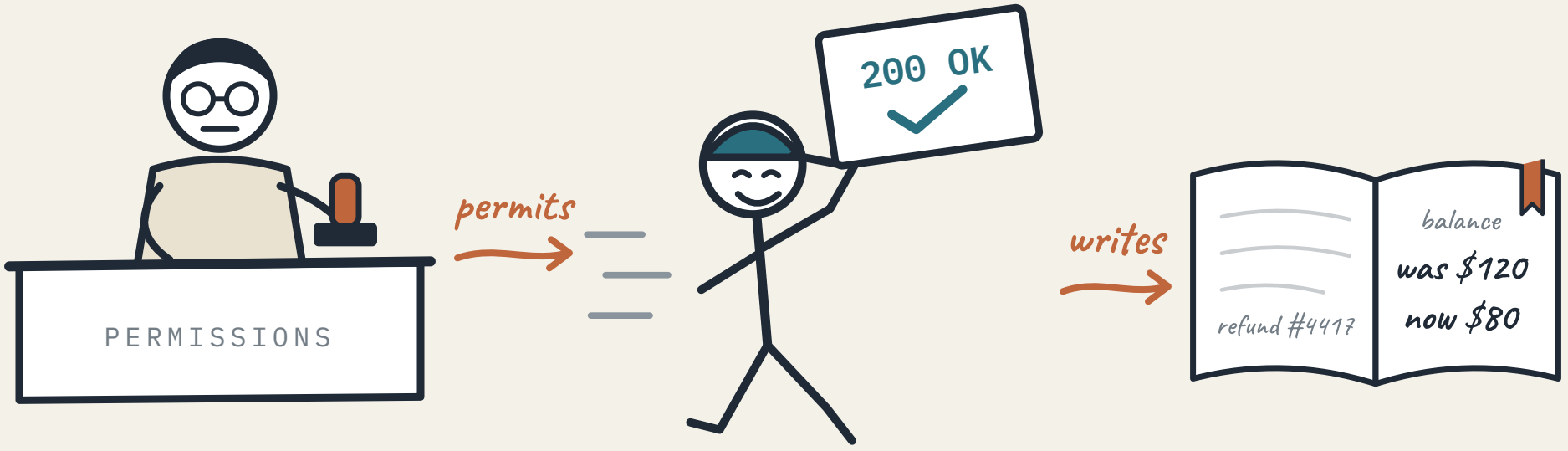
A STORYBOARD IN 9 FRAMES

From agent intent to governed execution

the model can only ask.

the model filled it in.

someone else signs



THE GATE

THE TOOL

THE RECORD

A model can ask for an action. Something else decides whether that ask becomes a real effect, and then whether it actually worked.

Four things happen before money moves.

1. Check the request

a real operation? right types? things that exist?

2. Decide if it's permitted

for this caller, in this context, right now

3. Run it

with whatever limits apply

4. Check what happened

not whether it returned. Whether it's true.

A sequence, not an architecture. Small systems collapse it into a few lines of code; it still helps to know which line is doing which.

Each step fails in its own way.

Valid, but not allowed

the gate said no. Rightly, or not.

Allowed, but it didn't run

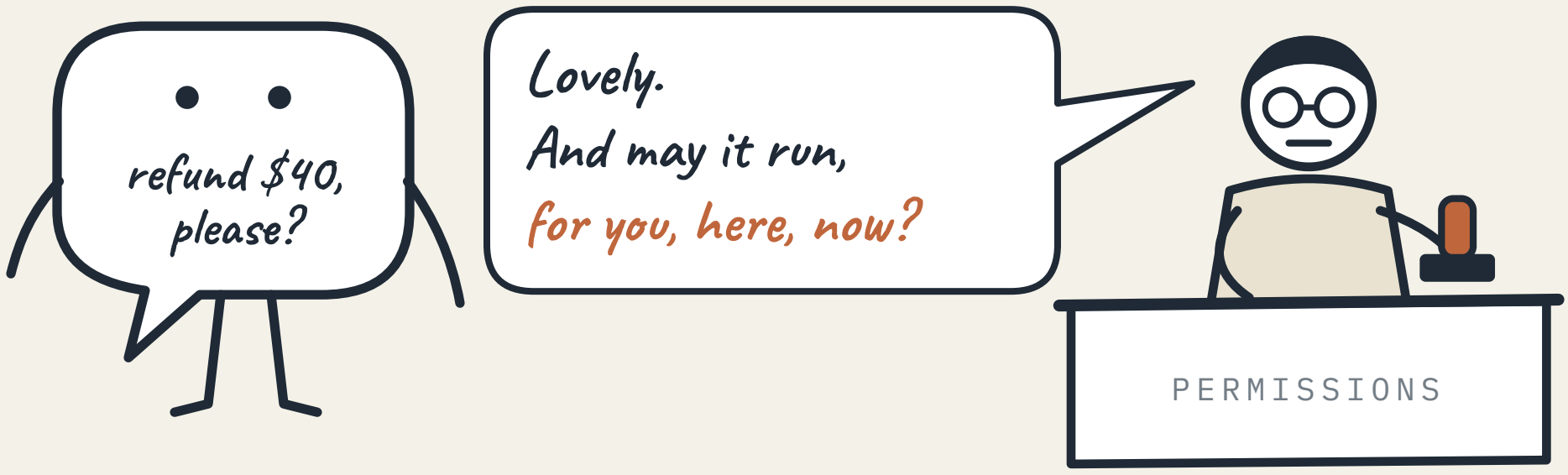
the tool errored, or timed out

It ran. Nothing changed.

the most confident failure of all

A request can be valid but not allowed, allowed but fail to run, and run successfully without producing the effect anyone wanted.

Four things that are not permission.



~~Logged in~~

tells me who.
Not what.

~~Well-formed~~

So is "delete production."

97%

~~"97% sure"~~

That's a mood, not a permit.

Dave: "fine!"

A human said yes

A way to answer.
Not the question.

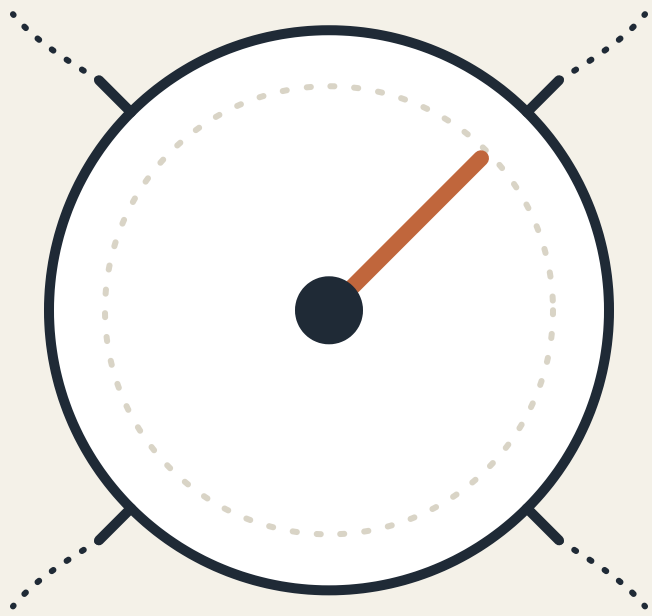
A rule in a prompt is a request. A rule at the gate is a check.

“Not sure” isn't one answer.

yes/no only: block the useful, or wave the rest through

ALLOW
permitted, and the evidence holds
leaves: a receipt

ASK
a person should decide
leaves: a question, and why



DENY
not permitted, or too risky
leaves: a reason code

DEFER
thin evidence, high stakes
leaves: a handoff note

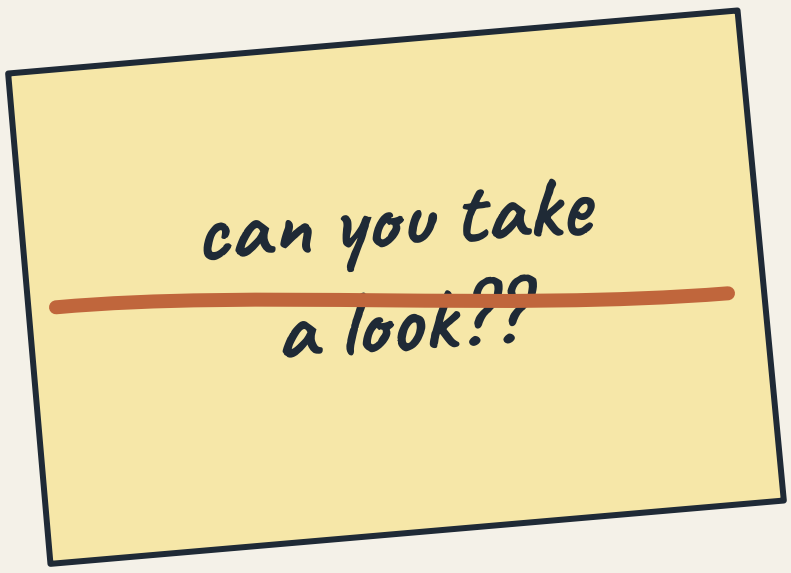
what moves the needle most: how hard is it to undo?

the bar for ALLOW rises →

undo: one click

undo: an apology email

“Ask a human” is a handoff, not a shrug.



at what, exactly?

REFUND $40 · ACCOUNT 88

Over the $25 auto-limit.
Order #4417 marked delivered.

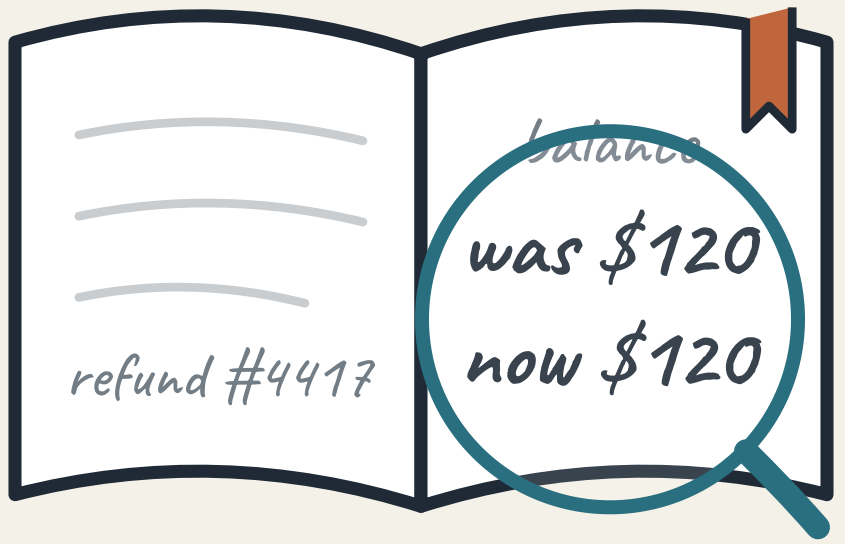
Approve or decline?

APPROVE DECLINE

a specific question, with the reason

Who reviews it, what they see, and whether they are positioned to say no is its own design problem: human judgment, the next stage of the map.

The tool says done. The record disagrees.

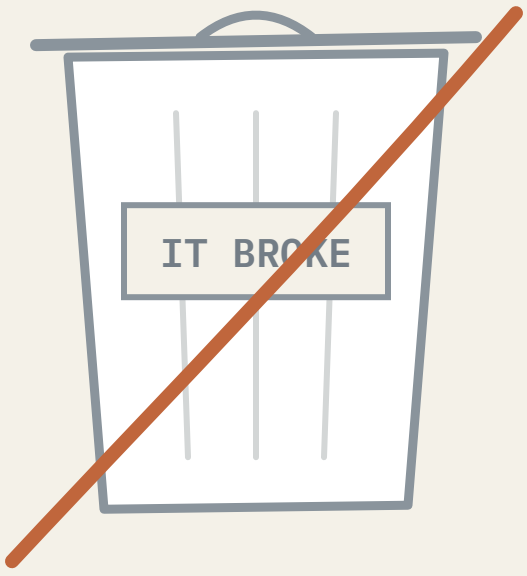


...is it, though?

~~reports done~~ → **shows done**

A 200 tells you what the interface said. If what you wanted was a change of state, go and read the state.

“It broke” is not a diagnosis.



“...and then the agent just...”

incident review, as storytelling

Policy denied it
was the rule right?

The tool errored
fix it, or retry it

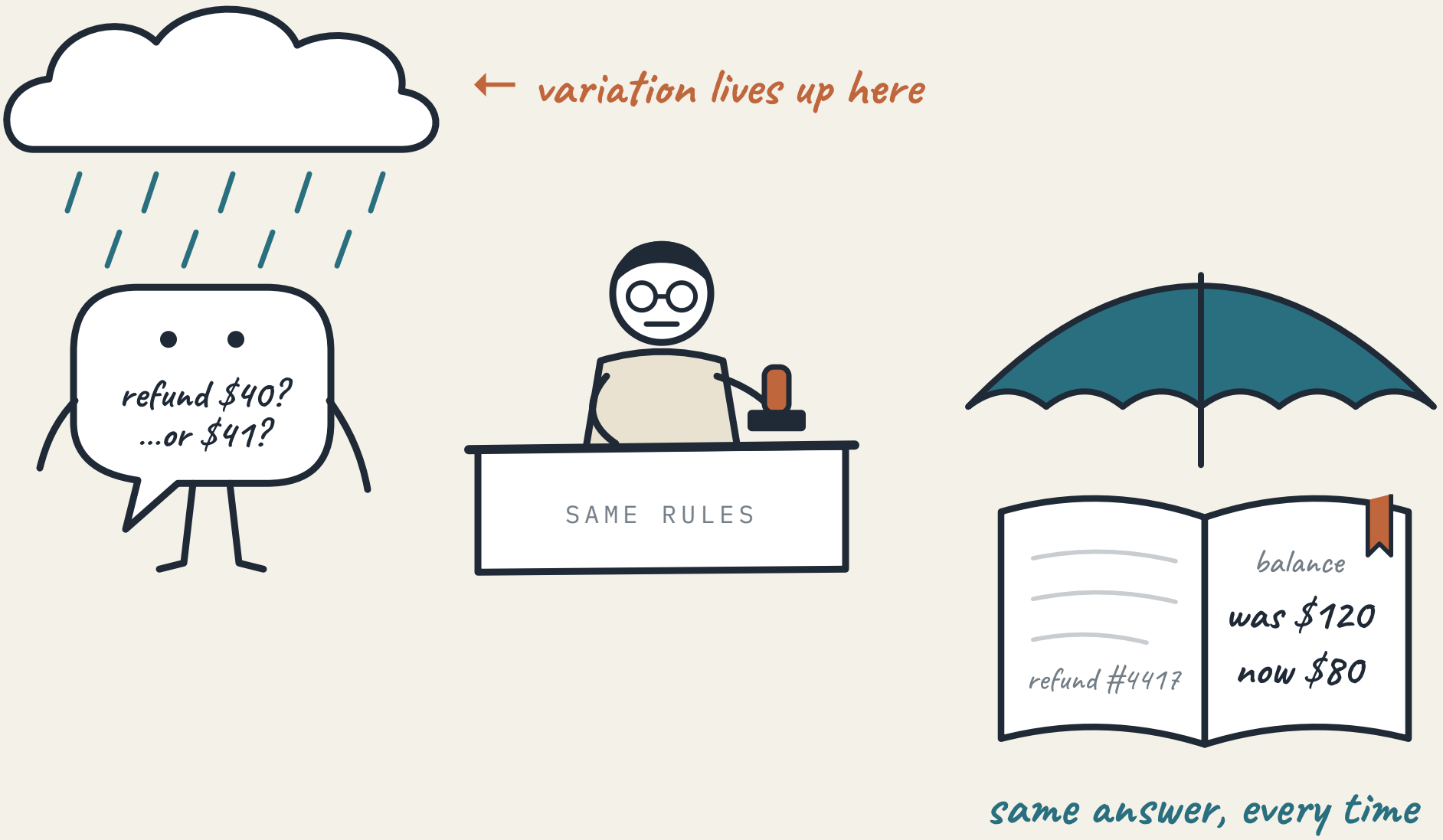
Verification failed
the tool fibbed, or you checked the wrong thing

Evidence was missing
go and get it

Over time or budget
raise the limit, or don't

Each tray points somewhere different. You can't sort a failure at a boundary you never made explicit.

The model still varies. Your effects needn't.



The checks don't make the model deterministic. They give you a dependable answer about one candidate action: permitted or not, every time.

Read the whole explanation

pruningmypothos.com/systems/
from-agent-intent-to-governed-execution/ →