Staying afloat on Neurath's boat

Guest lecture
INFR10054: Computational Cognitive Science

November 4th 2024

Neil Bramley

Ed Psych (& Friends) Movember 2024

Supporting Volunteer Hub at the Royal Edinburgh Hospital

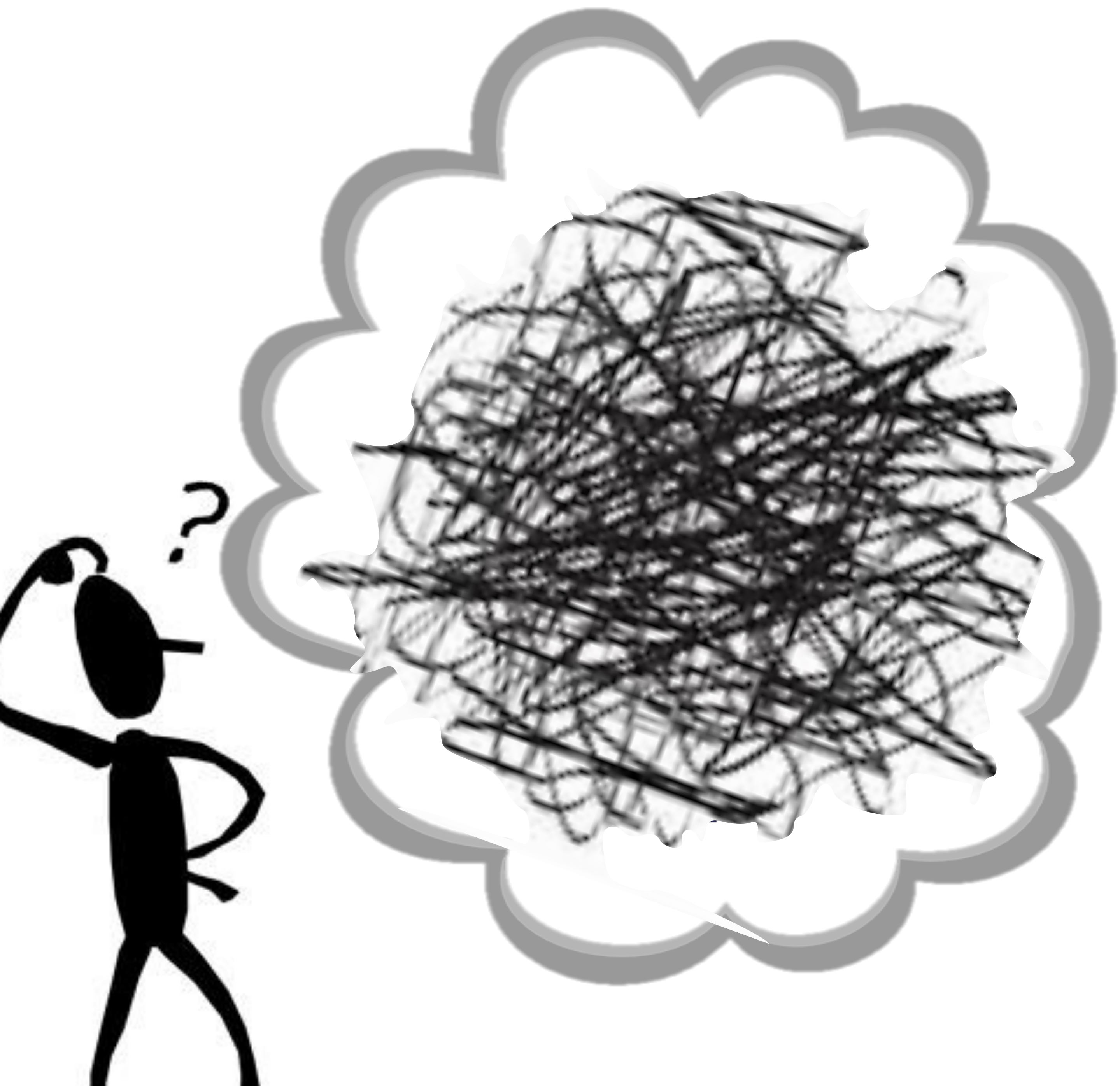
Before:

After…
(artist's impression)

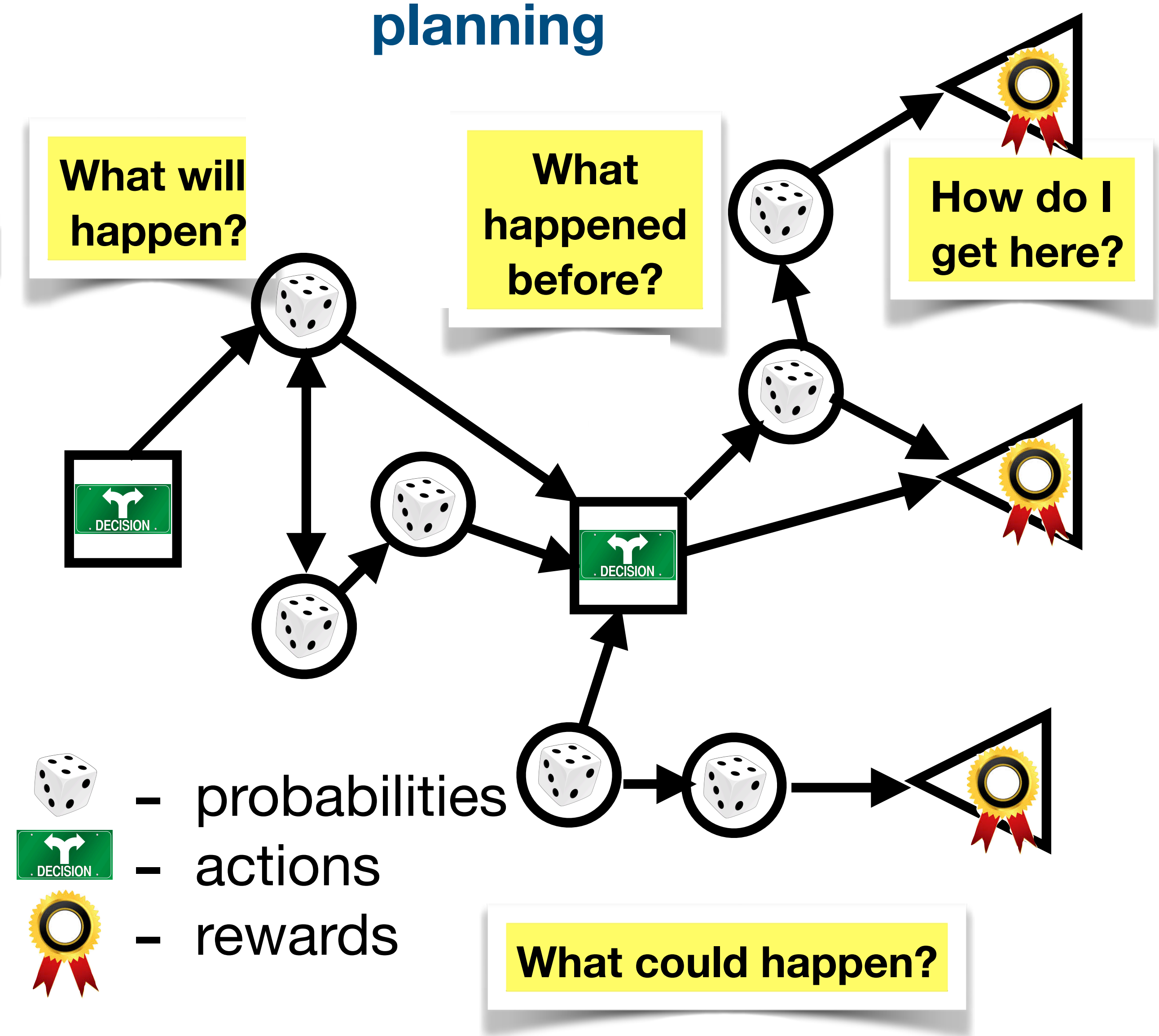
https://www.justgiving.com/page/edpsych-movember-2024

How do we learn how the world works?

We go from “blooming, buzzing confusion” (James, 1890)…



…to a world model that can support **prediction, diagnosis, imagination, planning**



Rational analysis approach to understanding how people come to understand the world
(Anderson, 1990; Marr 1978)

Probability & information theory

Frame problems of human learning

Generate candidate algorithms

Inspired by machine learning

Gather evidence

Adapt, refine, rethink

Experimental approach:

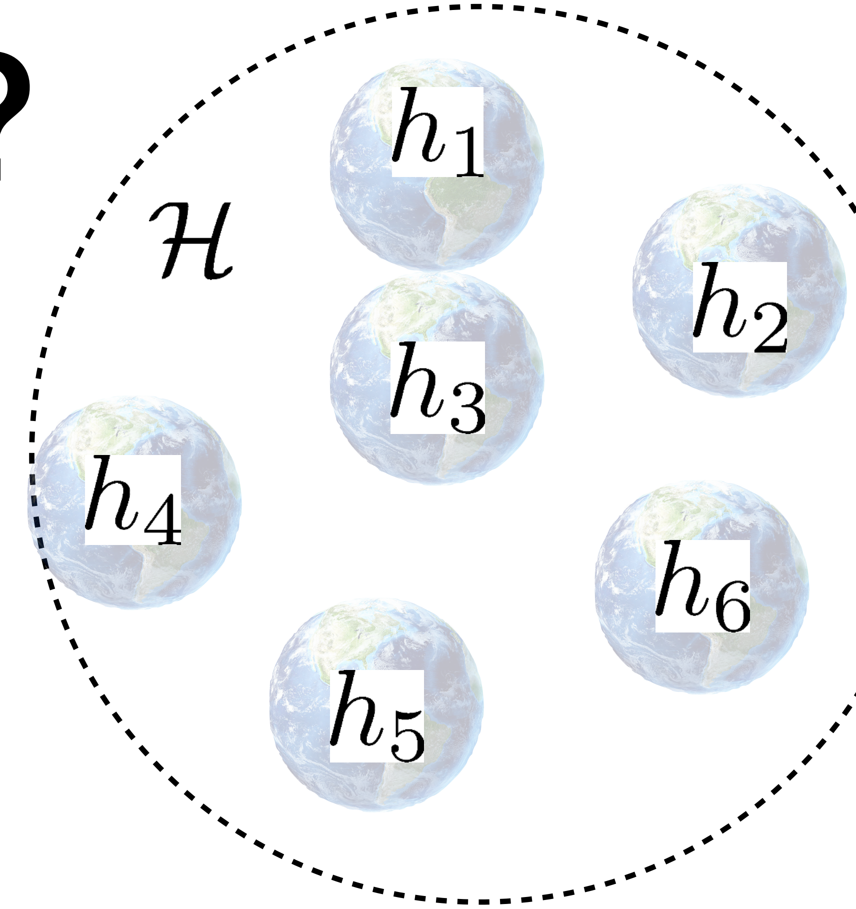
- Study tasks rich enough to reveal how people deal with complexity
- Build models that synthesise individual level behaviour

How do we learn how the world works?

- Can formalise probabilistic inference about world as the "ultimate" *inverse problem* (Bayes & Price, 1736; Earman, 1992; Pizlio, 2001)

$\mathbf{d}$

0010101010
100101110
1011010101
0001010001

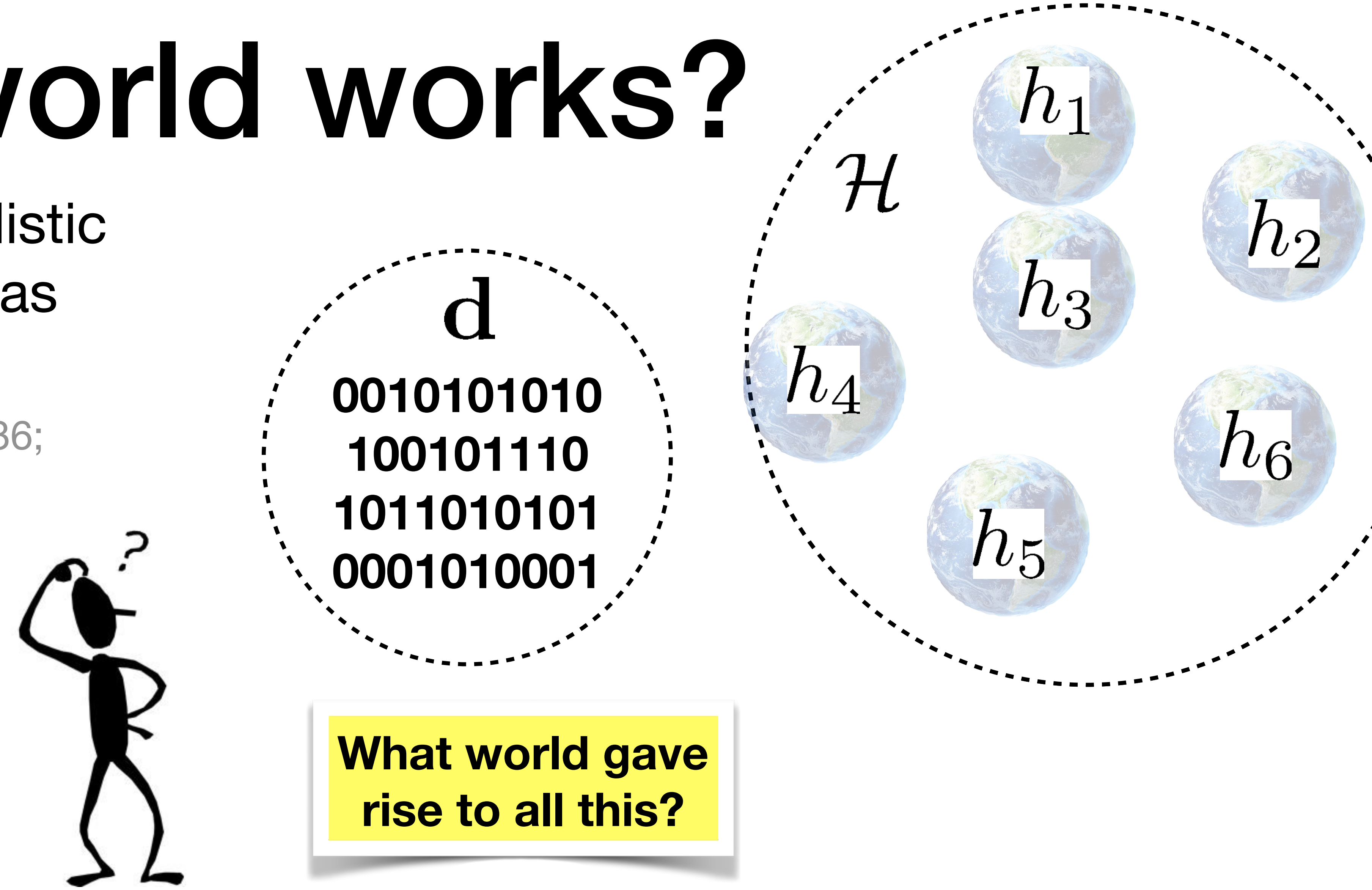


What world gave rise to all this?

$$P(\underbrace{h}_{\text{some hypothesis}}|\overbrace{\mathbf{d}}^{\text{evidence}}) = \frac{\overbrace{P(\mathbf{d}|h)}^{\text{likelihood}}\overbrace{P(h)}^{\text{prior}}}{\underbrace{\sum_{h'\in\mathcal{H}} P(\mathbf{d}|h')P(h')}_{\text{normalizing constant}}}$$

How do we learn how the world works?

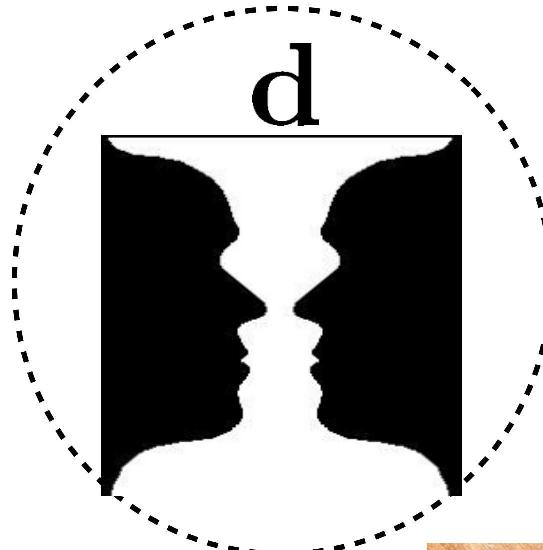
- Can formalise probabilistic inference about world as the “ultimate” *inverse problem* (Bayes & Price, 1736; Earman, 1992; Pizlio, 2001)

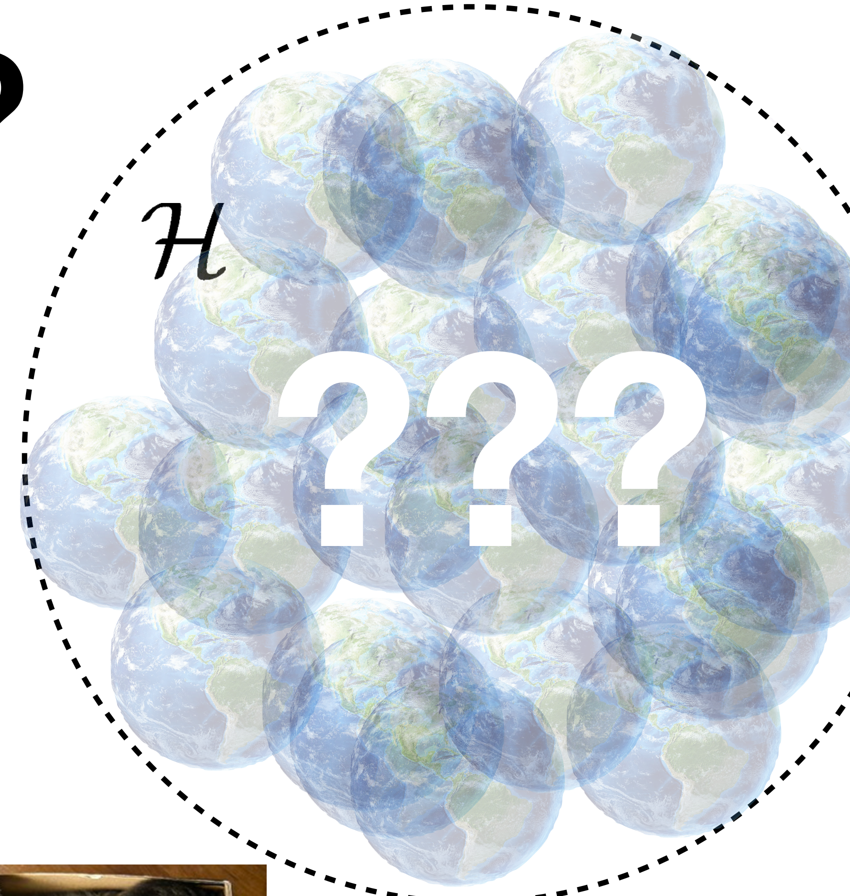


$$P(\underbrace{h}_{\text{some hypothesis}}|\overbrace{\mathbf{d}}^{\text{evidence}}) = \frac{\overbrace{P(\mathbf{d}|h)}^{\text{likelihood}}\overbrace{P(h)}^{\text{prior}}}{\underbrace{\sum_{h'\in\mathcal{H}} P(\mathbf{d}|h')P(h')}_{\text{normalizing constant}}}$$

How do we learn how the world works?

- Can formalise probabilistic inference about world as the “ultimate” *inverse problem* (Bayes & Price, 1736; Earman, 1992; Pizlio, 2001)





BUT:

- *H:* infinite/unknown
- P(*H) &* historical **d**: too large to store/compute over
- **d**: causally ambiguous

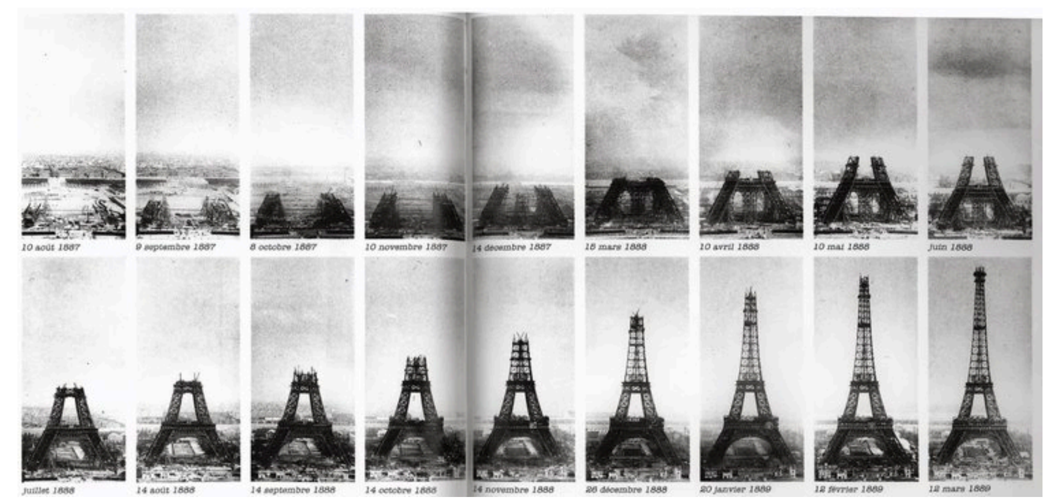
$$P(\underbrace{h}_{\text{some hypothesis}}|\overbrace{\mathbf{d}}^{\text{evidence}}) = \frac{\overbrace{P(\mathbf{d}|h)}^{\text{likelihood}}\overbrace{P(h)}^{\text{prior}}}{\underbrace{\sum_{h'\in\mathcal{H}} P(\mathbf{d}|h')P(h')}_{\text{normalizing constant}}}$$

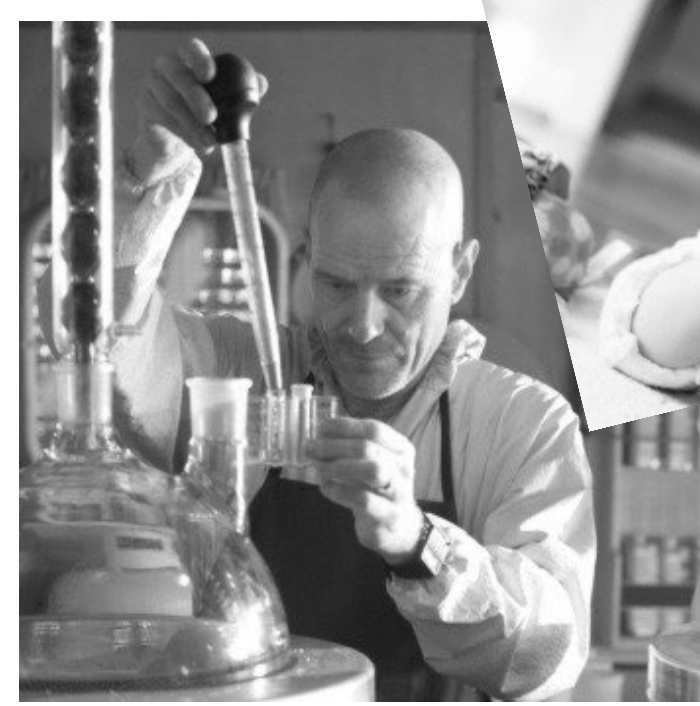
How do we learn how the world works?

I will argue humans use combination of tricks to make learning possible:

- Hypothesis space infinite/unknown → Explore stochastically/incrementally.
- Data large/weak → bootstrap accumulated knowledge
- Data causally ambiguous → *Act on world* in ways that distinguish causes from effects (c.f. Gopnik, Meltzoff & Kuhl, 1999)

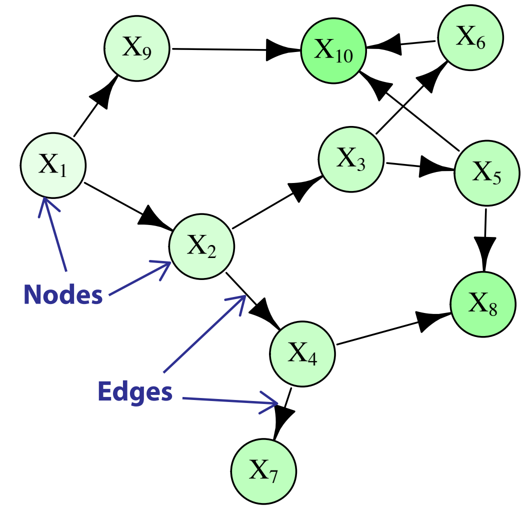
+ that if balanced appropriately, still results in boundedly rational belief formation

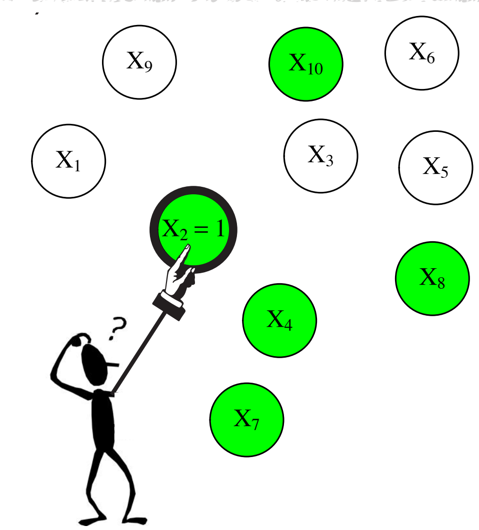




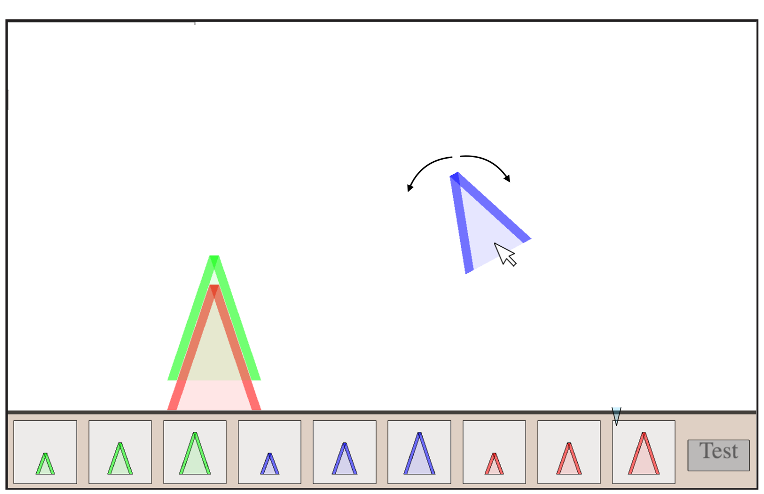
Various projects exploring bounded causal structure learning

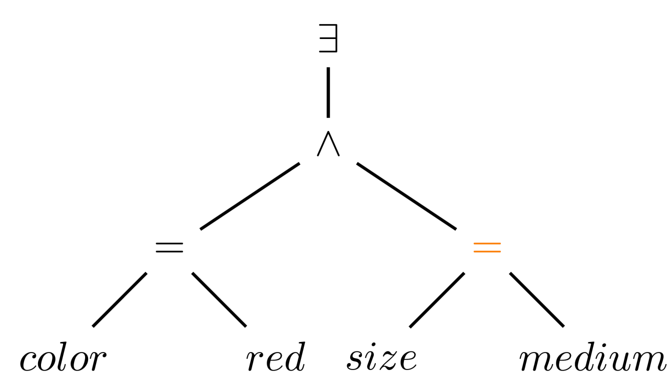
1. Contingencies: Active causal learning in probabilistic causal systems (CBNs)

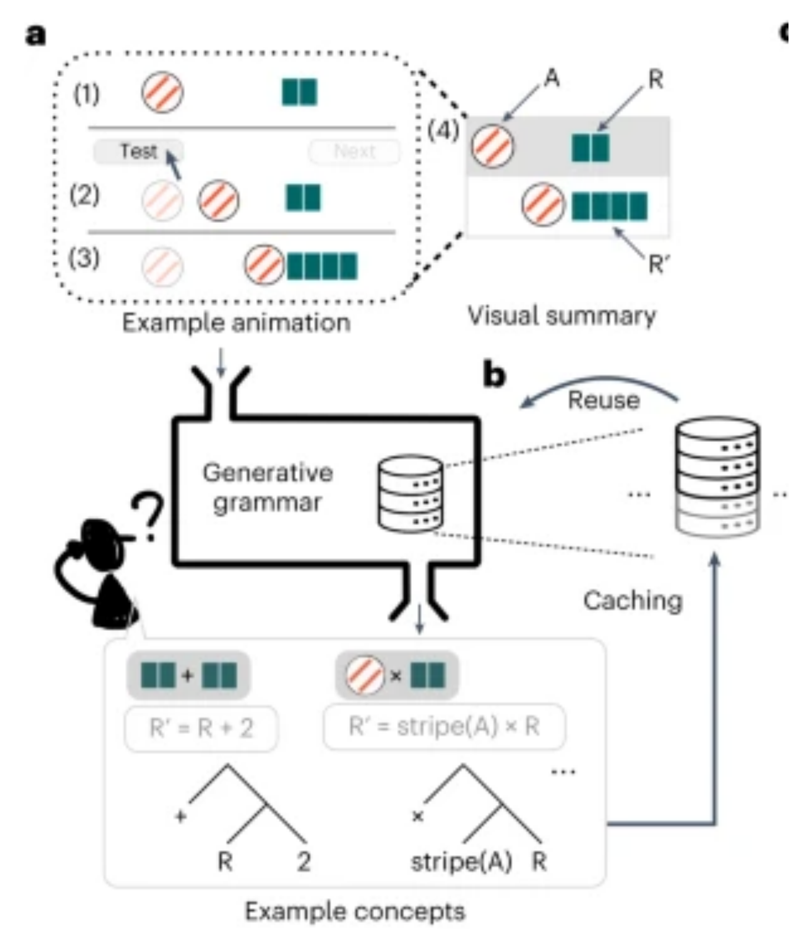




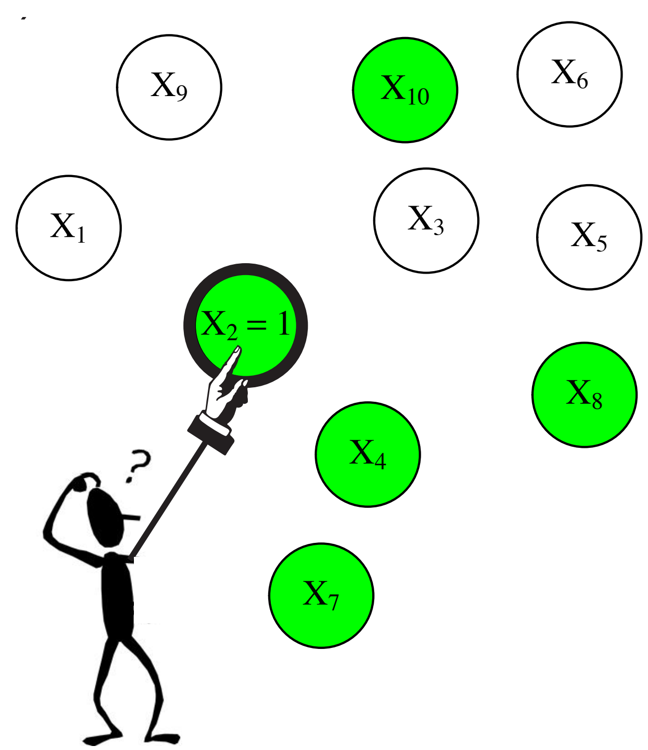
2. Compositional theories: Active inductive inference in open-ended contexts







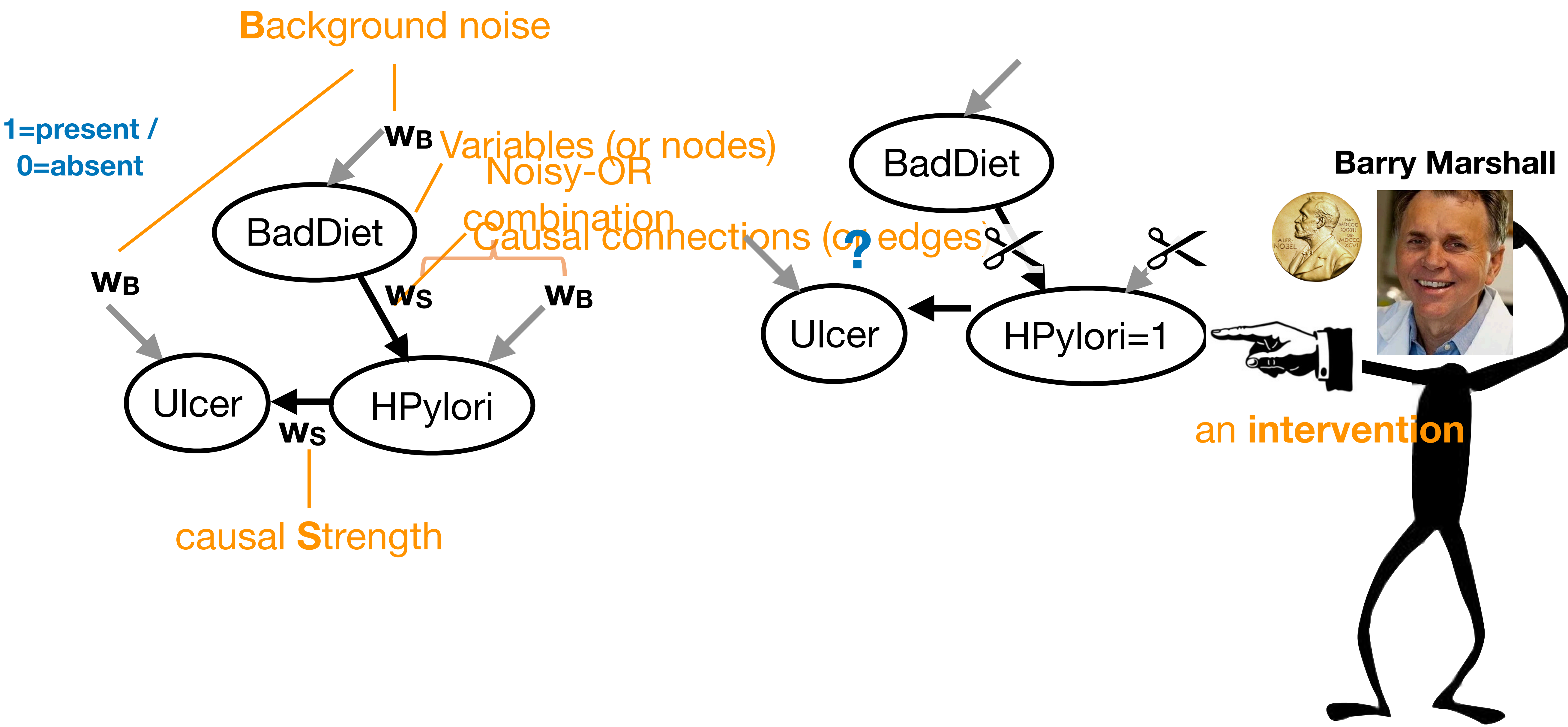
Today: Contingencies



Bramley, N. R., Dayan, P., Griffiths, T. L. & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*
Bramley, N. R., Lagnado, D. A. & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through interventions *JEP:LMC*
Coenen, A., **Bramley, N. R.,** Ruggeri, A. & Gureckis, T. M. (2019). Testing one or multiple: Beliefs about sparsity affect causal experimentation*. JEP:LMC*
McCormack, T., **Bramley, N. R.**, Frosch, C., Patrick, F. & Lagnado, D. A. (2016). Children's Use of Interventions to Learn Causal Structure. *JECP.*
Meng, Y., **Bramley, N. R.** & Xu, F. (2018). Children's causal interventions combine discrimination and confirmation.
Bramley, N. R., Dayan, P. & Lagnado, D. A. (2015). Staying afloat on Neurath's boat: Heuristics for sequential causal learning. *CogSci15*
Bramley, N. R. (2017). Constructing the world: Active causal learning in cognition. *PhD, UCL*

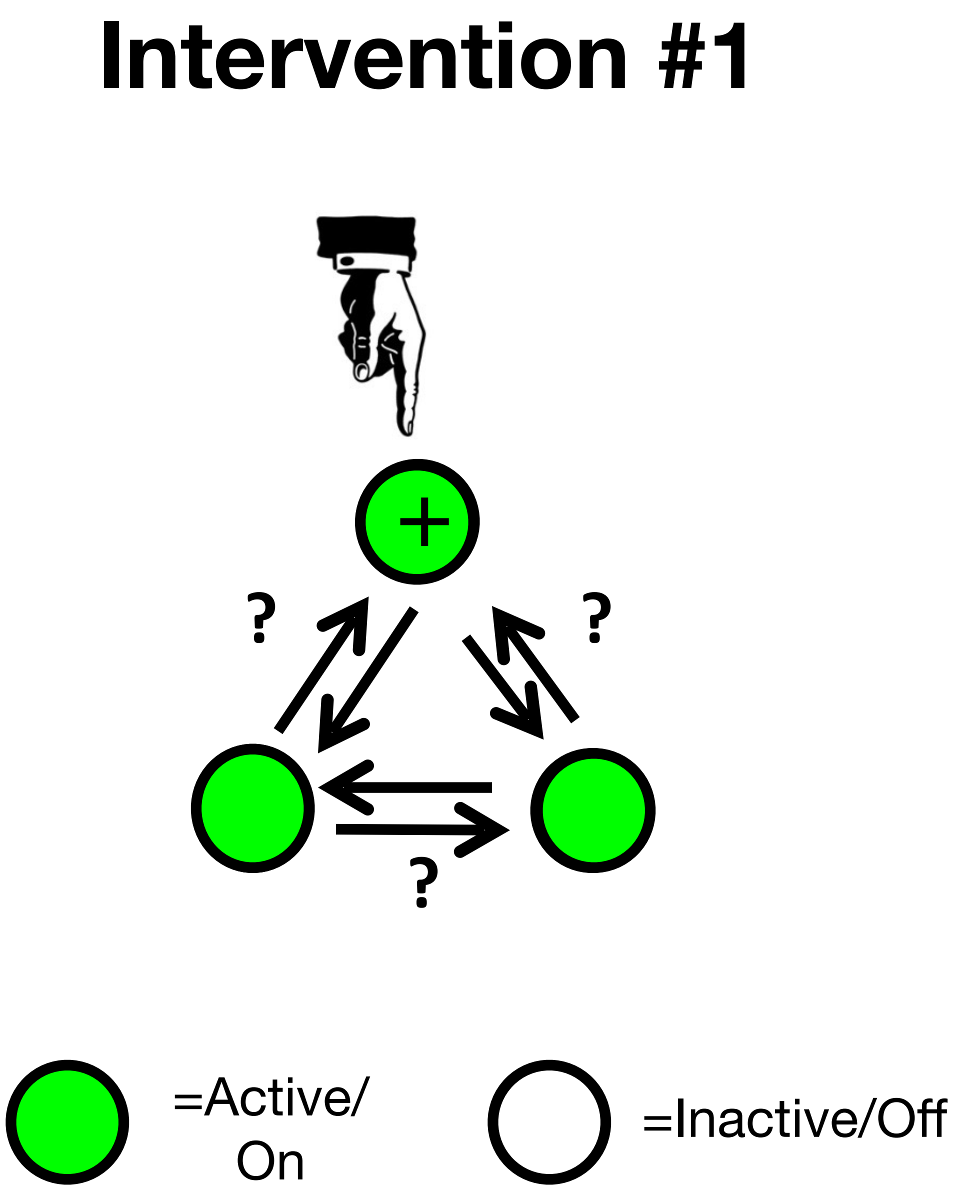
Causal Bayesian networks

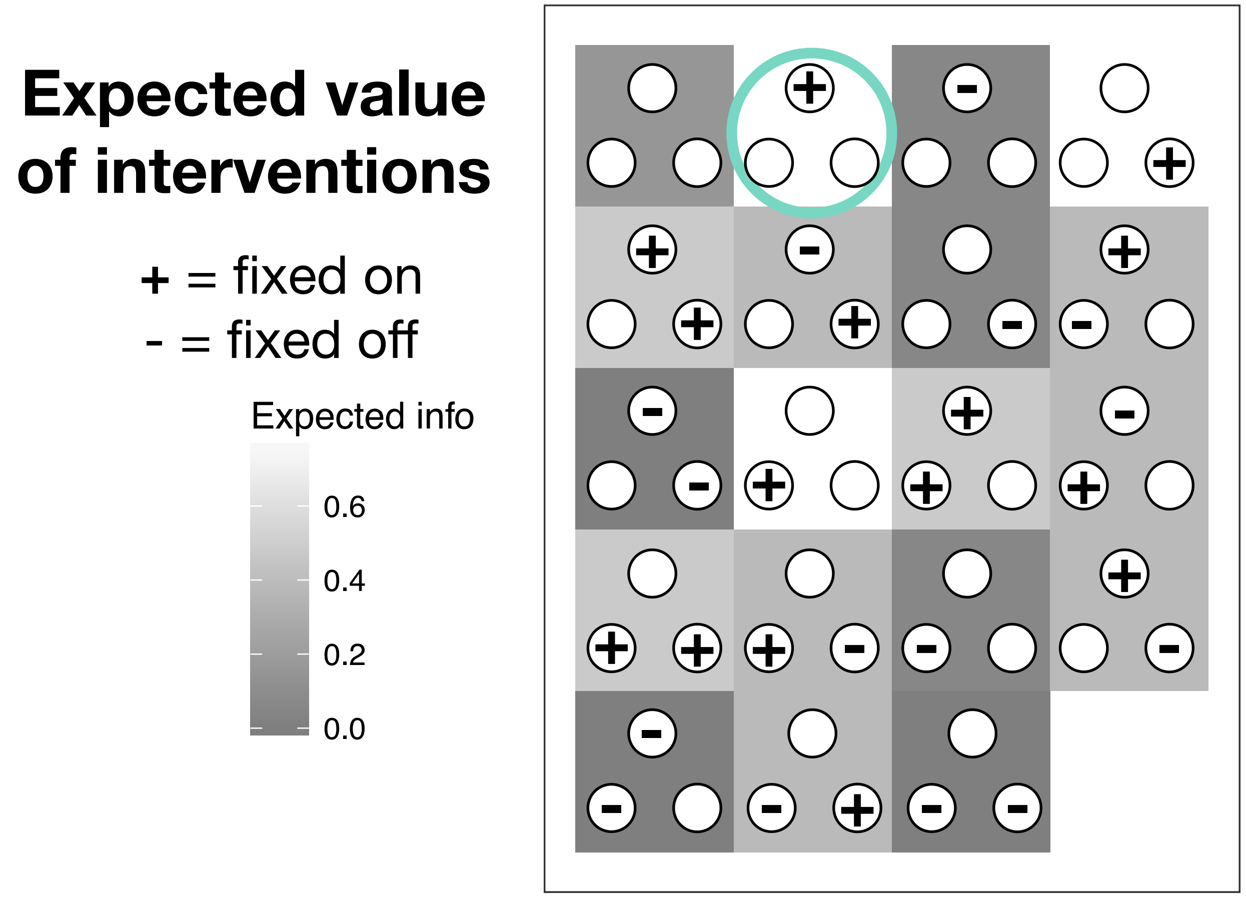
Cheng 1997; Coenen, Rehder & Gureckis, 2015; Griffiths & Tenenbaum, 2009; Gopnik et al 2004; Danks, 2014; Lagnado & Sloman, 2002; 2004; 2006; Pearl, 2000; Rehder, 2003; Steyvers et al, 2003; Waldmann & Holyoak 1992, Woodward, 2003

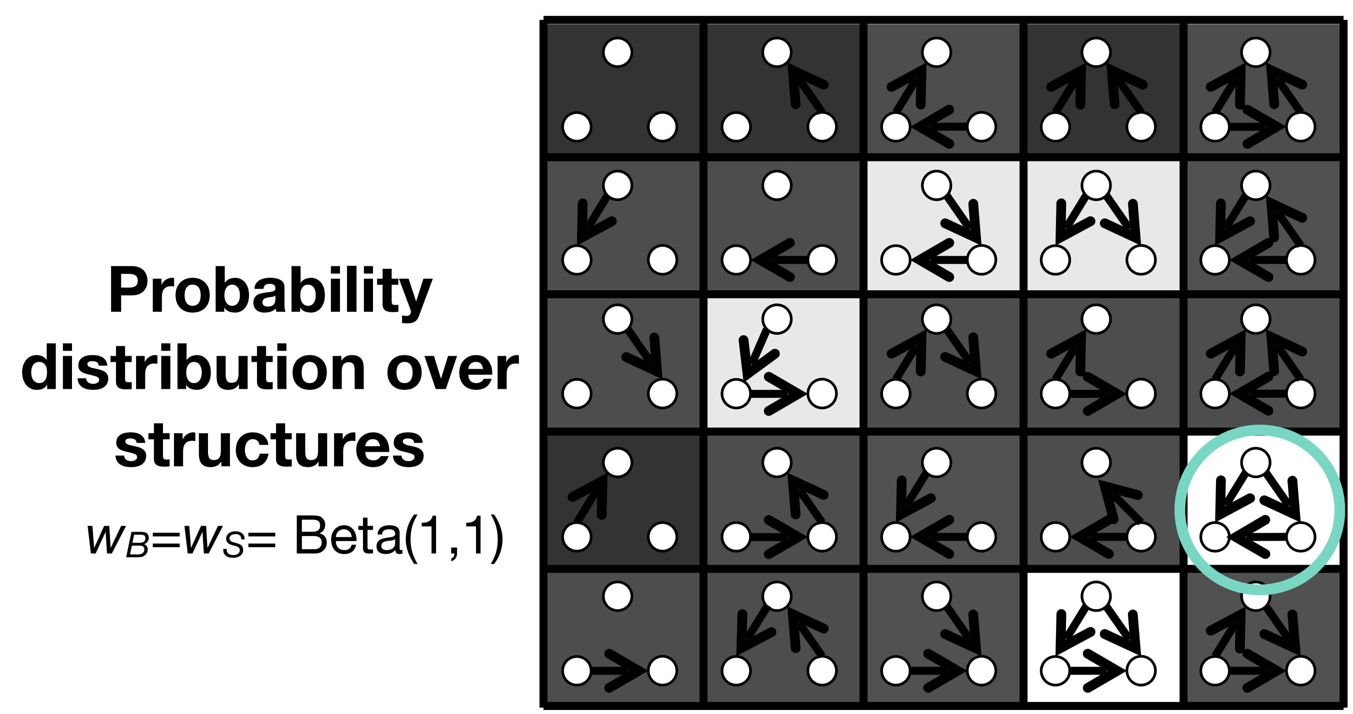


Bramley, N. R., Lagnado, D. A. & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through interventions *JEP:LMC*

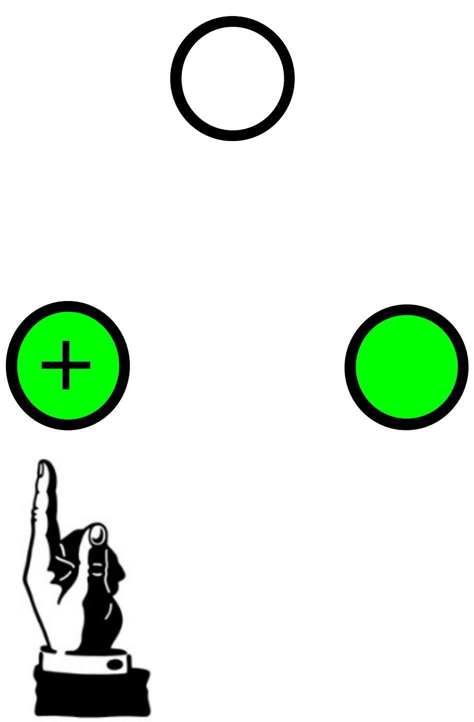
Idealised structure inference



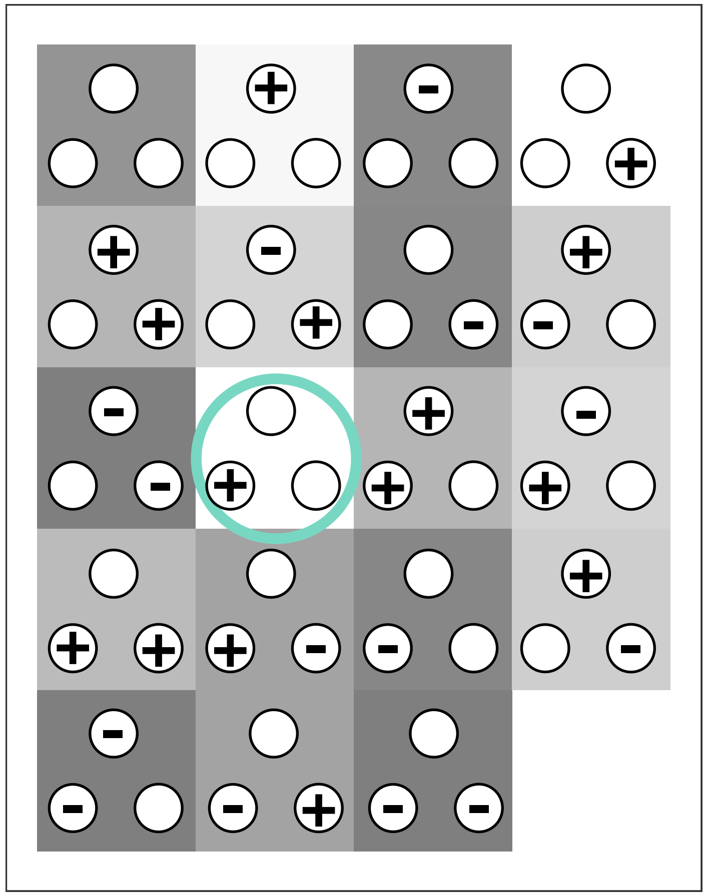




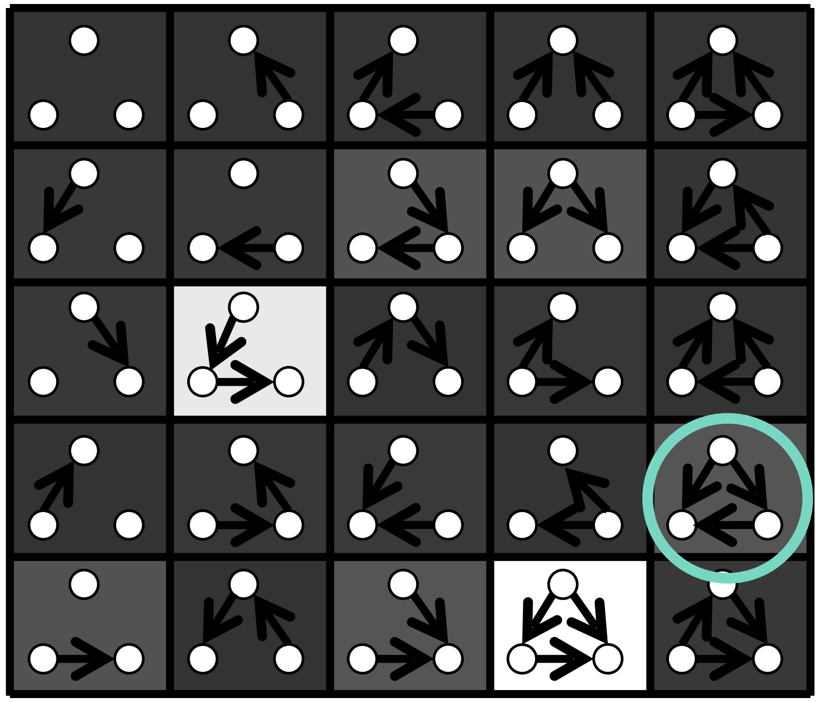
Intervention #2



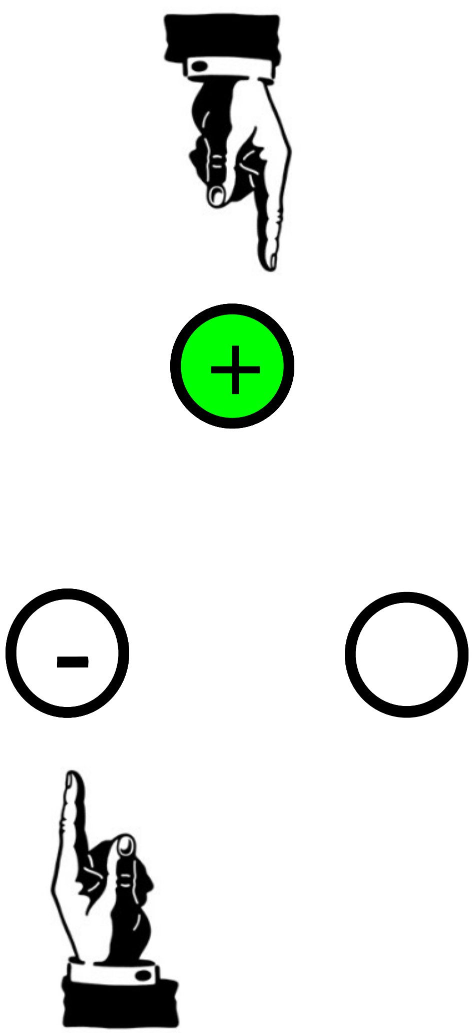
Expected value of interventions



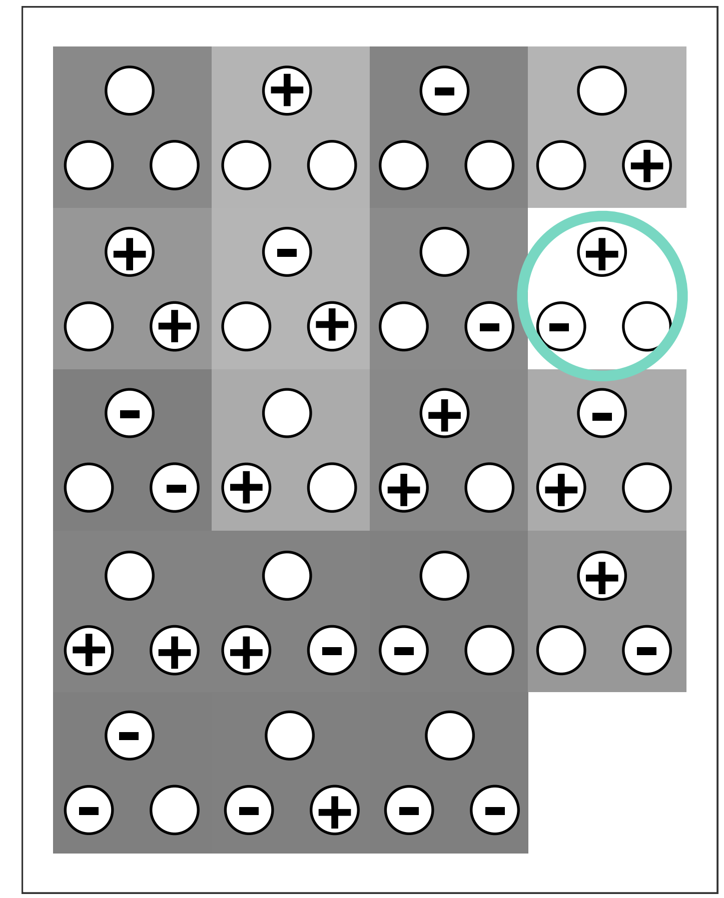
Probability distribution over structures



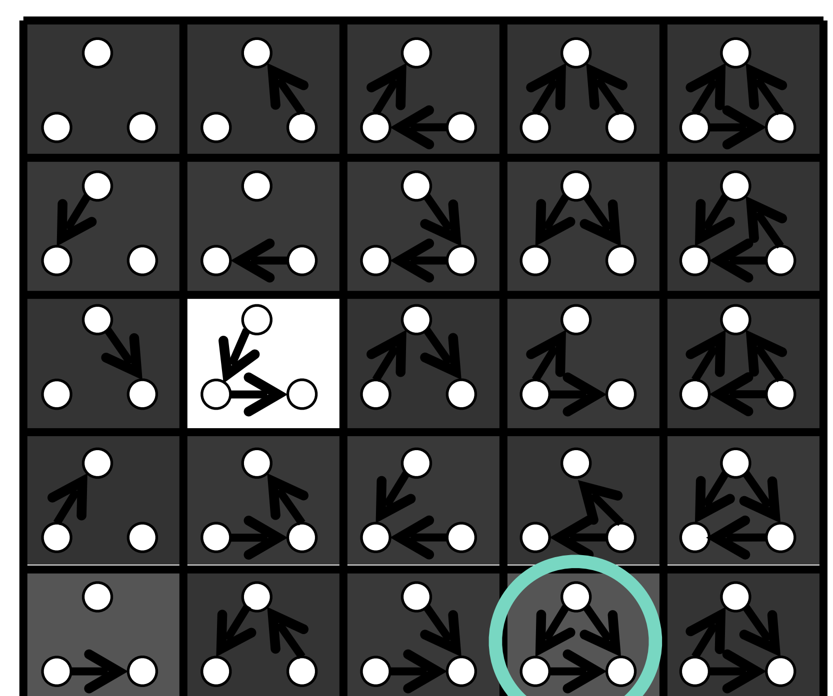
Intervention #3



Expected value of interventions



Probability distribution over structures

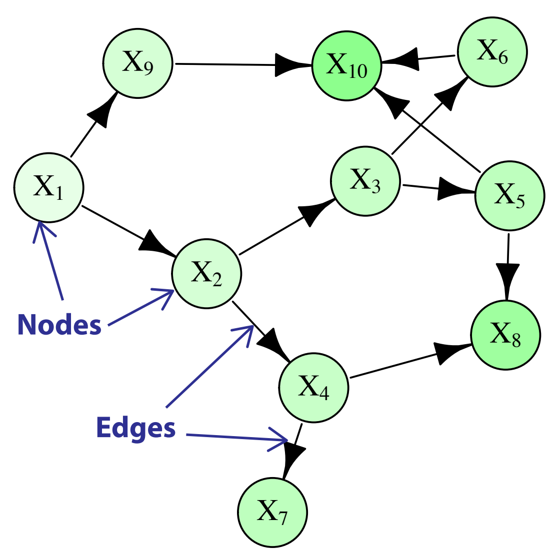


Hard because of the combinatorial explosion of possibilities…

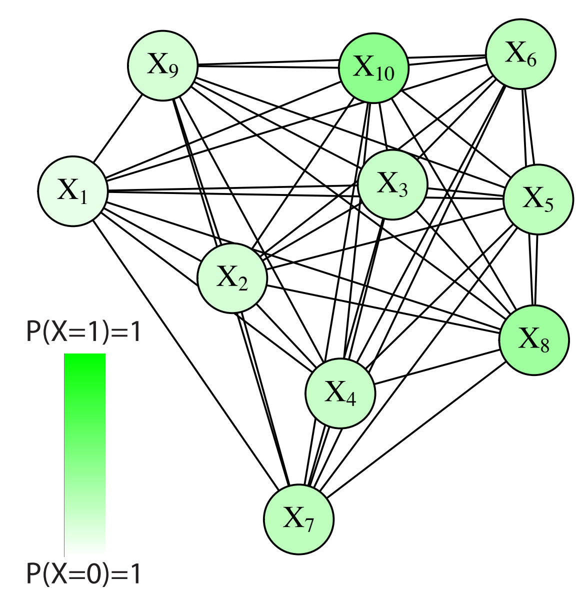
Variables	Structures	Interventions	Outcomes
1	1	3	1
2	3	9	2
3	25	27	4
4	543	81	8
5	29281	243	16
6	3781503	729	32
7	1138779265	2187	64
8	783702329343	6561	128
9	~1213442000000000	19683	256
10	~4175099000000000000	59049	512

…and long-range dependencies

True causal structure:



Marginal dependencies:

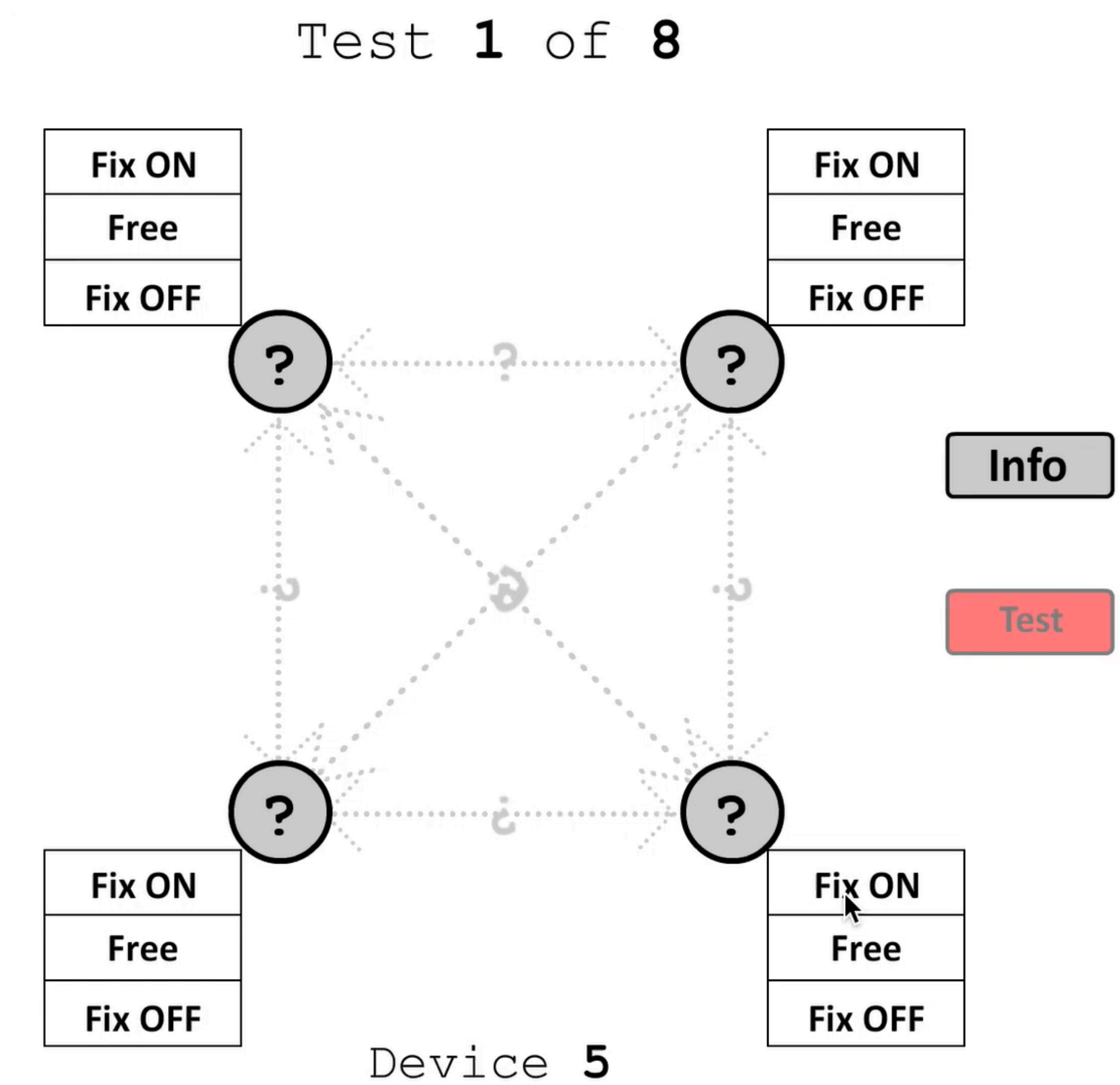


- Can people deal with this complexity?
- And if so, how do they do it?

Task

370 adults in **3** Exps

Interact repeatedly with causal systems governed by unknown CBN with goal of learning structure:



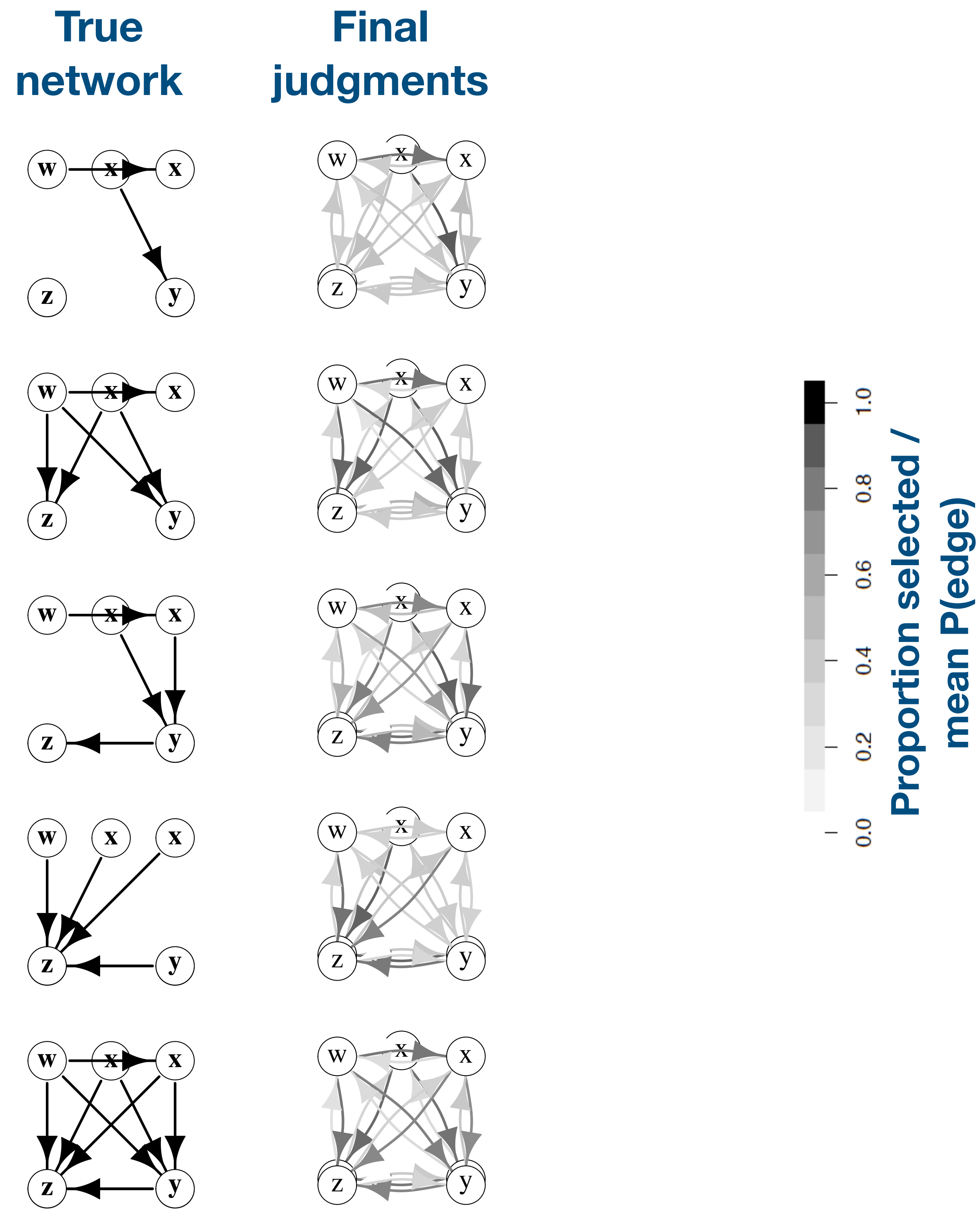
a) Perform interventions fixing variables
b) Observe resultant “activations” (green)
c) Guess the true network
d) Repeat… At end of trial, receive feedback

- Incentivized to be accurate
- Range of (un)known w_B and w_S

Bramley, N. R., Dayan, P., Griffiths, T. L. & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*

Behavioural patterns

- Aggregate judgment patterns similar to posterior probabilities
- Individuals “**probability match**” (Shanks et al, 2001) — i.e., sample from, but don’t maximise over, posterior?

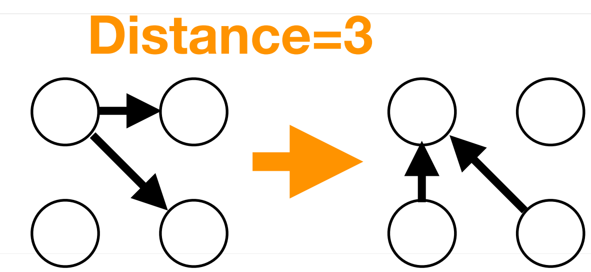


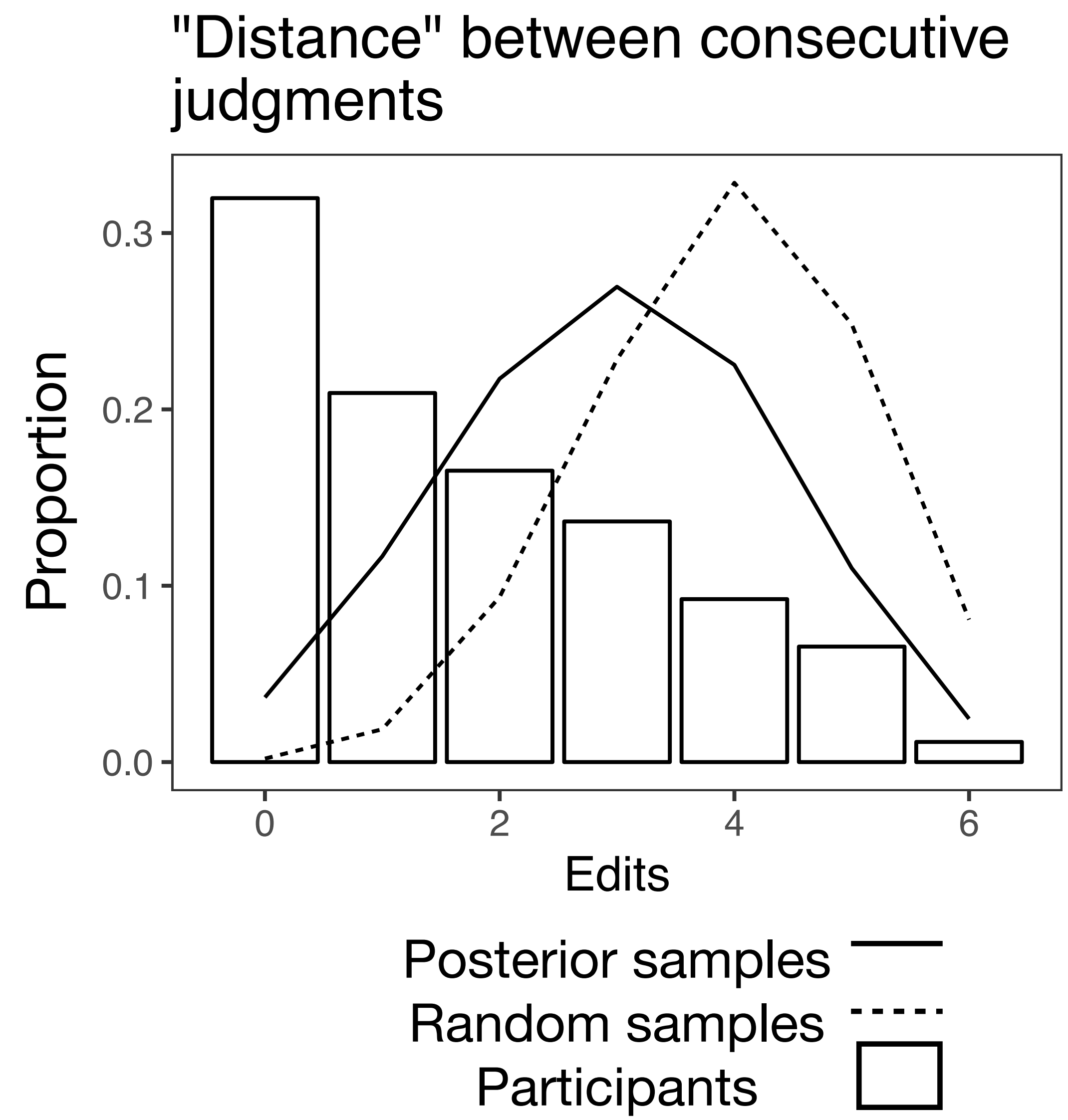
Bramley, N. R., Dayan, P., Griffiths, T. L. & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*

Behavioural patterns

- Aggregate judgment patterns similar to posterior probabilities
- Individuals probability match (Shanks et al, 2001) — i.e., sample from, but don't maximise over, posterior?
- **But** individuals' judgment sequences much too sequentially dependent
  - Too few edits between judgments (Bramley, Dayan & Lagnado, 2015b; Bramley, Dayan, Griffiths & Lagnado 2017)
  - Edits disproportionately reflect recent evidence (Bramley, Lagnado & Speekenbrink, 2015)

Types of "edit":
1. **Adding** a connection
2. **Removing** a connection
3. **Reversing** a connection





Bramley, N. R., Dayan, P., Griffiths, T. L. & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*

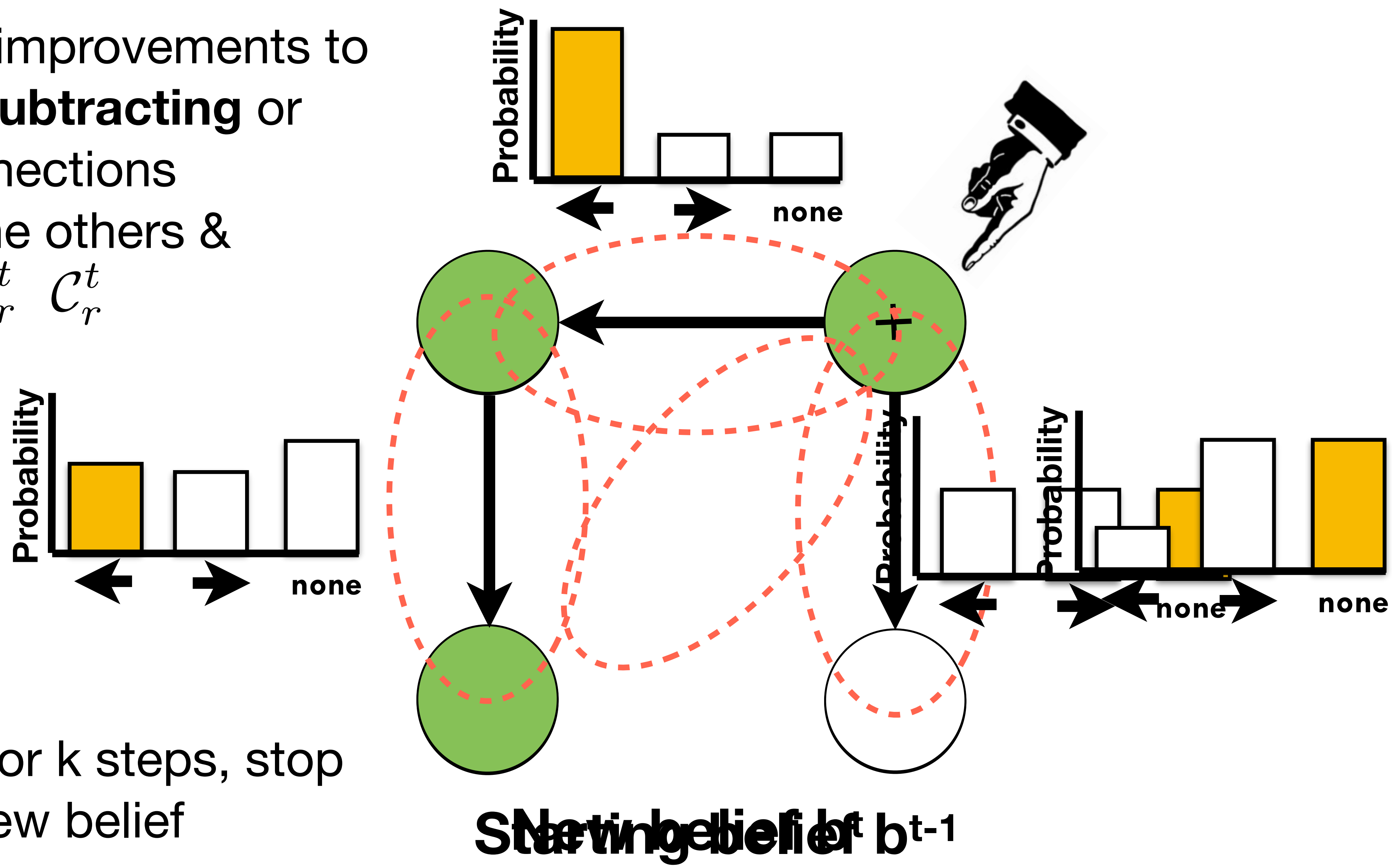
Goal

- **Process level** account of *group-level normativity* in face of *individual heterogeneity* and *sequential dependence* (cf. Courville & Daw 2007; Thaker, Tenenbaum & Gershman, 2017)

- Should show how people deal with *complexity* i.e., explain our *successes* in identifying pattern relating many variables

A model of local search

1. Search for local improvements to b^{t-1} by **adding, subtracting** or **reorienting** connections conditional on the others & "recent" data $\mathcal{D}_r^t$ $\mathcal{C}_r^t$



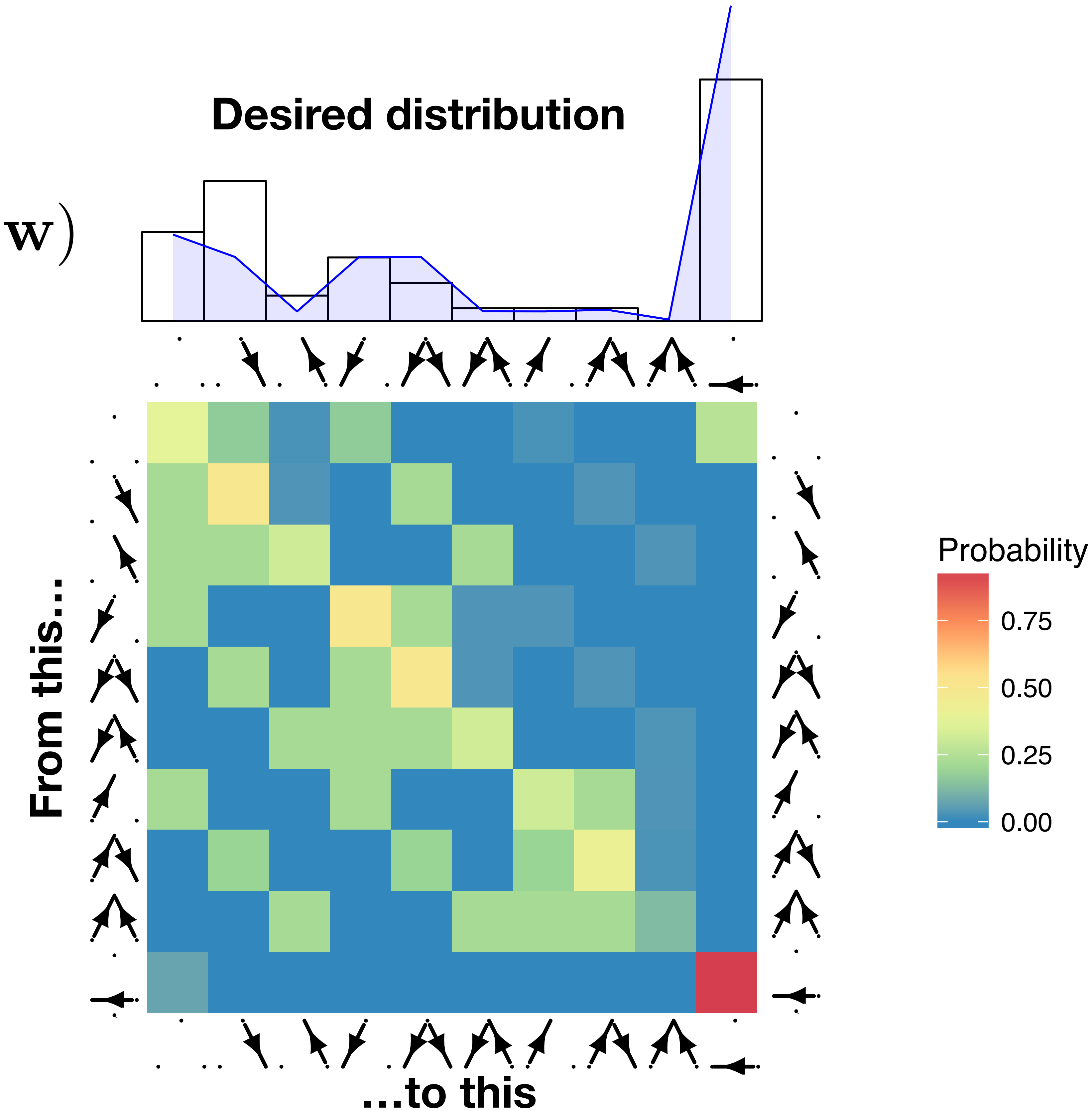
2. After searching for k steps, stop and take b^t as new belief

3. If b^t differs from b^{t-1} $\mathcal{D}_r^t = \mathcal{C}_r^t = \{\}$

Bramley, N. R., Dayan, P., Griffiths, T. L. & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*

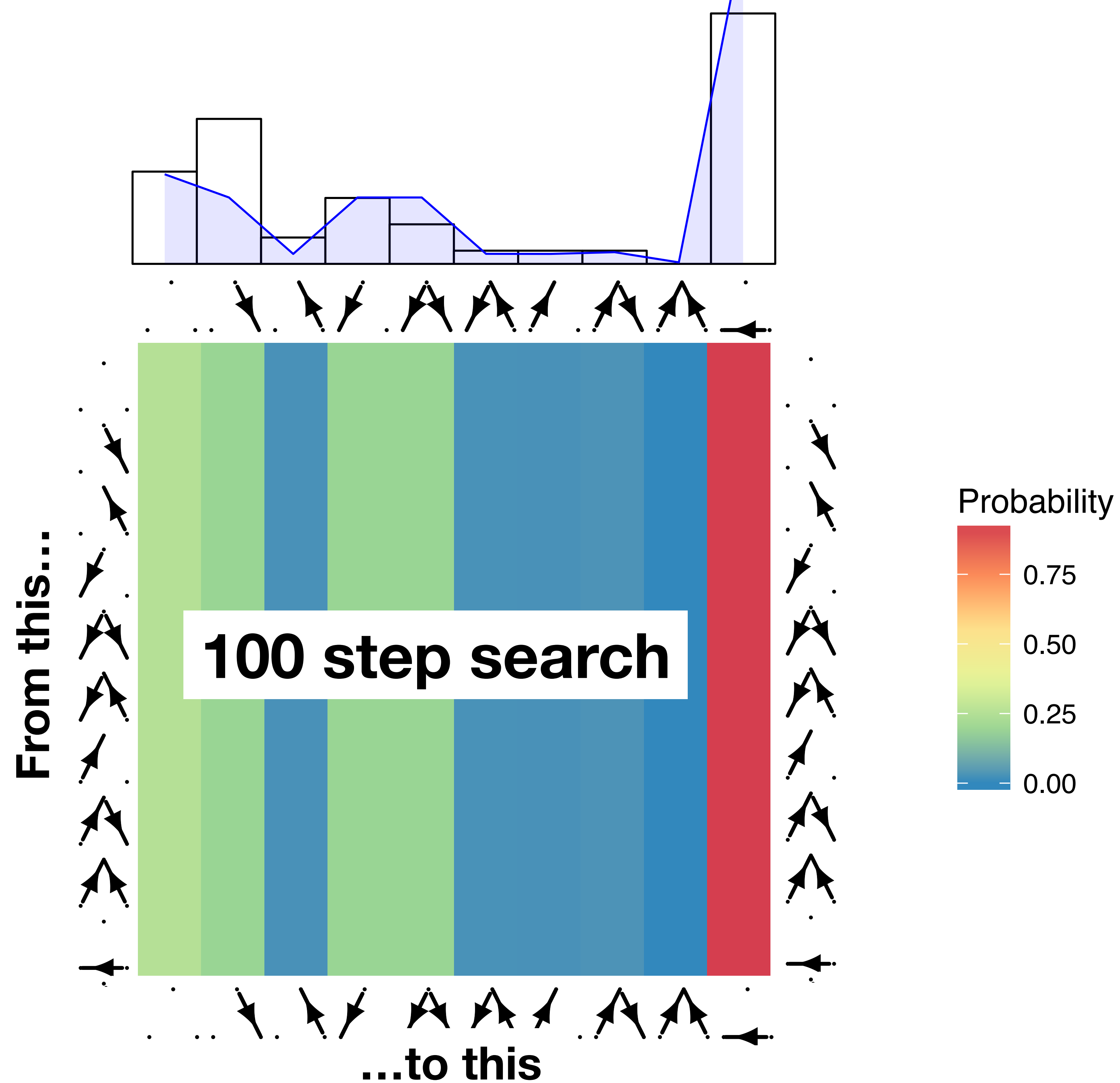
Asymptotic behaviour

$$P(M|\mathcal{D}_r^t;\mathcal{C}_r^t,\mathbf{w})$$



Aggregating over searches

$$P(M|\mathcal{D}_r^t;\mathcal{C}_r^t,\mathbf{w})$$



Connection to approximate Bayesian inference:

- Gibbs sampling to update
- Short search chain retains influence of starting point, i.e. a "single-particle particle-filter" (Vul et al, 2009; Sanborn et al, 2010; Courville & Daw, 2007)

A generalissed model of incremental belief adaption

Probability that local search terminates at model m:

Search length parameter *λ*
controls average over
search lengths

Matrix of Transitions

$$P(b^t = m | \underbrace{\mathcal{D}_r^t, \mathcal{C}_r^t,}\, b^{t-1}, \omega, \lambda) = \sum_0^\infty \overbrace{\frac{\lambda^k e^{-\lambda}}{k!}} [\underbrace{(R_t^\omega)}^k]_{b^{t-1}m}$$

Search based on "recent" evidence,
& interventions since last belief change

Behavior parameter: ω
interpolates between
0 = random
1 = Gibbs sampling
∞ = hill climbing

Predicts:

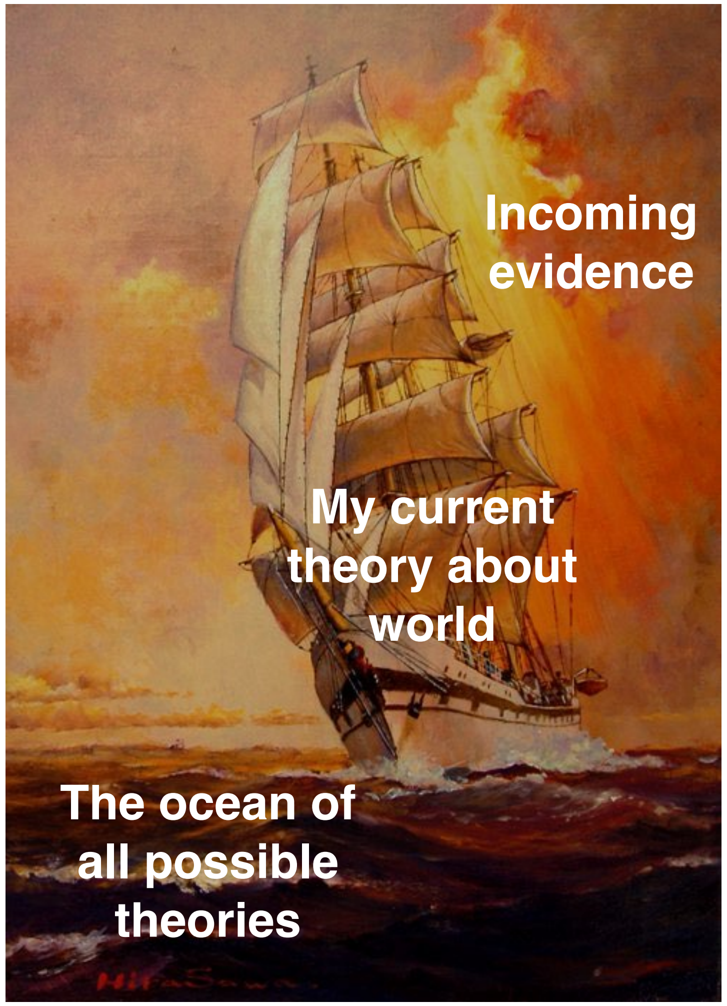
- Gradational sequential dependence between judgments due to limited search ***λ***
- Noisily normative aggregate behavior (depending on **ω**)

Bramley, N. R., Dayan, P., Griffiths, T. L. & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*

Neurath's ship

"We [theorists] are like sailors who on the open sea must reconstruct their ship but are never able to start afresh from the bottom. Where a beam is taken away a new one must at once be put there, and for this the rest of the ship is used as support."

– *W.V.O. Quine, (1960)*

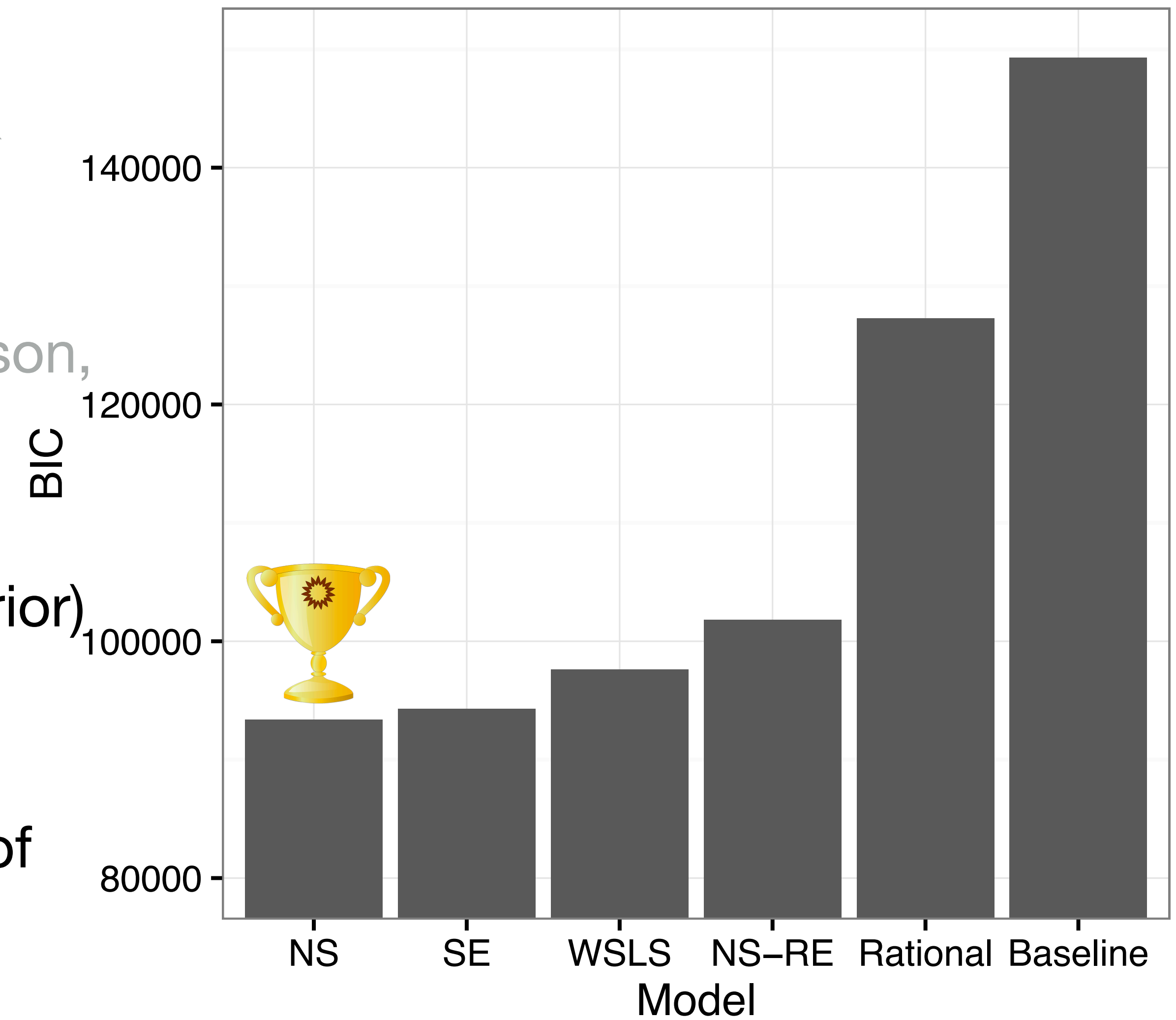


Bramley, N. R., Dayan, P., Griffiths, T. L. & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*

Judgment models

Compared Neurath's Ship (**NS**) against range of competitors:

1. Simple endorsement heuristic (**SE**) (Fernbach & Sloman, 2009; Waldmann, Cheng, Hagmayer & Blaisdell, 2008)
2. Win-stay, lose-sample (**WSLS**) (Bonawitz, Denison, Gopnik & Griffiths, 2014)
3. Random edits (**NS-RE)** (NS with ω fixed to 0)
4. Noisily **Rational** judgment (softmax over posterior)
5. Random judgments **Baseline**



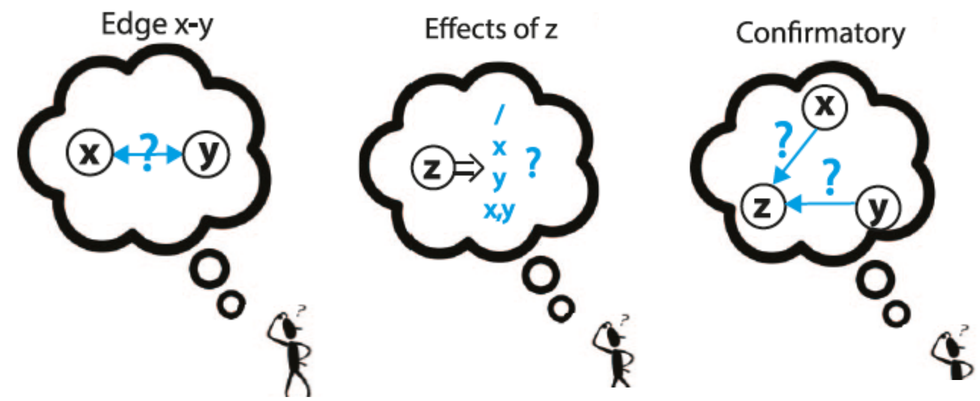
- **NS** lowest overall BIC & best fit largest number of individual participants 155/370
- Search lengths $\boldsymbol{\lambda}\approx 1.5$ & search behavior $\boldsymbol{\omega}\approx 6$ (moderate hill climbing)

Bramley, N. R., Dayan, P., Griffiths, T. L. & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*

Discussion

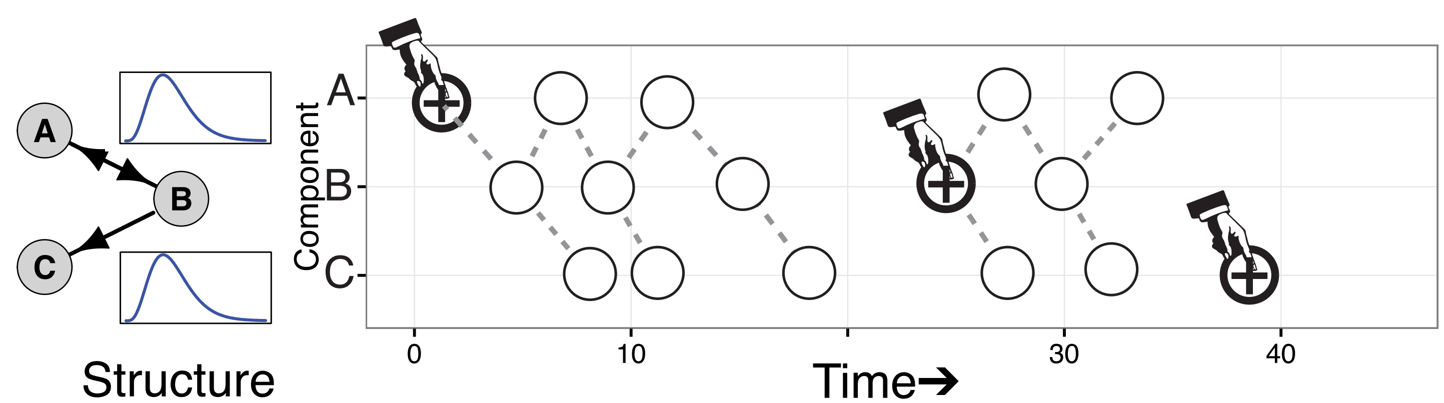
- **N**eurath's **S**hip model — captures incremental causal belief change

- **Locally focused interventions?**



Yes see Bramley et al (2017); Gong & Bramley (2023), but a story for another day…

- **Temporal information?**



Yes. See Davis, Bramley & Rehder (2019, 2018a, 2018b); Bramley, Gerstenberg↔Mayrhofer & Lagnado (2018; 2017; 2014) also a story for another day…

Discussion

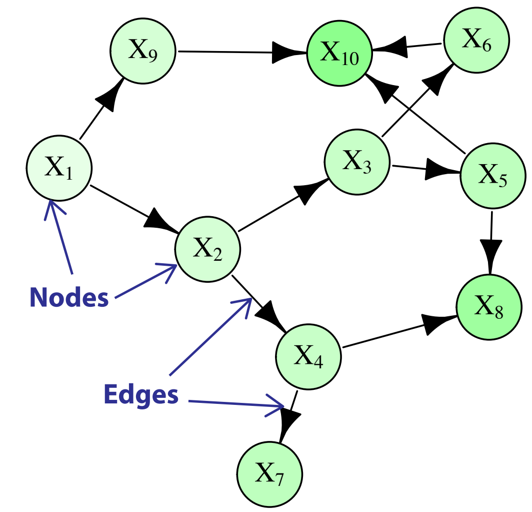
- While structure induction requires approximation…
- We still assumed a fixed, finite, hypothesis space
- But real settings are typically more open ended
- As is human thought, exhibiting **compositionally**, **systematicity,** *"infinite use of finite means"*

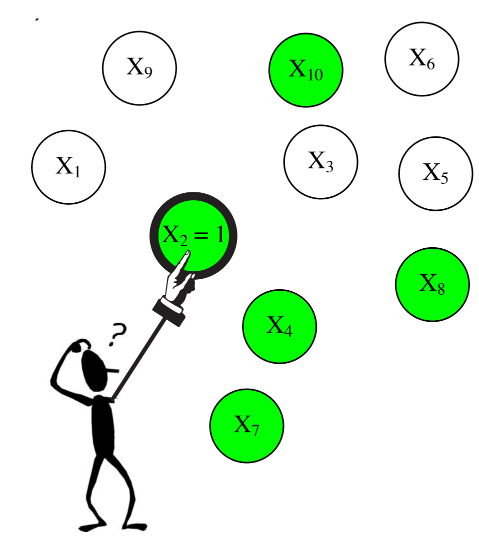
Variables	Structures
1	1
2	3
3	25
4	543
5	29281
6	3781503
7	1138779265
8	783702329343
9	∼1213442000000000
10	∼4175099000000000000

- Can we make sense of human learning in open-ended setting where hypothesis space is technically infinite?

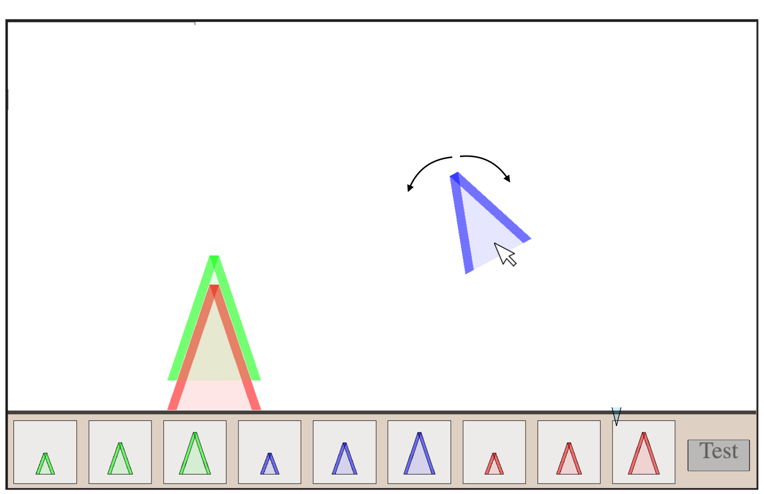
Projects exploring bounded causal structure learning

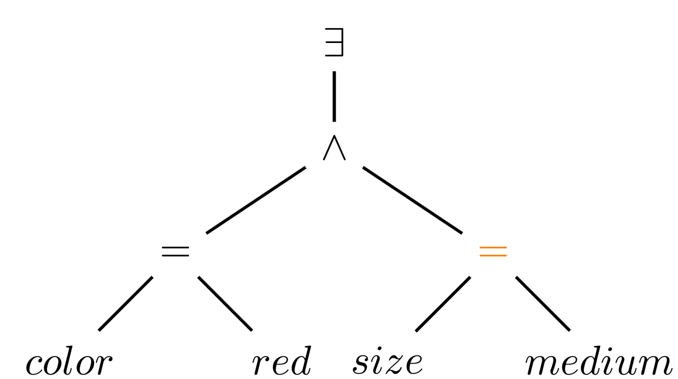
1. Contingencies: Active learning in probabilistic causal systems (CBNs)

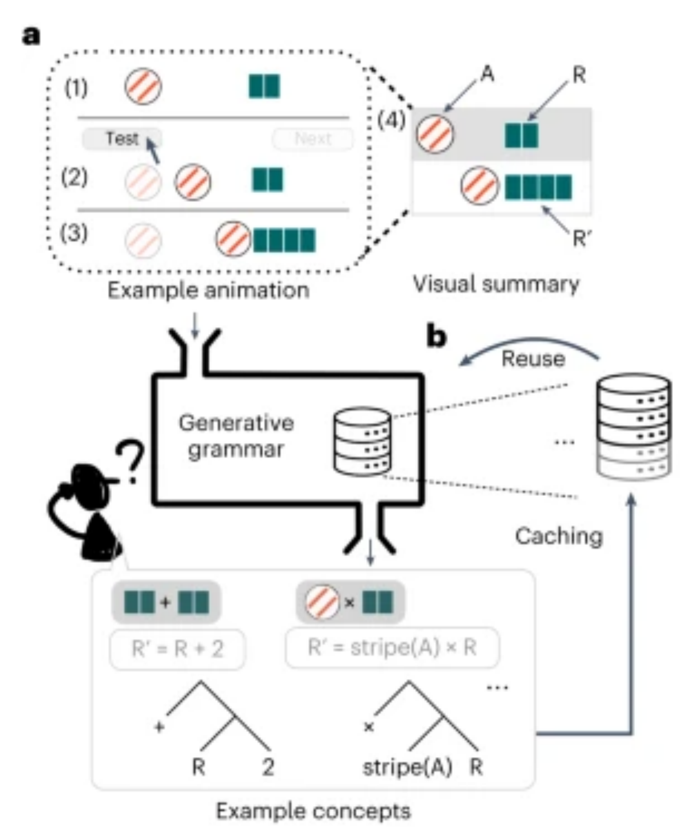




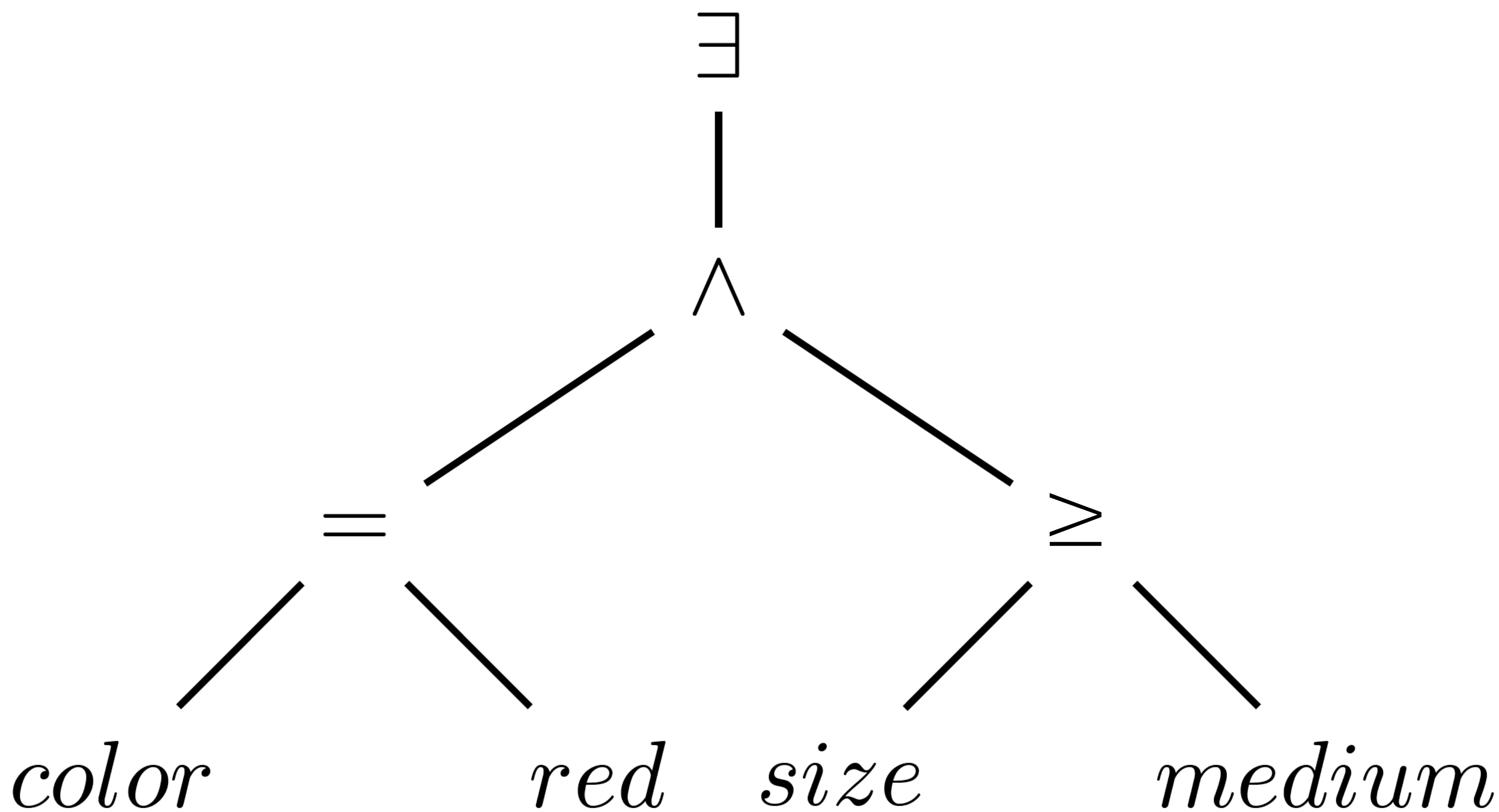
Compositional theories: Active inductive inference in open ended contexts







Taster, if time: Compositional theories



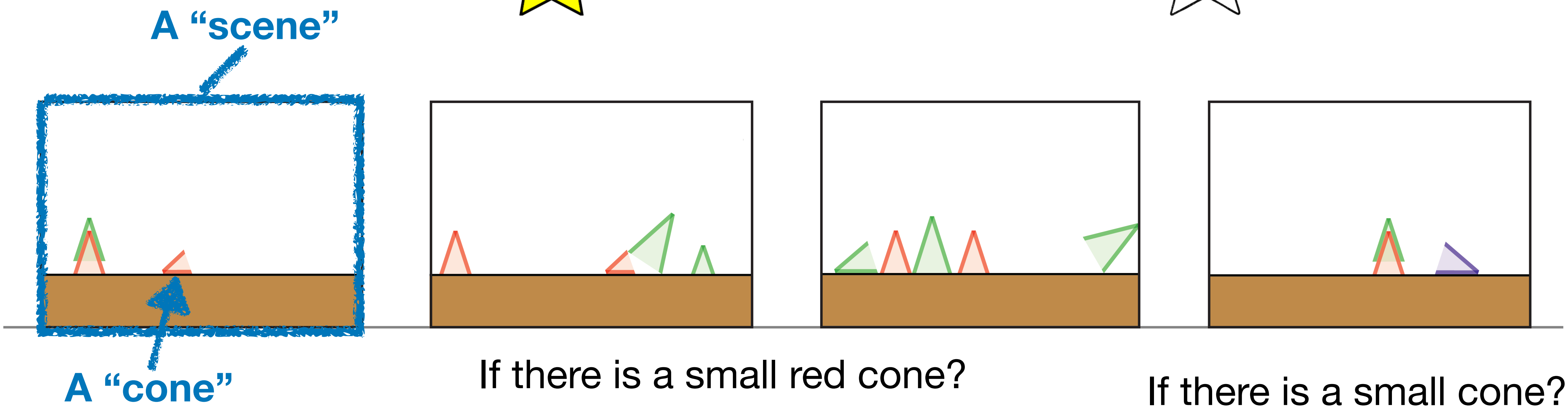
EPSRC New Investigator Grant (EP/T033967/1). Computational constructivism: The algorithmic basis of discovery

Zhao, B., Lucas, C. G., & Bramley, N. R. (2024). A model of conceptual bootstrapping in human cognition. *Nature Human Behaviour.*
Zhao, B., Lucas, C. G., & Bramley, N. R. (2022). How do people generalize causal relations over objects? A non-parametric Bayesian account. *Computational Brain & Behaviour*
Bramley, N. R., Zhao, B., Quillien, T., & Lucas, C. G. (2023). Local search and the evolution of world models. *Topics in Cognitive Science*.
Bramley, N. R., & Xu, F. (2023). Active inductive inference in children and adults: A constructivist perspective. *Cognition*.
Fränken, J. P., Lucas, C. G., Bramley, N. R., & Piantadosi, S. T. (2023). Modeling infant object perception as program induction. *CCN*
Fränken, J. P., Theodoropoulos, N. C., & Bramley, N. R. (2022). Algorithms of adaptation in inductive inference. *Cognitive Psychology*
Bramley, N. R., Rothe, A., Tenenbaum, J. B., Xu, F. & Gureckis, T. M. (2018). Grounding compositional hypothesis generation. *CogSci.*

A micro scientific induction task

Loosely based on Zendo™

Produces effect: ☆ **Does not produce effect:** ☆



If there is a small red cone?

If there is a small cone?

If a red cone is pointing left?

A red cone is the closest to the bottom left corner?

If there is a no blue cone and fewer than three green cones?

- Our guesses *seem* **symbolic & compositional**, *seem* to have language-like **productivity**
- Need mechanism for *generating*, *adapting, investigating* symbolic hypotheses…

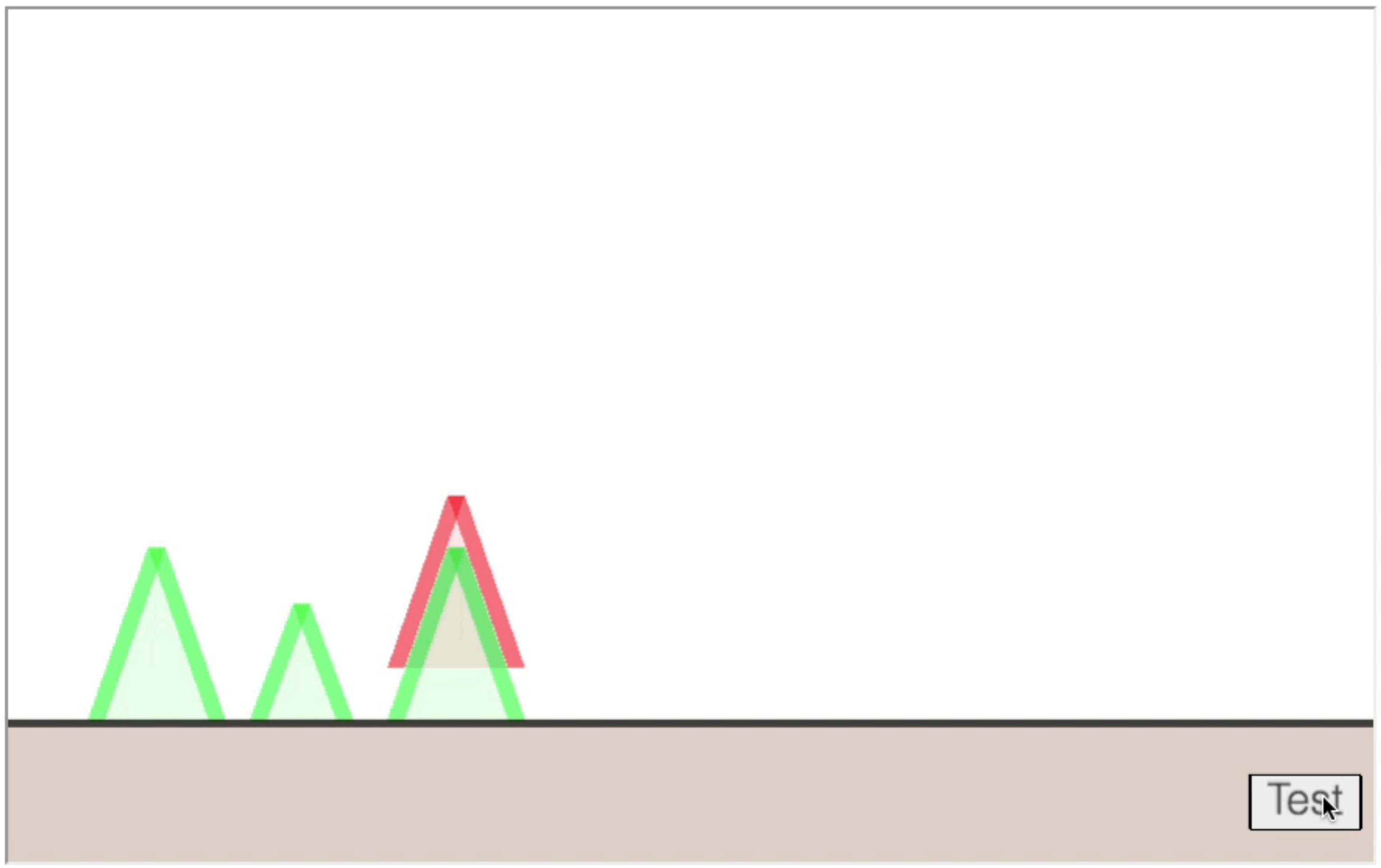
Compositional theories

- **Constructivism** (Carey, 2009; Piaget 1970; Piantadosi, 2017) influential idea in developmental psychology
  - In development/learning, we actively “construct” new ideas through **thinking**— recombining & modifying ideas…
  - …and **play**—exploring and discovering patterns

- **How could this work algorithmically? Is this an apt account of human cognition /development?**

Experiment

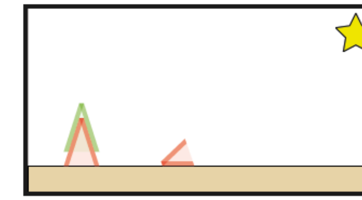
- **54 Children (8.9±1.1) & 50 Adult mTurkers**



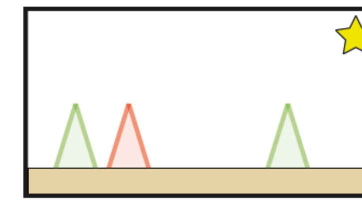
5 tasks:

Initial example

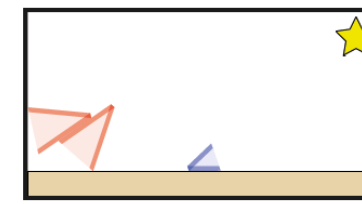
1. There is a red



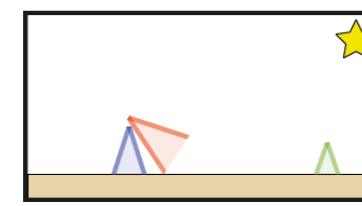
2. All same size



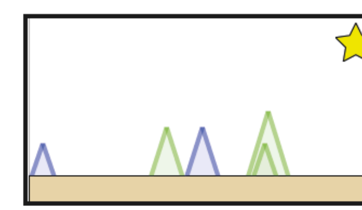
3. Nothing upright



4. Exactly 1 blue

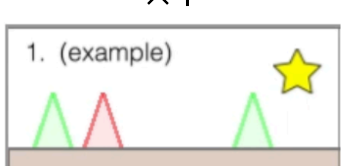


5. Something blue & small

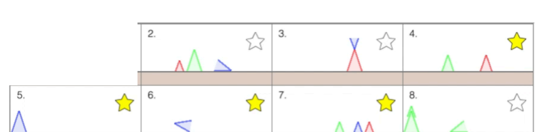


Procedure:

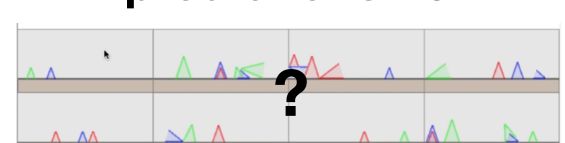
Positive example ×1



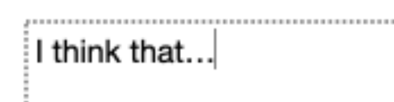
Active learning ×7



Generalisation predictions ×8



Free guess ×1



Example of kids protocol


Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

ELSEVIER

COGNITION

Original Articles

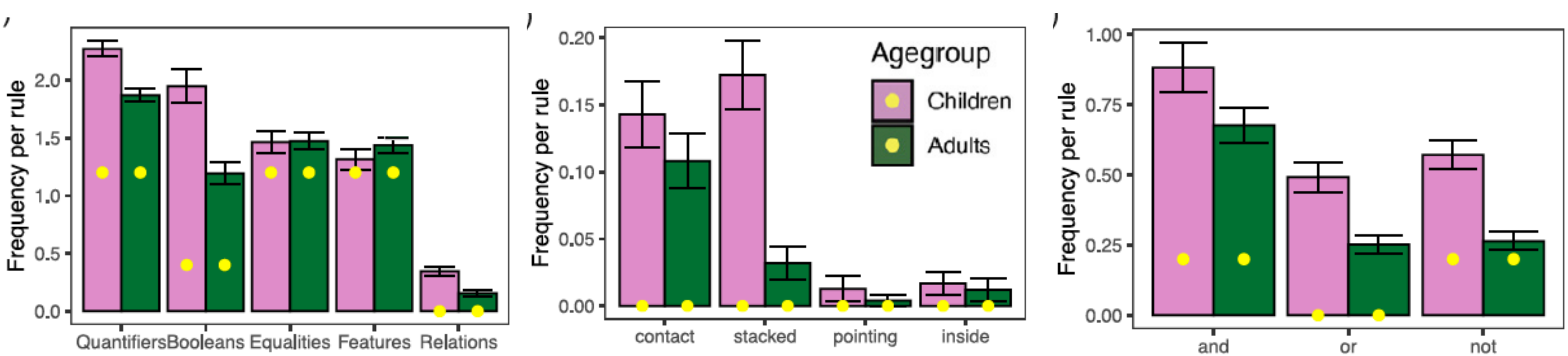
Active inductive inference in children and adults: A constructivist perspective

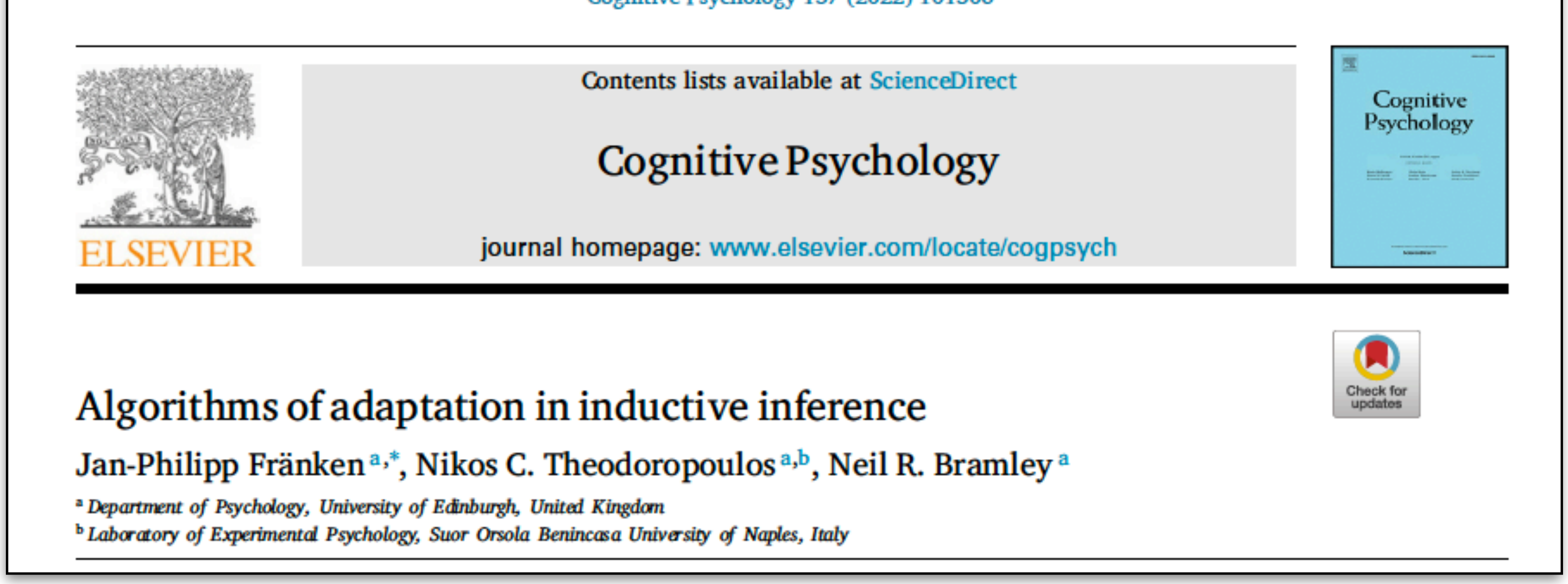
Neil R. Bramley[a,1,*], Fei Xu[b]

[a] Department of Psychology, University of Edinburgh, Scotland, United Kingdom
[b] Psychology Department, University of California, Berkeley, USA

Check for updates

- Counterintuitively: Children's hypotheses ***more diverse & elaborate*** than adults' —consistent with 'flatter', less tuned, generation mechanism (cf. Lucas, Griffiths, Bridgers & Gopnik, 2014) or perhaps also "hotter" search (Gopnik, 2020)
    - Mimicking developmental trajectory from flexibility/diversity to predictability/performance key to synthesising humanlike learning




Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Algorithms of adaptation in inductive inference

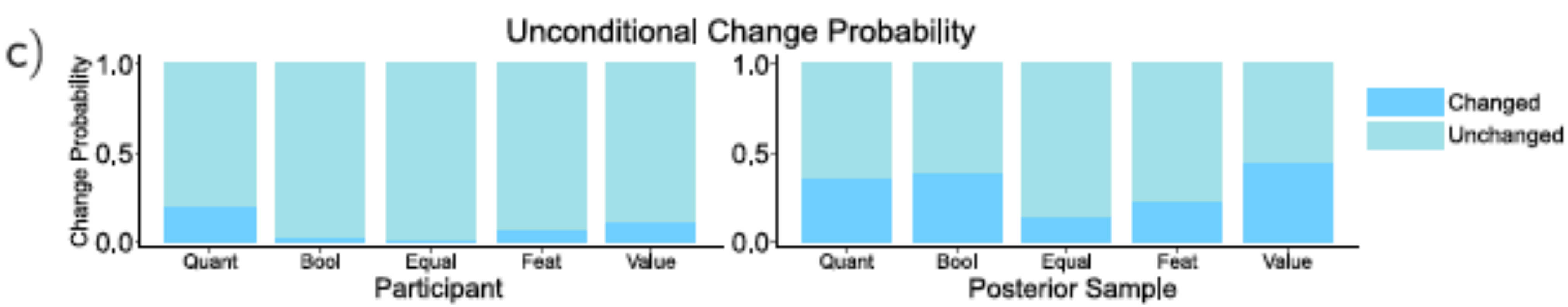
Jan-Philipp Fränken [a,*], Nikos C. Theodoropoulos [a,b], Neil R. Bramley [a]

[a] Department of Psychology, University of Edinburgh, United Kingdom

[b] Laboratory of Experimental Psychology, Suor Orsola Benincasa University of Naples, Italy

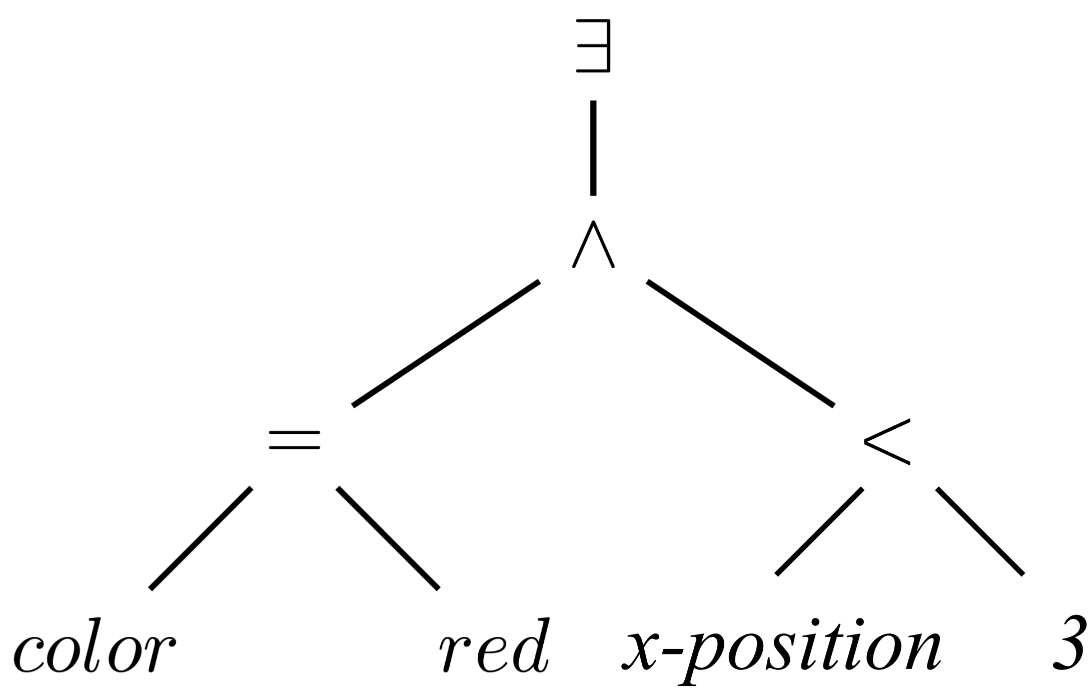
- Participants do Zendo task in two phases; we examine relationship between initial & revised guesses

- We find *syntax-level* anchoring on earlier guesses, (similar to structural anchoring in Neurath's ship paper)



Tree regrowth MCMC

- Established approach for program search/model estimation (Goodman et al, 2008; Piantadosi et al, 2016)

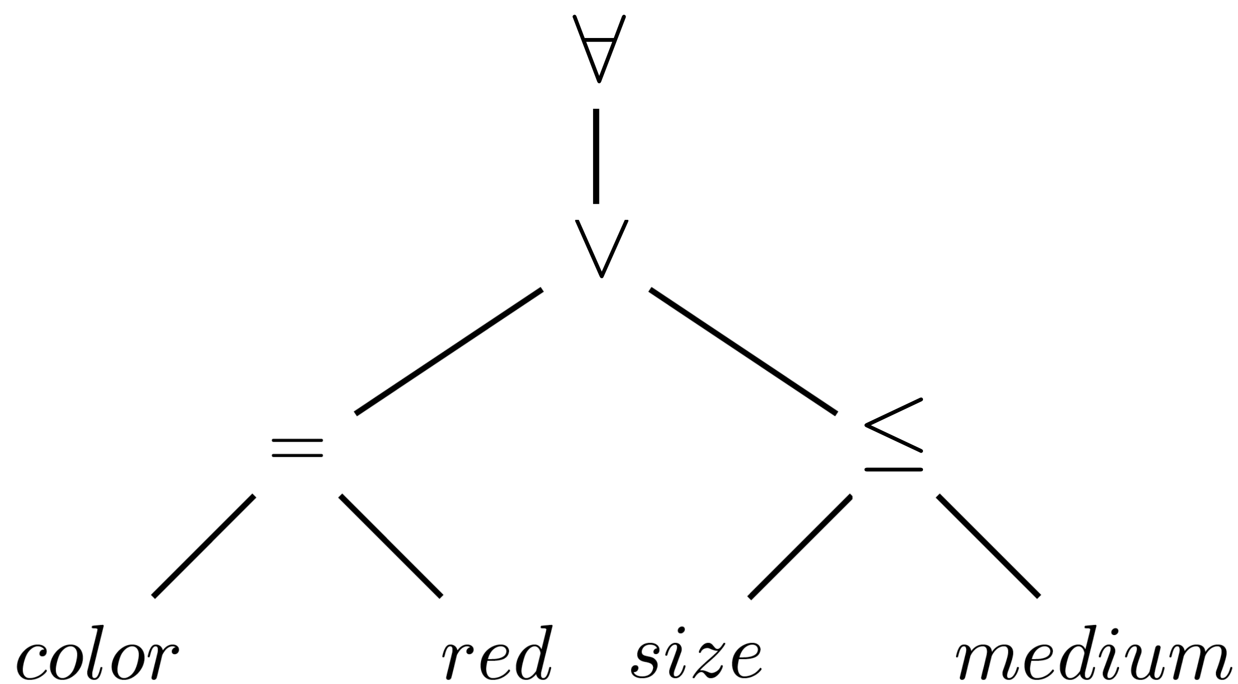


- Each new proposal deletes a random branch + regrows using *probabilistic context free grammar*
- Accepted according to function combining prior (complexity) and likelihood (fit)
- Many proposals constitute major edits, i.e. when regrown from near root

Tree “surgery” MCMC

- Novel MCMC proposal distribution (Fränken et al, 2022, *CogPsych*)

Step 3.



- Proposals limited to minimal “local” edits preserving downstream tree (unless they induce type-signature changes)
- Better model of participants’ revised generalisations and guesses
- Produces stronger sequential dependence + arguably more cognitively plausible model of symbolic search?


nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01719-1

A model of conceptual bootstrapping in human cognition

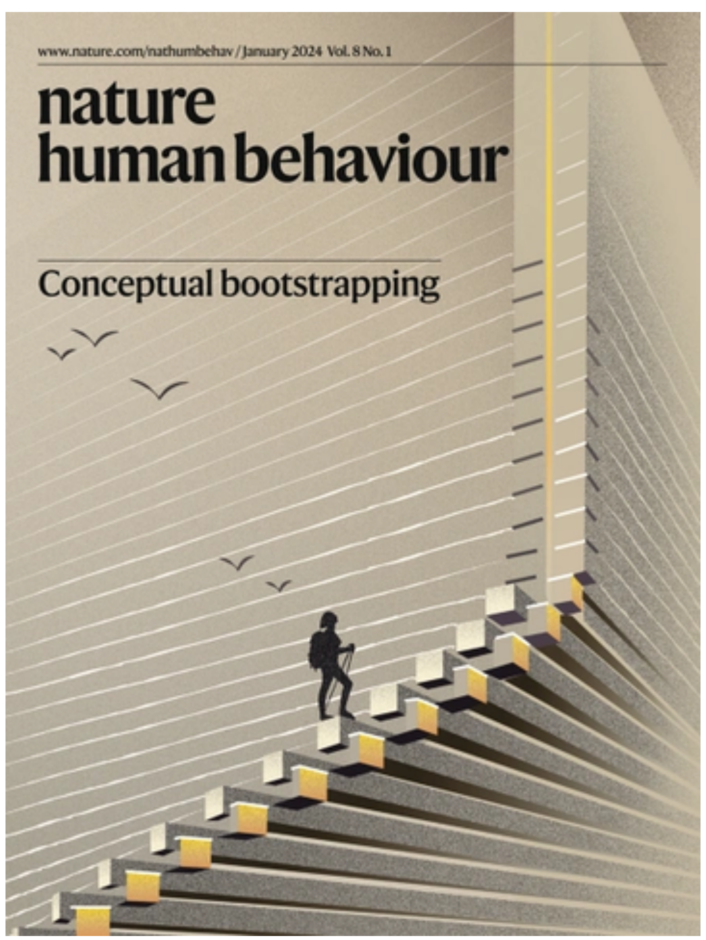
Received: 24 January 2023

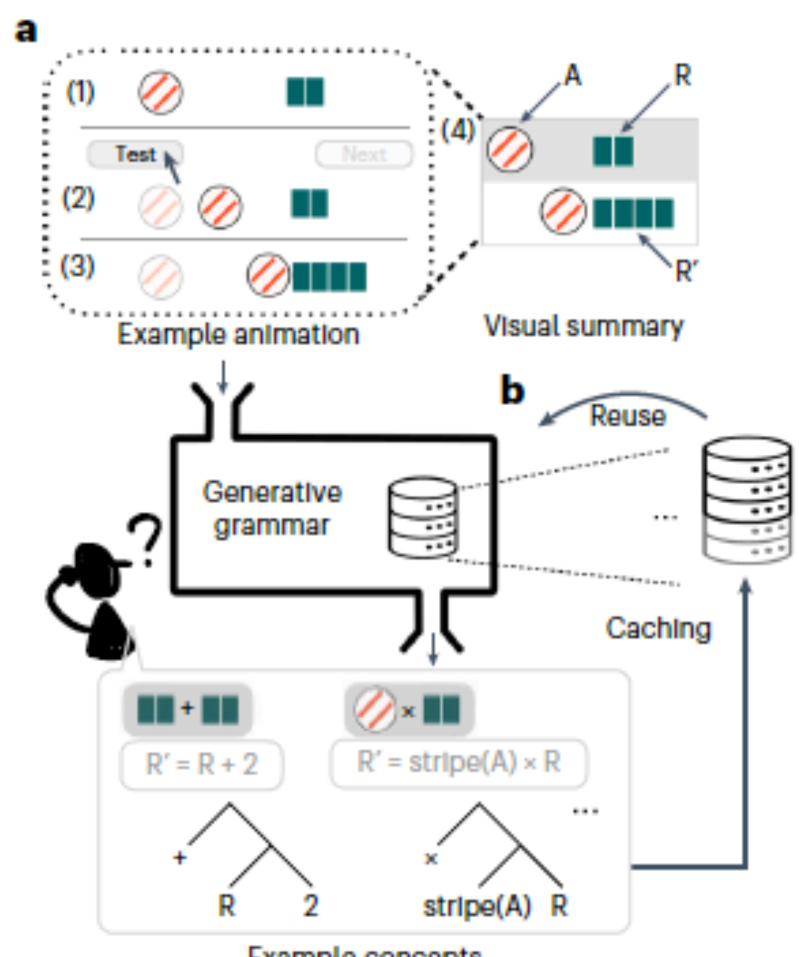
Bonan Zhao [1], Christopher G. Lucas [2] & Neil R. Bramley [1]

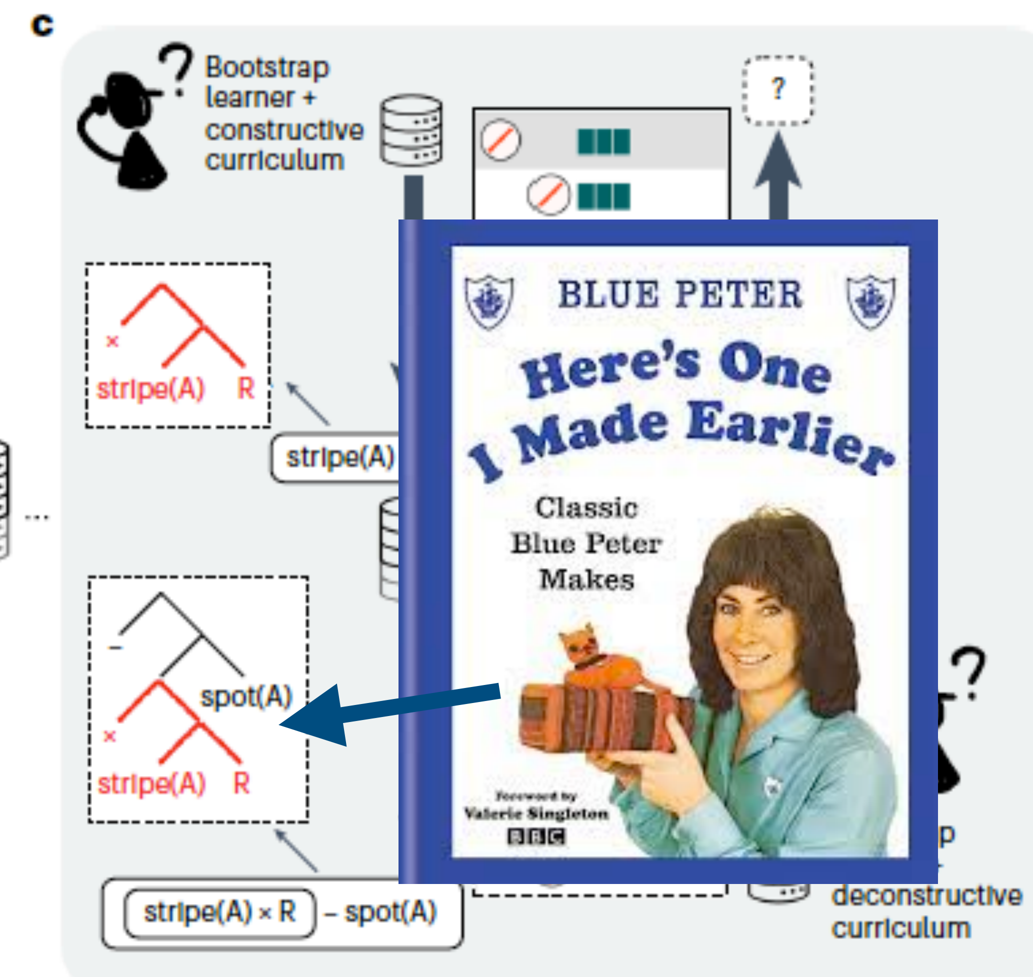
Accepted: 8 September 2023

We explore power of type-constrained generation mechanism with a **“flexible library”** of primitives

Show it captures dramatic order effects (very different conclusions from same evidence depending on order presented)







Discussion

- *Computational constructivist (or "learning as program induction")* framework ⇒ departure from fixed-dimension tasks & models

- Goal to explain *general* human **capacity** rather than explain *idiosyncratic* human **behaviours** (Bramley, Zhao, Quillien & Lucas, 2023, *TopICS*)

Overall perspective

- Hypothesis space infinite/unknown
  - Solve manageable subproblems while conditioning on surrounding beliefs
  - Chain hypothetical edits to search further for more radical changes
  - Balance search to retain connection with posterior
    → **MCMC-style search**
- Data large/weak Build on accumulated knowledge via iterative compositional search
  → **Adaptor Grammar bootstrapping**
- Data causally ambiguous Localise evidence to support adaptation and refinement
- Interact & intervene to resolve focal uncertainty
  → **Do[X]**

Conclusions

- Humans succeed at learning complex representations by striking a balance between exploring in the **world**…
  - actively interacting with their surroundings
- …and in the **mind**
  - actively adapting their theories and generating new hypotheses.

Thanks for listening!

https://bramleylab.ppls.ed.ac.uk

Thanks to collaborators on this stuff

Jan Philipp Fränken, Stanford

Maarten Speekenbrink, UCL

Todd Gureckis, NYU

David Lagnado, UCL

Peter Dayan, Tubingen

Tom Griffiths, Princeton

Fei Xu, Berkeley

Bonan Zhao, Princeton

https://www.justgiving.com/page/edpsych-movember-2024