform processes the payment, then the delivery estimate and return policy further reduce hesitation. Each layer lends confidence to the next. The user may believe they are making independent judgements about multiple objects. In reality, much of the trust has already been pre-assembled by the system's architecture.

The risk is that curation often feels more neutral than it is. It is easy to assume that if something was shown to you, some meaningful relationship exists between the platform and the Target. Sometimes that relationship is editorial or reputational. Sometimes it is behavioural, inferred from past clicks. Sometimes it is commercial. Sometimes it is simply probabilistic. Yet these different relationships are often presented in visually similar ways, and the Trustor responds to the cue before distinguishing among them.

The result is a world in which people increasingly trust by interface. They borrow confidence not from direct knowledge, but from fluent design, coherent sequencing, familiar layout, and the platform's apparent willingness to stand beside what it shows.

Answer Boxes as Compressed Platform Outputs

Answer boxes intensify this pattern because they do not just rank sources. They lift one out and present it as the answer. Traditional search left visible traces of the Source–Target Bridge: a title, a URL, a snippet, a list. An answer box compresses those traces into a single highlighted block at the top of the page. That can feel more helpful. It can also make the Trust Transfer less legible.

Google learned this in public. In March 2017, people who asked whether Barack Obama was planning a coup got a featured snippet, lifted from a conspiracy site, that said yes. Google Home read the same answer aloud. Further down the page sat an ABC News story debunking the rumour, but a reader who already had an answer had little reason to scroll down to it. Google removed the snippet and apologised, explaining that featured snippets were an automatic, algorithmic match drawn from third-party sites (Sullivan, 2017). The episode shows a

placement being read as a check. It does not show how often featured snippets fail, or that users generally cannot tell a snippet from an ordinary result.

The case shows where the trust came from. The conspiracy site had none of its own to lend. The box borrowed Google's. Placement above the ordinary results told the Trustor that this Source had been checked, when it had only been matched. The spoken version was worse. A voice answer has no link to follow, so even when the Source is named, the Trustor has little to check it with.

That is the cost of compression. The fewer traces of the Source a format leaves, the more the platform's own name has to carry, and the less the Trustor can do to check it.

The broader pattern holds here too. Users are not asking to trust the box directly. They want trustworthy intermediaries (human editors, fact-checkers, accountable institutions) standing behind whatever the box says. The more a format hides its Source, the more those external anchors matter.

Authority After Search

Search engines do more than rank sources. They teach publishers what counts as authority, and publishers respond. In the late 2000s a group of companies known as content farms learned to turn out thousands of cheap articles a day, each written to match a phrase people typed into search boxes. The pages were built to rank. Many were thin, copied, or simply unhelpful, yet they sat near the top of results because they matched what the ranking system rewarded.

In February 2011 Google changed its ranking in a single update that, by its own account, noticeably affected 11.8% of queries. The stated aim was to push low-quality sites down and lift original reporting and research (Singhal & Cutts, 2011). Several large content sites were widely reported to have lost ranking sharply. Nothing in their articles had changed. What changed was the Source standing behind their visibility.

The episode shows where search authority sits. A top result borrows credibility from the engine that placed it there, and the engine's standard can be gamed by publishers who are good at being found rather than good at being right. Careful sources that are hard to find can lose ground to thin ones that are easy to find. The Trustor sees only the position, never the contest that produced it.

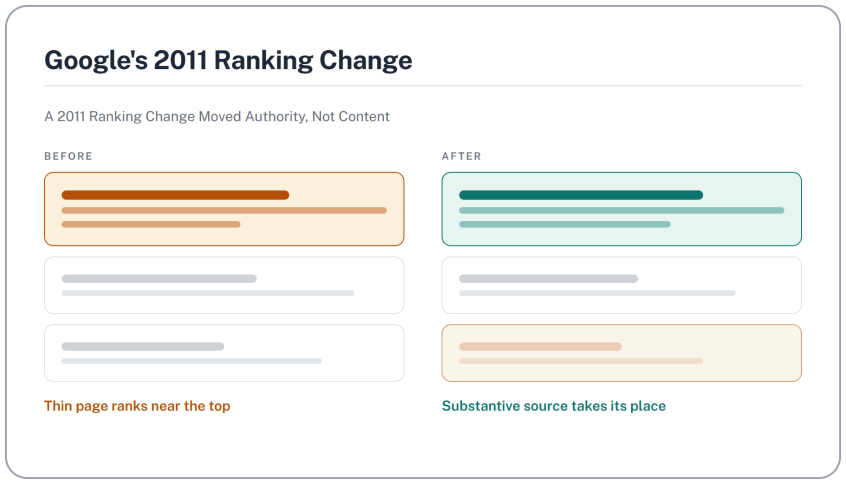


Figure 8.1 Google's 2011 Ranking Change. A single February 2011 update, affecting 11.8% of queries by Google's own account, aimed to push low-quality pages down and lift original reporting and research.

Platform Accountability

Regulators in multiple jurisdictions have begun treating algorithmic transparency and recommender architecture as a systemic safety and consumer-protection issue, not a purely technical one. They include the European Commission (under the Digital Services Act) and Australia's eSafety Commissioner (through statutory transparency notices) (European Commission, 2024; eSafety Commissioner, 2024). That regulatory attention is itself a trust signal: it tells the public that platform ranking and recommendation behaviour is now considered consequential enough to govern.

The underlying argument remains strong without them. Once platforms become the infrastructure through which people discover products, information, and public claims, the question is no longer whether they shape trust. They already do. The real question is whether their trust-routing power is governed visibly enough for users to borrow confidence from that governance.

People treat regulation this way everywhere. Few passengers know how a lift works. Many have noticed the inspection certificate posted inside it, and that certificate, not any knowledge of engineering, is what they are relying on. Translated into the platform context, the lesson is straightforward: governance can itself become a source in the trust-transfer chain, but only when it is visible and legible enough for users to perceive.

Invisible regulation can still protect people, but it transfers little trust. If users cannot see what rules exist, who enforces them, and how they shape the outputs

reaching them, the reassurance governance could offer is muted, even when the rules are real and working.

What Platforms Actually Lend

It is tempting to speak as though platforms lend one thing called trust. In reality, they lend different components at once. A search engine may lend competence: this result is probably relevant. A marketplace may lend integrity: this seller has probably been screened enough to transact. A social platform may lend benevolence, or the appearance of it: this recommendation feels socially validated and therefore less risky. These judgements are often bundled in the Eye of the Trustor, even when the evidence behind each is quite different.

Failures are rarely interpreted narrowly because of this bundling. If a search engine repeatedly surfaces wrong medical answers, the problem is not experienced only as an accuracy failure. It becomes a moral and institutional failure. If a social platform's recommendation system repeatedly drives vulnerable users toward harmful content, the damage attaches not only to the posts but to the system that routed them there. Trust was transferred across the Bridge, so mistrust returns across it too.

Platforms therefore occupy a strange position in modern life. Depending on the question being asked, they can look like publishers, utilities, or marketplaces. What they actually are, underneath all three, is a credibility environment. The practical consequence is that every ranking, label, recommendation, and default setting is also a governance choice about where trust will flow.

Platform Signal Boundaries

What a platform signal can make legible, and what it does not establish.

PLATFORM SIGNAL	WHAT IT CAN MAKE LEGIBLE	WHAT IT DOES NOT ESTABLISH
Presence on a known platform	The Target is available under that platform's stated inclusion rules. Any specific screening needs a named policy and date.	Competence, quality, screening, or fitness for the Trustor's specific use.
High search or ranking position	Some combination of relevance, freshness, or popularity signals.	Accuracy, authority, or truthfulness of the result.
Algorithmic recommendation or curation	That multiple signals have been pre-assembled into one routed suggestion.	Independent evaluation of the Target's suitability for this Trustor.
AI-generated answer or summary	A fluent, compressed synthesis across sources.	Which sources were weighted, or that the synthesis is correct.

▼ Text version of this table

Platform signal	What it can make legible	What it does not establish
Presence on a known platform	The Target is available under that platform's stated inclusion rules. Any specific screening needs a named policy and date.	Competence, quality, screening, or fitness for the Trustor's specific use.
High search or ranking position	Some combination of relevance, freshness, or popularity signals.	Accuracy, authority, or truthfulness of the result.
Algorithmic recommendation or curation	That multiple signals have been pre-assembled into one routed suggestion.	Independent evaluation of the Target's suitability for this Trustor.
AI-generated answer or summary	A fluent, compressed synthesis across sources.	Which sources were weighted, or that the synthesis is correct.

Table 8.1 **Platform Signal Boundaries.** What a platform signal can make legible, and what it does not establish.

Chapter Takeaway

Platforms, search engines, and algorithmic systems have become the hidden infrastructure of Trust Transfer. They shape what feels credible, what seems safe enough to click, and what appears worthy of action. Ranking, recommendation, verification, and interface design all function as Bridges that move trust from platform to Target.

Featured answers raise the stakes. They compress the Source–Target connection into a single box and reward whatever the ranking system can read most easily. Regulators' growing attention to ranking and recommendation tells the same story: in environments too complex to inspect directly, people look for governance systems they can borrow trust from.

When a platform's ranking, recommendation, or synthesis routes trust to a Target that fails, users may not blame the Target alone. The damage travels backward to the platform that vouched for it by showing it at all. Platforms already route trust, whether they mean to or not. The real question is whether they do it in ways accountable enough to deserve it. That question only gets sharper once the platform stops being a ranked list and starts acting on a person's behalf, which is where the next chapter turns.

Chapter 9

Machines We Never Inspect: Trust Through the Surroundings

In May 1854, at the Crystal Palace in New York City, a mechanic named Elisha Otis rode an open lift platform high above the crowd and ordered the rope holding it to be cut. His safety brake caught, and the platform held (Encyclopaedia Britannica, n.d.). Nobody watching could see the spring and ratchet that did the work. What they could see was a man willing to stand on his own machine.

That is how most people meet machines. They rarely judge the mechanism itself. They meet it inside something they already know: a bank, an airline, a carmaker, a hospital, an employer, a government office. Trust in a machine usually does not begin with the machine. It begins with the person, institution, or setting that stands next to it. In modern life, machine trust is often trust through the surroundings, which is broader than the Proxy Bridge Form from Chapter 2.

That dependence on surrounding context makes automation one of the clearest and most consequential domains for Trust Transfer. Most users cannot read the wiring diagram, check the tolerances, test the control logic, or verify the safety claims for themselves. They act anyway. They act because the machine appears inside a setting that already carries competence, legitimacy, or familiarity. A hospital buys a new infusion pump. An airline flies a new autopilot. A bank puts a cash machine in its wall. A carmaker adds anti-lock brakes. The machine inherits a starting credit line from the context around it.

That inheritance is not irrational. It is how people function in environments too complex to inspect directly. But it creates a structural problem. If the Source lending trust has stronger credibility than the machine deserves, users may rely too quickly. If the Source is weak or absent, even capable systems may be distrusted and underused. Machine trust therefore depends not only on what the system can do, but on what stands behind it, around it, and visibly above it.

The Adoption-Trust Gap

The world's first cash machine was unveiled at a Barclays branch in Enfield, north London, on 27 June 1967 (Barclays, 2017). Customers could not see how it counted out the notes. It is reasonable to suppose that the Barclays name on the wall did some of the reassuring, and that confidence in the machine itself came later, with use.

That gap is suggestive, but it does not by itself isolate Trust Transfer as the mechanism at work. Some of it is consistent with people moving ahead under partial confidence, borrowing reassurance from familiar institutions and social norms before direct proof is available. Some of it reflects reasons that have little to do with trust as such: workplace mandates, convenience, and the plain necessity of using tools that colleagues and competitors already use. Adoption is outrunning Earned Trust, though only part of that gap can be attributed to Borrowed Confidence specifically.

It also helps to separate trust in the provider from trust in the machine. A customer could trust Barclays completely and still doubt that a box in a wall would hand over the right amount and debit the right account. Those are different judgements. People may give the institution the benefit of the doubt while hesitating to let the machine handle the money itself.

Institutional Embedding

Trust reaches a machine through several recurring pathways. Each is one of the Bridge Forms from Chapter 2 at work: institutional embedding and interface familiarity are Embedding or Integration, provider reputation works as an Endorsement by the maker, and social proof works through Ranking or Placement and Co-presence. The first is institutional embedding. A machine feels safer when it appears inside a known institution with visible accountability. That first cash machine felt safer because it stood inside a Barclays wall, not on an unmarked street corner. Hospitals, banks, universities, and employers lend more than familiarity. They lend assumptions about vetting, oversight, standards, and consequences if something goes wrong.

Interface Familiarity

The second pathway is interface familiarity. Users can trust the environment before they trust the mechanism. Automatic lifts existed from about 1900, but many riders walked straight back out, looking for the operator. After New York's lift operators went on strike in 1945, the industry set out to change that: a recorded voice telling riders what to do, and a big red stop button, gave the empty car a familiar shape to step into (Henn, 2015). The machine felt less like a leap into the unknown and more like an extension of something already normal.

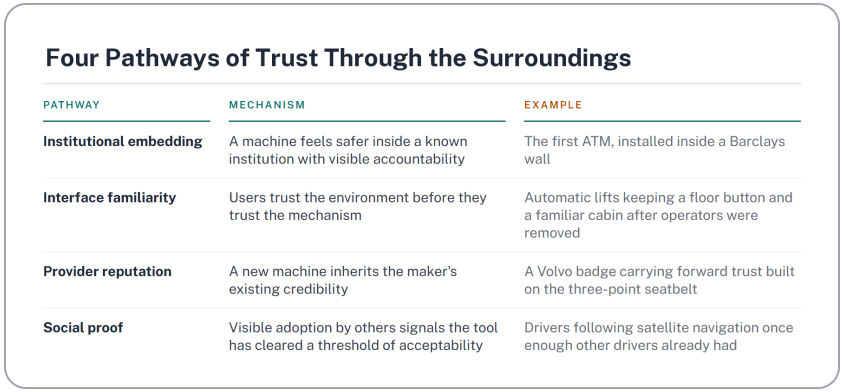
Provider Reputation

The third pathway is provider reputation. The trust literature repeatedly supports the idea that reputation, competence, and reliability are among the strongest drivers of trust judgements (Doney & Cannon, 1997; McKnight et al., 1998). That means a new machine often inherits credibility not from the category of machine it belongs to, but from the maker deploying, endorsing, or installing it. Volvo built its name on the three-point seatbelt; decades later, a Volvo badge on an unfamiliar safety feature still carries some of that original credibility forward.

Trust in a machine is usually not a direct judgement on the mechanism alone, following the same pattern. It is a layered judgement, made through the Eye of the Trustor, about the mechanism, the maker, the interface, the institution, and the use context all at once.

Social Proof

The fourth pathway is social proof. Satellite navigation spread this way: once enough drivers were visibly following a dashboard voice instead of a paper map, other drivers inferred that the system had crossed some threshold of practical acceptability. This does not prove the tool is reliable. It does make trust easier to extend.



▼ Text version of this table

Pathway	Mechanism	Example
Institutional embedding	A machine feels safer inside a known institution with visible accountability.	The first ATM, installed inside a Barclays wall.
Interface familiarity	Users trust the environment before they trust the mechanism.	Automatic lifts keeping a floor button and a familiar cabin after operators were removed.
Provider reputation	A new machine inherits the maker's existing credibility.	A Volvo badge carrying forward trust built on the three-point seatbelt.
Social proof	Visible adoption by others signals the tool has cleared a threshold of acceptability.	Drivers following satellite navigation once enough other drivers already had.

Table 9.1 Four Pathways of Trust Through the Surroundings. Machine trust rarely comes from the mechanism itself — it is usually borrowed through the institution, interface, provider, or crowd that surrounds it.

Task-Specific Risk

Trust in a machine is also highly task-dependent. Someone who will happily let an AI tool draft an email may want a doctor, not a chatbot, deciding their treatment. It may be much the same technology, but the tasks sit on opposite sides of a line: one is a wrong answer you can correct in seconds, the other is a wrong answer that is hard or impossible to undo. The line is an illustration, not an operational safety rule.

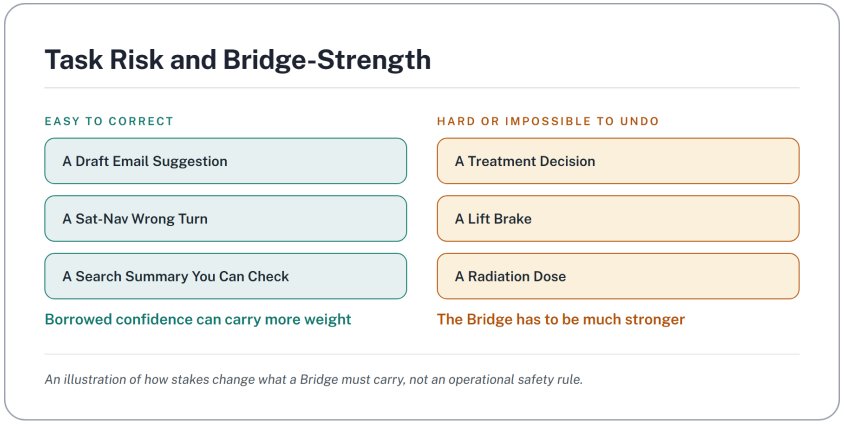


Figure 9.1 Task Risk and Bridge-Strength. Borrowed Confidence carries more weight for reversible, low-stakes tasks and far less once a decision becomes consequential or hard to undo.

This pattern is important because it shows that people do not trust automation in the abstract. They trust it in relation to recoverable versus irreversible risk. If the task is informational, reversible, or low-cost to correct, Borrowed Confidence can carry more weight. If the task is consequential, opaque, or hard to reverse, the Bridge has to be much stronger.

That logic aligns closely with the trust-transfer model in the rest of the book. The Source may be the brand, institution, or workflow through which the machine appears. The Target is the machine's output or action. The Bridge is the embedding itself. The Trustor is deciding not whether automation is good in general, but whether this system, in this setting, for this task, deserves action.

Automation and Calibration

The classic automation literature explains why this matters so much. Lee and See's 2004 review on trust in automation remains the central source for the idea of calibration: trust should match real capability. If trust runs ahead of capability, misuse becomes more likely. If trust lags behind capability, useful systems are ignored or underused.

That is the danger of trust through the surroundings in machine environments. A polished interface, a strong brand, or a respected institution can lend more confidence than the underlying mechanism has earned. The user may not be trusting the machine at all in any clean sense. They may be trusting the wrapper around it, and letting that wrapper carry the output farther than the evidence warrants.

Parasuraman and Riley's 1997 framework is still helpful here too. It names three ways automation goes wrong in human hands: misuse, relying on it too much; disuse, relying on it too little; and abuse, automating without enough regard for the people who must work with the system. These are not simply technical errors, and trust is part of each.

Air France Flight 447 shows the cost of getting the calibration wrong in the other direction. When the aircraft's pitot tubes iced over in 2009, the autopilot disconnected and handed control back to a crew who had little practice flying the aircraft by hand in those conditions, and the French investigators' final report sets out what followed (BEA, 2012). The case is often used to illustrate what has been called the automation paradox: the more reliable a system is in ordinary operation, the less practised the human may be at taking over when it fails.

More recent evidence suggests that even transparency can fail to solve this cleanly. A 2025 systematic review of automation bias in human–AI collaboration found that explanation features can sometimes reinforce misplaced trust in less experienced users. This happens particularly when the explanation is too technical or too simplified to actually inform judgement. The underlying studies were mixed, not uniform (Romeo & Conti, 2025).[3]

More information does not automatically produce better calibration. Explanations can reassure when they should caution. Confidence can be borrowed from the presence of explanation itself.

Familiarity Breeds Trust

One route from trust through the surroundings to Earned Trust is successful interaction. Go back to the lift. Elisha Otis convinced a crowd once, in public, with a cut rope. What actually built lasting confidence was different: millions of ordinary rides afterward, day after day, in which the brake never had to catch because nothing went wrong. Each uneventful ride was a small piece of Direct Evidence, and the accumulated weight of those rides is why most people today step into a lift without a second thought.

That pattern clarifies something important. Users do not rely forever on Borrowed Confidence from brand, institution, or category. A strong first-hand track record creates a new evidence base. Once a machine has been reliable for them personally, trust no longer depends only on what others say about it.

The practical lesson is simple. If organisations want trust in a new machine to deepen, the first stretch of ordinary, uneventful use matters more than any safety brochure. A run of successful early interactions converts Borrowed Confidence into something more direct and durable.

Why Institutions Still Matter

Even so, direct experience does not eliminate the institutional layer. Most complex machines remain too opaque, too dynamic, or too consequential for users to rely on experience alone. A passenger can feel that a flight was smooth, but still not know how the autopilot handled the descent, what its limits are, or what happens when it fails.

Governance therefore remains central to machine trust. No airline passenger personally checks an aircraft's airworthiness. They rely on a chain of aviation authorities, most visibly bodies like the FAA and EASA, that certify the aircraft type, the specific airframe, and the airline operating it before a single seat is sold. People are not merely asking whether the flight feels smooth. They are asking, even if they never phrase it this way, what institutional safeguards stand behind it.

Different countries run this certification chain with different degrees of rigour and transparency, and how quickly institutional trust transfers into a new machine varies accordingly.

When the Surroundings Fail

Because machine trust is so often borrowed, machine failure may not stay confined to the machine. Between 1985 and 1987, a computer-controlled radiation therapy machine called the Therac-25 delivered massive overdoses to at least six cancer patients, killing three of them. Investigators traced the cause to software race conditions that its manufacturer, Atomic Energy of Canada Limited, had not caught before removing the hardware safety interlocks earlier versions relied on (Leveson & Turner, 1993). When hospital staff first reported the burns, the manufacturer insisted the overdoses the machine reported were impossible, and treatments continued on other machines while the investigation dragged on. The institutions that had bought and trusted the Therac-25 absorbed the damage right alongside the manufacturer.

This is one reason deploying any complex, hard-to-inspect machine is a reputational and institutional decision, not just a product one. The stronger the Source lending trust to the system, the larger the backlash when the system violates that confidence.

It also means that responding to a machine failure requires more than a generic apology. The unsafe system has to be stopped or its use limited. The response must then match the violation type and make the correction visible. Competence failures call for acknowledgement, improvement, and verification that the mechanism or process has changed. Integrity failures require clarity about responsibility, truthfulness about culpability, and restored accountability.

That is exactly what AECL's early denials withheld. The technical and institutional corrections must then be observed over time, generating the Direct Evidence through which affected Trustors may judge trust to have been re-earned.

Design Implications

Trust should be transferred deliberately, not by accident. If an institution places a new machine inside a trusted workflow, it should be clear exactly what is being endorsed — usefulness, safety, oversight, limited scope, or only availability. Users need to know what the Bridge actually guarantees.

Calibration beats enthusiasm every time. A system that merely makes users comfortable has aimed at the wrong target; it should instead help them trust at the level the task and the evidence actually support. That means clear boundaries, visible uncertainty, and a legible path for human review — the lesson Air France 447 paid for in the worst possible way.

First experiences carry more weight than they should. A bounded, low-risk first interaction can build the kind of Direct Evidence that strengthens trust appropriately; a confusing or overpromising one can poison the relationship before it has properly begun.

Governance that isn't visible can still protect people, but it cannot reassure them. People cannot borrow trust from rules, audits, or accountability structures they cannot see. The more consequential the use case, the more those structures need to be legible to the user. A certification chain manages this for aviation and, until too late, failed to manage it for an isolated hospital device.

Every deployment of a complex machine spends institutional trust, whether the organisation notices or not. Each one draws on a reservoir built elsewhere — customer trust, employee trust, professional trust, civic trust, brand trust — and that reservoir can deepen adoption or drain away if the system asks for more confidence than it has earned.

Chapter Takeaway

Trust in a machine is rarely trust in the mechanism alone. Usually it starts as trust transferred from something surrounding the machine: an institution, a brand, a workflow, an interface, a provider, a social norm. That Borrowed Confidence is what gets most machine relationships started, from the first cash machine in a bank wall to an autopilot in an aircraft type nobody in the cabin has personally inspected.

This makes machine trust a design and governance problem as much as a performance one. Machines have always entered public life on borrowed credibility, and the real challenge is earning it.

Whether that Borrowed Confidence turns durable depends on what happens next. A system that performs well lets Borrowed Confidence give way to Earned Trust through experience. A system that fails, as the Therac-25 did, sends the damage backward, to the Sources that made it look safe in the first place. That is what the next chapter takes up: what happens once the Target becomes directly observable, and whether the confidence a machine's scaffolding created turns out to be earned. The staged version of that scaffolding — legitimacy borrowed rather than earned — is a problem in its own right, taken up later in Chapter 11, Borrowed Legitimacy.

3 Romeo, G., & Conti, D. (2025): a PRISMA-based systematic review of 35 studies on automation bias in human–AI collaboration, spanning high-stakes domains (healthcare, law, public administration); explanation and transparency features reinforced misplaced trust among less experienced users when explanations were too technical or too simplified, though findings were mixed across the underlying studies. Full discussion: Ch. 14.

Chapter 10

Outcomes of Target-Based Evidence

The preceding chapters examined where confidence is borrowed. This chapter examines what happens once the Target becomes directly observable and can begin to stand—or fail—on its own record. It is the hinge of the book: it closes Part II's tour through the domains where trust is borrowed, and sets up Part III's turn to what happens when a Bridge claims more legitimacy than the Target can support.

But borrowing is not the outcome. It is a provisional way of beginning.

The decisive change comes when the Target becomes directly observable. A customer sees how a company responds when something goes wrong. A patient learns whether the provider follows through. A citizen sees how an institution behaves when its conduct is tested in public. Direct Evidence does not simply add more confidence to what was already transferred. It changes the basis on which reliance is warranted.

This is the point at which the model produces its paired outcomes: Earned Trust and Earned Mistrust, with an honest third position, Unresolved, while the evidence is mixed or thin. All three are Target-based. They rest on what the Trustor has seen for themselves and on what the Target's Public Record documents. None should be confused with the confidence that first crossed the Bridge. This chapter uses target-based evidence for both routes: Direct Evidence from the Trustor's own encounters, and the Target's Public Record of documented conduct, which still reaches the Trustor through an intermediary.

The Crossing Is Not the Outcome

Early in a relationship, a person asked why they rely on a Target will often name the Bridge. A colleague recommended it. My doctor referred me. It is certified by an institution I trust. The answer locates confidence in the introduction rather than in the Target's demonstrated record.

Over time, direct encounter may change the answer. The person may begin to say: I have used it repeatedly; it works when I need it; it dealt fairly with a problem. The introduction has not been erased. It may still matter historically. But it is no longer carrying the full burden of confidence.

That shift is an authorial model proposition, not a survey result this book claims to have directly measured. Existing open confidence batteries ask respondents how much confidence they have; they do not ask who deserves credit for it. The

model therefore names the transition as an attribution shift and treats it as a useful interpretive question, not as a settled empirical finding.

Outcomes of Target-Based Evidence

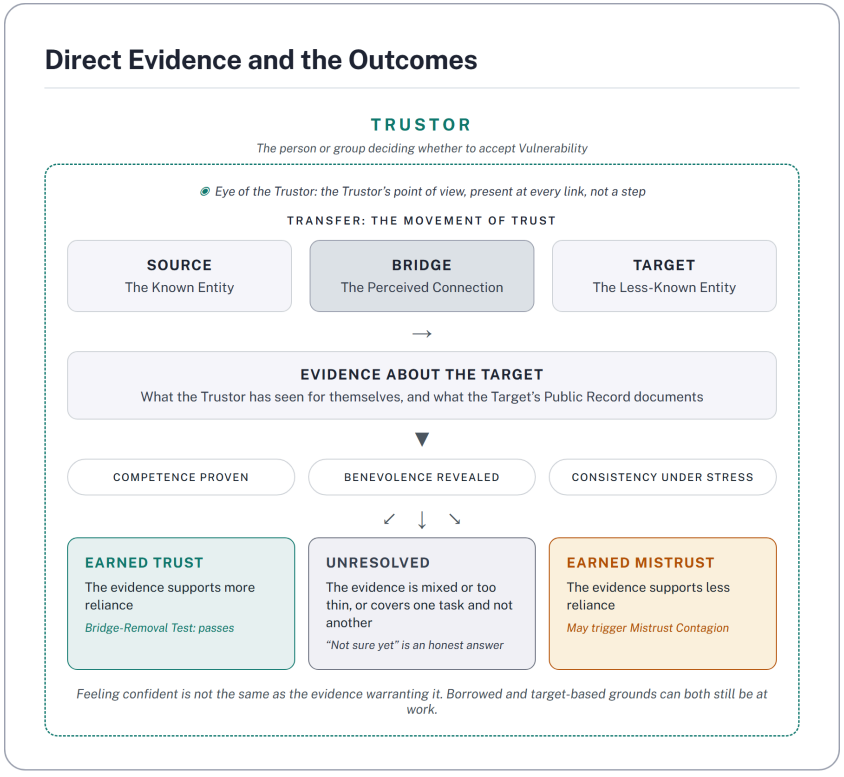


Figure 10.1 Direct Evidence and the Outcomes. Once a Target becomes directly observable, Direct Evidence can support Earned Trust or Earned Mistrust, or leave the judgement unresolved.

Earned Trust

Earned Trust is target-based confidence that survives without the Source that first introduced the Target. It is reached by the Trustor, through the same bounded lens that earlier weighed whether to accept Vulnerability at all, now reading the Target's own record instead of a borrowed signal. It is not simply trust that has lasted a long time, or trust that feels settled; a Trustor can feel sure and still be miscalibrated. A Bridge can be old and still depend on the original endorsement, ranking, platform, certification, or social proof that made it seem safe at the beginning.

Whether that record has actually taken over is worth testing directly. The Bridge-Removal Test, spelled out further on, is how.

Earned Mistrust

Earned Mistrust is the equally important paired outcome. It arises when target-based evidence justifies reduced reliance. The point is not that a person interpreting through the Eye of the Trustor has become emotional, cynical, or hostile. The point is that the same lens, reading Direct Evidence, has found that less reliance is warranted. The verdict belongs to the Trustor; the evidence belongs to the Target.

A Target may fail to deliver the competence it claimed, evade responsibility when it matters, or behave inconsistently under conditions that expose its real priorities. In those cases, reducing reliance is not the absence of a trust outcome. It is an outcome of direct encounter. It is as target-based as Earned Trust.

Now suppose the restaurant from the earlier example had gone the other way. On a later visit you notice a health inspection notice in the window, and learn the kitchen has been cited twice that year. When your colleague asks, you don't repeat your friend's original recommendation, but you also don't recommend the restaurant. You tell them what you saw. That reduced reliance is not suspicion for its own sake. It is the same Direct Evidence doing its work, just pointing the other way.

The same pattern shows up in a documented case, not just a hypothetical one. For more than a century, Wells Fargo's age and size worked as their own Bridge: a customer who had never tested the bank directly could reasonably assume an institution that old wouldn't risk its name on small-time fraud. That assumption stopped holding in September 2016. The Consumer Financial Protection Bureau, the Office of the Comptroller of the Currency, and the City of Los Angeles fined the bank $185 million. They found that employees had opened more than two million deposit and credit-card accounts without customers' permission, to meet sales targets (Consumer Financial Protection Bureau, 2016). The following year, after an expanded outside review, the bank itself raised that estimate to roughly 3.5 million potentially unauthorised accounts (Berliner, 2017). The bank fired 5,300 employees over the practice.

Customers did not need a lecture on trust theory to update on that evidence. Their own dealings with the bank may have shown nothing wrong. What they updated on was the bank's Public Record: the documented findings of the regulators. A survey of 1,500 bank customers, taken within weeks of the fine, found that negative perceptions of Wells Fargo had jumped from 15% to 52%. Positive perceptions fell from 60% to 24%. Of the bank's own customers, 14%

said they had already switched banks over it, and a further 30% were weighing whether to leave (CNN Money, 2016). The following year, the Harris Poll's annual Reputation Quotient recorded Wells Fargo's fall from 70th to 99th place out of the hundred most visible companies in the United States — the single largest year-on-year drop in the poll's eighteen-year history (The Motley Fool, 2017). None of that is the public turning cynical for its own sake. It is what Earned Mistrust looks like at scale: the institution's own conduct, not the century of reputation that had stood in for it, decided how much reliance customers were willing to extend.

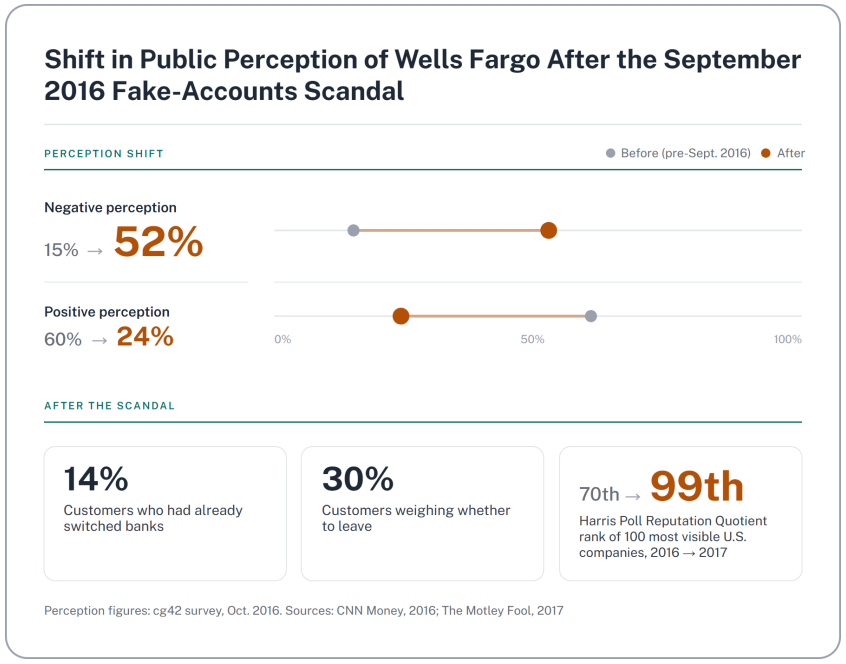


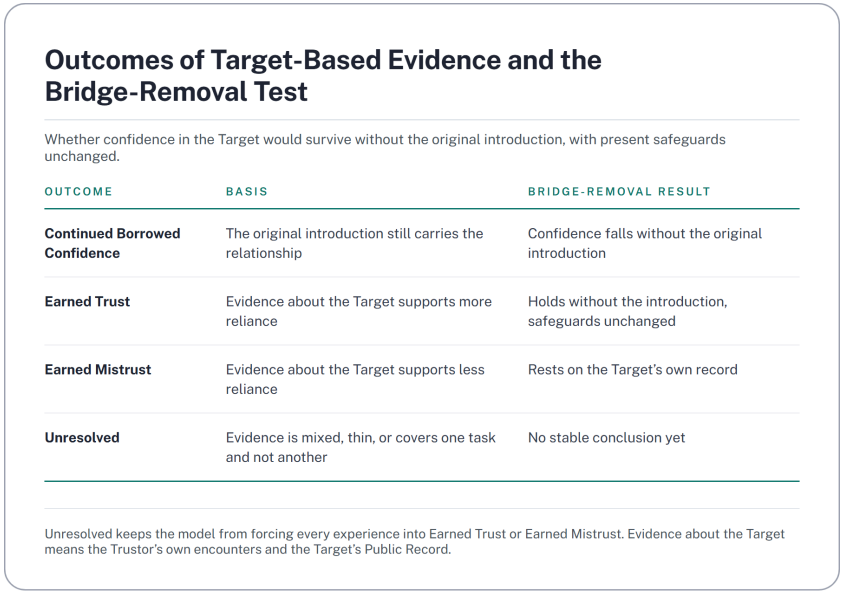
Figure 10.2 Shift in public perception of Wells Fargo following the September 2016 fake-accounts scandal (CNN Money, 2016; The Motley Fool, 2017).

This distinction matters because it stops the framework from treating trust as the desired result and mistrust as a mere malfunction. A well-calibrated relationship can lead to more reliance when the evidence warrants it and less reliance when the evidence warrants that instead. The central standard is not maximal trust. It is better-directed reliance.

Not every encounter settles the question. The evidence can be mixed, too thin, or good for one task and silent on another: a mechanic trusted for years with brakes has shown nothing yet about a gearbox. Table 10.1 gives that position its own row, Unresolved. It is not a failure to decide. It is an honest account of what

the evidence can carry so far, and while it lasts, Borrowed Confidence and the Target's own record can both still be doing some of the work.

Earned Mistrust is distinct from Mistrust Contagion. Contagion describes suspicion moving away from a failed link through a wider Bridge or network; it is neither gated by direct encounter nor limited to one direction. Earned Mistrust concerns what Direct Evidence warrants about the observed Target. Chapter 12 develops how failure can then travel through Sources, Bridges, Targets, and networks.



Outcomes of Target-Based Evidence and the Bridge-Removal Test

Whether confidence in the Target would survive without the original introduction, with present safeguards unchanged.

OUTCOME	BASIS	BRIDGE-REMOVAL RESULT
Continued Borrowed Confidence	The original introduction still carries the relationship	Confidence falls without the original introduction
Earned Trust	Evidence about the Target supports more reliance	Holds without the introduction, safeguards unchanged
Earned Mistrust	Evidence about the Target supports less reliance	Rests on the Target's own record
Unresolved	Evidence is mixed, thin, or covers one task and not another	No stable conclusion yet

Unresolved keeps the model from forcing every experience into Earned Trust or Earned Mistrust. Evidence about the Target means the Trustor's own encounters and the Target's Public Record.

▼ Text version of this table

Outcome	Basis	Bridge-Removal result
Continued Borrowed Confidence	The original introduction still carries the relationship	Confidence falls without the original introduction
Earned Trust	Evidence about the Target supports more reliance	Holds without the introduction, safeguards unchanged
Earned Mistrust	Evidence about the Target supports less reliance	Rests on the Target's own record
Unresolved	Evidence is mixed, thin, or covers one task and not another	No stable conclusion yet

Table 10.1 Outcomes of Target-Based Evidence and the Bridge-Removal Test. Whether confidence in the Target would survive without the original introduction, with present safeguards unchanged.

The Bridge-Removal Test

Earned Trust and Earned Mistrust are both verdicts on the Target's own record. But a reader still needs a way to tell, in a specific case, whether that record has actually done the work, or whether the original introduction is still quietly carrying the weight. The Bridge-Removal Test is that check.

The question is simple to ask and hard to dodge: if you had never received the original introduction, would what the Target has shown since still justify relying on it? Everything that still protects you today stays where it is while you ask it: the regulation, the certification, the oversight. The test removes the introduction, not the safeguards. Relying on a safeguard that is still working is not a sign that trust is unearned.

Take a tradesperson hired on a neighbour's recommendation. The first job goes well, and a second follows, and a third. At some point the homeowner would still call that tradesperson even if the neighbour who made the introduction moved to another city and was never heard from again. The neighbour's word got the relationship started; it is no longer what is keeping it going. That is a pass.

A pass is a diagnostic sign, not proof that trust is warranted. It means the Target's own record has become load-bearing: reliance is now grounded in what the Target has actually done, not in who vouched for it. A fail does not mean the relationship is unsound. It means reliance, however sensible, is still borrowed rather than earned, and still depends on the credibility of whoever made the introduction.

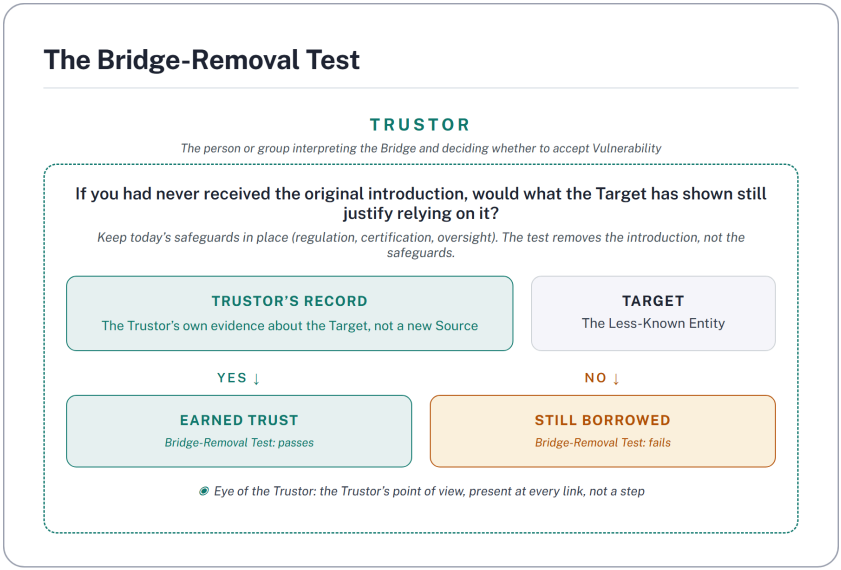


Figure 10.3 The Bridge-Removal Test. If confidence in the Target would survive without the introduction that first carried it, with today's safeguards left in place, trust has crossed from borrowed to earned.

What takes over is not a new Bridge. It is the Trustor's Record, the Trustor's own accumulated history of positive and negative interactions with the Target: evidence about the Target, not a new Source. Each further interaction adds to that record rather than to the original Borrowed Confidence. A new chain begins only when the Trustor recommends the Target to someone else. Then the Trustor is the Source, the recommendation is the Bridge, and the other person is the one deciding.

A Trustor's Record is not a simple running tally. Three features decide how much it can carry. The first is volume: two visits say less than twenty. The second is range: a Record built only in easy conditions has never been tested. A builder who has finished three small jobs on time has not yet shown what happens when the council inspector fails the footings or the supplier runs short of timber. The third is recency. An interaction from last month counts for more than one from five years ago, because the Target may have changed in between.

Nor are the entries weighed evenly. A courier who delivers forty parcels on time and then loses the one holding a passport has not built a Record of thirty-nine successes. In the customer's memory it is a Record of one loss. Chapter 12 returns to this asymmetry as a feature of whole systems. Within a single relationship, it means a Record is slow to build and quick to dent, and a Trustor who has been burned once may discount the long run of good conduct that came before.

A Record also has borders. It covers the tasks the Trustor has actually seen the Target perform. A mechanic trusted for years with brakes and tyres has earned nothing yet on a gearbox rebuild, let alone on advice about which car to buy. A bank that has held someone's savings without error has shown little about the quality of its financial planning. Stretching a Record past the ground it was built on is Borrowed Confidence in a new form: the Trustor lends their own past experience to a claim it never tested. The same limit applies when the Trustor becomes a Source. What passes to the next person is a recommendation, which is a new Bridge. It carries a summary of the Record, never the Record itself.

The test is deliberately diagnostic rather than empirical. No open survey examined for this book, including the General Social Survey, World Values Survey, or Latinobarómetro, asks the counterfactual required to measure it directly. The book offers it as a way to inspect a relationship, not as a claim that a dataset has already validated it.

How Direct Evidence Is Read

Three evaluative dimensions help make this outcome legible. They are not a universal psychological law, nor a checklist that mechanically produces trust. They are dimensions on which a Target's record becomes visible through the Eye of the Trustor.

Competence proven concerns whether the Target can do what it represents itself as able to do. A Source can predict competence; a Target must demonstrate it through outcomes.

Benevolence revealed concerns whether the Target takes the relying party's legitimate interests seriously, especially where opportunism would be easy or difficult to detect. A mistake acknowledged and repaired can reveal more than a polished assurance given in advance.

Consistency under stress concerns whether the Target's stated standard survives a moment when pressure makes a different choice tempting. Ordinary performance is informative. Conduct under public, financial, or operational strain often makes the basis of reliance clearer.

These dimensions do not eliminate interpretation. The same record can be read differently where people face different exposure, risk, or alternatives. That is why the model retains the Eye of the Trustor around the whole chain. The lens does not make evidence arbitrary; it identifies the situated conditions under which evidence is assessed.

What Stress Reveals, and What It Does Not

Institutional-confidence data can show how public assessments changed across a period of visible stress. They cannot, by themselves, show why every change occurred or prove that one historical event caused it. The distinction is important.

Large public surveys can show this kind of movement. Latinobarómetro's published reports have tracked confidence in institutions across Latin America through years of corruption scandals, recessions and political upheaval, including the period of Brazil's Operation Car Wash investigations (Corporación Latinobarómetro, 2018). An institutional shock cannot be represented as a single undifferentiated outcome. Direct Evidence is evaluated against the particular Target, the particular conduct at issue, and the institutional context in which that conduct is understood.

That is a caution too. Even where confidence fell, a single event like Operation Car Wash cannot be assumed to explain the movement. Corruption scandals, recessions, and political upheaval reshaped confidence in different countries in different years, and the survey numbers alone cannot separate what each one caused.

The appropriate inference is therefore limited but useful: when public institutions face serious, visible strain, confidence is not guaranteed to remain attached to their prior reputation. Direct Evidence can lower warranted reliance, and it can do so unevenly across institutions and settings.

A Second Temporal Test

A different dataset provides a related pattern through a different institution and population. Hout and Hastings' analysis of the General Social Survey's 2006–2010 three-wave panel follows 821 United States respondents who answered the relevant confidence question at all three waves (Hout & Hastings, 2016). The share reporting "hardly any confidence" in the people running financial institutions rose from 15% in 2006 to 23% in 2008 and 42% in 2010.

This is repeated panel evidence, not a causal experiment. It does not prove that the financial crisis alone caused every respondent's change. It does show that, across the same respondents, confidence was lower after the period in which financial institutions' conduct had become publicly consequential. The evidence is consistent with the chapter's narrower proposition: confidence is revisable when conduct becomes more directly observable under stress.

These survey results do not test the attribution shift. Nor do they test the Bridge-Removal Test. They do something more modest and still valuable: they show

that confidence is not a fixed asset bestowed permanently by an earlier Source. It remains exposed to evidence about the Target's conduct.

The Standard Is Better-Directed Reliance

The practical consequence is not that every Bridge should be distrusted. It is that Bridges should be understood as provisional where Direct Evidence is still thin.

A responsible Source can introduce. A strong Bridge can make a Target easier to assess. But neither can permanently substitute for the Target's record. The closer direct experience comes to the decision, the more clearly the basis of reliance should shift from what others said to what the Target has shown.

That shift can produce Earned Trust: confidence that remains without the original introduction, with present safeguards unchanged. It can produce Earned Mistrust: reduced reliance that the evidence justifies. Or it can leave the judgement unresolved. These judgements improve calibration only to the extent that the evidence supports the change in reliance.

Worked Example: Recommending a Restaurant

You try a new restaurant because a friend recommended it. The friend is the Source; the recommendation is the Bridge; the restaurant is a Target you have not yet judged for yourself.

You go back several times over the next year. Eventually, a colleague asks if the restaurant is any good. You don't mention your friend. You answer from your own visits: the food has been consistently good, and once, when an order came out wrong, the staff fixed it without an argument.

That is the test. You are no longer relying on the friend who introduced you; you are relying on what you have seen for yourself. Your own experience has become a Trustor's Record strong enough to stand on its own, and in the act of answering, you have become a Source for someone else.

Word of mouth is not a separate channel from the one this chapter has been describing. It is the same mechanism succeeding at scale, one Trustor at a time. Borrowed Confidence is a means, not the end goal. A brand's endorsement buys enough attention that the Trustor tries the Target directly. If that direct experience is good enough, often enough, the Trustor builds a record of their own that no longer needs the original endorser to hold it up. Once that record passes the Bridge-Removal Test, the trust rests on the Target's own conduct rather than on the transfer, and the relationship can run in reverse: a Trustor

with enough good direct experience becomes a Source in their own right, recommending the Target to people who have never used it.

The next question is design. If a Bridge should not conceal the distinction between borrowed and earned confidence, what should it make visible? How should organisations signal competence without overstating it, disclose incentives without destroying useful orientation, and remain accountable when the Target fails?

Chapter Takeaway

Crossing the Bridge is not the outcome. It is the beginning of the part Borrowed Confidence cannot finish: what the Target actually does once it is observed directly.

Direct Evidence can lead to Earned Trust, Earned Mistrust, or an honest Unresolved. Earned Trust is confidence that survives the Bridge-Removal Test on its own, because the Trustor's own record with the Target has become load-bearing. Earned Mistrust is reduced reliance that the same evidence justifies, not a failure of the model but an equally legitimate result of it. Unresolved is where the judgement sits while the evidence is mixed or thin.

None of these positions makes a Bridge worthless. It only means a Bridge should be treated as provisional until Direct Evidence is in. Part Three turns next to Bridges that claim more legitimacy than they have, and what happens when that borrowing is exposed.

Part III

The Breaks: Failure and Response

When borrowed credibility goes wrong: borrowing that is read as more than it warrants, the Mistrust Contagion that follows when a Bridge breaks, and what a credible response to failure requires over time.

Chapter 11

Borrowed Legitimacy

If Trust Transfer is the useful side of modern complexity, Borrowed Legitimacy is what happens when that useful borrowing is over-read. It describes a person, product, institution, or system that appears more trustworthy than its connection to something already trusted actually warrants, because being placed near, inside, beside, or beneath that Source is read as more than it is. The Target does not build confidence from Direct Evidence. It acquires more confidence by proximity than the proximity can justify.

Borrowing itself is not the problem. Much of social life depends on legitimate borrowing. A hospital lends credibility to a new treatment protocol. A school lends reassurance to a learning platform. A respected journal lends visibility to a researcher. A marketplace lends initial confidence to an unfamiliar seller. These are normal and often necessary shortcuts in a world too complex to inspect from first principles.

The problem begins when the Bridge carries more meaning than it can justify. The Trustor sees a real Bridge, but overreads what the Bridge guarantees. A badge is read as competence when it signals only identity. A platform placement is read as endorsement when it reflects popularity, payment, or optimisation. A partnership is read as deep due diligence when it may mean little more than access, affiliation, or co-marketing.

Borrowed Legitimacy is therefore a question of surplus inference rather than fraud. It concerns the extra trust people supply when a visible connection implies more checking, more responsibility, or more shared standards than actually exist.



Figure 11.1 Bridge Quality and Overreading. A verified badge, a top ranking, or a partnership line each state one narrow fact, but Trustors can read a far larger guarantee into them.

What Gets Borrowed

What moves in these situations is credibility, not merely attention, and it moves across multiple dimensions. Mayer, Davis, and Schoorman's (1995) classic trust model remains useful here because legitimacy borrowing often transfers inferred competence, integrity, and benevolence together. If the Source is seen as capable, ethical, and well-intentioned, the Target can inherit all three at once.

That bundling effect is what makes Borrowed Legitimacy so powerful. The Trustor does not merely think, this product may work. The Trustor often thinks, this seems like the kind of thing responsible people would allow, or this seems to sit inside a serious system. The Source is doing much more than introducing the Target. It is supplying moral and institutional cover.

Stewart's 2003 work on Trust Transfer traces why this happens so quickly online. When a trusted Source appears linked to an unknown Target, people often use the known Source as a cognitive shortcut rather than evaluating the Target independently. Metzger, Flanagin, and Medders (2010) show a related pattern in credibility research: people routinely use heuristics such as reputation, endorsement, consistency, and expectancy rather than verifying underlying substance.

Those shortcuts are what let Borrowed Legitimacy scale so effectively in digital environments: the platform, logo, ranking, recommendation, interface, affiliation line, or co-branded frame does the work before slower reasoning begins.

The Bridge Quality Problem

The core issue is not whether a Bridge exists. It is what kind of Bridge it is. The six Bridge Forms can also be compared by the accountability behind them, from institutional endorsements to reputational endorsements to weak proximity signals. This is a comparison of their basis, not a second set of Bridge Forms. Borrowed Legitimacy becomes dangerous when Trustors respond to weak Bridges as though they were strong ones.

An institutional Bridge usually carries real accountability. A medical approval, a legal certification, or a professional accreditation signals that a process with standards, documentation, and consequences exists behind the claim. A reputational Bridge carries softer but still meaningful weight: a respected editor, analyst, investor, or publication has placed some portion of their own credibility at stake.

A proximity Bridge is different. It may be visually persuasive while substantively thin. An as-seen-on logo collage, a featured label, a shallow partnership, or a placement high in a commercial interface can trigger the same trust response without carrying the same accountability. The form of the Bridge resembles endorsement, but the substance does not.

This asymmetry is one of the central vulnerabilities of modern trust environments. The mind is sensitive to the presence of connection, not always to the quality of the connection.

Why People Overread Signals

People overread these signals for reasons that are psychologically ordinary, not exotic. Under uncertainty, time pressure, and information overload, heuristics are efficient. Tversky and Kahneman's (1974) work on judgement under uncertainty identifies why people use available cues rather than complete investigation. Cialdini's (2021) authority and social-proof principles describe why visible endorsement, popularity, and institutional framing can carry disproportionate force.

McKnight, Cummings, and Chervany (1998) add another layer that matters in this chapter. Initial trust often forms before direct experience, through institutional familiarity and structural assurance. If something appears inside a familiar setting with visible order and procedural seriousness, people may treat that setting itself as evidence that someone upstream has already done the checking.

That is a rational adaptation much of the time. It is simply not always deserved. Borrowed Legitimacy works because it recruits useful mental shortcuts built for

ordinary cooperation and applies them in environments where the Bridge may be weak, strategic, or misleading.

This is also why the line between convenience and manipulation can be blurry. The cue is not always false. It is often merely incomplete.

Common Forms

Over-reading tends to happen in a recurring set of forms across domains. Each form can be warranted. Each becomes Borrowed Legitimacy when it is read as more than it is.

The first is brand adjacency. An unknown product feels safer because it appears inside a trusted store, app, or media environment. In e-commerce research, McKnight, Choudhury, and Kacmar (2002b) showed how website quality and reputation can create trust before any direct experience with the seller or offer.

The second is institutional embedding. A tool appears inside a school, hospital, employer workflow, or government service, and the surrounding institution lends assurance the Target has not independently earned. In workplaces this is particularly important, because staff often cannot evaluate the software or supplier they are handed and rely instead on the judgement of the employer that chose it.

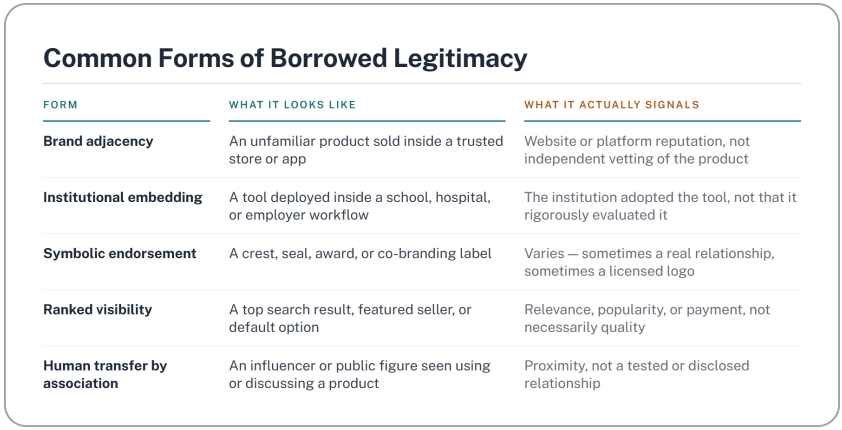
The third is symbolic endorsement. A crest, title, award, seal, partnership label, or co-branding arrangement signals seriousness. Sometimes that signal is deserved. Sometimes it exceeds the actual relationship.

The fourth is ranked visibility. Search results, featured sellers, recommended products, and default options often feel as though someone has vouched for them. But a ranking may reflect relevance, popularity, personalisation, commercial incentives, or machine optimisation rather than quality in the sense users imagine.

The fifth is human transfer by association. Influencers, experts, executives, and public figures move confidence toward Targets simply by being seen with them, discussing them, or appearing to use them. The relationship need not be deeply scrutinised for the transfer to happen.

These forms look different on the surface, but they share the same structure: the Source has Earned Trust on one basis, and the Target borrows that trust on another. They are not a new taxonomy. These five patterns are the same six Bridge Forms from Chapter 2, sorted here by how they tend to be abused rather than how they are built: brand adjacency is Co-presence, institutional embedding is Embedding or Integration, symbolic endorsement is

Endorsement or Certification, ranked visibility is Ranking or Placement, and human transfer by association is Endorsement or Co-presence.



▼ Text version of this table

Form	What it looks like	What it actually signals
Brand adjacency	An unfamiliar product sold inside a trusted store or app	Website or platform reputation, not independent vetting of the product
Institutional embedding	A tool deployed inside a school, hospital, or employer workflow	The institution adopted the tool, not that it rigorously evaluated it
Symbolic endorsement	A crest, seal, award, or co-branding label	Varies — sometimes a real relationship, sometimes a licensed logo
Ranked visibility	A top search result, featured seller, or default option	Relevance, popularity, or payment, not necessarily quality
Human transfer by association	An influencer or public figure seen using or discussing a product	Proximity, not a tested or disclosed relationship

Table 11.1 **Common Forms of Borrowed Legitimacy.** The same pattern recurs across five common forms, from brand adjacency to human transfer by association.

When Borrowing Becomes Abuse

The over-reading becomes deliberately exploitative when the Source–Target Bridge is arranged, framed, or allowed to be interpreted in ways that predictably exceed what the Source has actually warranted. The issue is not only factual falsity. It is foreseeable over-interpretation.

A platform may know that users read a verification badge as a quality signal even when it was intended only as identity confirmation. A publisher may know that sponsored placements are read as editorial selection. An institution may know that embedding a tool inside its workflow causes users to infer vetting standards the institution has not seriously applied.

At that point, the defence that no explicit false claim was made becomes weaker. Trust environments are built from implications as much as from statements. If a system consistently invites Trustors to infer more than is justified, it is still shaping a misleading Trust Transfer.

Design therefore carries a moral weight as well as a commercial one. Trust cues are not decorative. They route Vulnerability.

Institutional Wrappers as a Force Multiplier

Borrowed Legitimacy grows much stronger when it arrives inside a trusted institution. The institution adds polish and reach, and it hides the real limits of what sits inside it. Most people met Theranos not through its science but across a pharmacy counter.

In 2013 Theranos began offering finger-prick blood tests inside Walgreens stores in Arizona. Customers did not read validation studies. They saw a chain they had used for years, a pharmacist behind the counter, and a company board that included two former US secretaries of state. The Bridge ran from Walgreens to Theranos, and it carried far more than the device had earned.

Walgreens had not independently verified the technology it was hosting. After the Wall Street Journal reported in 2015 that most tests were not being run on Theranos's own machines, the damage ran both ways. Walgreens ended the partnership in 2016, and Theranos voided two years of results from its flagship device (Carreyrou, 2018).

The wrapper did the persuading. The pharmacy counter was a legitimacy cue, not evidence, and it reassured customers long before anyone had tested what stood behind it.

That is the signature danger of Borrowed Legitimacy inside institutions. Shopfronts, logos, and famous board members can reassure people well before they inform them.

Mistrust Travels Upstream

One reason Borrowed Legitimacy matters so much is that when the Target fails, the damage often moves upstream. The Source that lent the credibility is blamed as well. If a respected institution embeds a harmful tool, users do not confine the disappointment to the tool. They revise their view of the institution that seemed to stand behind it.

This is the reverse flow the next chapter treats in full as Mistrust Contagion: the same Bridges that carried confidence forward can carry suspicion backward.

Borrowed Legitimacy is therefore expensive even when it appears efficient: an institution spends reputational capital whenever it lets another Target stand inside its trust envelope. If that Target fails badly enough, the institution loses not just the transaction but part of the reservoir that made future Trust Transfer possible.

The danger is cumulative. Once users learn that Bridges are weaker than they look, they become less willing to cross even the good ones.

Why Response Is Harder Than Transfer

Confidence often moves faster than a failed trust relationship can be rebuilt. The shortcut that lets people extend confidence quickly does not have an equal and opposite mechanism after betrayal. Once a Trustor judges that Borrowed Legitimacy was abused, future signals are read more sceptically. Reassurance alone cannot reverse the original failure.

The appropriate response depends on the type of violation: competence failures and integrity failures do not respond to the same corrective strategy, and visible change tends to do more than apology or denial. Chapter 13 develops that response in full.

That matters here because legitimacy borrowing often creates mixed failures. The Target may fail on competence, but the Source may appear to have failed on integrity by allowing the Target to borrow too much confidence. The misleading Bridge must first stop carrying further confidence. The response must then address both the immediate failure and the Bridge that allowed overconfidence to form, and later evidence must show whether the revised relationship remains accurate and accountable.

An apology for the outcome is rarely enough. The Source has to show what relationship actually existed, what checks failed, what was stopped, what was corrected and verified, and what will be tracked over time. All of that has to pass through the Eye of the Trustor, now primed to weigh the response more warily, and no amount of sincerity substitutes for Direct Evidence the Trustor can actually use.

Chapter Takeaway

Borrowed Legitimacy is what happens when trust moves farther than the underlying Bridge can support. A Source with genuine credibility lends it, intentionally or not, and the Trustor reads the loan as covering more than it does.

Borrowing itself is often necessary and justified; that is Borrowed Confidence, and it is not a failure. Borrowed Legitimacy begins when weak Bridges are read as strong ones. Then legitimacy becomes a kind of unpriced subsidy. The Target gains confidence cheaply, the Trustor takes hidden risk, and the Source silently spends reputational capital it may not recover if the Bridge fails.

The dark side of Trust Transfer is rarely manipulation in the crude sense. It is the routine over-reading of connection, and in modern systems legitimacy can be borrowed at scale. When the Target fails, the damage moves upstream to the Source that lent it credibility: the same reverse flow the next chapter examines directly as Mistrust Contagion. Chapter 14 turns this into design tests: scope, basis, accountability, and the rule that a signal should mean only what it can bear.

Chapter 12

Mistrust Contagion

Trust Transfer has a dark reverse. If confidence can move forward from a known Source to an unfamiliar Target, suspicion can move backward and outward across the same links. A failure rarely stays where it began. It travels. It contaminates endorsers, platforms, adjacent brands, institutions, and sometimes entire categories.

This book calls that pattern Mistrust Contagion. It is not merely disappointment with one product, one spokesperson, or one system. It is the spread of doubt through the network that once made trust possible. The same Bridges that carried reassurance become channels for alarm.

That pattern follows directly from one fact: modern trust is networked. When people rely on indirect signals rather than direct proof, they are not only borrowing confidence from the Source. They are also exposing the Source to blowback if the Target fails. The more trust was transferred on the way in, the more mistrust can travel on the way out.

This is one reason modern scandals feel larger than their immediate facts. They are rarely judged as isolated events. They are read as evidence that upstream vetting failed, that surrounding institutions were weaker than assumed, or that an entire environment has become less reliable than it looked.

Why Mistrust Spreads

Mistrust spreads for the same reason trust spreads: people cannot investigate everything directly. When a failure occurs, Trustors again use shortcuts. They ask what else is connected to this? Who vouched for it? Who allowed it? What other things that looked safe may rest on the same weak Bridge?

These are not paranoid questions. They are efficient ones. If a respected Source lent confidence to something that failed, the failure is informative not only about the Target but also about the Source's judgement, oversight, incentives, or competence.

This is exactly what the trust-transfer framework predicts. The Source, Target, Bridge, and Trustor were linked on the way forward. Once the Target violates expectations, the Trustor updates not only the Target but the Bridge and the Source as well. Suspicion travels upstream because the original confidence did too.

The psychology is familiar. Expectancy violation drives it: people react strongly when a trusted pattern breaks. Reputation heuristics also reverse direction.

What once signalled safety can become a warning sign: if this institution missed that, what else might it be missing?

Four Directions of Contagion

The direction doubt travels depends on which link failed. Four cases cover the ground, and they are the four failure types set out in the front matter and shown in Figure 0.3.

Source failure sends doubt forward. When the Source is discredited, everything it introduced is reappraised, whether or not those Targets did anything wrong. The endorsement that once opened doors now invites the question of what else was waved through on the same authority.

Target failure sends doubt backward. When the Target fails on its own merits, the Trustor revises their view of whoever vouched for it, because the failure is evidence about the vouching as much as about the thing vouched for. This is the flow traced through the FTX collapse below, and again through the resignations in the illustration at the end of this chapter.

Bridge failure sends doubt both ways at once, and sideways as well. When the Bridge proves weaker than it looked, the damage reaches the party that issued the signal, the party that carried it, and the next signal of the same kind the Trustor meets. A badge found to certify less than users assumed devalues every badge that resembles it.

Network failure spreads the cost across a whole class. When many Targets lean on the same Bridge type or the same source of legitimacy, the failure of one reprices confidence in all of them. That is the category collapse discussed later in this chapter: the weakness sits in shared structure rather than in any single link.



Figure 12.1 Mistrust Contagion. Doubt can move from a failed Source toward its Targets, from a failed Target back toward its Source, both ways when the Bridge fails, or outward across a shared network.

Real events often fire more than one of these at once. FTX began as a Target failure, became a Bridge failure as the vetting around the company was re-examined, and ended as a Network failure across an entire asset class. That compounding is why some failures read as far larger than their own facts.

When the Bridge connecting two people is personal rather than institutional, doubt can skip formal channels altogether. A Trustor who has recommended a Target from their own experience has become a Source, as Chapter 10 describes, and can hear about a bad interaction straight from the person they recommended it to. That Trustor can carry the warning straight back to whichever Source or Target first earned their own confidence. None of this needs a review, a complaint form, or any of the infrastructure a Bridge failure usually has to travel through. The same direct channel that carries a recommendation forward can carry doubt backward, faster and with less friction than any formal Bridge. It is still a Bridge, a personal one; what it skips is the public route the original confidence travelled.

The Asymmetry Problem

Mistrust is often more contagious than trust because losses weigh more heavily than gains. The mind usually needs less evidence to become cautious after a betrayal than it needed to become open in the first place. Trust can be extended on partial information. Once broken, it often demands stronger proof to return.

Repair is therefore harder than formation. A good experience can build confidence gradually. A sharp violation can collapse it quickly. In networked settings, that collapse is not confined to the immediate point of failure.

The asymmetry matters for institutions acting as trust Bridges. They may spend years building a reputation for competence and integrity, then weaken it through one badly handled endorsement, one unsafe integration, or one public demonstration that the visible Bridge was stronger in appearance than in substance.

The result reaches beyond local disappointment. It is a reduced willingness to rely on similar signals later. Future badges, recommendations, and affiliations all carry less weight because the Trustor has learned that the Bridge type itself may be unreliable.

The FTX Lesson

Chapter 1 used FTX because the associations around it lowered the threshold for reliance before the company's own record was available. The collapse then ran the other way. The fraud was treated as evidence about market discipline and elite judgement, not only about one founder. That is the compounding described above. The details are already in Chapter 1.

The pattern is consistent. When confidence is transferred on the way in, blame is shared on the way out.

How Platforms Amplify Contagion

Digital platforms intensify Mistrust Contagion because they compress distance between Source and Target while hiding how the connection was made. Rankings, recommendations, featured labels, creator partnerships, and embedded tools all make weak or opaque Bridges feel stronger than they may be.

When something fails inside that environment, users often do not know whether the error came from the seller, the ranking system, the recommendation model, the moderation process, or the platform's incentives.

They only know that the whole experience felt endorsed. So the platform absorbs distrust beyond the immediate failure.

This is one reason platform trust is fragile even when usage remains high. Once users suspect that recommendations are shaped by payment, manipulation, or weak controls, they begin reading many outputs through that lens. A single incident becomes a theory about the system.

Stewart's (2003) original model helps here too. Trust moves through contextual links. In platforms, those links are everywhere and often invisible. That makes them efficient when they work and highly vulnerable when they fail.

Opaque Supply Chains and System-Wide Suspicion

Mistrust spreads fastest where the Bridge is opaque at every level. A parent buying infant formula cannot see the dairy farm, the collection station, the factory, or the test lab. They rely on a brand and a regulator to have seen those things for them.

In 2008 that reliance broke in China. Milk suppliers had been adding melamine, an industrial chemical, to watered-down milk so it would pass protein tests. Sanlu, one of the country's best-known formula brands, was at the centre, but inspectors found the chemical in products from 22 companies. Around 300,000 infants fell ill, and at least six died (Gossner et al., 2009).

The mistrust did not stay with Sanlu. It moved to the other domestic brands, to the regulator that had exempted leading firms from routine inspection, and to Chinese-made formula as a whole. Parents turned to imported brands. Demand was still so heavy five years later that Hong Kong limited travellers to two cans each in 2013.

That is why failures in opaque systems are rarely read as isolated faults. People read them as signals about oversight, incentives, and whether institutions have been asking for more reliance than they earned.

From Incident to Category Collapse

Contagion becomes especially dangerous when people generalise from one failure to an entire class. This is not always irrational. If multiple Targets appear to share similar incentives, weak controls, or common sources of legitimacy, a single failure can be a reasonable basis for wider caution.

But category collapse can also overshoot. A scandal in one hospital system can weaken comfort with digital health tools more broadly. A failure in one autonomous system can taint public perceptions of automation beyond the

specific use case. A manipulative influencer campaign can reduce trust in creator endorsements generally, including legitimate ones.

This is the systemic cost of cheap borrowing. When too many Targets lean on the same Bridge type, one failure can reprice the value of the Bridge for everyone.

The Network Contagion Research Institute's (2021) report, which traced distrust spreading through online anti-vaccination communities, offers one illustration. It does not show that every trust failure spreads virally in the same way. It clarifies the basic logic: distrust can move across adjacent institutions when people perceive shared responsibility, shared incentives, or shared identity.

What Response Requires After Contagion

Responding after Mistrust Contagion is more demanding because the damage is layered. The immediate Target failure has to stop, and the pathways carrying doubt through the network have to be interrupted. The response must address the Target's conduct and the Source or Bridge that made overconfidence possible. Later observation must then establish whether both corrections endure across the affected relationships.

The corrective strategy must still match the violation. Kim, Ferrin, Cooper, and Dirks (2004) showed that how well a response repairs trust depends on whether the alleged violation is read as competence-based or integrity-based. In contagious failures, the Target and Source may be judged on different dimensions: a product may fail on competence while the institution that endorsed it is judged on integrity for allowing the endorsement to stand.

Johnson & Johnson's response to the 1982 Tylenol poisonings, which Chapter 13 examines in detail, sharpens the sequence. The nationwide withdrawal stopped further exposure. Tamper-resistant packaging corrected and visibly verified the product-level change. Continuing public observation allowed consumers to judge whether it endured. The response did not rely on reassurance alone; it generated Direct Evidence that Trustors could evaluate.

Fuoli and Paradis (2014) add a useful communication layer. Response discourse works when it clarifies responsibility, acknowledges the breach, and makes corrective action legible. In contagion settings, Trustors need to know what was stopped, how the Bridge was made safer, and what will continue to be tested. That evidence still has to register through the Eye of the Trustor, now altered by Mistrust Contagion itself.

Why Visibility Matters

Mistrust Contagion spreads most aggressively where relationships were influential but poorly understood. If users never knew what a badge meant, what a ranking optimised for, or what due diligence an institution had actually performed, they will supply their own explanation after failure. That explanation is often harsher than the truth.

Clarity on the way in is therefore valuable on the way out. A trust environment with legible Bridges can contain disappointment more effectively because people understand the limits of what was promised. A trust environment built on ambiguity invites broader contamination.

The principle applies well beyond technology. Public trust in media, finance, health, and government depends partly on whether institutions clearly signal what they are and are not guaranteeing. Ambiguous endorsement is efficient in the short run and expensive in the long run.

The broader point is simple. Unclear Bridges create wider blast radii.

Chapter Takeaway

Mistrust Contagion travels wherever the failed link connects. Often that means the Bridges that carried confidence in the first place, but where the connection is personal, doubt can bypass the original public Bridge and reach the Source along a new person-to-person one. A failure at the Target does not stay local. It moves backward to the Source and outward across related systems, especially when people relied on indirect signals rather than direct proof.

That makes modern trust environments both efficient and fragile. The same networks that let credibility move quickly also let suspicion spread quickly when a Bridge fails.

The practical lesson is not to avoid Trust Transfer. It is to respect its blast radius. Whenever credibility is lent, contagion risk is created. The more invisible the Bridge, the wider the damage when it breaks. The next chapter develops the Contain–Repair–Monitor Model: how to stop the failure and its spread, correct the cause and failed trust relationship, and test whether the correction endures.

Chapter 13

The Contain–Repair–Monitor Model

In 1982, seven people in the Chicago area died after taking Extra-Strength Tylenol capsules laced with cyanide. Investigators concluded that someone had tampered with the product after it left the factories, most likely by taking bottles from shop shelves, putting cyanide into some capsules, and returning them; the poisoner was never identified (Gaydou, n.d.). Tylenol had held about 35 per cent of the pain-reliever market. After the deaths, its share fell to 7 per cent, and a brand built over decades looked finished (Knowledge at Wharton, 2012).

What the company did next became the textbook case for a reason. It recalled 31 million bottles of Tylenol, at a cost reported at 100 million dollars or more (Knowledge at Wharton, 2012; Gaydou, n.d.). It opened up to the press and the public rather than waiting for a complete picture. Within a few months it brought the product back in triple-sealed, tamper-resistant packaging (Gaydou, n.d.). None of this undid the deaths. None of it could.

The case earns its place in every textbook because it makes a distinction this book has mostly left unexamined. A Bridge can transfer confidence cheaply, in a sentence or a badge. A Target that has just failed in public cannot buy its way back the same way. Johnson & Johnson could not talk its way to the trust it once had. It had to generate Direct Evidence strong enough to sit alongside the worst evidence anyone had ever seen from it, and let that new evidence outweigh the old.

Chapter 11 noted, in passing, that responding to a trust failure is harder than transferring confidence. Mistrust Contagion went further and showed what the response requires once a failure has spread across several Bridges at once. This chapter begins at the level of a single trust relationship, because before a response can extend across a network, it has to work between one Trustor and one Target. It then applies the same logic to platforms and to technical failures the public cannot inspect directly, where a failure is rarely confined to one relationship.

Why Responding to Failure Is Harder Than Transfer

Trust builds slowly and breaks quickly. Daniel Kahneman's work on loss aversion offers one reason: people weigh losses roughly twice as heavily as equivalent gains (Kahneman, 2011). A hundred ordinary transactions with a bank build a slow, low-grade confidence. One fee applied in bad faith can undo most of it in an afternoon. That is not a personality flaw in the customer who reacts that way. It is closer to how evidence gets weighted by default.

Dirks, Lewicki, and Zaheer, writing about repair inside and between organisations, make a related point: trust forms on an accumulated pattern of small evidence, but it can collapse on a single reinterpretation of that evidence (Dirks, Lewicki, & Zaheer, 2009). A traveller who has trusted an airline through a hundred uneventful flights does not need a hundred bad flights to withdraw that trust. One flight can be enough, not because it is statistically decisive, but because it makes the traveller reread the other ninety-nine differently: not as proof of reliability, but as luck that eventually ran out.

This is why repair is a harder problem than transfer, not merely a slower one. Transfer borrows a signal that has not yet been tested. Repair has to work against a signal that has already been tested and found wanting. The Bridge that once made a Target look credible is now competing with the Target's own recent history, rather than standing in for it.

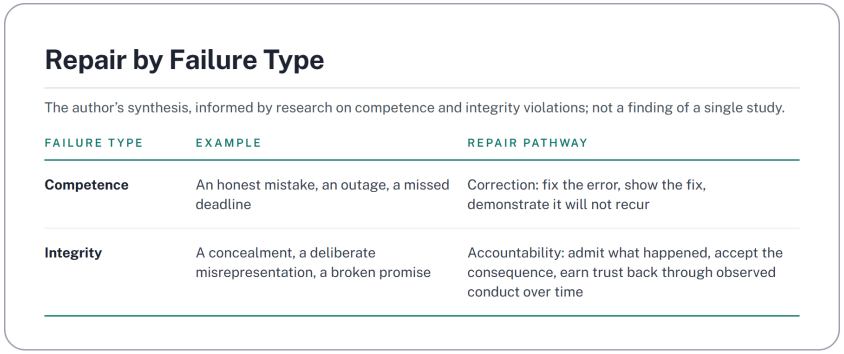
Competence Failure Versus Integrity Failure

Mistrust Contagion introduced Kim, Ferrin, Cooper, and Dirks' finding that repair strategy depends on how a violation is read: as a failure of competence, or a failure of integrity (Kim, Ferrin, Cooper, & Dirks, 2004). That distinction deserves the fuller treatment it did not get there, because it is the single most useful diagnostic in this chapter.

In two studies, participants judged a person accused of an earlier violation. For some participants the allegation concerned competence; for others it concerned integrity. The accused person then either apologised or denied culpability, and later evidence pointed either to guilt or to innocence.

The results ran in opposite directions. When the allegation concerned competence, apologising repaired trust better than denying. When it concerned integrity, denying repaired trust better than apologising. Later evidence mattered too: an apology worked better when the evidence went on to show guilt, and a denial worked better when it showed innocence. That is an experimental finding about how people react to responses to an allegation. It is not advice to deny wrongdoing that has been established, and it does not show that an apology on its own repairs a real organisational failure. What this book takes from it is narrower: people judge competence failures and integrity failures differently, so a repair has to answer the question people are actually asking. My own view, which goes beyond what the study tested, is that competence failures respond to correction, and integrity failures respond to accountability, and only slowly to anything else. Stripped to its simplest form, the distinction is this: a competence failure says you could not do what you promised. An integrity failure says you did not act honestly or responsibly. The

first is a capability problem; the second is a character problem, and readers intuitively demand different repairs for each.



▼ Text version of this table

Failure type	Example	Repair pathway
Competence	An honest mistake, an outage, a missed deadline	Correction: fix the error, show the fix, demonstrate it will not recur
Integrity	A concealment, a deliberate misrepresentation, a broken promise	Accountability: admit what happened, accept the consequence, earn trust back through observed conduct over time

Table 13.1 Repair by Failure Type. The author's synthesis, not a finding of Kim et al. (2004): competence failures respond to correction and evidence of change; integrity failures respond to accountability, and only slowly to anything else.

This distinction explains why so many corporate apologies fail. A company caught lying and responding with "we made a mistake" is applying a competence-failure remedy to an integrity failure, and the mismatch is usually obvious to an audience even when they cannot name why the apology felt hollow. Repair has to match the kind of failure, or it reads as evasion rather than accountability.

Repair Follows the Bridge

Chapter 2 set out six ways a Bridge can carry confidence from Source to Target: Endorsement, Certification, Ranking or Placement, Embedding or Integration, Proxy, and Co-presence. Each Bridge form fails differently, so each requires a different correction. This is the book's own addition to the repair literature, and it follows from treating trust as something that moves through specific structures rather than something a person simply has or lacks.

Endorsement. Sometimes what broke is the endorser's own credibility. A celebrity may be attached to a fraudulent product, or an executive may have vouched for a partner that later collapsed. The correction options are then

narrow. One is withdrawal, where the endorser publicly steps back. Another is clarification, where the endorser explains the limits of what they knew or vouched for. The third is re-endorsement, where a more credible endorser replaces the damaged one. Silence rarely works, because an Endorsement left unaddressed reads as continued vouching.

Certification. When a certifying body's mark turns out to have covered something it should have caught, the correction is recertification: a visible, preferably independent, re-examination that either withdraws the mark or reissues it under tightened criteria. A quiet renewal of the same certificate, with no visible change to the standard, tends to make matters worse because it looks like the certifier protecting itself rather than the standard.

Ranking or Placement. When a Ranking or Placement is revealed to be paid, gamed, or otherwise not what it appeared to be, the correction is ranking transparency: disclose how the ordering is produced, who pays for what, and what a given position does and does not mean. This will not instantly re-earn trust, but it converts an implicit promise the signal could never keep into an explicit one it can.

Embedding or Integration. When trust breaks because a product or service was embedded inside a larger trusted system—a plugin inside a platform or a supplier inside a supply chain—the correction runs through institutional accountability. The host institution has to demonstrate that it can find and remove the failure, not merely that the embedded party has apologised. Until the host shows it can govern what it contains, the Embedding or Integration remains a live wound.

Proxy. When an intermediary appointed to represent the Trustor's interests instead represents someone else's, the correction is intermediary reform: a demonstrable change to whom the Proxy answers, how the Proxy is paid, or which conflicts the Proxy must disclose. Fixing the specific outcome without fixing the incentive that produced it only delays the next failure.

Co-presence. When trust was borrowed from proximity to something credible—sharing a stage, conference, or panel—the correction is association clarification: a clear public statement of what the relationship was and was not. Because Co-presence never certified anything, the fastest way to contain over-reading is usually to state that boundary clearly and early.

The pattern across all six Bridge forms is the same. The response does not run through the Trustor's feelings. It runs through the structure the original Bridge used to move confidence, and it must visibly close the specific channel that allowed the failure through.

Across these Bridge forms, the Contain–Repair–Monitor Model applies in the same way. Contain stops the failed signal from carrying further confidence. Repair corrects the specific Bridge and verifies the correction at implementation. Monitor tests whether the revised Endorsement, Certification, Ranking or Placement, Embedding or Integration, Proxy, or Co-presence remains accurate and accountable over time.

Applying the Model to Platforms

Platforms fail in their own particular way, because the thing under repair is usually a system rather than a single actor.

Consider verification badges. When a platform's blue check meant "this account belongs to who it says it belongs to," it was a certification Bridge with a narrow, checkable claim. When platforms began selling the same badge as a subscription perk, the certification broke: the badge no longer certified identity, but its visual meaning had not caught up. Users kept reading it as verification long after it had become something else. The repair a platform owes here is not a better explainer page. It is deciding which claim the badge actually makes, and making the badge look different depending on which claim applies.

Moderation failures follow the embedding pattern above. A platform accused of hosting harmful content is rarely accused of creating it; it is accused of failing to police what it embeds. The repairs that have actually worked look less like statements of values and more like visible, auditable enforcement: content removed, an account suspended, a policy change dated and specific. A values statement without enforcement reads, correctly, as the same failure with better public relations.

Misinformation is the hardest platform case because the platform is rarely the Source of the false claim; it is the Bridge that gave the claim reach. Contain may require slowing or removing the claim while it is spreading. Repair requires making ranking and distribution decisions transparent enough to identify how the Bridge amplified it. Monitor then tests whether the revised controls continue to reduce recurrence without merely hiding the same failure behind a new rule or label.

Applying the Model When the Fix Is Invisible

Some repairs carry a problem the others in this chapter do not have: the person deciding whether to trust the system again cannot inspect what actually changed inside it.

When a human professional makes a mistake and then claims to have fixed the underlying cause, the claim is at least checkable in principle. A colleague can ask what changed, watch subsequent conduct, or ask someone who knows them. Aircraft software offers no such checks. Two Boeing 737 MAX crashes, Lion Air Flight 610 in Indonesia in October 2018 and Ethiopian Airlines Flight 302 in March 2019, killed 346 people, and regulators around the world grounded the plane. The failures centred on MCAS, software designed to push the aircraft's nose down automatically in certain conditions (House Committee on Transportation and Infrastructure, 2020). After design changes and new pilot training requirements, the US Federal Aviation Administration cleared the aircraft to return to service on 18 November 2020 (Federal Aviation Administration, 2020). No passenger could check that fix. Boarding a MAX meant borrowing trust from the same regulator whose original certification had been called into question.

Lee and See's foundational work on trust in automation offers a partial way through: calibration, not restoration (Lee & See, 2004). Their argument was never that people should trust automation fully or not at all, but that trust should track the system's actual reliability as closely as possible. Applied to repair, a system failure does not need to be followed by a claim of restored perfection. It needs to be followed by a narrower, honestly calibrated claim: what kind of error was found, what kind of error is still possible, and what the user should watch for in the meantime. A narrow claim that turns out to be true does more for repair than a broad claim that turns out to be optimistic.

The competence-versus-integrity distinction from earlier in this chapter applies here too, with a twist. A software fault first presents as a competence failure: the system got something wrong. The MAX case did far more damage once it began to read as an integrity failure. A congressional investigation found that pilots were largely unaware MCAS existed, and that Boeing had withheld crucial information from the FAA, its customers and 737 MAX pilots (House Committee on Transportation and Infrastructure, 2020). Failures of that kind do not repair on a patch. They repair, if they repair at all, only after a sustained period in which the maker's disclosed conduct is checked against its actual conduct, a slow process no software update can shortcut.

In cases like the MAX, grounding the aircraft contains exposure; correcting and independently approving the design repairs the immediate competence failure; and monitoring later operations, disclosures, training, and organisational conduct tests whether the technical and integrity repairs endure. The phases solve different problems. A verified software change is not evidence that the wider institutional correction will persist.

The Contain–Repair–Monitor Model

Responding to a trust failure is not a single act. The Contain–Repair–Monitor Model has three phases: Contain the failure, Repair the cause, harm, and trust relationship, and Monitor whether the correction holds. The phases are sequential but not final. If Monitor reveals recurrence or a new Mistrust Contagion pathway, the process returns to Contain. The Correct step changes according to whether the failure occurred in the Source, Bridge, or Target. An organisation can complete corrective actions; only the Trustor can decide whether the resulting Direct Evidence is sufficient to re-earn trust. Each phase also needs a named owner, a trigger for stopping or restricting use, and a date for the next review. Without them, the steps describe what should happen but not who will make it happen. The decision sheet in the Appendix gives one way to record them.

Phase 1: Contain

C1. Identify — establish what failed and identify the Source, Bridge, Target, and affected Trustors.

C2. Map — trace the active harm and the pathways through which Mistrust Contagion may spread.
C3. Isolate — suspend, withdraw, or restrict the failed Target, Bridge, Endorsement, Certification, or process.

Phase 2: Repair

R1. Acknowledge — name the failure clearly, without minimising language, before being forced to.

R2. Explain — describe what happened and why in terms the affected Trustors can evaluate.
R3. Correct — remedy the harm and fix the Source, Bridge, or Target that failed:

Source failure: sever or narrow the failed Endorsement, clarify responsibility, and establish a credible new basis for any future connection. A replacement Source can help only if the Target itself did not fail.

Bridge failure: repair the connection through the remedy appropriate to its form—recertification, ranking transparency, institutional accountability, intermediary reform, or association clarification—as described earlier in this chapter.

Target failure: correct the competence or integrity failure and remedy the specific harm visibly.

R4. Verify — obtain credible, preferably independent, confirmation that the correction works at the point of implementation.
R5. Notify — tell the people affected, through channels they actually use, and check how far the notice reached. No organisation can reach every Trustor a failure touched; the duty is to reach those affected in proportion to the harm, and to keep track of who has not yet heard.

Phase 3: Monitor

M1. Observe — track whether the correction continues to work over time and under changing conditions, including recurrence, performance, complaints, unintended effects, and the movement of Mistrust Contagion through the network.
M2. Demonstrate — make continuing evidence of changed conduct and structural improvement visible.
M3. Adapt — revise controls, standards, responsibilities, or communications when Monitor reveals weakness.
M4. Accumulate Direct Evidence — sustain conduct that allows the Trustor's Record to be rebuilt over time. If the Trustor judges that record sufficient, trust may be re-earned; the organisation cannot announce that outcome for itself.

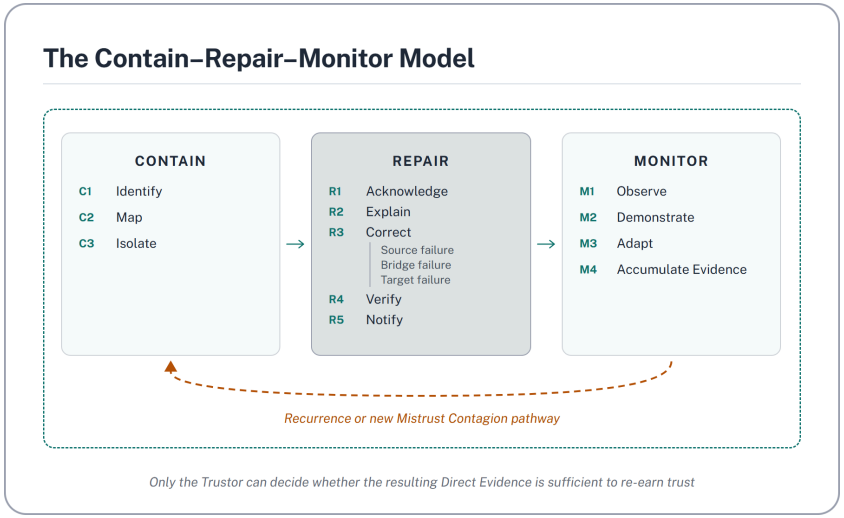


Figure 13.1 The Contain–Repair–Monitor Model. Contain (C1–C3) stops the failure and limits its spread; Repair (R1–R5) addresses the cause, harm, and failed trust relationship and verifies the correction at implementation; Monitor (M1–M4) tests whether the correction endures and accumulates the sustained Direct Evidence through which the Trustor may judge trust to have been re-earned.

Read back through this book's model, which did not exist in 1982, Johnson & Johnson's Tylenol response falls into recognisable phases. The nationwide withdrawal contained further exposure. Public communication and tamper-

resistant packaging repaired both the immediate product risk and the Bridge between the company and consumers; testing and regulatory scrutiny verified the packaging change at implementation. Continued observation, visible packaging practices, and later experience then allowed the public to judge whether the correction endured. The governing logic is simple: Contain the failure, Repair the cause and the trust relationship, and Monitor whether the correction holds. If Monitor reveals recurrence or a new Mistrust Contagion pathway, the process returns to Contain.

Chapter Takeaway

Responding to a trust failure is not the reverse of failure. The Contain–Repair–Monitor Model follows a three-phase logic. Contain stops the failure and limits its spread. Repair addresses the cause, harm, and transfer structure through which confidence originally moved, matching the response to both the failure type—competence or integrity—and the Bridge involved. Monitor tests whether the correction continues to work over time and under changing conditions.

Verify can confirm that a correction works at implementation; Monitor establishes whether it endures. An organisation can complete corrective actions, but it cannot declare trust restored. It can only accumulate Direct Evidence. The Trustor decides whether that evidence is sufficient to rebuild the Trustor's Record and re-earn trust, one observable piece of evidence at a time.

Part IV

Building Better Trust: Design and the Machine Frontier

How to build trust deliberately: designing honest signals, the institutions that verify them at scale, a field guide for reading them as a Trustor, and the frontier where the Target is a machine.

Chapter 14

Designing Trust Signals

By this point, the pattern should be clear. Trust does not only live inside people or institutions. It is built into environments. It is routed by labels, rankings, interfaces, defaults, affiliations, recommendations, and workflows. The modern trust problem is therefore less about whether trust exists than about whether the signals that move it are designed well.

Trust cues are never neutral, which is what makes this question unavoidable. A badge does not merely decorate a page. A ranking does not merely organise options. A partnership line does not merely add context. Each signal tells the Trustor something about what has already been checked, who is standing behind the Target, and how much Vulnerability is reasonable. Even when no explicit promise is made, the signal is doing interpretive work.

This chapter turns from diagnosis to design. If Trust Transfer is unavoidable, how should systems be built so that the movement of trust is proportionate to the strength of the Bridge? How should organisations signal reassurance without inviting more confidence than they have earned?

The answer begins with a simple principle: the people a trust signal is meant for should read it as meaning what it actually means, and no more. Different people will read the same cue differently, so this is a goal to test with real users, not a guarantee design can give.

Why Design Matters

In many organisations, trust signals are treated as communication details. They sit with interface design, brand, compliance, or conversion optimisation. That is too small a frame. Trust signals are part of the governance of Vulnerability.

When a product is marked verified, recommended, top rated, featured, or integrated with a trusted institution, the user is receiving more than information. The user is receiving permission to rely. The signal is lowering the felt need for independent investigation.

That is why weak signals can do disproportionate damage. If they invite strong inferences without strong substance behind them, they create overtrust on the way in and backlash on the way out. The cost is not limited to one interaction. It spreads to the Source, to similar signals, and eventually to the broader environment.

Design, then, is not cosmetic. It determines how much Borrowed Confidence a system is asking people to spend.

Worked Example: Evaluating a Vendor Badge

A small business owner is choosing a commercial cleaning company from her coworking space's vendor directory. The listing leads with a badge reading "Preferred Vendor" and a line noting the company has "cleaned 50+ offices in the building."

Running a few of the model's diagnostic questions, which Chapter 16 sets out as a checklist, changes what she asks next. What is the Source of the trust she is feeling? Not the cleaning company itself, but the coworking space's directory and the badge it displays. What exactly is the Bridge? She asks the building manager how vendors earn "Preferred" status and learns it means the company paid a listing fee and passed a basic insurance check, nothing about the quality of its work. Is that Bridge real or merely implied? The badge implied the building had vetted the company's work; the actual relationship was a paid placement with an insurance checkbox. What has she been assuming was checked? That other tenants were satisfied with the cleaning; no feedback had actually been collected. What would happen if the Bridge is weaker than it looks? A poor cleaning job, with her own lease, not the coworking space, on the line if anything gets damaged.

She books a one-month trial with a cancellation clause instead of the annual contract the directory promoted. The badge was real. The Bridge it implied was not.

The First Rule: Semantic Precision

The first rule of trust-signal design is semantic precision. Signals should communicate exactly what they certify, rank, or imply. If a badge means identity verification, it should not look like a quality seal. If a recommendation is paid placement, it should not borrow the visual language of editorial curation. If a model is embedded in a respected workflow, the interface should distinguish clearly between integration and endorsement.

This sounds obvious, but much of the modern trust problem comes from semantic slippage. Users see one visual cue and infer a whole package of meanings: competence, safety, fairness, oversight, and approval. Often the system is only signalling one of those things.

The gap between signal meaning and user interpretation is where Borrowed Legitimacy grows. Good design narrows that gap. It does not depend on users reading fine print after the impression has already formed.

In trust-transfer terms, semantic precision makes the Bridge legible. It tells the Trustor what kind of Bridge this is.

The Second Rule: Match Signal Strength to Bridge-Strength

A strong-looking cue should rest on a strong underlying Bridge. That is the second rule. If a signal visually resembles formal endorsement, then formal accountability should stand behind it. If a ranking looks authoritative, its basis should be inspectable. If a system claims to be trustworthy, the processes that justify that claim should exist independently of the marketing.

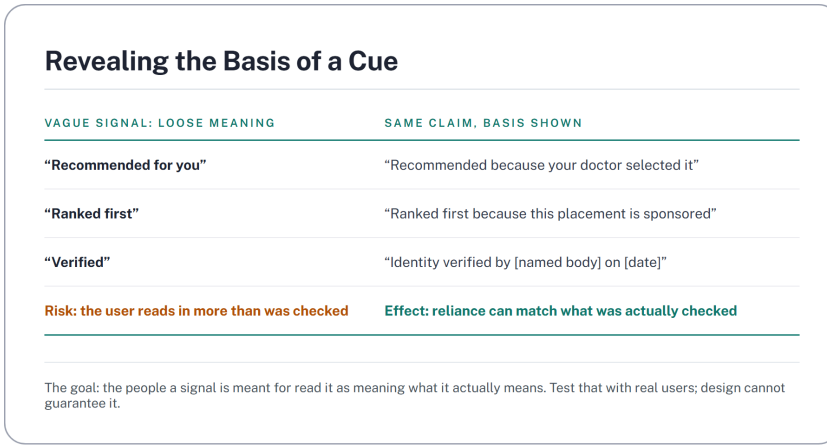
The reason is that people react first to the presence of a Bridge, not to its fine-grained architecture. The interface may trigger authority, familiarity, or social-proof heuristics long before the user stops to ask what is really being guaranteed.

McKnight, Choudhury, and Kacmar (2002b) showed how website quality and reputation build initial trust before any direct experience. Metzger, Flanagin, and Medders (2010) showed that users routinely rely on heuristics such as reputation, endorsement, and expectancy rather than deep verification. Those findings make the design burden heavier, not lighter. If users predictably read cues quickly, systems have a duty not to overstate what the cue warrants.

A fragile Bridge should not be dressed as a load-bearing one.

The Third Rule: Reveal the Basis of the Cue

A trust signal is more ethical and more durable when it shows its basis. Recommended because your doctor selected it is different from recommended for you. Ranked first because this is sponsored is different from ranked first. “Verified” is different from “identity verified by a named body on a stated date”: the second makes the same claim, but says what was checked and by whom.



▼ Text version of this figure

Vague signal: loose meaning (says)	Precise signal: shown basis (shows)
“Recommended for you”	“Your doctor selected this”
“Ranked first”	“Ranked first, sponsored”
“Verified”	“Independently safety-audited”
Risk: user overtrusts	Effect: trust matched to the real basis

Figure 14.1 Revealing the Basis of a Cue. “Recommended because your doctor selected it” tells the Trustor something “recommended for you” never does.

The important move is not verbosity. It is causal clarity. Users do not need a dissertation. They need to know why the signal exists and what process produced it.

This is especially important in algorithmic systems. A result, seller, tool, or answer may appear with an aura of institutional confidence even when the ranking or recommendation was generated by optimisation criteria the user would not consider trust-relevant. If the cue reflects popularity, payment, personalisation, or model prediction rather than human review or formal certification, that distinction should be visible.

Trust Transfer works through a perceived Bridge, read by the Eye of the Trustor. Revealing the basis of the cue makes that Bridge inspectable rather than mythical.

The Fourth Rule: Distinguish Availability From Endorsement

One of the most common trust failures occurs when users read availability as endorsement. A product appears on a respected platform. A tool appears inside

a hospital portal. A model is available in a work system. A result appears at the top of a search interface. In each case, people often infer that someone upstream has judged the Target to be good, safe, or appropriate.

Sometimes that inference is justified. Often it is only partially justified. The Target may have been made available because it met a minimum threshold, because it was integrated technically, because it matched predicted relevance, or because it paid for position.

Design should not blur these categories. A system that merely hosts something should not visually mimic the language of recommendation. A workflow that includes a tool should make clear whether the institution has deeply evaluated that tool, lightly reviewed it, or simply made it accessible.

This distinction matters here: institutional embedding is one of the strongest trust Bridges in the book. The more credible the Source, the more carefully it must separate access from endorsement.

The Fifth Rule: Design for Calibrated Reliance

The goal of trust design is not maximum trust. It is appropriate trust. Lee and See's (2004) work on trust in automation remains one of the clearest guides here. The main failure modes are overtrust and undertrust. A good system helps users rely when reliance is justified and remain sceptical when it is not.

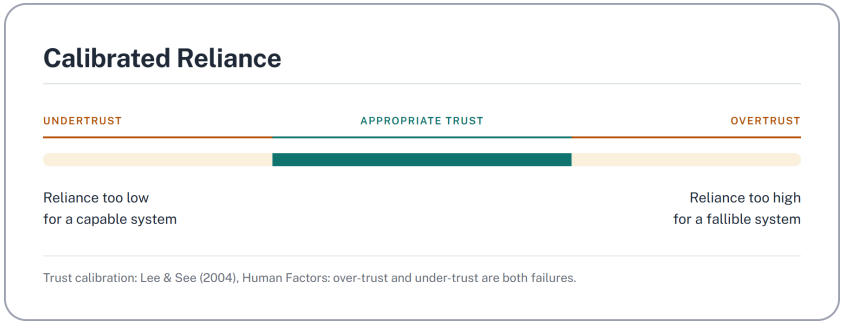


Figure 14.2 Calibrated Reliance. The goal is not maximum trust but appropriate trust — a design that invites reliance when it is justified and scepticism when it is not.

That principle is easy to say and hard to implement, because many commercial systems are rewarded for immediate uptake rather than calibrated use. But if the cue structure pushes people toward more confidence than the system deserves, the design is not trust-building. It is trust-extracting.

Parasuraman and Riley's 1997 framework on misuse, disuse, and abuse sharpens the point. Systems fail not only when people refuse useful automation,

but also when they rely on it too quickly or too broadly. Design has to shape the level, timing, and scope of reliance.

This is where warning design, uncertainty design, escalation paths, and human override become necessary. They are not signs of weakness. They are mechanisms for keeping trust proportional to capability.

The Disclosure Trap

Disclosure makes trust-signal design harder because more information can itself become a misleading cue. In most product settings, more context improves judgement. With conflicts of interest, that is not always true.

Cain, Loewenstein, and Moore showed this with a jar of coins (Cain, Loewenstein, & Moore, 2005). Advisers estimated the jar's value for people who had to guess it, and the advisers were paid more when those guesses ran high. When advisers had to disclose that conflict, they did not become more honest. They exaggerated further, and the people they advised did not discount the advice nearly enough. Disclosure left the advised worse off than silence did.

That result surprises people because disclosure is often treated as a universal solution. But a statement that says "I have a conflict" without saying how large it is, or how to correct for it, may deepen overtrust rather than correct it. A disclaimer, a fine-print box, or a "sponsored" tag can become just another Source cue.

The design lesson is subtle but important. Disclosure should not merely make the Source feel more honest. It should make the limits of the advice more legible too.

The Governance Layer

The strongest trust signals are often not interface elements at all. They are governance structures made visible. Independent audit, clear accountability, opt-out rights, documented review standards, human override, and public enforcement all act as high-quality Bridges because they place real consequences behind reassurance.

Restaurant hygiene grades show how this works. In 1998 Los Angeles County introduced letter grades posted in restaurant windows. Jin and Leslie (2003) found that inspection scores rose, diners became more responsive to hygiene, and hospitalisations for foodborne illness fell. A later re-analysis with more data questioned the size of the health effect (Ho, Ashwood, & Handan-Nader, 2019). The narrower point survives: a single letter in the window turned a

private inspection into a public Bridge that diners could read and restaurants could not ignore.

The same logic reaches well beyond restaurants. When trust stakes rise, cosmetic signals lose power and accountability signals gain it. A badge is stronger when it points to a process. A recommendation is stronger when it exposes its standard. A platform is stronger when it makes enforcement visible.

Good design therefore has two layers: the surface cue and the institutional structure beneath it. The cue should never claim more than the structure can support.

Testing Trust Signals

Most organisations test signals for click-through, conversion, and engagement. Far fewer test them for interpretive accuracy. That is a mistake. A trust cue can perform well commercially while still being misleading.

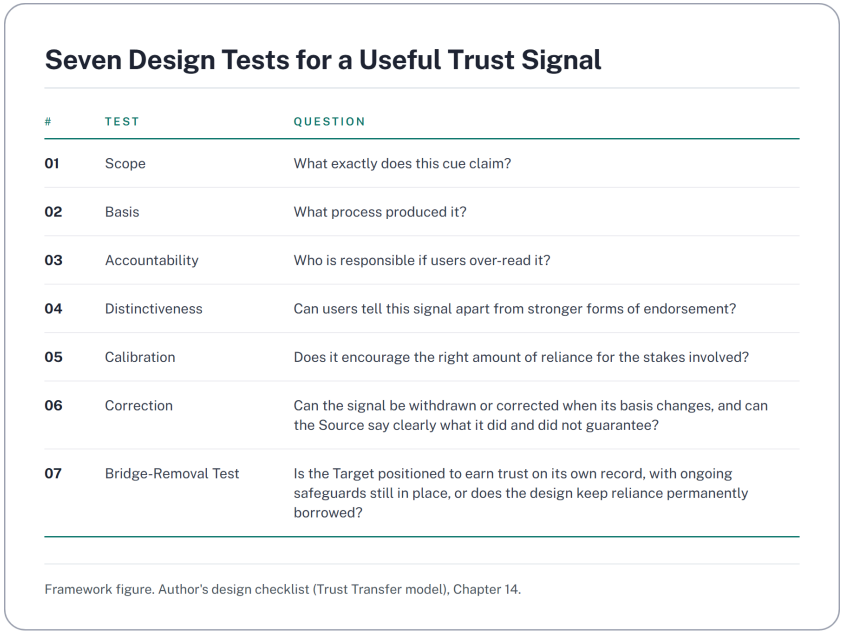
The better question is not only whether the signal increased action. It is what users think the signal meant. Did they infer quality when it indicated identity? Did they infer endorsement when it indicated payment? Did they infer oversight when the Source had only provided distribution?

Trust-signal testing should therefore include comprehension, calibration, and failure simulation. Ask users what they believe has been checked and how much responsibility they think the Source has accepted. Then test what would have to stop if the Target failed, what would need to be corrected and verified, and what later evidence would show that the correction endured. If the backlash exceeds what the institution believes it promised, the signal was too strong for the Bridge underneath it.

This kind of testing is especially important for high-trust institutions. Schools, hospitals, governments, financial firms, and major platforms all have large reservoirs of borrowed credibility to lose.

Practical Design Checklist

A useful trust signal should pass seven tests, set out in Figure 14.3: scope, basis, accountability, distinctiveness, calibration, correction (the signal can be withdrawn or corrected when its basis changes), and the Bridge-Removal Test.



▼ Text version of this figure

#	Test	Question
01	Scope	What exactly does this cue claim?
02	Basis	What process produced it?
03	Accountability	Who is responsible if users over-read it?
04	Distinctiveness	Can users tell this signal apart from stronger forms of endorsement?
05	Calibration	Does it encourage the right amount of reliance for the stakes involved?
06	Correction	Can the signal be withdrawn or corrected when its basis changes, and can the Source say clearly what it did and did not guarantee?
07	Bridge-Removal Test	Is the Target positioned to earn trust on its own record, with ongoing safeguards still in place, or does the design keep reliance permanently borrowed?

Figure 14.3 Seven design tests for a useful trust signal. A useful trust signal identifies its scope and basis, assigns accountability, remains visually distinct, supports calibrated reliance, can be withdrawn or corrected, and allows the Target eventually to earn trust on its own record.

If a signal fails these tests, it may still boost action in the short run. But it is probably drawing down trust capital faster than the organisation realises.

From Trust Design to Trust Infrastructure

The deeper shift this chapter argues for is conceptual. Trust design should not be treated as a set of persuasive tricks placed at the edge of a product. It should

be treated as infrastructure. It determines whether people can make good Vulnerability decisions under uncertainty.

A healthy trust environment does not maximise comfort. It improves interpretation. It helps Trustors distinguish between institutional endorsement, reputational endorsement, and mere proximity. It reduces the gap between what a cue suggests and what the underlying Bridge can actually bear.

That is the design standard worth aiming for. Not more trust. Better-directed trust.

Chapter Takeaway

Trust signals are not decorations. They are Bridges. They tell people what has been checked, who is standing behind a Target, and how much Vulnerability is reasonable.

When those signals are vague, inflated, or semantically slippery, they create overtrust, Borrowed Legitimacy, and eventual backlash. When they are precise, accountable, and matched to real underlying structures, they make Trust Transfer safer and more deserved.

The design goal is therefore not to make systems feel trustworthy. It is to make the meaning of trust cues accurate enough that people can rely in the right amount, for the right reasons, at the right time. The next chapter asks a related question: once a signal is well designed, what makes it trustworthy at the scale of a whole market, not just one product.

Chapter 15

Verification at Scale

In 1893, at the World's Columbian Exposition in Chicago, a young electrical engineer named William Henry Merrill was sent to inspect the wiring behind the fair's dazzling new displays of alternating current. Fire insurers were nervous. The exposition ran without a major electrical fire, and Merrill saw an opening: a laboratory that tested devices before they reached the public, not after they burned a building down. He opened it above a fire patrol station in 1894 with two employees and $350 of equipment (UL Research Institutes, n.d.). That laboratory became Underwriters Laboratories, and the small circle it stamps on a toaster or a phone charger has become a widely recognised trust mark.

Nobody buying that toaster tests it themselves. Nobody reads the underlying standard, which specifies exact wire gauges and clearances most consumers could not evaluate. They see a circle with two letters in it and reach for their wallet. That circle is a Bridge doing an enormous amount of work, connecting a product nobody has personally tested to a household willing to plug it into the wall.

This chapter is about how that Bridge gets built at scale. Chapter 14 asked how one organisation designs one honest signal. This one asks a harder question: what makes a signal trustworthy across an entire market of thousands of products, from a manufacturer the buyer has never heard of, verified by a body the buyer has never met.

Why One Bridge Doesn't Scale

A single referral, of the kind Chapter 3 described, works between two people who can check back with each other. It does not work for a market. No shopper can personally test the wiring in every phone charger sold this year, and no employer can personally audit every bank holding its payroll account. At some point the number of Targets outruns any one Trustor's capacity to gather Direct Evidence, and something has to stand in for that evidence at scale.

That something is usually a third party built for exactly this job: a testing lab, an auditor, a licensing board, a ratings agency. It checks the Target on behalf of everyone who will never inspect it themselves, sometimes once and sometimes on a continuing schedule, and issues a mark that travels with the product or the institution. The mark is a Bridge built to be reused millions of times without wearing out.

That reuse is also the risk. A referral that turns out wrong damages one relationship. A verification mark that turns out wrong damages every Target that has ever carried it, all at once.

What a Verification Mark Actually Certifies

A mark like this certifies less than people assume. The UL mark on an appliance says the product was assessed against specific safety requirements. Under UL's ordinary ongoing certification programme, that first assessment is followed by factory inspections and audits meant to confirm the product keeps meeting those requirements (UL Solutions, n.d.). Even so, the mark covers those requirements and nothing more. It says nothing about whether the appliance is well made, reasonably priced, or likely to last. An annual bank audit works the same way: it says the numbers in the financial statements fairly represent the bank's position as of a stated date, under standards written well before anyone anticipated the bank's specific risks. It is not a forecast, and it is not a guarantee that the bank will still be solvent next month.

The gap between what a mark certifies and what people think it certifies is where Borrowed Confidence usually overshoots. The Trustor reads the seal as "safe, full stop." The Source only ever meant "meets these specific requirements, within the limits of what this programme checks."

When the Verifier Fails

On 24 February 2023, KPMG signed off a clean audit of SVB Financial Group, the parent of Silicon Valley Bank. Fourteen days later, regulators seized the bank after a run on deposits. Two days after that, Signature Bank failed too; KPMG's clean audit of it, signed 1 March, was eleven days old (Accounting Today, 2023; PYMNTS, 2023). According to Bloomberg Tax's report on a later Senate subcommittee inquiry, KPMG had considered dozens of risk factors that might signal a bank in trouble, and concluded that none applied to Silicon Valley Bank in early 2023 (Bloomberg Tax, 2025).

The short gap between a clean opinion and a collapse raises questions, but it does not answer them. A clean audit opinion is not a guarantee of future solvency. Under the standards auditors work to, it answers a narrower question: do the financial statements fairly represent the bank's position as of a stated date. Nor does a collapse, on its own, show that an audit was done badly. The opinion, the failure, and any documented findings about the audit have to be looked at separately. But few people reading a headline about a "clean audit" hold that narrow question in mind. They hear: this bank is fine.

The damage did not stop at the two banks. Once one Big Four firm's clean opinion turned out to precede a collapse within two weeks, it invited a second look at clean opinions the same firm had issued elsewhere. The doubt did not stay inside one Bridge. It could reach any Target the same verifier had vouched for.

The same pattern shows up outside banking. Asset managers increasingly attach an ESG label to a fund, signalling that environmental, social, and governance factors shaped the underlying investment decisions. An investor who chooses that fund over an otherwise identical one is relying on a Bridge: the manager's own claim that a stated process was followed.

In 2021, Desiree Fixler, the outgoing head of sustainability at DWS, the asset-management arm of Deutsche Bank, said the firm's public claims overstated how much ESG integration actually shaped its investment process. Two years later, the U.S. Securities and Exchange Commission charged DWS over those claims. In September 2023, DWS paid a $19 million penalty for making misleading statements about its ESG controls, on top of a separate penalty for unrelated anti-money-laundering failures (U.S. Securities and Exchange Commission, 2023). The firm did not admit wrongdoing, but the label itself, "ESG integrated," had already done its work on every investor who read it as a settled fact rather than a claim still to be checked.

The failure mode is related to the audit case, though not the same. No independent verifier stood behind this label: the Bridge was the manager's own claim about its own process. A label meant to travel across thousands of funds substituted for the Direct Evidence no individual investor could gather. When the substitution turned out to be false, the damage was not confined to one fund. It reached every product that had carried the same claim from the same manager.

What Verification Marks Actually Certify

What a mark claims, what people read into it, and what happens when the verifier, or the reading, is wrong.

MARK	WHAT IT CLAIMS → WHAT PEOPLE ASSUME	WHAT HAPPENED
Baseline: UL mark	Met specified safety requirements, with initial assessment and ongoing factory inspections → "Safe, full stop"	Covers those requirements only, not quality, price or durability
Clean audit opinion (KPMG)	Financial statements fairly presented as of a stated date → "The bank is fine"	SVB and Signature Bank failed days after clean opinions; the collapse alone does not show the audits were faulty
ESG label (DWS)	A stated ESG process was followed (the manager's own claim) → "ESG shaped the holdings"	The SEC charged DWS over misleading ESG statements; DWS paid a $19M penalty in 2023 without admitting wrongdoing

▼ Text version of this table

Mark	What it claims → what people assume	What happened
Baseline: UL mark	Met specified safety requirements, with initial assessment and ongoing factory inspections → "Safe, full stop"	Covers those requirements only, not quality, price or durability
Clean audit opinion (KPMG)	Financial statements fairly presented as of a stated date → "The bank is fine"	SVB and Signature Bank failed days after clean opinions; the collapse alone does not show the audits were faulty
ESG label (DWS)	A stated ESG process was followed (the manager's own claim) → "ESG shaped the holdings"	The SEC charged DWS over misleading ESG statements; DWS paid a $19M penalty in 2023 without admitting wrongdoing

Table 15.1 What Verification Marks Actually Certify. What a mark claims, what people read into it, and what happens when the verifier turns out to be wrong.

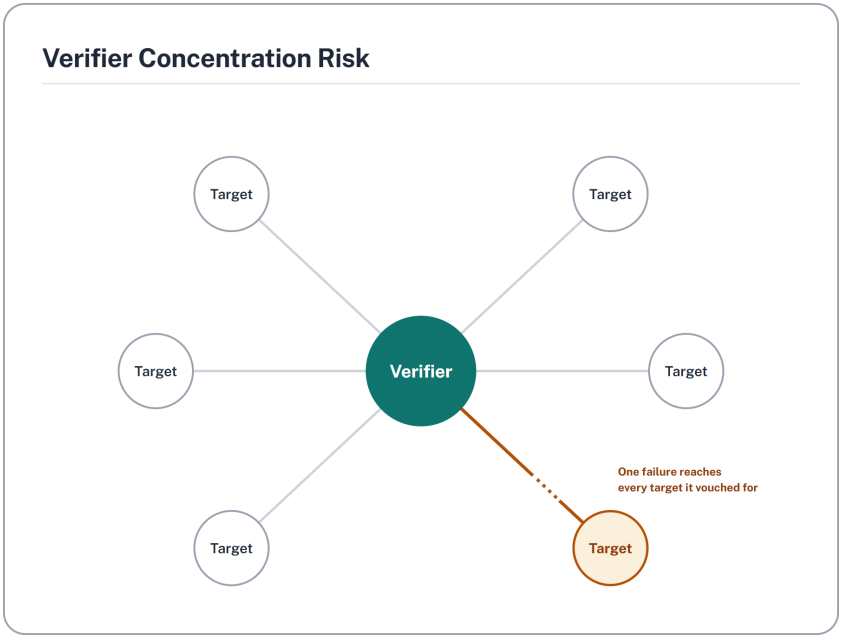


Figure 15.1 Verifier Concentration Risk. A single verifier stands behind many Targets at once, so one failure can reach every Target it ever vouched for.

Design Implications

For institutions that build or rely on verification systems, the same pattern points to a few practical rules.

State the scope out loud. A mark that says "tested against Standard X, as of date Y" invites less overreading than one that simply implies "safe" — the narrower the claim looks, the less it can be blamed for promising something it never said.

Expect concentration risk. Any single verifier that certifies enough Targets becomes, in effect, a second Source standing behind an entire category. A prudent institution asks not just "is our own Bridge sound" but "what happens to everyone we've vouched for if we turn out to be wrong."

And keep the verifier's own incentives visible. An auditor paid by the bank it audits, a lab paid by the manufacturer it tests, faces the same conflict Chapter 14 flagged for disclosure generally. That doesn't make every verification worthless — it means a mark is only as strong as the arrangement that produced it, and that arrangement is worth knowing before the mark is trusted.

Chapter Takeaway

A market cannot run Direct Evidence on every Target, so it builds verifiers to stand in for it: testing labs, auditors, licensing boards, ratings agencies. These marks let a stranger's product or balance sheet earn provisional trust before anyone personally checks it.

But a verification mark certifies less than people read into it, and it reuses one Bridge across thousands of Targets at once. When the verifier is right, that scale is the whole point. When a verifier is wrong, or is read as having promised more than it did, the doubt does not stay inside one relationship. It can reach everything the verifier has vouched for, as the questions about the clean audits signed days before Silicon Valley Bank and Signature Bank collapsed showed.

The next chapter turns from institutions back to the individual reader, with a practical question: given everything this book has argued, what should you actually do the next time you are asked to rely on a Bridge you cannot personally inspect?

Chapter 16

A Field Guide for the Trustor

A Checklist, Not a Lecture

Here is an illustration, with the details changed. A friend once picked a hand surgeon in eleven minutes. Her hand hurt, a search turned up three names, one had more reviews than the other two combined, and she booked the appointment. She did not read a single study on outcomes. She did not look the surgeon up on the medical register. She saw a number and a photo of a waiting room that looked clean, and she decided that was enough.

For most decisions, that would probably have been enough. A hand surgeon is not most decisions, and the checklist later in this chapter will say so directly. But the instinct behind her eleven minutes is not the problem this book is after. Borrowed Confidence is not a shortcut people should feel embarrassed about. It is how anyone gets through a day with thousands of small decisions and no time to investigate any of them properly.

Chapter 14 asked how an organisation designs an honest signal. Chapter 15 asked how a whole market keeps its verifiers honest. This chapter asks a smaller, more useful question: on an ordinary Tuesday, faced with a badge, a recommendation, a ranking, or a review score, what should you personally do with it? Not a theory of trust. A checklist you can actually run in your head in under a minute.

Name the Source, Name the Bridge

The first move is the simplest and the one people skip most often: say out loud who is actually vouching for what.

A review score is not the surgeon vouching for herself. It is hundreds of past patients, filtered through a platform's ranking, vouching for an experience that may or may not resemble yours. A badge on a product page is not the retailer testing the product. It may be a manufacturer's own claim, or a paid placement, dressed up to look like a verdict, and unless the page says which, you cannot tell. Once you can say the sentence "X is vouching for Y, through Z," you have already done most of the analytical work this book has been building toward since Chapter 2.

This is not a demand for suspicion. It is a habit of specificity. Many Bridges wobble under this question, not because they are dishonest, but because nobody, including the people who built them, had bothered to say precisely what they were claiming.

Ask What the Bridge Actually Certifies

Once you have named the Source and the Bridge, ask the narrower question Chapter 15 raised: what, exactly, does this Bridge claim, and as of when?

A five-star average says people who bothered to leave a review were satisfied enough to do so. It does not say the product is well made, only that it did not provoke enough complaint to drag the average down. A professional licence establishes what the licensing scheme requires. Depending on the profession and jurisdiction, that may include renewal, continuing education and current conditions, so check what the register actually confirms. Registration alone does not show that someone suits your particular case, or how any complaints against them were resolved.

Most signals are honest about a narrow claim and let the reader's imagination supply the broader one. Your job is to hand the imagination back its narrower assignment before you rely on it.

Watch for the Common Tells

A few patterns recur often enough to be worth naming, because Chapter 11 described exactly how they get made.

A badge with no visible standard behind it is usually decoration, not certification. A recommendation with no disclosed relationship between the recommender and the recommended is worth a second look, particularly when the recommender benefits from the placement. A high ranking on a platform tells you the platform's algorithm favoured this result; it does not tell you why, and "why" is exactly what a ranking hides. And a familiar name attached to an unfamiliar product tells you the name-holder was paid, was fooled, or genuinely checked. Those three explanations look identical from the outside.

None of these tells prove a Bridge is weak. They just tell you where to spend your remaining two minutes of scrutiny, instead of spreading it evenly across a decision that does not deserve equal attention everywhere.

When to Slow Down

Most of this checklist is meant to run in seconds, because most decisions deserve seconds. A wrong choice of coffee shop costs you a bad coffee. A small number of decisions deserve minutes or hours instead, and Chapter 9 already gave the test: ask whether the decision is reversible, and ask what it costs if the Bridge turns out to be wrong.

Choosing a surgeon, a mortgage broker, a childcare provider, or a financial adviser sits on the expensive, hard-to-reverse side of that line. So does anything where the Target will hold your money, your health, or your legal standing after the transaction ends, not just during it. For those decisions, run the full checklist, not the eleven-minute version. Name the Source and Bridge precisely. Find out what the certification actually covers and when it was last renewed. Look for evidence about the Target itself, rather than relying on Borrowed Confidence alone: its Public Record, such as a regulator's register or a published finding, or your own experience of it. A person you know who has used this Target is worth asking, but their account is testimony, another Bridge, not your own Direct Evidence.

The skill is not scrutinising everything. It is knowing which handful of decisions in a given month deserve it.

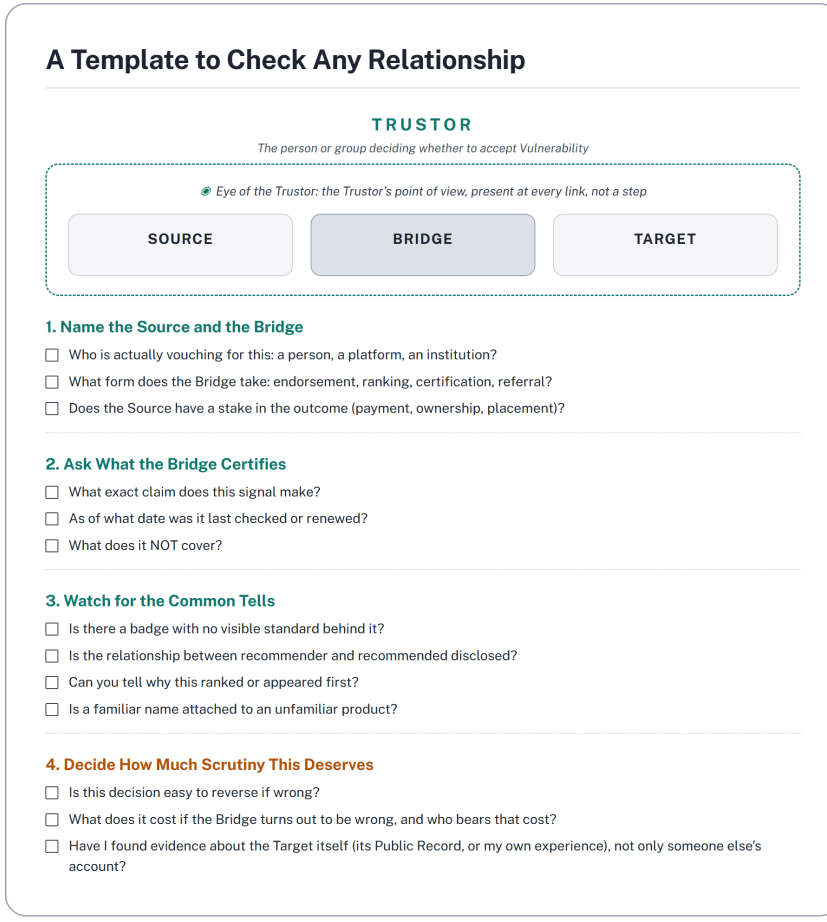


Figure 16.1 A fillable checklist for naming a Bridge, checking what it certifies, watching for common tells, and deciding how much scrutiny a relationship deserves.

A Worked Example

Back to the hand surgeon. Running the checklist on that eleven-minute decision looks like this.

Name the Source and Bridge: the recommendation came from a review platform's star average, not from a doctor, a friend, or a referring physician. Ask what it certifies: a higher star count than two competitors, built from patients who chose to leave a review, which says more about who bothers to write reviews than about surgical outcomes. Watch for tells: no visible credentialing badge, no disclosed relationship between the platform and the practice, a ranking whose logic is invisible. On the reversibility test: hand surgery is not

trivially reversible, and a bad outcome is expensive in every sense, so this decision sits on the side of the line that deserved more than eleven minutes.

That does not mean the choice was wrong. The surgeon may well have been excellent. It means the process, run honestly against the checklist, would have flagged this as a moment to slow down: check the surgeon's registration and any conditions on it (in Australia, on the Ahpra public register; elsewhere, with the relevant medical board), read one or two of the actual written reviews rather than just the average, and ask her GP. None of that proves a surgeon is right for her; it only makes the choice better informed. Booking a first consultation is one decision; agreeing to surgery is another, and the second is where the surgeon's own advice, her GP and, if she wants one, a second opinion come in. Borrowed Confidence got her to a name. It should not have been the only thing that got her onto the operating table.

Four Audit Tools, Stress-Tested

The checklist above catches a Bridge in the moment. These four tools go further. Each one forces a decision on something specific, and each tells you upfront where it will mislead you if you lean on it past its limit.

1. The Source–Bridge–Target Audit

Before you act on someone else's confidence, say the chain out loud — the same three boxes as Figure 16.1. Name the Source — whoever this confidence is actually coming from. Name the Bridge — the specific recommendation, badge, ranking, or relationship doing the connecting. Then ask what the Target itself has actually shown, separate from anything the Source or Bridge implied.

Where this slips: roles are not fixed. The same organisation can be a Source in one sentence and a Target in the next, and a single case can hide one Bridge inside another. If the first pass feels too clean, that is the signal to run it again.

2. The Bridge-Strength Scorecard

Orientation is whether a Bridge points you toward real information about the Target, or just back toward the Source's own reputation. Strength is how much weight it can carry even when it points the right way — a certification tied to a defined standard carries more than a vague affiliation that could mean almost anything. Score each from 1 to 5 and take the lower of the two: a Bridge can win on one and lose on the other, and the weak side is the one that should set the score. Orientation scores 1 if the Bridge points only back to the Source's own name, 3 if it points to some information about the Target but its scope or date is unclear, and 5 if it points to specific, current, checkable information that bears on this decision. Strength scores 1 for Co-presence with no stated basis, 3 for a

stated basis that is only partly documented or has no clear owner, and 5 for a defined standard with a named party accountable for it, an independent check, and all of it current.

The catch: both scores depend on information you might not have. If you cannot find out what a badge actually certifies, write "?" instead of a number. Unknown is the honest answer, and it is not the same as weak. A scorecard with a "?" in it is not finished: it tells you what to find out before you lean on the Bridge. The score is deliberately diagnostic, not a validated measurement: it organises a judgement call, it does not replace one.

3. The Vulnerability Check

Before accepting a Bridge, size up the Vulnerability you are about to take on. Score how much is actually on the table (Size) and whether a bad outcome can be undone (Reversibility), each 1 to 5, and multiply them, so a higher number always means more risk. Size scores 1 if a loss would be small and easily absorbed, 3 if it would cost weeks of money, time or trouble, and 5 if it touches health, safety, savings or someone else's wellbeing. Reversibility scores 1 if a bad outcome can be fully undone, 3 if it can be partly undone at real cost, and 5 if it cannot be undone: a large stake that can be undone is not the same bet as a large stake that cannot. What kind of cost it would be — money, health, time, reputation, someone else's safety — stays a label you note beside the number, not a third figure folded into it.

One wrinkle: the person accepting the Vulnerability is not always the one who carries it. A purchasing officer can accept little Vulnerability for themselves while signing off on real Vulnerability for a patient who never got a vote. Work out whose Vulnerability you are actually measuring. If reversibility can't be determined, score it 5, irreversible, until proven otherwise, and note that you did — the cautious assumption, not a guess in the middle. Like the scorecard above, this index is deliberately diagnostic, not a validated measurement: the number organises a judgement call, it does not replace one.

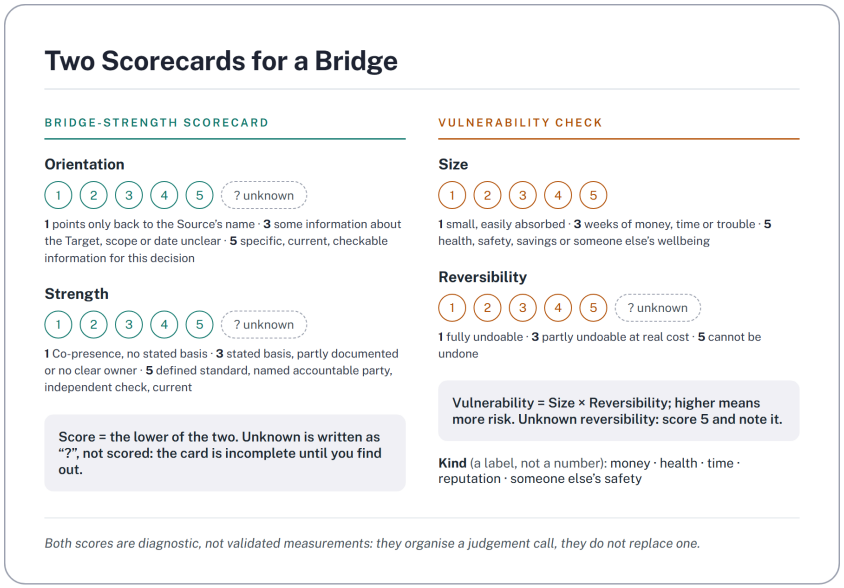


Figure 16.2 Two Scorecards for a Bridge. The Bridge-Strength Scorecard and the Vulnerability Check, each scored from the formulas introduced above.

4. The Bridge-Removal Test, as a Standing Habit

Periodically ask of any ongoing reliance the same question Figure 10.3 put to a single case: if you had never received the original introduction, would what this Target has shown since still justify relying on it, with today's safeguards left in place? A pass suggests the Trustor's Record has taken over; it is a diagnostic sign, not proof. A fail means the original introduction is still doing the work.

The limit: this test only grades the past. A pass today describes the record built so far — it says nothing about ground the Target has not yet been tested on. A long run of good conduct in easy conditions does not predict what happens when conditions get hard; repetition was never renewal.

Chapter Takeaway

Borrowed Confidence is not a flaw to be trained out of yourself. It is how anyone functions in a world with far more decisions than time to investigate them. The skill worth building is not suspicion. It is a habit of naming what a Bridge actually claims before you lean on it: who is vouching, for what, as of when.

Most of the time, that habit runs in seconds and changes nothing about what you decide. It just makes the decision an informed use of Borrowed Confidence rather than an accidental one. The rest of the time, on the handful of decisions each month where the cost of a wrong Bridge is high and the outcome is hard to undo, that same habit is what tells you to slow down and look for Direct Evidence before you act.

The next chapter turns to the hardest version of this problem: a Target that cannot explain itself, built by a company you may or may not trust, doing work you increasingly cannot check by hand. Everything this chapter just asked you to do gets harder when the Target is a machine.

Chapter 17

The Machine Problem: Can Trust Transfer to AI?

Every chapter in this book has followed one pattern across one setting after another. Trust moves from a known Source, across a perceived Bridge, to a less-known Target, and a Trustor accepts Vulnerability on the strength of that link. The pattern held in endorsements and in institutions. It held in social networks, in influencer markets, in platforms, and in search. It held when trust was earned and when it was merely borrowed. It held when the whole thing reversed into mistrust and spread through the same Bridges that once carried confidence, moving forward, backward, or outward depending on which link failed.

This chapter asks whether the same pattern survives contact with the hardest case the modern world has produced: autonomous, generative Artificial Intelligence. This Target does not just execute fixed rules, as the autopilot or the lift controller in Chapter 9 do. It generates its own outputs and increasingly acts as an agent in its own right.

Artificial Intelligence is more than another domain to add to the list. It is a stress test. If the Source → Bridge → Target model is useful beyond the human settings it grew from, it should still help describe what happens even when the Source cannot be inspected, the Bridge is invented by an algorithm, and the thing doing the recommending is not a person at all. This chapter examines whether the framework can also be applied to machine-mediated trust. It also marks where AI stretches the model, and where this book's evidence runs out.

The Gap Between Using and Trusting

The defining fact about AI trust in the mid-2020s is a gap. People use these systems far more than they trust them.

A global study by the University of Melbourne and KPMG (Gillespie et al., 2025; referred to below as KPMG and University of Melbourne), summarised in Table 17.1, records a twenty-point gap between AI use and AI trust (66% versus 46% globally) that has widened rather than closed since 2022, and falls further to 39% in advanced economies.

Australia sits at the low end of that range measured in Table 17.1. Only 36% of Australians say they are willing to trust AI systems, against the 46% global average. Only 17% report high acceptance of AI, against a global average of 33%. Yet organisational adoption in Australia nearly tripled in two years, from 23% in

2022 to 65% in 2024. The machines are being deployed faster than the trust to support them is being built.

One interpretation of that gap, in the language of this book, is a transfer problem: people may use a system before they have found assurances they regard as sufficient, such as certifications, accountability or audit trails. But use and trust are different measures, and their aggregate difference does not show why any individual uses AI.

It is worth pausing on what makes AI such an extreme case of this. With most Targets in this book, a determined person could, in principle, do the checking themselves. They could read the contract, inspect the product, or research the seller. The shortcut is a convenience. With a large AI model, one kind of checking is out of reach. No user can audit a system trained on billions of documents and tuned by processes its own makers describe only in approximation. That opacity is built into the thing, and better disclosure will not fully remove it. But auditing the whole model is not the only kind of checking. A user can still test a bounded output: compare a claim with its primary source, run the code, check the sum. Some uses are easy to check this way. Others, such as a medical or legal judgement the user cannot assess, are hard or unsafe to check alone. That is why the trust question matters so much here: where the output cannot be checked, more of the weight rests on transferred trust and on the institutions standing behind the system.

Trusting the Company Is Not the Same as Trusting the System

There is a second gap hiding inside the first.

The KPMG and University of Melbourne (2025) study found a split between how trustworthy AI is seen to be and how willing people are to rely on it. Edelman's 2025 flash poll on AI, an online survey of about 5,000 people in Brazil, China, Germany, the United Kingdom and the United States, found a related gradient in who people are comfortable seeing use AI: on average, 56% of employees were comfortable with their own employer's use of AI, 46% with business overall, and 34% with government (Grossman, 2025). Table 17.1 gives the exact figures, populations, and years. The two gaps should not be read as one effect measured twice, since the studies use different populations and different survey questions. The Edelman gradient deserves a closer look: the closer and more concrete the institution, the more comfortable people said they were.

One way to read these numbers is through this book's model: confidence sits with institutions people know, and has not yet been carried across to the systems themselves, because the Bridge that would carry it is weak or missing.

That is an interpretation, not something the surveys measured. The same numbers fit other explanations too: doubts about what the systems can do, worries about whose interests they serve, being required to use them at work, unequal access, and the level of risk in a given task. Differences between groups of respondents do not show any one person's trust moving from a Source to a Target.

If the Bridge reading is right, the structure is the one described in earlier chapters, only now the Target is a sentence of generated text or a ranked answer rather than a product or a seller.

The distinction has a direct consequence for anyone building or deploying these systems. If that reading holds, closing the gap is less a matter of persuading people that AI companies are credible and more a matter of constructing the visible, inspectable Bridge that lets a person move from "I trust this company" to "I trust what this system just told me." It is a Bridge an AI product can make visible, or leave hidden.

Trust Is Calibrated to How Much You Can Recover

People do not extend trust to AI in a single block. They ration it by task, and the rationing follows a logic the rest of this book would predict.

Chapter 9 already showed the shape of this: trust holds for reversible, low-stakes tasks and collapses for high-stakes ones. The point worth returning to here is not the percentages again, but what that pattern implies once the Target is a machine rather than a person or a brand.

The pattern is risk-calibrated. People extend Borrowed Confidence to AI in proportion to how recoverable the decision is if the machine is wrong. Gathering information is recoverable; you can check it, ignore it, or try again. A bad legal or medical decision often is not. This is the same instinct that makes a shopper comfortable trusting an unfamiliar product on a familiar marketplace but cautious about a large irreversible purchase. The Eye of the Trustor, the same bounded perceptual lens at work in every earlier chapter, is quietly asking how much exposure the Bridge actually justifies.

Using the Thing May Build Trust

Institutional signals are one path to trust. Direct experience is another, and with AI it may be a quick one.

Edelman's November 2025 flash poll points the same way, with a caution attached. Respondents who said AI had helped them were far more likely to express trust in it than respondents who said it had made no difference: among

employees, the gap between those who said AI had helped them find solutions at work and those who said it had made no impact ranged from 26 to 46 percentage points across four of the five markets surveyed (Grossman, 2025). That is a difference between groups of respondents at one point in time, not a measurement of anyone's trust rising after an experience. It is consistent with experience building trust. It is also consistent with people who already trust AI using it more, and noticing its benefits more.

Edelman also reports that, in its analysis, trust was one of the two strongest predictors of enthusiastic adoption, just behind how informed people felt about AI (Grossman, 2025). A predictor is not a cause, but trust and use do appear to travel together.[4]

There is a deeper point buried in that finding. Direct experience does not merely add to Borrowed Confidence; over time it can replace the need for it. The first time a person uses an unfamiliar AI tool, much of their confidence is transferred, whether from the brand that offers it, the platform it sits inside, or the friend who suggested it. After a handful of successful interactions, a direct basis for trust may begin to form, one that depends less on the original Source, and only for the tasks the user has actually seen the system do. If that happens, it is the same arc by which a stranger becomes a trusted colleague, compressed into days. It also means the Borrowed Confidence of the first encounter is doing enormous work: it carries much of the weight until direct experience can take over.

The practical lesson for anyone deploying AI is blunt. A safety page or a governance statement is not enough on its own. A first interaction that works, on a task the user can check, matters too. Design for that moment, and for the user's ability to check it.

What People Are Really Asking for Is Governance

When people are asked what would actually make them trust AI, they do not ask for cleverer models. They ask for institutions.

The KPMG and Melbourne (2025) study found that far more people believe AI regulation is necessary than believe existing laws are adequate to make AI safe. The sharpest version of that gap is awareness itself: most people are unaware of any law, regulation, or policy that applies to AI in their country (see Table 17.1 for the exact figures)[5]. Safeguards can protect people even when those people cannot see them. But a safeguard nobody knows about cannot help anyone judge how much to rely on a system, and on these figures most people know of none (KPMG & University of Melbourne, 2025).

The specific mechanisms people want read like a blueprint for transfer infrastructure. The most popular is the right to opt out of having personal data used by AI, which 86% say would increase their trust. Close behind sit accuracy monitoring, the ability for a human to intervene and override an output, and laws governing responsible AI, each endorsed by around 84%. Independent third-party audit, the machine-age equivalent of professional accreditation, was endorsed by 74%.

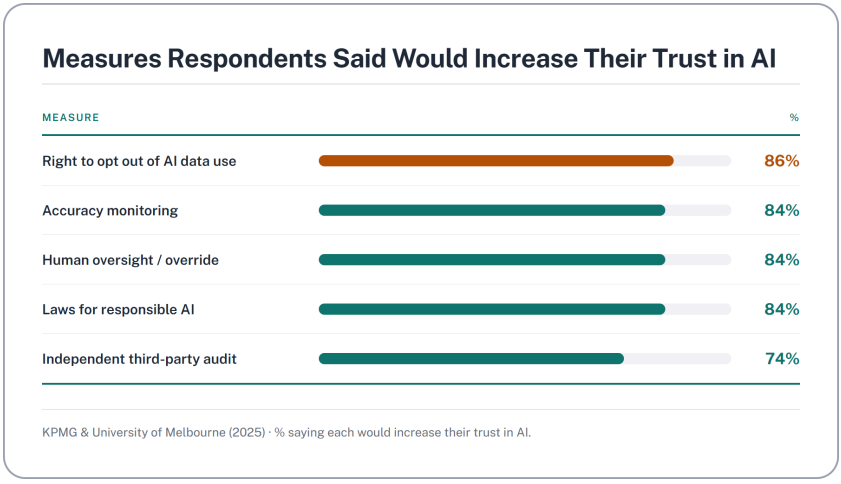


Figure 17.1 Measures respondents said would increase their trust in AI. Percentages are not necessarily mutually exclusive and should be read as reported survey preferences, not as a ranked causal model of trust.

The same study points to why this matters so urgently. Eighty-seven per cent of people want stronger laws against AI-generated misinformation, and 70% say they now find it hard to trust online information at all because they cannot reliably tell what is real from what a machine produced. This is the environment in which AI is being deployed: not one of general hostility, but one of specific, well-articulated demand for visible accountability. When people cannot tell a genuine signal from a synthetic one, the cheap, ambient trust that once made platforms feel efficient begins to drain away. A credible institution standing visibly behind the output is one of the things that can refill it.

None of this is a demand to ban AI. It is a demand for the Bridges. People are describing the regulatory and accountability frameworks that would let trust flow from well-designed governance to a specific deployment. Where those Bridges exist and are visible, people have more to go on. Where they are missing, even a capable system has to earn trust some other way, usually slowly, one checkable task at a time.

Four Ways Trust Transfers to a Machine

Read through this book's model, the survey numbers suggest four pathways. Mapped to the Bridge Forms from Chapter 2, brand-to-AI transfer is Embedding or Integration, regulator-to-AI transfer is Certification, explainability is not a separate form but a way of making any Bridge legible, and multi-source trust is several Bridges at once. Each has some evidence behind it, of differing strength.

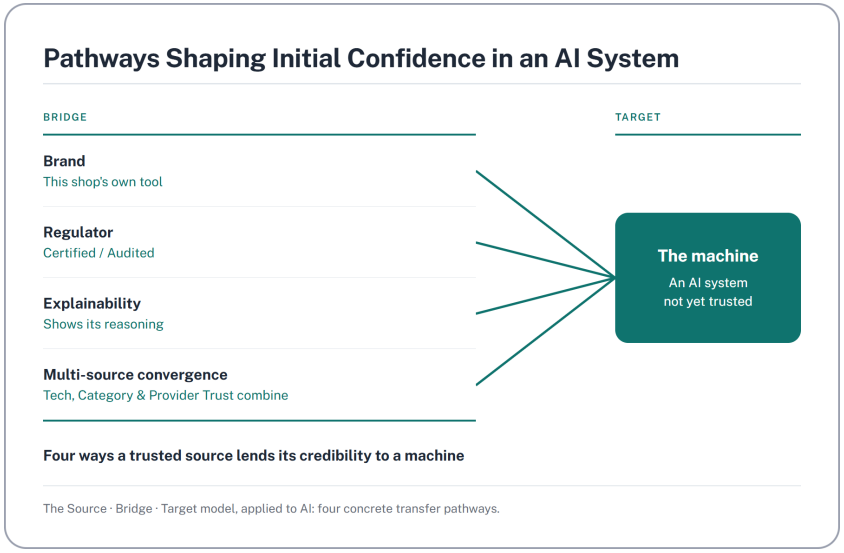


Figure 17.2 Pathways shaping initial confidence in an AI system. Initial confidence may be influenced by brand-to-AI transfer, regulator-to-AI transfer, explainability, and multi-source convergence. These pathways differ in what they establish and should not be treated as interchangeable guarantees of system quality or safety.

The first is brand-to-AI transfer. When an AI tool is embedded in a brand a person already trusts, the brand's credibility flows to the tool's outputs. An Ipsos US sample found that a majority trusted AI product recommendations from a specific store or site using their purchase history (Ipsos, 2024). Edelman's gradient, earlier in this chapter, points the same way: people were more comfortable with their own employer's use of AI than with business in general (Grossman, 2025). Different questions and populations: not one measured gap, but the same pattern Chapters 4 and 7 already named. A concrete Source seems to carry more confidence than an abstract category.

The second is regulator-to-AI transfer. When a system is certified, audited, or approved by a trusted body, the institution's credibility passes to the system. This is the same mechanism by which a patient trusts a medication because a regulator approved it, not because they read the trial data. The demand for this is overwhelming, with around 84% of people wanting laws for responsible AI.

On the same survey, most people know of no AI law at all, which suggests that this Bridge, where it exists, is not yet visible to them (KPMG & University of Melbourne, 2025).

The third is explainability as relationship-building. When a system shows its reasoning ("recommended because you saved similar items," or "based on guidelines from a named institution"), it makes the Bridge between Source and output visible to the Trustor. This is the same move as a "featured seller" label or a university crest, except the machine is constructing the link in real time. Transparency here is more than an ethical nicety. It is one of the clearest mechanisms by which the Bridge becomes visible to the Trustor, though not the only one this book has described: ambiguous or even invisible Bridges have already been shown, in earlier chapters, to carry confidence across.

The fourth is multi-source AI trust. People rarely trust a new AI system from one Source alone. Research on AI-capable technologies, including autonomous vehicles, found that users draw at once on trust in the underlying technology, trust in the AI category from prior exposure, and trust in the provider building it (Renner et al., 2021; see also Lee & See, 2004). Each Source contributes to confidence in the new system. This is the same multi-source pattern the framework was used to interpret in the FTX collapse in the opening chapter, now operating in a car that drives itself. The same framework, used earlier to interpret how many people came to trust a fraudulent exchange, can also be applied to how a person decides whether to let a machine take the wheel.

AI Trust Survey Sources

The studies, populations, and years behind the AI trust figures cited in Chapter 17.

STUDY	POPULATION	YEAR	ITEM MEASURED	FIGURE
KPMG & University of Melbourne	Global (47 countries)	2025	AI use vs. AI trust	66% vs. 46%
KPMG & University of Melbourne	Advanced economies	2025	AI trust	39%
KPMG & University of Melbourne	Australia	2022 → 2024	Organisational AI adoption	23% → 65%
KPMG & University of Melbourne	Australia	2025	Willing to trust AI (vs. 46% global)	36%
KPMG & University of Melbourne	Global	2025	AI seen as trustworthy vs. willing to trust	58% vs. 46%
KPMG & University of Melbourne	Global	2025	Believe AI regulation necessary / existing laws adequate	70% / 43%
KPMG & University of Melbourne	Global	2025	Unaware of any AI law	83%
KPMG & University of Melbourne	Global	2025	Want data opt-out / accuracy monitoring, human override, laws / third-party audit	86% / ~84% / 74%
Edelman flash poll (Grossman)	Online, 5 countries, ~5,000	Oct 2025	Comfortable with AI use by own employer / business / government	56% / 46% / 34%
Edelman flash poll (Grossman)	Employees, 4 of 5 countries	Oct 2025	Trust gap: AI helped at work vs. no impact (group difference, not change over time)	26–46 points
Ipsos Consumer Tracker (Wave 93)	US adults	March 2024	Trust AI product recommendations from a familiar store or site	~51%

A tilde (~) marks an approximate figure as reported by the source. Survey differences between groups are associations, not measured changes within people.

▼ Text version of this table

Study	Population	Year	Item measured	Figure
KPMG & University of Melbourne	Global (47 countries)	2025	AI use vs. AI trust	66% vs. 46%
KPMG & University of Melbourne	Advanced economies	2025	AI trust	39%
KPMG & University of Melbourne	Australia	2022 → 2024	Organisational AI adoption	23% → 65%
KPMG & University of Melbourne	Australia	2025	Willing to trust AI (vs. 46% global)	36%

Study	Population	Year	Item measured	Figure
KPMG & University of Melbourne	Global	2025	AI seen as trustworthy vs. willing to trust	58% vs. 46%
KPMG & University of Melbourne	Global	2025	Believe AI regulation necessary / existing laws adequate	70% / 43%
KPMG & University of Melbourne	Global	2025	Unaware of any AI law	83%
KPMG & University of Melbourne	Global	2025	Want data opt-out / accuracy monitoring, human override, laws / third-party audit	86% / ~84% / 74%
Edelman flash poll (Grossman)	Online, 5 countries, ~5,000	Oct 2025	Comfortable with AI use by own employer / business / government	56% / 46% / 34%
Edelman flash poll (Grossman)	Employees, 4 of 5 countries	Oct 2025	Trust gap: AI helped at work vs. no impact (group difference, not change over time)	26–46 points
Ipsos Consumer Tracker (Wave 93)	US adults	March 2024	Trust AI product recommendations from a familiar store or site	~51%

Table 17.1 **AI Trust Survey Sources.** The studies, populations, and years behind every AI trust statistic cited in this chapter.

Where the Model Is Tested by Machines

The four pathways above show that the framework's mechanics still apply to machines: trust still moves from a perceived Source, across a perceived Bridge, to a Target. But AI raises four questions in a new combination, and each one changes what "perceived" can mean once the perceiving is partly done by the system itself.

- The Source is difficult to identify, decompose, or inspect. When a friend recommends something, the Trustor can weigh the friend's track record, motives, and expertise. When an AI recommends something, the Source is a training set, a weighting function, and a retrieval step that even the system's builders cannot fully explain. The Trustor is asked to send trust across a Bridge whose mechanism is largely invisible and whose Source is, in any ordinary sense, hard to pin down.
- The Bridge is algorithmically constructed. In most of the human cases in this book, the link was something that had happened: an endorsement, a partnership, an affiliation, even when it was over-read or, as with paid reviews, manufactured. In AI, the link is often generated at the moment of output. "Recommended because of your history" is not a Bridge that two parties formed; it is a pattern inference. The Trustor perceives a Bridge that the Source never intended and the Target never negotiated.

- Trust can be transferred to synthetic Sources. AI can manufacture the very signals that carry trust (endorsements, reviews, testimonials, authority markers) and make them hard to tell from genuine ones. Earlier chapters already met fake signals: paid reviews, manufactured consensus, misleading affiliations. AI does not invent that problem. It changes its scale, speed and cost, and makes the fakes easier to personalise and harder to spot.
- Decay rates are unknown. For human relationships, there are decades of research on how trust erodes after a betrayal or a scandal. For AI, the evidence this book reviewed does not settle how fast transferred trust collapses after a single hallucination, a confident error, or a disclosed manipulation. It remains an open question, and it is one of the most important questions for anyone designing governance, because a system whose trust collapses instantly behaves very differently from one whose trust degrades slowly.

Taken together, these four properties do something the rest of this book has not had to confront. Earlier chapters met each of these problems on its own: Sources that were hard to examine, Bridges that were over-read or faked, signals that were manufactured. With AI, they can arrive together, at scale, and in the same moment. The Source is difficult to inspect, the Bridge is generated, inferred, or otherwise mediated in the instant rather than actually formed, the signals can be wholly synthetic, and the rules of collapse are unwritten. The model still describes what is happening. Trust is still moving from a perceived Source, across a perceived Bridge, to a Target. But more of that perception can now be shaped by the system itself. That is not a verdict that the model has failed. It is an open question, and this book does not settle it.

The Bridge to What Follows

Applying a framework to a case is not the same as testing it. On the readings offered here, the framework laid out in this book (Source, Bridge, and Target, linked by Transfer) has been a useful way to describe what happens in endorsements, in institutions, in platforms, in social networks, in search, and in commerce. But it now meets conditions it was not built for: systems that have no reputation in the traditional sense, Bridges that are algorithmically constructed rather than actually formed, and Sources that may be synthetic rather than real. A companion volume takes up whether those loosenings change the model, or only the evidence a Trustor can get.

Chapter Takeaway

On this book's reading, trust moves toward machines through the same Source → Bridge → Target pattern that governs human trust, along four pathways: from a trusted brand, from a regulator, through visible explanation, and from several Sources at once. On that reading, the mechanics still apply: people still come to trust a machine by moving confidence from a Source, across a Bridge, to a Target, and all three can still be seen in how they do it. Whether that trust is well-founded is a harder question, because the Source, the Bridge, and the signals themselves can now be manufactured by the system doing the asking. Those are the four open questions this chapter surfaces rather than settles.

The defining fact of AI in this decade, set out in Table 17.1, is the gap between use and trust: 66% use it, 46% trust it, and only 36% in Australia. People are not asking only for better models. Respondents also place substantial weight on institutional Bridges, among them audits, oversight, and the right to opt out, that would let them borrow trust from authorities they already believe in and extend it to systems they cannot inspect.

When trust transfers to a machine across a Bridge that is invisible, invented, or faked, the Trustor may be lending confidence to a Source they have no way to check. If that confidence collapses, the mistrust will travel backward through the same opaque Bridges. We do not yet know how quickly it does. It may take the trusted brands, platforms, and institutions that deployed the machine down with it.

4 Edelman flash poll (Nov 2025): trusters vs. distrusters are 4.5x more likely to accept AI for job-search decisions, 6x for financial decisions, 5.5x for healthcare decisions.

5 KPMG & University of Melbourne (2025): 83% of people are unaware of any law, regulation, or policy that applies to AI in their country.

In Closing: What to Carry Forward

This book began with a café a friend recommended. Your friend was the Source, the recommendation was the Bridge, the café was the Target, and you were the one deciding. Everything since has been that same small decision, carried into brands, platforms, institutions, verifiers and machines, where the Bridges get longer, the Sources harder to see, and the stakes larger.

The model does not tell you whom to trust. It tells you where to look. Borrowed Confidence is how anyone gets started in a world too complex to check from first principles, and it is not a mistake. The mistake is letting a Bridge carry more than it was built for: reading a badge as a guarantee, a ranking as a recommendation, a famous name as a warranty, or a fluent answer as a checked one.

So the next time something feels trustworthy, name what is doing the reassuring. Ask what the connection actually establishes, what remains unknown, and who bears the cost if it is wrong. Look for evidence that fits the decision, not simply a stronger-looking signal. Then keep your judgement open to what the Target actually does: more reliance if it earns it, less if it does not, and “not sure yet” for as long as the evidence is thin.

Trust Transfer is not a reason to reject every shortcut. It is a way to see what the shortcut carries before you ask it to carry you.

Appendix: A Decision Sheet for Organisations

This sheet turns the book's questions into one page an organisation can fill in before it relies on a supplier, a partner, a platform or a badge, and again if something goes wrong. Use it to record a decision, not to score one. It is a diagnostic aid, not a validated compliance instrument.

Decision and scope: What exactly are we allowing this person, product or organisation to do, and for what?

Who decides, and who bears the risk: Who makes the call, and who would carry the money, health, legal, privacy or reputational consequences if it goes wrong? They are not always the same people.

Source, Bridge, Target: Who or what supplies our first confidence, through which specific connection, towards which claim or service?

What was actually checked: What was checked, by whom, against what standard, on what date, and what continues to be checked?

What we know, and how we know it: Which evidence comes from our own experience of the Target, which from its Public Record, which from testimony, and what is still unknown? Write "unknown" where it is unknown; do not turn it into a score.

Action: Proceed, run a limited trial, find out more, restrict use, or decline. Give the reason.

Owner and stop rule: Who is accountable, and what event or evidence would make us restrict or withdraw?

Review: When will we look at this again, what will we watch in the meantime, and what evidence would change the decision?

If it fails: What do we contain straight away, what cause and harm must be repaired, who checks the fix, who needs to be told, and how will we watch for recurrence? Chapter 13 sets out the full Contain–Repair–Monitor sequence.

Key Terms and How to Use Them

Most trust in modern life is not built from direct experience. It moves from something known toward something less known through a Bridge the Trustor interprets.

This book uses that sequence to examine how trust travels through brands, recommendations, institutions, platforms, rankings, certifications, partnerships, and other social arrangements. The terms below give that pattern a precise language. You do not need to read this section straight through. Return to it whenever a distinction becomes important. This book makes the Trustor's bounded perceptual position explicit through a concept it calls the Eye of the Trustor: the bounded perceptual lens through which every link in the chain is read. This book keeps the person doing that reading explicitly in view, and asks why the same signal can land differently depending on whose eye it passes through.

Trust

In everyday life, trust can mean confidence, comfort, belief, loyalty, or optimism. In this book, trust means accepting Vulnerability on the basis of a positive expectation about a person, institution, product, platform, claim, or system.

Trust is never free. It is a judgement — not a feeling — that something is capable, honest, and safe enough to rely upon, given what is at stake.

When you encounter trust in this book, ask: Is this confidence earned, or is it still borrowed?

Trustor

The Trustor is the person or group interpreting the Bridge and deciding whether to accept Vulnerability.

The Trustor sits outside the directional formula, not at the end of it — the party whose judgement the whole model exists to describe. Put simply, the Trustor is the person doing the trusting.

When you encounter a Trustor in this book, ask: Whose judgement is being shaped, and what makes this Bridge persuasive to them?

Eye of the Trustor

The Eye of the Trustor is the bounded perceptual lens through which the Trustor reads every link in the chain: the Source, the Bridge, the Target, and, once Direct Evidence is available, the Target's own record.

It does not switch off once the crossing is made — the same lens later reads the Target's own conduct, which is how Earned Trust and Earned Mistrust are reached, or how a judgement stays unresolved.

When you encounter the Eye of the Trustor in this book, ask: Whose perspective is doing the reading, and what might it be missing?

Vulnerability

Vulnerability is the exposure to risk that makes trust necessary in the first place: the possibility of being wrong, disappointed, or harmed if the Target does not deliver what the Bridge implied.

When you encounter Vulnerability in this book, ask: What, exactly, is being risked here, and by whom?

Source

The Source is the known entity from which confidence originates.

A Source can be a friend, expert, brand, institution, marketplace, media outlet, regulator, school, hospital, employer, platform, or familiar interface. It is the recognised reference point from which a person begins to extend confidence.

When you encounter a Source in this book, ask: What is the trusted starting point here?

Bridge

The Bridge is the perceived connection between a Source and a Target.

That connection may be formal, commercial, institutional, social, technical, or symbolic. It may be strong or weak, direct or indirect, warranted or misleading. What matters is that the Trustor interprets the connection as meaningful enough for confidence in the Source to affect judgement of the Target. A Bridge only becomes consequential once it takes one of the recognisable forms described in Chapter 2.

When you encounter a Bridge in this book, ask: Why does this connection make trust feel transferable, and what form makes it consequential here?

Bridge Forms

Bridge Forms are the six recognisable ways a Bridge carries confidence from a Source to a Target, set out in Chapter 2: Endorsement, Certification, Ranking or Placement, Embedding or Integration, Proxy, and Co-presence. The visible form of a connection does not by itself establish its scope, basis, or accountability.

When you encounter a Bridge Form in this book, ask: What kind of connection is this, and what does that kind actually guarantee?

Target

The Target is the unknown or less-known person, product, claim, service, organisation, or system receiving confidence.

The Target need not be completely unfamiliar. It may simply be known mainly through the Source rather than through direct experience. A reader may trust a book because of an author's reputation, a seller because of a marketplace guarantee, or a product because it appears inside a trusted platform.

When you encounter a Target in this book, ask: What is the thing people are being asked to trust?

Trust Transfer

Trust Transfer is the movement of confidence from a Source toward a Target through a perceived Bridge, made consequential by one of its recognisable forms.

That Borrowed Confidence can be useful — no one can inspect every product, institution, or system from first principles — but it isn't Earned Trust. A familiar Source can make a Target look more credible than the evidence actually supports.

When you encounter Trust Transfer in this book, ask: Where did this confidence actually originate, and what carried it here?

Borrowed Confidence

Borrowed Confidence is confidence in a Target that arrives through a Bridge from a Source, before the Target's own record exists. It is not naive or corrupt by definition; it is a way of beginning under uncertainty, and one way a complex society stays usable.

When you encounter Borrowed Confidence in this book, ask: What is carrying this confidence, and what would the Target's own record add?

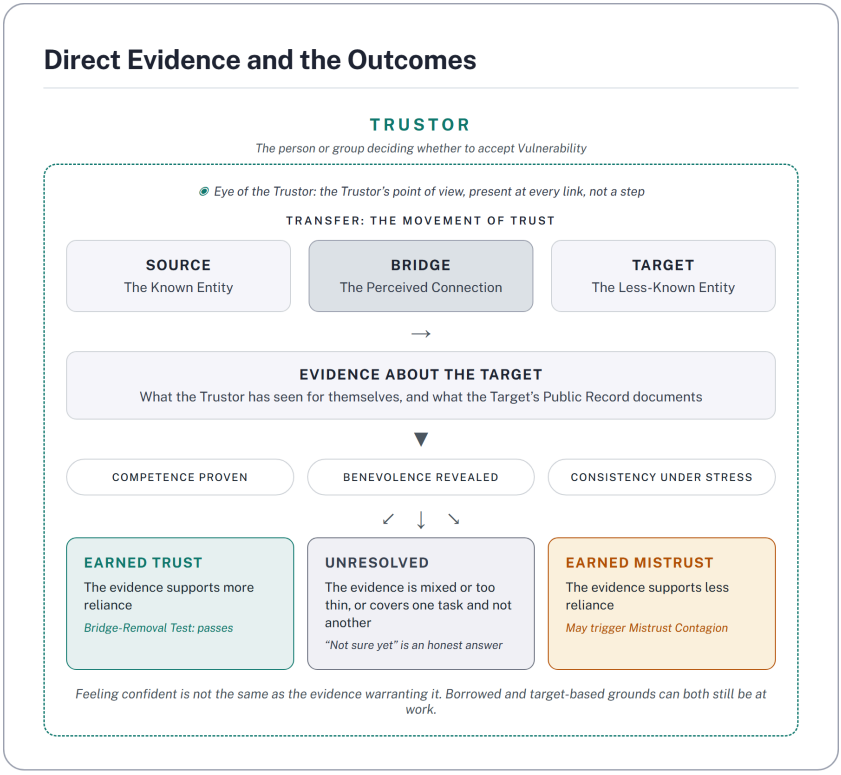


Figure 0.2 Direct Evidence and the Outcomes. Once Direct Evidence becomes available, the same lens reads the Target directly, and the result may be Earned Trust, Earned Mistrust, or a judgement still unresolved.

Direct Evidence

Direct Evidence is what the Trustor observes about the Target itself, once the Target is being read directly rather than through a Bridge. A friend's account of the Target is not your Direct Evidence: it was first-hand for them, but it reaches you through another Bridge, as testimony.

When you encounter Direct Evidence in this book, ask: What has the Target itself shown, apart from what the Bridge claimed?

Earned Trust

Earned Trust is what's left once reliance stops depending on the Source that introduced it — the point where confidence rests on the Target's own track record, not on borrowed credibility. "Earned" means the evidence supports it, not just that it feels settled: a Trustor can feel sure and still be miscalibrated.

When you encounter Earned Trust in this book, ask: If I had never heard of the original Source, would what the Target has shown still justify this trust?

Earned Mistrust

Earned Mistrust is what happens when Direct Evidence about the Target justifies less reliance. It's the paired outcome to Earned Trust — not the absence of one.

When you encounter Earned Mistrust in this book, ask: What has the Target itself done to justify less reliance, rather than more?

Unresolved

Unresolved is the honest position while the evidence about the Target is mixed, too thin, or good for one task and silent on another. It is not a failure to decide. Borrowed Confidence and the Target's own record may both still be doing some of the work.

When you encounter Unresolved in this book, ask: What would I need to see before I could say more?

Bridge-Removal Test

The Bridge-Removal Test asks one question: if you had never received the original introduction, would what the Target has shown since still justify relying on it? Everything that still protects you stays in place while you ask it, such as regulation, certification and oversight. The test removes the introduction, not the safeguards.

When you encounter the Bridge-Removal Test in this book, ask: Has the Trustor's own record taken over, or is the original introduction still carrying the weight?

Trustor's Record

The Trustor's Record is the Trustor's own accumulated history of positive and negative interactions with the Target, once Direct Evidence starts building up. It is evidence about the Target, not a new Source.

When you encounter the Trustor's Record in this book, ask: What has this relationship actually shown, on balance, and is it strong enough to carry someone else's reliance?

Target's Public Record

The Target's Public Record is what official or published sources document about the Target's own conduct: regulator findings, court outcomes, inspection results. It reaches the Trustor through a Bridge, but it records what the Target did, so together with the Trustor's Record it counts as evidence about the Target. A friend's account of their own experience is neither record: it is testimony, another Bridge.

When you encounter the Target's Public Record in this book, ask: Who documented this, about what conduct, and when?



Figure 0.3 Mistrust Contagion. Doubt can move from a failed Source toward its Targets, from a failed Target back toward its Source, both ways when the Bridge fails, or outward across a shared network.

Mistrust Contagion

Mistrust Contagion is suspicion moving through Bridges and networks, not always in the same direction trust originally moved.

When you come across Mistrust Contagion in this book, ask: Which Bridge is now carrying doubt, and who else becomes implicated?

Borrowed Legitimacy

Borrowed Legitimacy is confidence in a Target that goes beyond what its connection to something already trusted actually warrants: being placed near, inside, beside, or beneath that Source is read as more than it is. It is a question of surplus inference rather than fraud. Ordinary, warranted borrowing is Borrowed Confidence, not Borrowed Legitimacy.

When you encounter Borrowed Legitimacy in this book, ask: What does this connection actually guarantee, and what am I reading into it?

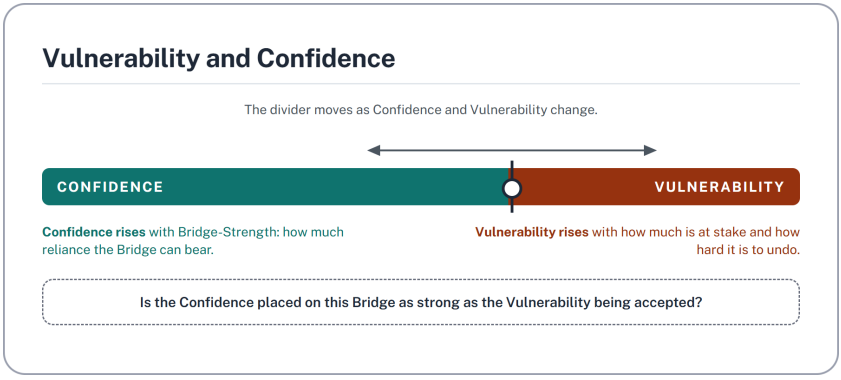
Contain–Repair–Monitor Model

The Contain–Repair–Monitor Model is this book's three-phase response to a trust failure: Contain the failure, Repair the cause, harm, and trust relationship, and Monitor whether the correction holds. An organisation can complete the corrective actions; only the Trustor can judge whether the resulting Direct Evidence is enough to re-earn trust.

When you encounter the Contain–Repair–Monitor Model in this book, ask: Which phase is this response in, and what would show that it has worked?

Vulnerability and Confidence

The Bridge lends Confidence; the Trustor takes on Vulnerability. Trust Transfer is sound when the Confidence a Bridge deserves is at least as great as the Vulnerability the Trustor accepts.



▼ Text version of this figure

Confidence is lent by the Bridge and rises with Bridge-Strength, how much reliance the Bridge can bear. Vulnerability is what the Trustor risks by accepting it and rises with how much is at stake and how hard it is to undo. The two share one bar and the divider between them moves as either changes. The test: is the Confidence placed on this Bridge as strong as the Vulnerability being accepted?

Figure 0.4 Vulnerability and Confidence. Confidence, lent by the Bridge, sits on one side of a bar and Vulnerability, what the Trustor risks by accepting it, on the other. The divider moves as either changes.

References

Accounting Today. (2023, March 15). KPMG faces questions over SVB and Signature audits. https://www.accountingtoday.com/news/kpmg-faces-questions-over-svb-and-signature-audits

Barclays. (2017, June 27). From the archives: The ATM is 50. https://home.barclays/news/2017/06/from-the-archives-the-atm-is-50/

Berliner, U. (2017, August 31). Wells Fargo admits to nearly twice as many possible fake accounts — 3.5 million. NPR. https://www.npr.org/sections/thetwo-way/2017/08/31/547550804/wells-fargo-admits-to-nearly-twice-as-many-possible-fake-accounts-3-5-million

Berwick, A. (2022, November 12). Exclusive: At least $1 billion of client funds missing at failed crypto firm FTX. Reuters. https://www.reuters.com/markets/currencies/exclusive-least-1-billion-client-funds-missing-crypto-firm-ftx-sources-2022-11-11/

Bloomberg Tax. (2025, September 17). KPMG dismissed red flags at regional banks, Senate review finds. https://news.bloombergtax.com/financial-accounting/kpmg-dismissed-red-flags-at-regional-banks-senate-review-finds

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (BEA). (2012). Final report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris (Report f-cp090601). Ministère de l'Écologie, du Développement durable et de l'Énergie.

Cain, D. M., Loewenstein, G., & Moore, D. A. (2005). The dirt on coming clean: Perverse effects of disclosing conflicts of interest. The Journal of Legal Studies, 34(1), 1–25. https://doi.org/10.1086/426699

Cao, Z., & Belo, R. (2024). Effects of explicit sponsorship disclosure on user engagement in social media influencer marketing. MIS Quarterly, 48(1), 375–392. https://doi.org/10.25300/MISQ/2023/17944

Carreyrou, J. (2018). Bad blood: Secrets and lies in a Silicon Valley startup. Alfred A. Knopf.

Cialdini, R. B. (2021). Influence: The psychology of persuasion (New and expanded ed.). Harper Business.

CNN Money. (2016, October 24). Wells Fargo's reputation is tanking, survey finds. https://money.cnn.com/2016/10/24/investing/wells-fargo-fake-accounts-angry-customers/index.html

Consumer Financial Protection Bureau. (2016, September 8). Hundreds of thousands of accounts secretly created by Wells Fargo Bank employees leads to historic $100 million fine from the CFPB [Press release]. https://www.consumerfinance.gov/archive/blog/hundreds-thousands-accounts-secretly-created-wells-fargo-bank-employees-leads-historic-100-million-fine-cfpb/

Corporación Latinobarómetro. (2018). Informe 2018. https://www.latinobarometro.org/

Dirks, K. T., Lewicki, R. J., & Zaheer, A. (2009). Repairing relationships within and between organizations: Building a conceptual foundation. Academy of Management Review, 34(1), 68–84. https://doi.org/10.5465/AMR.2009.35713285

Doney, P. M., & Cannon, J. P. (1997). An examination of the nature of trust in buyer–seller relationships. Journal of Marketing, 61(2), 35–51. https://doi.org/10.1177/002224299706100203

Edelman. (2025). 2025 Edelman Trust Barometer. Edelman Trust Institute. https://www.edelman.com/trust/2025/trust-barometer

Encyclopaedia Britannica. (n.d.). Elisha Otis. Retrieved 14 November 2026, from https://www.britannica.com/biography/Elisha-Otis

eSafety Commissioner. (2024). Transparency report: Recommender systems and algorithmic ranking on online platforms. Australian Government.

European Commission. (2024, May 16). Commission opens formal proceedings against Meta under the Digital Services Act (Press release IP/24/2662).

Federal Aviation Administration. (2020, November 18). FAA statement on Boeing 737 MAX return to service. FAA Updates on Boeing 737 MAX. https://www.faa.gov/newsroom/faa-updates-boeing-737-max-0

Federal Trade Commission. (2016a, March 15). Lord & Taylor, LLC; Complaint (FTC File No. 152-3181, Docket No. C-4576). https://www.ftc.gov/legal-library/browse/cases-proceedings/152-3181-c4576-lord-taylor-llc-matter

Federal Trade Commission. (2016b, May 23). Lord & Taylor, LLC; Decision and order (FTC File No. 152-3181, Docket No. C-4576). https://www.ftc.gov/system/files/documents/cases/160523lordtaylordo.pdf

Financial Crisis Inquiry Commission. (2011). The financial crisis inquiry report: Final report of the National Commission on the Causes of the Financial and Economic Crisis in the United States. U.S. Government Printing Office.

Fuoli, M., & Paradis, C. (2014). A model of trust-repair discourse. Journal of Pragmatics, 74, 52–69. https://doi.org/10.1016/j.pragma.2014.09.001

Gaydou, E. (n.d.). Seven people die after taking cyanide-laced Tylenol. EBSCO Research Starters. Retrieved 14 November 2026, from https://www.ebsco.com/research-starters/history/seven-people-die-after-taking-cyanide-laced-tylenol

Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hassed, G. (2025). Trust, attitudes and use of artificial intelligence: A global study 2025. The University of Melbourne & KPMG. https://doi.org/10.26188/28822919

Gossner, C. M.-E., Schlundt, J., Ben Embarek, P., Hird, S., Lo-Fo-Wong, D., Beltran, J. J. O., Teoh, K. N., & Tritscher, A. (2009). The melamine incident: Implications for international food and feed safety. Environmental Health Perspectives, 117(12), 1803–1808. https://doi.org/10.1289/ehp.0900949

Grossman, G. (2025, November 18). When familiarity breeds trust: How experience turns AI skeptics into believers. Edelman. https://www.edelman.com/insights/familiarity-breeds-trust-in-ai

Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect. Cognitive Research: Principles and Implications, 6, Article 38. https://doi.org/10.1186/s41235-021-00301-5

Henn, S. (2015, July 31). Remembering when driverless elevators drew skepticism [Radio transcript]. NPR. https://www.npr.org/2015/07/31/427990392/remembering-when-driverless-elevators-drew-skepticism

Ho, D. E., Ashwood, Z. C., & Handan-Nader, C. (2019). New evidence on information disclosure through restaurant hygiene grading. American Economic Journal: Economic Policy, 11(4), 404–428. https://doi.org/10.1257/pol.20180230

House Committee on Transportation and Infrastructure. (2020, September 16). After 18-month investigation, Chairs DeFazio and Larsen release final committee report on Boeing 737 MAX [Press release]. https://democrats-transportation.house.gov/news/press-releases/after-18-month-investigation-chairs-defazio-and-larsen-release-final-committee-report-on-boeing-737-max

Hout, M., & Hastings, O. P. (2016). Reliability of the core items in the General Social Survey: Estimates from the three-wave panels, 2006–2014. Sociological Science, 3, 971–1002. https://doi.org/10.15195/v3.a43

Ipsos. (2024, March). Ipsos Consumer Tracker: Wave 93 [Topline results]. https://www.ipsos.com/sites/default/files/ct/news/documents/2024-03/Topline%20Ipsos%20Consumer%20Tracker%20Wave%2093.pdf

Jin, G. Z., & Leslie, P. (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. Quarterly Journal of Economics, 118(2), 409–451. https://doi.org/10.1162/003355303321675428

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: The effects of apology versus denial for repairing competence- versus integrity-based trust violations. Journal of Applied Psychology, 89(1), 104–118. https://doi.org/10.1037/0021-9010.89.1.104

Knittel, C. R., & Stango, V. (2014). Celebrity endorsements, firm value, and reputation risk: Evidence from the Tiger Woods scandal. Management Science, 60(1), 21–37. https://doi.org/10.1287/mnsc.2013.1749

Knowledge at Wharton. (2012, October 2). Tylenol and the legacy of J&J's James Burke. https://knowledge.wharton.upenn.edu/article/tylenol-and-the-legacy-of-jjs-james-burke/

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human Factors, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

Leveson, N. G., & Turner, C. S. (1993). An investigation of the Therac-25 accidents. IEEE Computer, 26(7), 18–41. https://doi.org/10.1109/MC.1993.274940

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. Proceedings of the National Academy of Sciences, 108(22), 9020–9025. https://doi.org/10.1073/pnas.1008636108

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734. https://doi.org/10.2307/258792

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002a). Developing and validating trust measures for e-commerce: An integrative typology. Information Systems Research, 13(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002b). The impact of initial consumer trust on intentions to transact with a web site: A trust building model. Journal of Strategic Information Systems, 11(3–4), 297–323. https://doi.org/10.1016/S0963-8687(02)00020-3

McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. Academy of Management Review, 23(3), 473–490. https://doi.org/10.5465/amr.1998.926622

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. Journal of Communication, 60(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

The Motley Fool. (2017, February 23). Wells Fargo sees its reputation hit a new low. https://www.fool.com/investing/2017/02/23/wells-fargo-sees-its-reputation-hit-a-new-low.aspx

Network Contagion Research Institute. (2021). A contagion of institutional distrust. Rutgers Miller Center. https://millercenter.rutgers.edu/wp-content/uploads/2021/03/NCRI_Anti-Vaccination_v5.pdf

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. Human Factors, 39(2), 230–253. https://doi.org/10.1518/001872097778543886

PYMNTS. (2023, March 14). SVB and Signature Bank passed KPMG audits days before collapses. https://www.pymnts.com/digital-first-banking/2023/svb-and-signature-bank-passed-kpmg-audits-days-before-collapses/

Reel v. Fyre Festival, LLC, No. 1:17-cv-21683-FAM (S.D. Fla. filed May 5, 2017) (complaint).

Renner, M., Lins, S., Söllner, M., Thiebes, S., & Sunyaev, A. (2021). Achieving trustworthy artificial intelligence: Multi-source trust transfer in artificial intelligence-capable technology. Proceedings of the 42nd International Conference on Information Systems (ICIS 2021), Paper 1120, Austin, TX.

Romeo, G., & Conti, D. (2025). Exploring automation bias in human–AI collaboration: A review and implications for explainable AI. AI & Society, 41, 259–278. https://doi.org/10.1007/s00146-025-02422-7

Singhal, A., & Cutts, M. (2011, February 24). Finding more high-quality sites in search. Official Google Blog. https://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html

Stewart, K. J. (2003). Trust transfer on the World Wide Web. Organization Science, 14(1), 5–17. https://doi.org/10.1287/orsc.14.1.5.12810

Stewart, K. J., & Zhang, Y. (2003). Effects of hypertext links on trust transfer. Proceedings of the 5th International Conference on Electronic Commerce, 235–239. https://doi.org/10.1145/948005.948037

Sullivan, D. (2017, March 5). Google's "one true answer" problem: When featured snippets go bad. Search Engine Land. https://searchengineland.com/googles-one-true-answer-problem-featured-snippets-

270549

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

U.S. Department of Justice, Office of Public Affairs. (2024, March 28). Samuel Bankman-Fried sentenced to 25 years for his orchestration of multiple fraudulent schemes [Press release]. https://www.justice.gov/archives/opa/pr/samuel-bankman-fried-sentenced-25-years-his-orchestration-multiple-fraudulent-schemes

U.S. Food and Drug Administration. (2021, December 8). Coronavirus (COVID-19) update: FDA authorizes new long-acting monoclonal antibodies for pre-exposure prevention of COVID-19 in certain individuals [Press release]. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-new-long-acting-monoclonal-antibodies-pre-exposure

U.S. Food and Drug Administration. (2023a, January 26). Fact sheet for healthcare providers: Emergency use authorization for Evusheld (tixagevimab co-packaged with cilgavimab). https://www.fda.gov/media/154701/download

U.S. Food and Drug Administration. (2023b, January 26). Letter of authorization: Evusheld emergency use authorization (revised scope). https://www.fda.gov/media/154704/download

U.S. Securities and Exchange Commission. (2022, December 13). SEC charges Samuel Bankman-Fried with defrauding investors in crypto asset trading platform FTX [Press release]. https://www.sec.gov/newsroom/press-releases/2022-219

U.S. Securities and Exchange Commission. (2023, September 25). Deutsche Bank subsidiary DWS to pay $25 million for anti-money laundering violations and misstatements regarding ESG investments [Press release]. https://www.sec.gov/newsroom/press-releases/2023-194

UL Research Institutes. (n.d.). Our history. Retrieved 14 November 2026, from https://ul.org/about/our-history/

UL Solutions. (n.d.). Follow-up services: Ongoing onsite certification inspections. Retrieved 14 November 2026, from https://www.ul.com/resources/follow-up-services

Winfrey, O. (Executive Producer). (2006, January 26). The Oprah Winfrey Show [Television broadcast]. Harpo Productions.

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgment. Organizational Behavior and Human Decision Processes, 103(1), 104–120. https://doi.org/10.1016/j.obhdp.2006.05.006

Further reading

Binzer, B., Kendziorra, J., Witte, A.-K., & Winkler, T. J. (2024). Trust in public and private providers of health apps and usage intentions. Business & Information Systems Engineering, 66(3), 273–297. https://doi.org/10.1007/s12599-024-00869-4

Celestin, M., & Sujatha, S. (2024). Impact on brand trust and consumer behavior: Leveraging social media influencers to drive credibility and engagement across industries. International Journal of Applied and Advanced Scientific Research, 9(2), 81–87. https://doi.org/10.5281/zenodo.13887149

Chaudhuri, A., & Holbrook, M. B. (2001). The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty. Journal of Marketing, 65(2), 81–93. https://doi.org/10.1509/jmkg.65.2.81.18255

Chien, S.-Y., Lewis, M., Semnani-Azad, Z., Wang, W., & Sycara, K. (2016). Towards a model of cultural and cognitive influences on human–robot trust. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60(1), 228–232. SAGE Publications.

Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. International Journal of Human–Computer Interaction, 39(9), 1727–1739. https://doi.org/10.1080/10447318.2022.2050543

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L series, 2024/1689.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. Annual Review of Psychology, 59, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, N. J., Phua, J., Lim, J., & Jun, H. (2017). Disclosing Instagram influencer advertising: The effects of disclosure language on advertising recognition, attitudes, and behavioral intent. Journal of Interactive Advertising, 17(2), 138–149. https://doi.org/10.1080/15252019.2017.1366885

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association, 19(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human–robot interaction. Human Factors, 53(5), 517–527. https://doi.org/10.1177/0018720811417254

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. Psychiatry, 19(3), 215–229. https://doi.org/10.1080/00332747.1956.11023049

Influencer Marketing Hub. (2023). The state of influencer marketing: Benchmark report 2023. Influencer Marketing Hub. https://influencermarketinghub.com/ebooks/Influencer_Marketing_Benchmark_Report_2023.pdf

Khamitov, M., Rajavi, K., Huang, D.-W., & Hong, Y. (2024). Consumer trust: Meta-analysis of 50 years of empirical research. Journal of Consumer Research, 51(1), 7–18. https://doi.org/10.1093/jcr/ucad065

Khurana, T., Pannu, S., Dalal, G., Vyas, P., & Rani, P. (2025). How do social media influencers' credibility and brand trust drive purchase intentions for green cosmetics? NMIMS Management Review, 33(3), 171–185. https://doi.org/10.1177/09711023251352318

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. Nature, 435(7042), 673–676. https://doi.org/10.1038/nature03701

Lee, J. D., & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies, 40(1), 153–184. https://doi.org/10.1006/ijhc.1994.1007

Madsen, M., & Gregor, S. (2000). Measuring human–computer trust. Proceedings of the 11th Australasian Conference on Information Systems (ACIS 2000), Paper 53, Brisbane, Australia.

Najar, A. H., Wani, I. S., & Rather, A. H. (2024). Impact of social media influencers' credibility on destination brand trust and destination purchase intention: Extending the meaning transfer model. Global Business Review. https://doi.org/10.1177/09721509241225354

Nielsen. (2013). Global trust in advertising and brand messages. Nielsen Global Survey of Trust in Advertising. https://www.pnrc.net/wp-content/uploads/2014/02/Nielsen-Global-Trust-in-Advertising-Report-September-2013.pdf

OECD. (2019). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449). https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. Journal of Advertising, 19(3), 39–52. https://doi.org/10.1080/00913367.1990.10673191

Sokolova, K., & Kefi, H. (2020). Instagram and YouTube bloggers promote it, why should I buy? How credibility and parasocial interaction influence purchase intentions. Journal of Retailing and Consumer Services, 53, Article 101742. https://doi.org/10.1016/j.jretconser.2019.01.011

Turel, O., & Cui, T. (2026). Apologizing artificial intelligence: Designing and evaluating effective AI apologies after errors. AI & Society. Advance online publication. https://doi.org/10.1007/s00146-026-03067-w

Ulfstedt, V. (2025). Trust and authenticity in influencer marketing [Bachelor's thesis]. Theseus. https://www.theseus.fi/handle/10024/903287

United Nations Department of Economic and Social Affairs. (2025). The erosion of trust: A threat to social progress (Policy Brief No. 184). https://desapublications.un.org/policy-briefs/un-desa-policy-brief-no-184-erosion-trust-threat-social-progress

U.S. Securities and Exchange Commission, Division of Trading and Markets. (2012). Report to Congress on assigned credit ratings, as required by Section 939F of the Dodd-Frank Wall Street Reform and Consumer Protection Act. https://www.sec.gov/news/studies/2012/assigned-credit-ratings-study.pdf

Zak, P. J. (2008). The neurobiology of trust. Scientific American, 298(6), 88–95. https://doi.org/10.1038/scientificamerican0608-88

Zak, P. J. (2018). The neuroscience of high-trust organizations. Consulting Psychology Journal: Practice and Research, 70(1), 45–58. https://doi.org/10.1037/cpb0000076

Zak, P. J., Stanton, A. A., & Ahmadi, S. (2007). Oxytocin increases generosity in humans. PLoS ONE, 2(11), Article e1128. https://doi.org/10.1371/journal.pone.0001128

Zhou, X., Huang, Y., & Inoue, Y. (2024). Parasocial interactions and parasocial relationships on Instagram: An in-depth analysis of fashion and beauty influencers. Heliyon, 10(21), e39708. https://doi.org/10.1016/j.heliyon.2024.e39708

List of Figures and Tables

Figures

0.1 The Trust Transfer Model

0.2 Direct Evidence and the Outcomes

0.3 Mistrust Contagion

0.4 Vulnerability and Confidence

1.1 FTX, by the Numbers

1.2 The Trust Transfer Model, Applied to FTX

2.1 A recommendation as a worked example of the framework

4.1 Two Directions of Failure

5.1 Gains from Additional Advice

5.2 Three Conditions of Convergence

6.1 A Certification Bridge with a Time Condition: The Evusheld Authorisation

7.1 The Commercial Endorsement Model

8.1 Google's 2011 Ranking Change

9.1 Task Risk and Bridge-Strength

10.1 Direct Evidence and the Outcomes

10.2 Shift in public perception of Wells Fargo following the September 2016 fake-accounts scandal

10.3 The Bridge-Removal Test

11.1 Bridge Quality and Overreading

12.1 Mistrust Contagion

13.1 The Contain–Repair–Monitor Model

14.1 Revealing the Basis of a Cue

14.2 Calibrated Reliance

14.3 Seven design tests for a useful trust signal

15.1 Verifier Concentration Risk

16.1 A fillable checklist for naming a Bridge, checking what it certifies, watching for common tells, and deciding how much scrutiny a relationship deserves

16.2 Two Scorecards for a Bridge

17.1 Measures respondents said would increase their trust in AI
17.2 Pathways shaping initial confidence in an AI system

Tables

2.1 Bridge Forms
2.2 Core Terms and Diagnostic Questions
3.1 Referral Boundaries
3.2 The Weight of a Single Link
4.1 Public Signal Boundaries
5.1 Independence Checks
6.1 Temporal Checks
7.1 Commercial Signal Boundaries
8.1 Platform Signal Boundaries
9.1 Four Pathways of Trust Through the Surroundings
10.1 Outcomes of Target-Based Evidence and the Bridge-Removal Test
11.1 Common Forms of Borrowed Legitimacy
13.1 Repair by Failure Type
15.1 What Verification Marks Actually Certify
17.1 AI Trust Survey Sources

Acknowledgements

I acknowledge the Dharawal people, the Traditional Custodians of the land on which this book was written, and pay my respects to their Elders past, present and emerging.

This book carries more hands in it than the ones that typed it.

Anoushka, my daughter — your love has been unconditional in a way I notice more with every year, and it held me steady through every version of this.

Soja Madhavan, my mother — you sat with more half-formed frameworks than anyone should have to, asked the questions that mattered, and never let me settle for an easy answer.

Tomas Kubacki, my partner — you carried the cost of this book without complaint and carried me through the frustration of writing it.

To my father, G. Mohan Das, for being there, and to my brother, Dr Sidharth Mohandas — the two of you taught me, long before I had a framework for it, that trust is never something to take for granted.

And to Rani, Lara, and Kali, who offered their own quiet, four-legged kind of support throughout.

To Dr Samanda Bhadran and Mrs Geetha Bhadran, for kindly agreeing to read the final version and offering their generous feedback.

A note on AI use: parts of this book — drafting, editing, and structural passes — were developed with the assistance of AI tools. The framework, the arguments, and the conclusions are mine; the tools helped me say them more clearly and catch what I had missed along the way. Any errors that remain are my own.

We trust what we don't know.

No one can audit every balance sheet, inspect every supply chain, or test every system. So we lean on signals from elsewhere: a referral, a brand, a platform, a regulator.

This book shows what those signals actually establish, who stands behind them, and where their assurance ends.

VIVEKA MOHAN DAS

Mindful Machines Press

ISBN 978-1-7649186-1-9