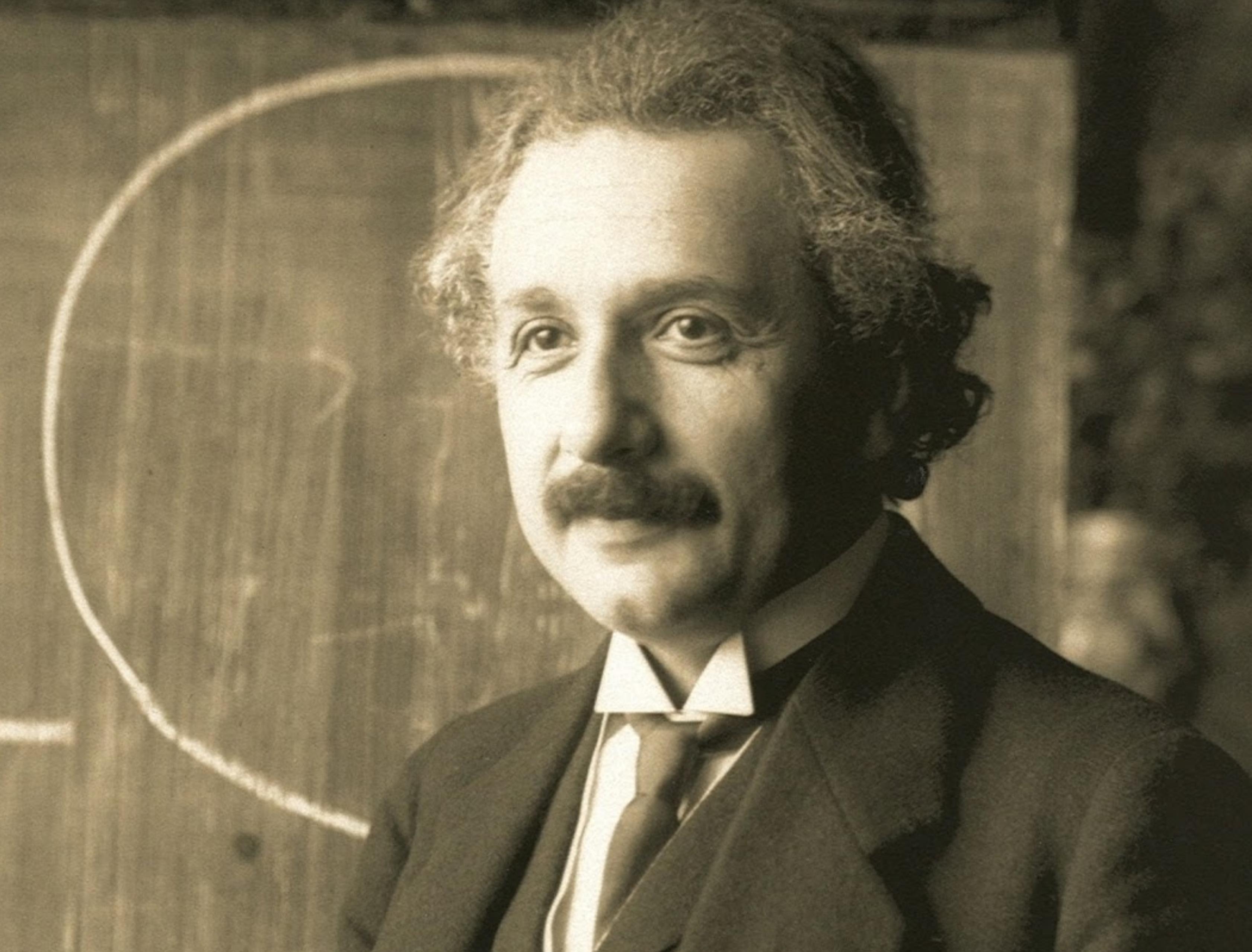


XPE | AUG. 25, 2022 | MATTHEW PHILIP

# NO (LAB) JACKET REQUIRED

Designing Experiments for Organizational Learning

K.



# WHAT WE'LL COVER

## Activity

Introductions

Experiment-Thinking Activities

Improvement Kata

Identifying Experiments

Experiment Canvases

Questions (and Answers?)

\* 10-minute break approximately every hour

# WHO'S TALKING ABOUT EXPERIMENTS?

IMPROVE COLLABORATIVELY, EVOLVE **EXPERIMENTALLY**.

Kanban Method

**EXPERIMENT AND LEARN RAPIDLY.**

Modern Agile

Lean Startup

PIVOTING .... A STRUCTURED COURSE CORRECTION DESIGNED TO TEST ... **HYPOTHESIS** ABOUT THE PRODUCT, STRATEGY, AND ENGINE OF GROWTH.

EACH FEATURE INCLUDES A BENEFIT **HYPOTHESIS**.

SAFe

IN COMPLEX ENVIRONMENTS, YOU CAN'T FOLLOW RECIPES OR CONDUCT DETAILED ANALYSIS TO UNDERSTAND THE SITUATION. RATHER, YOU MUST **EXPERIMENT (PROBE)**.

Cynefin

DevOps

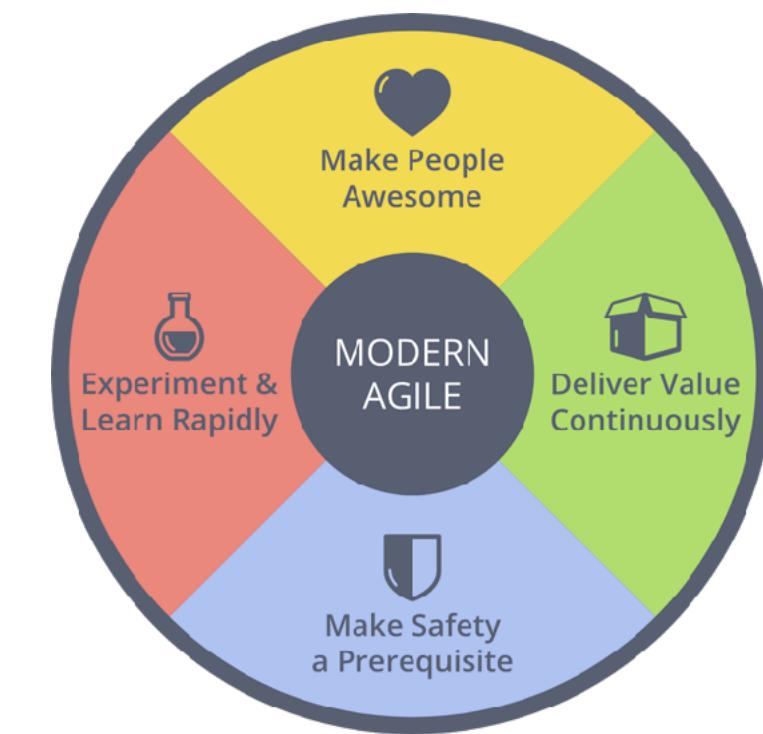
...A CULTURE THAT FOSTERS ... **CONTINUAL EXPERIMENTATION, TAKING RISKS AND LEARNING FROM FAILURE...**

PLAN, DO, CHECK (STUDY), ACT

Deming

SCALE EXPERIMENTING-LEARNING-ADAPTING-EVOLVING GUIDED BY COMPLEXITY, LEAN AND AGILE.

Generative Scaling



**BUT IN SPITE OF THE LIP SERVICE THAT IS PAID TO...  
‘LEARNING FROM FAILURE,’ TODAY’S ORGANIZATIONS,  
PROCESSES, AND MANAGEMENT OF INNOVATION  
OFTEN IMPEDE EXPERIMENTATION.**

— STEFAN THOMKE, **EXPERIMENTATION MATTERS: UNLOCKING THE POTENTIAL OF NEW TECHNOLOGIES**

**YOU KEEP USING THAT WORD**

**I DO NOT THINK IT MEANS WHAT YOU THINK IT MEANS**

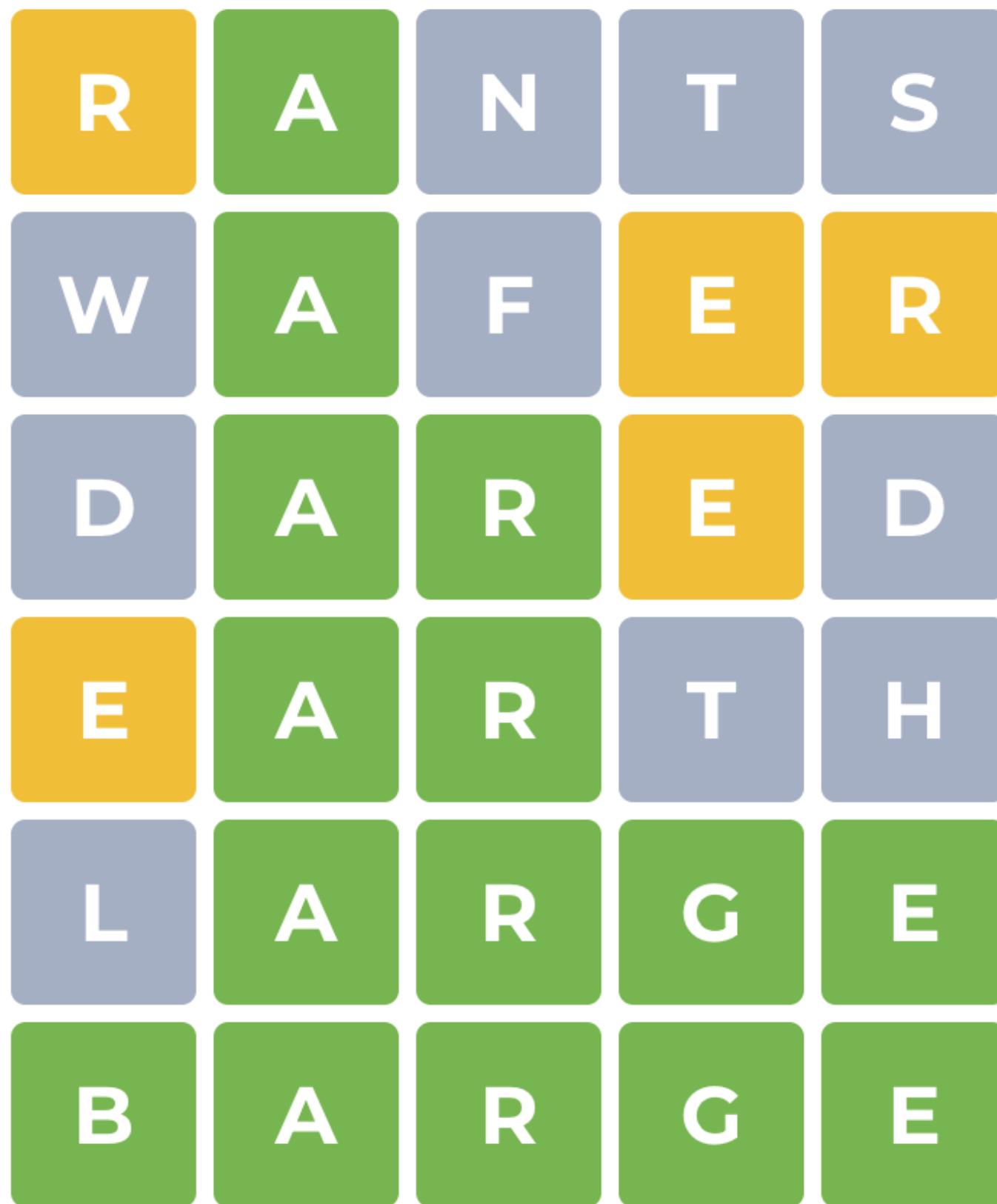
**WHAT IS AN EXPERIMENT MINDSET?  
WHAT IS SCIENTIFIC METHOD?**

# WHAT IS AN EXPERIMENT MINDSET? SCIENTIFIC THINKING?

Scientific thinking is a skill – a habit – that empowers us to better navigate complexity:

1. *Acknowledging* that our comprehension is always incomplete and possibly wrong.
2. *Assuming* that answers will be found by test rather than just deliberation. (You make predictions and test them with experiments.)
3. *Appreciating* that differences between a prediction and what actually happens can be a useful source of learning and adjustment.

# **WARMUP: WORDLE**



You Won! 🏆



# **WHAT DID WE LEARN?**

**WHAT HAPPENED IN THAT GAME?  
HOW DID YOU APPROACH SOLVING IT?  
ANYTHING UNEXPECTED HAPPEN?**

# ELEUSIS EXPEDITIOUS: A GAME ABOUT HYPOTHESES

# OBJECT

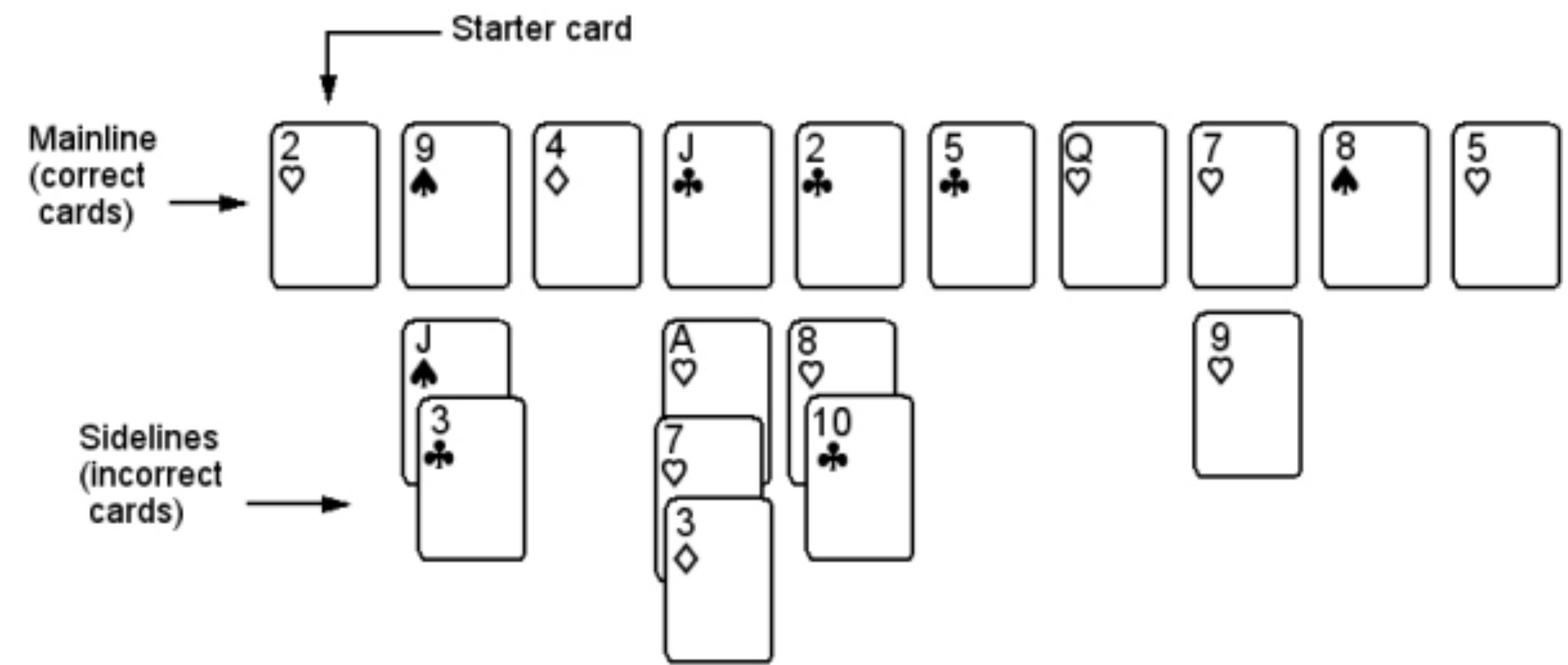
- Each round, one player is chosen as the oracle, who has a secret rule for which the cards must be played. (The oracle does not play any cards.)
  - For example: *Each card has to be a different color from the card before it.*
- Other players lay down cards to discover the pattern. If a player lays down a card correctly, he or she can try to guess the rule.



— Based on Eleusis Express, which itself is a modification of Robert Abbott's game Eleusis

# GAMEPLAY

- The oracle places the pile of cards face down, turns over the top card, and puts it on the table.
- The player to the left of the oracle goes first, then the play continues clockwise.
- A player turns the top card from the deck and lays it on the table. He or she asserts whether it follows the rule or not, and the oracle accepts or rejects:
  - If the card follows the rule, the card is placed horizontally to the right of the last card on the mainline.
  - If the card does *not* follow the rule, the card is placed below the last correct card (it either starts a new sideline or it adds to a sideline).
- Whenever a player makes a correct assertion, he or she can guess aloud the rule.
- All cards are played to a central layout that grows as the round progresses.
- The round ends when a player guesses the rule.
- Ace is 1 (odd), jack is 11 (odd), queen is 12 (even), and king is 13 (odd)



# **WHAT DID WE LEARN?**

**HOW DID IT FEEL TO BE WRONG? RIGHT?  
WHAT IF WE MAPPED THE RESULTS OF OUR LEARNING?**

# SCIENTIFIC METHOD AND ELEUSIS EXPEDITIOUS

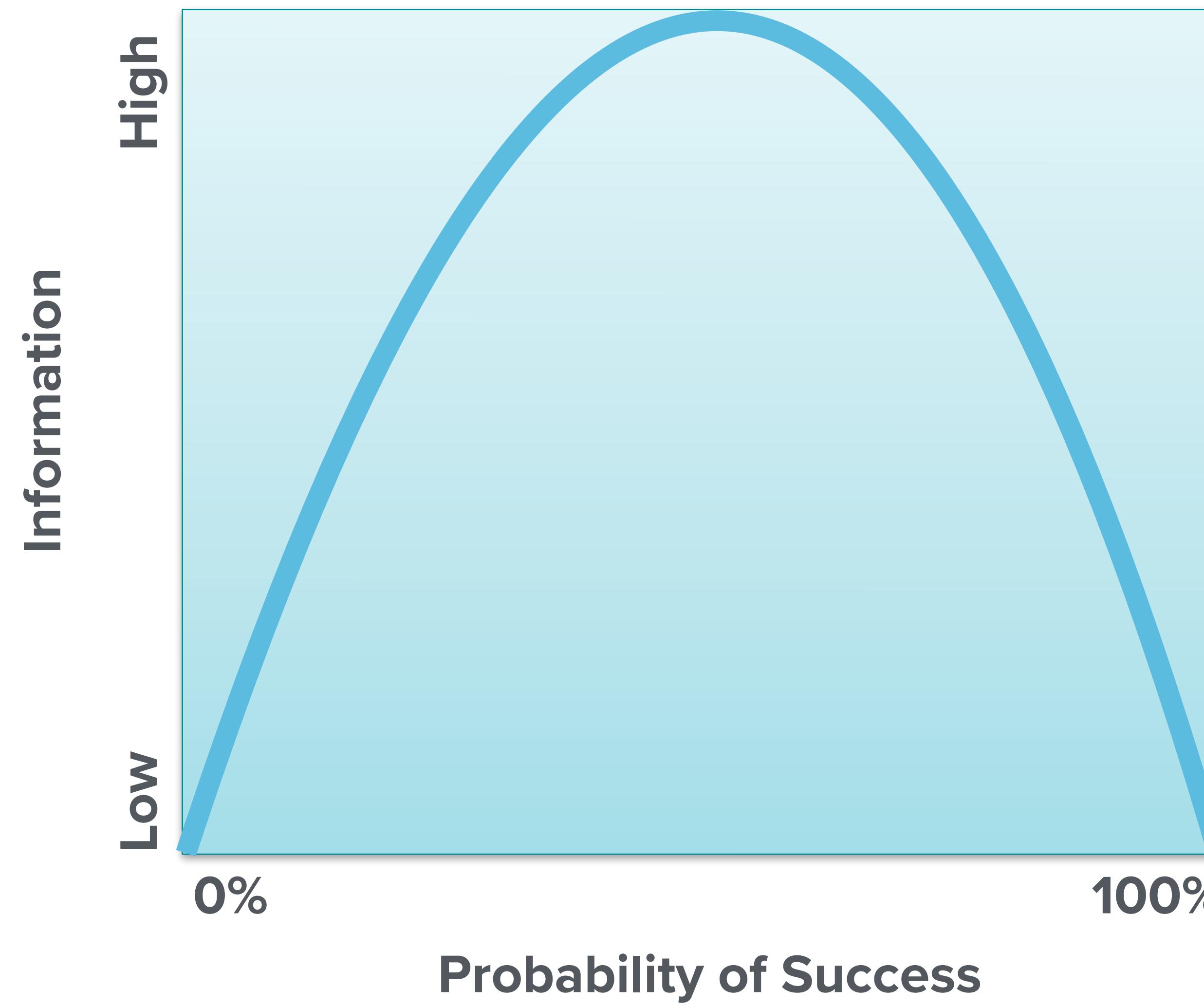
Define a question	“What is the rule?”
Gather information and resources (observe)	“This card apparently follows the rule, this other one doesn’t.”
Form an explanatory hypothesis	“The rule is only number cards.”
Test hypothesis by experimenting (replicating)	“Numbers 1-10 would follow rule, face cards would not.”
Analyze the data	“2 and 6 followed the rule, but 5 did not.”
Interpret and form new hypothesis	“The rule is only even numbers.”
Share results and learnings	

A black and white photograph of the Grateful Dead band members. From left to right, they are: Phil Lesh (bassist), Bob Weir (guitarist), Jerry Garcia (lead singer/guitarist), Mickey Hart (drummer), and Bill Kreutzmann (percussionist). They are all smiling and looking towards the camera. The background is a brick wall.

**“WELL I AIN’T OFTEN RIGHT,  
BUT I’VE NEVER BEEN WRONG.”**

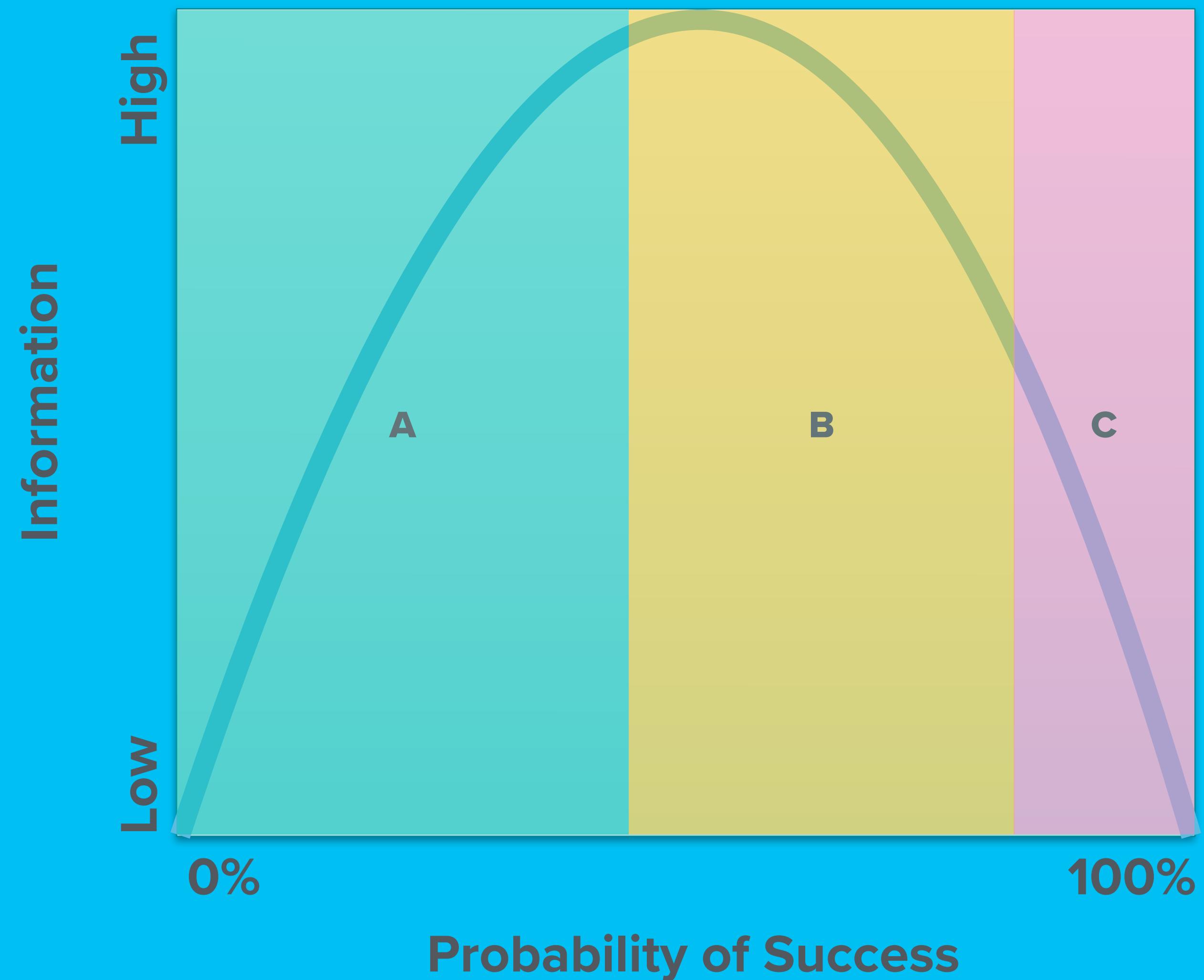
GRATEFUL DEAD

# INFORMATION THEORY



# THINKING IN BETS

- Belief -> bet -> [Set of outcomes]
- Belief -> bet ->
  - outcome A (probability of likelihood %)
  - outcome B (probability of likelihood %)
  - outcome C (probability of likelihood %)



**TRUNCATE UNPRODUCTIVE PATHS  
QUICKLY.**

**DON REINERTSEN**

# REINERTSEN'S FRONT-LOADED LOTTERY OPTION

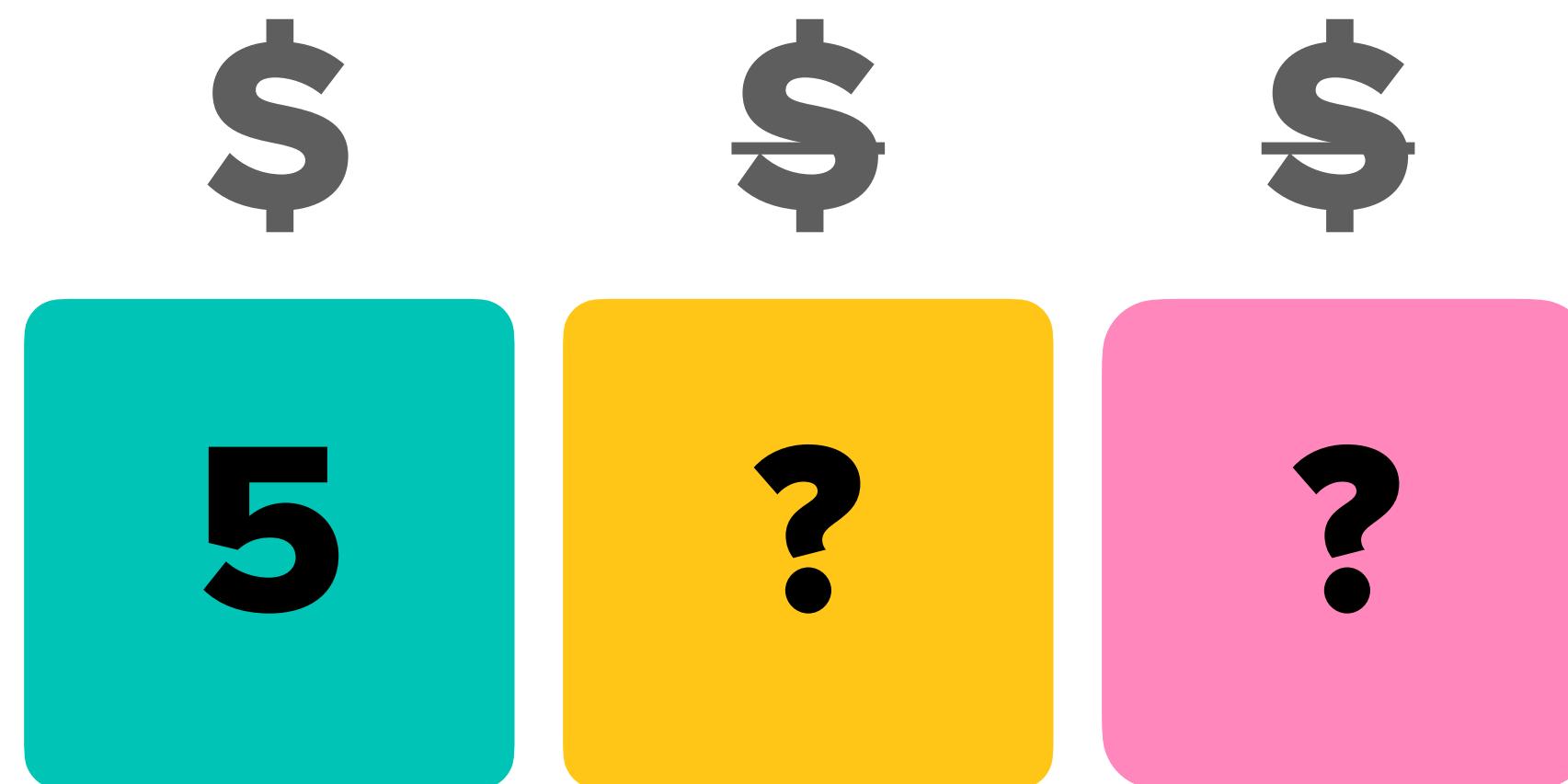
## OPTION 1

Pay \$3 to select all three digits at once



## OPTION 2

Pay \$1 for the first digit, find out if it is correct, then choose if you wish to pay \$1 for the second digit, and then choose if you wish to pay \$1 for the third digit.



# **WHAT THINGS ARE GOOD FOR EXPERIMENTATION?**

# START WITH OUTCOMES

- Quickly write a dozen or more outcomes onto sticky notes, spread out in front of you such that they all can be read easily

# BOTTOM RIGHT

- Start at the bottom right hand corner of your square.
- Carefully choose from your list the strongest single example of an outcome where there's an approach you can *all see and agree on*.
- If you're not 100% confident in the reliability of the approach or if there's not complete consensus that's ok; choose the closest example you can find.
- Now place that single best example in the bottom right corner.



# TOP RIGHT

- Select the strongest single example of an outcome where you're confident that an *expert or some research will determine a good approach.*
- As with the first corner, “good” implies confidence that the outcome will be delivered, and it's predicated on the availability (by whatever means) of someone who has the necessary expertise.



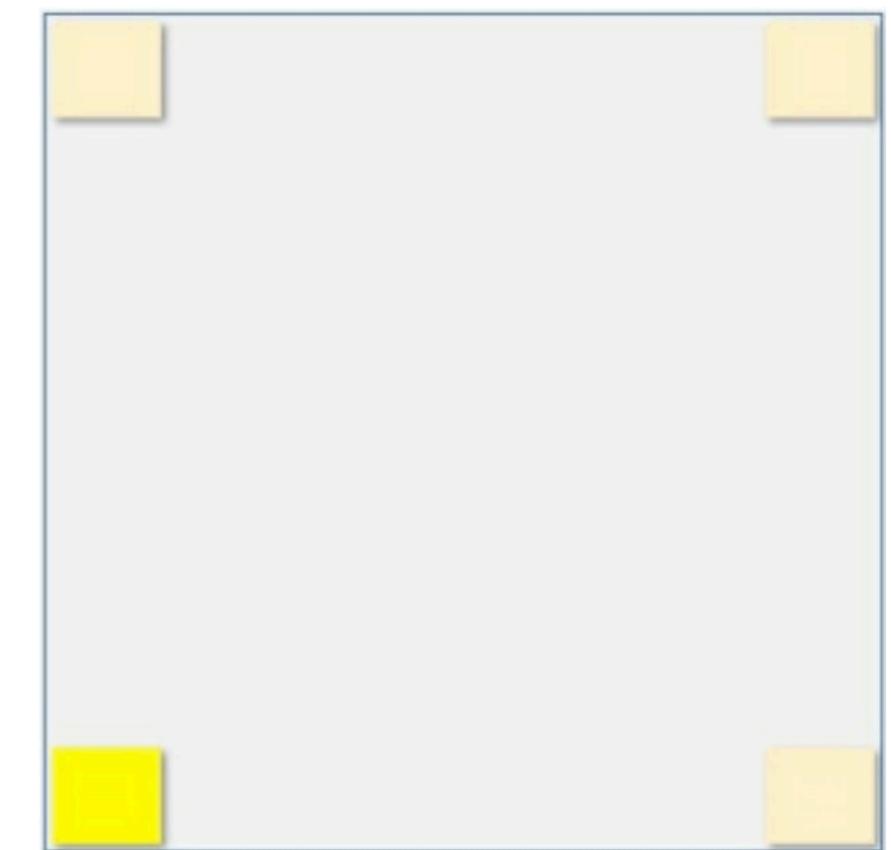
# TOP LEFT

- Carefully choose the strongest single example of an outcome where *there's no one right approach* – experts will disagree. 10 experts will give you 20 different ideas, and even though some of them might be really great, typical of this kind of outcome is the sense that no single idea is sure to get you the whole way to your outcome.
- Much as you might wish it to be otherwise, you couldn't just delegate this outcome to an expert and reasonably expect a reliable plan to be laid out for you.



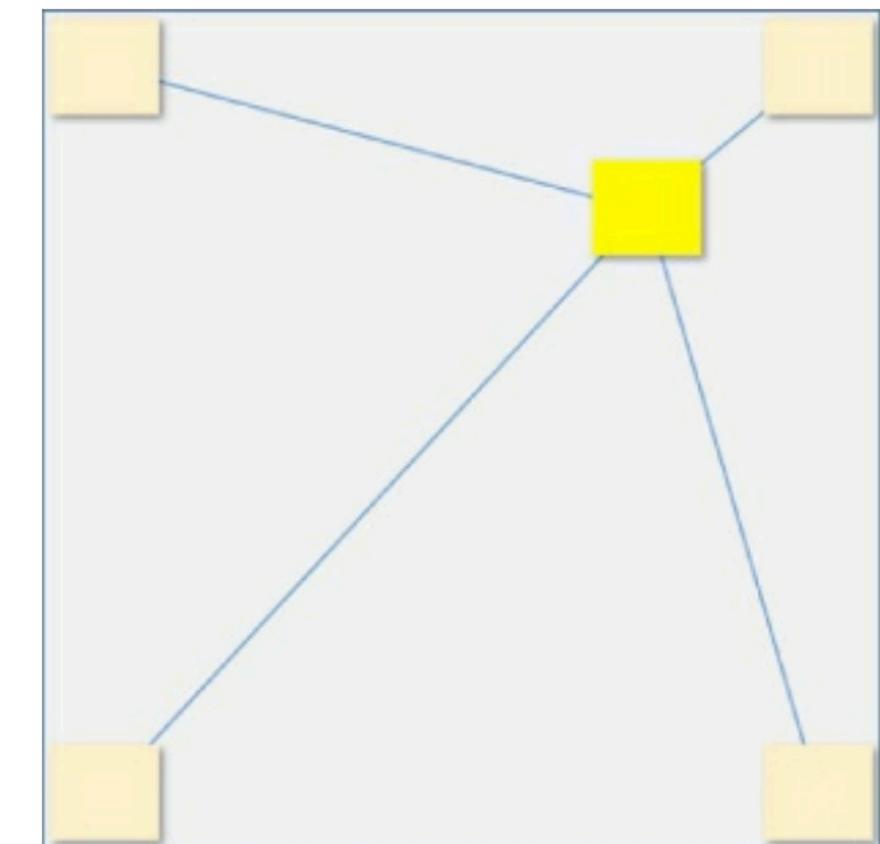
# BOTTOM LEFT

- Choose the strongest single example of an outcome where *no known approach is immediately apparent*.
- Beyond symptomatic fixes (stem the bleeding, put out the fire, etc) you're not sure that an expert will help you in the time available. You don't know which of your ideas will stick. You don't know how things will settle down after they have been disrupted – whether the source of disruption is accidental or deliberate, internal or external. You don't know where to start!



# YOUR NEXT STICKY...

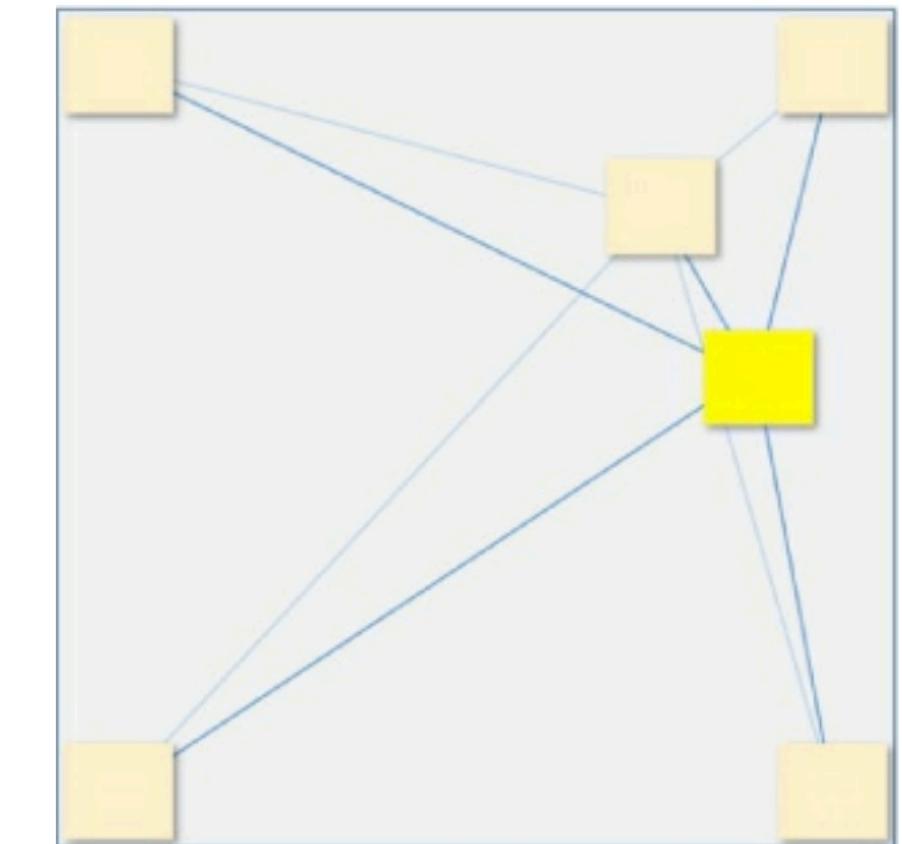
- Now choose another sticky at random – just one.
- Imagine this new sticky being drawn by imaginary lines of force towards the corner stickies to which it is most similar in terms of the kind of approach.
- It might end up next to one of them, between two, or somewhere towards the middle, drawn in various degrees towards three or four corners.



— Agendashift: Outcome-oriented change and continuous transformation

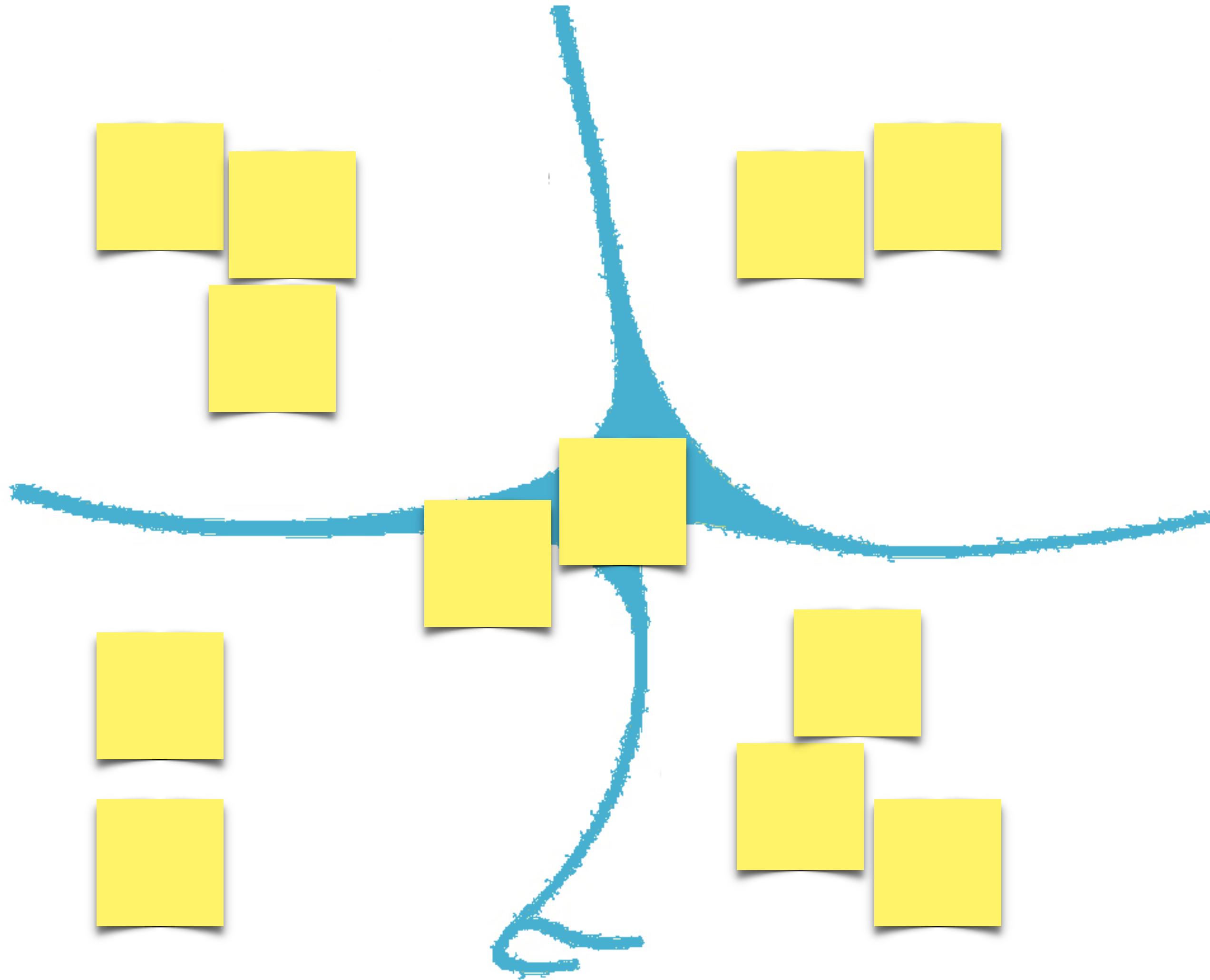
# REMAINING STICKIES

- Now add your remaining stickies one at a time, trying to place them relative to those you've placed already, rearranging as necessary.
- Some will have a strong affinity approach-wise with one of your four exemplars and will be easy to place near one of the corners. Others will seem to have a natural position somewhere relative to two or three existing stickies.
- Any that really won't be placed authentically this way should be arranged in a central region.



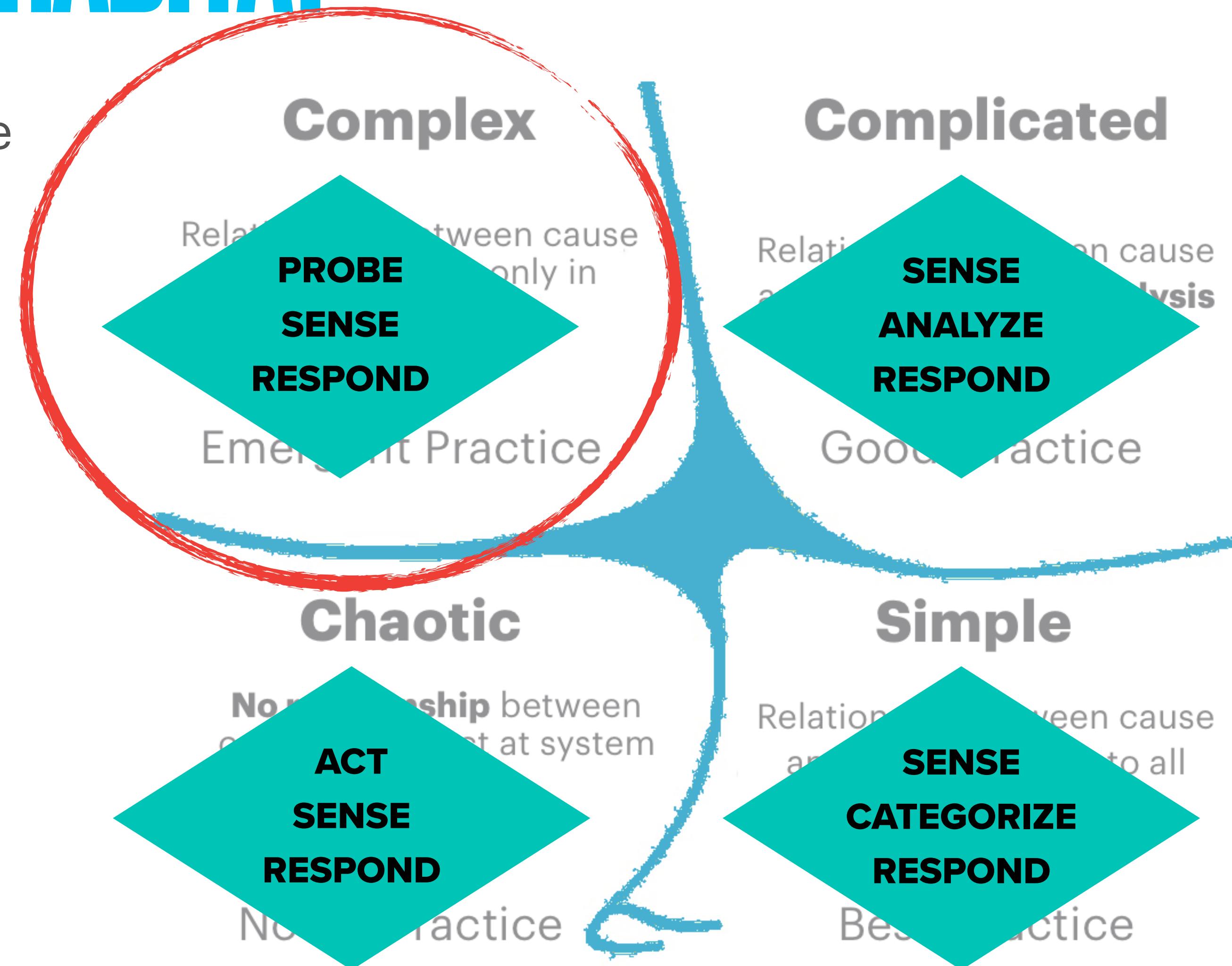
— Agendashift: Outcome-oriented change and continuous transformation

# SENSEMAKING



# CYNEFIN: “HABITAT”

- Allow ideas that are not useful to fail in small, contained and tolerable ways (Dave Snowden)
- Make decisions in situations of high uncertainty.



# WHAT DID WE LEARN?

THOUGHTS?  
OBSERVATIONS?

# EXPERIMENTS AND PRODUCT MANAGEMENT

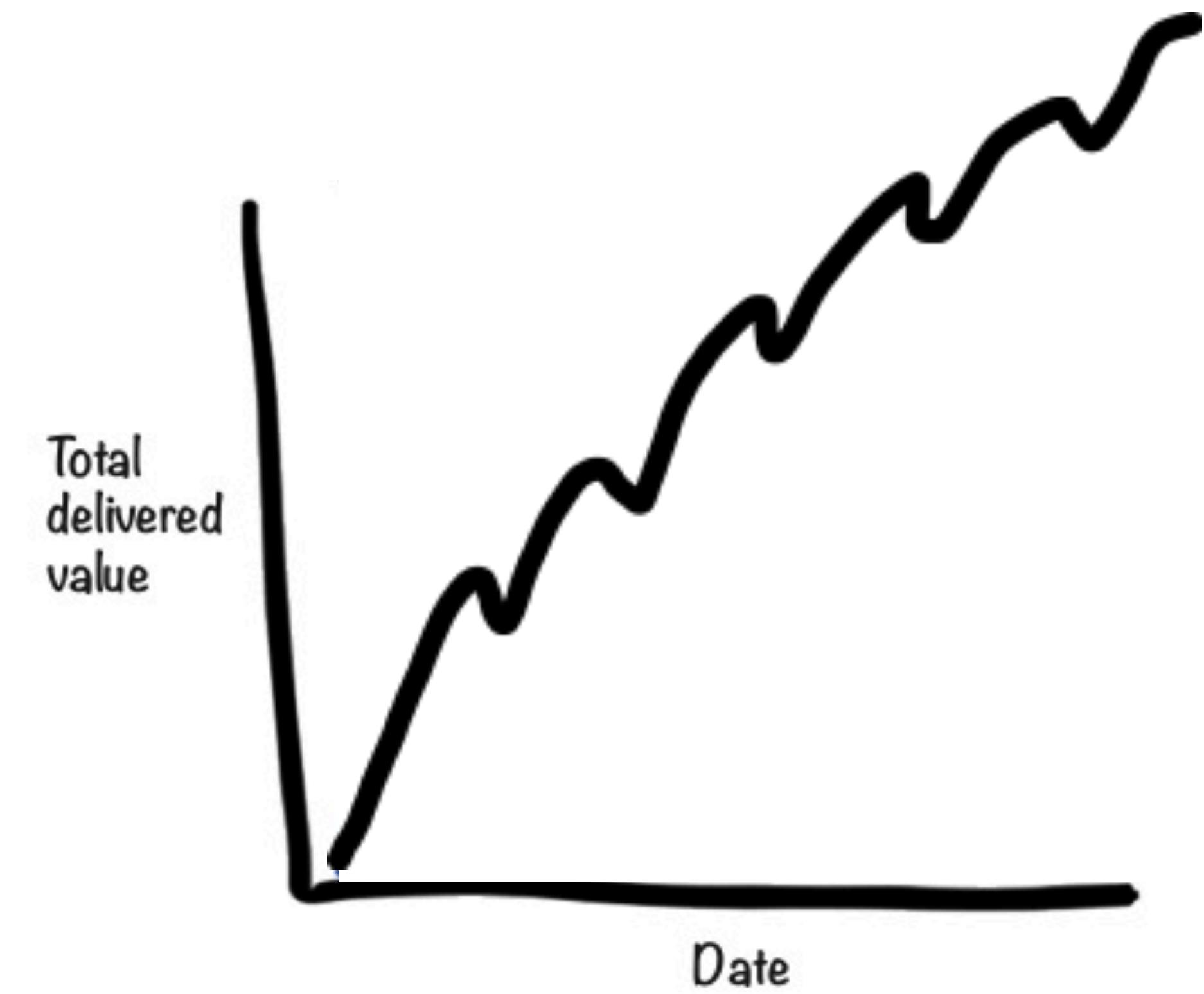
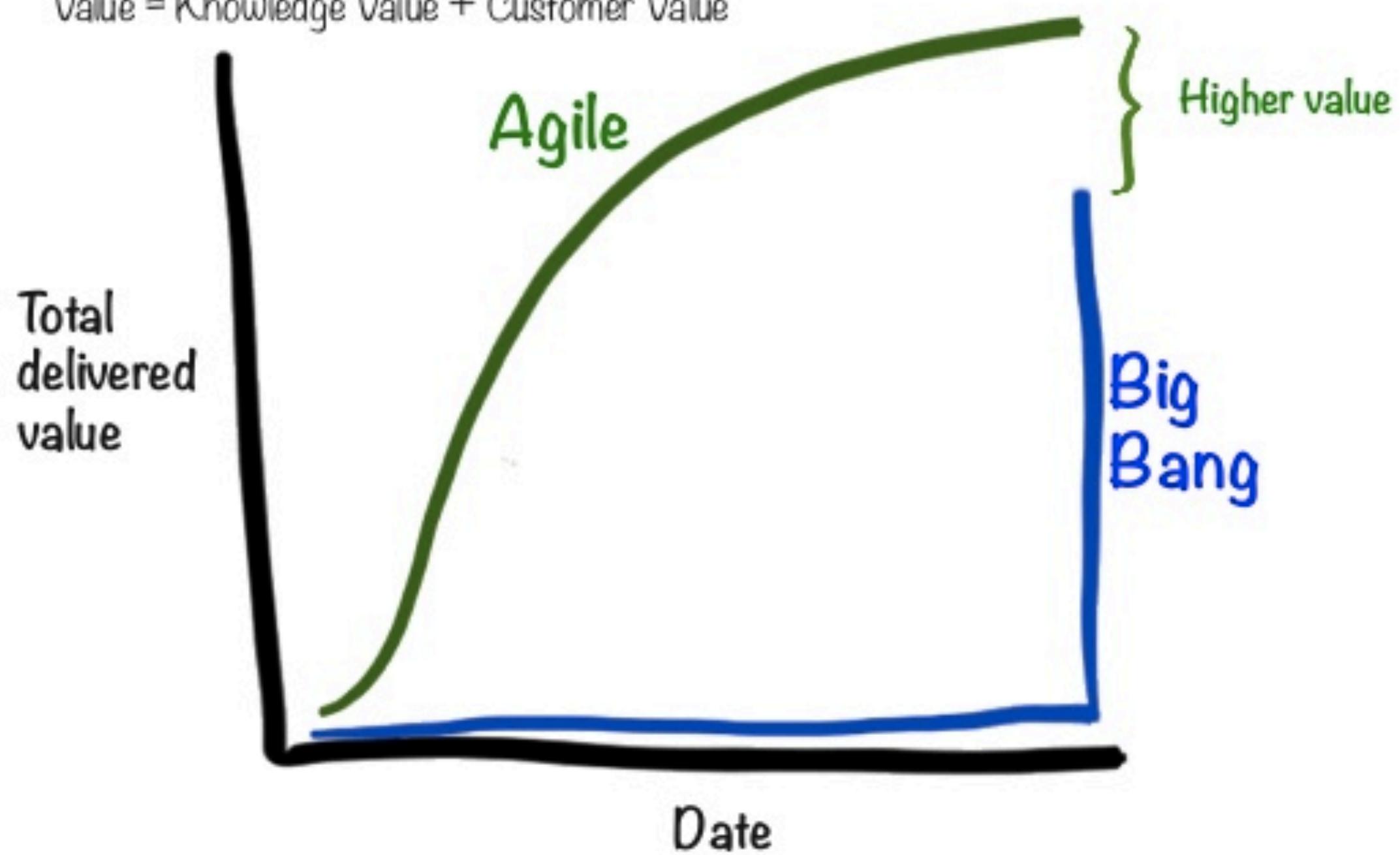
# SIGNS OF BLACK-HOLE AGILE

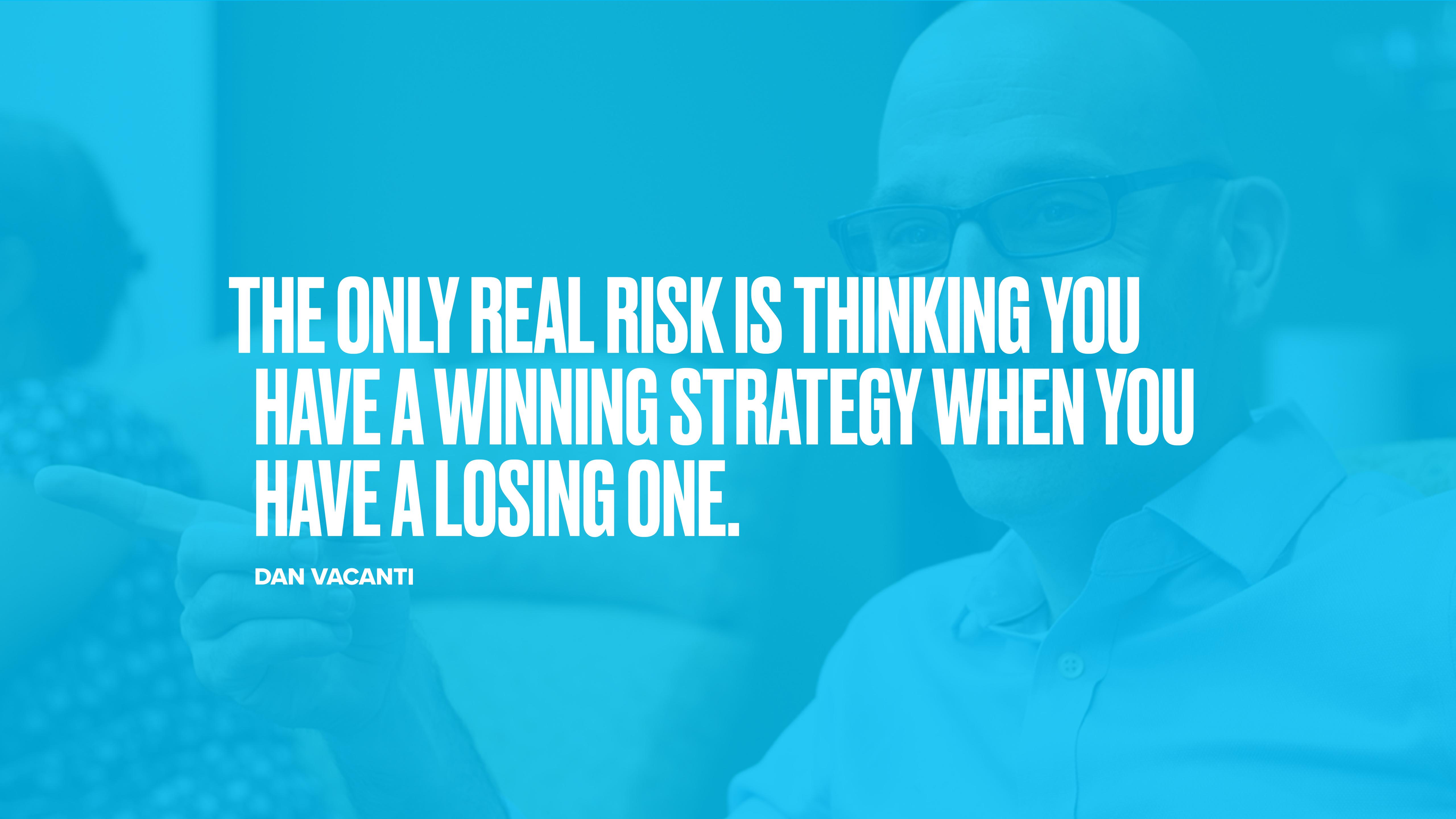
- Done means coded/tested/released (not Validated)
- Releases are celebrated more than user success
- No feature ever fails
- Planning starts with feature ideas

# AGILE VALUE PROP WHEN WE VALIDATE

Faster learning = Higher value

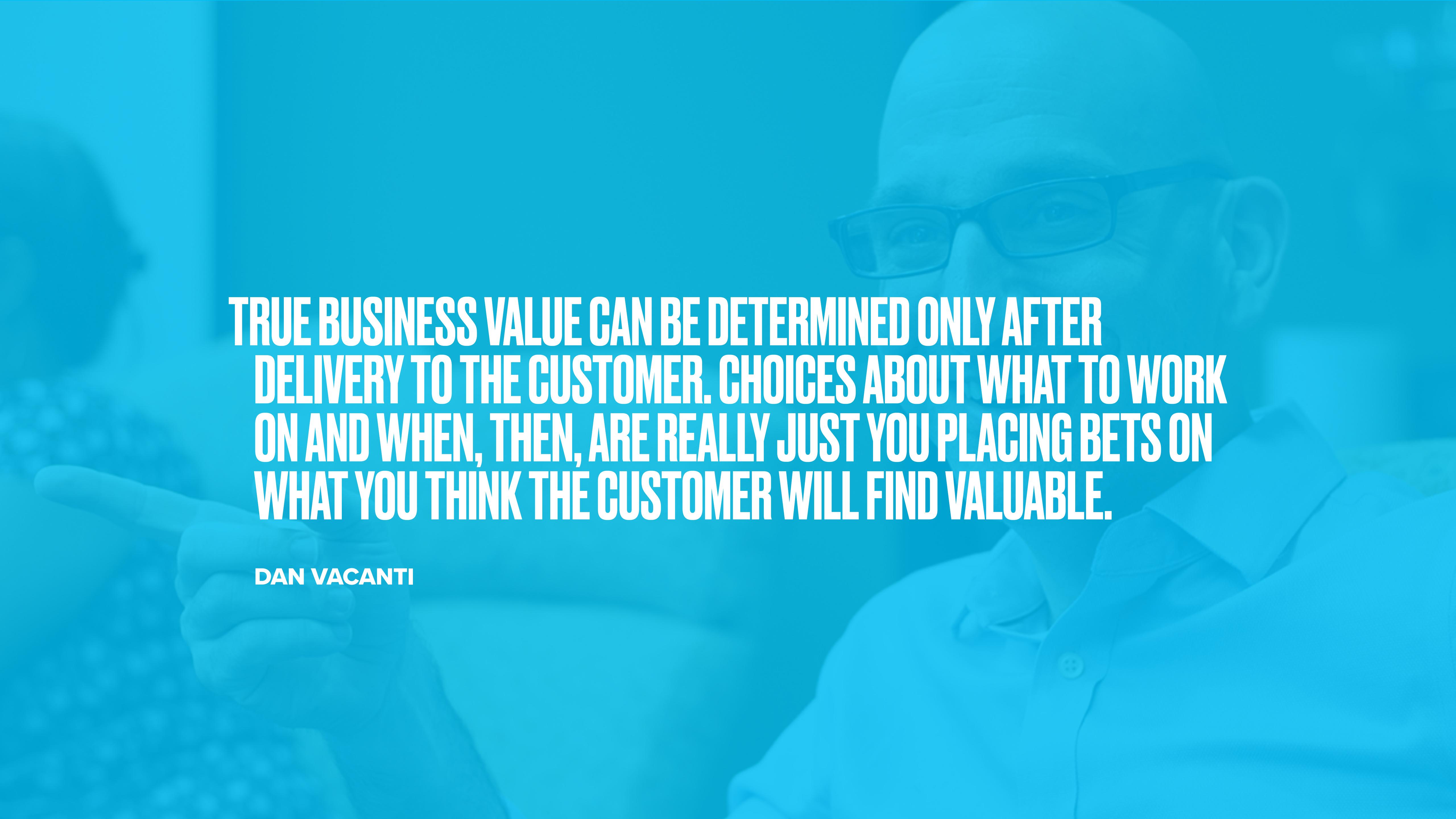
Value = Knowledge Value + Customer Value



A soft-focus background image of a man with dark hair and glasses, wearing a light blue button-down shirt. He is looking slightly to his left with a contemplative expression. The lighting is warm and focused on his face.

**THE ONLY REAL RISK IS THINKING YOU  
HAVE A WINNING STRATEGY WHEN YOU  
HAVE A LOSING ONE.**

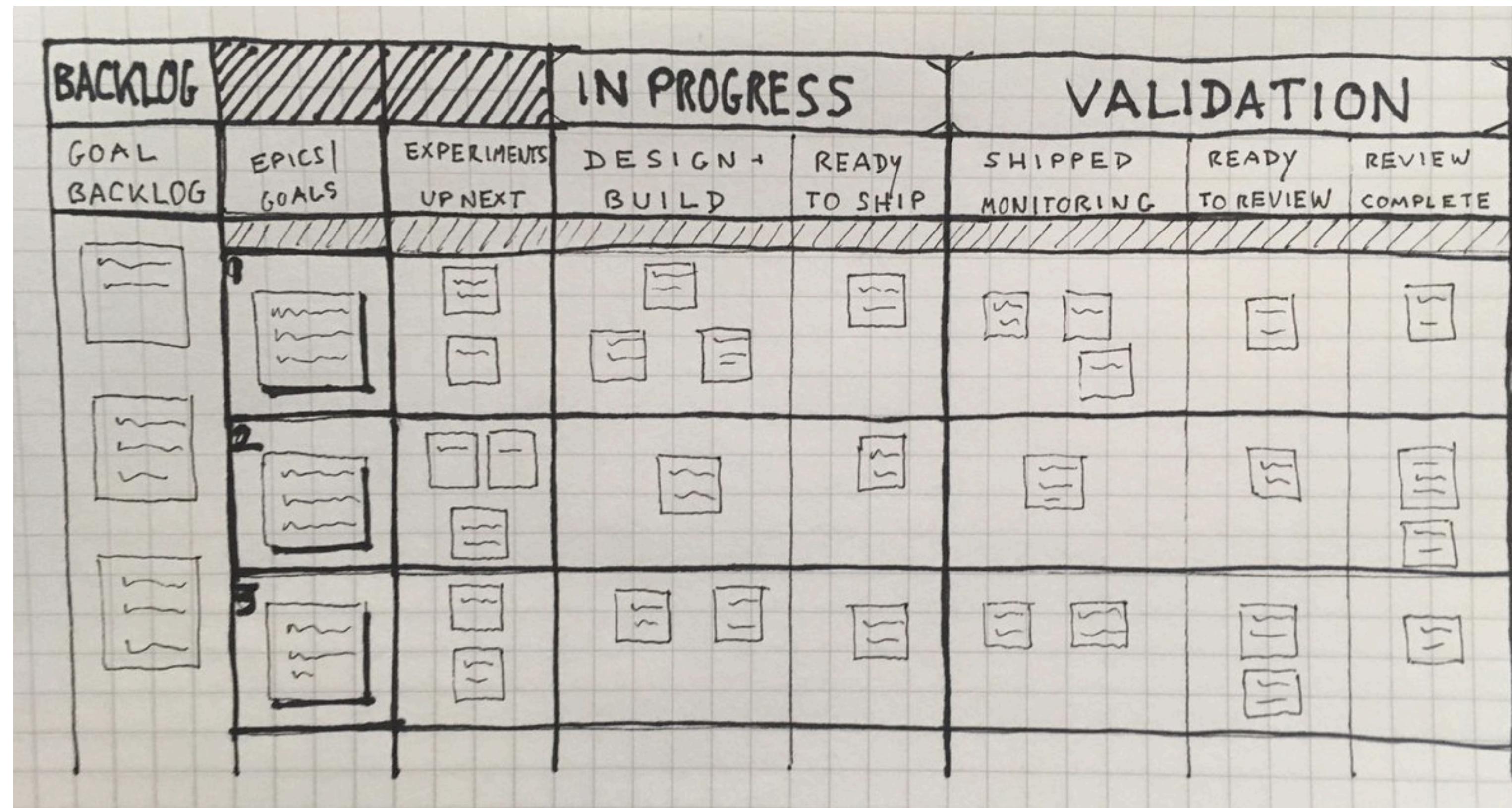
DAN VACANTI

A soft-focus background image of a man with glasses and a blue shirt, looking thoughtful.

**TRUE BUSINESS VALUE CAN BE DETERMINED ONLY AFTER  
DELIVERY TO THE CUSTOMER. CHOICES ABOUT WHAT TO WORK  
ON AND WHEN, THEN, ARE REALLY JUST YOU PLACING BETS ON  
WHAT YOU THINK THE CUSTOMER WILL FIND VALUABLE.**

**DAN VACANTI**

# WHAT IS “DONE” FOR YOU?



-- John Cutler

# Geeking with Greg

Tuesday, April 25, 2006

## Early Amazon: Shopping cart recommendations

I have talked about a couple fun projects ([\[1\]](#) [\[2\]](#)) I did at Amazon even though I was supposed to be working on other things. This story is more extreme, a project I was explicitly forbidden to do and did anyway.

I loved the idea of making recommendations based on the items in your Amazon shopping cart. Add a couple things, see what pops up. Add a couple more, see what changes.

The idea of recommending items at checkout is nothing new. Grocery stores put candy and other impulse buys in the checkout lanes. Hardware stores put small tools and gadgets near the register.

But here we had an opportunity to personalize impulse buys. It is as if the rack near the checkout lane peered into your grocery cart and magically rearranged the candy based on what you are buying.

Health food in your cart? Let's bubble that organic dark chocolate bar to the top of the impulse buys. Steaks and soda? Get those snack-sized potato chip bags up there right away.

### About Me



GREG LINDEN

[View my complete profile](#)

### Subscribe to the Feed

 [Subscribe in a reader](#)

### More Geeking with Me

[Tweeting with Greg](#)

### Blog Archive

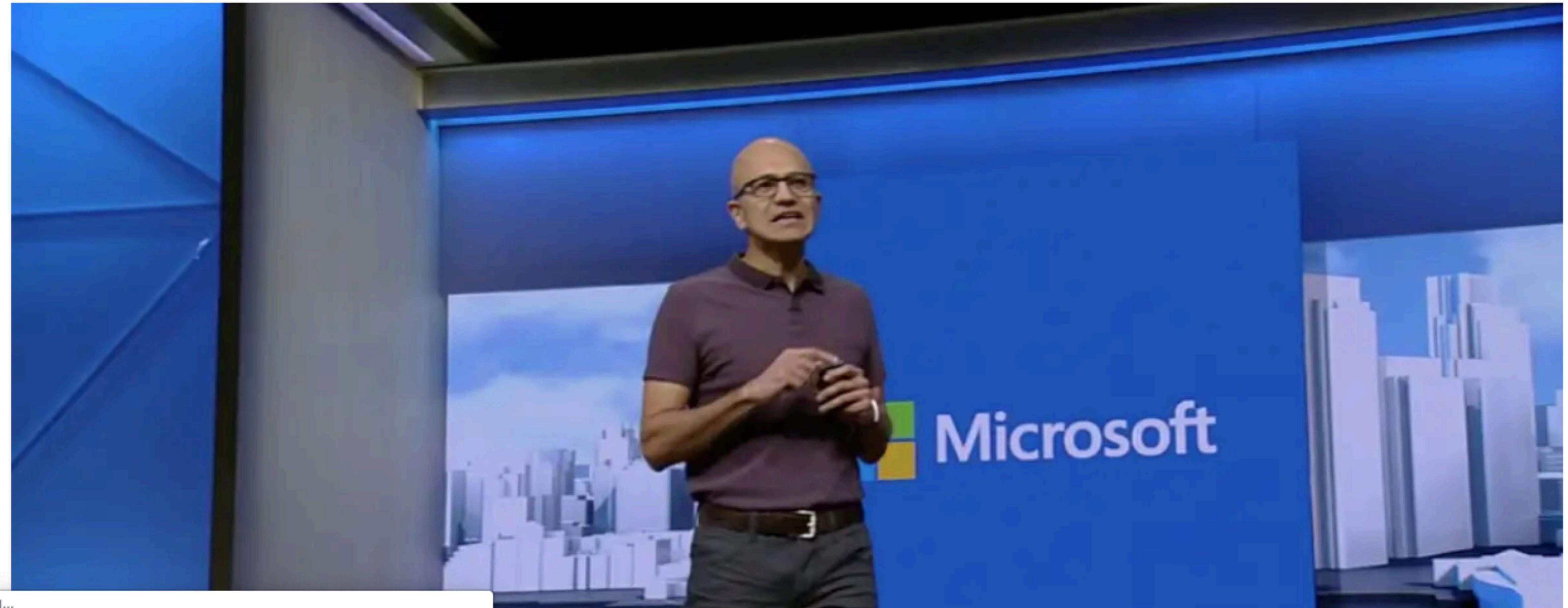
- [2021 \(1\)](#)
- [2020 \(2\)](#)
- [2019 \(2\)](#)

**THE REALITY IS THAT YOU DON'T KNOW IN ADVANCE WHICH PROJECTS OR FEATURES WILL SUCCEED: THE AVERAGE FAILURE RATE IS ANYWHERE BETWEEN 50 AND 80%.**



# Microsoft experimented with a 4-day workweek, and productivity jumped by 40%

Lisa Eadicicco Nov 4, 2019, 8:39 AM



**EVALUATING WELL-DESIGNED AND EXECUTED  
EXPERIMENTS THAT WERE DESIGNED TO IMPROVE A  
KEY METRIC, ONLY ABOUT 1/3 WERE SUCCESSFUL AT  
IMPROVING THE KEY METRIC!**

**RONNY KOHAVI, “ONLINE EXPERIMENTATION AT MICROSOFT”**

# HYPOTHESIS-DRIVEN DEVELOPMENT

Hypothesis-driven development

We believe *<this capability>*

Will result in *<this outcome>*

We will have confidence to  
proceed when

*<we see a measurable signal>*

@barryoreilly, <http://barryoreilly.com/2013/10/21/how-to-implement-hypothesis-driven-development/>

# GENERATING EXPERIMENTS

# ELEMENTS OF GOOD EXPERIMENTS

- Time-based
- Measurable
- Single variable (Control!)
- Provable (and disprovable) hypothesis
- Psychological safety to fail
- Learning is the outcome

# HOW WOULD YOU DESIGN THESE AS EXPERIMENTS?

- Company policy to give a bonus to employees who refer new hires
- Pairing on work or ensemble work
- A new team-leadership model
- Work-from-home policy
- [Your own example]

# EXPERIMENT CANVAS

**Title:**

**Owner:**

<b>Context</b>	Our problem is...	<b>Actions</b>	To prove/disprove the hypothesis we will...
<b>Hypothesis</b>	We believe we can solve it by...	<b>Results</b>	We will declare success or failure when...
<b>Rationale</b>	We believe this because...	<b>Follow-up</b>	As a result of success or failure we will...



Karl Scotland Ltd

Experiment A3

<https://availagility.co.uk/templates>

# WHAT DID WE LEARN?

HOW DID THAT GO?

# TYPES OF FAILURE

	Preventable	Complex	Intelligent
Definition	Deviations from known process that produce unwanted outcomes	Unique and novel combinations of events and actions that give rise to unwanted outcomes	Novel forays into new territory that lead to unwanted outcomes
Common Causes	Behavior, skill, attention deficiencies	Complexity, variability, novel factors imposed on familiar situations	Uncertainty, experimentation, risk taking
AKA	Process deviation	System breakdown	Unsuccessful trial
Contexts Where Each Is Most Salient	Production line manufacturing Fast-food services Basic utilities and services	NASA shuttle program Aircraft carrier Nuclear power plant	Drug development New product design

— Amy Edmondson, *The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth*

# DESTIGMATIZING FAILURE

	Traditional Frame	Destigmatizing Frame
Concept of Failure	Failure is not acceptable.	<b>Failure is a natural by-product of experimentation.</b>
Beliefs About Effective Performance	Effective performers don't fail.	Effective performers <b>produce, learn from and share the lessons</b> from intelligent failures.
Goal	Prevent failure.	Learn fast.
Impact of Frame	People hide failures to protect themselves.	Open discussion, fast learning and innovation.

— Amy Edmondson, The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth

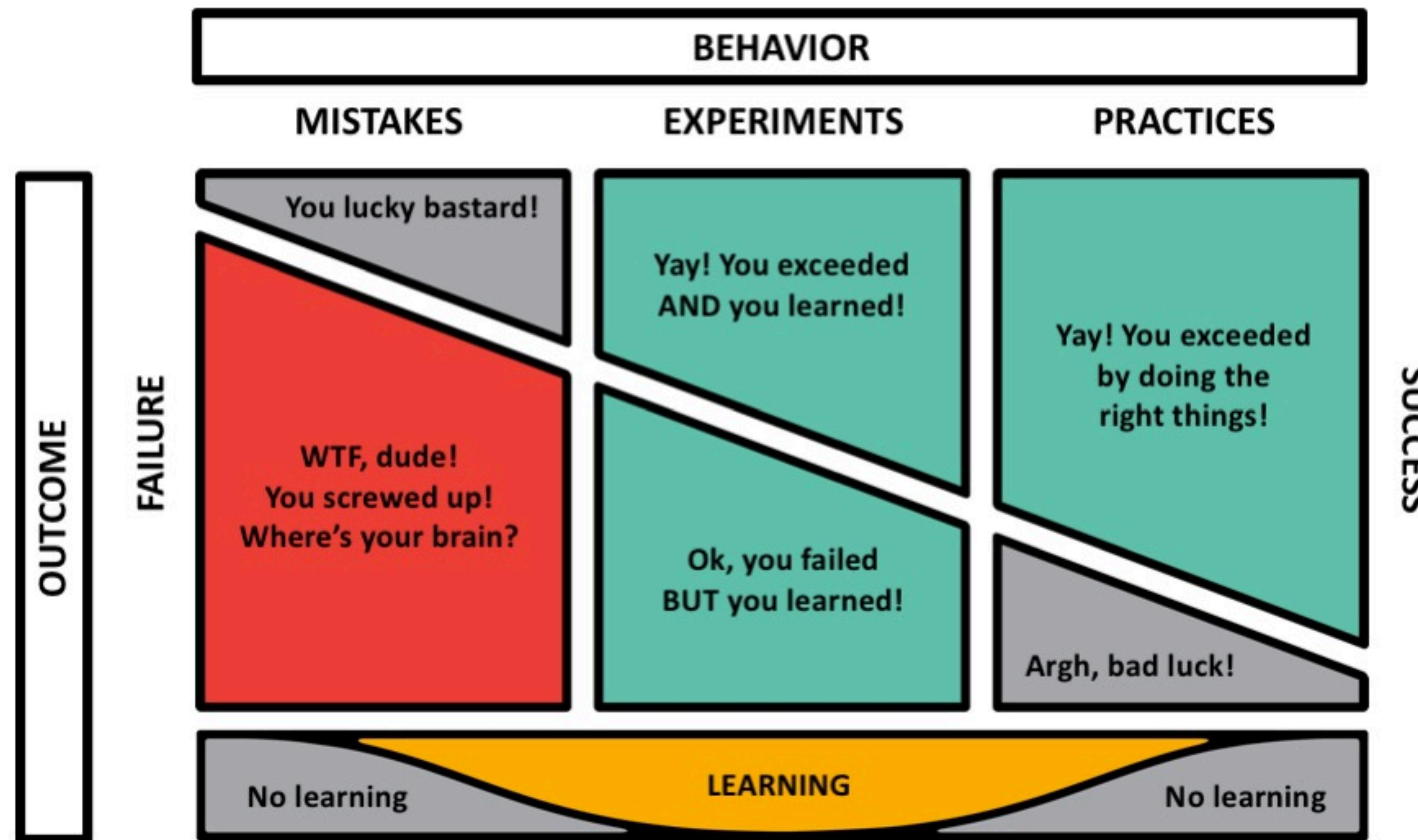


# PRODUCTIVE RESPONSES TO FAILURE

Preventable Failure	Complex Failure	Intelligent Failure
<ul style="list-style-type: none"><li>- Training</li><li>- Retraining</li><li>- Process improvement</li><li>- System redesign</li><li>- Sanctions, if repeated or otherwise blameworthy actions are found</li></ul>	<ul style="list-style-type: none"><li>- Failure analysis from diverse perspectives</li><li>- Identification of risk factors to address</li><li>- System improvement</li></ul>	<ul style="list-style-type: none"><li>- Failure parties</li><li>- Failure awards</li><li>- Thoughtful analysis of results to figure out implications</li><li>- Brainstorming of new hypotheses</li><li>- Design of next steps or additional experiments</li></ul>

— Amy Edmondson, *The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth*

# CELEBRATION GRID



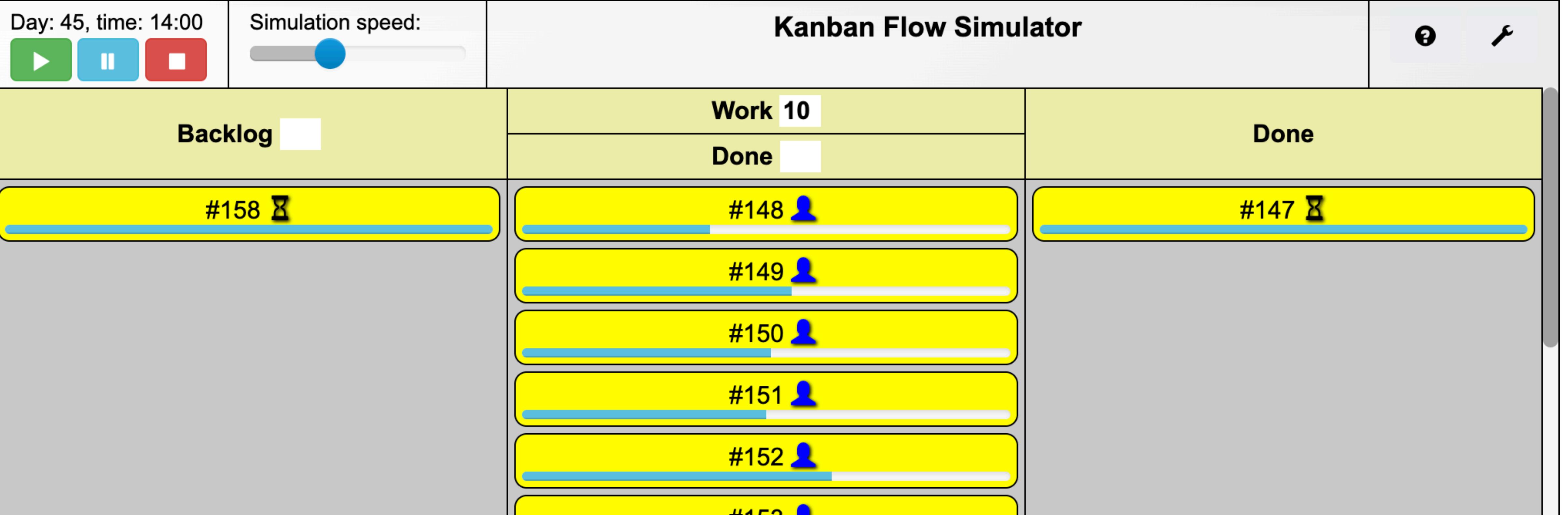
# Why Etsy engineers send company-wide emails confessing mistakes they made



# OTHER CONSIDERATIONS

- Language: “Fail Fast” → “Prove/disprove hypothesis”
- Leaders SETT the boundaries (“Safe Enough To Try”)
- Organizational experiments board
- “Goldratt User Stories”
- Finding your optimal work-in-progress limit
- Retrospective actions
- A/B tests

# FLOW SIMULATION



**BEWARE**



# HAWTHORNE EFFECT

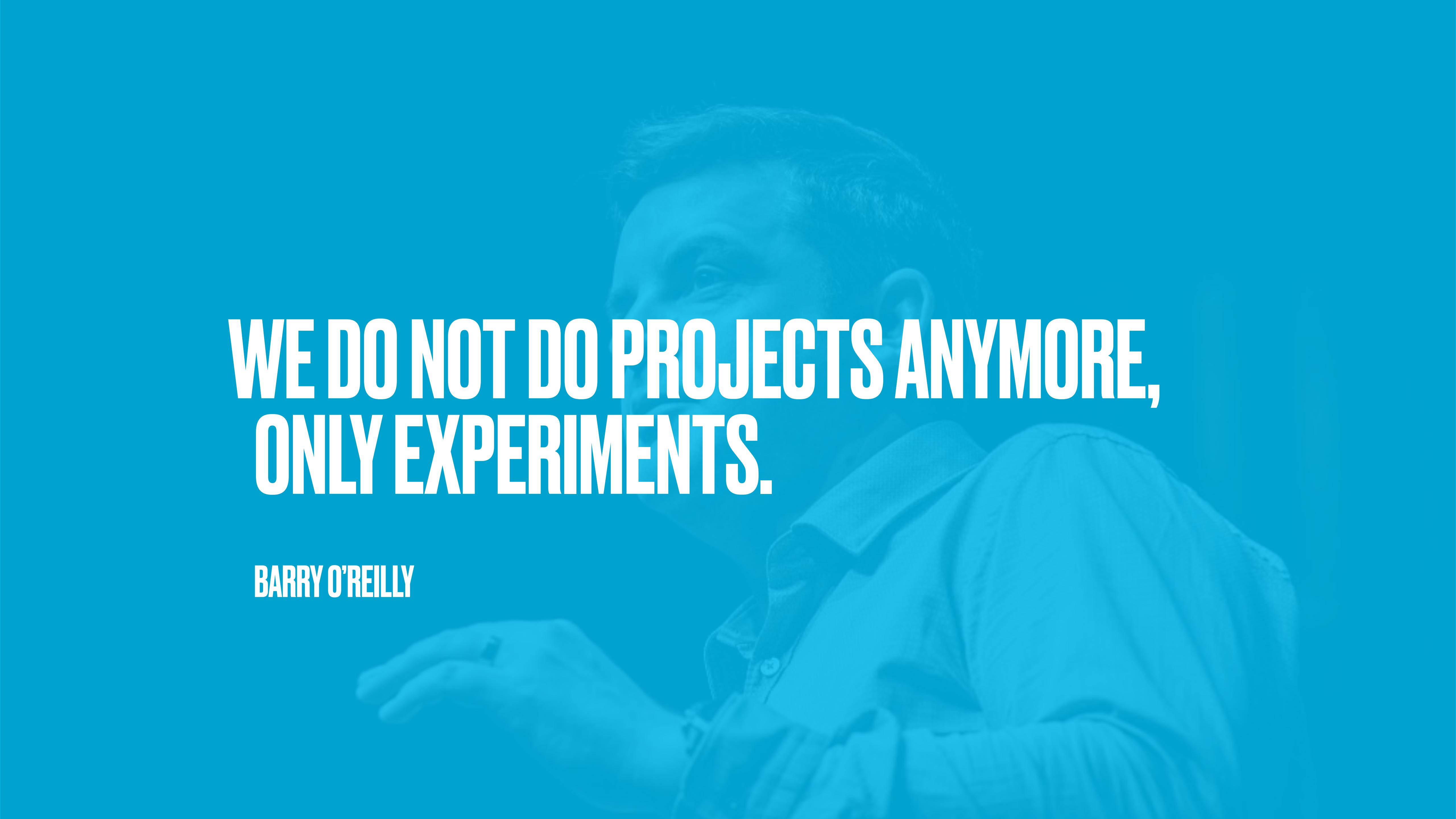
AKA Observer Effect

# SELECTING SAFE-TO-FAIL EXPERIMENTS

- 1. Any experiment must be something you can do something about and that you believe stands a chance of having a positive effect.
- 2. Secondly, it has to be a change with an observable or measurable effect – so that you can see if the change was good or bad.
- 3. The experiment must be something you believe you can dampen down if it goes wrong (ie safe to fail) or amplify the effect if it goes well.

# WHO CAN BENEFIT FROM EXPERIMENTS?

- **Product Owners:** Measurable outcomes from user stories
- **Coaches:** Impact on people, teams, departments
- **Transformation Agents:** Evolutionary organizational fitness
- **Developers**
- **Teams:** Retrospective actions
- **Organizations:** Leverage scale by running multiple experiments

A black and white photograph of a man with dark hair and a beard. He is wearing a light-colored shirt and a dark suit jacket over it. His right hand is resting against his chin, with his fingers partially hidden in his pocket. He is looking slightly to the left of the camera with a contemplative expression.

**WE DO NOT DO PROJECTS ANYMORE,  
ONLY EXPERIMENTS.**

**BARRY O'REILLY**

1. Eleusis Expeditious and Eleusis Express (<https://availability.co.uk/resources/games/eleusis-expeditious/>)
2. A3 Experiment Canvas from Karl Scotland (<https://availability.co.uk/>)
3. Safe-to-Fail Probes (<http://cognitive-edge.com/methods/safe-to-fail-probes/>)
4. How to Fail like a Pro, Freakonomics Radio: 370.
5. Unlearn by Barry O'Reilly
6. Company-wide Agility with Beyond Budgeting, Open Space & Sociocracy by John Buck and Jutta Eckstein
7. Principles of Product Development Flow by Don Reinertsen
8. How to Measure Anything by Douglas W. Hubbard
9. pretty much anything from John Cutler
10. Jurgen Appelo: <https://management30.com/practice/celebration-grids/>
11. The Surprising Power of Online Experiments (<https://hbr.org/2017/09/the-surprising-power-of-online-experiments>)
12. Online Controlled Experiments: Lessons from Running A/B/n Tests for 12 years (<http://bit.ly/KDD2015Kohavi>)
13. Thinking in Bets by Annie Duke

14. How to Implement Hypothesis-Driven Development (<https://barryoreilly.com/how-to-implement-hypothesis-driven-development/>)
15. Beyond the Black Hole: Product Management for Continuous Delivery, Elizabeth Ayer, FlowCon 2019
16. The Fearless Organization by Amy Edmondson
17. Don't Be a Ditka, Dan Vacanti (<https://www.infoq.com/presentations/prioritization-scarcity-stress-uncertainty/>)
18. Goldratt User Stories (<https://yorkesoftware.com/2018/03/09/goldratt-user-stories/>)
19. "Online Experimentation at Microsoft" Kohavi et al (<http://stanford.io/130uW6X>)
20. Robert Hooke, first Curator of Experiments (<https://royalsociety.org/science-events-and-lectures/2003/summer-science/mr-hooke/>)
21. <https://itrevolution.com/the-three-ways-principles-underpinning-devops/>
22. Complexity Thinking Dimitar Bakardzhiev
23. Mike Burrows, Agendashift
24. Adam Light
25. <https://www.katatogrow.com>
26. <https://docondev.com/blog/2019/11/8/the-experiment-canvas>
27. [mattphilip.wordpress.com](http://mattphilip.wordpress.com)

