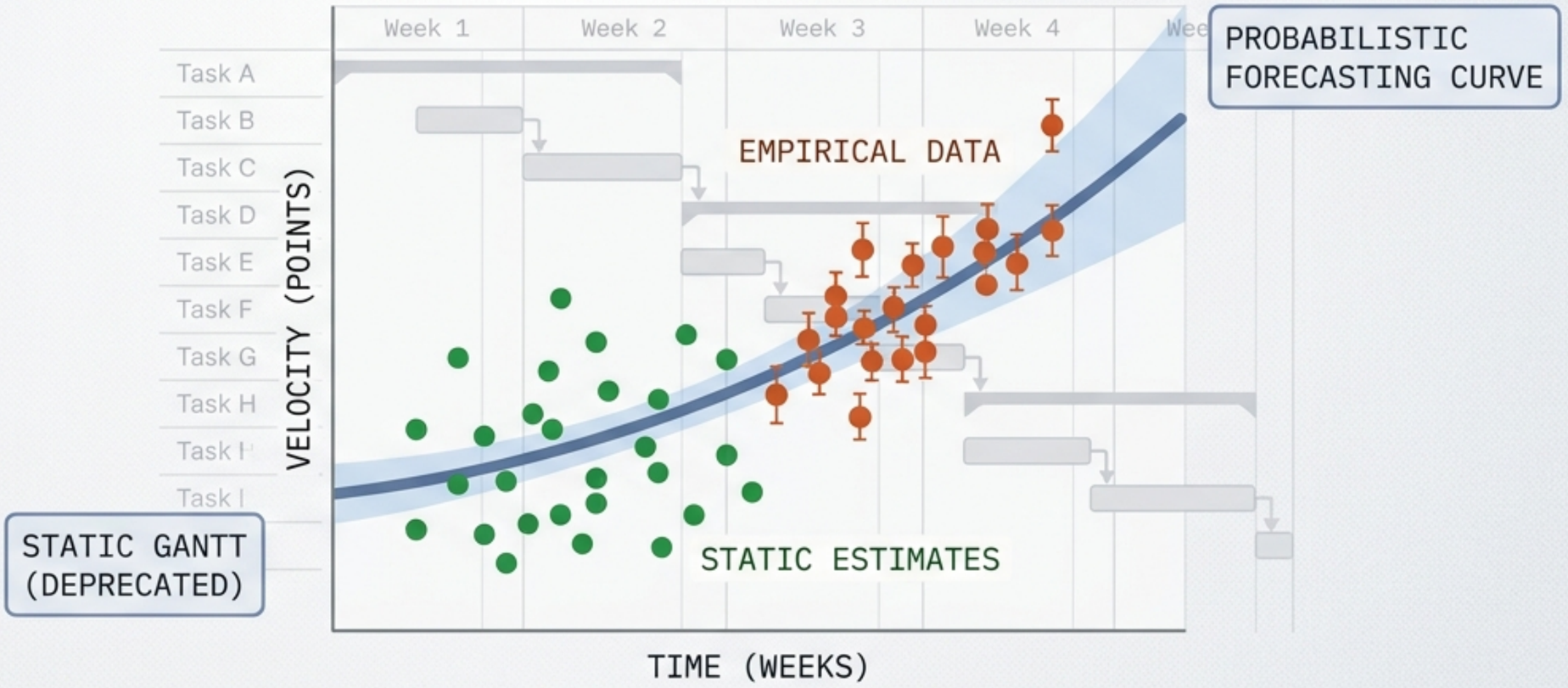
Ways of Working — S07E03: Planning, Estimation, Flow Metrics & Probabilistic Forecasting

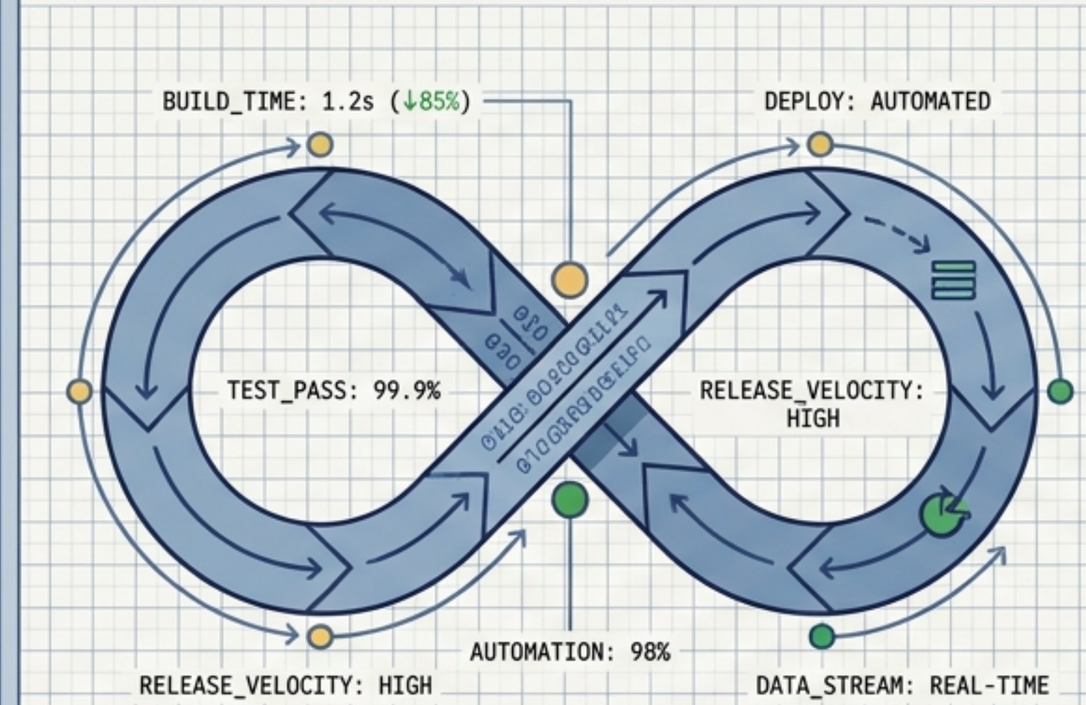
Separating empirical evidence from practitioner opinion.



Gemini Notebook

S07E02 Recap:

We optimized our CI/CD pipelines for speed, but our planning process remains anchored in guesswork.



"Building software is nothing like building houses." — Allen Holub, #NoEstimates

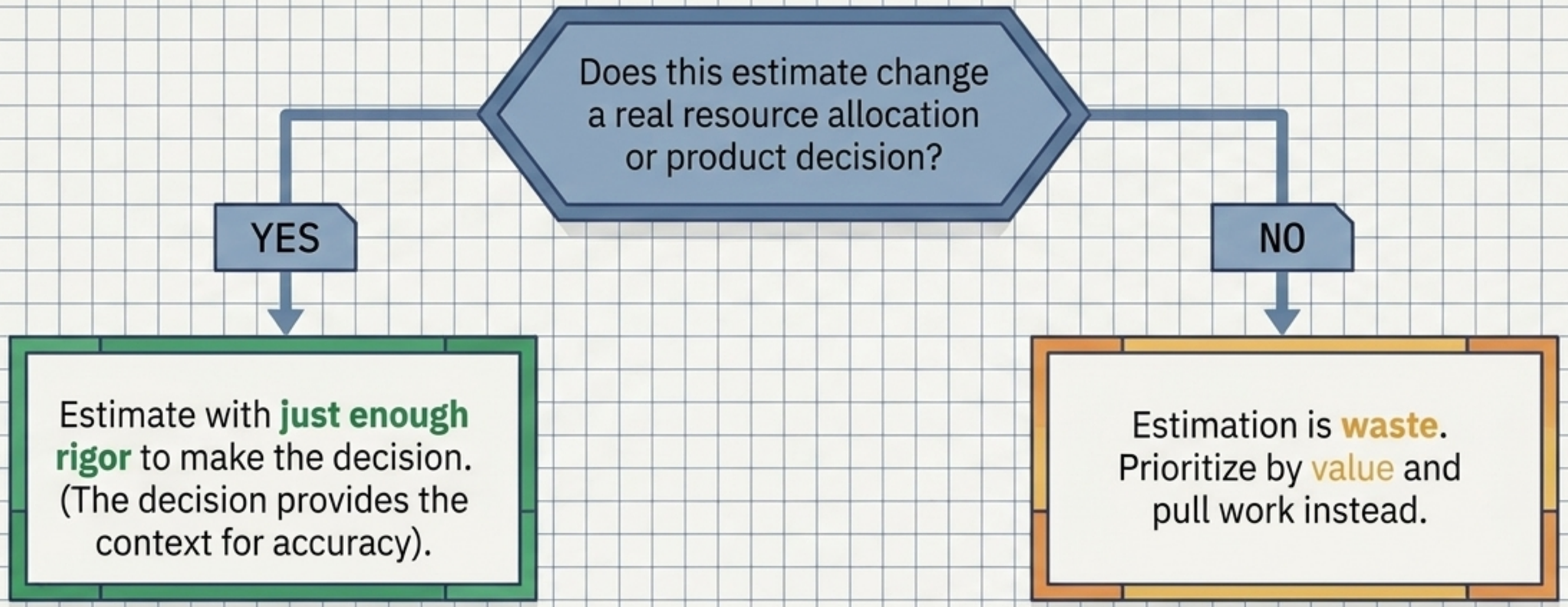
The Estimation Paradox



→ We cannot use deterministic planning to manage highly stochastic discovery work.

Gemini Notebook

Martin Fowler’s Purpose of Estimation Test



If it’s worth doing well, it’s worth asking why on earth you’re doing it at all.

Gemini Notebook

The Danger of Deadline Hardening

STAGE 1 (LOW FIDELITY):
Developers “throw” a rapid consensus estimate (e.g., 3 fingers) to gauge relative size.





STAGE 2 (TRANSLATION):
Project managers translate abstract sizing into a calendar date.

STAGE 3 (LOSS AVERSION):
Stakeholders anchor to the date. Any deviation is perceived as a critical loss.



STAGE 4 (THE TRAP):
The baseline forecast crystallizes into a quality-crushing, fixed-scope blood oath.



Gemini Notebook

Story Points: Original Intent vs. Corporate Misuse

Original XP Intent



- **Yesterday's Weather:** A capacity planning tool to see what fits in an iteration.

CAPACITY: PREDICTABLE FLOW



- **Obfuscation:** Intentionally abstract to prevent managers from pressuring the team over time estimates.

ESTIMATES: ABSTRACT UNITS



- **Focus:** Pacing and rhythm.

FOCUS: TEAM RHYTHM

Modern Corporate Misuse



- **The KPI Trap:** Used as a performance metric for individual developers.

METRIC: INDIVIDUAL PERFORMANCE



- **Translation:** Points are directly equated to developer-hours (e.g., 1 pt = 4 hours).

CONVERSION: 1 PT = 4 HRS (WARNING)



- **Focus:** Earning points and increasing arbitrary 'velocity'.

FOCUS: INCREASE "VELOCITY" (RISK)

Gemini Notebook

The Velocity Trap & Goodhart's Law

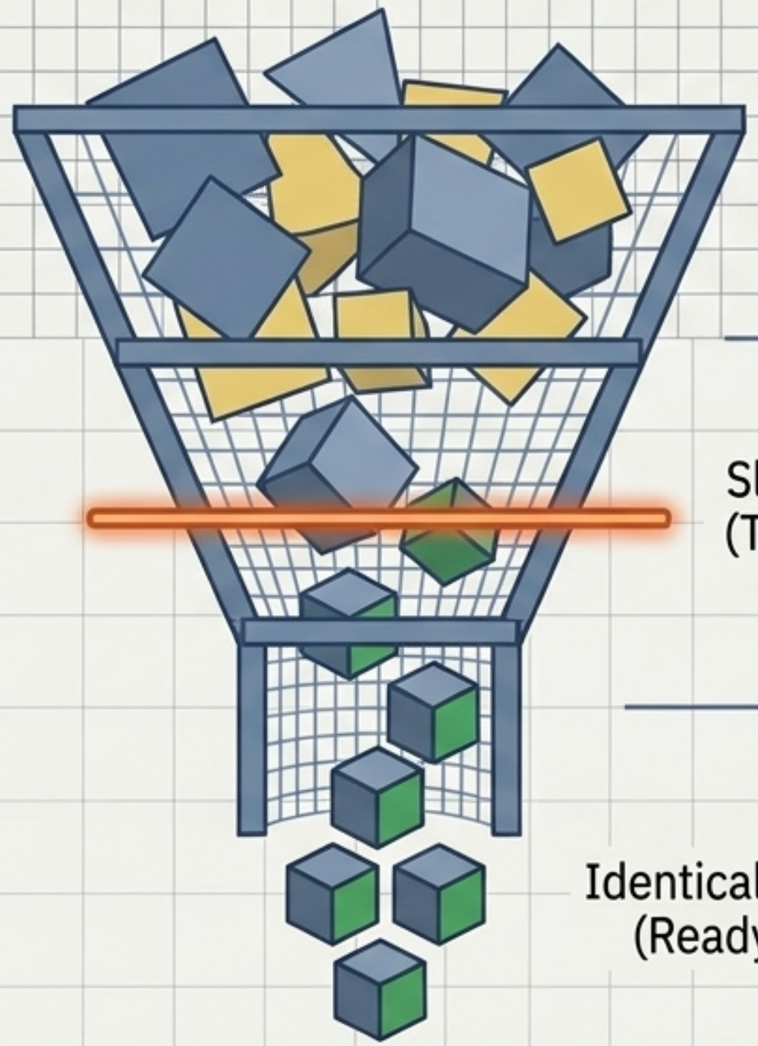
> I may have invented story points, and if I did, I'm sorry now. – Ron Jeffries

1. Management demands higher **Velocity** (Points/Sprint).
2. Teams **subconsciously** (or consciously) **inflate point estimates** for the same amount of work.
3. Velocity numerically **increases**.
4. The metric **loses all informational value** for actual capacity planning.

⚠ **Takeaway:** Cross-team velocity comparison is mathematical nonsense.

Gemini Notebook

Slicing, Counting, and #NoEstimates



Bulky, Irregular Stories

Slicing Indicator (Testable Slices)

Identical, Uniform Cubes (Ready for Counting)

Alternative 1 (Bill Wake's INVEST):

Focus obsessively on "**Small**" and "**Testable.**" Slice work down until it requires just a single acceptance test.

Alternative 2 (Fowler's Story Counting):

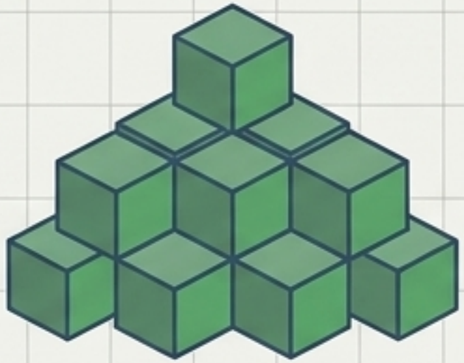
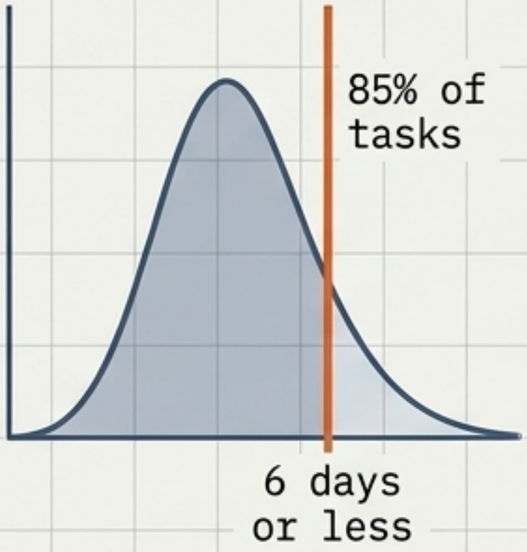
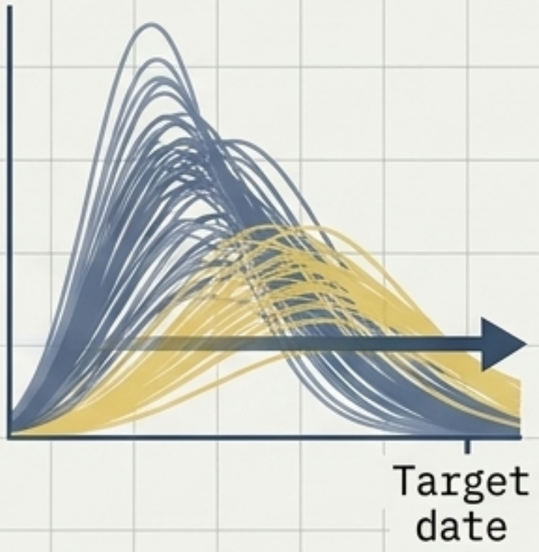
Ignore relative size completely. Count the number of stories. Over time, the Law of Large Numbers ensures bigger and smaller stories cancel each other out.

Alternative 3 (Allen Holub's #NoEstimates):

Ditch the guesswork entirely. Use projections based on real historical throughput to dynamically adjust scope.

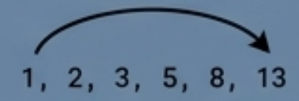
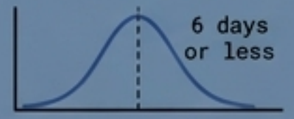
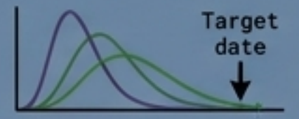
Gemini Notebook

The Estimation & Forecasting Taxonomy

Fibonacci Story Points	Right-Sized Story Counting	Historical Cycle-Time SLEs	Monte Carlo Simulation
			
Guessing relative size using arbitrary sequence numbers.	Slicing work into small, roughly equal batches and counting volume.	Service Level Expectations (“85% of our tasks finish in in 6 days or less”) based on past flow.	Running 10,000+ randomized mathematical models against historical throughput to generate probability curves.

Gemini Notebook

Evaluating the Methodologies

	Fibonacci Story Points 	Right-Sized Story Counting 	Historical Cycle-Time SLEs 	Monte Carlo Simulation 
Meeting Overhead (Time wasted arguing)	High	Medium	Low	Low
Gaming Vulnerability (Goodhart's Law)	High	Medium	Low	Low
Input Requirements (Data needed)	Low	Medium	High	High
Stakeholder Output Format	Deterministic Date	Rough Range	Probabilistic	Confidence Interval

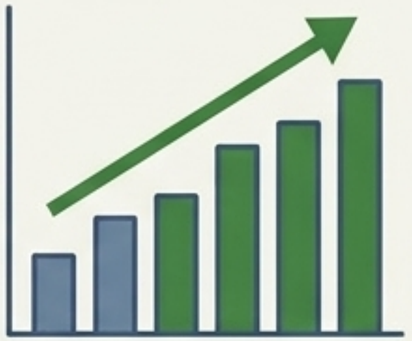
Gemini Notebook

The Four Core Flow Metrics



1. WIP (Work in Progress)

The number of work items currently active. (The ultimate constraint on speed).



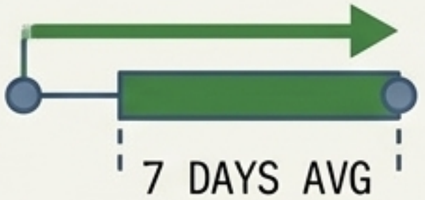
2. Throughput

The number of items completed per unit of time. (Measures output, not effort).



3. Work Item Age

How long active items have been in progress. (The vital *leading* indicator of cycle time).



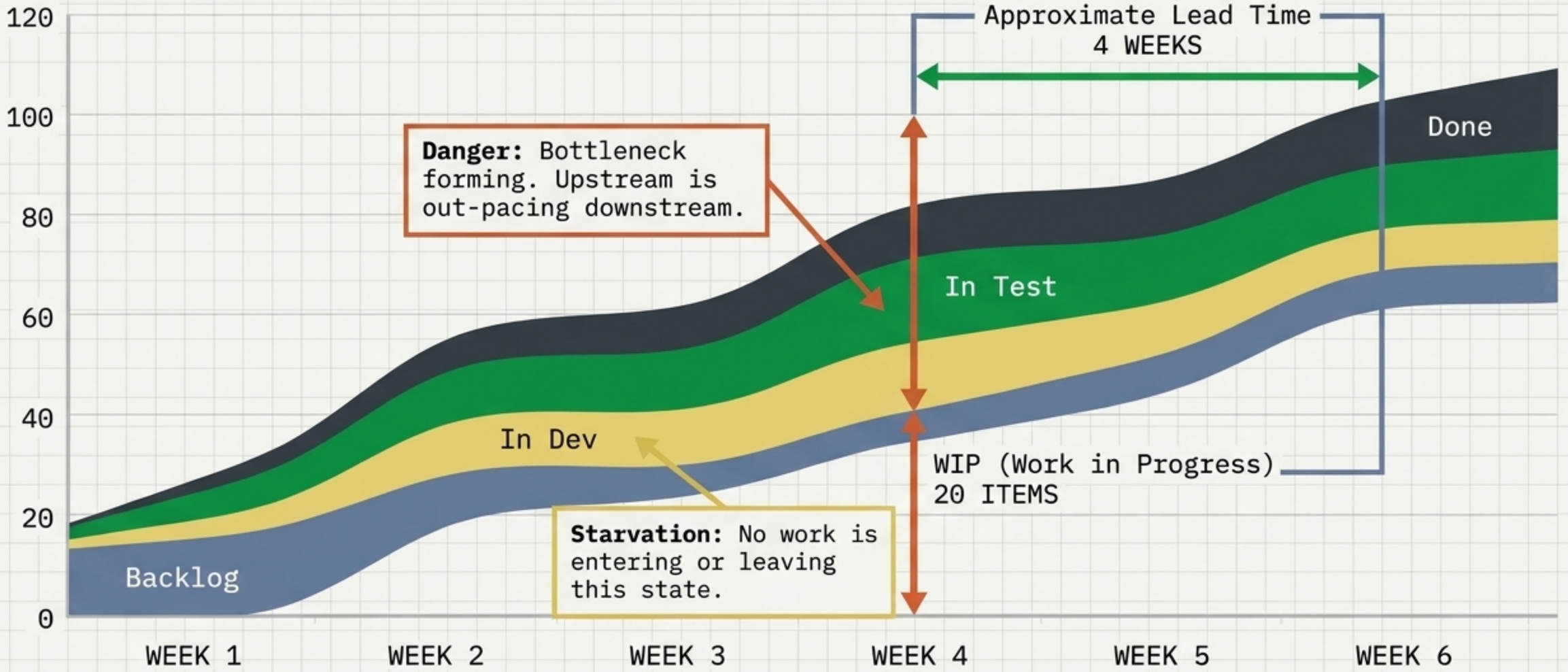
4. Cycle Time

How long it took to complete an item from start to finish.

(The *lagging* indicator of delivery capability).

Gemini Notebook

Reading the Cumulative Flow Diagram



Gemini Notebook

Managing Uncertainty with **Hill Charts**

The Concept: 50% done means nothing.
Instead, **track work based on unknowns.**

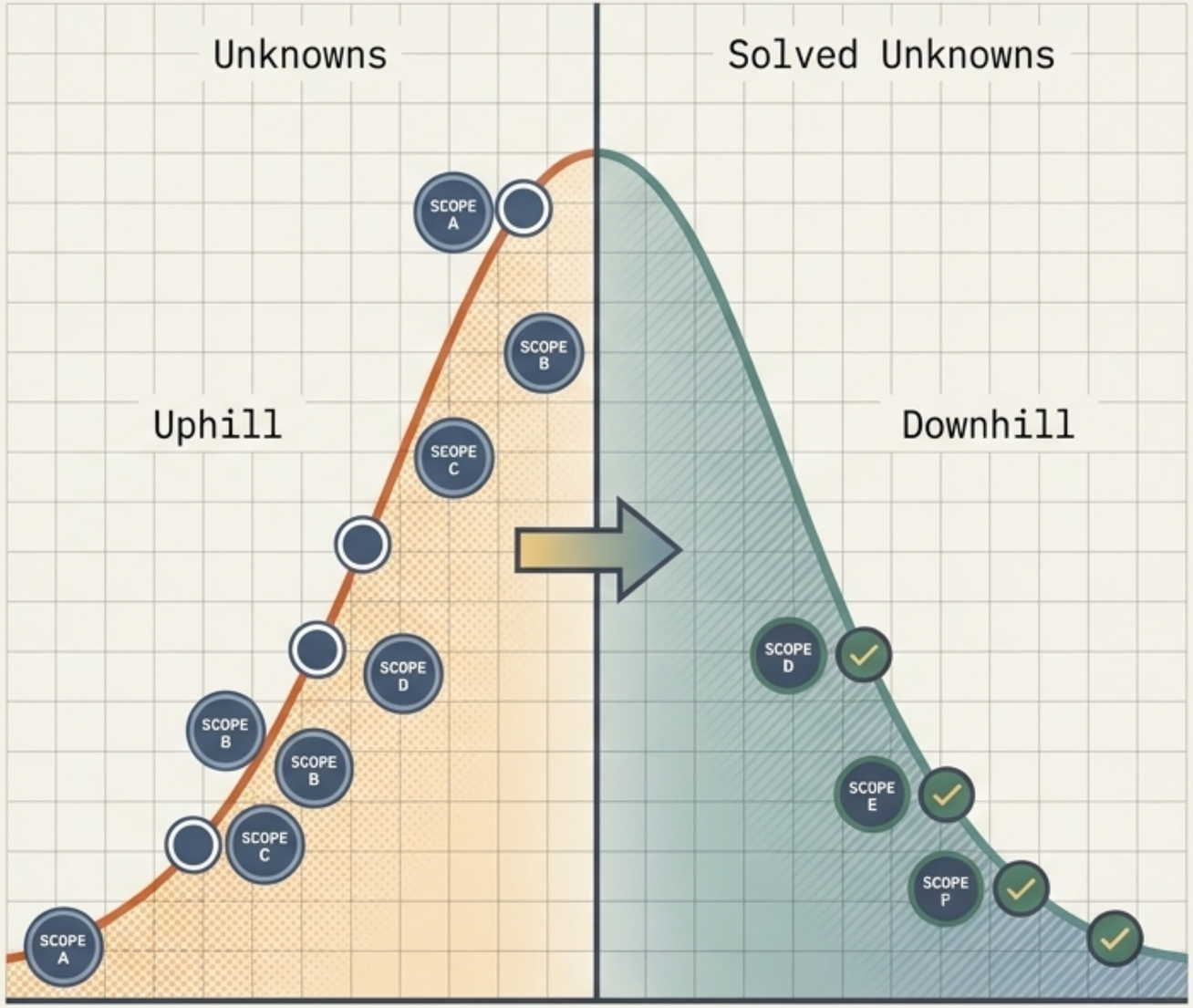
Uphill (Figuring out what to do):

High uncertainty. “Imagined tasks” vs real tasks. You are discovering the discovering the problem space.

Downhill (Execution):

Certainty. All unknowns are solved. Only execution remains.

Takeaway: **Never** report a deterministic status while work is still on the **uphill slope** of discovery.



Gemini Notebook

Probabilistic Forecasting: The Monte Carlo Engine

The Mechanics

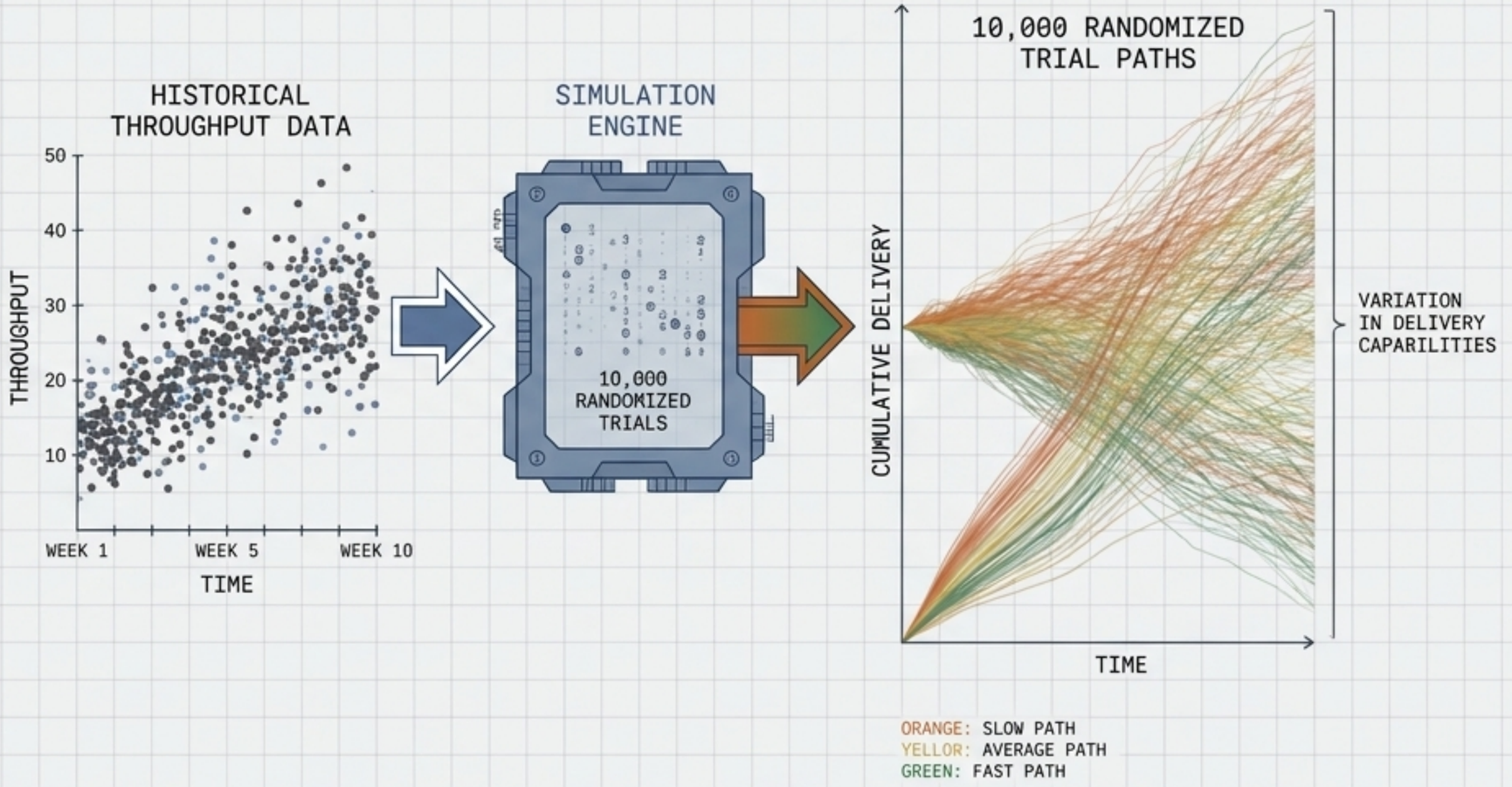
Stop guessing. Take your team's actual historical throughput data.

The Simulation

Run 10,000 randomized trials of future delivery capabilities by sampling from your own past performance.

The Result

An aggregate model that accounts for the natural variation, bottlenecks, and sick days that human estimation ignores.



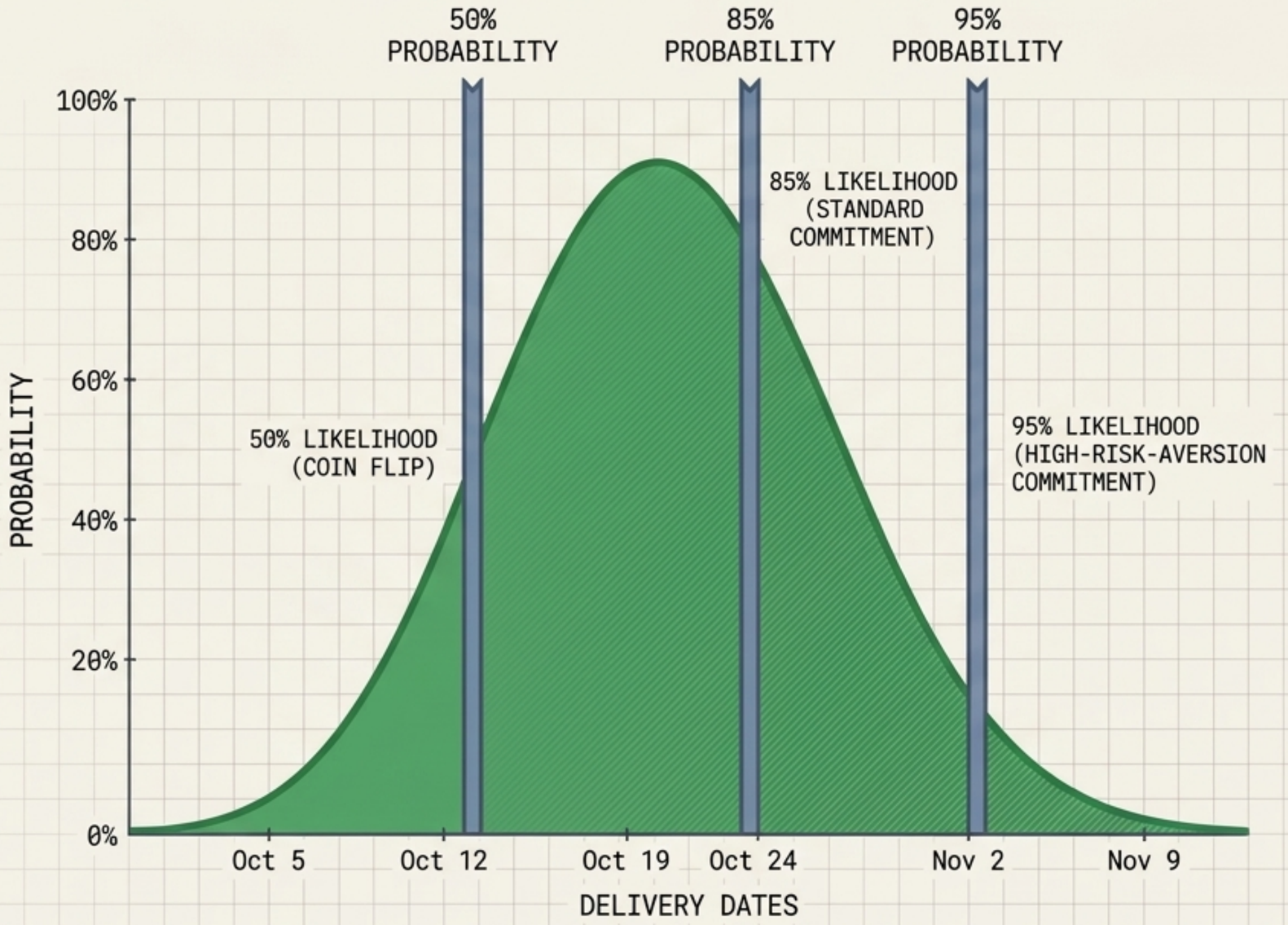
Gemini Notebook

Eradicating the Single-Date Deterministic Lie

How to Communicate:

- Based on our historical data, we are 50% likely to deliver by Oct 12. (Coin flip)
- We are 85% likely to deliver by Oct 24. (Standard commitment)
- We are 95% likely to deliver by Nov 2. (High-risk-aversion commitment)

The Rule: Never provide a single date. Always provide a probability range. Allow the business to decide how much risk they want to assume.



Gemini Notebook

The Flawed Quest to Measure Productivity

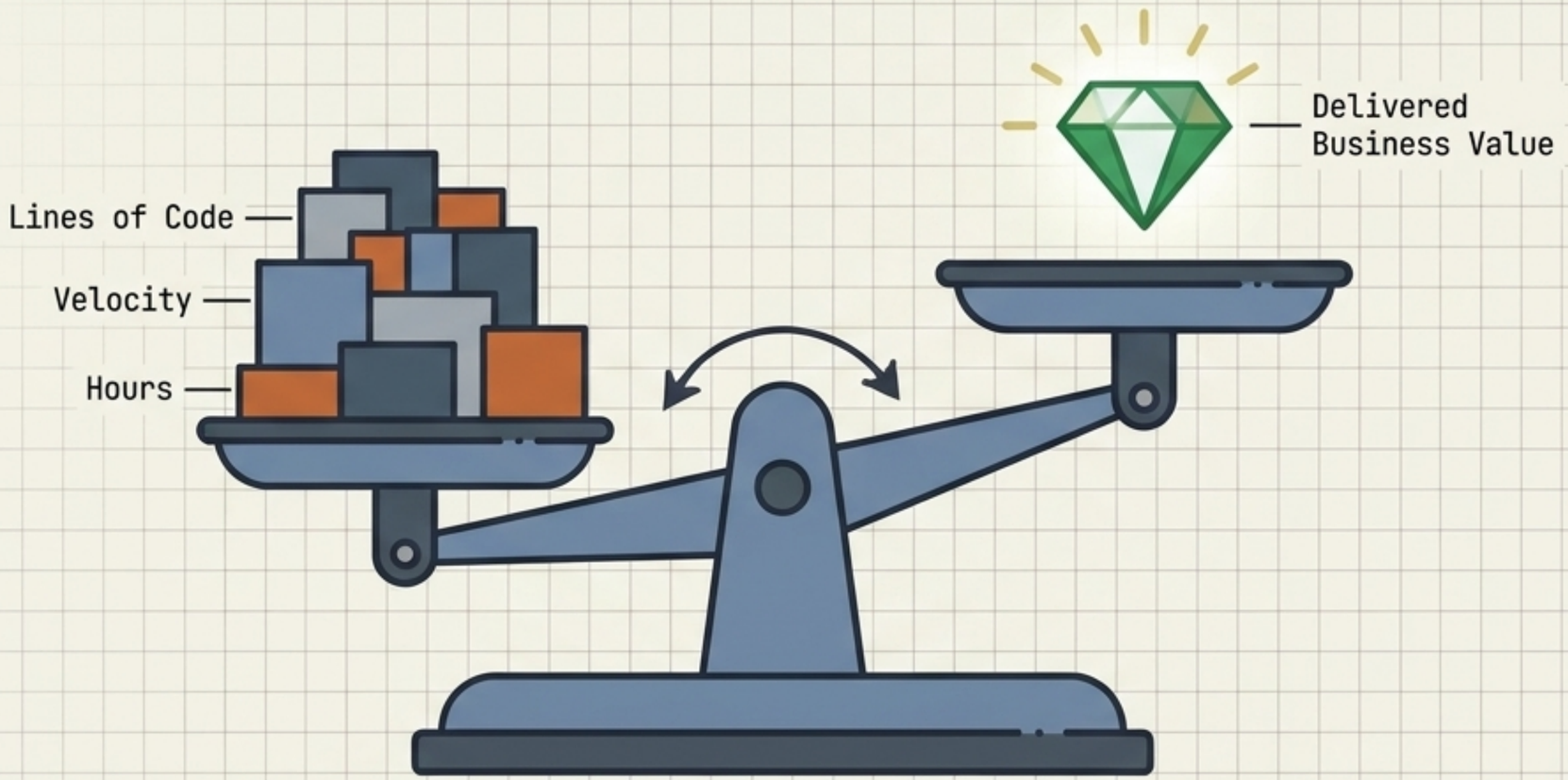
The Illusion

Productivity cannot be measured by inputs or outputs (LOC, Function Points, Velocity).

A 100K LOC system built in a year is not more productive than a 10K LOC system built in a year if the design is worse.

The Reality

Individual output metrics fail. True productivity is solely measured by delivered business value at the team level.



Gemini Notebook

Empirical Science: The DORA Research

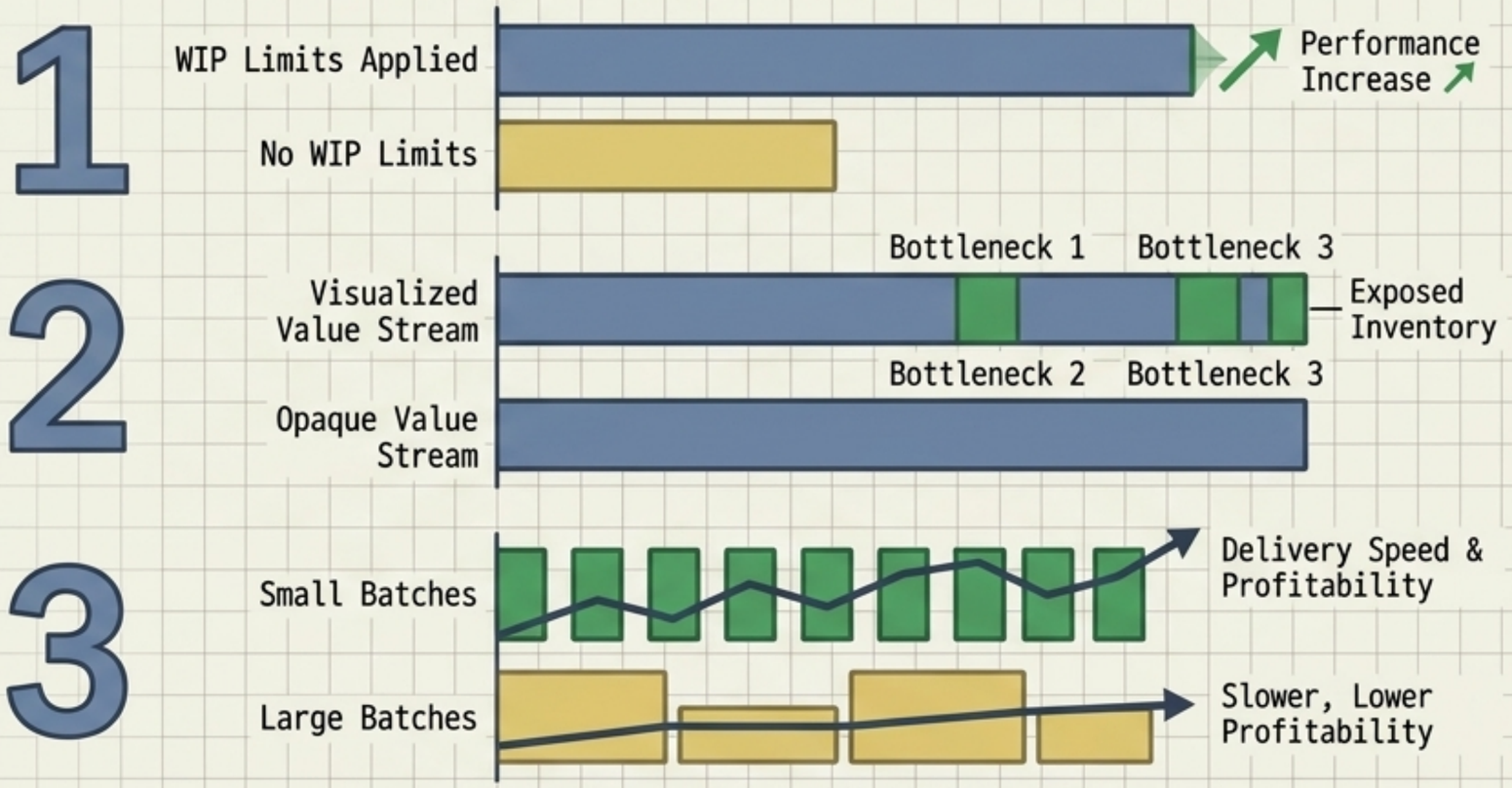
The Evidence: DORA's peer-reviewed research proves that manufacturing-style WIP limits work in knowledge work.

Key Findings:

1. Limiting WIP directly drives improvements in software delivery performance.

2. Visualizing work in the value stream exposes bottlenecks (invisible inventory).

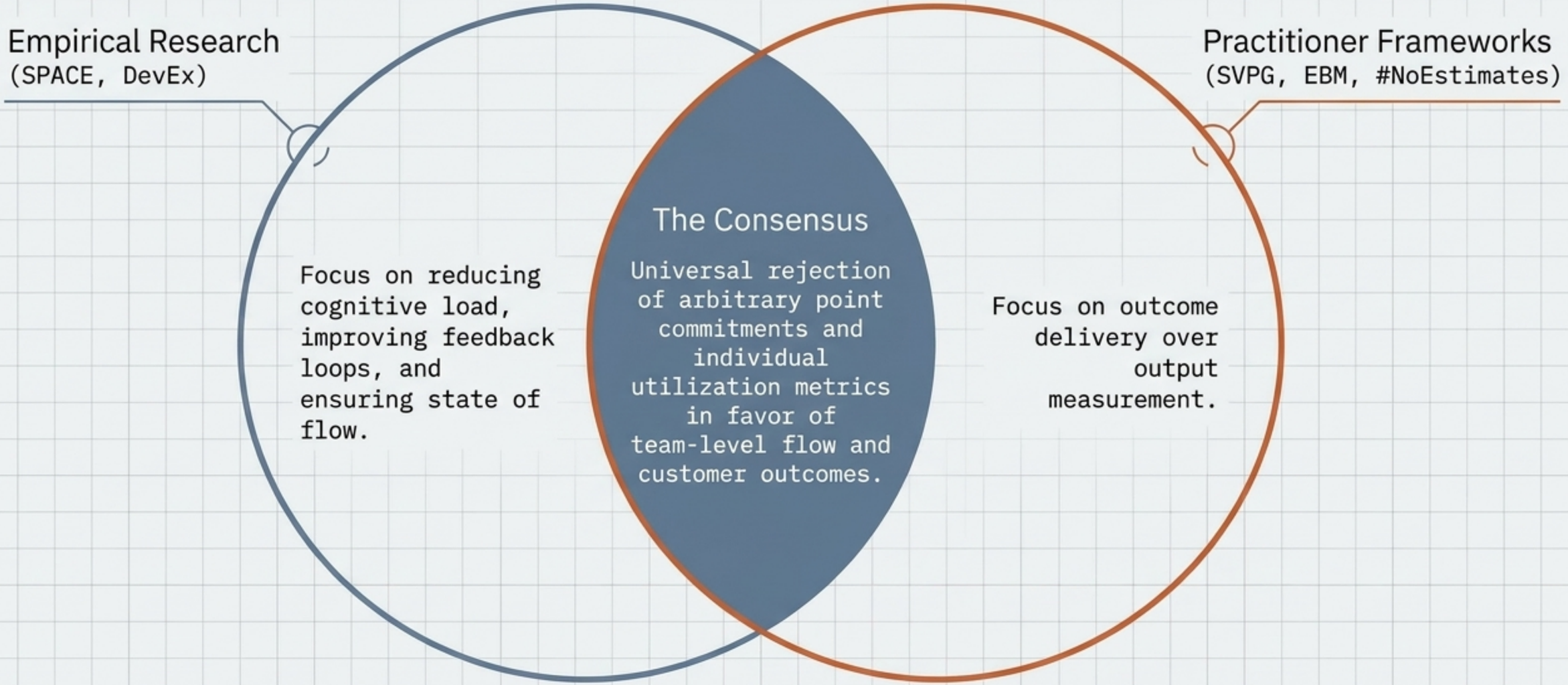
3. Working in small batches predicts not only delivery speed but overall organizational profitability.



Conclusion: Flow metrics are not just Agile theory; they are scientifically validated predictors of commercial success.

Gemini Notebook

When Science Meets Practice



Gemini Notebook

Planning Theatre Diagnostic

The Exchange Rate

Converting abstract story points directly into developer-hours (e.g., 1 pt = 4 hrs).

The Leaderboard

Executives comparing 'Velocity' across entirely different teams to gauge performance.

The Utilization Trap

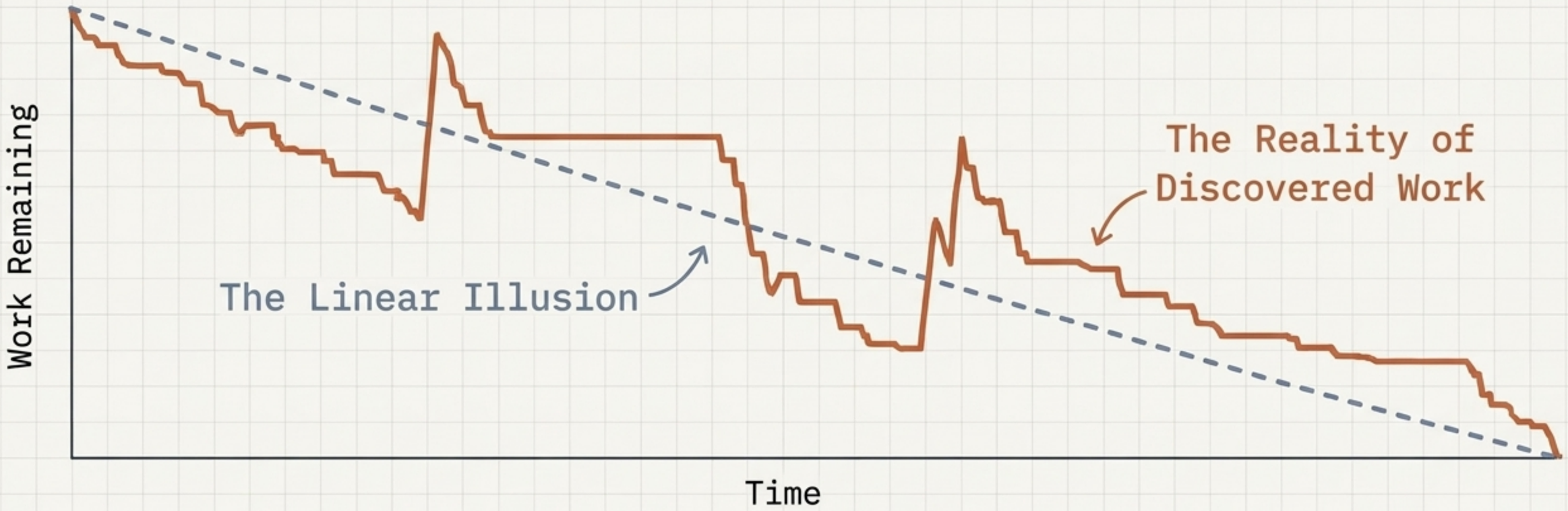
Assigning developers to 4 different active tasks simultaneously to ensure no one is 'idle'.

The Schedule Variance

Measuring success purely by how closely the team adhered to an upfront guess.

Gemini Notebook

The Linear Burndown Illusion



The Trap:

Linear burndowns assume zero discovery risk. They assume the list of tasks is perfectly known on day one.

The Reality:

As a team gets involved in the problem space, discovered tasks emerge. The to-do list actually grows as progress is made.

The Consequence:

Forcing teams to adhere to a linear burndown guarantees quality-cutting death marches.

Gemini Notebook

The 4-Quarter Gantt Illusion

The Problem:

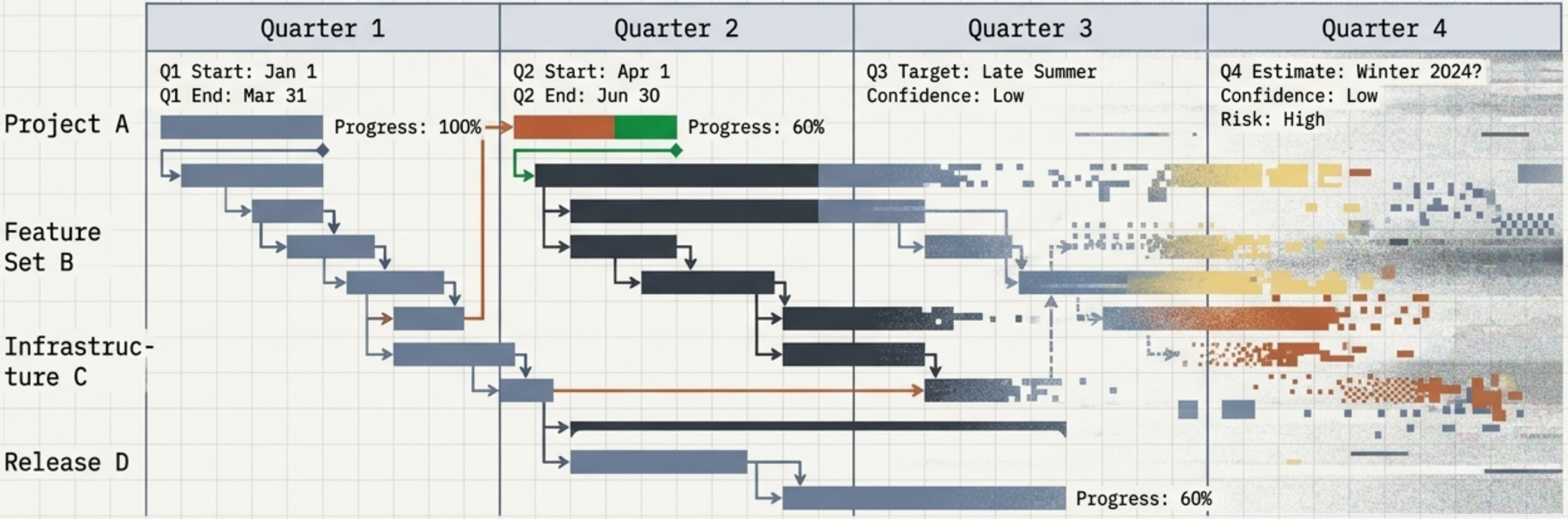
Pretending zero discovery risk exists 12 months out.

The Reality:

Promising specific features on specific dates a year in advance guarantees failure. Customer needs change, technologies shift, and initial assumptions are almost always wrong.

The Cost:

Building the wrong thing, exactly on time and on budget, buys the business nothing.



Gemini Notebook

SVPG Empowered Teams & Outcome Objectives

Mercenaries vs. Missionaries:

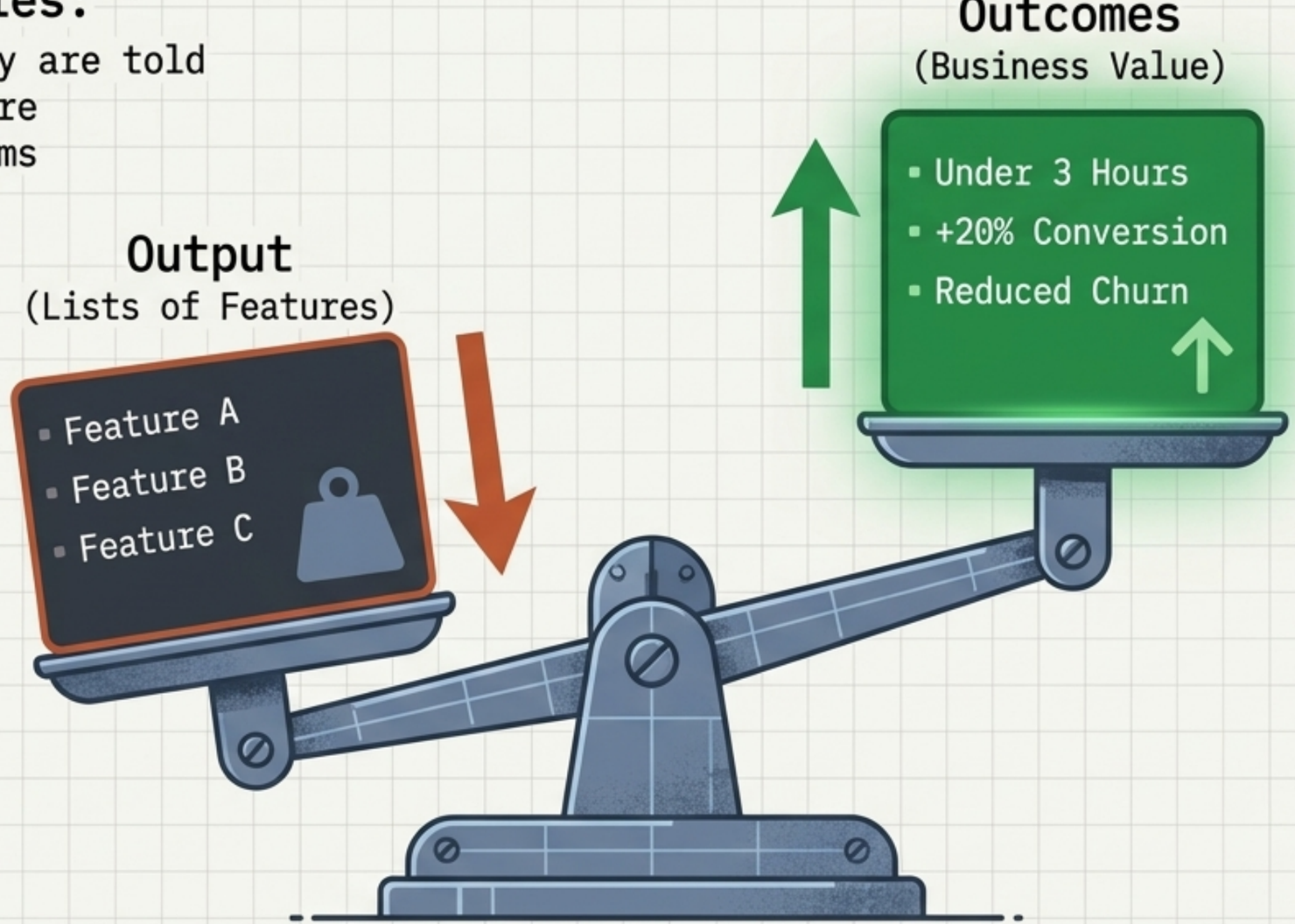
Mercenaries build exactly what they are told (Feature Roadmaps). Missionaries are empowered to solve customer problems (Outcome Roadmaps).

The Shift:

Roadmaps must dictate objectives (e.g., 'Reduce onboarding time to under 3 hours'), not features (e.g., 'Build the new wizard').

The Litmus Test:

If a feature ships but fails to move the business metric, the team's job isn't done.



Gemini Notebook

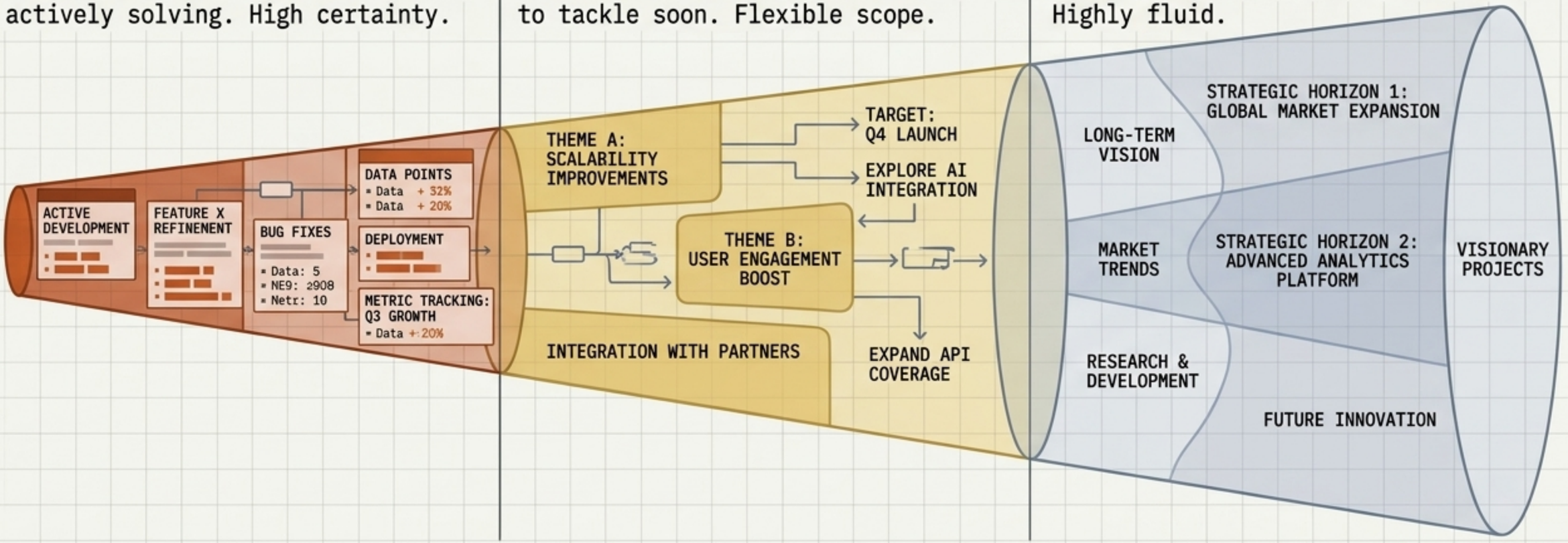
The Now–Next–Later Roadmap

The Strategy: Communicate strategic intent to stakeholders without making fraudulent, deterministic long-term commitments.

NOW (High Fidelity):
Granular, specific problems we are actively solving. High certainty.

NEXT (Medium Fidelity):
Broad themes and objectives we intend to tackle soon. Flexible scope.

LATER (Low Fidelity):
Strategic goals and distant horizons. Highly fluid.



Gemini Notebook

Monday Morning Upgrades

```
> sudo stop-velocity-comparisons
Eradicate cross-team velocity metrics immediately.
Stop measuring performance by guesses.

> invest slice --count
Slice stories small enough that you can just count them.
Discard Fibonacci sizing if it causes meeting waste.

> track work-item-age
Start visualizing Work Item Age in daily standups. Hunt
down blocked work before it poisons your cycle time.
```

Gemini Notebook

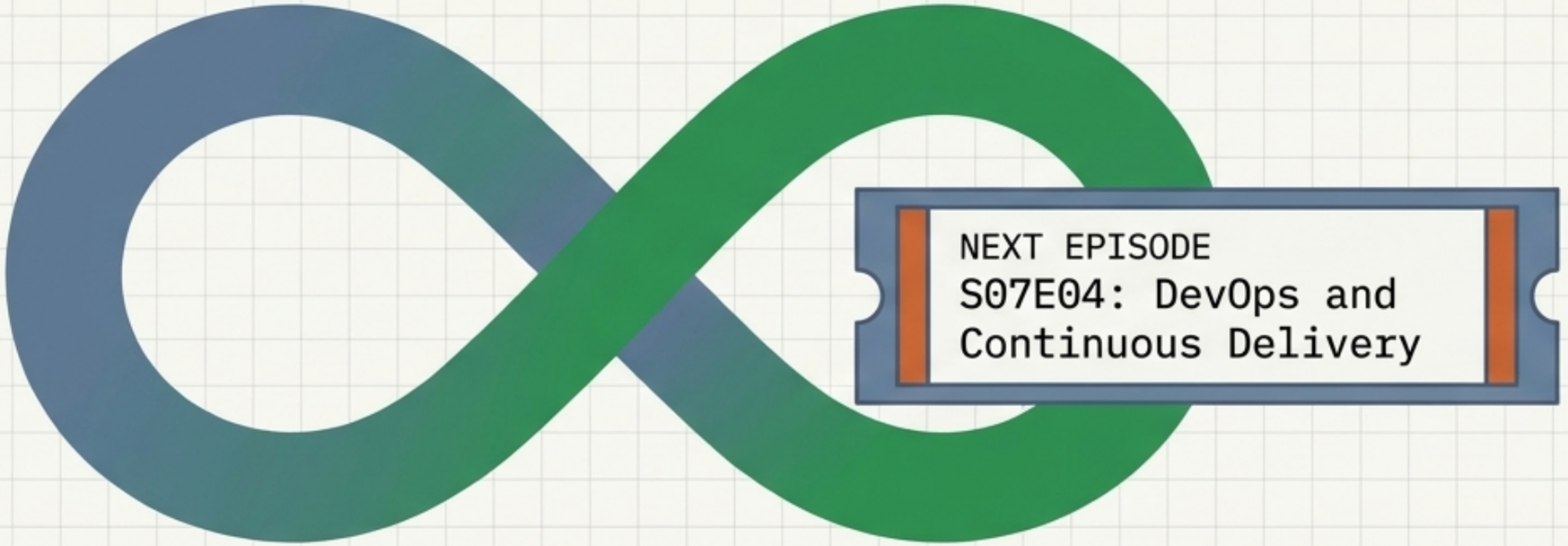
Closing the Loop

Summary:

We've dismantled planning theatre and replaced it with empirical flow metrics, probabilistic forecasting, and outcome roadmaps.

Preview — S07E04: DevOps and Continuous Delivery:

Now that we've fixed the planning pipeline, how do we automate the delivery pipeline? Coming next: DevOps, Continuous Delivery, and closing the value loop.



Gemini Notebook