Lecture 12: Computational Cognitive Modeling

Causal Interventions, Active Learning, and Bounded Rationality

email address for instructors:
instructors-ccm-spring2022@nyuccl.org

course website:
https://brendenlake.github.io/CCM-site/

Today: Learning by doing!

Three Parts:

1. **Interventions** - The logic of experimentation
2. **Active learning** - learning by doing!
3. **Resource Rational Models** - how limited cognitive minds and implement complex inference schemes

Part 1: Interventions and causation

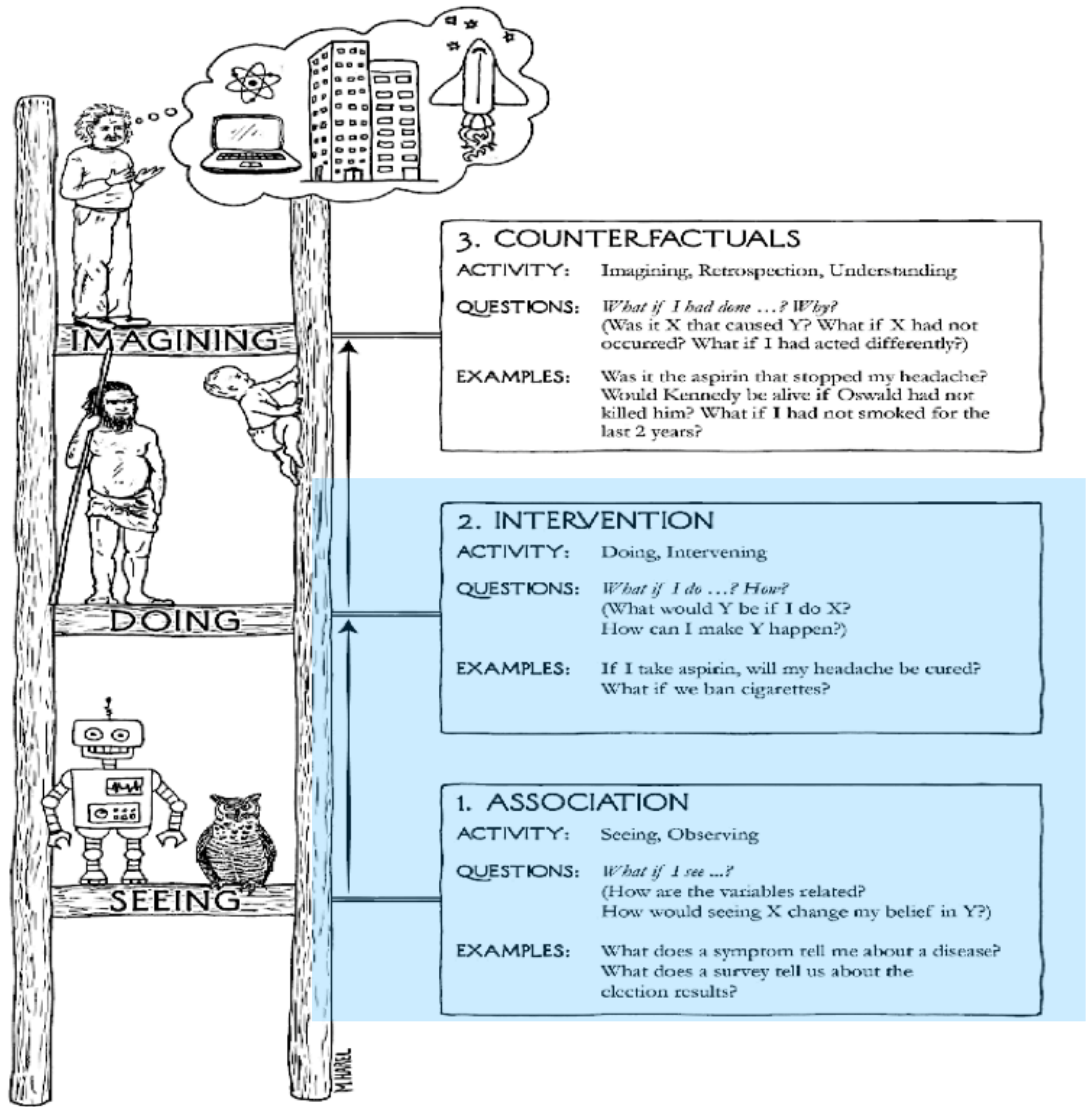
The Ladder of Causation

- Pearl (2018) articulates value of causal models for reasoning in terms of enabling a “ladder” (or hierarchy) of forms of inference going beyond associative inferences

Bottom rung:

- Traditional statistical methods

- Associative inferences e.g.:
  - The *probability* I have a symptom given that I have a disease
  - The *probability* I have a disease given that I have a symptom
  - The *chance* of a particular election result given the recent poll
  - The *odds* that a sentence contains the word “be” given that it contains the word “the”
  - The *association* between a personality measure performance in some task



Intervention

- Causal judgments are fundamentally about "difference making" (Lewis, Woodward, Hume)
  - That is, about what will happen *if you do something* (that might not have otherwise occurred)
- Not “Do joggers have lower blood pressure on average?” but “Will I lower my blood pressure **if** I take up jogging?”
- It’s a different question. **The answer turns on a different kind of evidence.**
- Not data obtained by comparing natural joggers to natural non-joggers…
  - ….but obtained by *forcing* some non-joggers to jog, and/or *forcing* some joggers to not jog
  - Manipulating something in the situation and seeing what difference it makes
  - In other words, running an experiment!

Intervention in CBNs

- All of these variables are pairwise associated ☞

- But each variable actually has only 0-2 causes

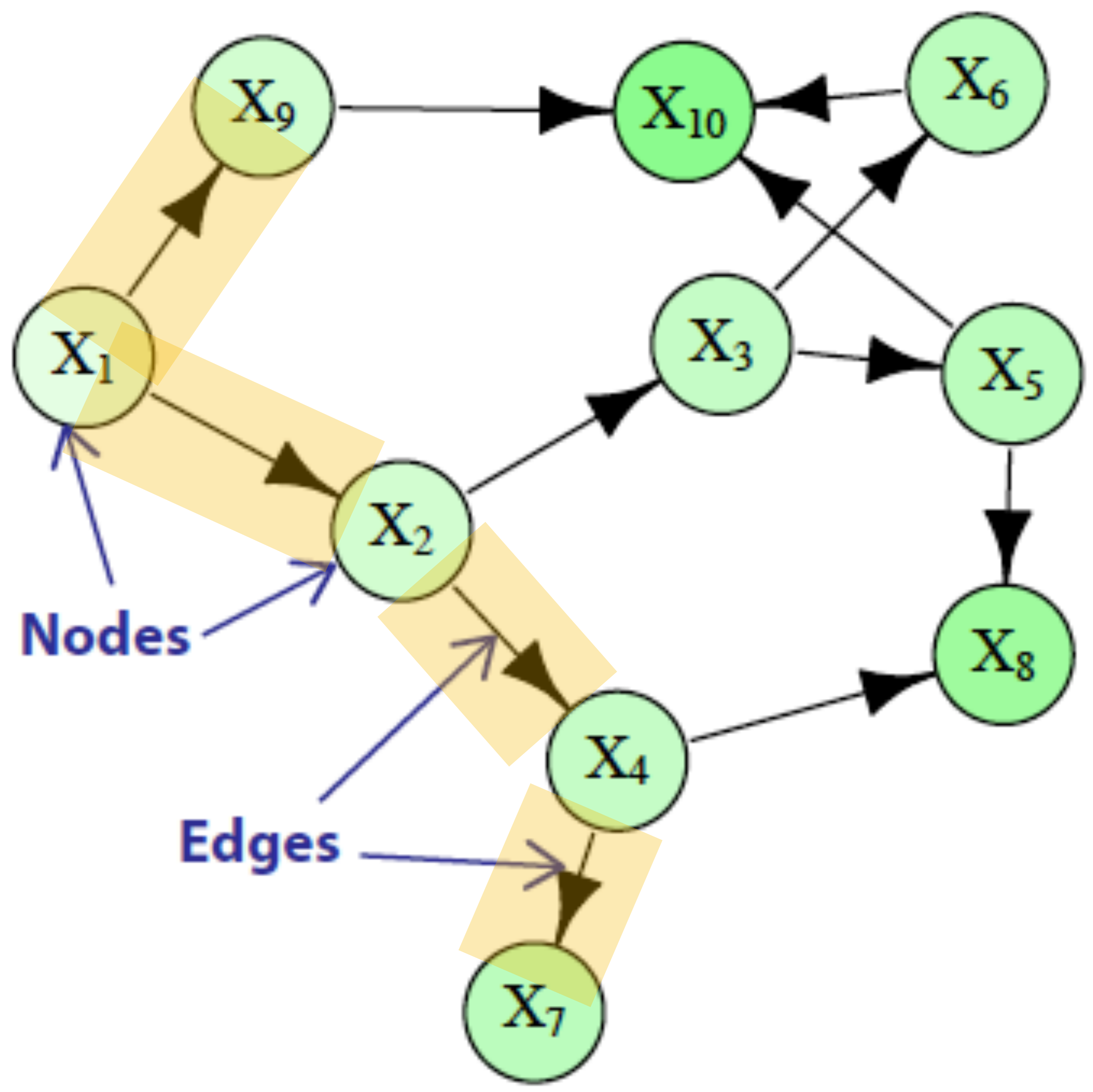
  - *Ceterius paribus,* the more variables in a CBN, the *greater* the chance that an an observed (unconditional) association is due to shared ancestor rather than a direct connection

    - E.g. X_9's parent is X_7's great grandparent

- Markov condition: *Variables are (only) independent of their non descendants conditional on their parents*

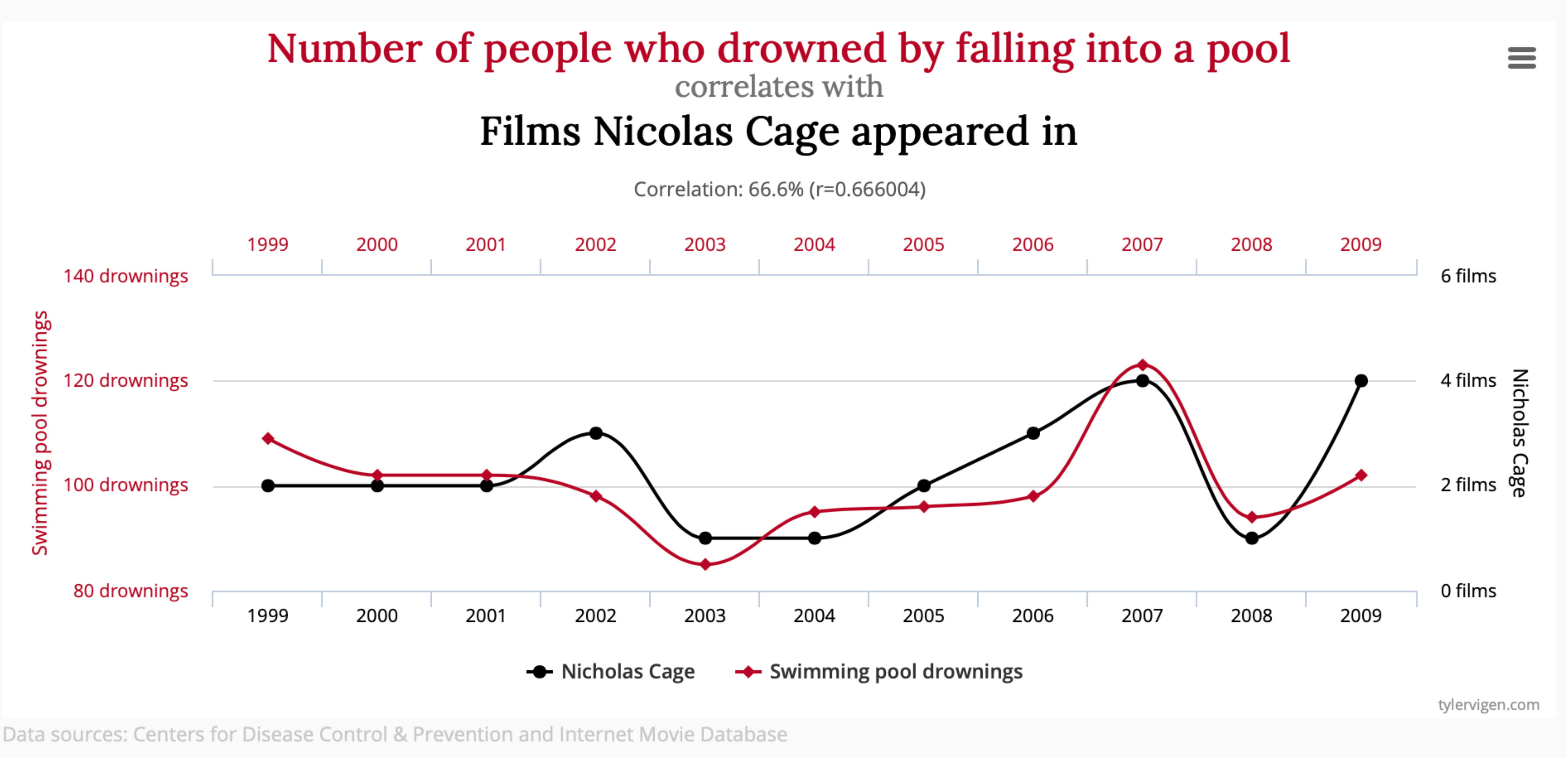
  - Question: What would it take to make X_9 and X_7 independent here?

  - Answer: Observing X_1, X_2 or X_4



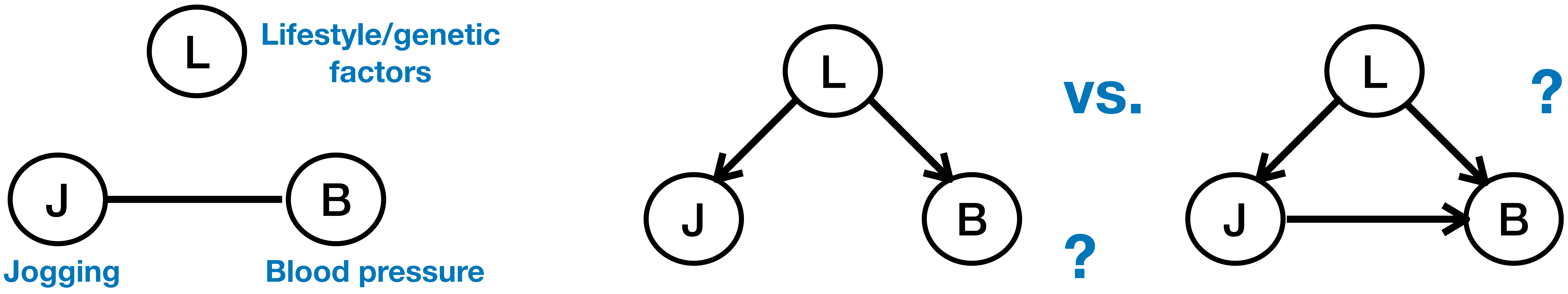
Intervention in CBNs

- **The vast majority of associations we observe are spurious!**
  - Divorce rates in Main ~ Per capita consumption of margarine (r=.993***)
  - Annual deaths by drowning in swimming pools ~ Nicholas Cage Movies (r=.666***)
  - Letters in Scripp's National Spelling Bee ~ Number of people killed by venomous spiders (r=.806***)
  - US crude oil imports from Norway ~ Drivers killed in collisions with trains (r=.955***)



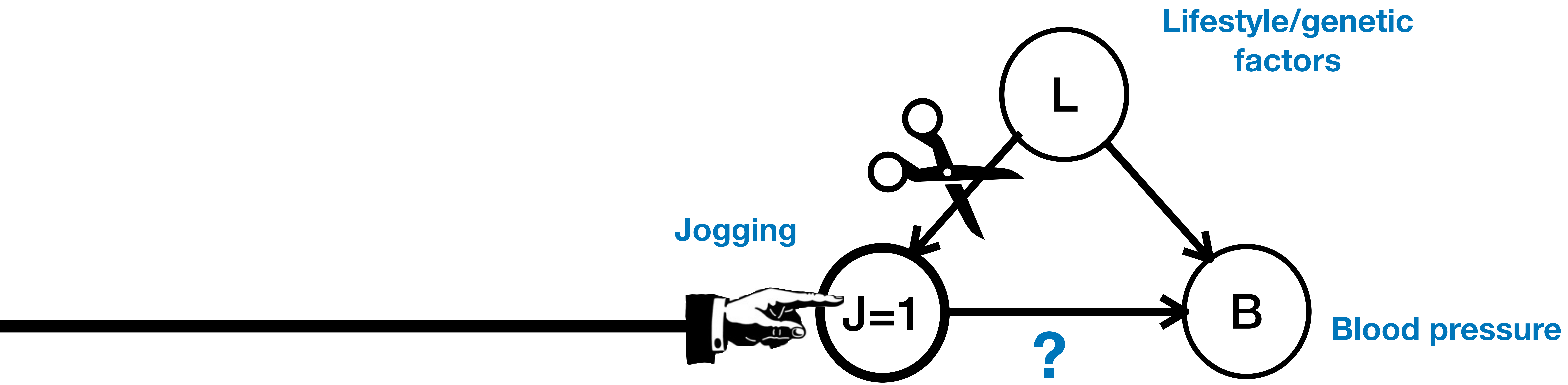
http://www.tylervigen.com/spurious-correlations

Interventions



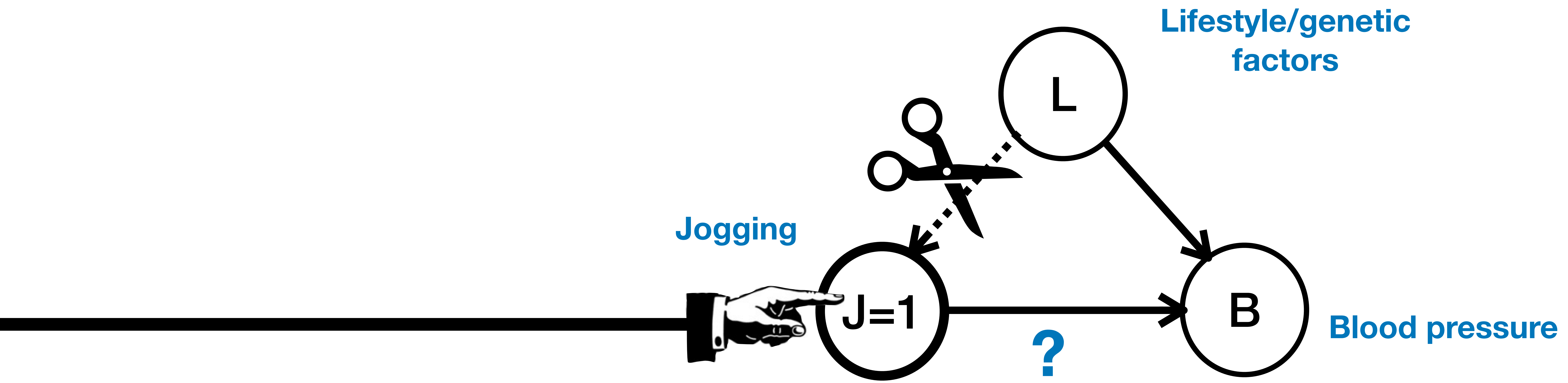
- Pearl (2000) incorporated idea of *evidence produced by our actions* through notion of an “intervention”
- **Intuition:** When you manipulate something, roughly speaking, you are reaching into the system and changing a variable’s value
  - By setting a variable to a specific value, you override whatever value it would have taken naturally
- This temporarily *disconnects* “intervened on” variables from their normal causes

Interventions as graph surgery



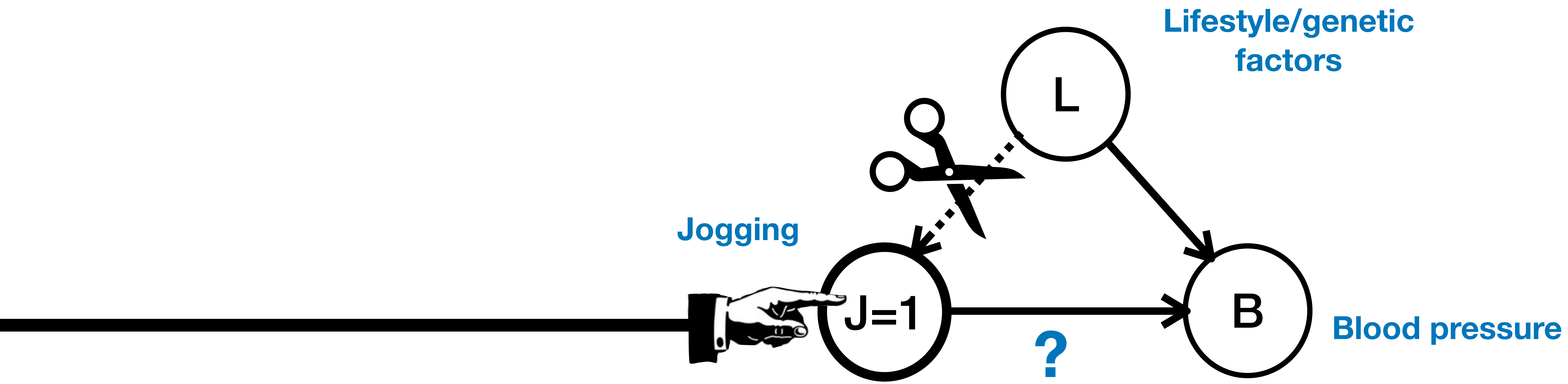
- If we want to know if jogging affects blood pressure
- We can make someone (or many people) jog and see if their blood pressure changes relative to baseline
  - or equally, compare people we have forced to jog against people we have forced not to jog
- We express this experiment in the CBN framework by drawing a new edge from outside the graph
  - indicating our influence comes from outside or “*exogenous*” to the model (as opposed to “*endogenous*" = within the model)

Interventions as graph surgery



- Our intervention temporarily severs all normal incoming connections (here the influence of our lifestyle and genetics on probability of being a jogger)
- Graphically this is represented by “graph surgery” — i.e. we have bypassed jogging’s normal causes
- **Downside**: this jogging behaviour is now uninformative about lifestyle & genetic factors (because we made them do it)
- <u>**Upside:** Now we can interpret any change to blood pressure as a causal effect!</u>

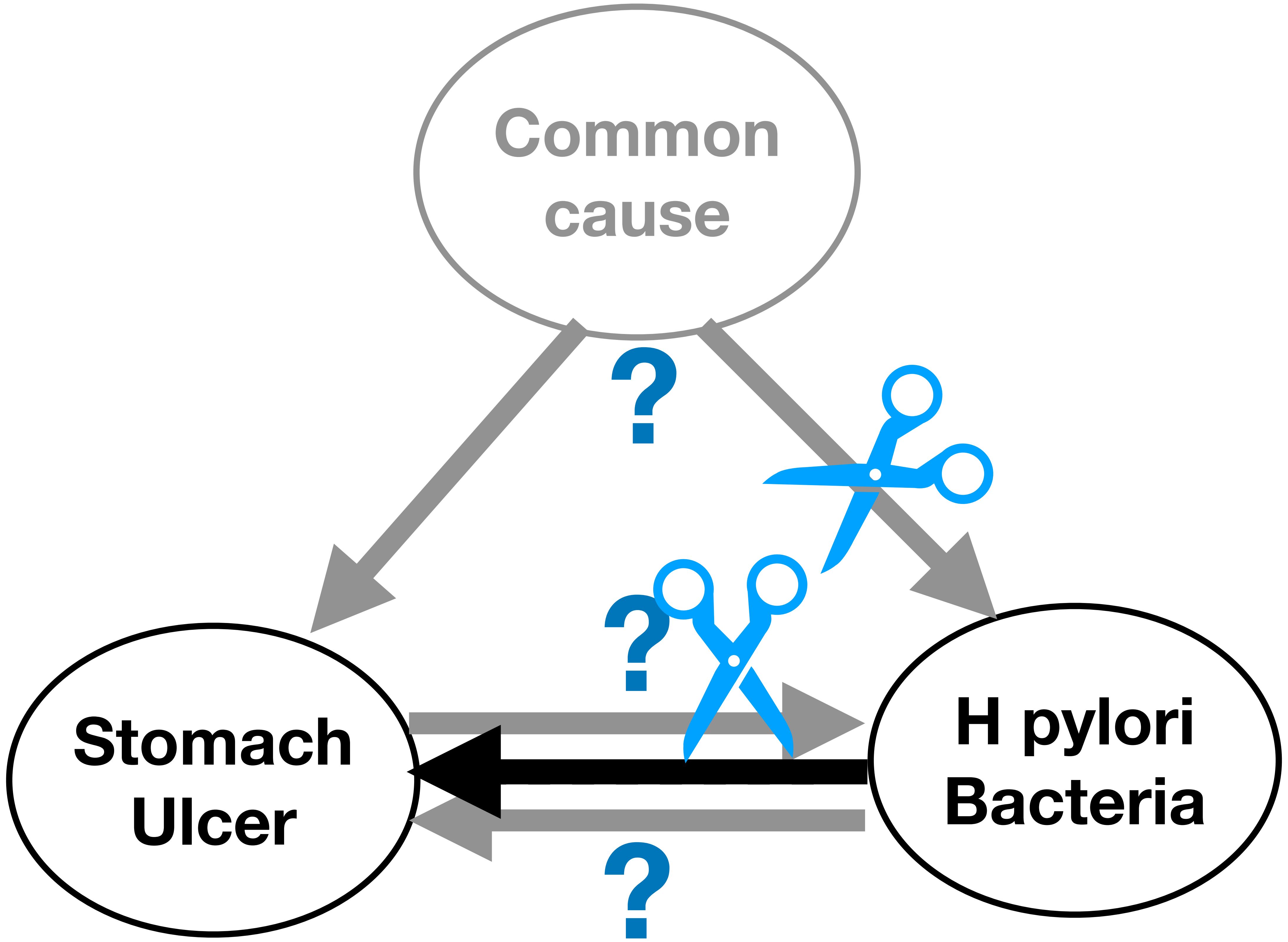
'Do' calculus

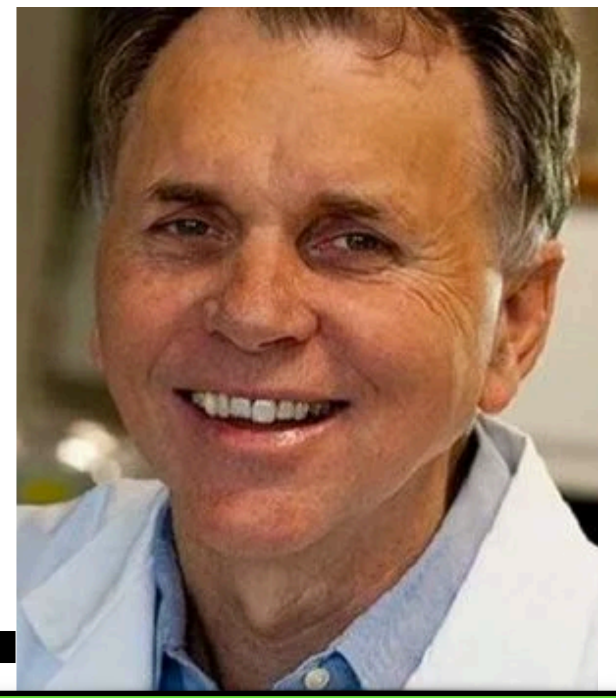


- Formally we write this using Pearl's 'Do[.]' operator
- Instead of observing a sample of $P(B|J)$ we are now observing a sample of $P(B|\mathrm{Do}[J])$
- If we find a statistical difference, i.e. $P(B|J) \neq P(B|\mathrm{Do}[J])$ for any level of B or J
  - Or equally if $P(B|\mathrm{Do}[J]) \neq P(B|\mathrm{Do}[\neg J])$ for any level of B or J
- Then we can conclude that $J \rightarrow B$
  - i.e. that jogging *causally influences* blood pressure

Real world example

Example: Established correlation between having ‘h pylori’ bacteria and stomach ulcer **but what is the causality?**





Barry Marshall
Australian physiologist

Experimental interventions

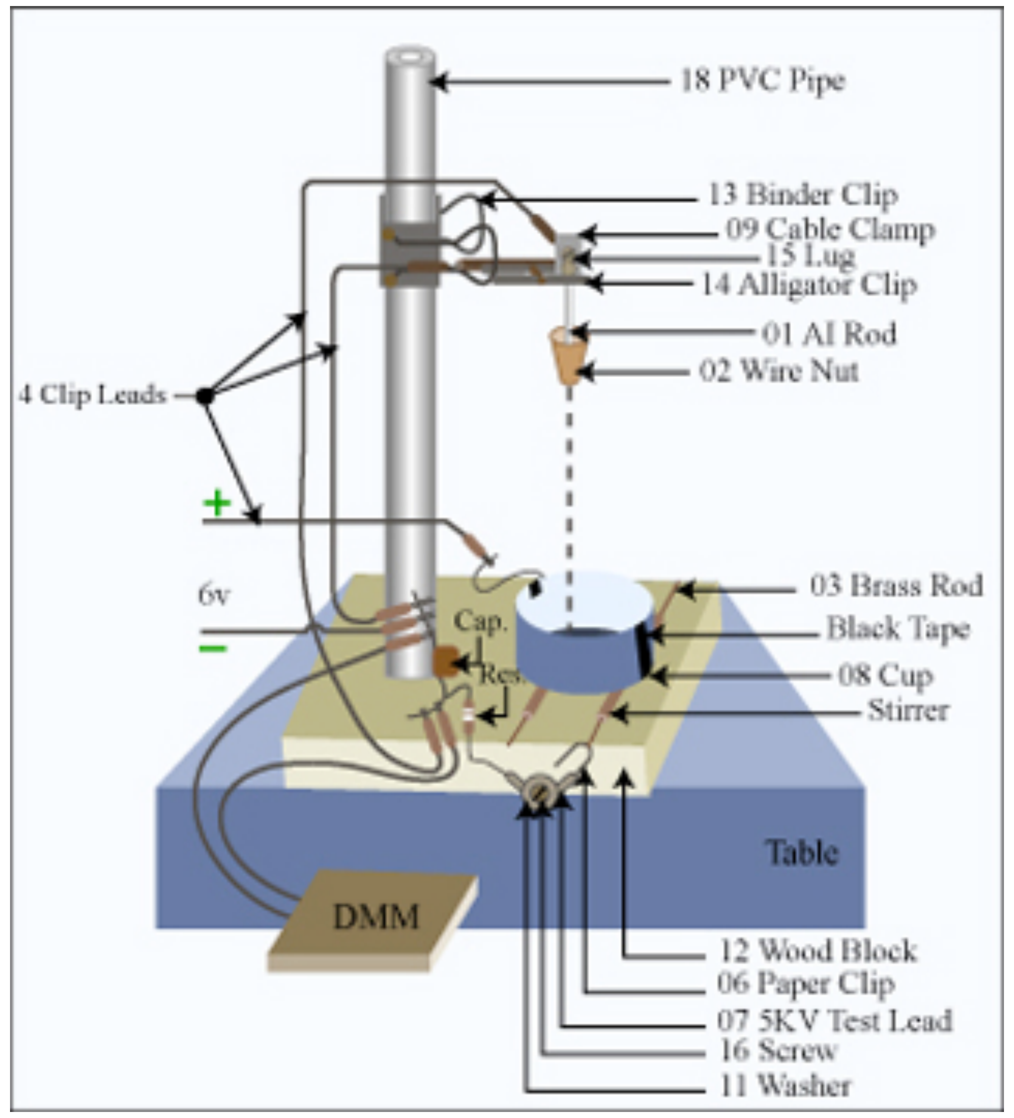
- Primary mechanism of science - Intervene systematically on world, bringing about atypical situations that reveal causality
  - Fire some particles at one another
  - Mix some stuff together
  - Assign subjects to different groups/conditions
- Repeat (or control) procedure enough to overwhelm statistical noise
- What was unusual about Barry Marshall's experiment?
  - N=1

CONTROL GROUP

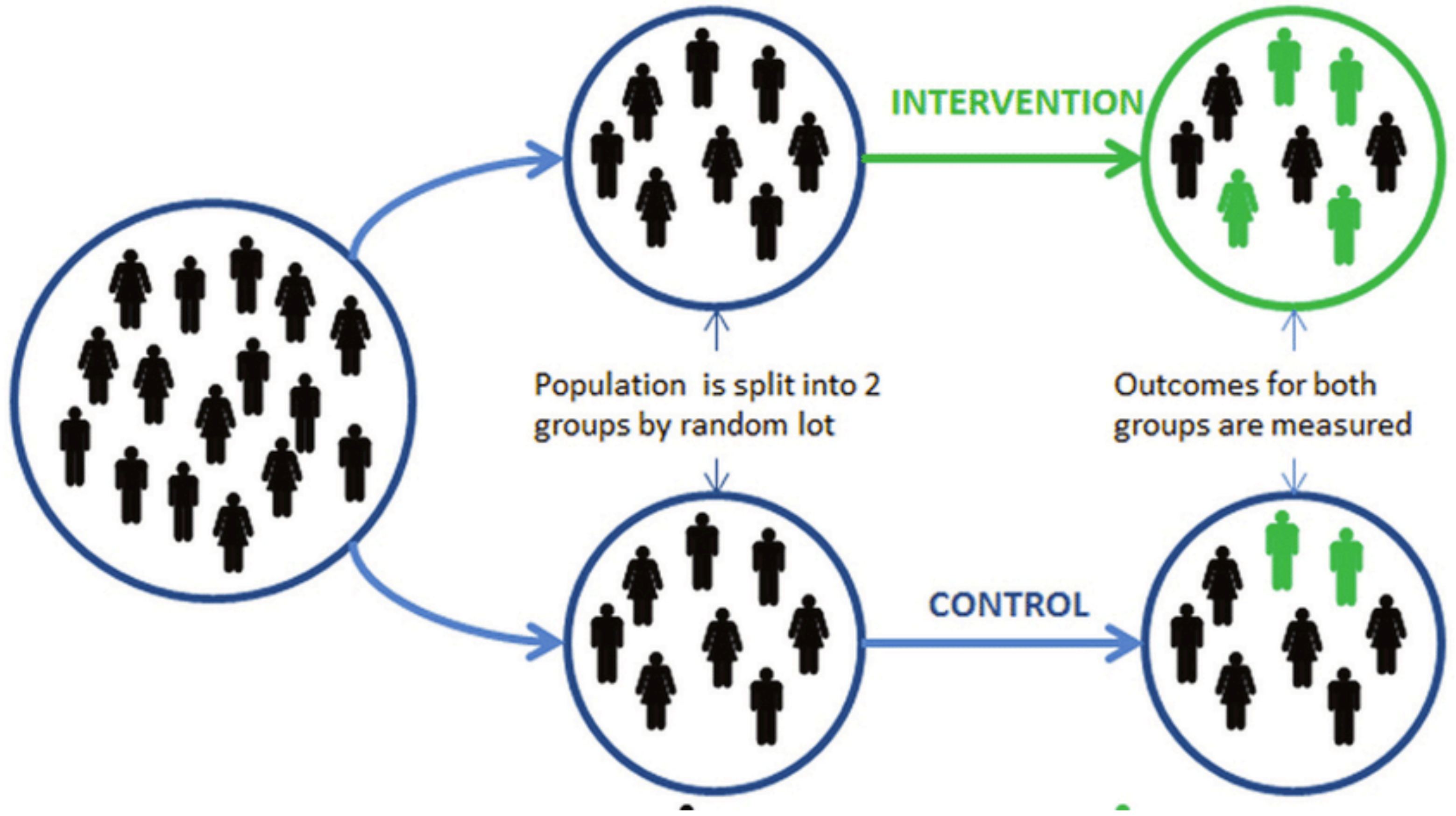
OUT OF CONTROL GROUP.

Interventions as experiments

- Psychology and medicine are particularly tough domains for causal inference
  - Universal causal effects often small & noisy relative to individual differences
  - And numerous factors that likely to produce spurious effects if allowed to “corrupt” interventions
- A typical “natural science" protocol seeks to hold every conceivable confounding factor constant *except* what is being manipulated and what is being observed
- This is not generally possible for psychology experiments
- Nor is a homogenous sample desirable, since we want results that generalise to the heterogeneous population



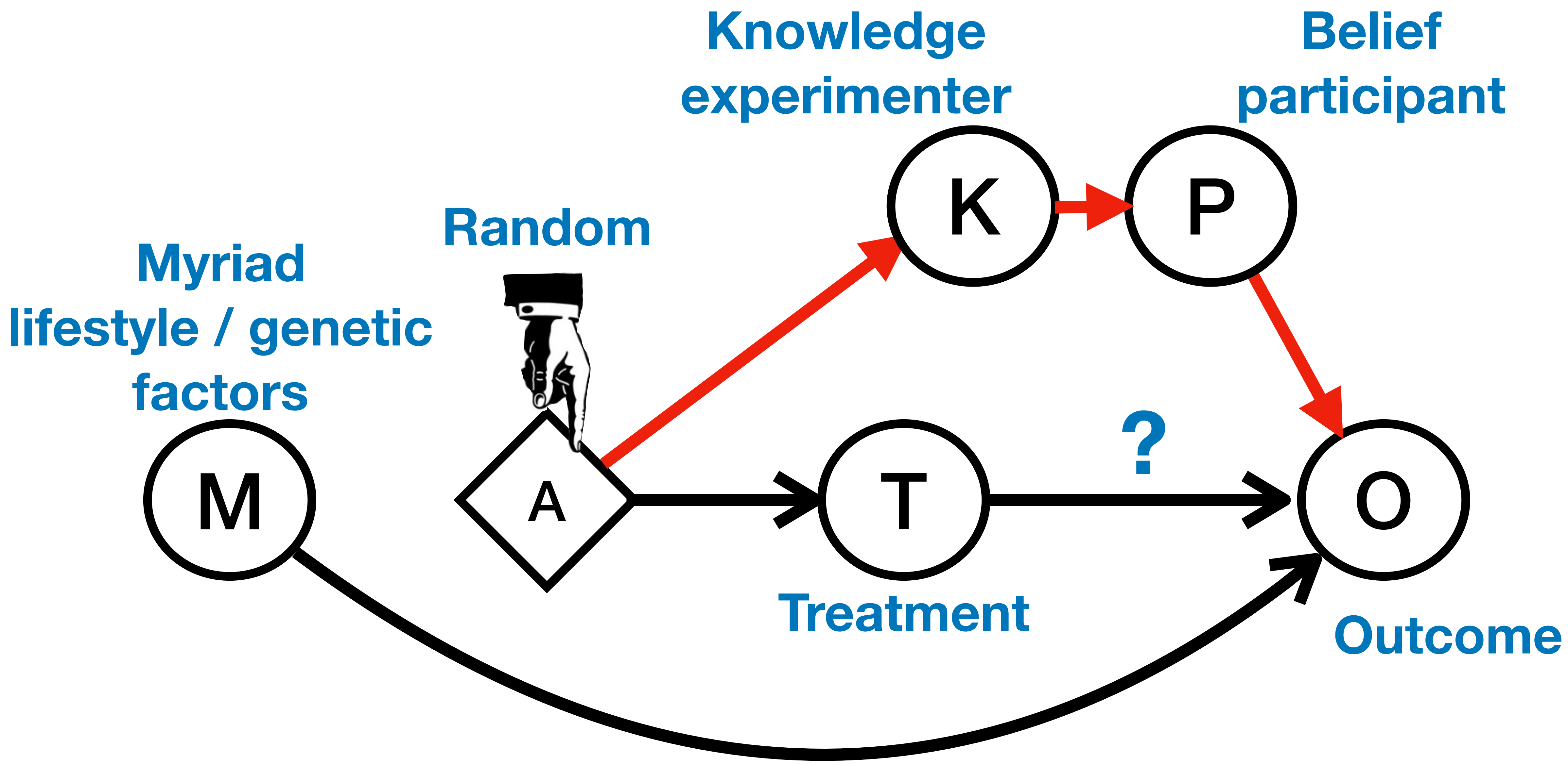
A physics experiment



A psychology experiment

Interventions as experiments

- **Problem:** Intervention protocols can easily be leaky
- **Example:** You want to know if a new treatment is effective
  - You randomly assign your participants ✔, you blind them to which condition they are in ✔ but you are aware of their assignment as you introduce the study to them
  - This introduces another potential leak — Experimenters' beliefs affect participants' beliefs affecting outcomes & systematically correlated with the treatment



- **How can we ensure that experimenters' beliefs about condition assignment do not affect experiment outcomes**
- **Answer:** <u>Double blind, so condition is hidden from both experimenter and participant until after the study</u>

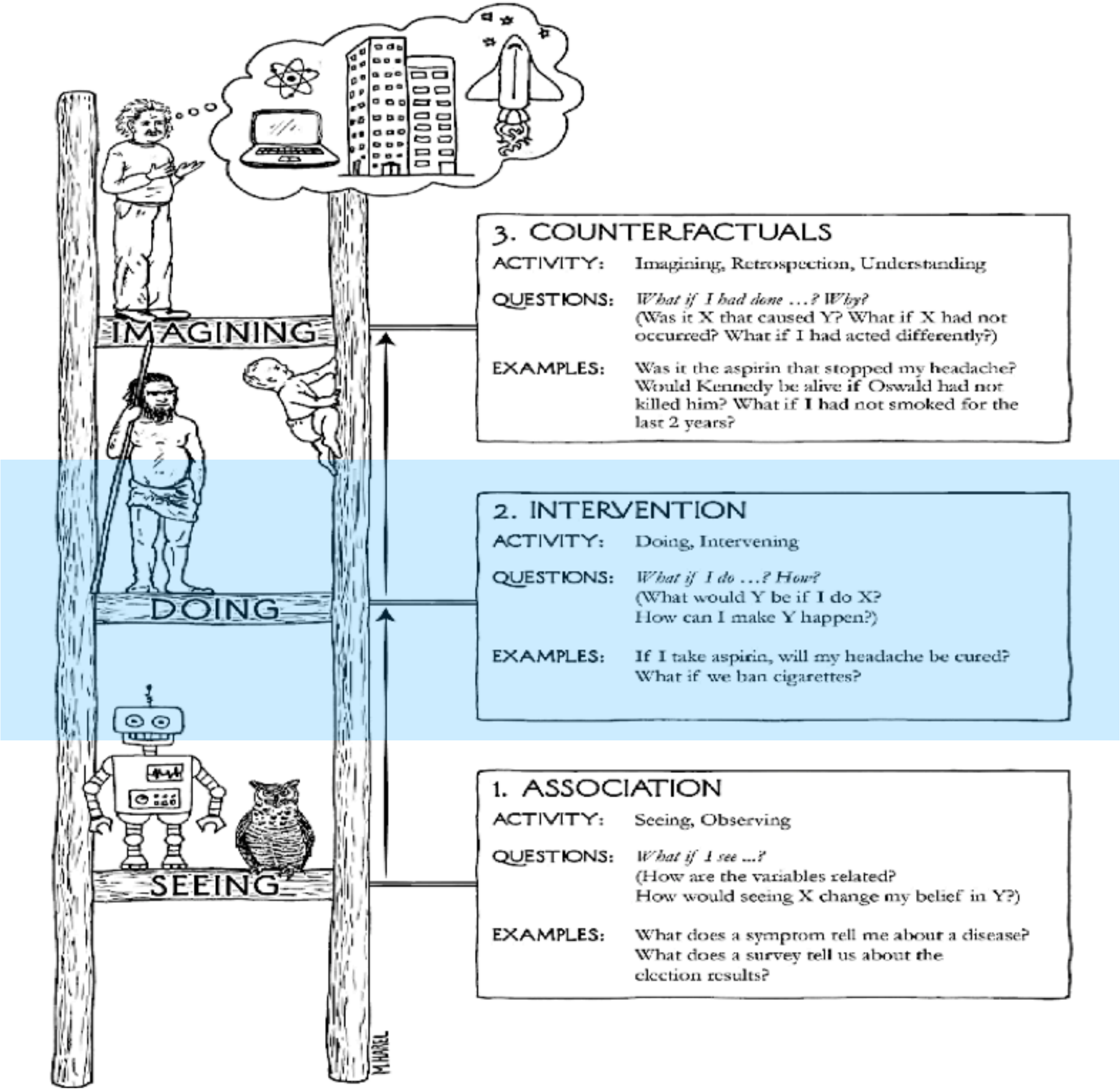
Interim Summary

- *Interventions* are manipulations of a system that can reveal causal directionality by disconnecting variables from their normal causes
- *Causal Bayesian Networks* + *‘Do*' *calculus* provide a handy way to formalise this
- They clarify why interventions must be “surgical” to be informative
  - I.e. must “set” the relevant variable without disturbing normal causes or introducing “leaks”

Interim Summary

- In science, we normally call our interventions "experiments"
  - Next we'll touch on how CBNs allow us to achieve *Optimal Experimental Designs* (OED, Atkinson & Donev, 1992)
- Experiments use protocols that reproduce setting many times with minimal variability apart from the factor being manipulated
- In psychology experiments, typically there are many factors that cannot be fixed
  - but we can temporarily surgically detach from them by randomisation
- Randomised Controlled Trials (RCTs) are held to be the "gold standard" of scientific evidence (Cartwright, 2011)

Interventions in individual cognition



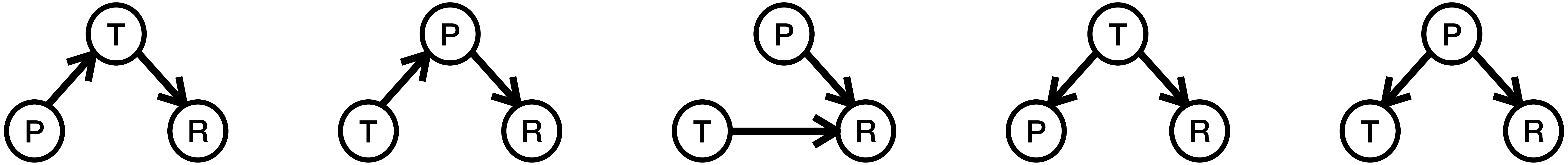
- Interventional evidence tells us the consequences of “doing” rather than just observing
- Seems important… We constantly “do” things!
- Conceptually/theoretically related to Reinforcement Learning
- Analogous to how causal assumptions (i.e. powers and base rates) drive human judgments of structure from contingencies
- **Do causal intervention principle drive learning from our own actions?** First study to look at this was Lagnado & Sloman (2002)…

Recall:

- **Classical conditioning:** Building associations from observed contingencies
- **Operant conditioning / Reinforcement Learning (RL):** Associating outcomes with actions

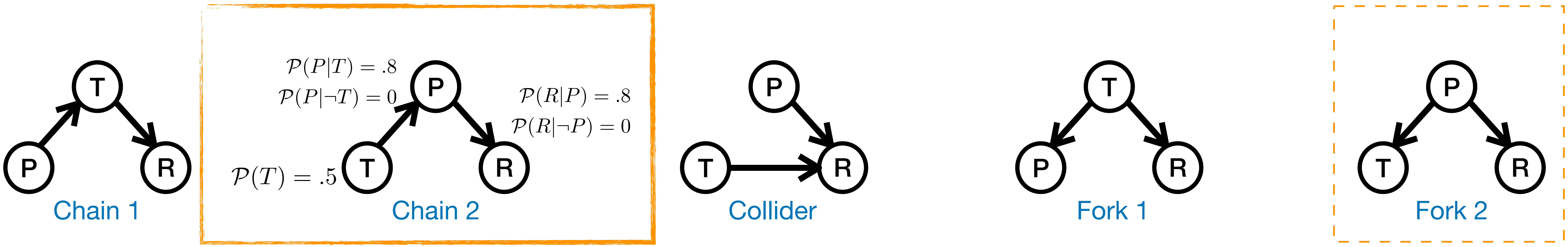
Lagnado & Sloman (2002)

- Participants (N=33) infer the causal structure relating 3 variables in 2 within-subjects conditions:
  - Condition 1: based on 50 observations
  - Condition 2: based on 50 freely chosen interventions
- Task order and cover story counterbalanced
  - Cover story 1: **T**emperature (low/high), **P**ressure (low/high), **R**ocket launch (no/yes)
  - Cover story 2: Acid level (low/high), Ester level (low/high), Perfume produced (no/yes)
- Participant learning probed through conditional probability judgments + forced choice between 5 possible causal structures



Chain 1 Chain 2 Collider Fork 1 Fork 2

Stimuli and predictions



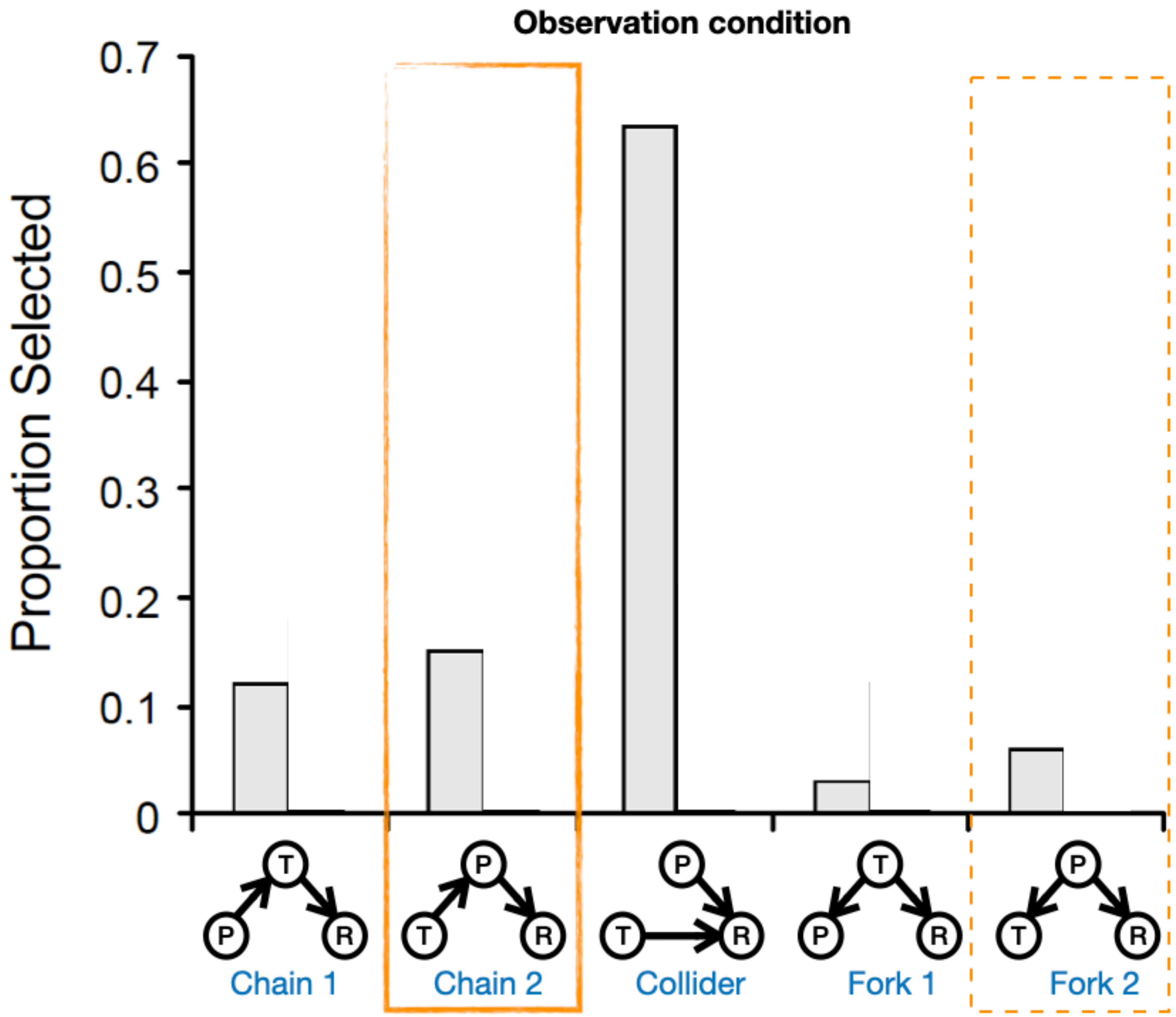
T	P	R	Count / 50	$P(T,P,R)$
0	0	0	25	0.5
1	0	0	5	0.1
0	1	0	0	0
1	1	0	4	0.08
0	0	1	0	0
1	0	1	0	0
0	1	1	0	0
1	1	1	16	0.32

Observational condition data

- Data was actually produced by Chain 2, so $T \perp\!\!\!\perp R|P$
- Which is (observationally) *Markov equivalent* to Fork 2
  - But all are distinguishable with interventional evidence (Condition 2)
  - i.e. $\mathcal{P}(P|\mathrm{Do}[T]) \neq \mathcal{P}(P|\mathrm{Do}[\neg T])$ but $\mathcal{P}(T|\mathrm{Do}[P]) = \mathcal{P}(T|\mathrm{Do}[\neg P])$
  - What did participants think?

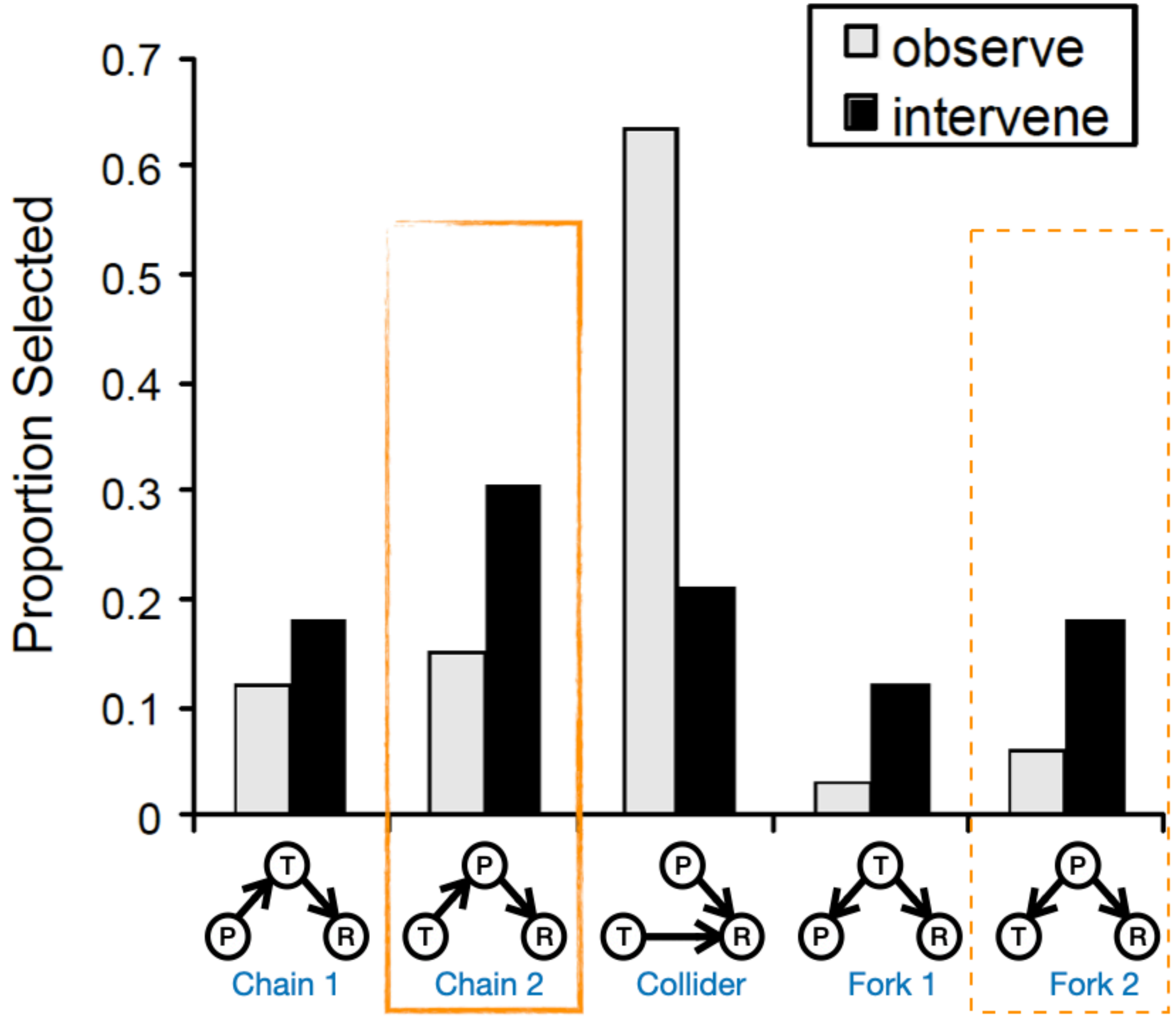
Results - Observation condition

- In **observation** condition, structure judgments were poor!
- Participants overwhelmingly favoured collider, despite opposite statistical dependencies!
- Suggests insensitivity to subtle observational statistics (dominant data were [0,0,0] 50% and [1,1,1] 32%)
  - Although see Rothe et al (2018) for recent demonstration of successful observational structure induction
- Do participants find *simultaneous presentations* (i.e. P&T at same time) inconsistent with chain?



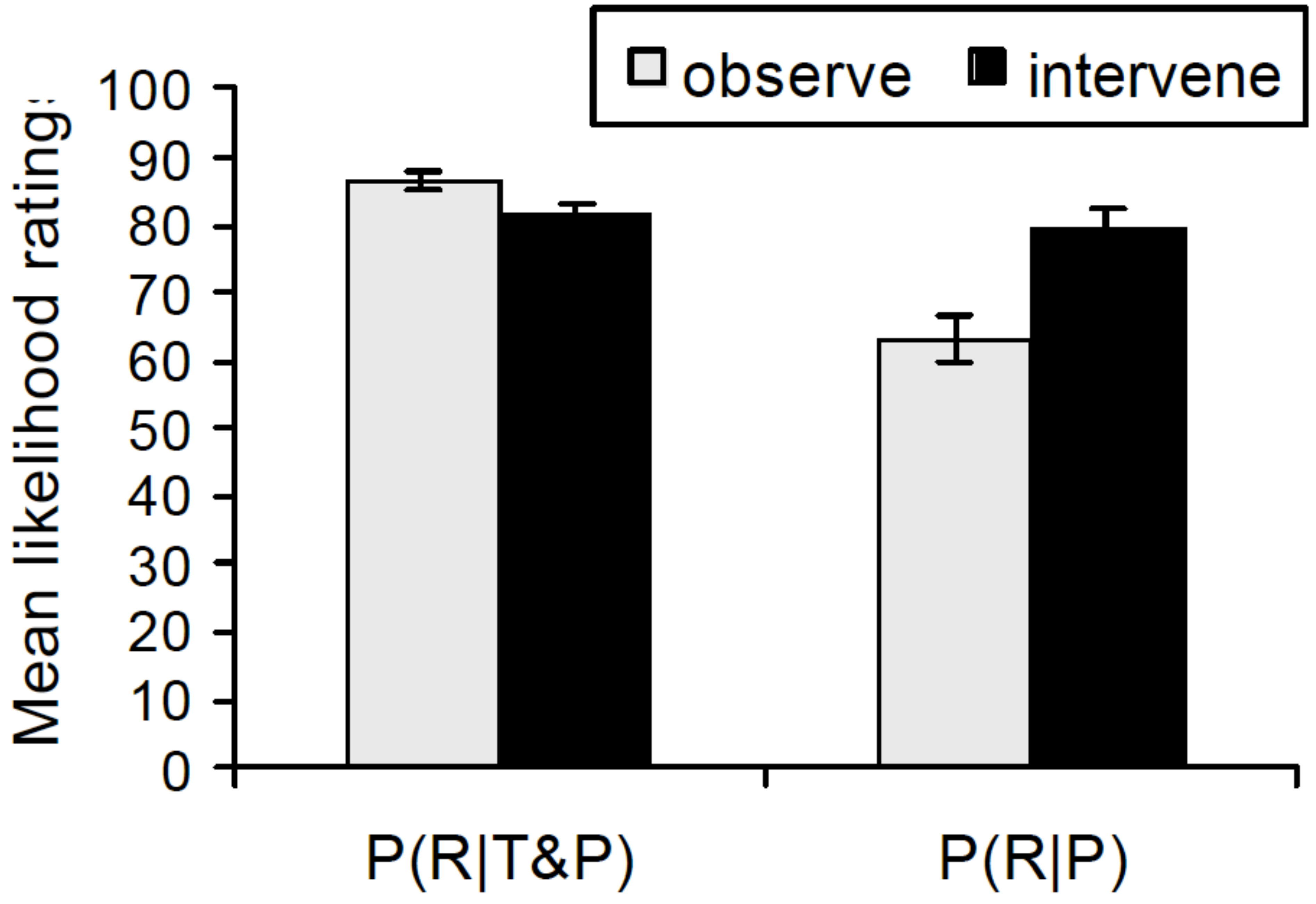
Results - Intervention condition

- In **intervention** condition, structure judgments improved
- Modal judgment now correct
  - Although… intervention judgment pattern statistically indistinguishable from random responding
  - However, plenty of evidence for stronger, more normative, interventional inferences



Results - Conditional probability judgments

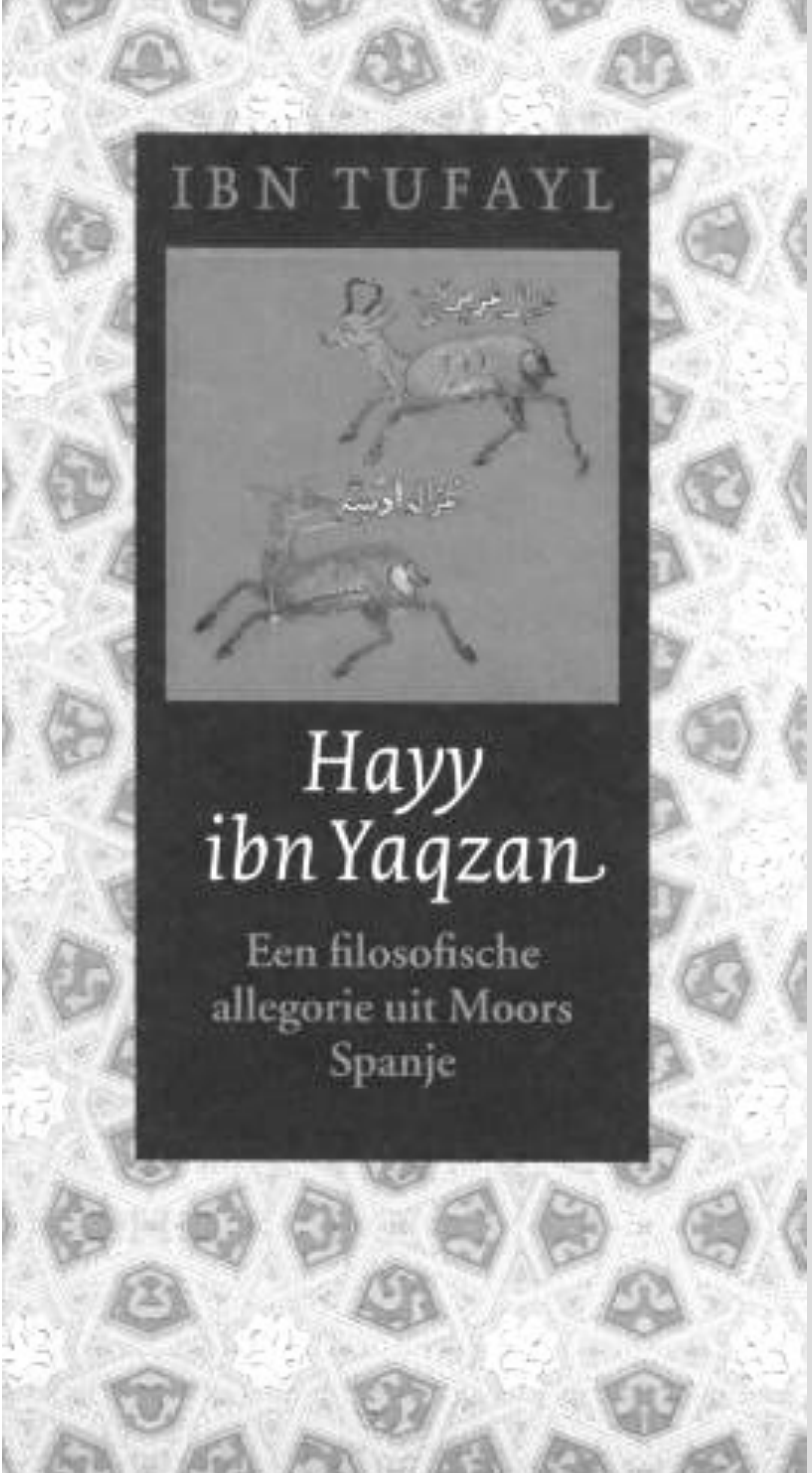
- Recall that P screens T off from R
  - i.e. $T \perp\!\!\!\perp R|P$
  - $\mathcal{P}(R|T,P) = \mathcal{P}(R|P)$
- Participants do not seem to realise this in observation condition
  - i.e. give significantly different conditional probability estimates for $\mathcal{P}(R|T,P)$ and $\mathcal{P}(R|P)$, violating the Markov condition
- But they do realise this in intervention condition
  - i.e. give approximately same probability estimate, consistent with Markov condition



Summary

- *Interventions* are manipulations of a system that can reveal causal directionality by disconnecting variables from their normal causes
- *Causal Bayesian Networks* + *‘Do*' *calculus* provide a handy way to formalise this
- People seem to learn more about a simple causal system when they make interventions themselves than when the passively observe

Part 2: Active Learning


IBN TUFAYL
Hayy ibn Yaqzan
Een filosofische allegorie uit Moors Spanje

Active Learning and **Optimal Experiment Design (OED)**

- Dominant framework for modeling many kinds of inquiry in Psychology/Cognitive Science
- e.g., categorization, logical reasoning, causal learning, spatial search, eye movements, rule learning, …
- Core metaphor/hypothesis

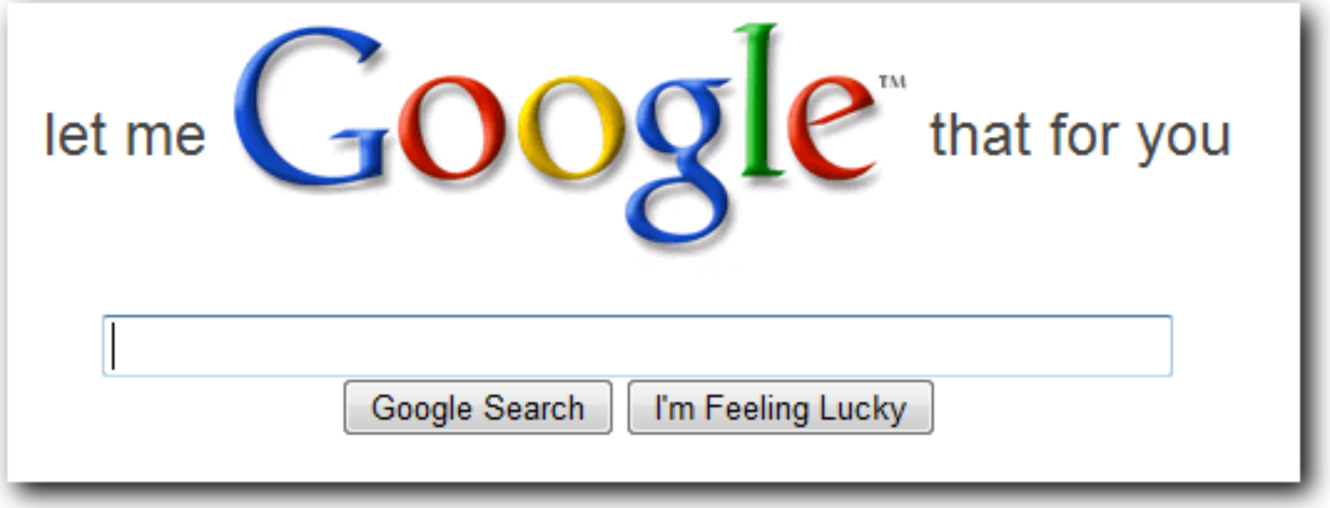
People are intuitive scientists and their information-seeking actions are optimal experiments

- Inspired by statistical work on ‘actual’ experiment design (e.g., Fedorov, 1972; Good, 1950; Lindley, 1956)

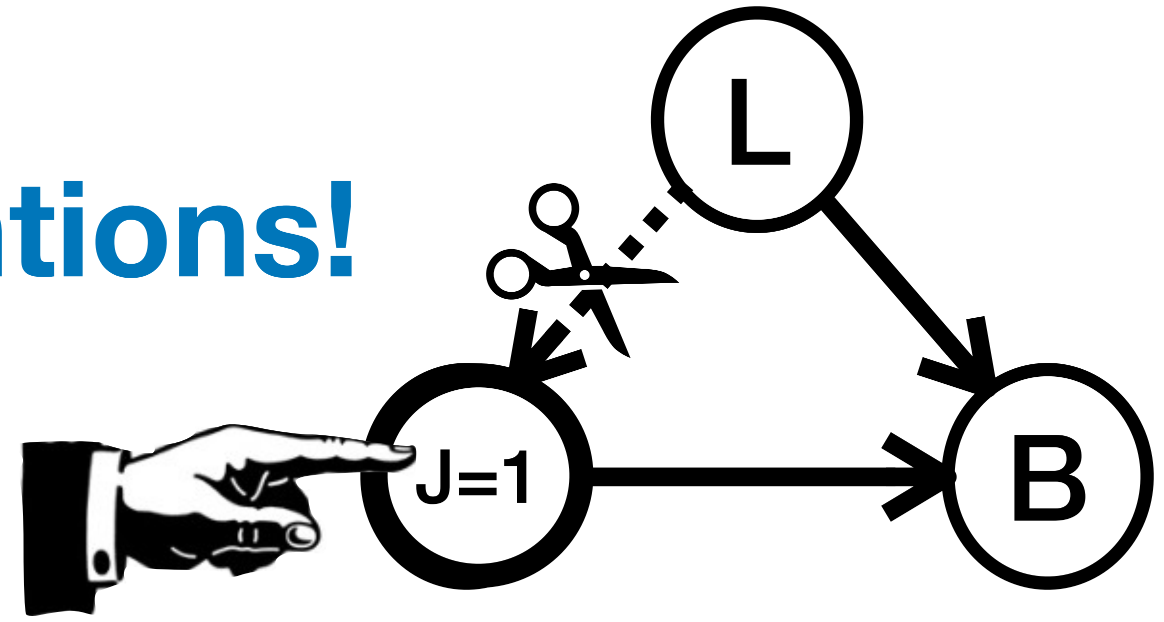
What is active learning?

The study of situations in which people have control over the information they see

- <u>Higher level cognition:</u>
  - Asking pertinent questions (Rothe et al, 2018)
    - Querying a category
    - Googling stuff
    - Emailing your lecturer about the midterms
    - Playing "20 questions", "Guess who" or "Battleship"
  - Designing an informative experiment a.k.a. *Optimal Experimental Design* (OED, *Atkinson & Donev,1992*)
  - Choosing what test to run next (e.g. medical diagnosis, fault finding)
  - Taking an action to see what its effects are…

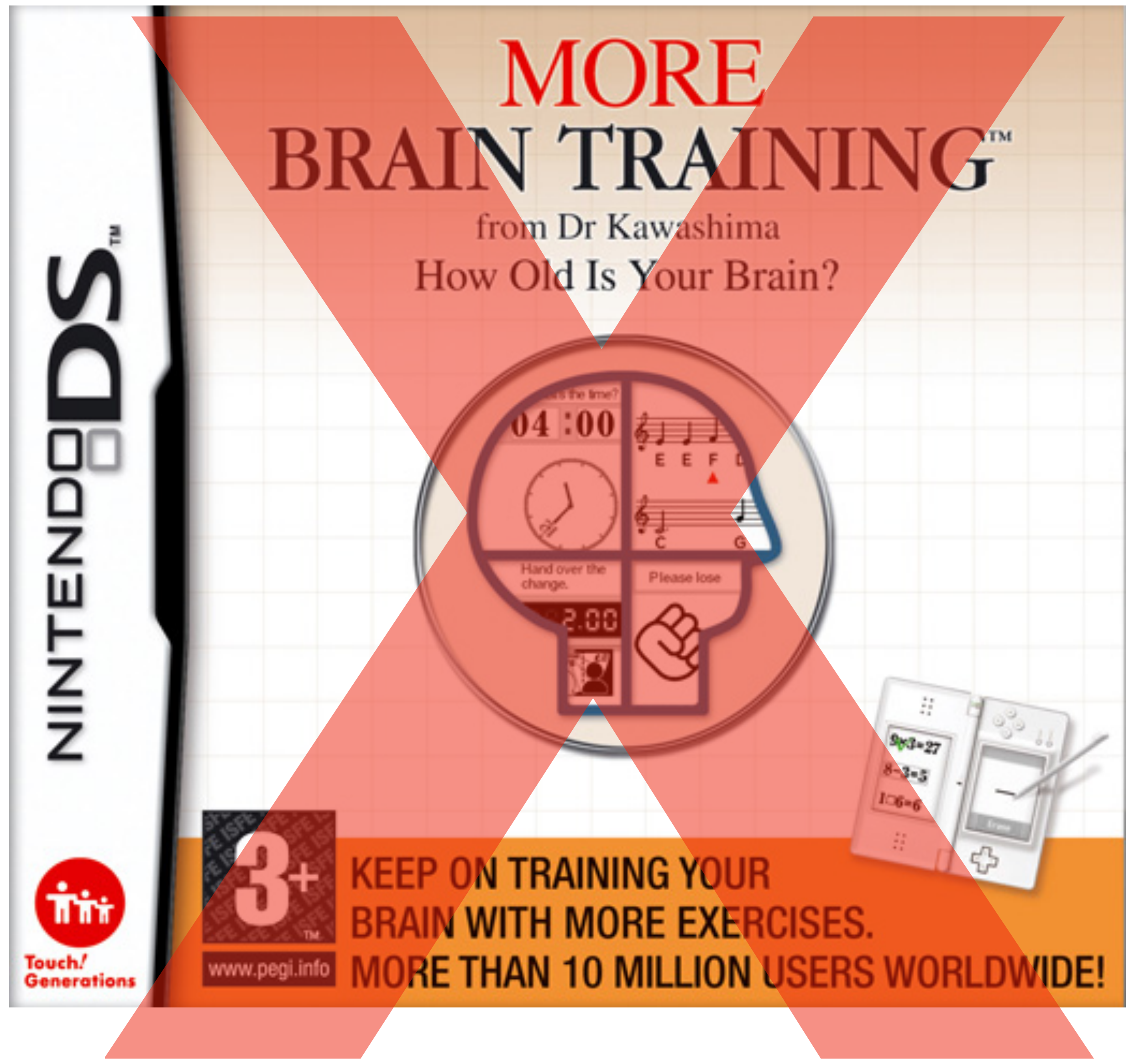


Interventions!



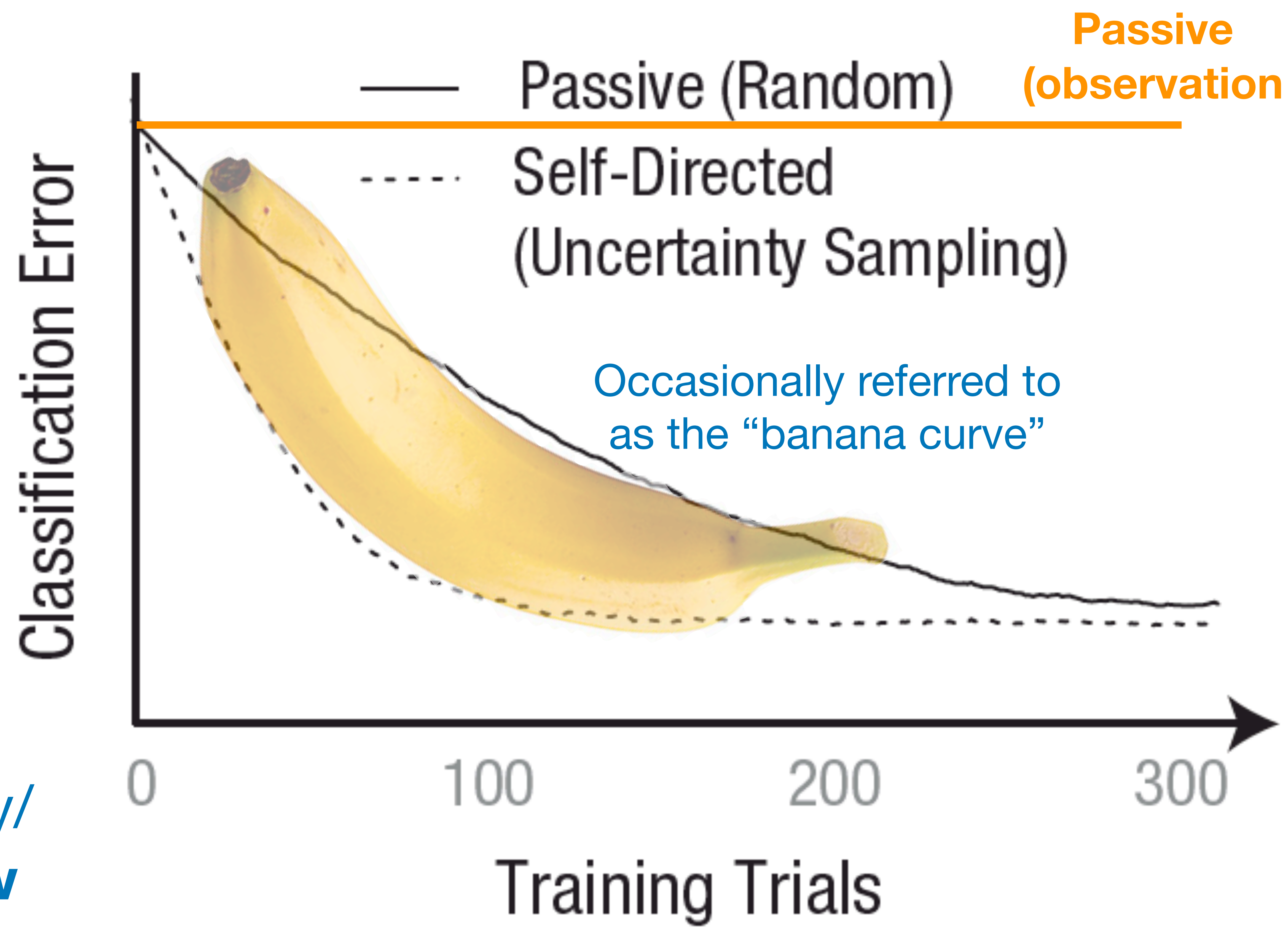
What isn't active learning?

- Being active while learning (e.g. Hillman et al, 2008)
- Brain training (e.g. Ball et al, 2002)



The advantages of active learning

- If done well, active learning speeds up learning
- Learners focus on what they're unsure about so experience less redundant evidence & accuracy increases more rapidly
  - I.e. if you moved your eyes at random it would take longer to establish what is infront of you
  - If you took random actions it would take a long time to discover relevant causal relations
- In causal context, choosing right interventions **necessary** to make progress

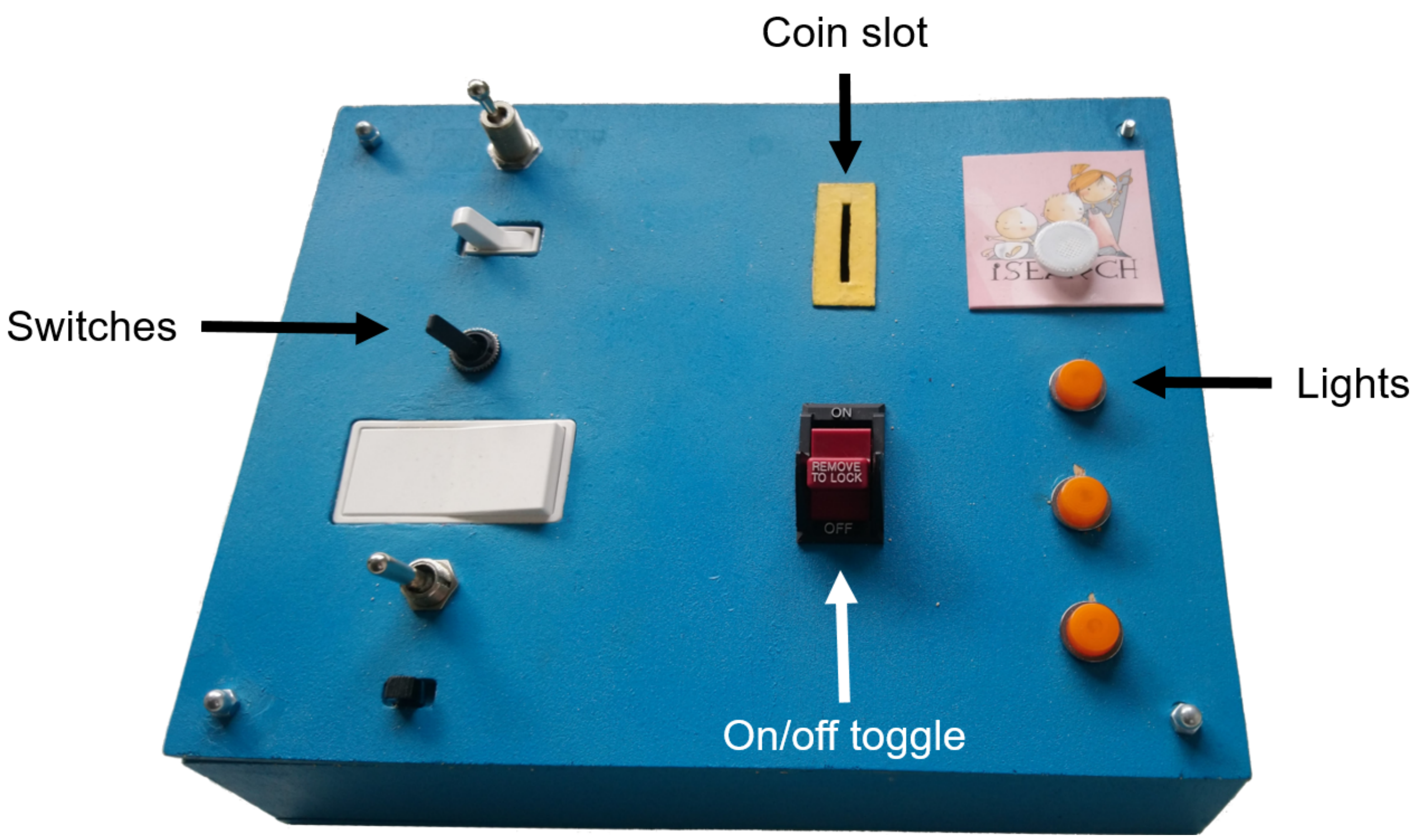


(From Gureckis & Markant, 2012)

- **But what makes one intervention** (question/query/test/action) **more informative than another? How might we measure this?**

Choosing interventions

- **(Foreshadowing) answer**: We can use *information theory*
  - But best illustrated via an experiment / example…
- Coenen, Ruggeri, Bramley & Gureckis (2019)
  - Participants interact with a mysterious magic switch box with:
    - Several switches that be set *on* (1) or *off* (0)
    - A light bank that might turn *on* (1) or *not* (0)
    - A testing toggle and coin slot for paying for tests

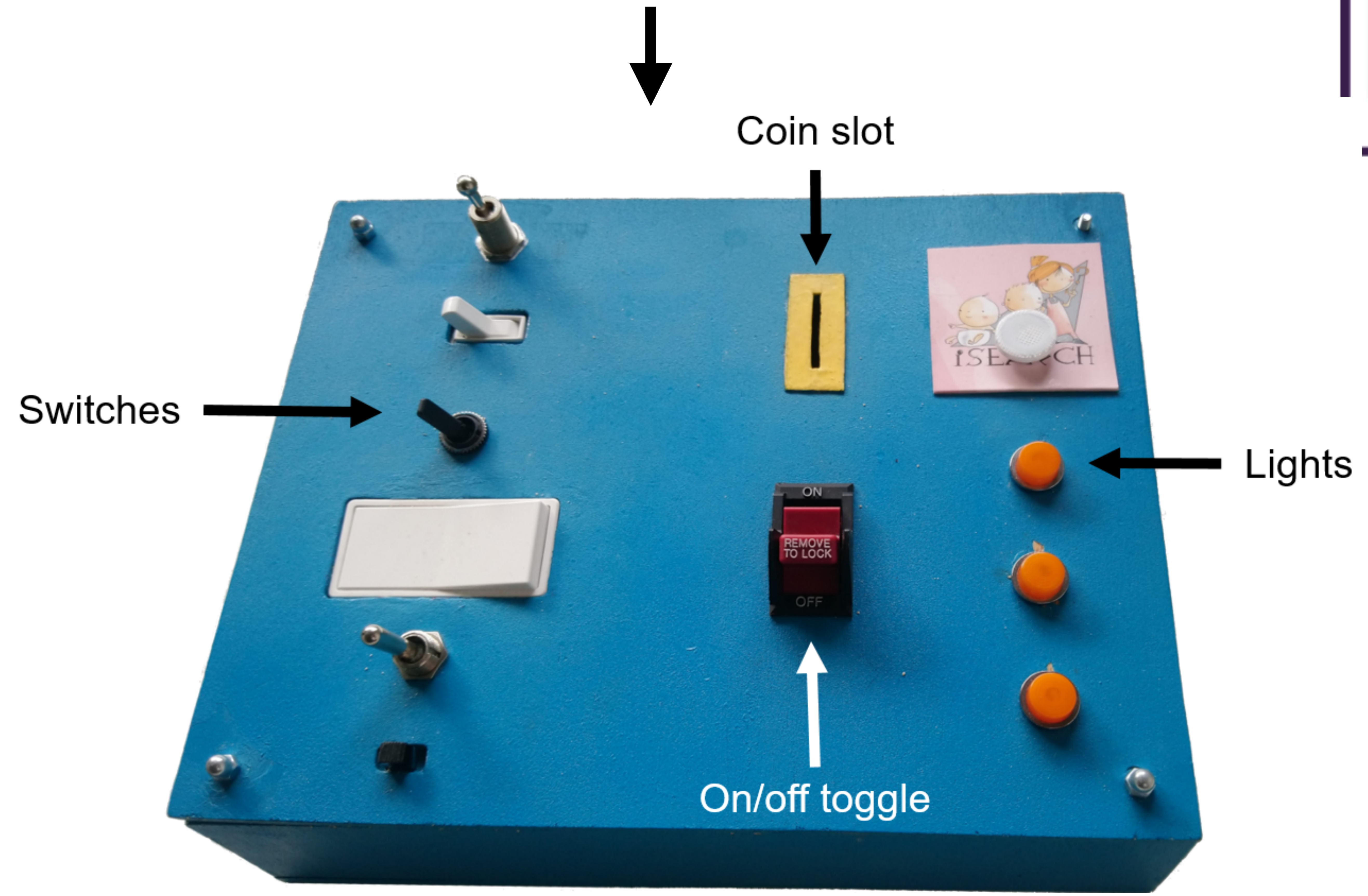


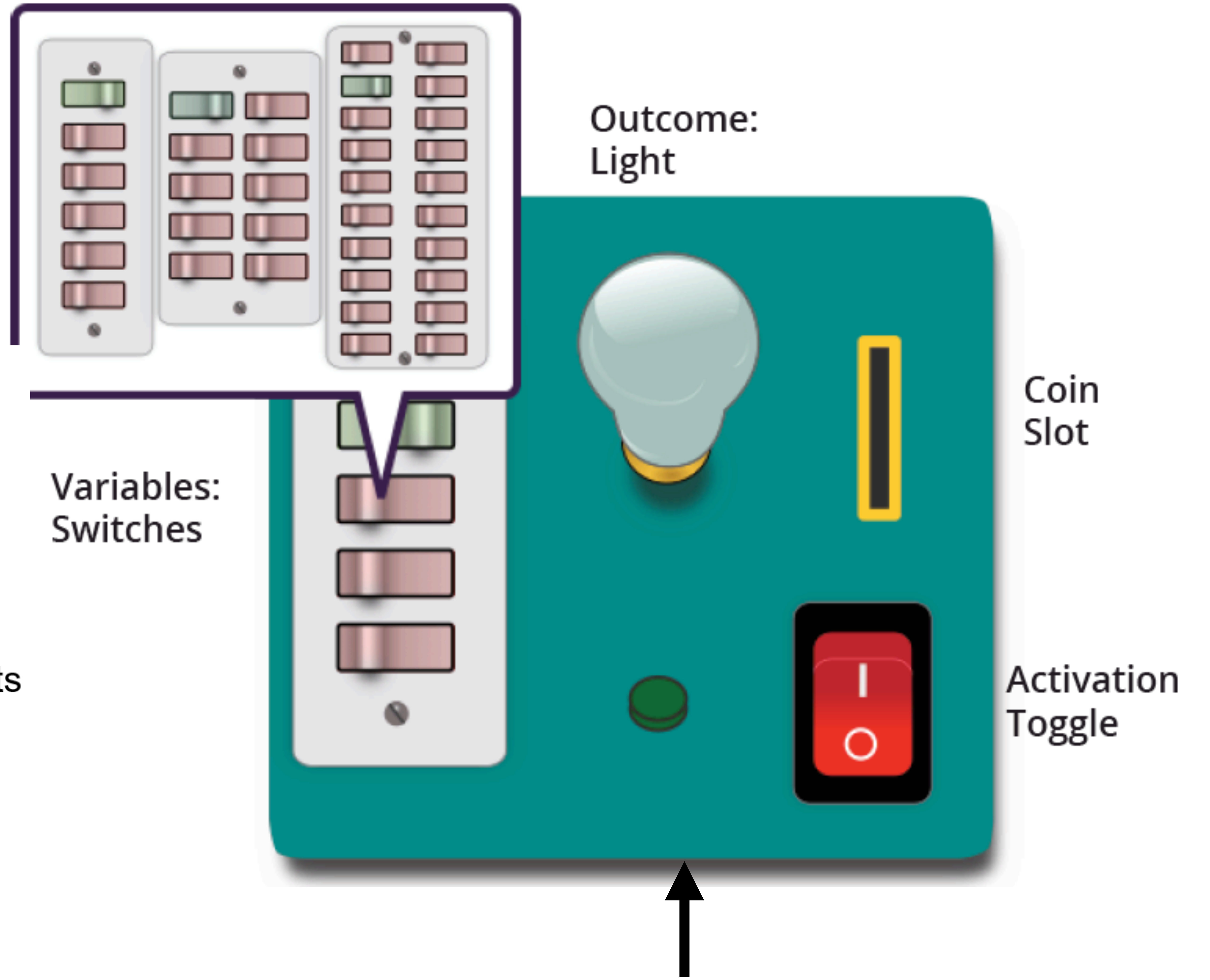
Choosing interventions

Original switch box

(Coenen et al, 2019 Exp 1
developmental tasks next week,
Built by Neil & Todd)

+
+





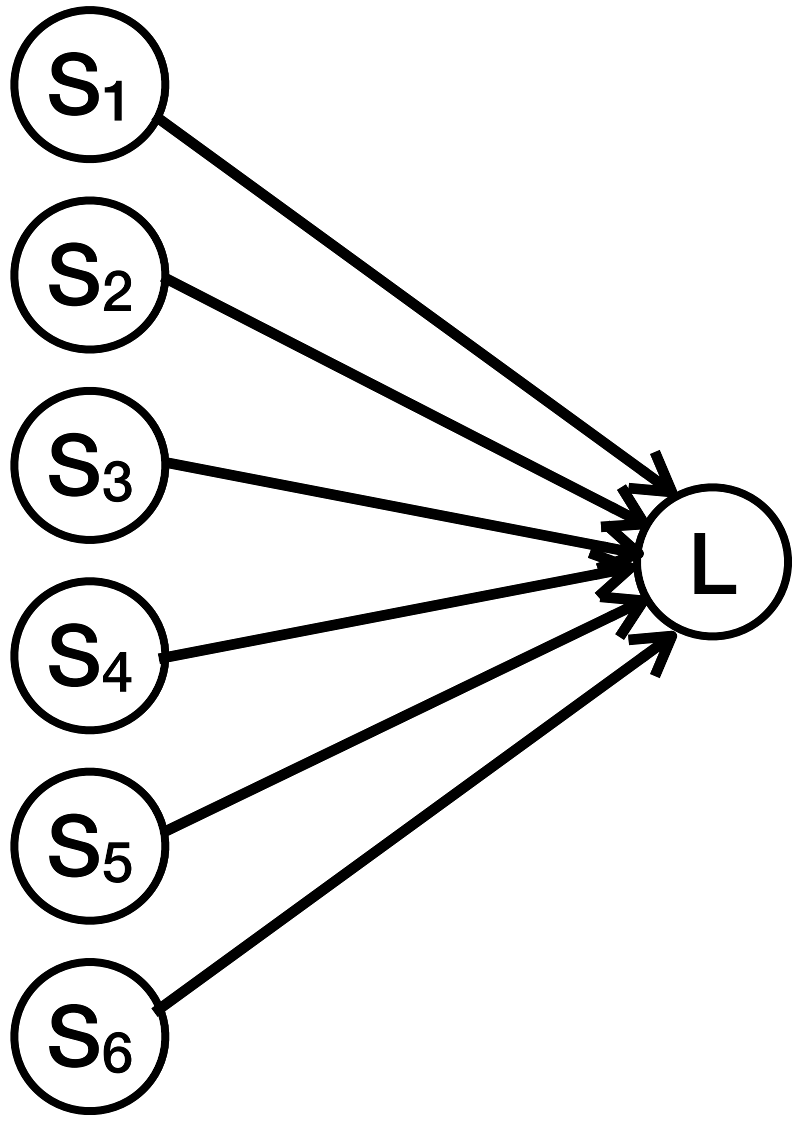
Online task version

(Coenen et al, 2019 Exp 2-4)

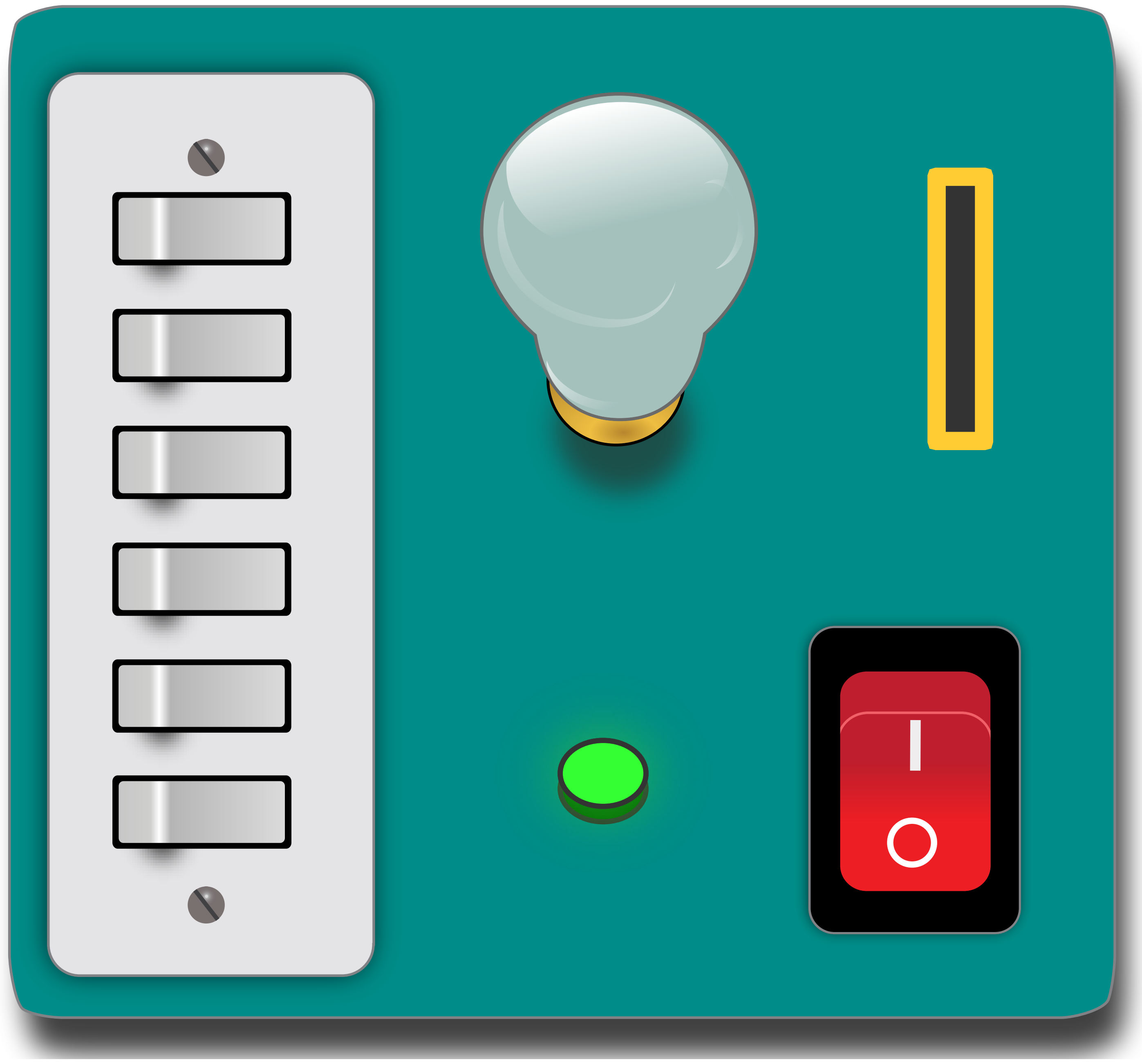
“Sparse” condition

“only one of the switches works”

- So 6 possible causal hypotheses (+: working, -: not working):
  - h_1:[+,-,-,-,-,-]
  - h_2:[-,+,-,-,-,-]
  - h_3:[-,-,+,-,-,-]
  - h_4:[-,-,-,+,-,-]
  - h_5:[-,-,-,-,+,-]
  - h_6:[-,-,-,-,-,+]



- And 64 possible interventions:
  - Do[0,0,0,0,0,0]
  - Do[1,0,0,0,0,0]
  - Do[0,1,0,0,0,0]
  - Do[0,0,1,0,0,0]
  - Do[0,0,0,1,0,0]
  - Do[0,0,0,0,1,0]
  - Do[0,0,0,0,0,1]
  - Do[1,1,0,0,0,0]
  - …
  - Do[1,1,1,1,1,1]

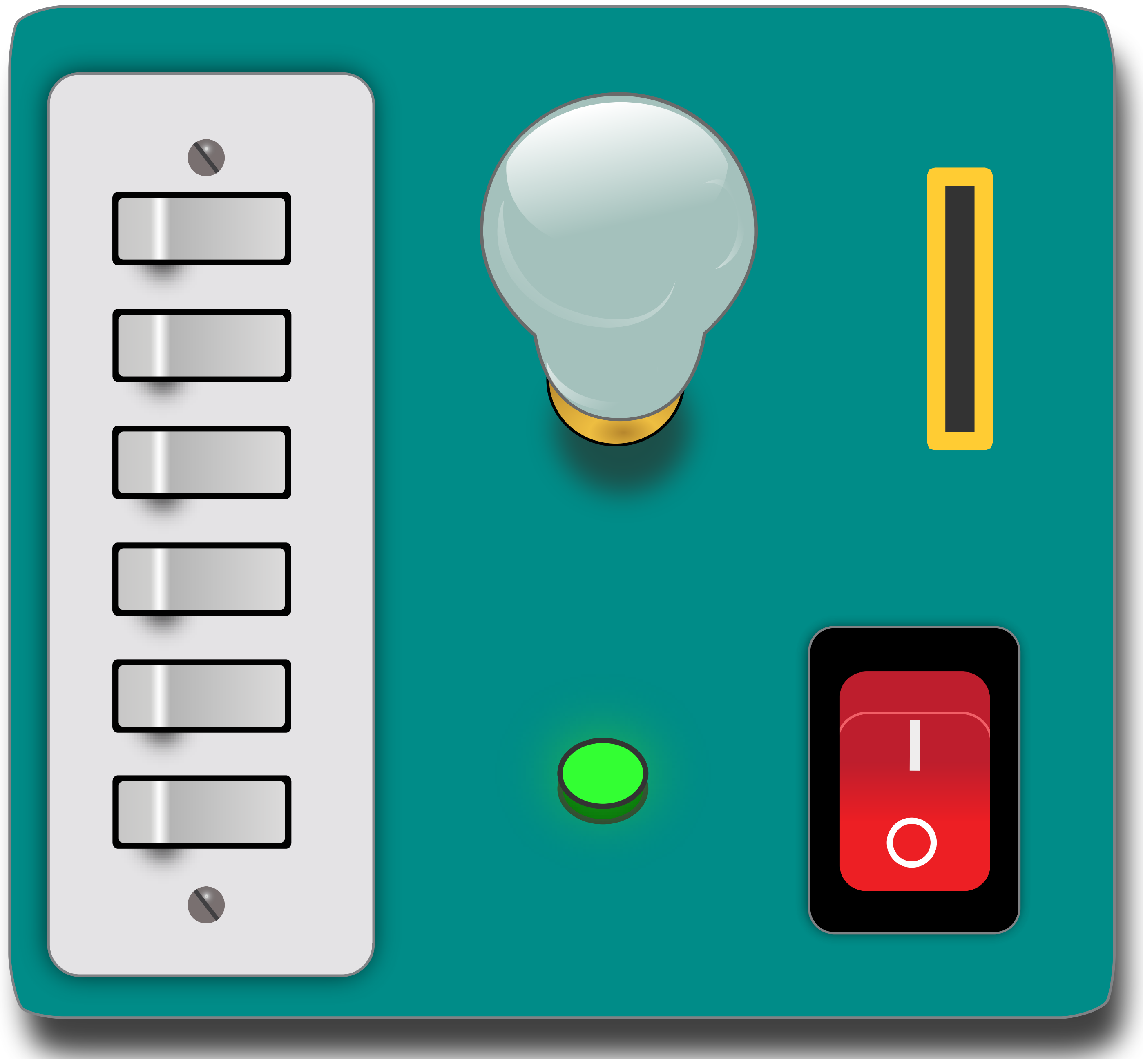


“Sparse” condition

“only one of the switches works”

- What would you do?
- **Try one switch at a time?**
  - e.g. Do[1,0,0,0,0,0], then Do[0,1,0,0,0,0], then Do[0,0,1,0,0,0] until you observe light (💡)
  - When 💡, you’ve found the working switch…
  - This will work…
  - But much can you expect to win with this approach?
    - Best case £5 (if its the 1st switch you test)
    - Worst case £1 (if its the 5th switch you test)
      - Or if its none of the first 5 (then it must be the 6th)
      - On average… £2.50
- Can you do better?

h_1:[+,-,-,-,-,-]
h_2:[-,+,-,-,-,-]
h_3:[-,-,+,-,-,-]
h_4:[-,-,-,+,-,-]
h_5:[-,-,-,-,+,-]
h_6:[-,-,-,-,-,+]

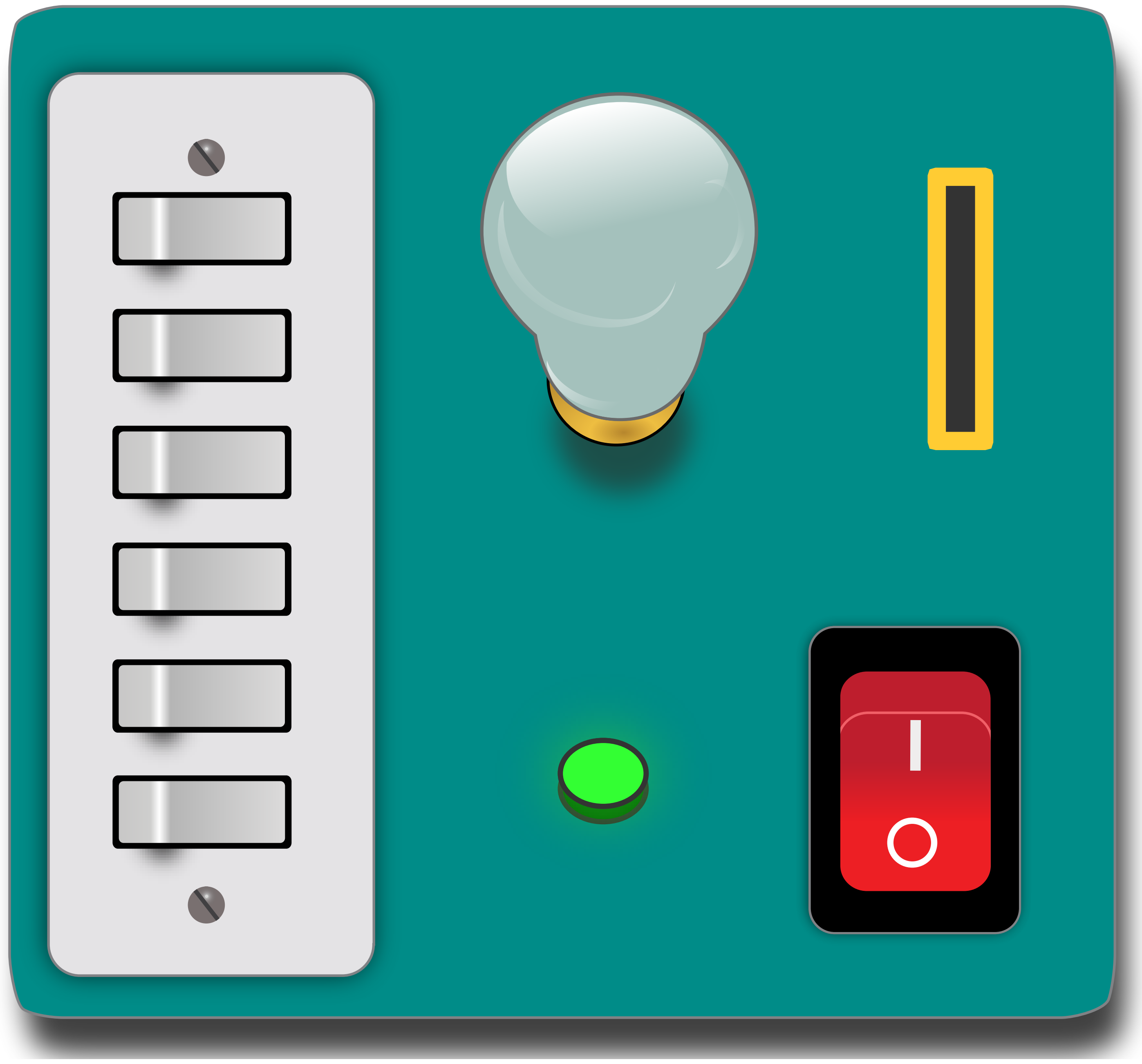


"Sparse" condition

"only one of the switches works"

h_1:[+,-,-,-,-,-]
h_2:[-,+,-,-,-,-]
h_3:[-,-,+,-,-,-]
h_4:[-,-,-,+,-,-]
h_5:[-,-,-,-,+,-]
h_6:[-,-,-,-,-,+]

- **Try half the "remaining" switches each time?**
  - e.g. Do[1,1,1,0,0,0]
    - If 💡, $\cancel{h_4}, \cancel{h_5}, \cancel{h_6}$. Then Do[1,0,0,0,0,0]
      - If 💡, $\cancel{h_2}, \cancel{h_3}$. You're done! (it must be h_1)
        Otherwise Do[0,1,0,0,0,0]
        - If 💡, $\cancel{h_3}$. You're done (it must be h_2)
          Otherwise $\cancel{h_4}$ you're also done (it must be h_3)
    - If not, $\cancel{h_1}, \cancel{h_2}, \cancel{h_3}$. Then Do[0,0,0,1,0,0]
      - If 💡, $\cancel{h_5}, \cancel{h_6}$. You're done! (it must be h_4)
        Otherwise Do[0,0,0,0,1,0]
        - If 💡, $\cancel{h_6}$. You're done (it must be h_5)
          Otherwise $\cancel{h_5}$ you're also done (it must be h_6)



“only one of the switches works”

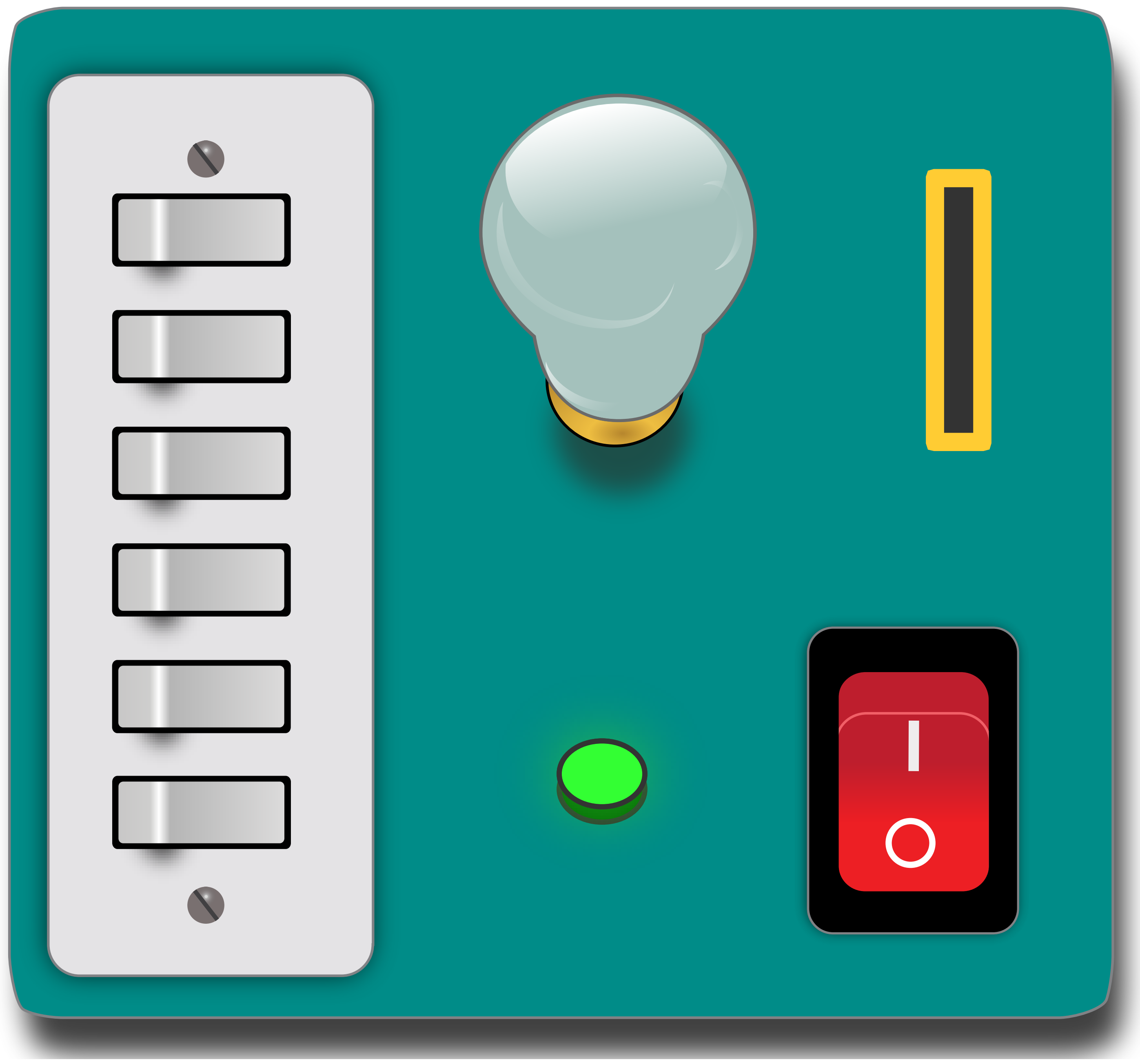
- **Try half the “remaining” switches each time?**
  - But much can you expect to win with this approach?
    - Best case £4 (if you isolate the working switch in 2 tests)
    - Worst case £3 (if you need a third test)
      - On average you'll make **~£3.40!**
- A kind of “divide and conquer” strategy
- Known as the “split half heuristic” (Nelson et al, 2013)
  - Also optimal approach here + in games like *Guess Who* (i.e. ask about gender first since it cuts the field down by half)

“Dense” condition

“all but one of the switches works”

- So 6 possible causal hypotheses (+: working, -: not working):
  - h_1:[-,+,+,+,+,+]
  - h_2:[+,-,+,+,+,+]
  - h_3:[+,+,-,+,+,+]
  - h_4:[+,+,+,-,+,+]
  - h_5:[+,+,+,+,-,+]
  - h_6:[+,+,+,+,+,-]

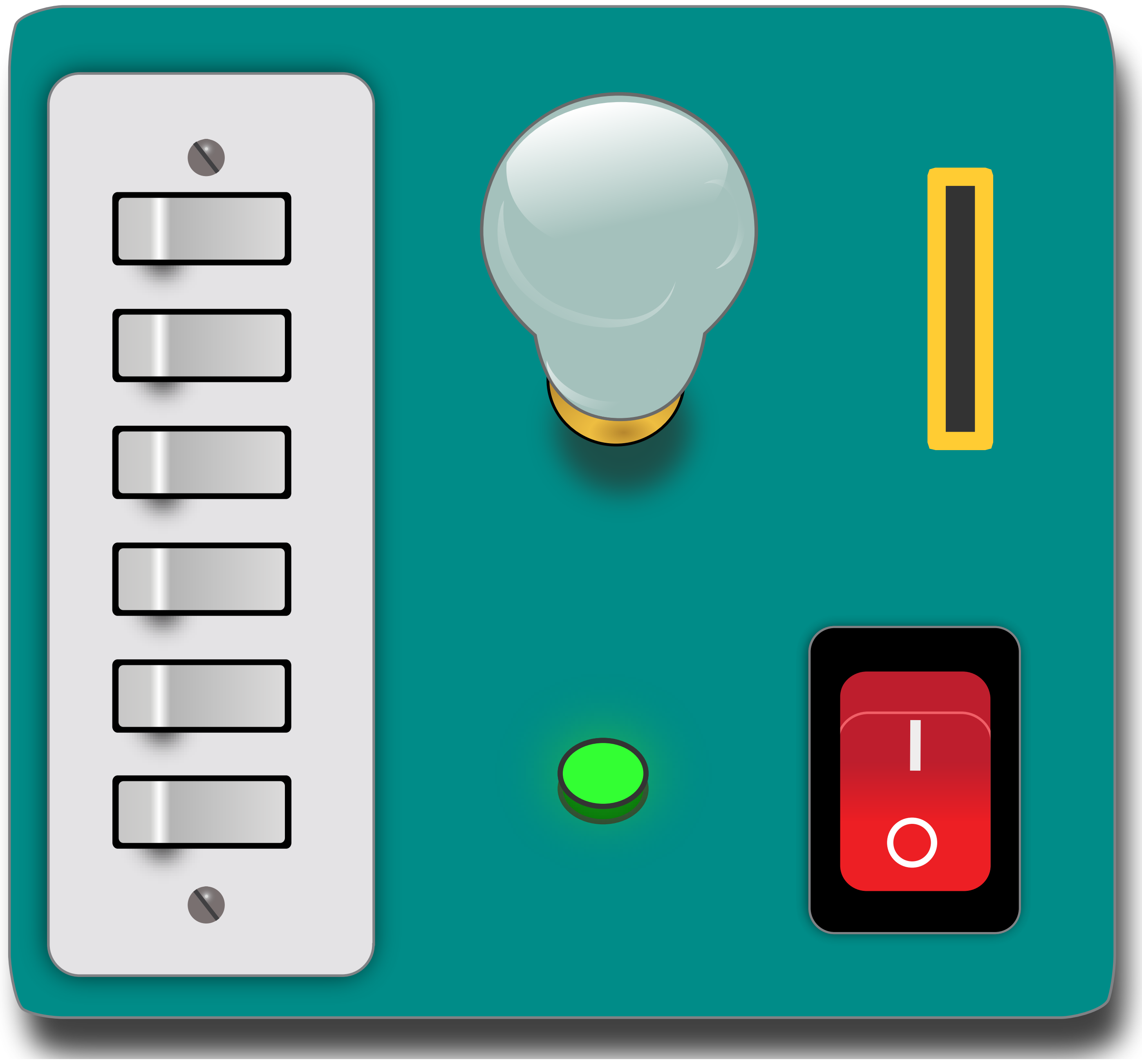
- And 64 possible interventions:
  - Do[0,0,0,0,0,0]
  - Do[1,0,0,0,0,0]
  - Do[0,1,0,0,0,0]
  - Do[0,0,1,0,0,0]
  - Do[0,0,0,1,0,0]
  - Do[0,0,0,0,1,0]
  - Do[0,0,0,0,0,1]
  - Do[1,1,0,0,0,0]
  - …
  - Do[1,1,1,1,1,1]



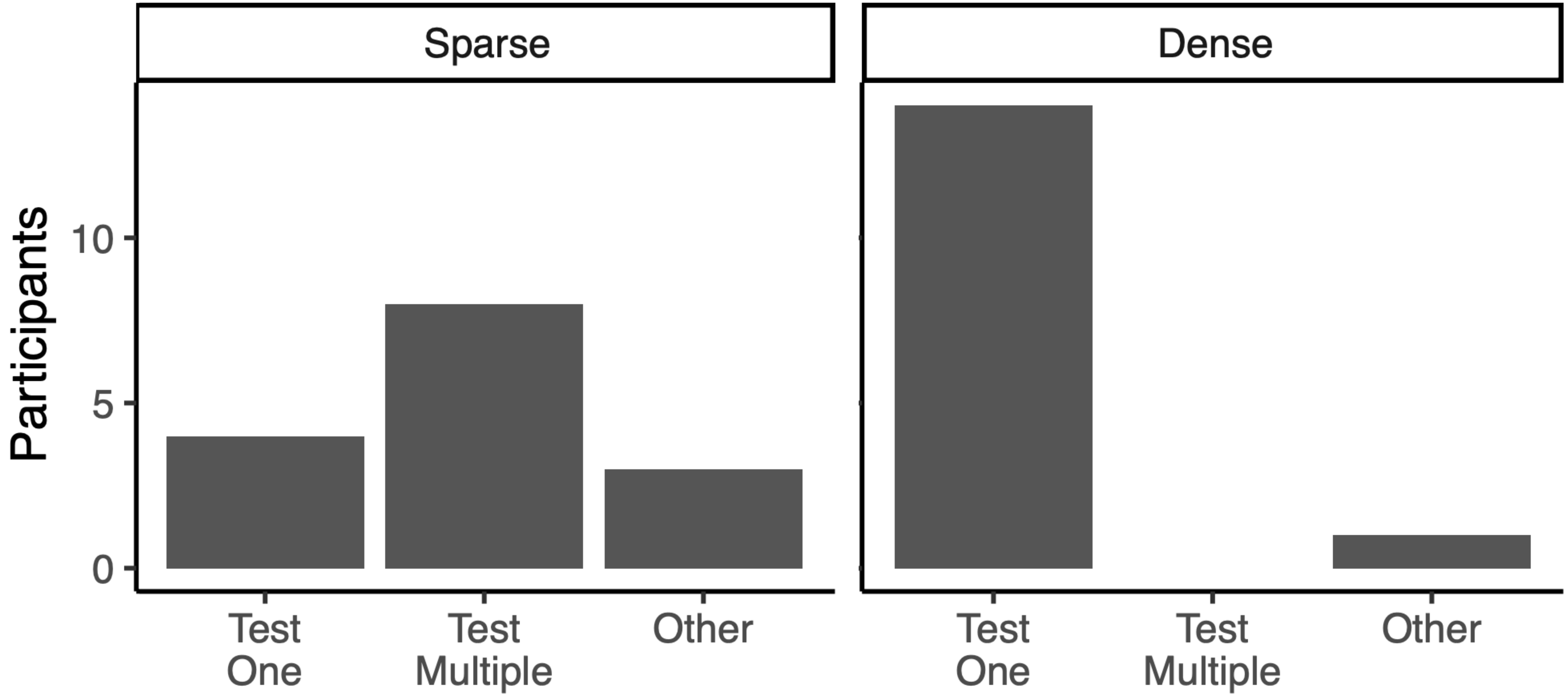
"Dense" condition

"all but one of the switches works"

- What would you do?
- **Try one switch at a time?**
  - e.g. Do[1,0,0,0,0,0], then Do[0,1,0,0,0,0], then Do[0,0,1,0,0,0] until you observe *no lights* (💡)
  - When 💡, you've found the broken switch…
  - This will work…
- Can you do better?
- **No!** This is actually the only strategy that will work at all here.
- If you turn on more than one switch then the lights *always* come on, no matter which hypothesis is true, so you will learn nothing…

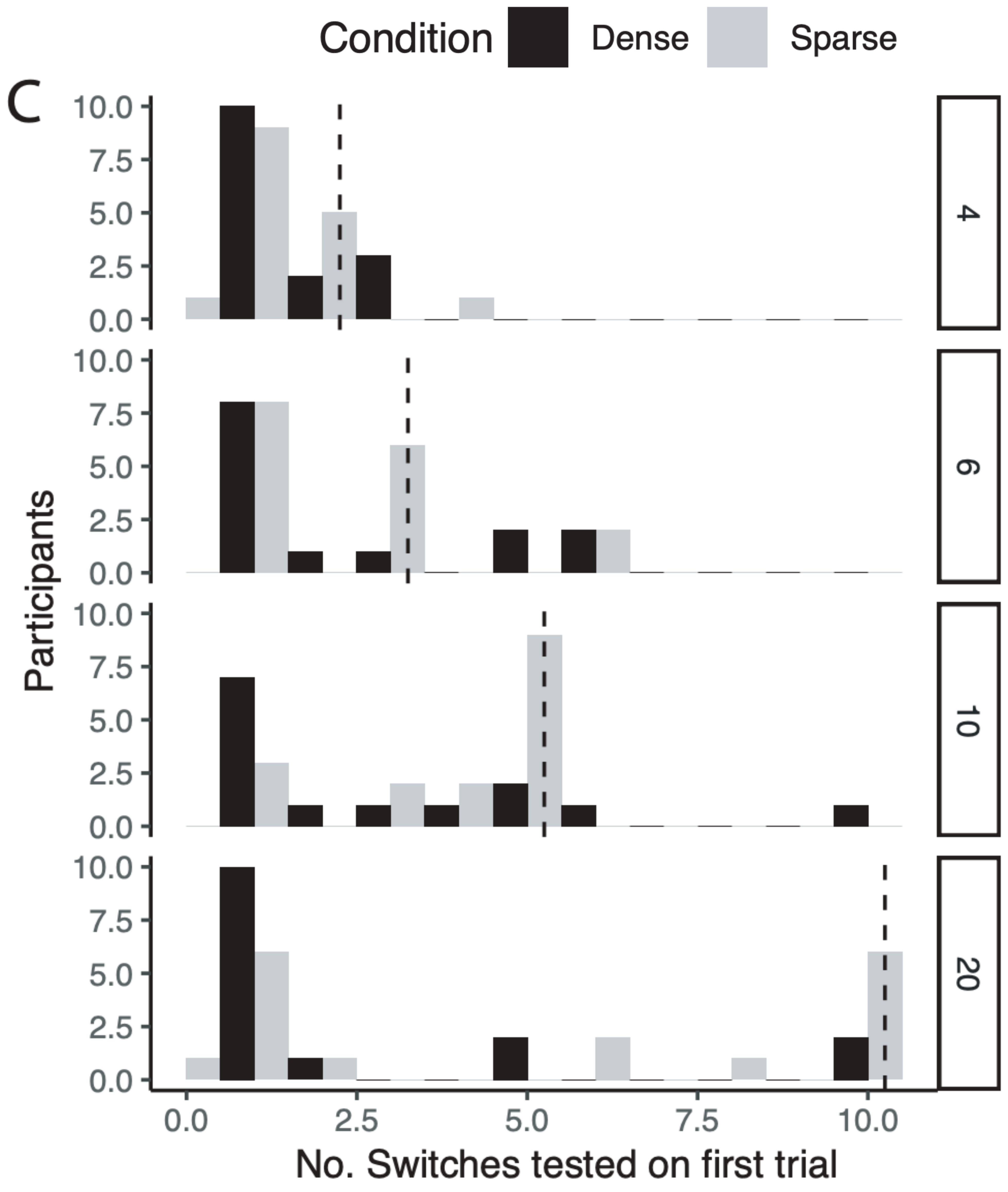


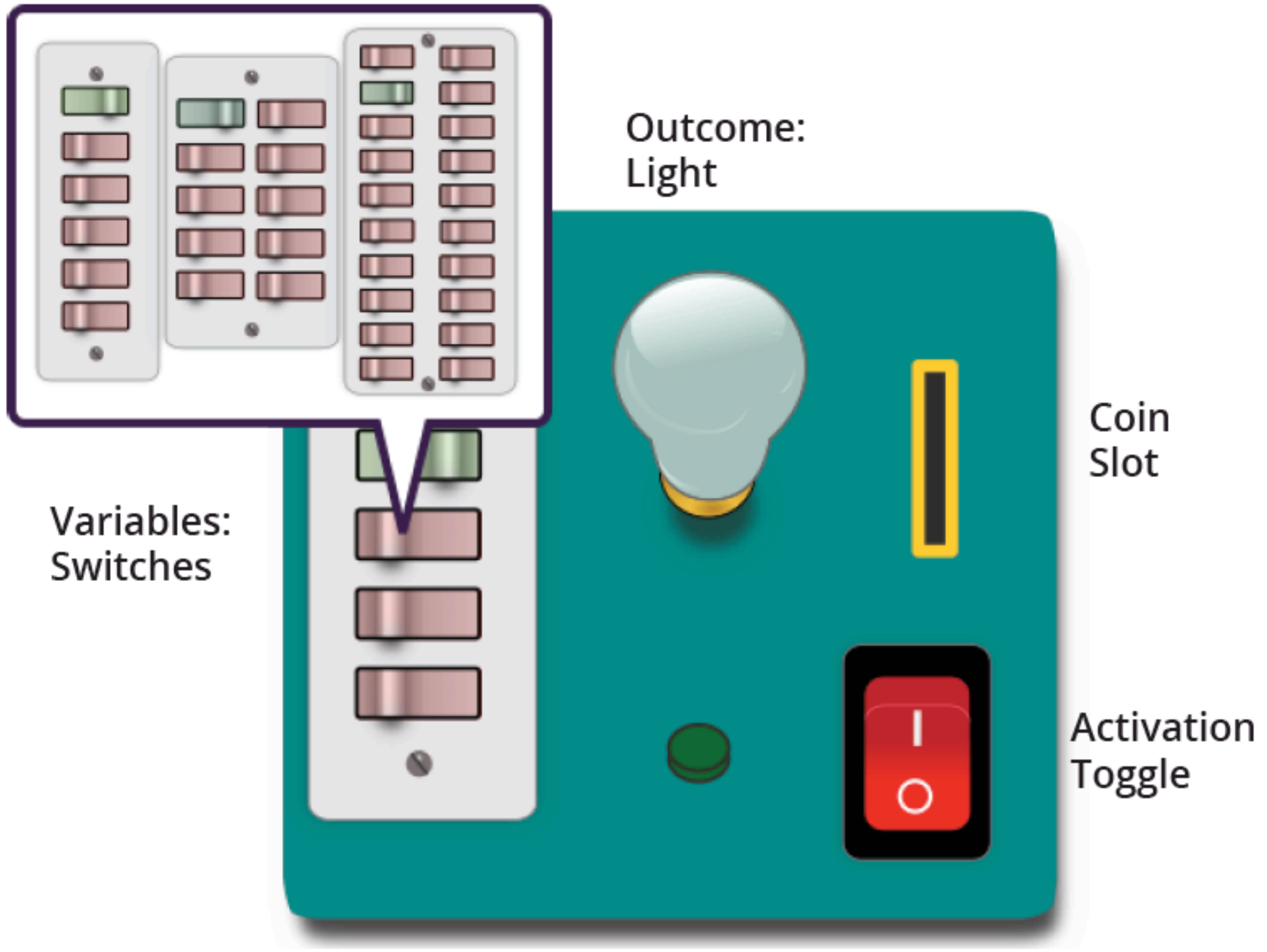
What did people do?



- Exp 1 (N=30). Accuracy very high: 100% in sparse, 80% in dense condition
- Most participants in **Sparse** condition switch at least 2-5 switches on first trial and then 1 or 2 of the remainder on next trial (see paper for more complex “strategy classification”)
- All but one participant in **Dense** condition switches 1 on first trial

What did people do?

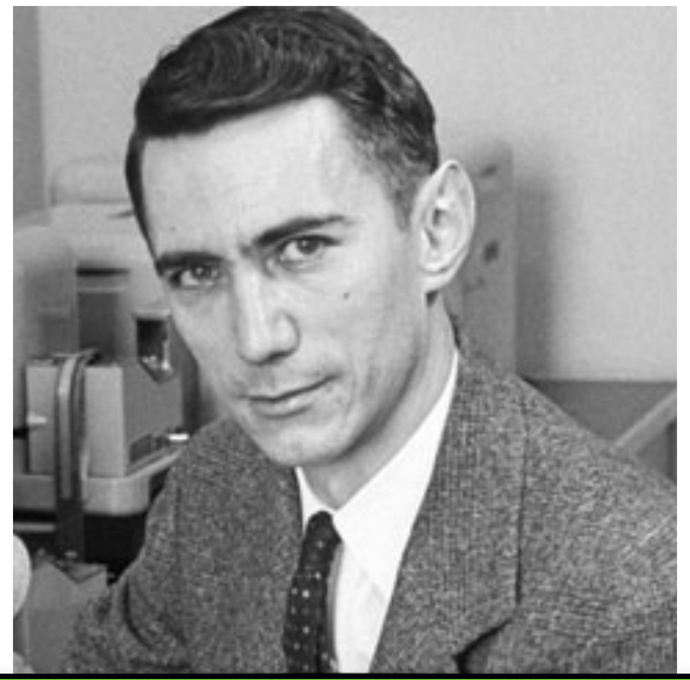




- Exp 2 (N=130). Similar patterns as number of switches increases (increasing pressure to be efficient with interventions)

Information

- Why was the Test Multiple strategy more efficient in the sparse condition?
  - It narrowed the option set more rapidly…
  - In other words: It **reduced uncertainty** about the true hypothesis more quickly
- How can we measure uncertainty?
- We often use a measure called “Information Entropy” (Shannon, 1948)
  - Based on loose analogy to thermodynamic entropy in physics
  - Where high entropy means disorder



Claude Shannon
Bell Telephones Employee / Cryptographer

Low thermodynamic entropy = order

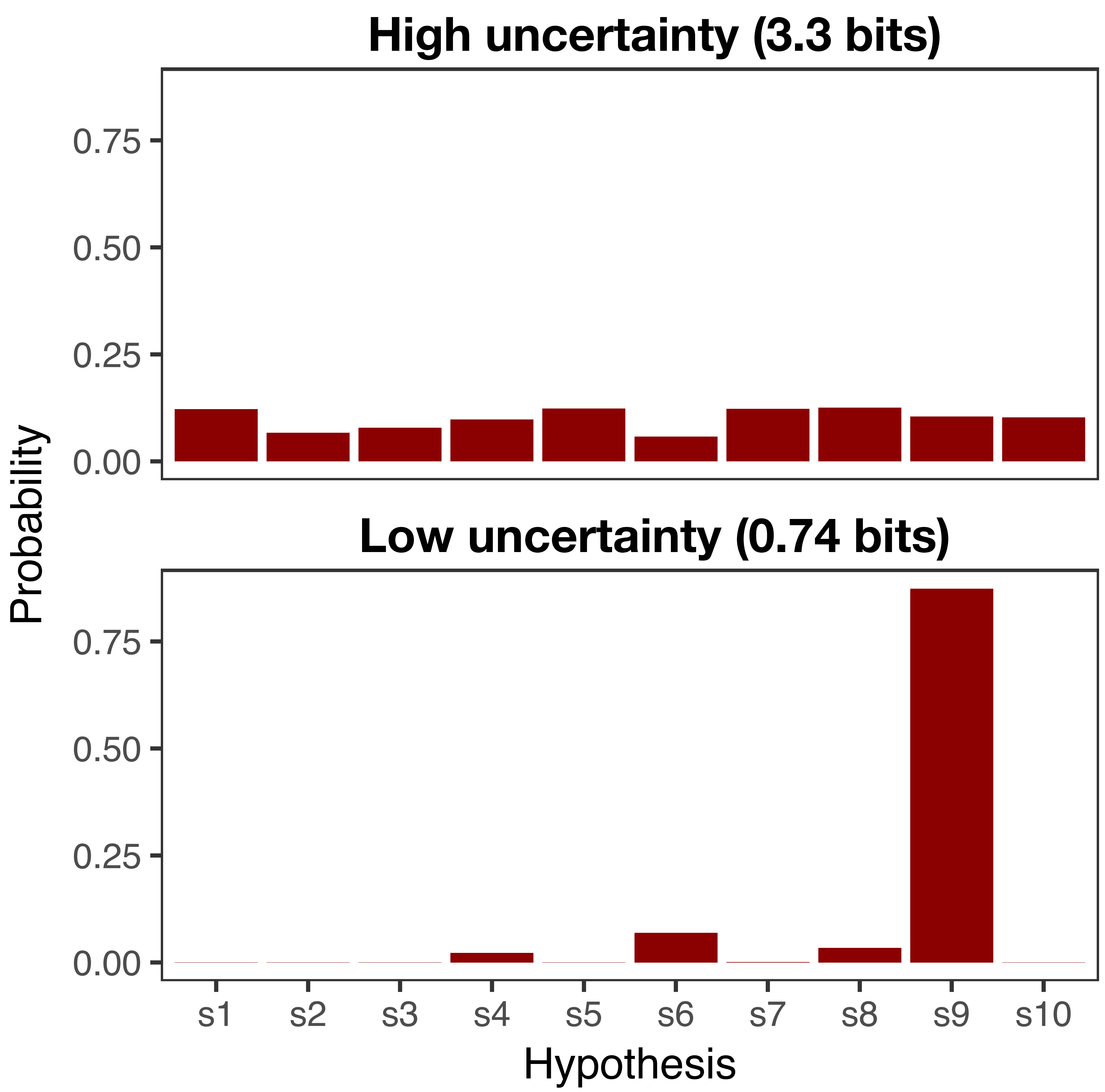
High thermodynamic entropy = chaos

Information

- **Entropy** (or **uncertainty**): A measure of how unsure you are about the state of a random variable (i.e. something represented by a probability distribution)

$$\mathcal{H}(S) = -\sum_{s \in S} P(s) \log_2 P(s)$$

- **Technical interpretation:** How "surprised" you'd be, on average, when you find out the true value
  - We've met lots of random variables already
  - You could measure the entropy of any of them…

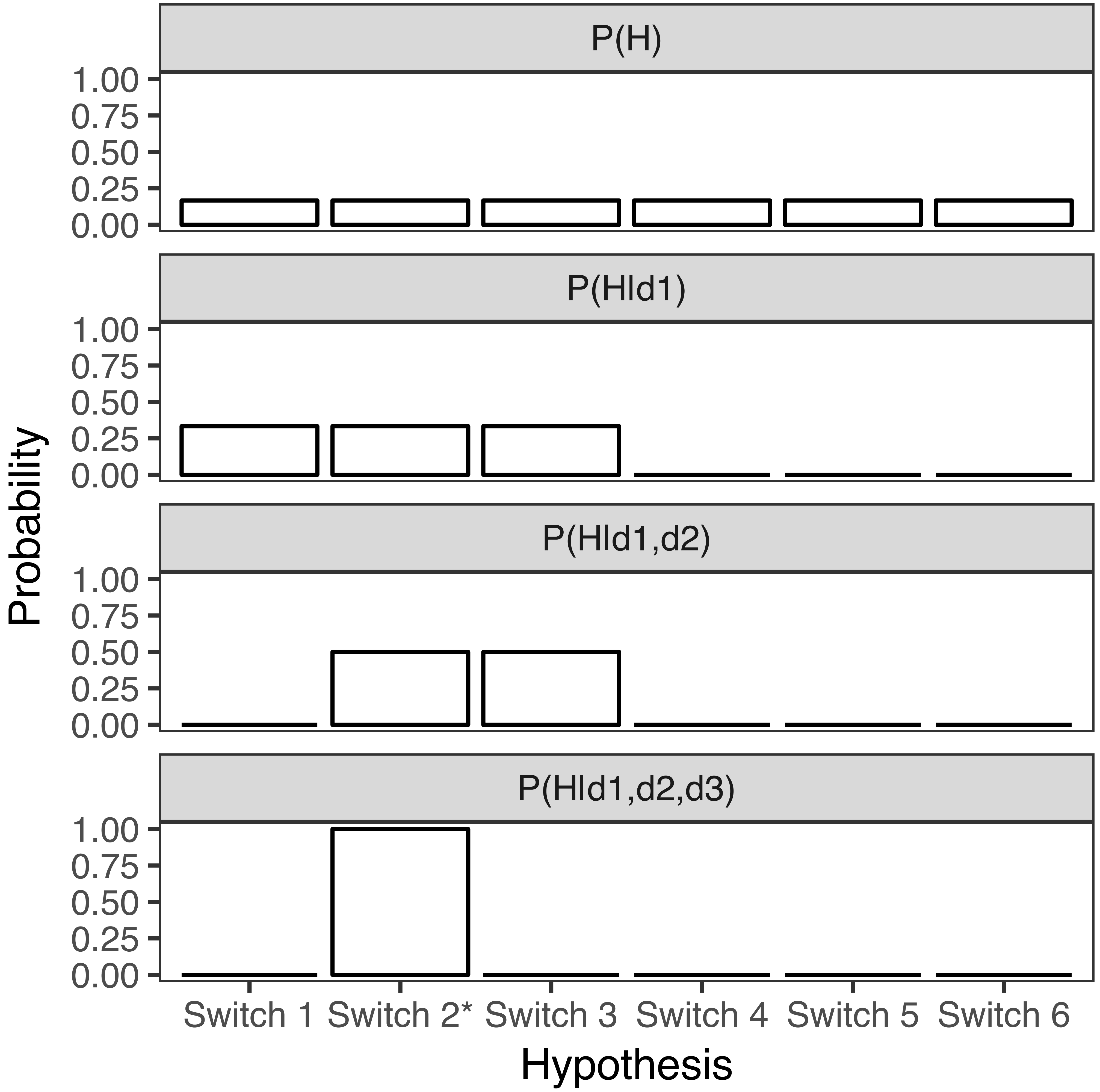


Information

- **Information**: Current **uncertainty** - absolute **uncertainty**
- **Information gain:** The difference in **uncertainty** from before to after receiving some evidence
- Measured in “bits” (+ “bytes”, “Mb” etc)
- Every memory slot in your computer stores 1 bit of information (either a 0 or a 1)
  - Looking at a memory slot takes you from complete ignorance P(“0”)=.5 or P(“1”)=.5 to knowing that its i.e. a P(“1”)=1
- Similarly with a (fair) coin flip, by looking at outcome learn 1 bit of information

Measuring information

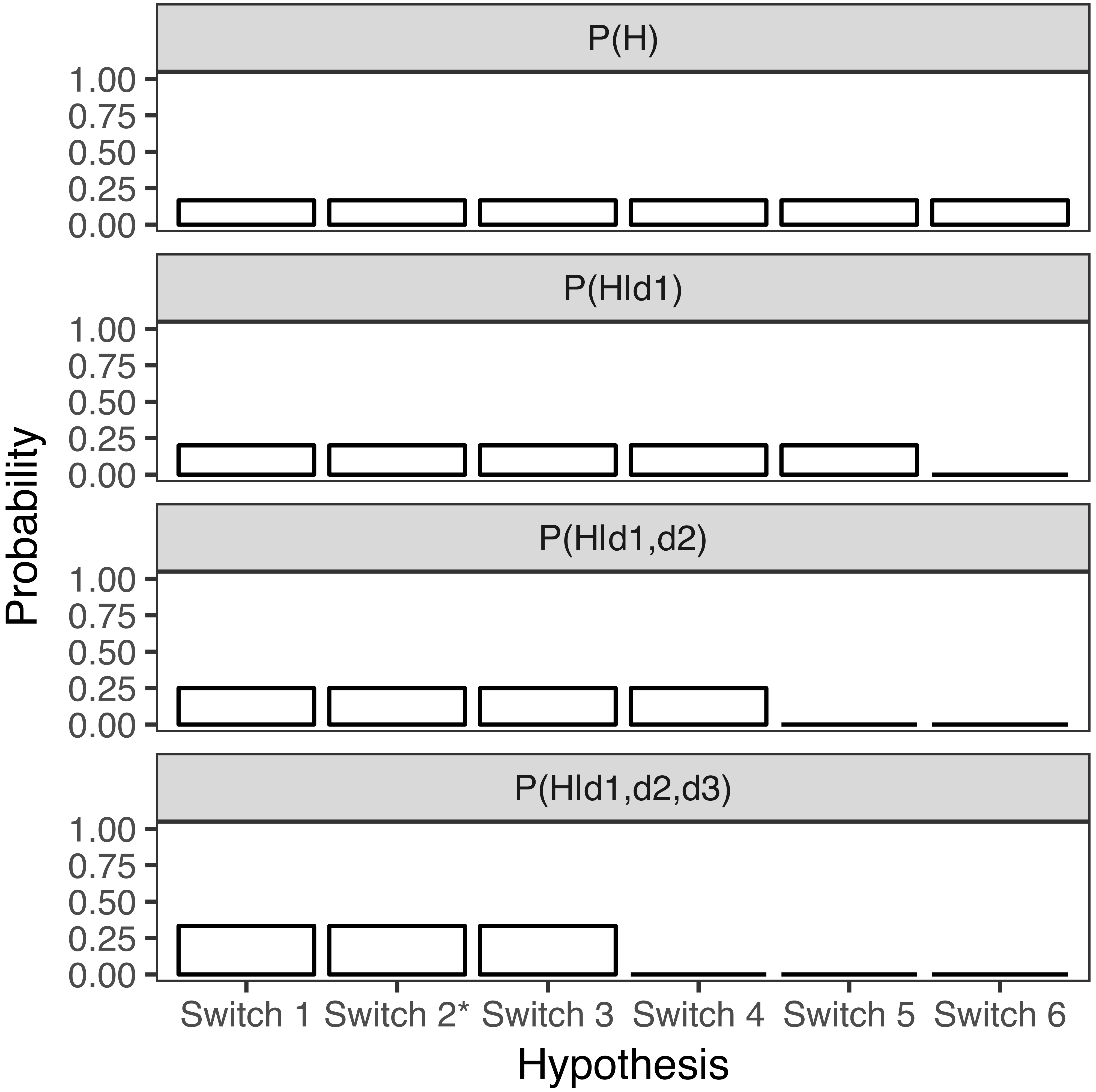
Test multiple strategy



Information Entropy	Information Gain
2.58	
	+1.00
1.58	
	+0.58
1.00	
	+1.00
0	

Measuring information

Test one strategy



Information Entropy

2.58

2.32

2.00

1.58

Information Gain

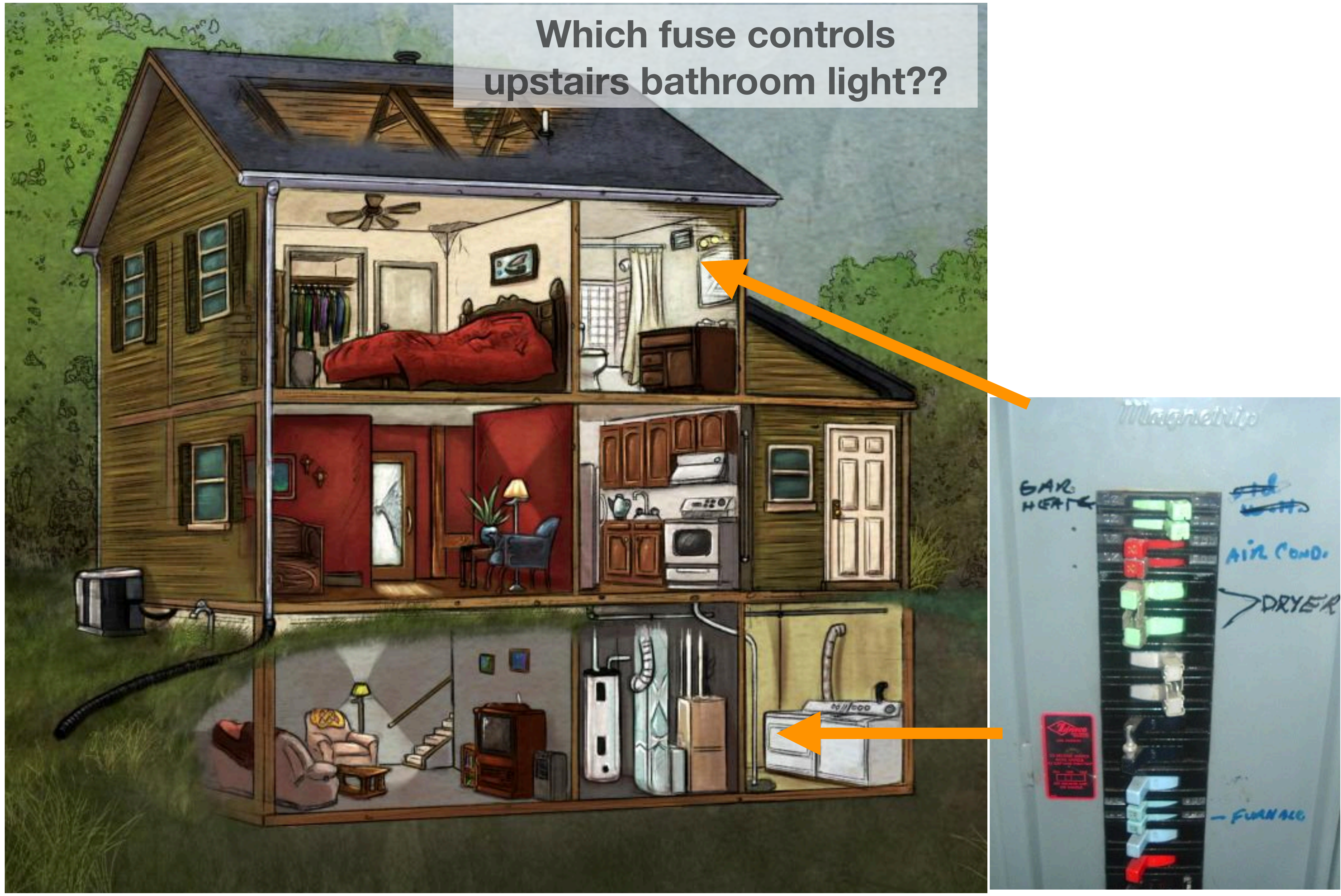
+0.26

+0.32

+0.42

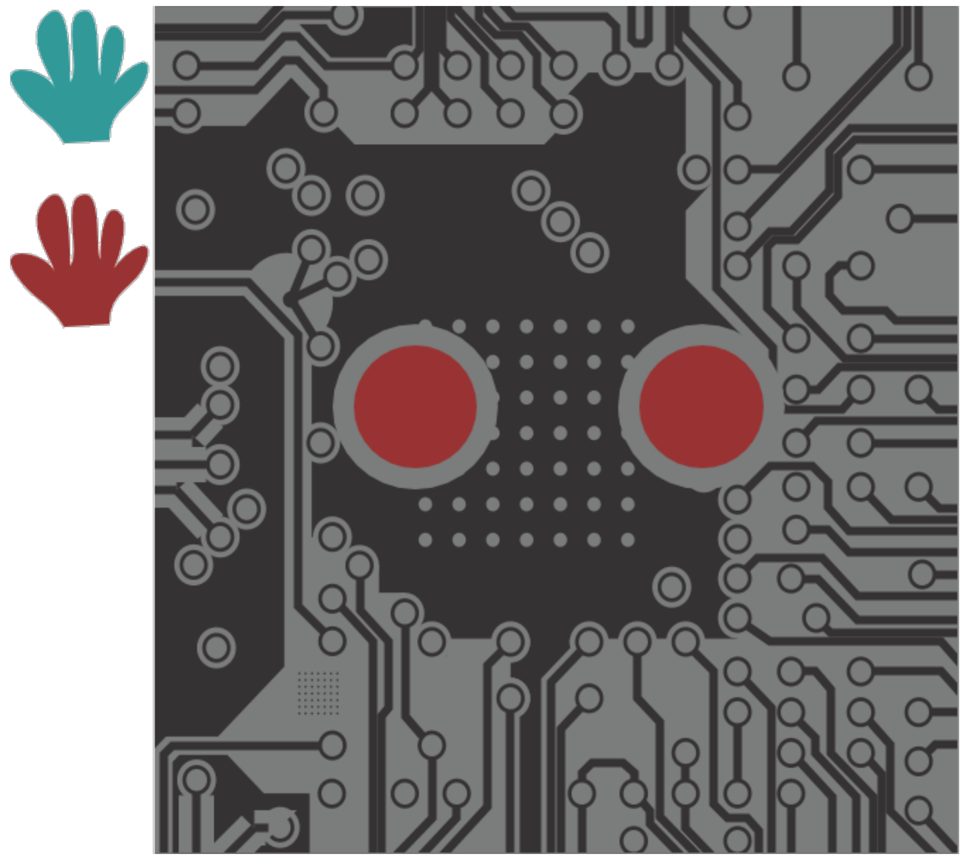
Interim summary

- **So information theory captures why some intervention strategies are more or less efficient at resolving a learner’s uncertainty**
- The switch box seems like an unusually idealised case
  - Deterministic binary relationships
  - Known ‘candidate causes’ and a single effect
- How often do we face problems like this in real life?
- But what about experiments looking at more general causal inference cases?



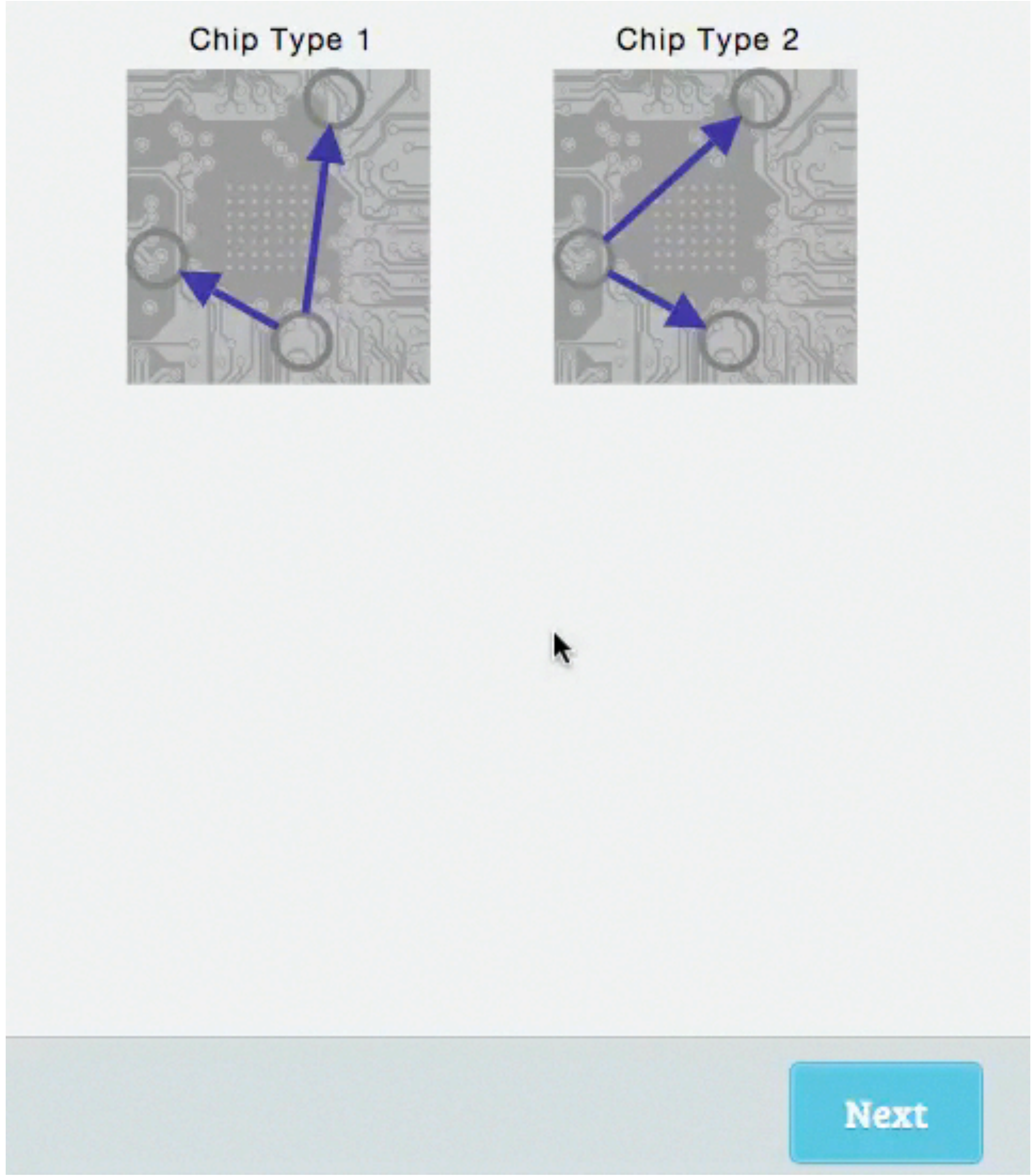
Strategies of learning (Coenen, Rehder, & Gureckis, 2015)

- **Cover story:** *You work in a computer chip factory. There was an accident and the chips got mixed up. Test them to help work out which is which. Each chip could be from one of two possible areas in factory, corresponding to two possible wiring diagrams…*



the **chip factory task**

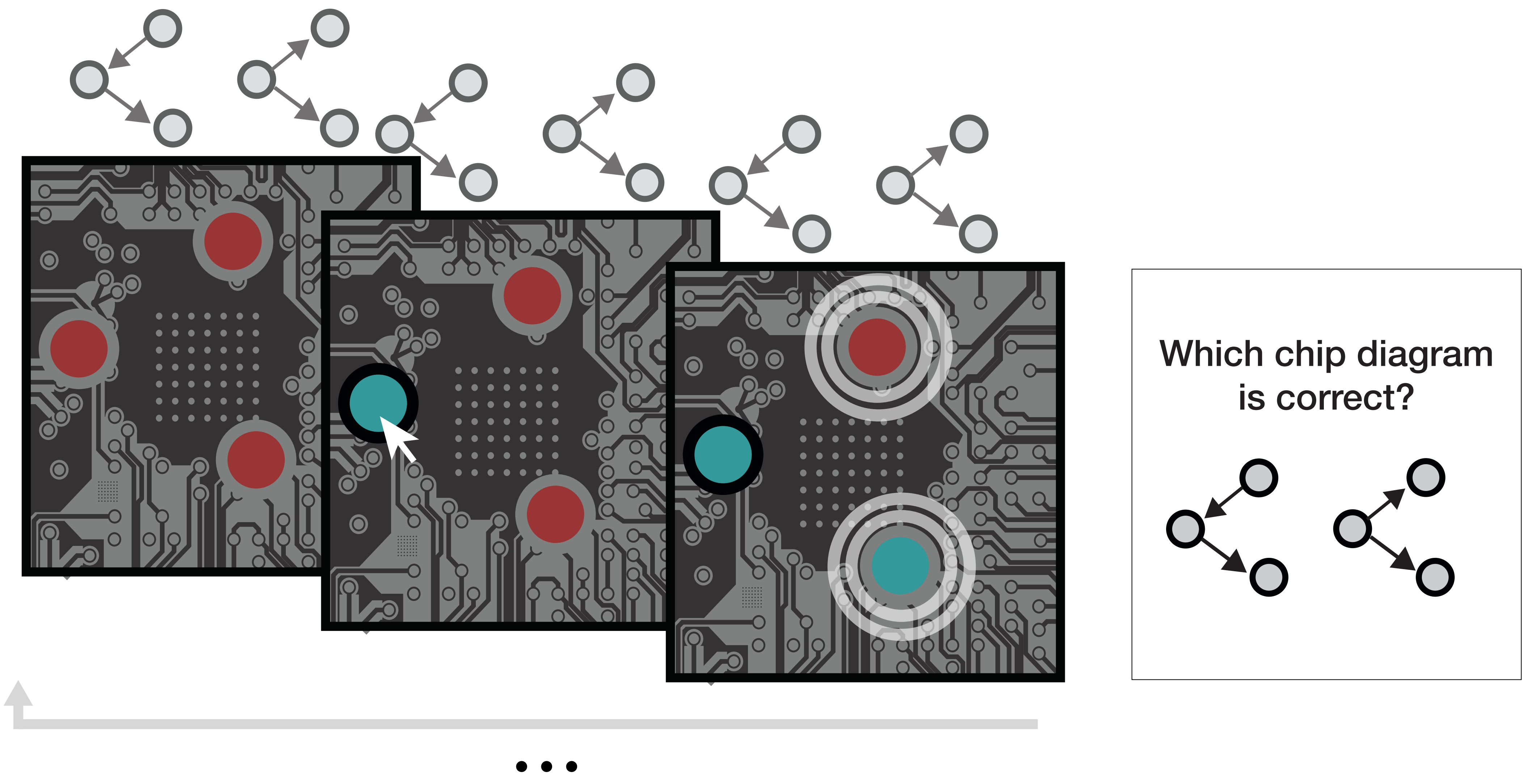
Strategies of learning (Coenen, Rehder, & Gureckis, 2015)



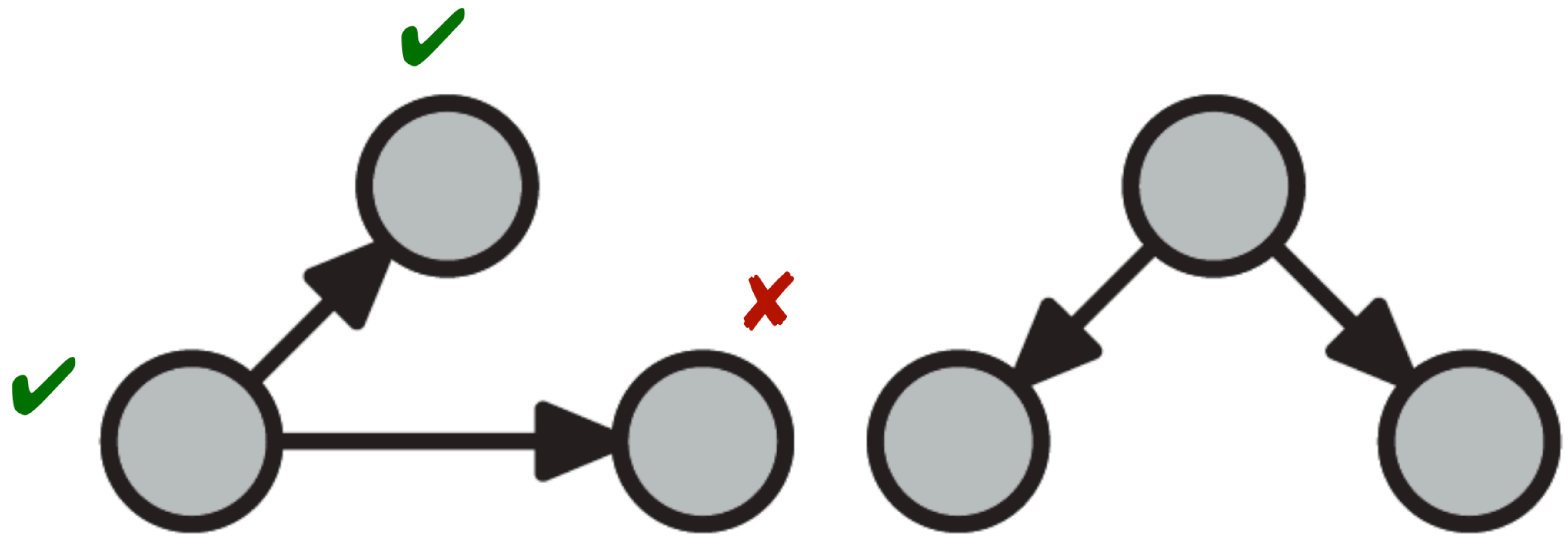
video of **experimental trial**

experiment 1 procedure

- 105 participants on Amazon Mechanical Turk
- test “computer chips” given two hypotheses or **wiring diagrams**
- **intervene** by clicking on chip components
- **incentive for efficiency** is penalty for every intervention after the first
- **network dynamics** are no background causes, p(effect I active cause)=0.8
- 27 problem types (pairs of hypotheses)

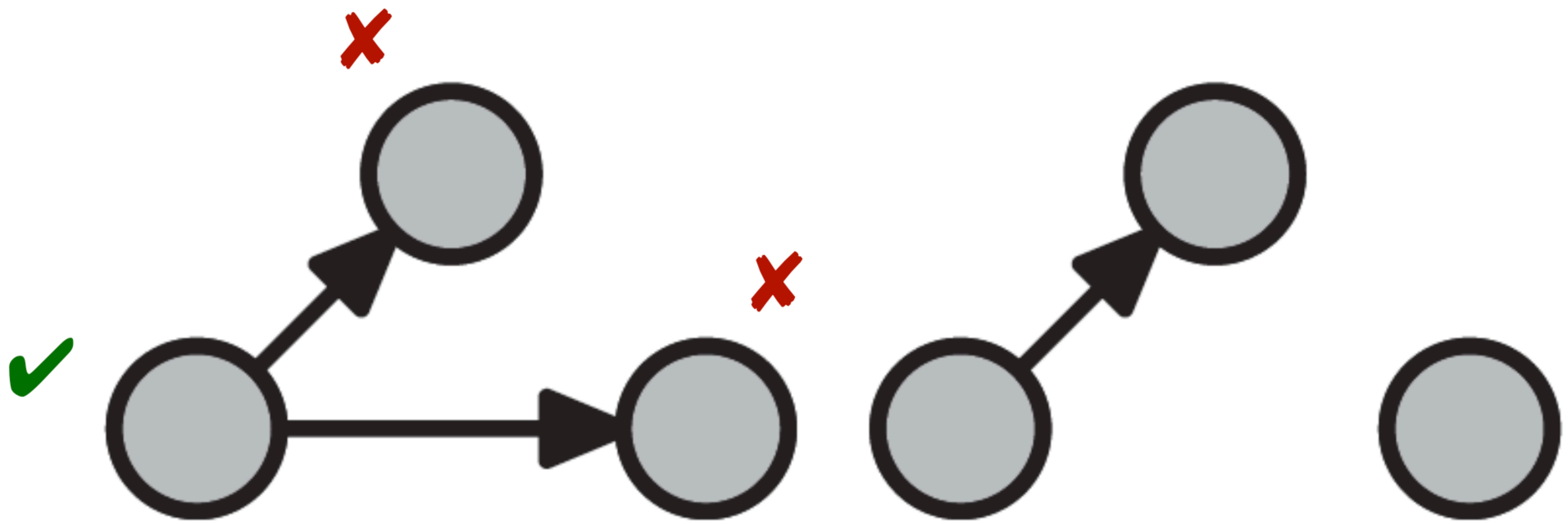


Strategies of learning (Coenen, Rehder, & Gureckis, 2015)



Which would you intervene on?

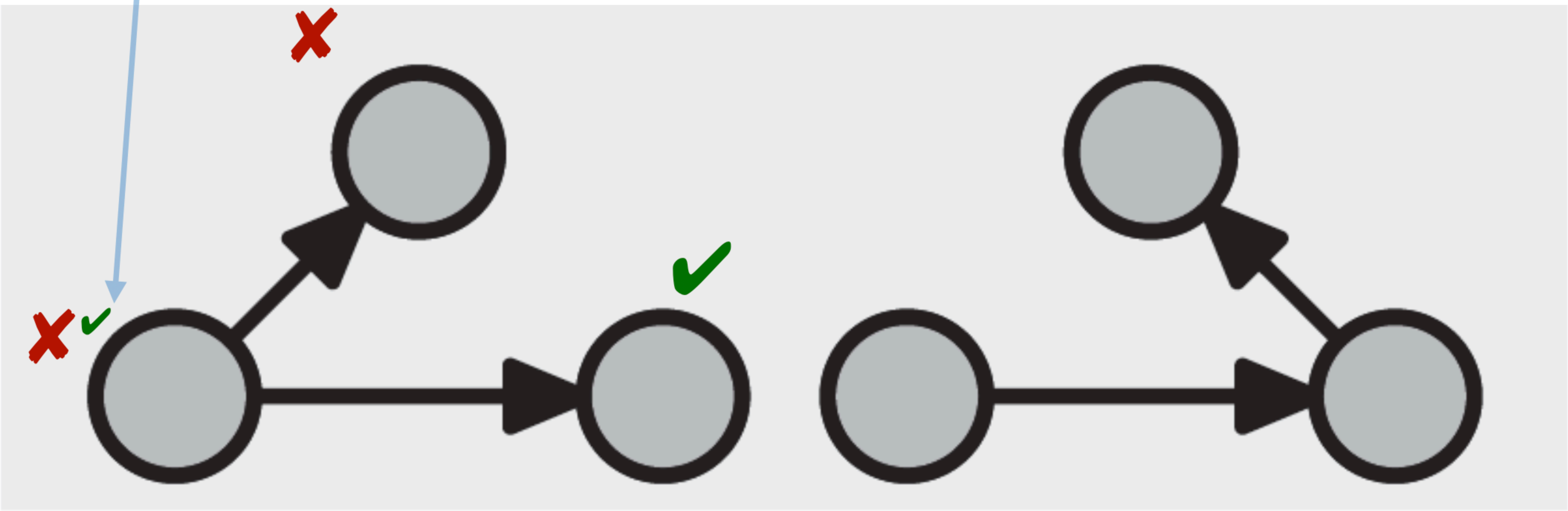
Strategies of learning (Coenen, Rehder, & Gureckis, 2015)



Which would you intervene on?

Strategies of learning (Coenen, Rehder, & Gureckis, 2015)

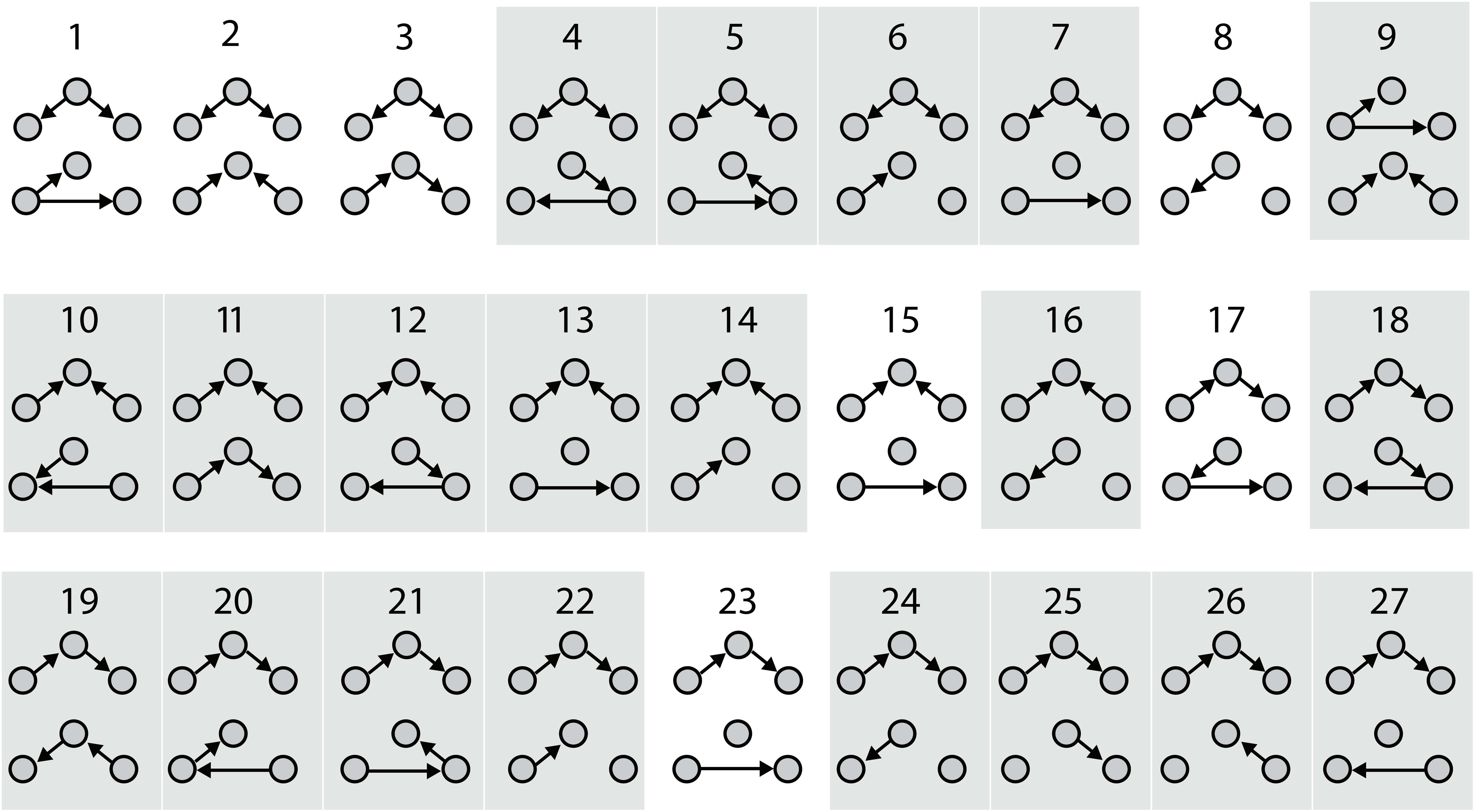
- Top component slightly more likely to activate if LHS model is correct (power=.8)
- Also could activate in absence of bottom right component unlike RHS (base rate=0)
- But intervention on bottom right component is much more informative on average



Which would you intervene on?

Strategies of learning (Coenen, Rehder, & Gureckis, 2015)

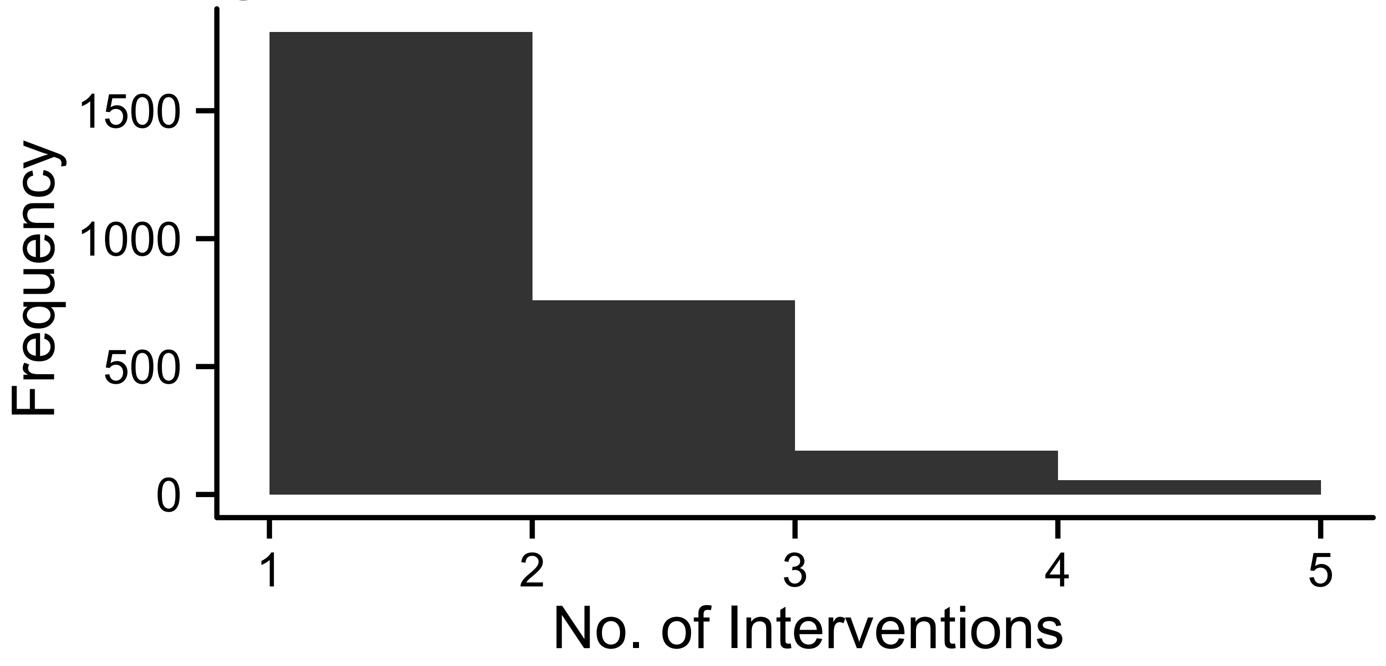
experiment 1 structure tuples



grey boxes = discriminate **PTS** and **IG** predictions

high structure identification **accuracy** (*M = 87%*)

frugal interventions (*M = 1.6*)

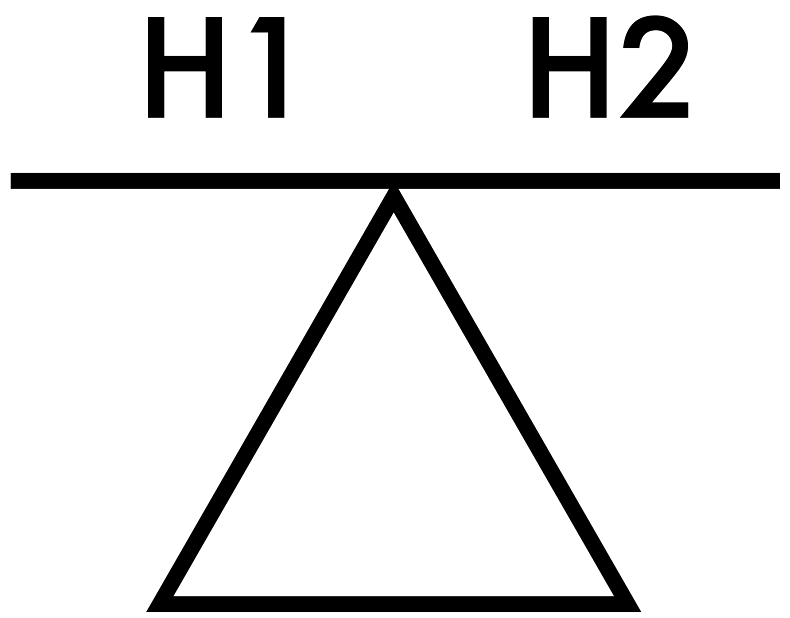


Strategies of learning (Coenen, Rehder, & Gureckis, 2015)

- **Intuitive strategies for learning to learn**

1. Information Gain:

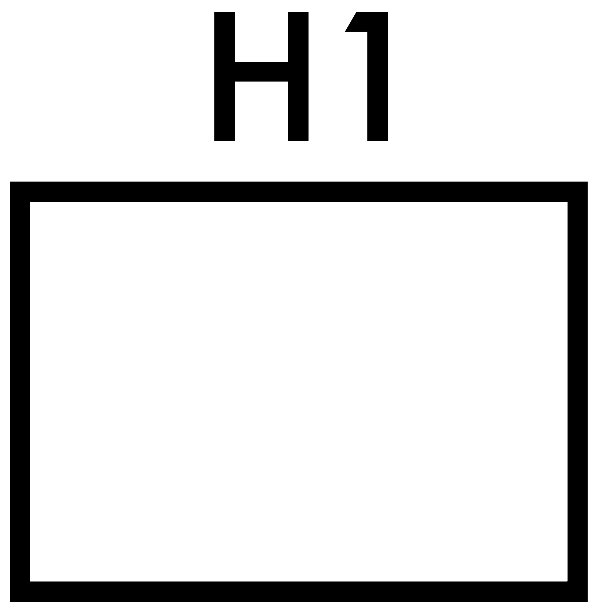
   Compare between multiple hypotheses



(Murphy, 2001; Steyvers et al., 2001)

2. Positive Testing Strategy:

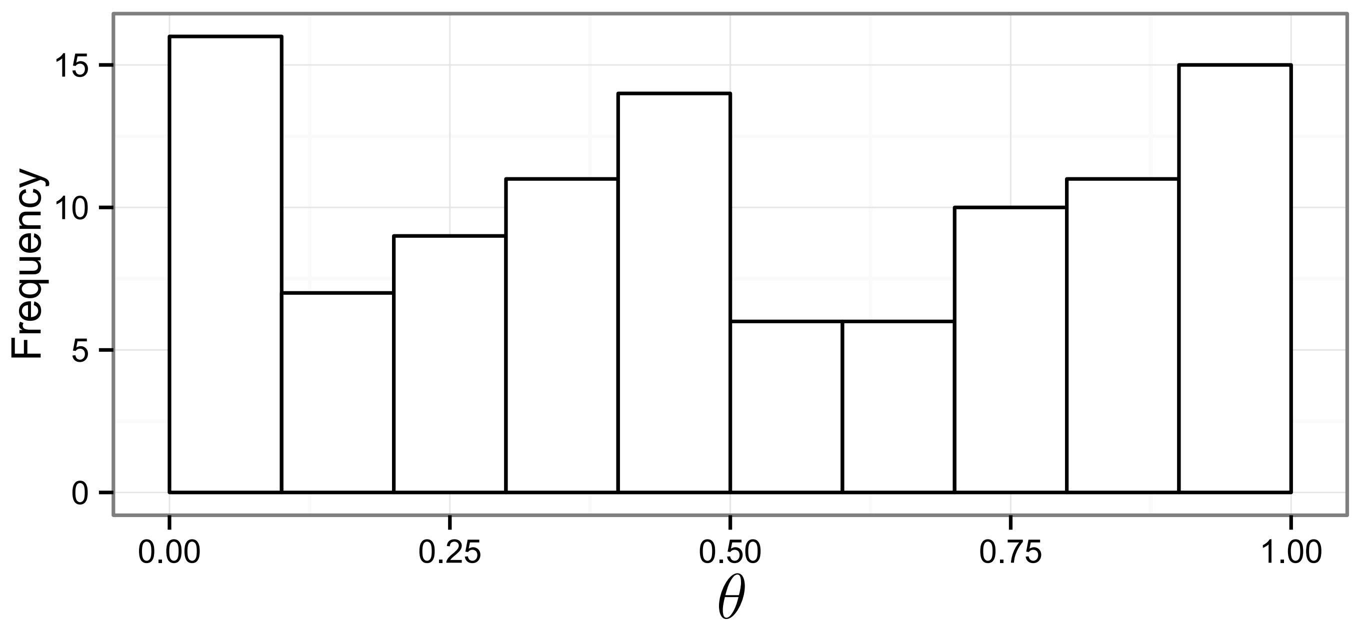
   Verify a single hypothesis



(Klayman & Ha, 1998; Wason, 1960)

Strategies of learning (Coenen, Rehder, & Gureckis, 2015)

Participants fall on a continuum between IG and PTS:

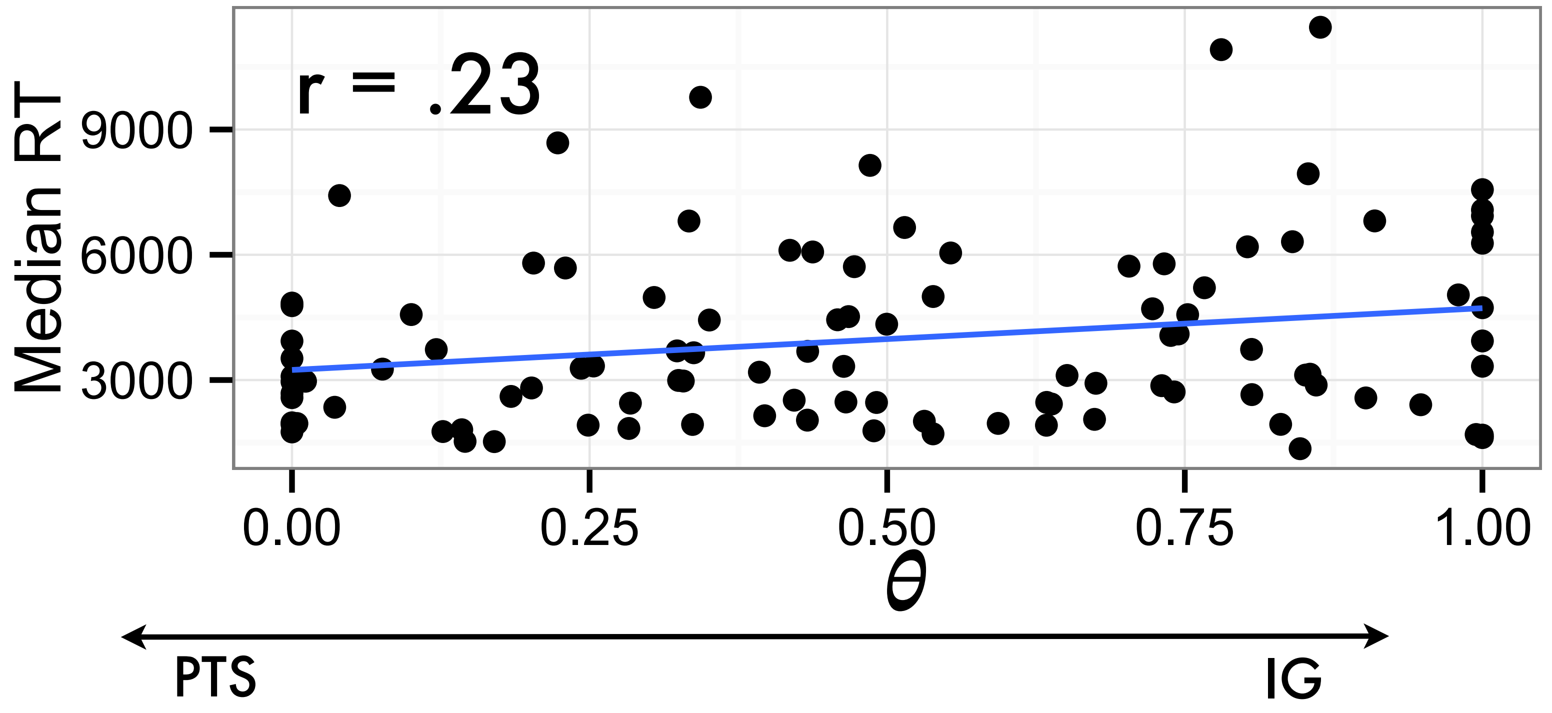


PTS ⟷ **IG**

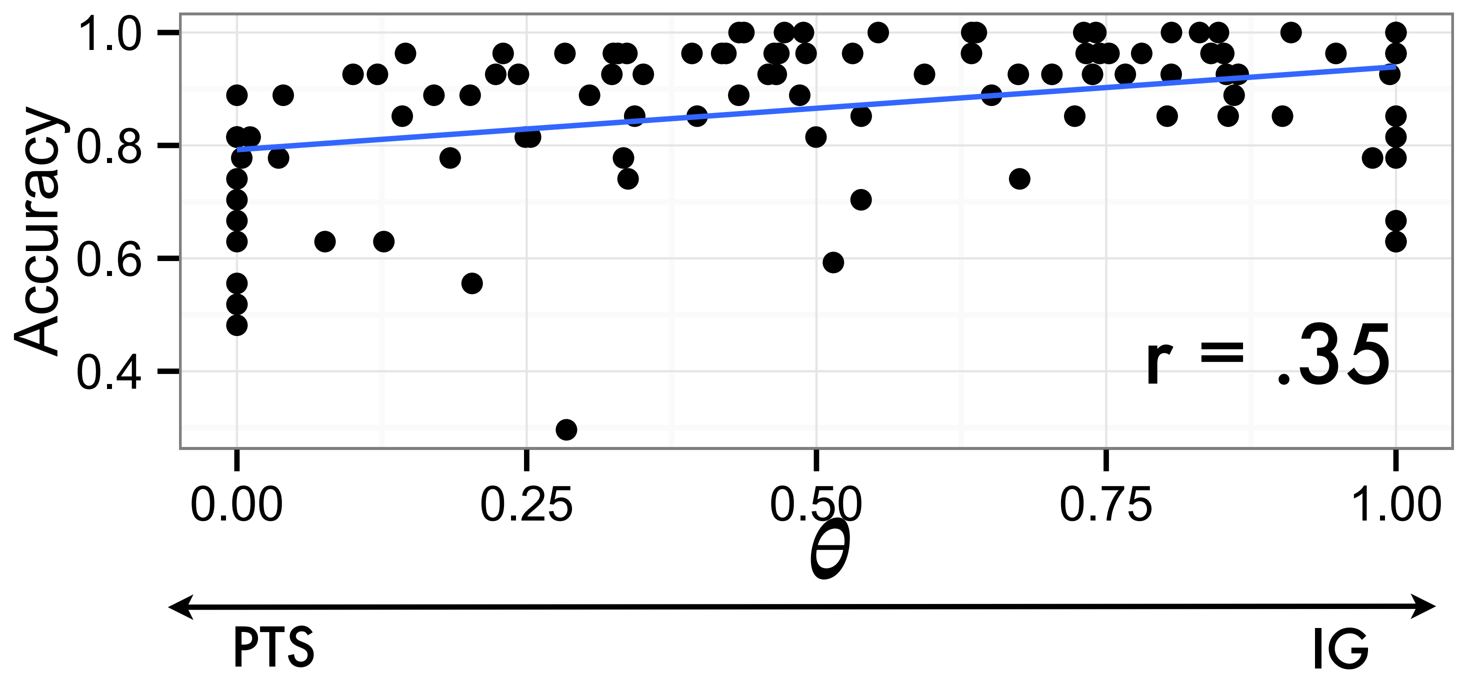
these fits are <u>**real**</u>

θ correlates with:

longer reaction times
(i.e., thinking)



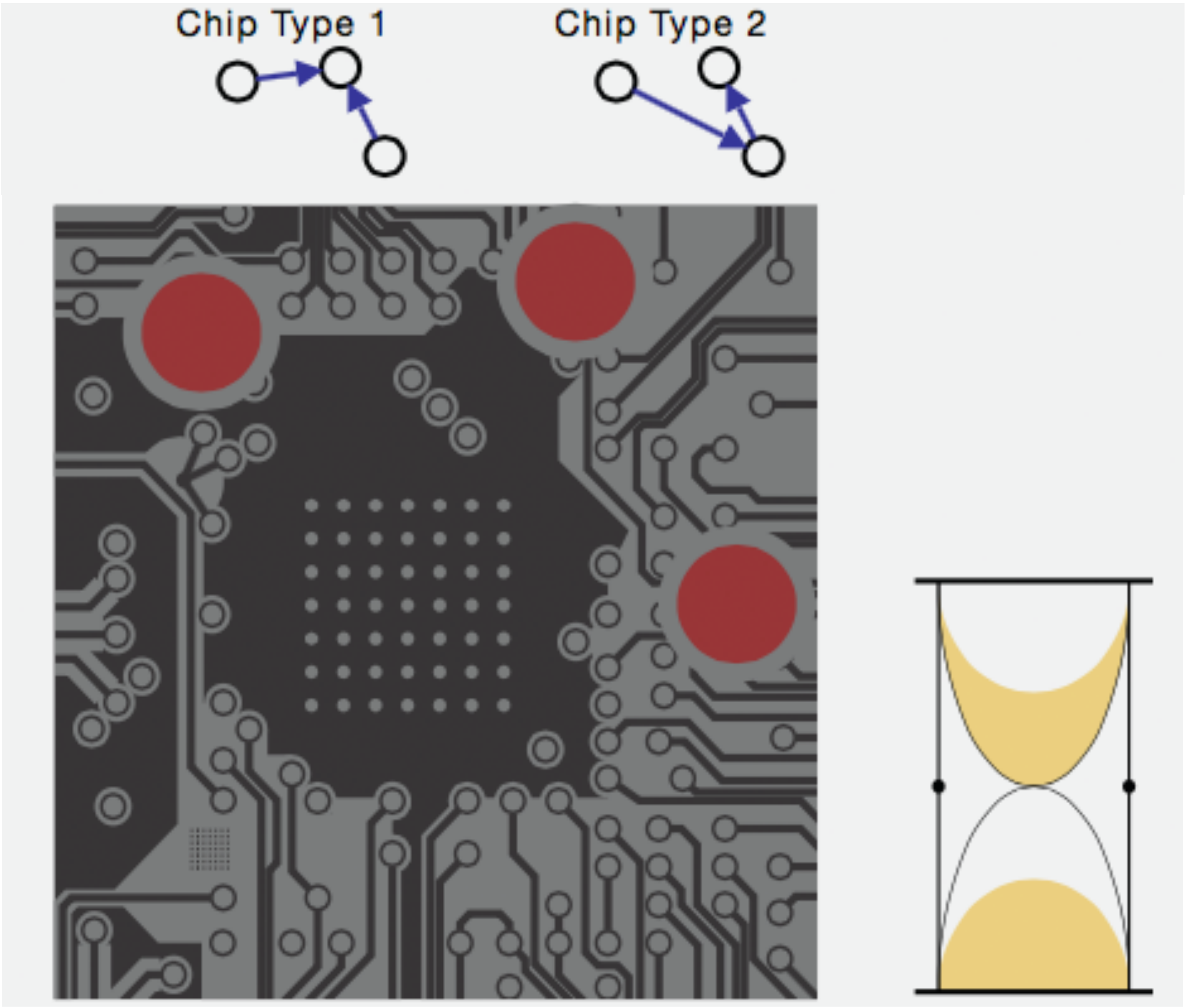
more correct structure choices
(i.e., optimality)



Strategies of learning (Coenen, Rehder, & Gureckis, 2015)

experiment 3: impact of time pressure

- **3 conditions: 4 seconds, 8 seconds, & 60 seconds**



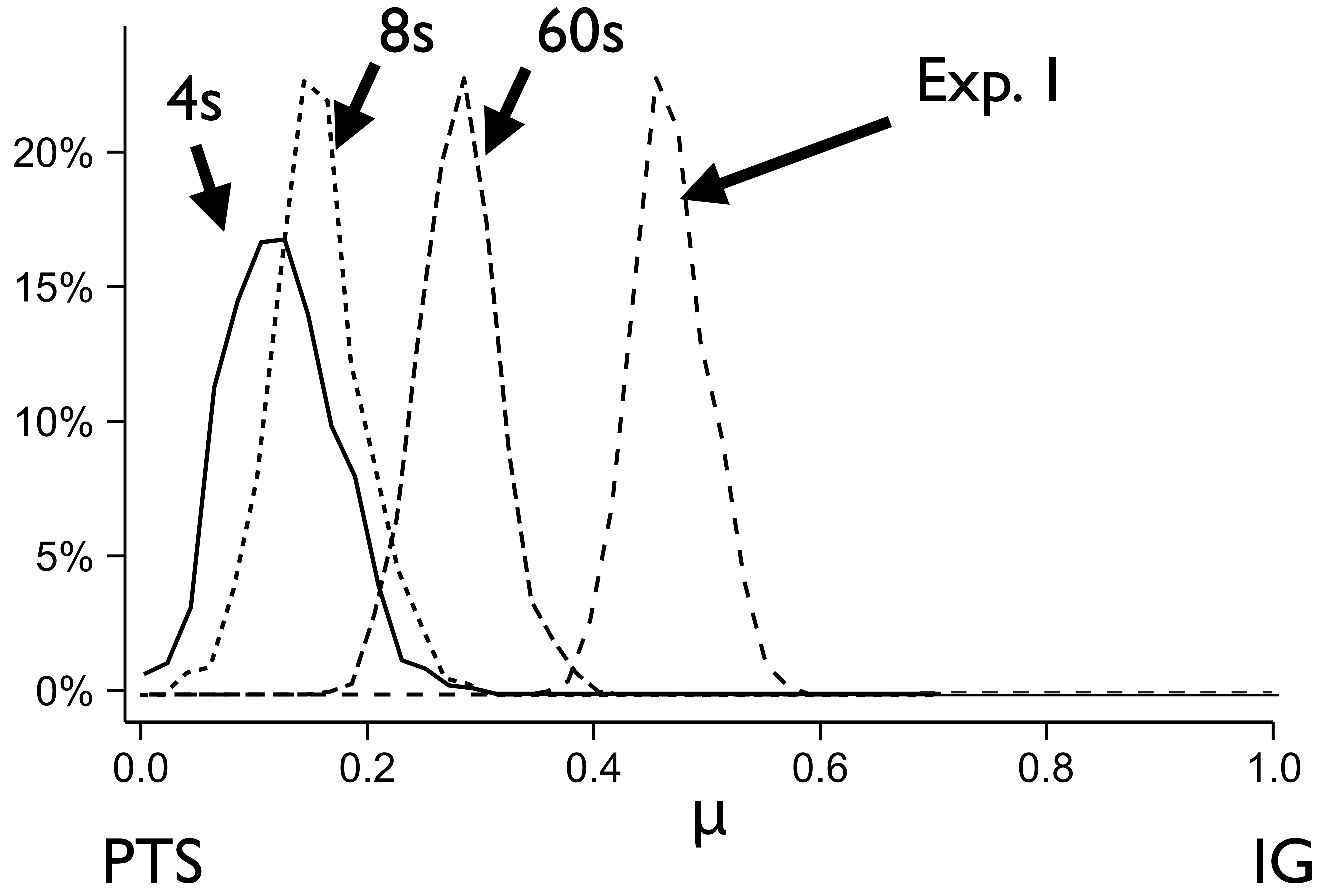
Hourglass to indicate time (and bonus) remaining

Hypothesis:

If computational capacity influences strategy choice, time pressure should increase use of PTS compared to IG

Strategies of learning (Coenen, Rehder, & Gureckis, 2015)

experiment 3 results



Summary

- Active Learning is a framework for deciding how best to act in order to support learning
- It is related to reinforcement learning but the objective function is not to earn immediately reward but to reduce uncertainty about the world.
- Often leads to more efficient information generating actions
- People seem to use a number of information gathering strategies and these may be related to the availability of cognitive resources.

Part 3: Sampling and Resource Rational Cognition

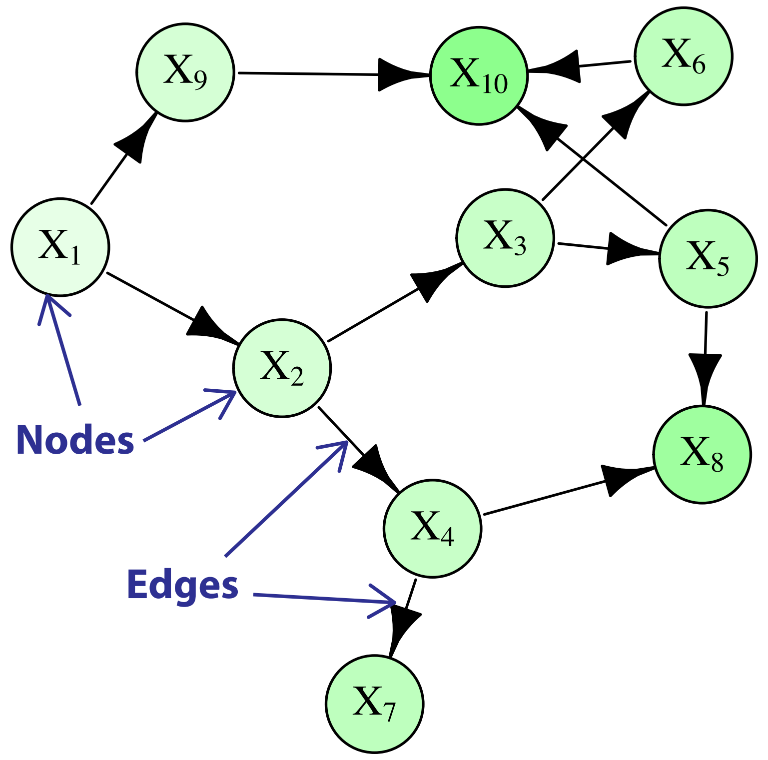
Bayesian inference is hard

The goal: A *predictive* account of behavior that can *emulate* both the general successes and the occasional failures of human performance, and predict which will happen when.

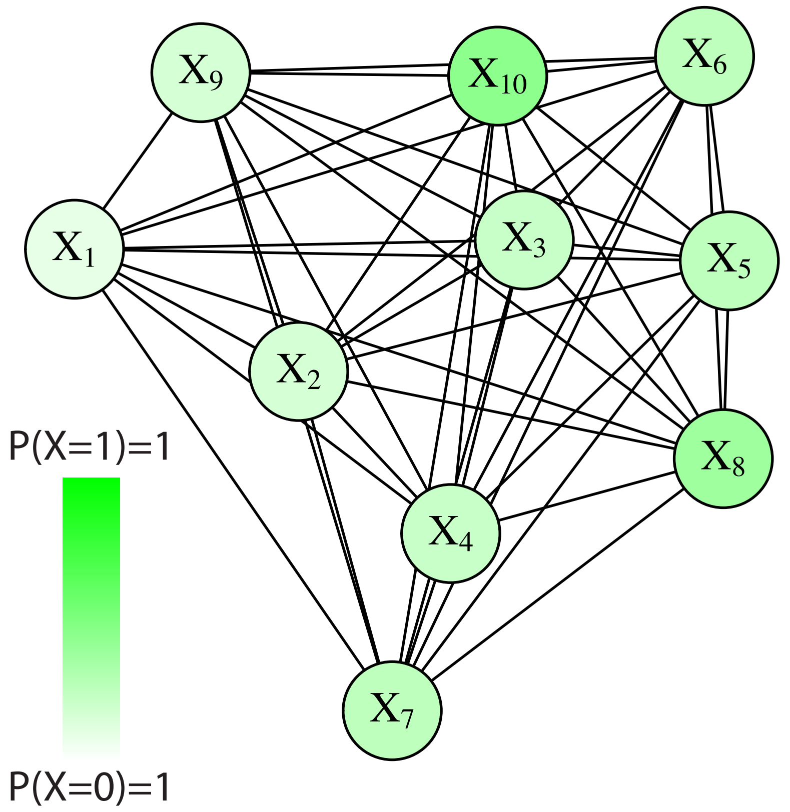
Why *predict?* Why *emulate?* Why emulate *failures?*

- In realistic problems, the number of possible hypotheses can be huge
  - e.g., more than 100,000 clusterings of 10 objects
- In the worst case, the time required to perform exact Bayesian inference increases linearly in the number of hypotheses

True causal structure:

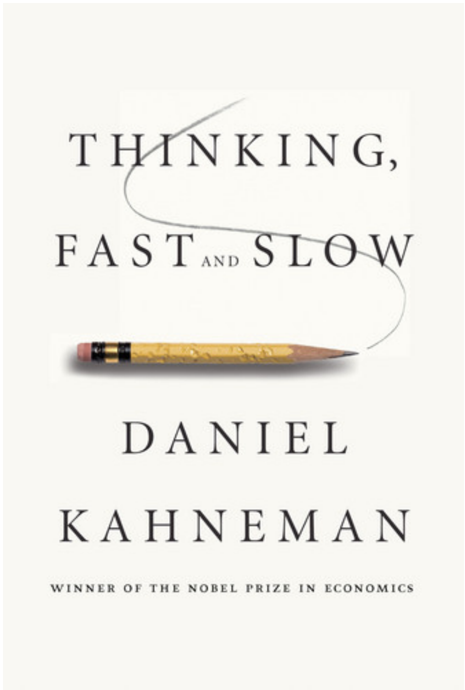


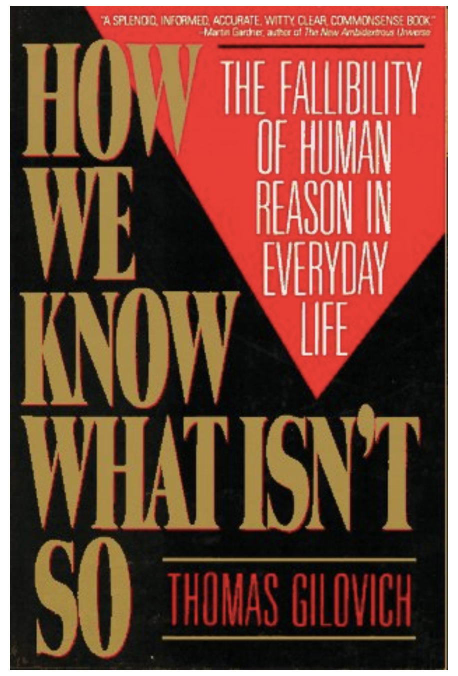
Marginal dependencies:

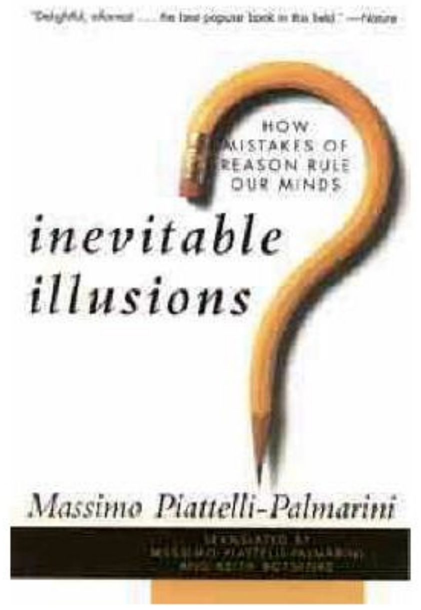


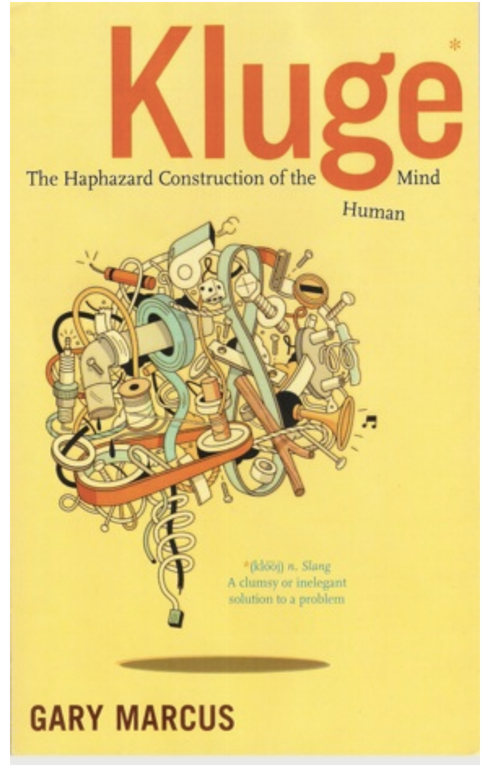
Variables	Structures	Interventions	Outcomes
1	1	3	1
2	3	9	2
3	25	27	4
4	543	81	8
5	29281	243	16
6	3781503	729	32
7	1138779265	2187	64
8	783702329343	6561	128
9	~1213442000000000	19683	256
10	~4175099000000000000	59049	512

Rationality vs. Heuristics



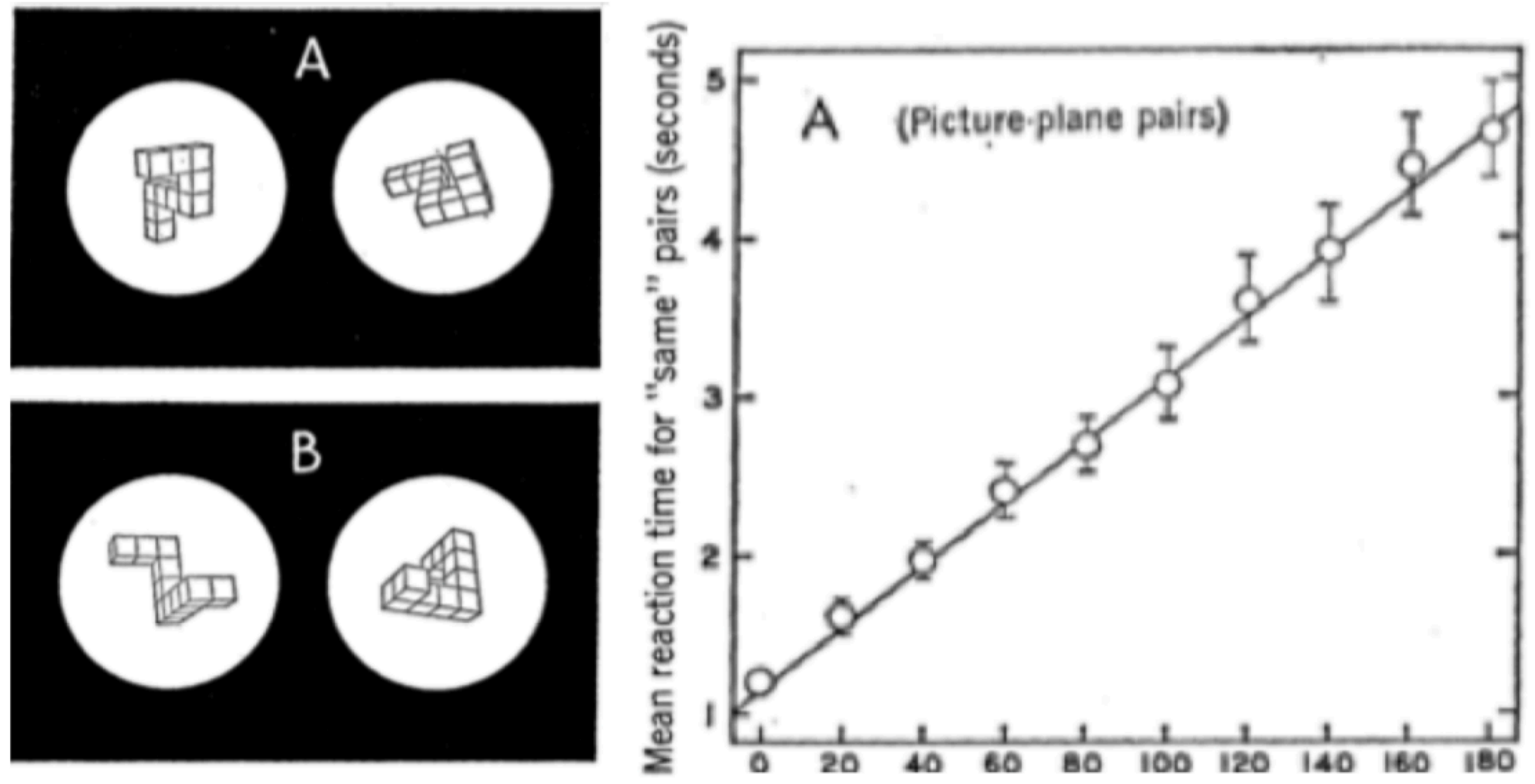






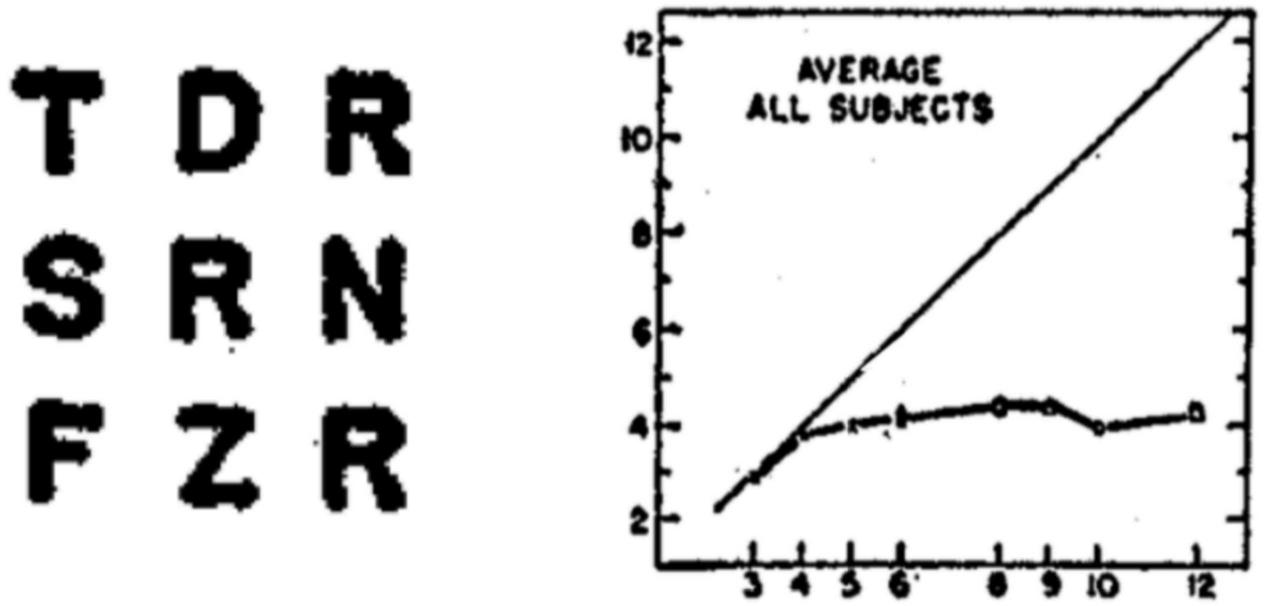
Resource constraints on cognition

Thinking takes time



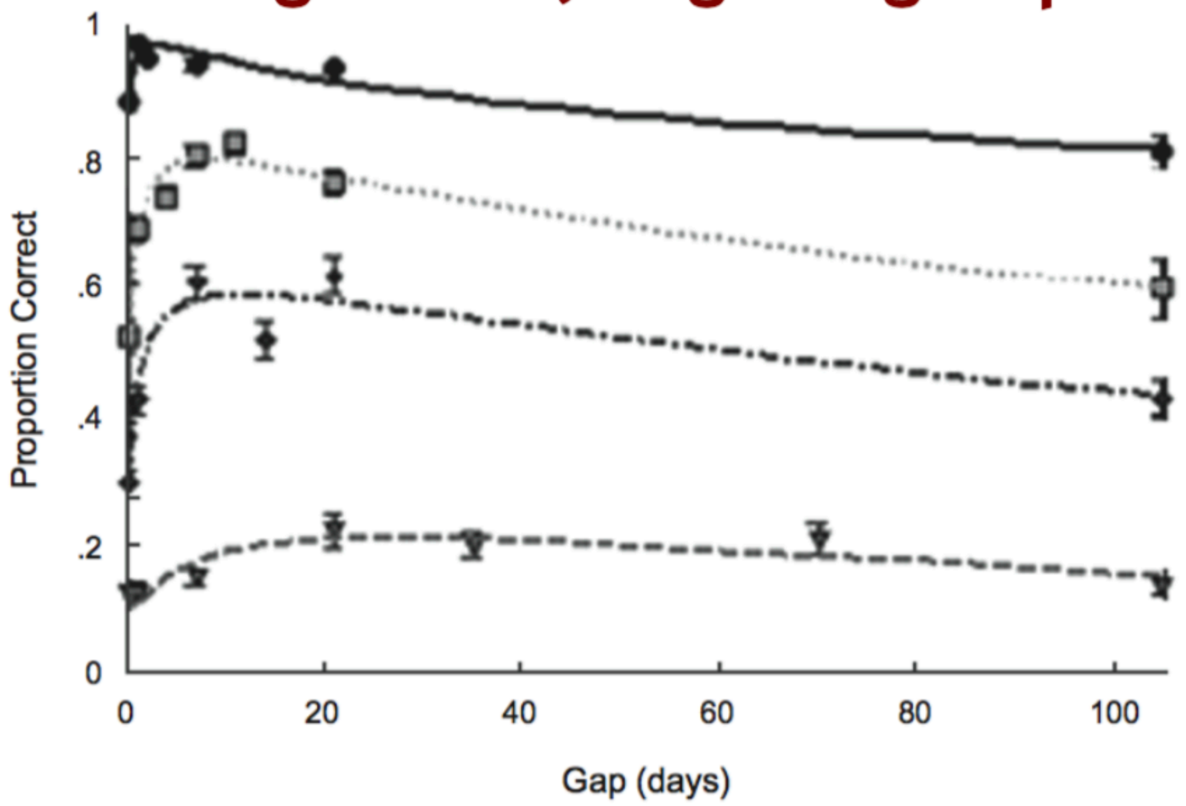
Shepard & Metzler (1971)

Memory stores are limited



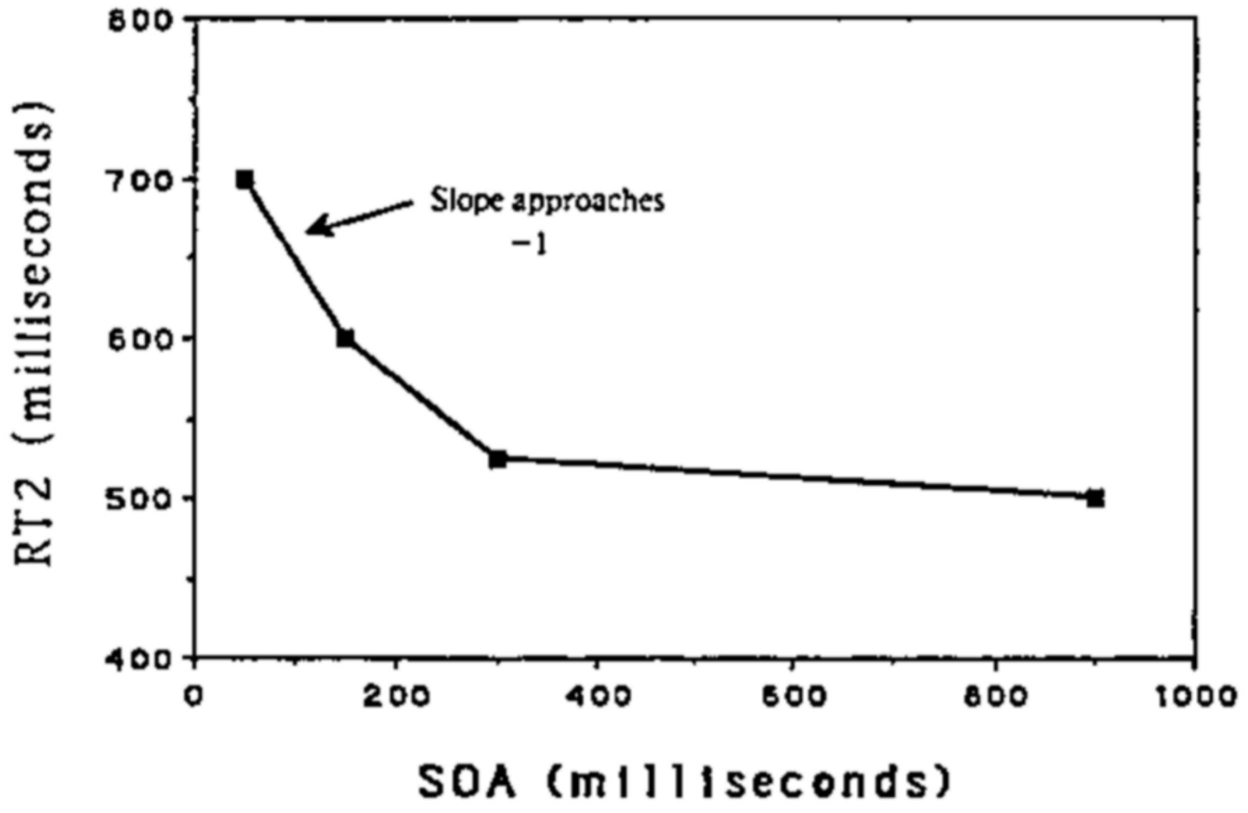
Sperling (1960)

Learning is slow, forgetting is quick



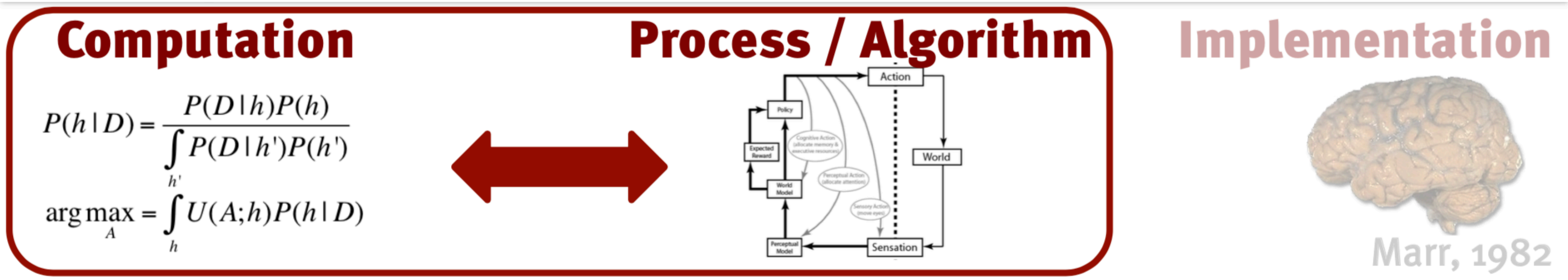
Cepeda, Vul, Rohrer, Wixted, Pahler, 2008

Thinking has a central bottleneck



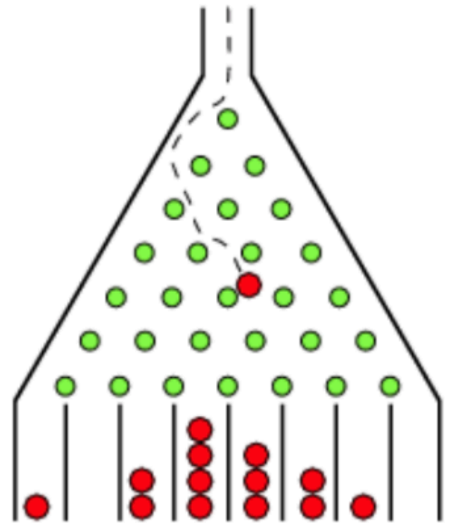
Pashler (1994)

Computation-algorithm interface

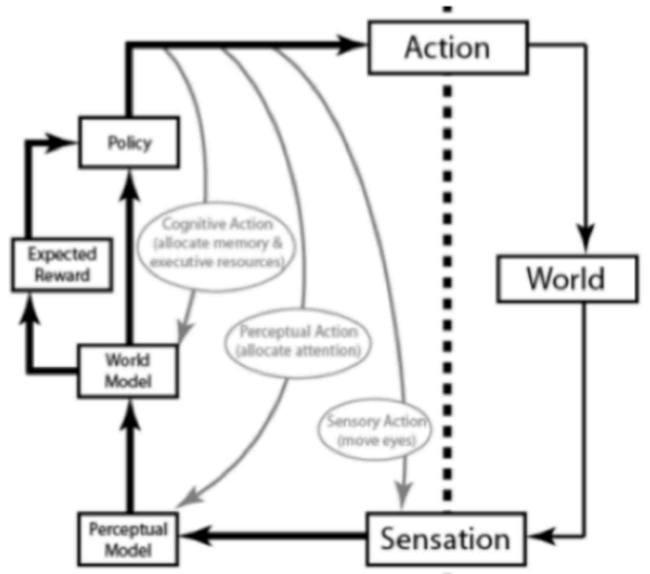


Challenges at the intersection of probabilistic reasoning and cognitive resource constraints:

- How does the mind represent uncertainty?



- How do we use limited cognitive resources?

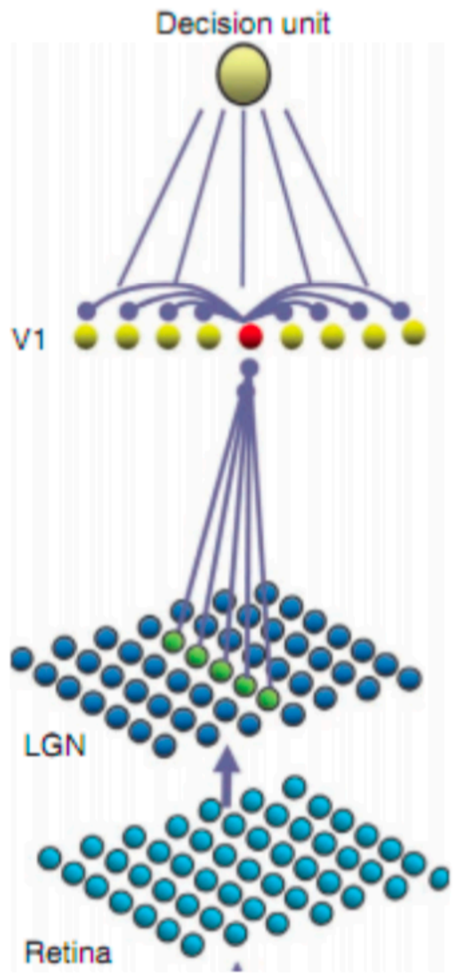


Computation-algorithm interface

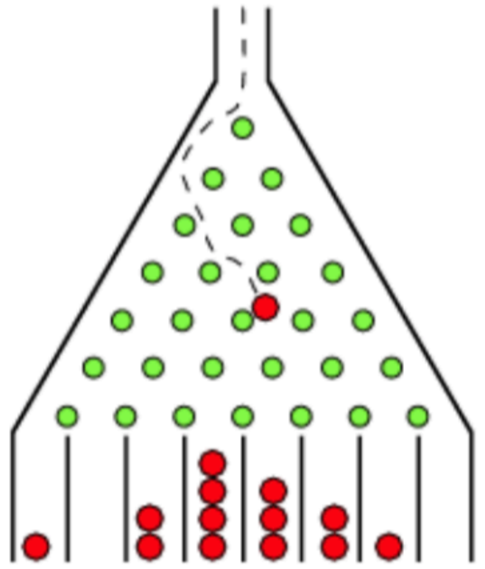
- **Analytical**
  i.e., equations, including variational methods (Friston)

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

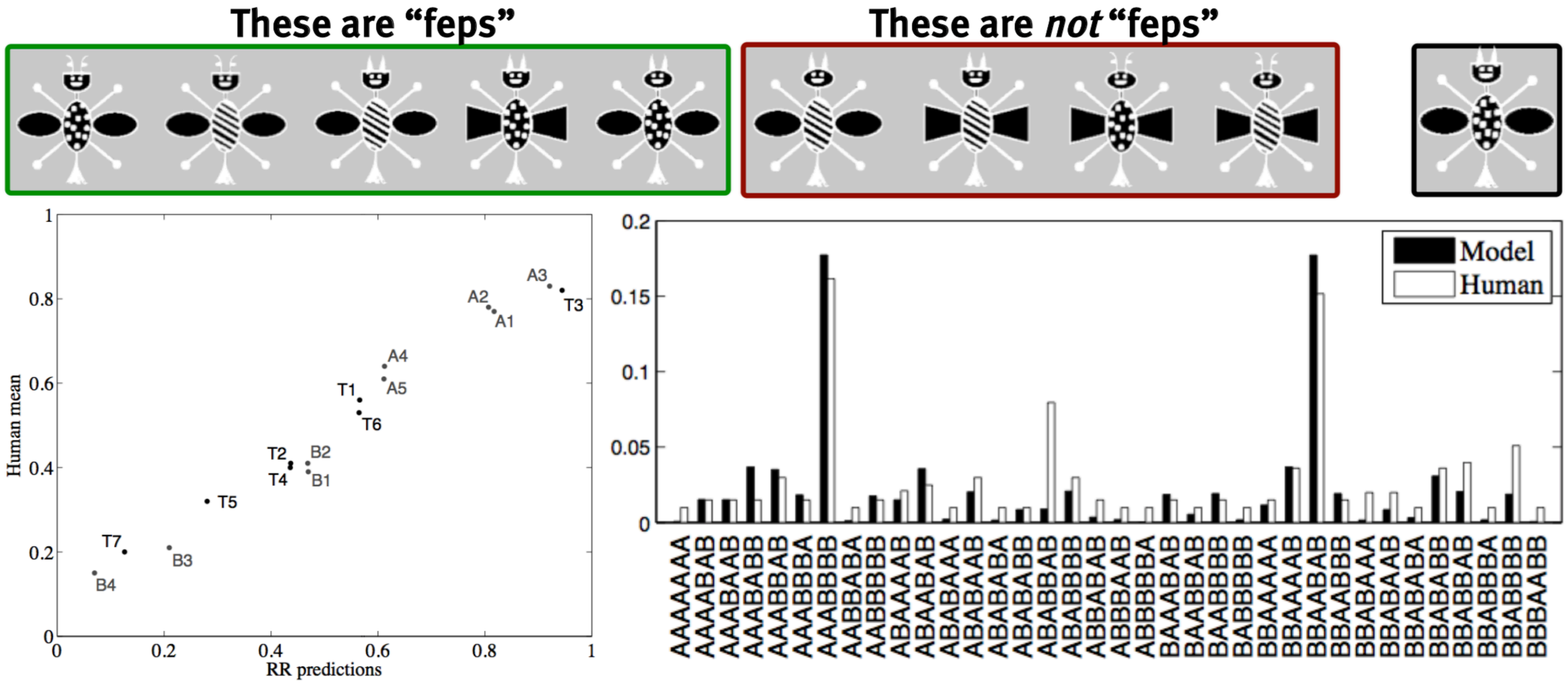
- **Tabular / grids**
  e.g., probabilistic population codes (Pouget/Ma/Beck)



- **Sampling**
  (Goodman/Griffiths/Sanborn/Tenenbaum/Vul, etc.)



Computation-algorithm interface

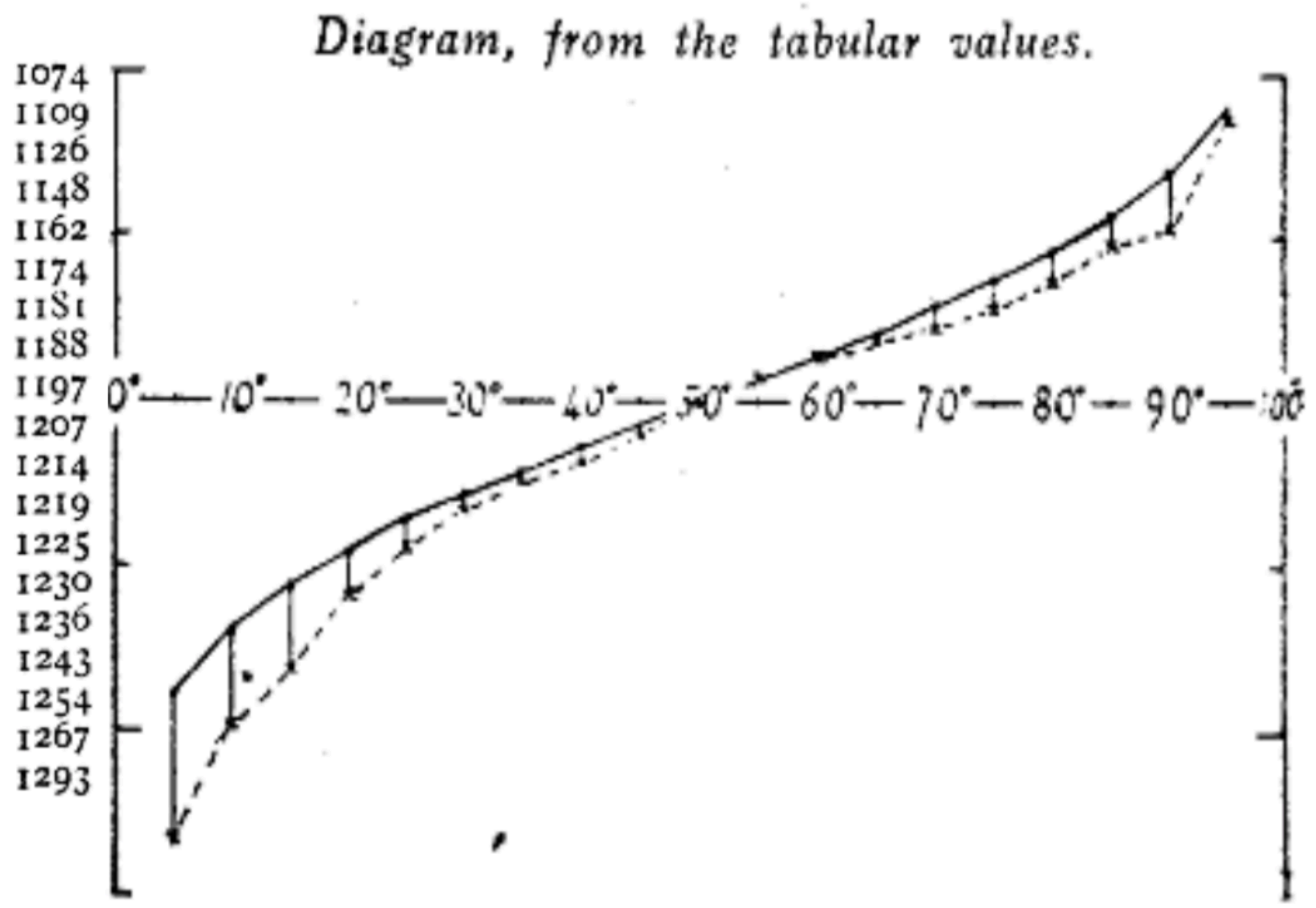


Generalization patterns suggest that each individual was not using a probability distribution over rules, but just one sampled rule.

Goodman, Tenenbaum, Feldman, & Griffiths, 2008

Wisdom of the crowds

Galton, 1907: Vox Populi
How much does an ox weigh?



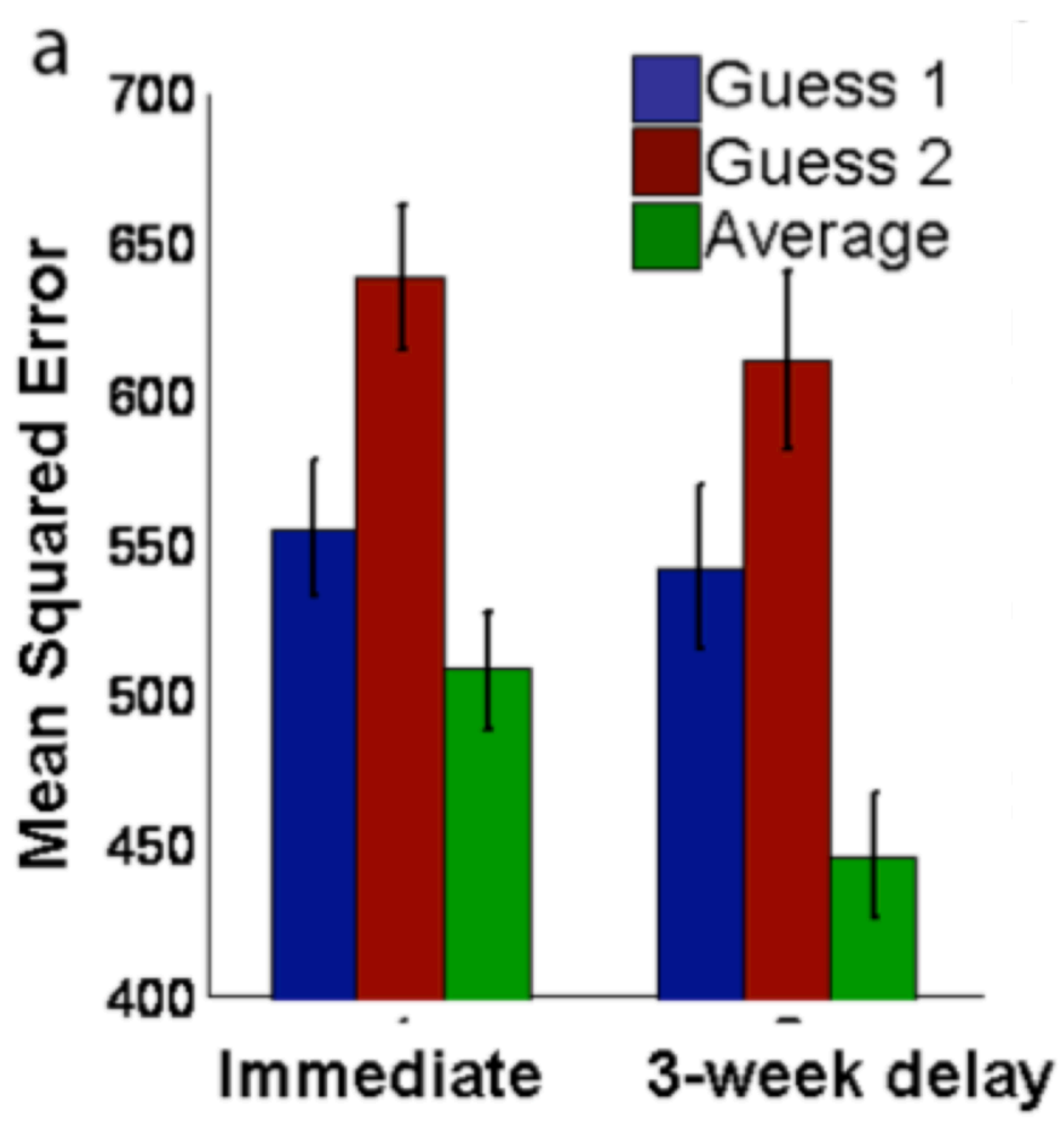
Mean was (1206) was closer to the correct answer (1207) than any one guess.

Benefit of averaging multiple guesses holds so long as errors are independent samples.

Do we get the same effect *within* individuals?

Wisdom of the crowd within

- What percent of the world's airports are in the United States?
- Saudi Arabia consumes what percentage of the oil it produces?
- What percentage of the world' s countries have a higher life expectancy than the United States?



Benefit of averaging multiple guesses from a single individual:

Estimation errors do not arise only from individual biases, but reflect *sampling* under uncertainty.

Vul & Pashler (2008)

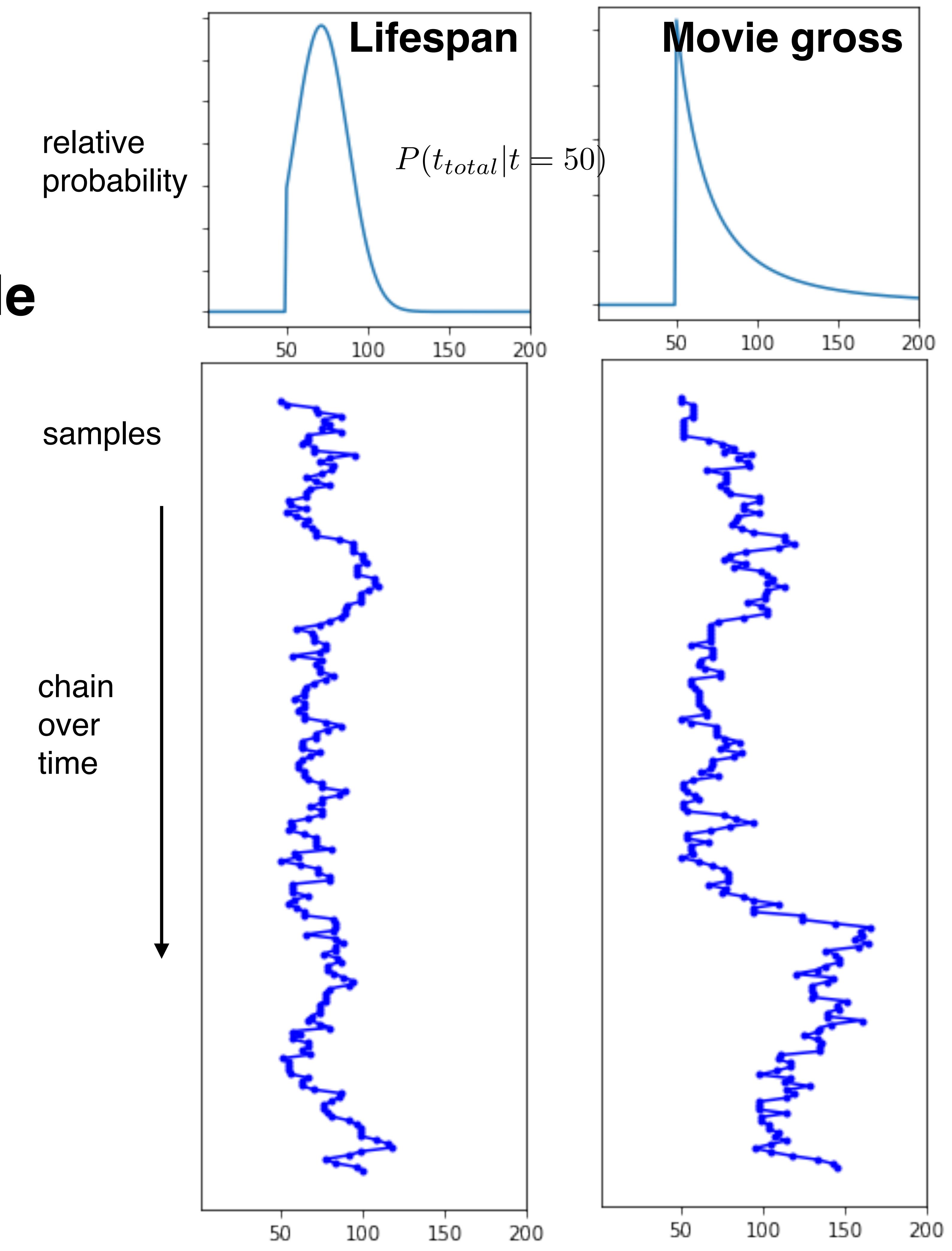
The monte carlo principle

The Monte Carlo principle

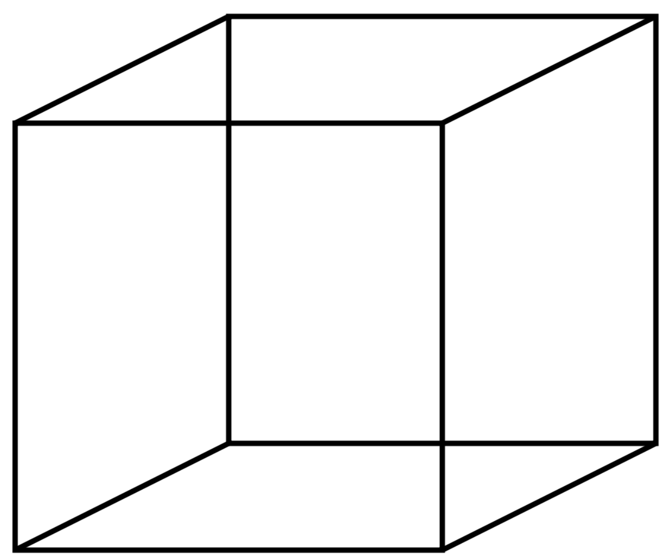
- The expectation of f with respect to P
- can be approximated by

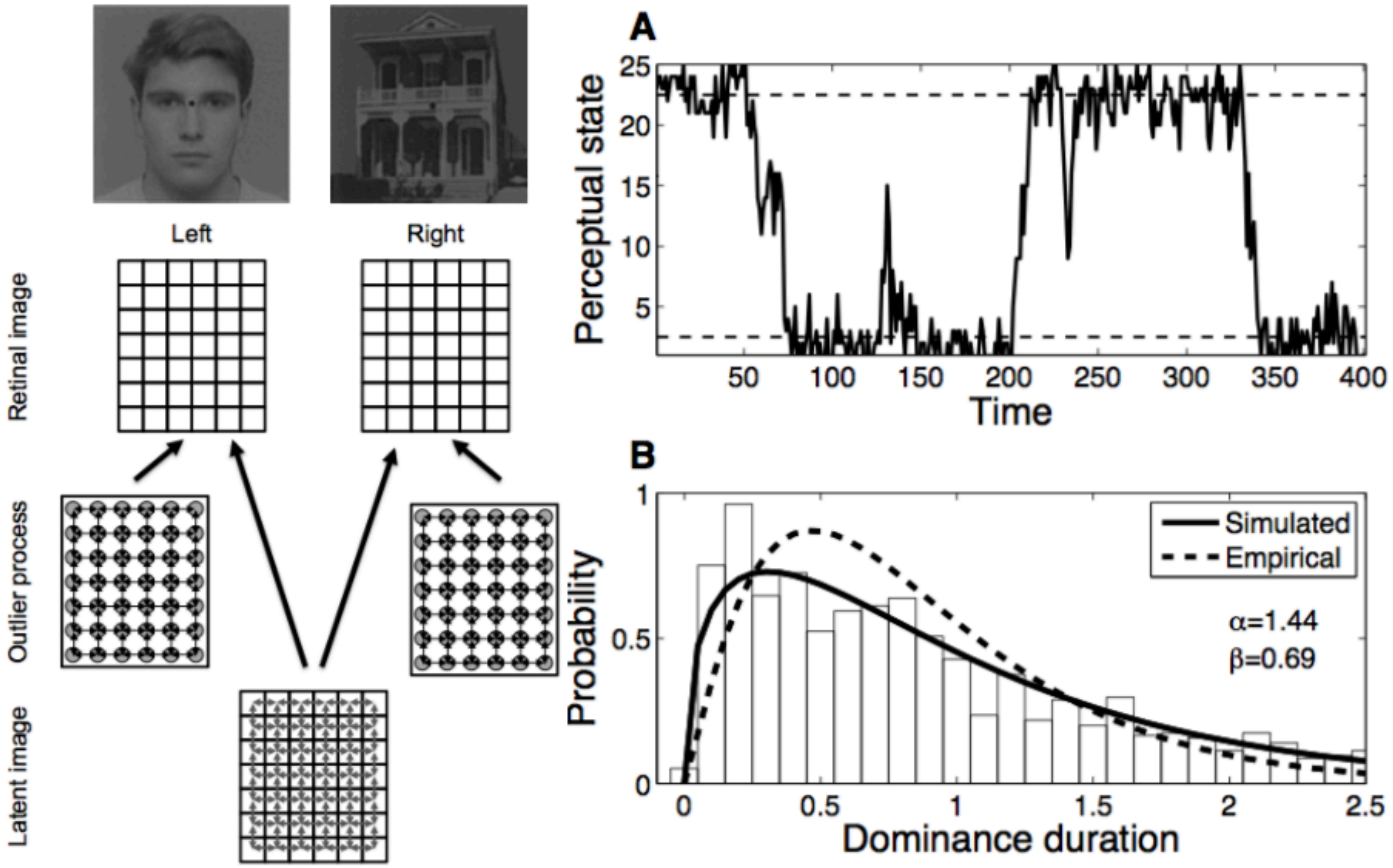
$$E_{P(x)}[f(x)] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

where the x_i are sampled from $P(x)$



The monte carlo principle





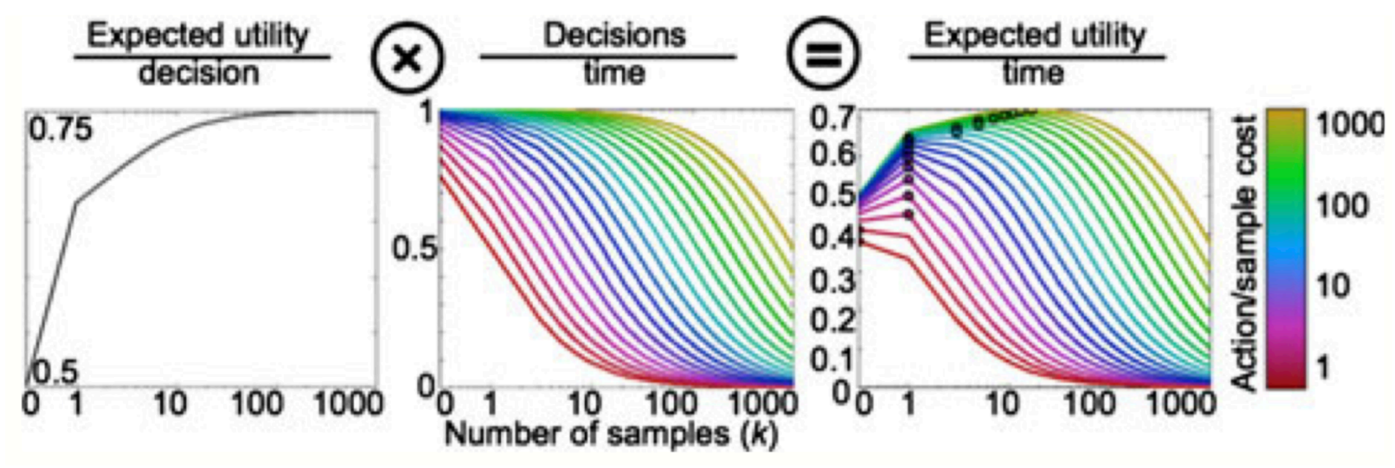
Gershman, Vul, & Tenenbaum; 2011

How many samples?

How many samples?

$$A^* = \arg\max_A \sum_S U(A;S)P(S|D).$$

What does adaptive behavior look like if there is a cost (at least of time) to drawing new samples (computation)?



Vul, Goodman, Griffiths, & Tenenbaum (2014)

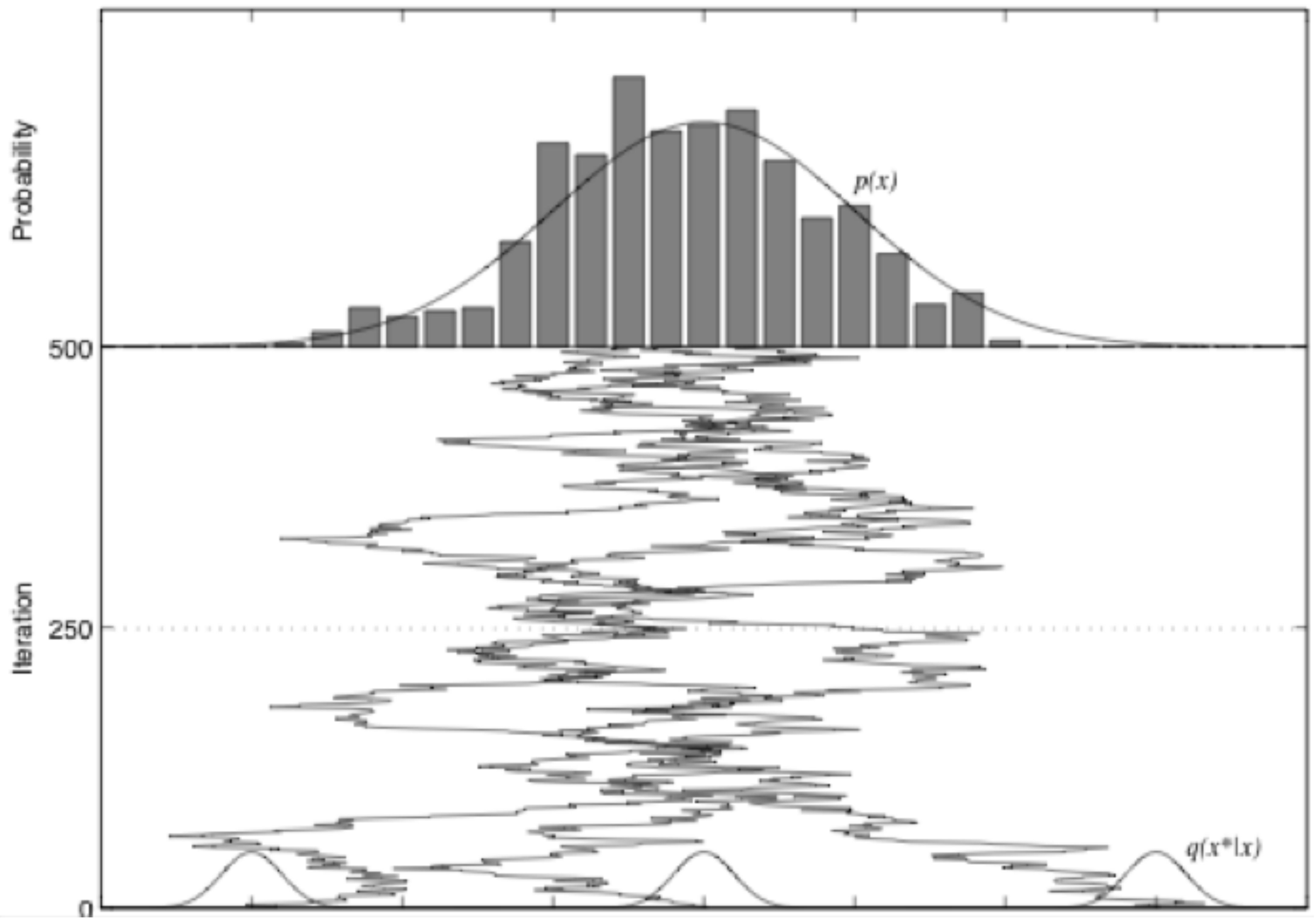
cost of computation
= cost of thinking

Anchoring and adjustment/Metropolis-Hasting

- Is the population of Chicago greater or less than 200,000 people?

- Now guess the population of Chicago

- People give lower estimates when given a lower anchor (200,000) than a higher anchor (5 million) (Jacowitz & Kahneman, 2005)



Sampling as a mechanism

- **Each sampling solution yields different samples:**
  Variability across decisions;
  Variability across trials;
  Variability across participants

- **Solutions outside of the "convergent" regime:**
  Systematic deviations from optimal decisions that will reflect the biases of the sampling algorithm.

Biases

- Probability matching
  Vul et al. 2009
- Anchoring and adjustment
  Lieder et al. 2013
- Sequential effects
  Sanborn et al. 2010
- Garden-pathing
  Levy et al. 2009
- Dynamics of belief change
  Gershman et al. 2012
- Memory reproduction
  Shi et al. 2010

We now face tradeoffs between the effort of computation, variability, and bias of answers. These tradeoffs yield tricky meta-cognitive decisions.

Sampling as a mechanism

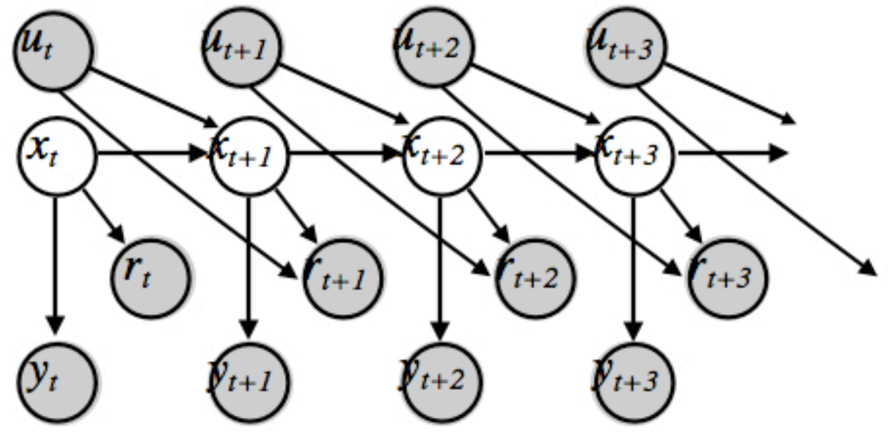
Three levels of description *(David Marr, 1982)*

Computational
Why do things work the way they do?
What is the goal of the computation?
What are the unifying principles?

maximize:

$$R_t = r_{t+1} + r_{t+2} + \cdots + r_T$$

Algorthmic
What representations can implement such computations?
How does the choice of representations determine the algorithm?



Implementational
How can such a system be built in hardware?
How can neurons carry out the computations?