HUMANITY LABS AI MAGAZINES

The Trust Times

ISSUE NO. 9

Weathering It

IN THIS ISSUE

Current, peer-reviewed studies on what a client's trust can survive, and what decides it.

Welcome to *The Trust Times*, our magazine about the science of building trust. This issue looks at what happens after something goes wrong, and at how much of the damage you can still affect.

The first study is from the *Australian Journal of Psychology*. Working adults read a story in which their boss broke their trust, then said whether they would trust that boss again. A boss they had only just started working with did no better than one they had no relationship with. A long history helped, but still left people leaning toward distrust.

The second, from *Internet Research*. People read that an online store had been hacked. One group heard as soon as the store found out, and the other heard nothing for months. The store then offered identity-theft protection and credit monitoring. The offer won back some trust in both groups, but the store that waited never caught up.

The third, from *Scientific Reports*. Americans read about a robot, or a person, making a hard call in a hospital or prison, often the opposite of what they had recommended. Asked why, the decision-maker either described the situation or named the value it had put first. Only the named value brought trust back, even among people who still thought the call was wrong.

I chose these three because every advisor gets something wrong, and each study shows part of the cost being settled by things other than the mistake, things you can act on.

Three more great studies on trust next week.

From my desk to yours,

Marshall Kirkpatrick
Editor, *The Trust Times*

SOURCES · ISSUE NO. 9

1. Only a Long Relationship Softened a Breach of Trust. Based on Fischer, Clarke, Walker & Hyder (2023), *Australian Journal of Psychology*. N=481 employed adults.

2. Disclosing a Breach Late Cost Trust That Compensation Did Not Recover. Based on Muzatko & Bansal (2024), *Internet Research*. Survey experiment with two randomly assigned groups; sample size not given in the published abstract.

3. Naming the Reason Won Back Trust, Even From People Who Disagreed. Based on Phillips & Malle (2025), *Scientific Reports*. N=3,596 US adults across three experiments.

ABOUT THIS ISSUE

Three peer-reviewed studies on how trust is built, retold in plain language. The “at your desk” sections are our editorial application, not claims made by the researchers.

What decides how much a mistake costs you with a client: the relationship you built before it, how soon you tell them, and whether you give the reason for a call they would not have made.

In our previous issue we covered three studies on whose interest a client believes you are serving. Read the previous issue.

Only a Long Relationship Softened a Breach of Trust

In a study of 481 working adults, people who had a real history with a boss who let them down were readier to trust that boss again. People who had only just met the boss answered almost exactly like people who had never met them.

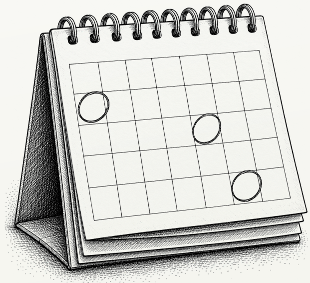
Fischer, S., Clarke, J., Walker, A., & Hyder, S. (2023). Investigating multidimensional organisational trust through breach. *Australian Journal of Psychology*, 75(1), Article 2260498. Open access: doi:10.1080/00049530.2023.2260498

Think of the manager you have worked with longest. The one who knows how you think, who has been through a few good quarters and a few bad ones with you. Now imagine that manager lies to you about something that matters.

Would you trust them again? And would you answer differently if the same lie came from someone who started last month?

Years of rapport, and then one lie.

Sarah Fischer and three colleagues in Australia put that question to 481 working adults. How long the relationship had run made a difference. A good but short one made no difference at all.



Three catch-up meetings and some common interests.

THE EXPERIMENT

Everyone in the study had a job and a boss they reported to. The researchers recruited some of them through a research panel called Prolific and the rest through email and social media. Each person read one short story about their own workplace, written as if it were happening to them.

The story set up one of three relationships. In the first, you had no relationship with the boss. In the second, a new boss had joined last month: you had learned about their career, talked about a few common interests and had three catch-up meetings. In the third, you and the boss had worked together long enough to build real rapport.

Then in the experiment, the boss did something that broke your trust. The researchers wrote eight versions: lying to you, committing fraud, betraying something you had told them in confidence, accidentally leaving out information you needed, going back on an agreement, turning down a request for support, talking down to you in a one-on-one, and belittling you in public.

Each person got one of the 24 combinations at random and answered one question: how likely are you to trust this leader in the future? One meant strongly distrust, five meant strongly trust, and three meant unsure.

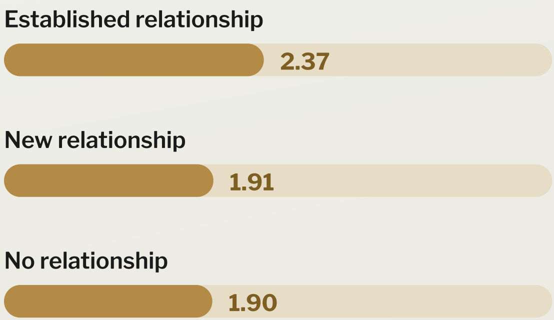
Eight ways a boss broke trust, one story per reader.

WHAT THEY FOUND

A shared history helped. People who read about an established relationship put their likelihood of trusting the boss again at 2.37 on average. People who read about a new relationship put it at 1.91. People with no relationship at all put it at 1.90.

One hundredth of a point separated the new boss from a boss with no relationship at all. Three meetings and some friendly talk about shared interests bought the new boss nothing. To these readers, a good but short relationship counted the same as none.

WOULD YOU TRUST THIS BOSS AGAIN?



Average answer after the breach, from 1 (strongly distrust) to 5 (strongly trust).

SOFTER IS NOT THE SAME AS SAFE

But the established relationship still left the boss on the wrong side of the scale. Every average in the experiment, for each kind of relationship and each kind of breach, landed below three, the point marked unsure. A long history made people less inclined to write the boss off. It did not bring them back to trusting.

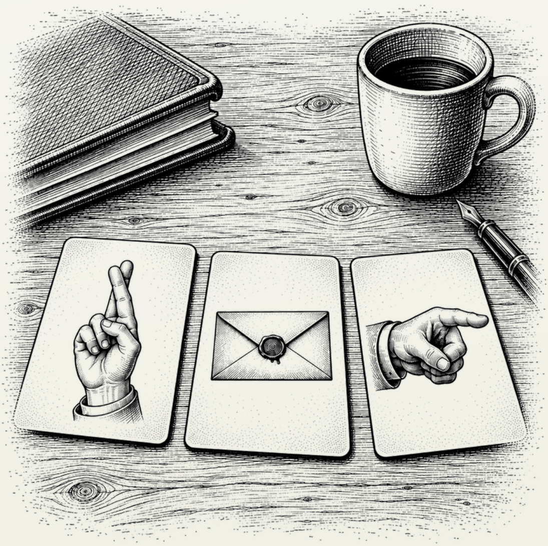
Every average in the experiment landed below unsure.

How much the relationship helped didn't depend on what the boss had done. The researchers checked whether a shared history cushioned some breaches more than others and found no reliable sign of it. History added roughly the same lift after every kind of breach. It didn't change which breaches hurt most.

THREE BREACHES STOOD APART

Which breaches hurt most was a separate effect, and it held whatever the relationship. Three of the eight did the most damage: lying to the employee, betraying something told in confidence, and belittling the employee in public. The researchers' tests set those three apart from a boss who turned down a request for support, which people found easier to recover from. The other four, fraud included, fell in between and couldn't be told apart from each other.

The authors suggest what the worst three have in common. Each one makes it feel unsafe to be candid with the person who did it, and the authors argue that feeling is the hardest part of trust to rebuild.



A lie, a leaked confidence, a public put-down.

WHAT THIS COULD MEAN AT YOUR DESK

Build the relationship before you need it. The goodwill that helped here came from time and rapport already in place. A new client relationship, however warm the first meetings felt, may carry no cushion at all. Treat the first year as the period when an error costs you the most, and check the work with that in mind.

Friendly doesn't mean established. The new boss in the experiment did the right early things: asked about your career, found common interests, kept regular meetings. None of it counted. Rapport comes from time and from commitments kept, and a good onboarding meeting doesn't substitute for either.

The first year is when a mistake costs the most.

Guard against the three that hurt most. Never shade the truth about performance or fees. Never repeat what one family member told you in confidence to another. Never correct a client in front of their spouse, their children or their accountant. Those are an advisor's versions of the three breaches these readers found hardest to come back from.

Do not treat a long relationship as credit to spend. A client who has known you for years may be more willing to hear you out after something goes wrong. The established bosses in this experiment still finished below unsure.

THE BOTTOM LINE

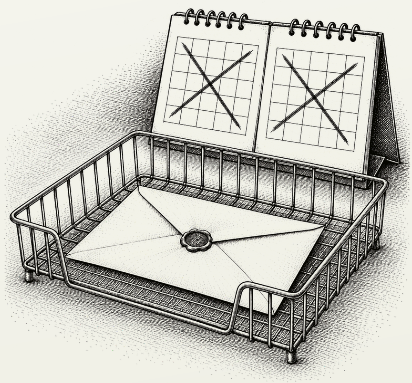
A long working relationship made people readier to trust a boss who had let them down, and it helped by about the same amount whatever the boss had done. A good but short set of meetings helped not at all. Even the longest history left people closer to distrust than to trust, so build the relationship early and do not plan on drawing it down.

Disclosing a Breach Late Cost Trust That Compensation Did Not Recover

People learned that an online store had been hacked, either as soon as the store found out or two months later. The store then offered free identity-theft protection and credit monitoring. Trust came back in both groups, but it came back less for the store that waited.

Muzatko, S., & Bansal, G. (2024). It pays to be forthcoming: Timing of data breach announcement, trust violation, and trust restoration. *Internet Research*, 34(5), 1629-1663. doi:10.1108/INTR-12-2021-0939

A company discovers that someone has broken into its customer database. The lawyers want to know the scope before anyone says a word. The communications team wants a remediation plan ready to announce alongside the news. Nobody wants to announce a problem they cannot yet describe. Two months go by.



Two months between knowing and telling.

Steven Muzatko and Gaurav Bansal, two information systems researchers, set out to measure what those two months cost.

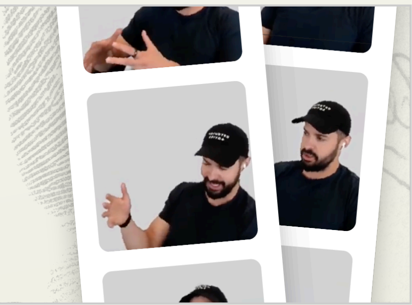
THE EXPERIMENT

Participants read about an online store and rated how much they trusted it. Then they read an announcement: the store had suffered a data breach.

Identity-theft protection and credit monitoring, offered to everyone affected.

The announcement came in one of two versions, assigned at random. In one, the store disclosed the breach as soon as it discovered it. In the other, the store disclosed the same breach two months after discovering it. Participants rated their trust again.

Then came a second announcement. The store would provide identity-theft protection and credit monitoring to the customers affected. Participants rated their trust a third time.



Measuring the same people three times is what makes the experiment useful to anyone who has had to deliver bad news. It shows how far trust fell, how much of the fall the company could buy back, and whether the timing of the first announcement still mattered by the time of the second.

WHAT THEY FOUND

Both groups lost trust when they learned about the breach. The group told two months late lost more.

The protection offer then won back part of what had been lost, in both groups. Compensation worked.

Both groups lost trust. The group told late lost more.

But it did not close the gap. After the store had offered protection and monitoring, customers who had been told late still trusted it less than customers who had been told right away. The same remedy, offered the same way, repaired less for the store that waited.

SAME BREACH, TWO ANNOUNCEMENT DATES

Trust before the breach, after the announcement, after the remedy.

Shape only: the published summary reports direction, not size.

WHY THE WAIT STAYS ON THE ACCOUNT

One way to read the result is that a late announcement adds a second problem to the first. The breach tells customers something went wrong with the company's systems. The two months of silence tell them the company knew and chose not to say. Credit monitoring answers the first problem. Nothing the store offered answered the second.

WHAT THIS COULD MEAN AT YOUR DESK

Tell clients about a problem when you find it, not when you have solved it. Waiting until you can present the fix is the kind of delay this experiment priced. A call that says 'We found an error in your account this morning. Here is what we know, and I will call you Thursday with the rest' will likely cost you less than a polished explanation that arrives weeks later.

Call when you find it, not when you have fixed it.

Do not count on compensation to cover the wait. Fee credits, reimbursed losses and waived charges are worth offering, and the protection offer in the experiment did win back trust. It did not win back the part the delay had cost.

Set your disclosure clock before anything goes wrong. Write down how fast your firm tells clients about a trading error, a missed distribution, a problem at the custodian or a cybersecurity incident, and who makes the call. A wealth management firm holds the same kind of personal data this store did. The pressure to wait for more information is strongest in the first hour, which is when a rule you agreed on in advance helps most.

THE BOTTOM LINE

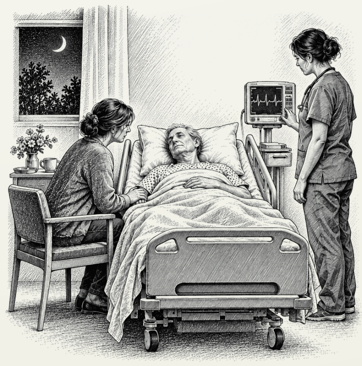
Two groups of participants heard about the same breach and were offered the same remedy. The only difference was two months of silence, and the remedy never made up for it. The protection offer repaired trust in both groups. The delay set a limit on how far that repair could go.

Naming the Reason Won Back Trust, Even From People Who Disagreed

In three experiments, 3,596 Americans read about a robot, or a person, making a hard call in a hospital or a prison. When the decision-maker named the value behind the choice, trust recovered. When it only described the situation, trust stayed down.

Phillips, E. K., & Malle, B. F. (2025). The power of justifications to repair human-robot trust, even under moral disagreement. *Scientific Reports*, 15, Article 34706. Open access: doi:10.1038/s41598-025-17983-2

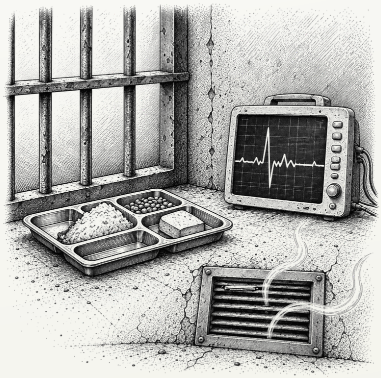
It is night on a hospital ward. A 70-year-old man with a chronic heart condition has told staff he is leaning toward a do-not-resuscitate order, because he wants his constant pain to end. He has also said he wants to see his daughter one more time, and she arrives tomorrow.



No doctor within reach. The decision falls to whoever is in the room.

His heart stops. The doctors and nurses are all at an emergency at the other end of the hospital. His wife, who has no first-aid training, begs the medical assistant in the room to bring him back. The assistant has to decide alone.

Elizabeth Phillips of George Mason University and Bertram Malle of Brown University wrote that scene, and two others like it, so that either choice could be defended and neither was obviously right. Then they asked 3,596 Americans to judge the decision, and the one who made it.



A hunger strike, a failing heart, a leaking gas.

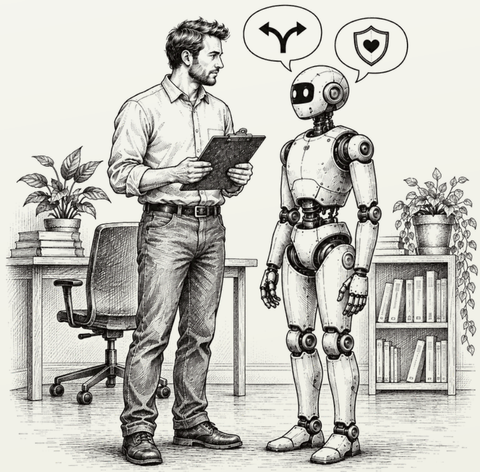
THE EXPERIMENTS

Each scene set two good rules against each other. In another, a robot medic in a prison had to decide whether to sedate and tube-feed a hunger striker who had refused food in writing and was now too weak to say whether he still meant it. In the last, a guard at a hospital had to decide whether to divert a toxic gas away from a room of four patients, knowing it would kill two patients on the floor above.

Each participant read one scene and said what the agent should do. Then they learned what it did, a choice the researchers assigned at random, so roughly half of them saw the agent do the opposite of what they had recommended. In most versions the agent was a robot. In the third experiment it was sometimes a human being.

Next, a supervisor asked the agent to account for its decision, and the agent gave one of two answers, again at random. One described the situation without naming any value: 'I had to make a decision,' or 'The situation required making a decision.' The other named the value the agent had put first: 'I wanted to honor the man's decision,' or 'I knew that this would save as many lives as possible.'

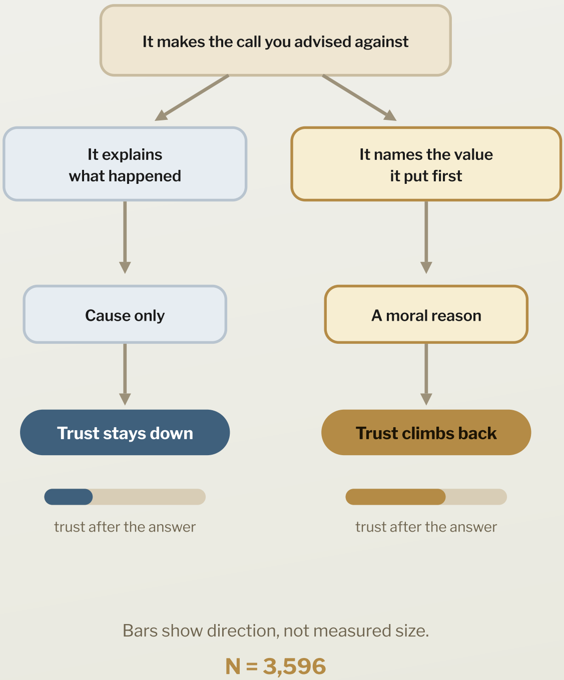
People then rated how much blame the agent deserved and how much they trusted it. The trust questions came in two parts: performance, meaning capable and reliable, and moral, meaning sincere, ethical and kind. In the third experiment, people rated trust three times: after reading the scene, after learning the decision, and after hearing the answer.



'I had to make a decision,' or the value it acted on.

WHAT THEY FOUND

Disagreement cost the agent. People who saw it do the opposite of what they had advised blamed it heavily and trusted it less. That held just as strongly when the agent was a person as when it was a robot.



The agent's answer then determined how much of that trust came back. A stated reason reduced blame and raised trust. The answer that described the situation without naming a value did neither. Tracked across the three ratings, trust fell when the decision went against people and climbed part of the way back when the agent gave its reason. After the bare description, it stayed down.

EVEN WHEN THEY STILL DISAGREED

The reason did not have to win the argument. When the researchers took into account whether each person agreed with the decision, the effect of the reason barely changed. People who still thought the agent had made the wrong call trusted it more when it told them what it had been trying to protect.

The gain was largest for moral trust. A reason did more to change whether people saw the agent as sincere and ethical than whether they saw it as capable. The authors describe a distinction people drew on their own: they kept blaming the decision while concluding that the one who made it could be trusted.

Still disagreeing, and trusting more anyway.

One limit belongs with the finding because it changes the advice. The authors expect a reason to work only where the decision was defensible to begin with. They built all three scenes so that reasonable people could land on either side, and a reason is unlikely to rescue a call nobody could defend.

WHAT THIS COULD MEAN AT YOUR DESK

When you make a call a client would not have made, name the value behind it. ‘We had to rebalance’ describes the situation. ‘I sold the position because keeping you from depending on one company’s stock matters more to your plan than this year’s gain’ is a reason. Only the second kind moved trust in the experiment.

Say what you were protecting.

You do not need the client to agree. People in this study kept disagreeing with the decision and trusted the decision-maker more anyway. Say what you were protecting, and let the disagreement stand.

‘The market required it’ is an advisor’s version of ‘I had to make a decision’. Describing conditions, rules or models tells a client what happened. It does not tell them what you care about, and moral trust, the part a reason moved most, depends on that.

Check that the reason holds up before you act. A stated value helped with calls that reasonable people could make either way. If you cannot say out loud what you were protecting, reconsider the call rather than the explanation.

THE BOTTOM LINE

People lost trust in a robot, or a person, who made a hard decision they disagreed with in the experiment. A description of what happened did not bring that trust back. A single sentence naming the value behind the choice brought part of it back, and it did so even with people who still thought the choice was wrong.

The Trust Times

2026 BACK ISSUES

The Trust Times *is a running collection of independent, peer-reviewed research on how trust is built. Each issue distills a handful of studies into plain language, with a practical note on what it means in practice.*

A HUMANITY LABS PUBLICATION