Mapping Deliberation · No. 1

I went looking for the recipe for good deliberation. I found *seven rival ones.*

And one uncomfortable possibility that every one of them has to beat.

Michael Levinson

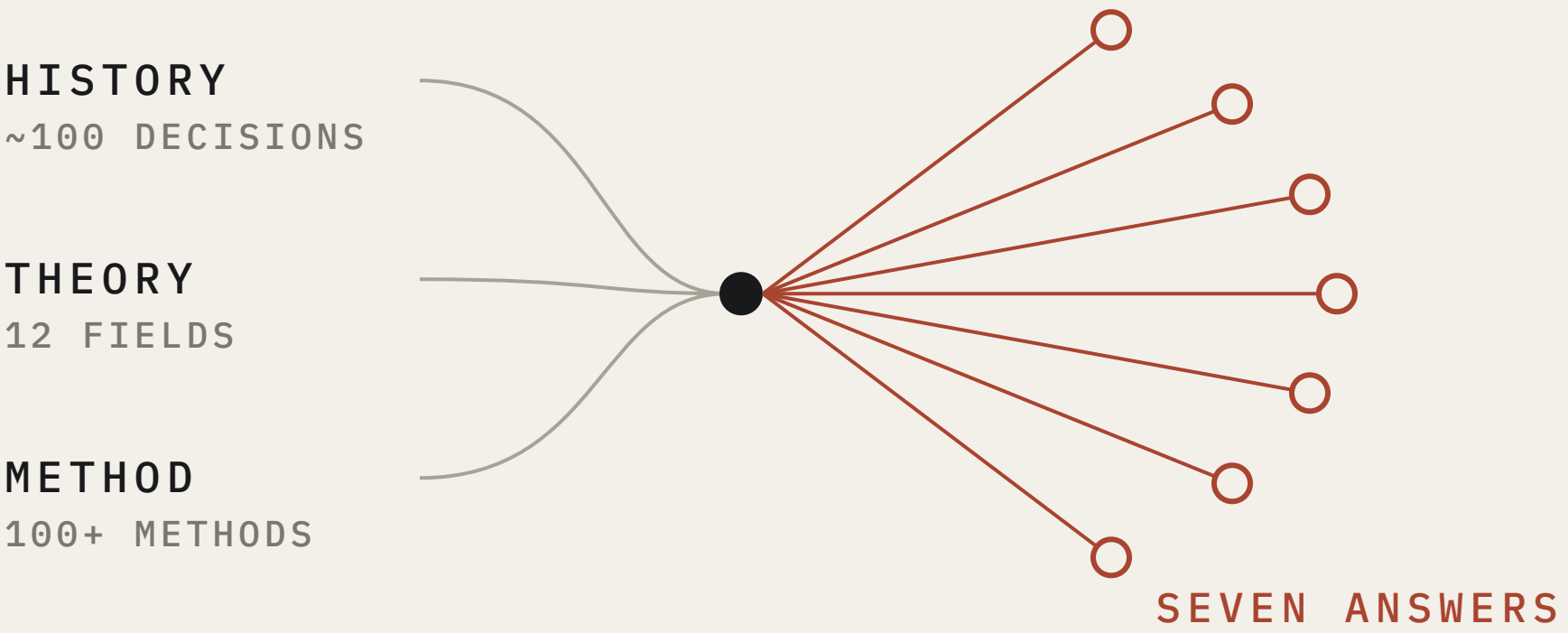
COMPUTATIONAL DELIBERATION LAB

Swipe →

Where this started

In August I wrote that we know how to think better.

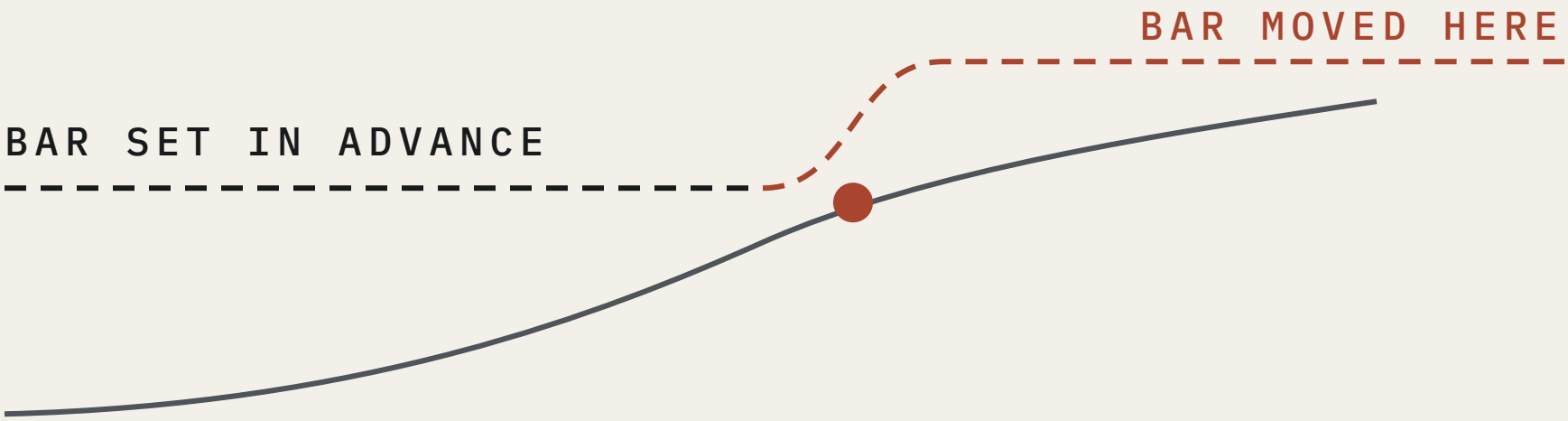
It's just too expensive.



Then I tried to write down what “better” means, working from what AI models recall about history, theory and practice, kept blind to my hypotheses.

The sources didn't agree on what better means.

Decisions break when the bar moves.



The bet: quality lives in what you fixed before anyone was invested.

SEEN IN Everest, 1996. Rob Hall's team kept climbing past its pre-set turnaround time.

WRONG IF

errors are no more common after a mid-course change of criteria.

Ordinary judges, well arranged, may beat better ones.



SEEN EACH OTHER

LOCKED, THEN COMBINED

The bet: lock each judgment before anyone sees the others, then combine by a rule set in advance.

SEEN IN N-rays, 1903-04. Dozens of researchers saw a ray that didn't exist, sharing one expectation.

WRONG IF

the strongest single judge, with the same budget, does just as well.

The right question beats the general one.

BROAD PROMPT

"What could go wrong?"

→ A LONG, VAGUE LIST

TRIGGERED QUESTION

"Which units does the other team use?"

→ THE FLAW THAT KILLED IT

The bet: specific questions, triggered by specific conditions, catch what a broad review misses.

SEEN IN Mars Climate Orbiter, 1999. One team's software used pound-force seconds; navigation assumed newtons.

WRONG IF

a generic prompt catches as much, or the specific questions fire just as often on sound plans.

Pick the better judge, and process matters less.

SCORED TRACK RECORD



The bet: where a scored record exists, choosing who deliberates can beat any change of process.

SEEN IN Forecasting tournaments, where scored records show who is reliably right.

WRONG IF

a weaker participant, given the same time, gets close to the stronger one.

The warning was in the room. It couldn't stop anything.



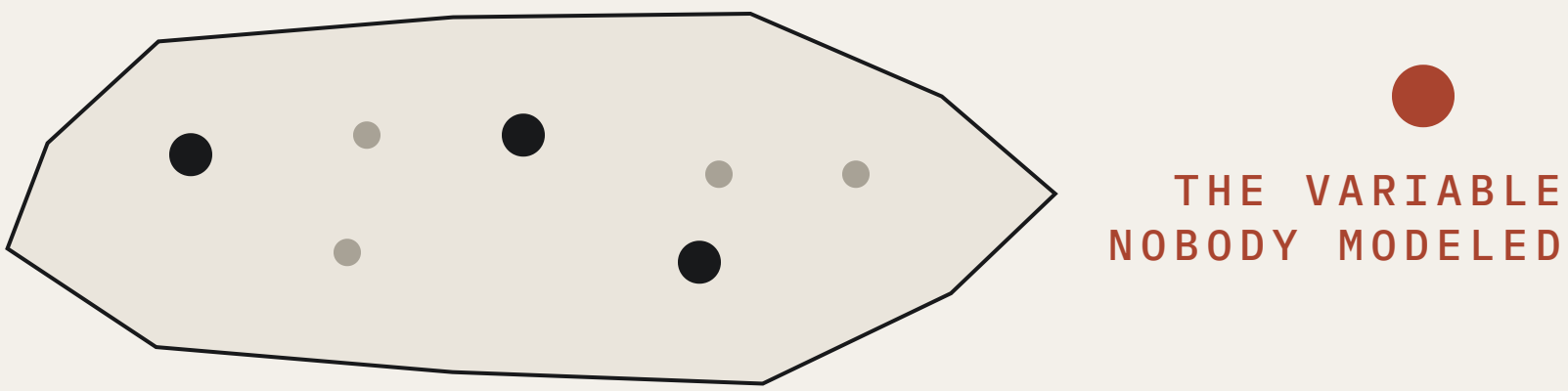
The bet: failures concentrate where a correct signal was present and had no power to govern.

SEEN IN Hubble, 1990. Tests that disagreed with the main instrument were set aside; the mirror flew flawed.

WRONG IF

most final errors turn out to be ones no stage ever found.

Ask only what could change the act.



KEPT: ONLY WHAT COULD CHANGE IT

The bet: screening inquiry by decision-relevance saves effort, and fails silently outside the model.

SEEN IN Shoreham. A finished nuclear plant that never ran commercially; evacuation planning decided it.

WRONG IF

screening saves nothing on ordinary decisions, or fails no more often on unusual ones.

Count the tests that could have failed.

TESTS PASSED



MANY

COULD HAVE FAILED



FEW

The bet: how hard a plan was put at risk may predict outcomes better than how many checks it passed.

SEEN IN Cold fusion, 1989. Announced at a press conference before the paper appeared or anyone could replicate it.

WRONG IF

passed-check counts predict outcomes just as well.

What this is not

None of these seven has been tested yet.

The material is what one family of AI models recalls, not read from sources

Every historical case was chosen because it ended badly

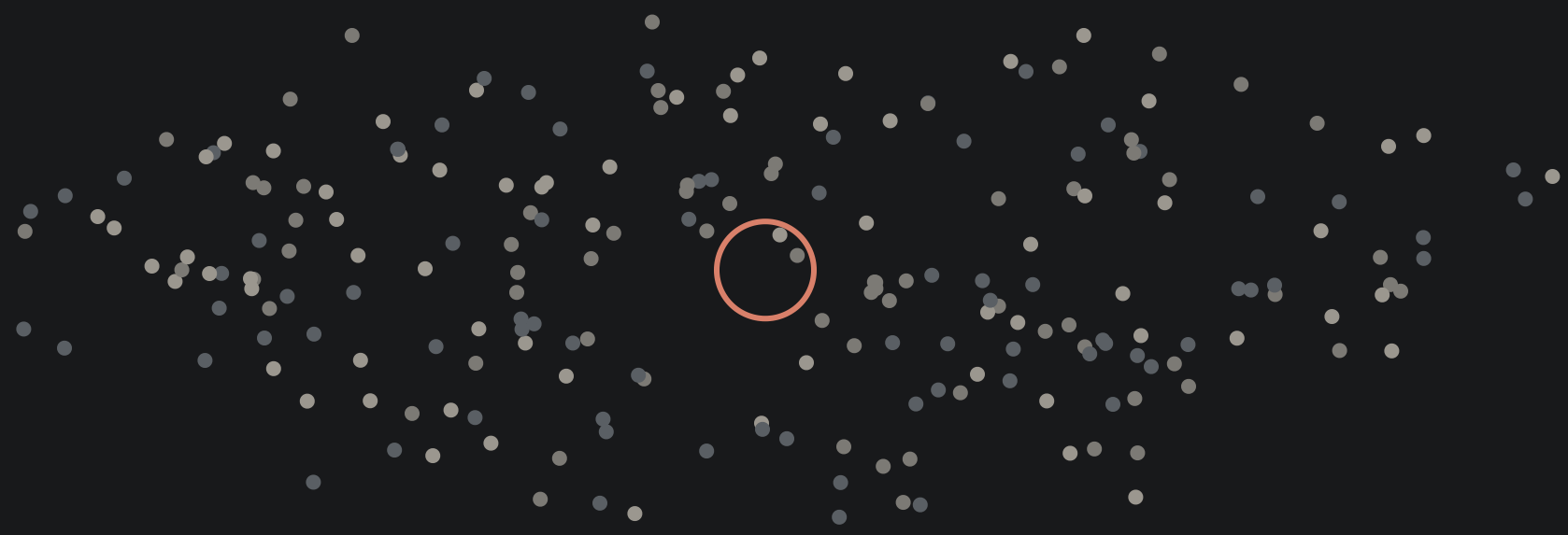
The evidence streams share literature, so agreement counts for less

No outcome has been scored against any of them

This is a map of bets, each with a stated way to be wrong.

The baseline every theory has to beat

Maybe process doesn't decide outcomes at all.



PROCESS, INSIDE EVERYTHING ELSE

Luck, information and the world may swamp whatever a group does in the room. I can't rule that out yet.

Any honest theory of good deliberation has to show it does better than this.

Putting it together

Seven places quality might live.

01 What you fixed in advance

02 How judgments are kept independent

03 What gets asked

04 Who deliberates

05 Who can stop it

06 Asking only what could change the act

07 How hard the plan was put at risk

They collide, and averaging them hides the one variable that would decide between them.

An open question

If experts can't agree what makes deliberation good, *what are we training machines to do?*

These are the theories I could map. Some will die when tested, and others are still missing.

Which one would you bet on?

Full method: computational-deliberation.com