Algorithms for decision-making and decision-making for algorithms

Fred Callaway
Princeton University

Rational decision-making

Perfect rationality

- Take the action with maximal expected utility.

$$\arg\max_a \mathbb{E}\left[U(a)\right]$$

"Do the right thing"

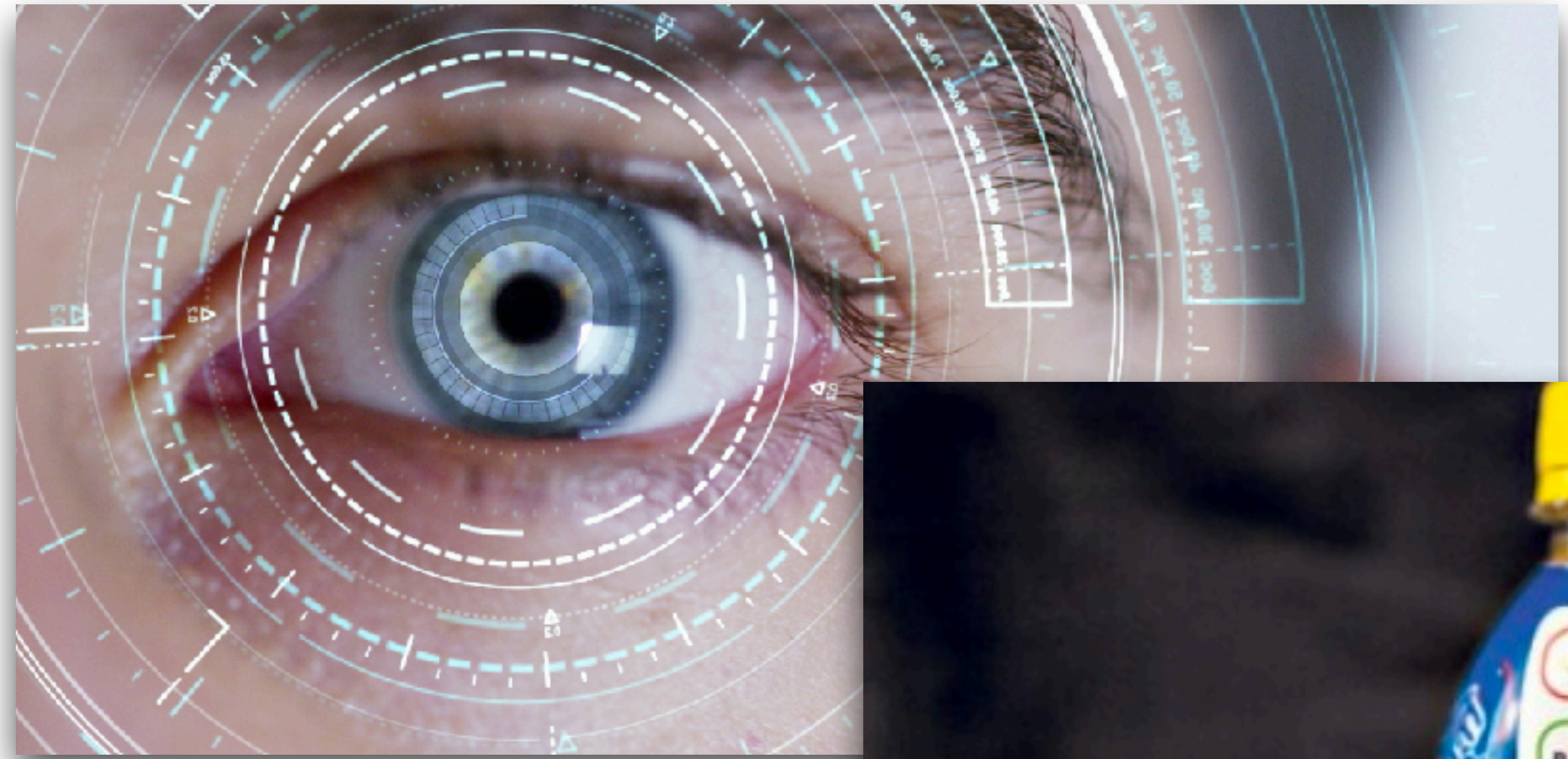
Metalevel rationality

- Use the *cognitive strategy* that best trades off utility and computational cost.

$$\arg\max_\pi \mathbb{E}\left[\max_a \mathbb{E}\left[U(a)|B_T\right] - \sum_{t=0}^{T-1} \text{cost}(B_t, C_t) \,\middle|\, C_t \sim \pi(B_t)\right]$$

"Do the right *thinking*"

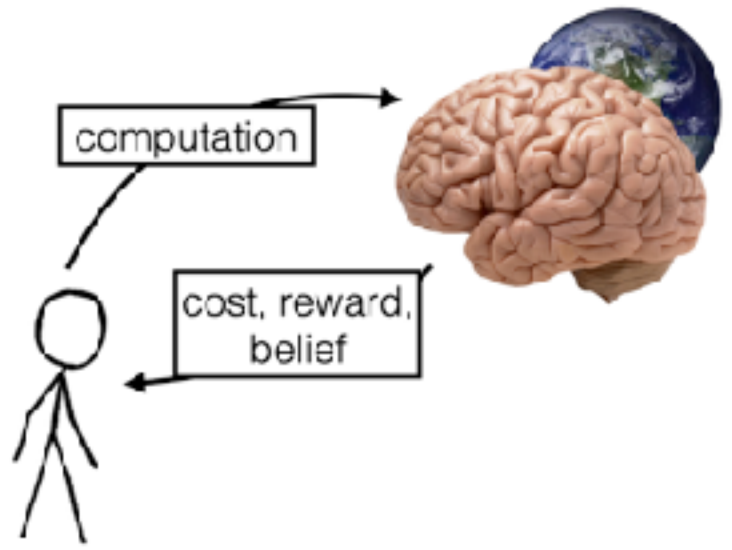
Russell & Wefald (1991)



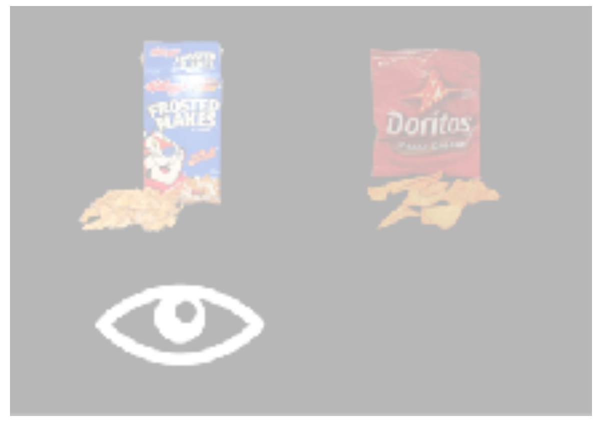

ALTO en AZÚCAR
BAJO en SAL
no contiene GRASA
ALTO en AZÚCAR
MEDIO en GRASA
BAJO en SAL
MEDIO en SAL
BAJO en AZÚCAR


the rematch

Metalevel MDPs



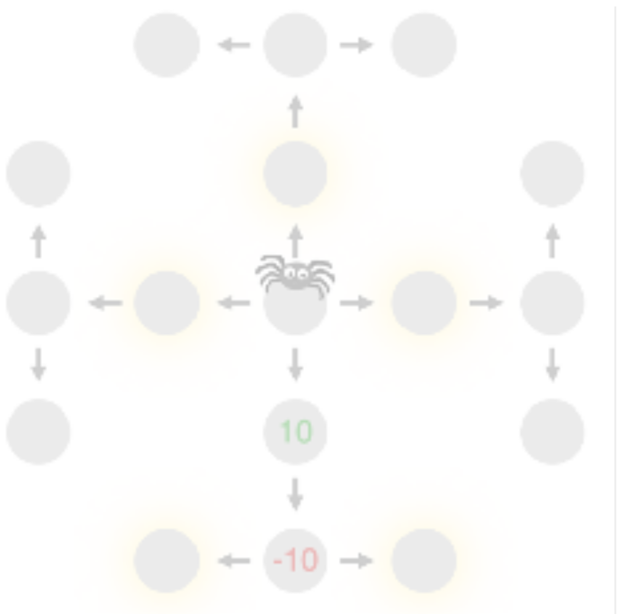
Simple decisions



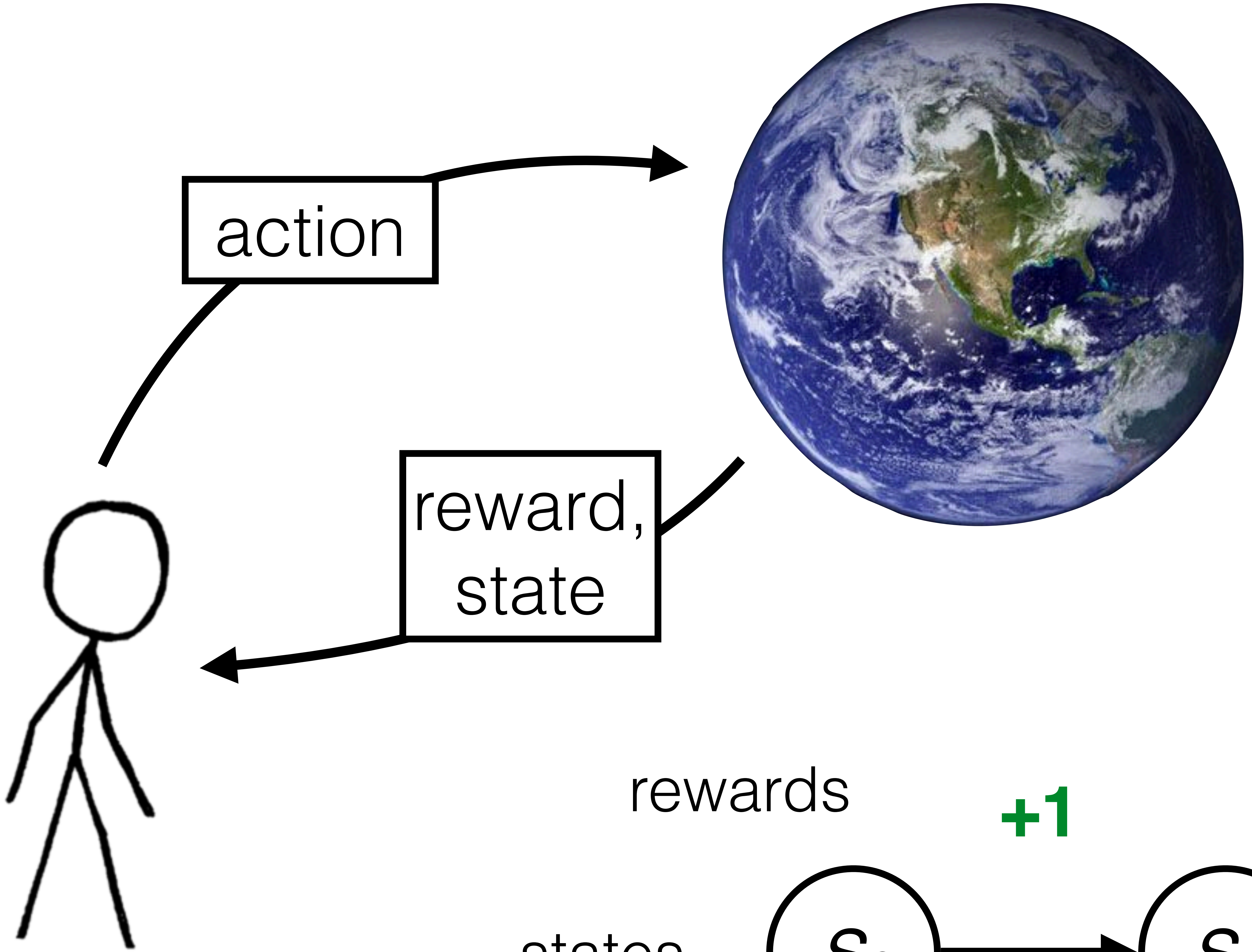
Multi-attribute decisions

Prizes	Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
A: 3 points	2		3	4	
B: 2 points	7				7
C: 2 points	7	4		2	
D: 21 points	7		8	6	
E: 2 points	9				6

Sequential decisions

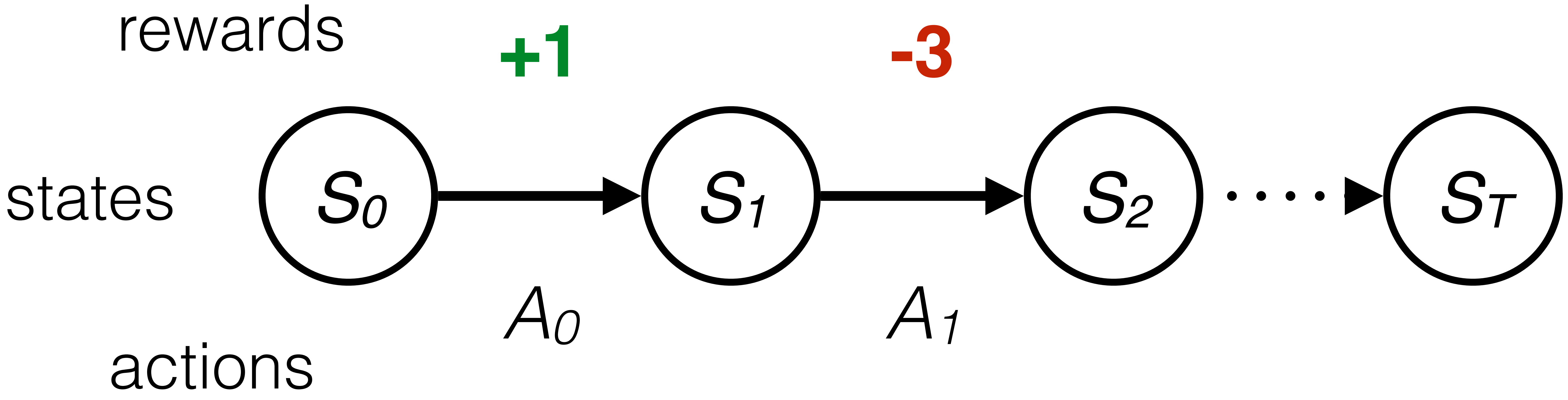


Markov decision process (MDP)

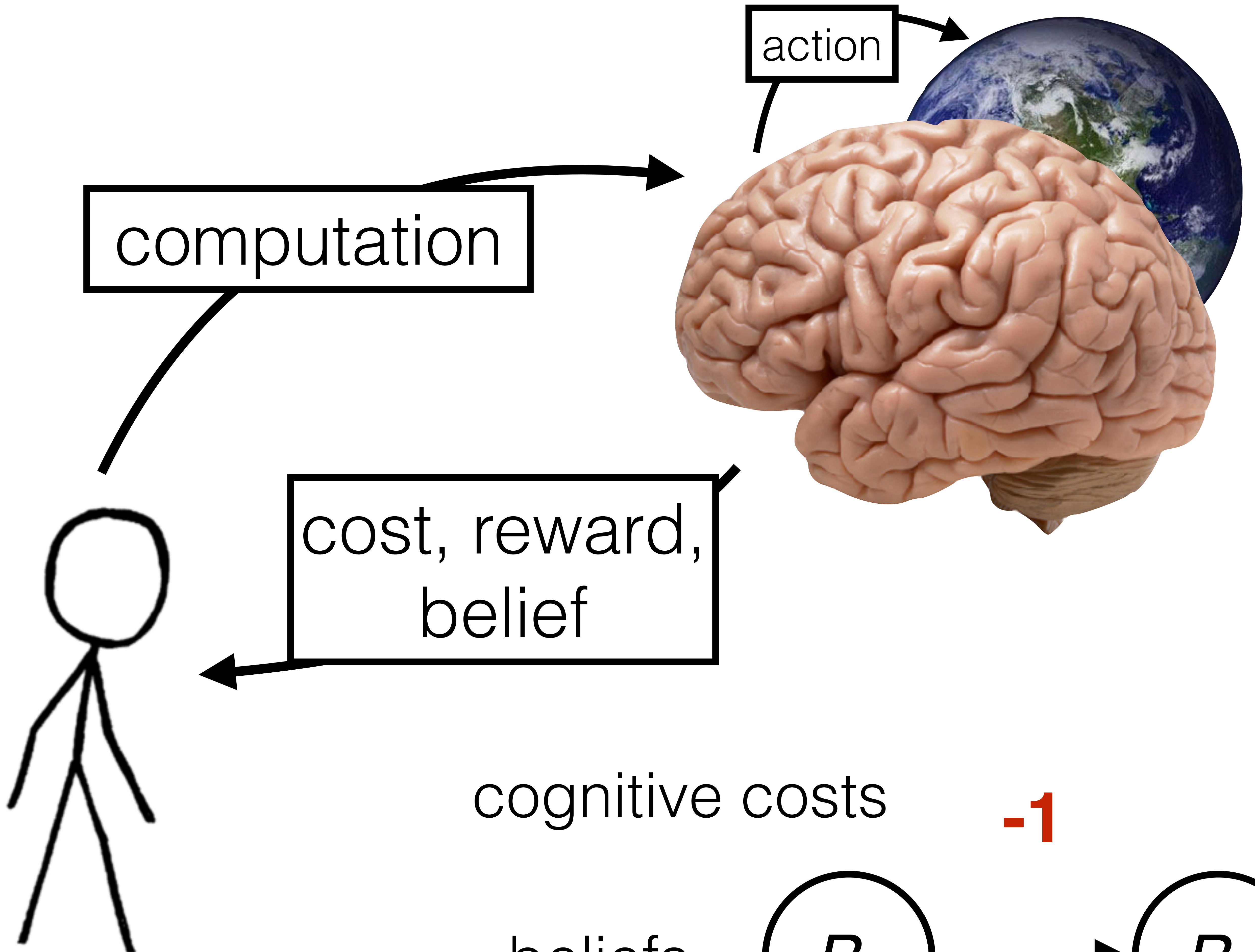


policy $\pi : \mathcal{S} \to \Delta(\mathcal{A})$

optimal policy $\arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T} R_t \mid A_t \sim \pi(S_t) \right]$



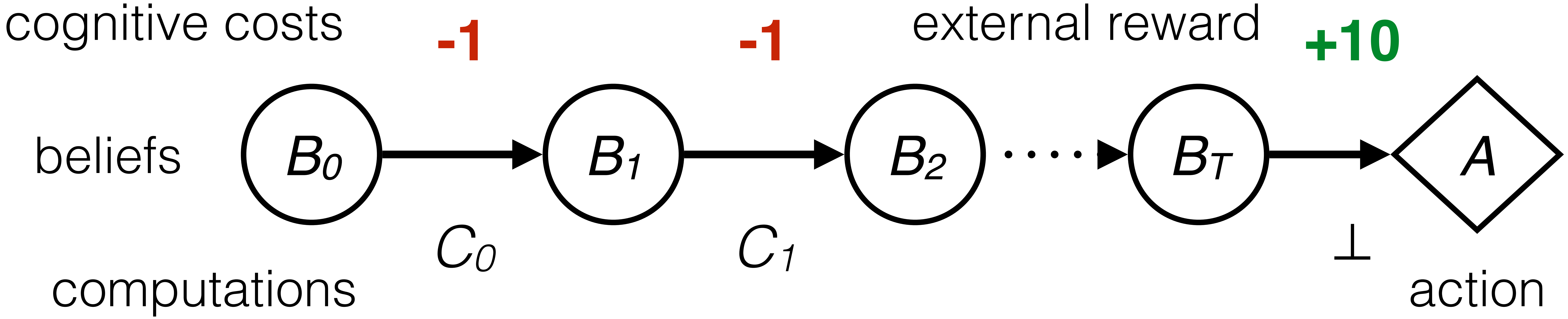
Metalevel Markov decision process (meta MDP)



cognitive strategy $\pi : \mathcal{B} \rightarrow \Delta(\mathcal{C})$

optimal cognitive strategy $\arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T} R_t \mid C_t \sim \pi(B_t)\right]$

$$\max_{a} \mathbb{E}\left[U(a)|B_T\right] - \sum_{t=0}^{T-1} \mathrm{cost}(B_t, C_t)$$

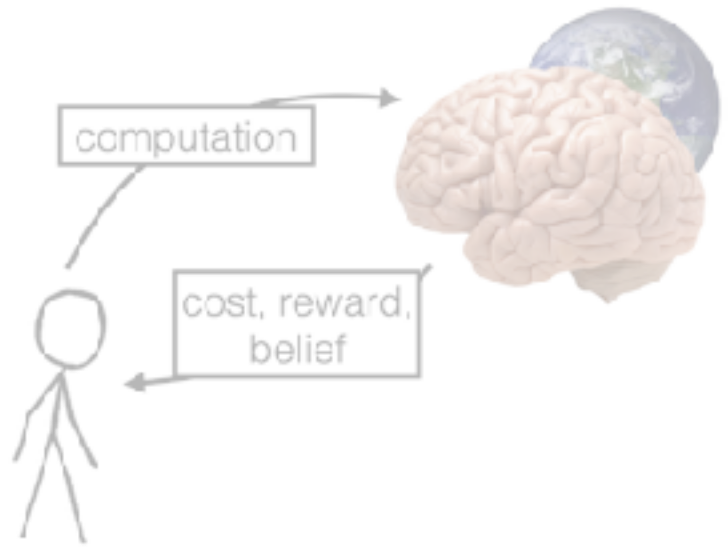


Hay et al. (2012)

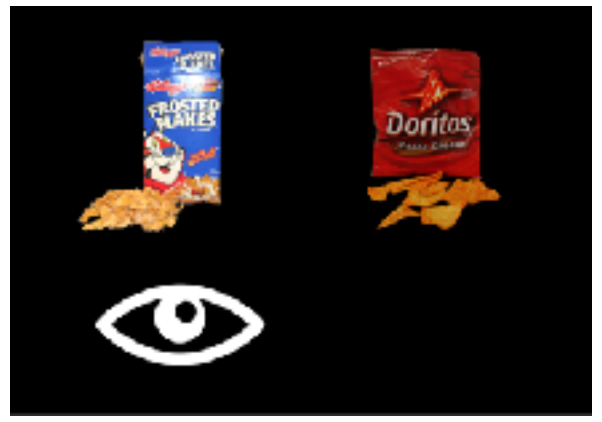
We must be prepared to accept the possibility that what we call "the environment" may lie, in part, within the skin of the biological organisms.

Simon (1955)

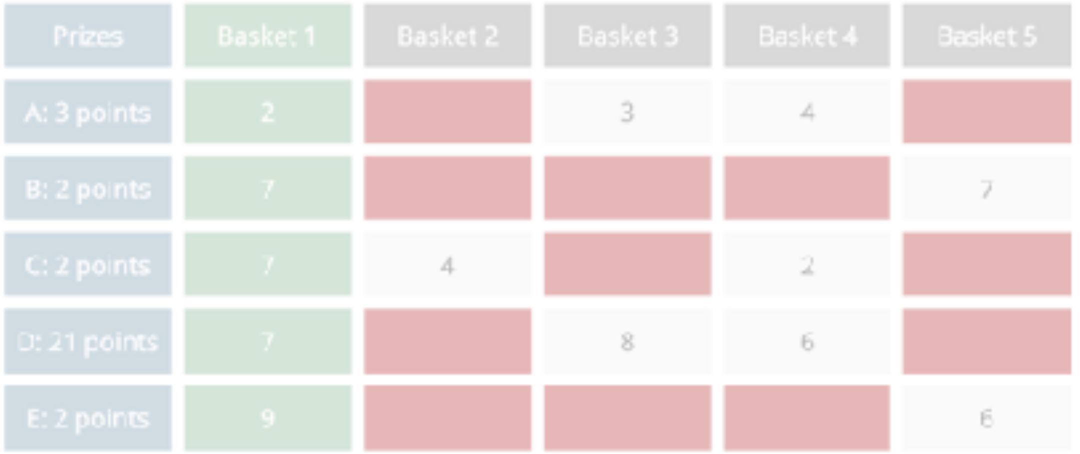
Metalevel MDPs



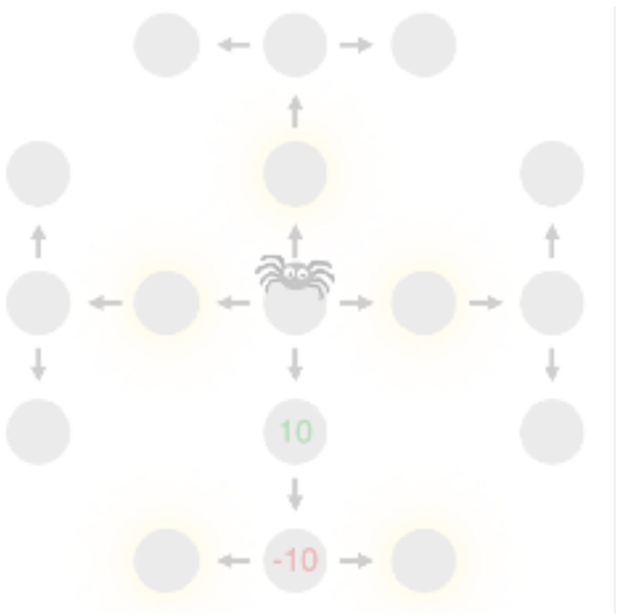
Simple decisions



Multi-attribute decisions



Sequential decisions

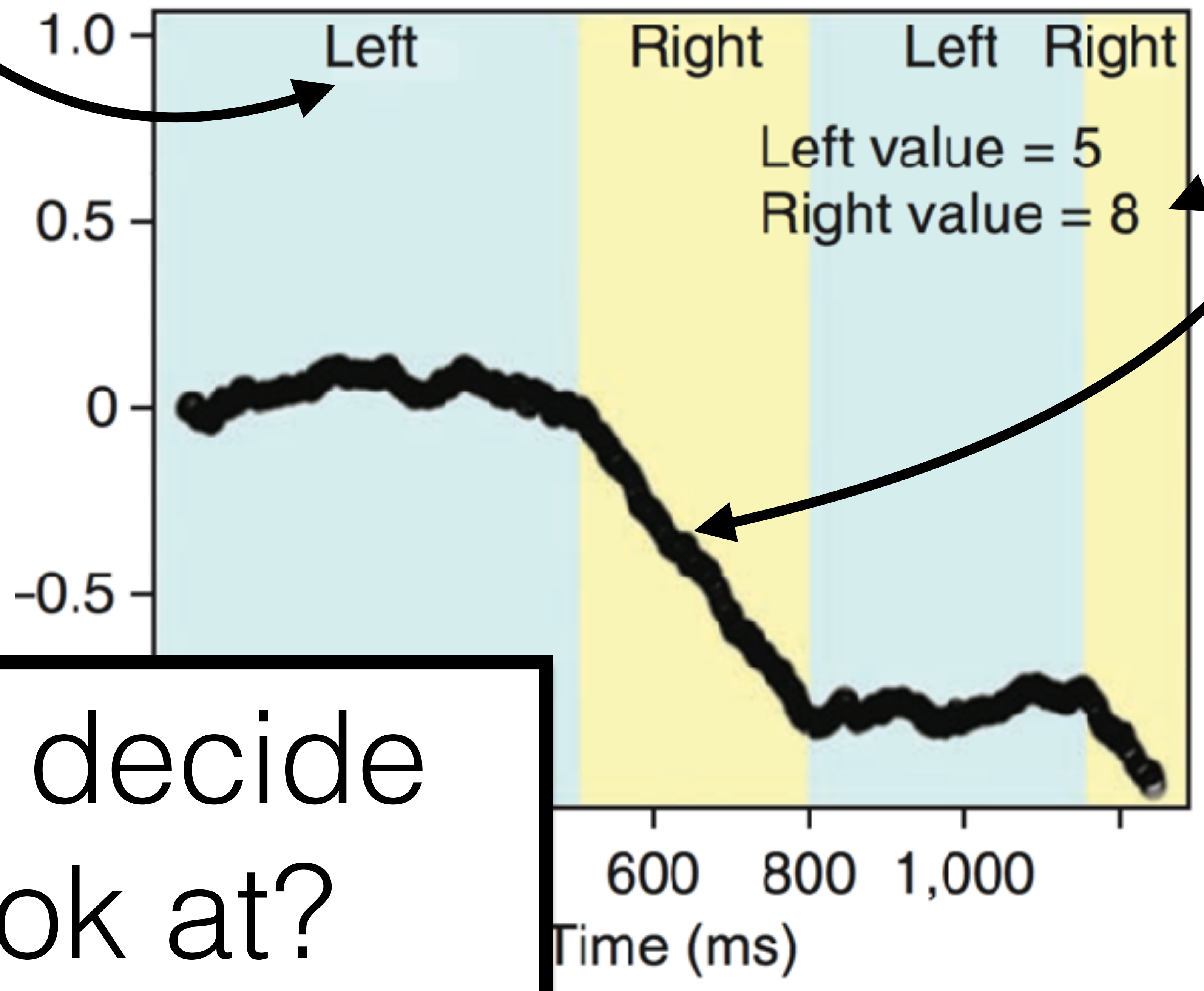


Background: Attention in preferential choice

evidence biased in favor of fixated option

aDDM

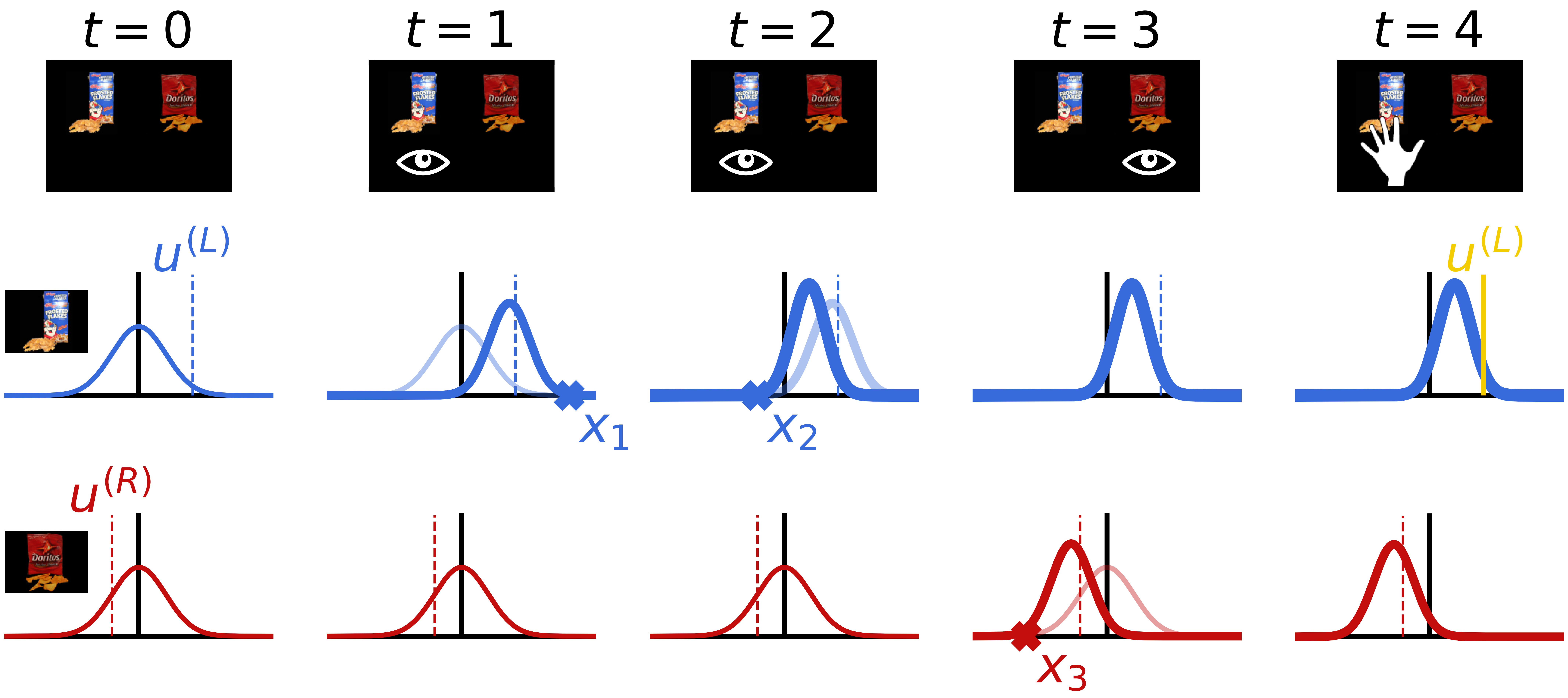
accumulate noisy evidence about the value of the options



How do we decide what to look at?

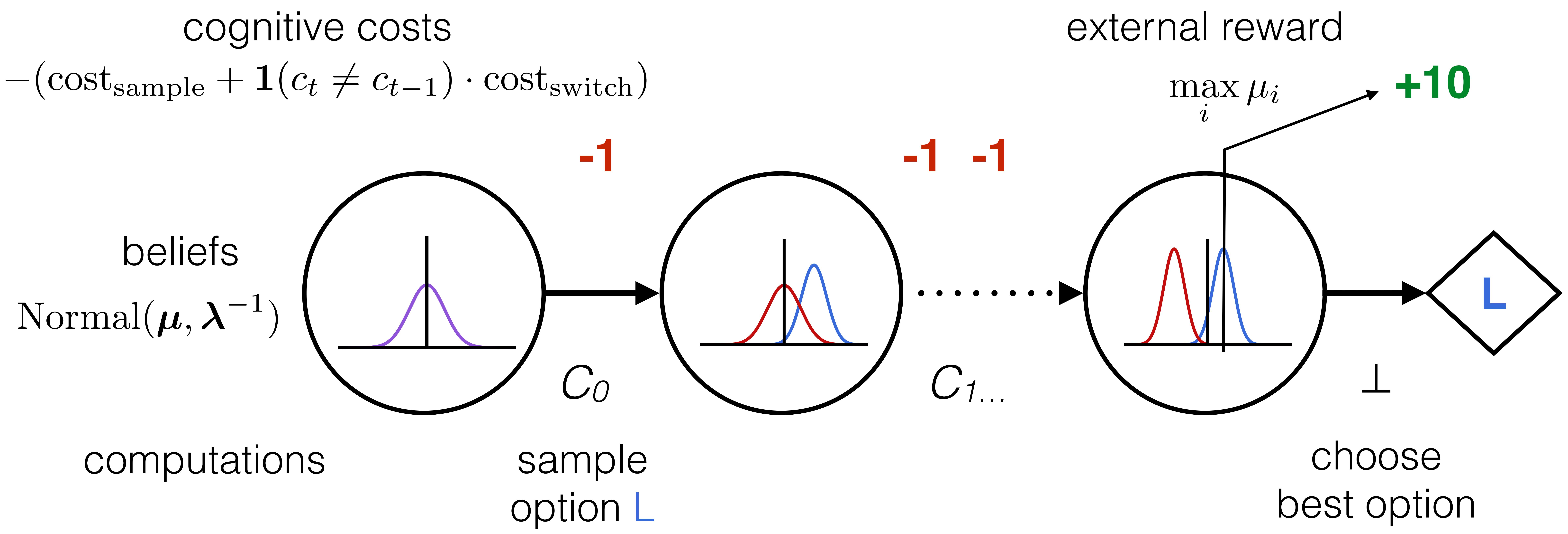
Krajbich, Armel, Rangel (2010)

Model: Bayesian evidence accumulation



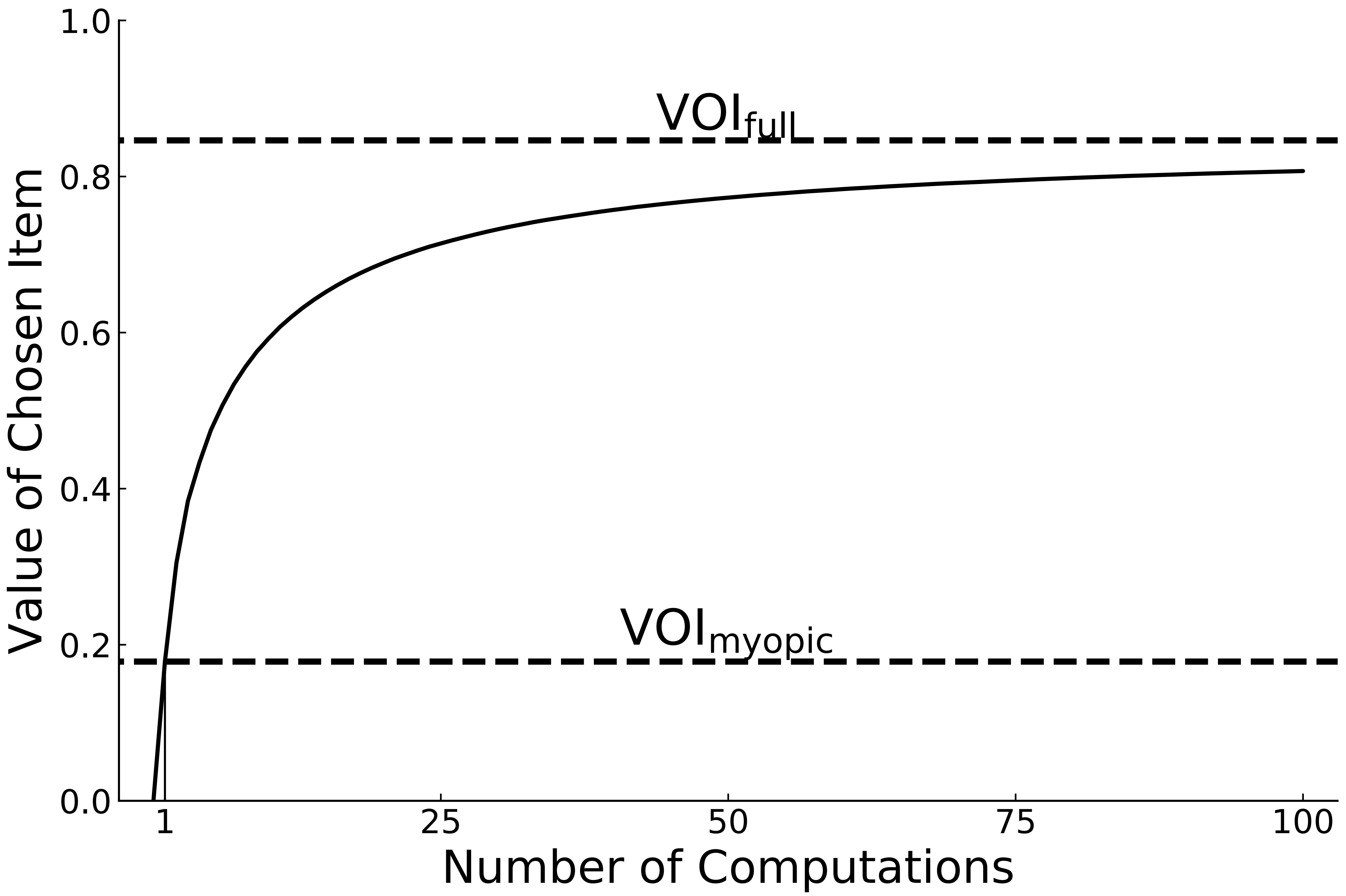
Callaway, Rangel, Griffiths (2021)

Model: Meta MDP



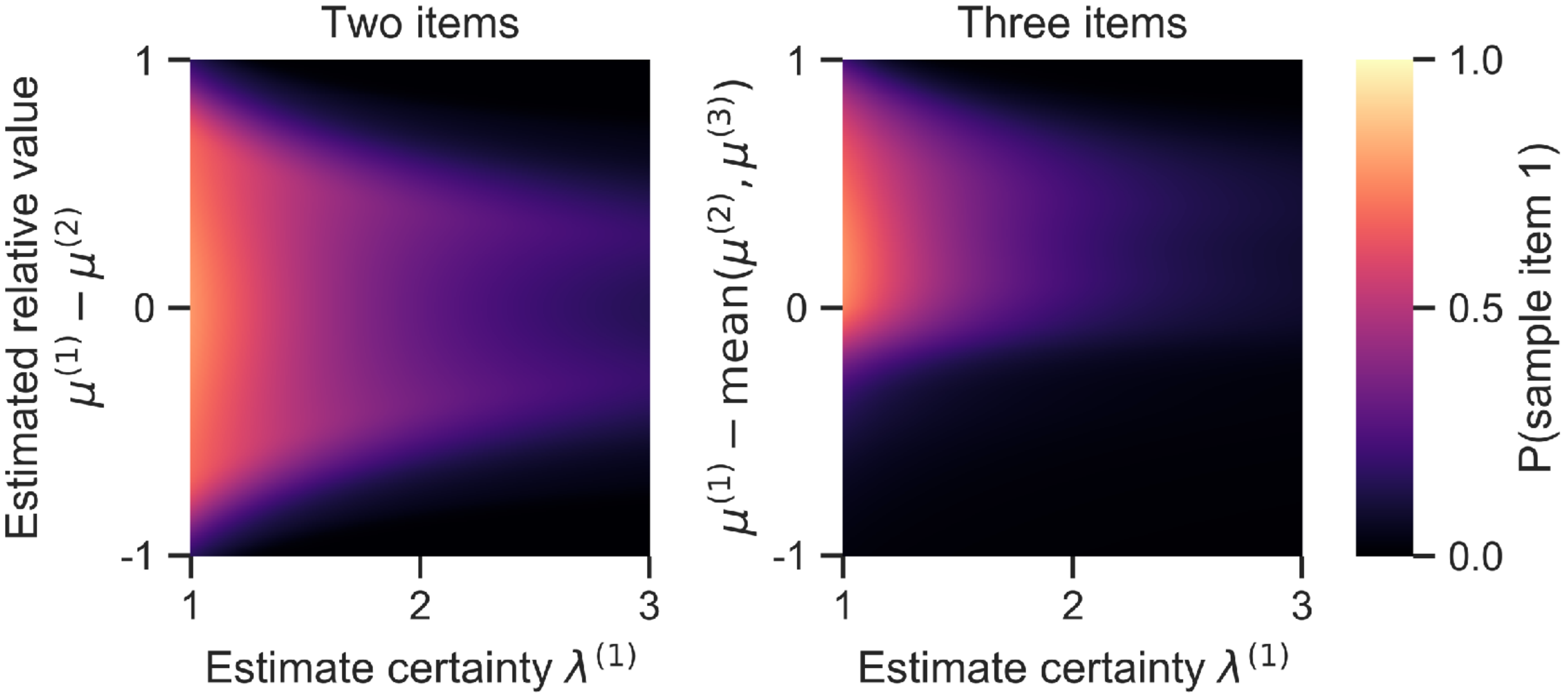
Model: Optimal policy

- Approximate value of computation with a linear combination of *value of information* features.
- Find weights that maximize meta-level reward.

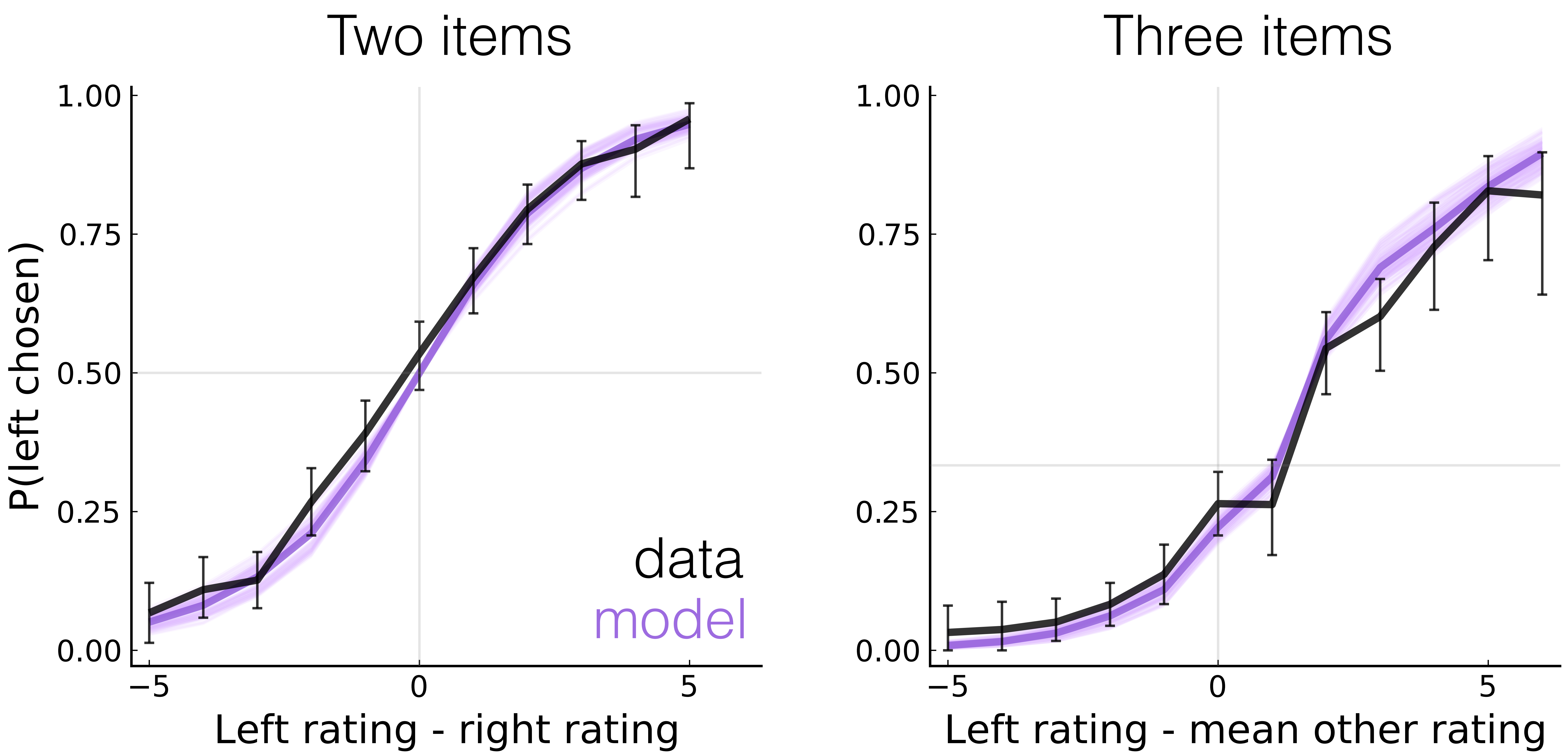


Callaway et al. (2018)

Model: Optimal policy

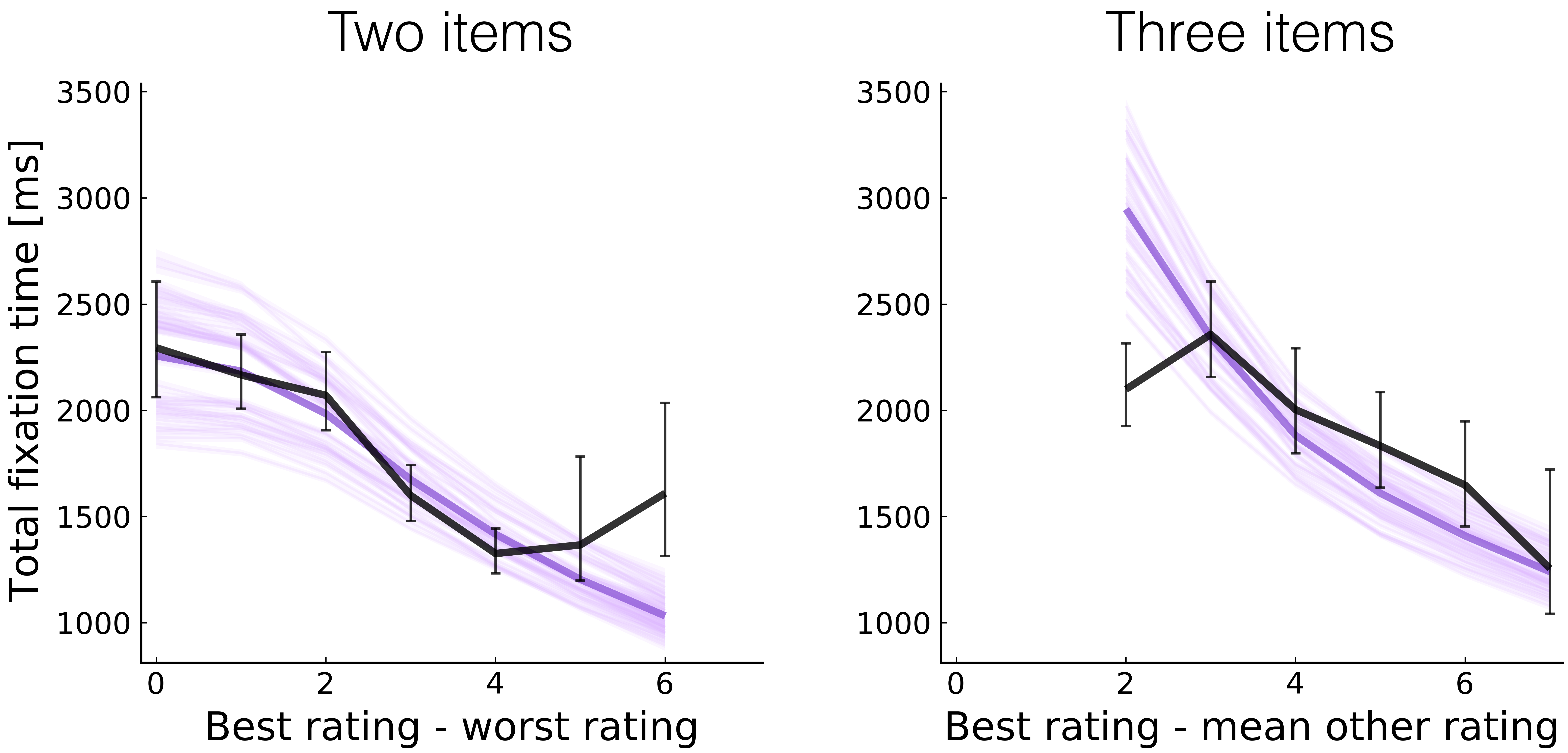


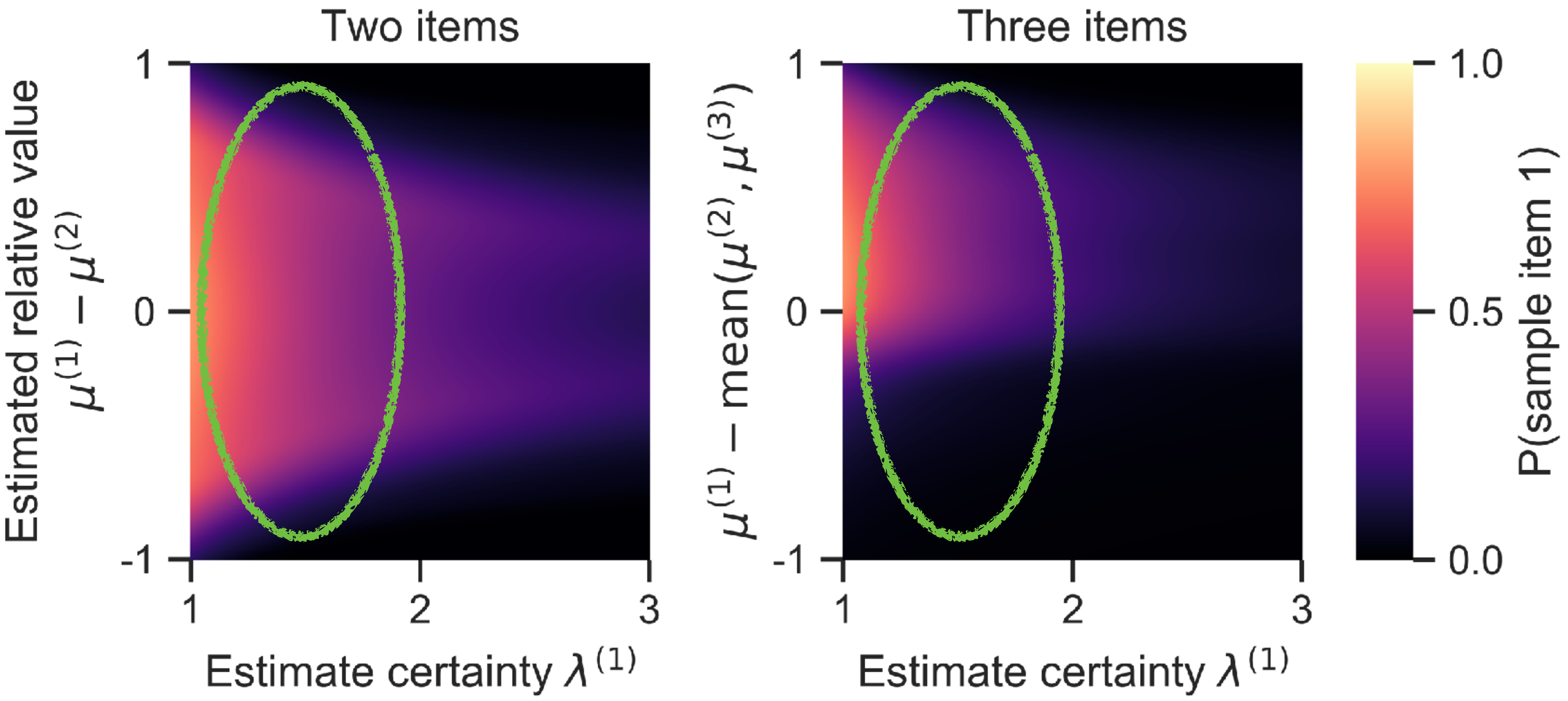
Results: People choose things they like more



data: Krajbich, Armel Rangel (2010), Krajbich & Rangel (2011)

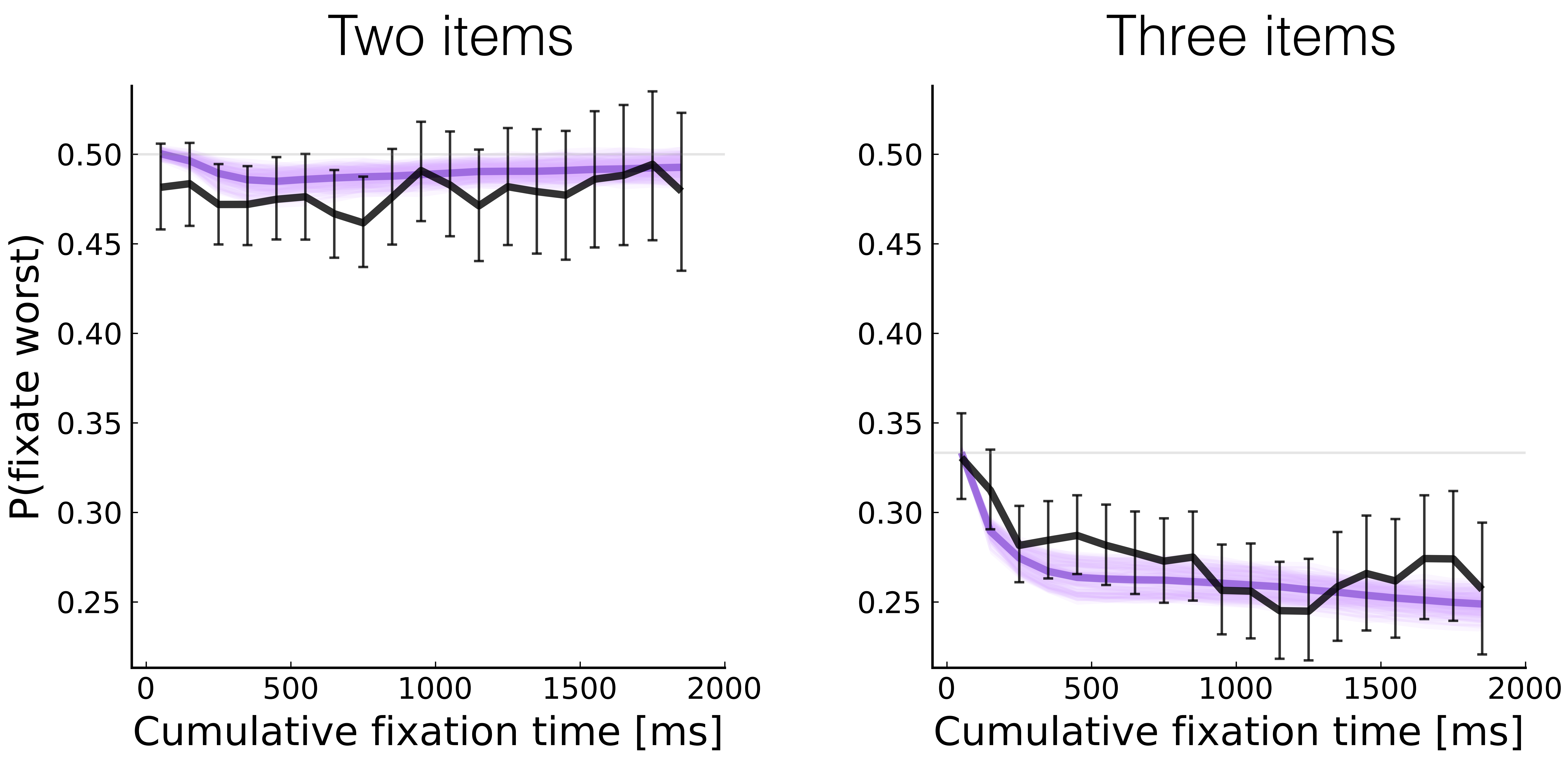
Results: People *quickly* choose things they like *a lot* more



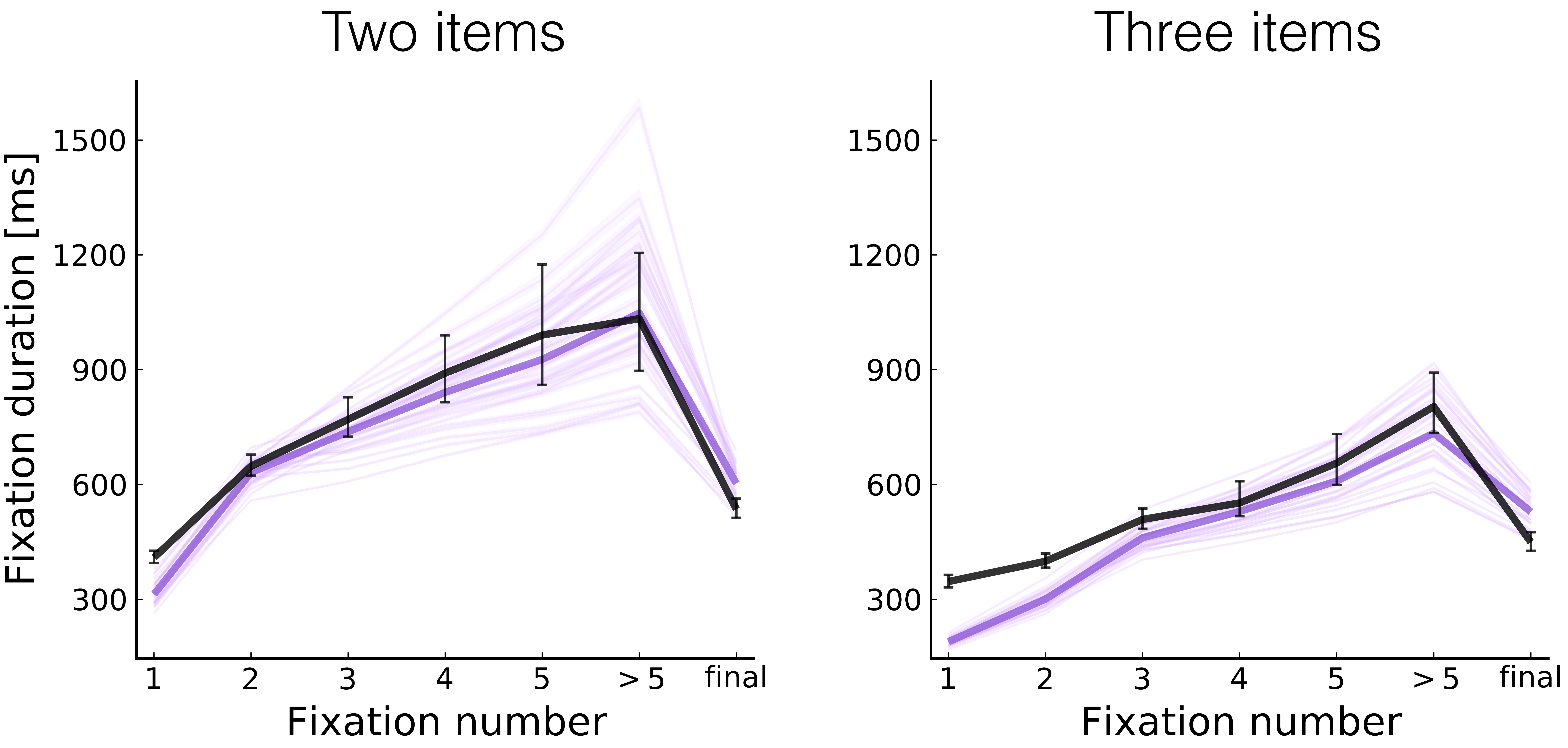


Two items
Three items
Estimated relative value
$\mu^{(1)} - \mu^{(2)}$
$\mu^{(1)} - \text{mean}(\mu^{(2)}, \mu^{(3)})$
1
0
-1
1
2
3
Estimate certainty $\lambda^{(1)}$
Estimate certainty $\lambda^{(1)}$
1.0
0.5
0.0
P(sample item 1)

Results: Least valuable item fixated less later in trial



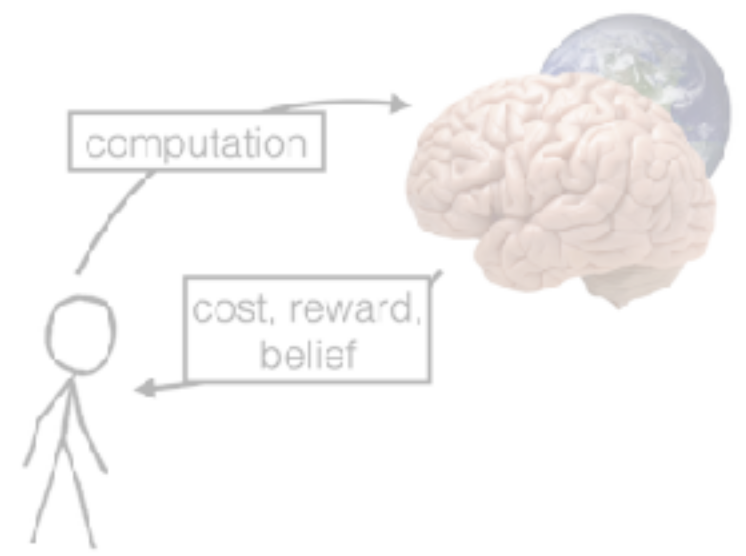
Results: Fixations are longer later in the trial



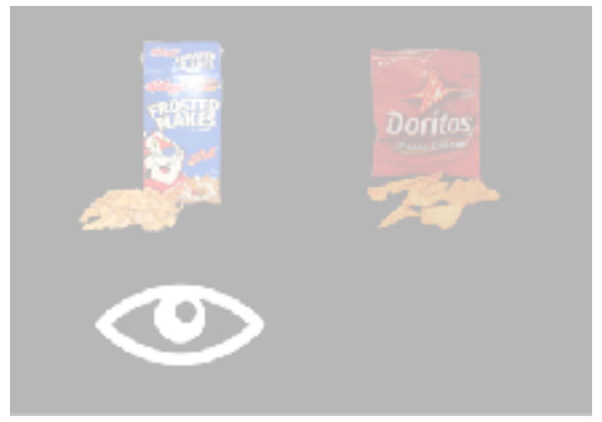
Summary: Rational attention in simple choice

- Directing one's attention when making a decision can be modeled as a meta MDP where an agent estimates the value of each choice option from a sequence of noisy signals.
- Human fixations in simple choice tasks are consistent with a near-optimal solution to that meta MDP.
- Like the optimal model, people selectively attend to options they think are valuable, but only when there are more than two options.
  - ➡ People might be only partially sensitive to the qualification.

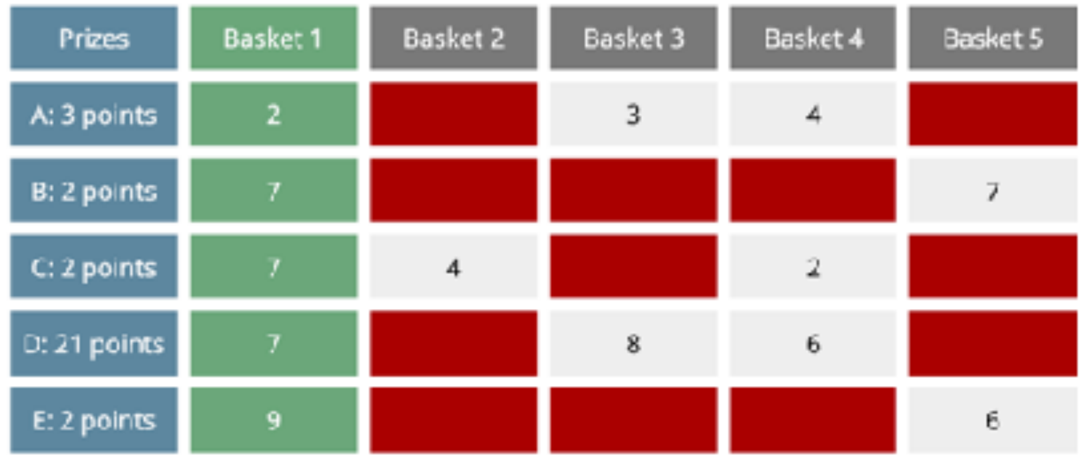
Metalevel MDPs



Simple decisions

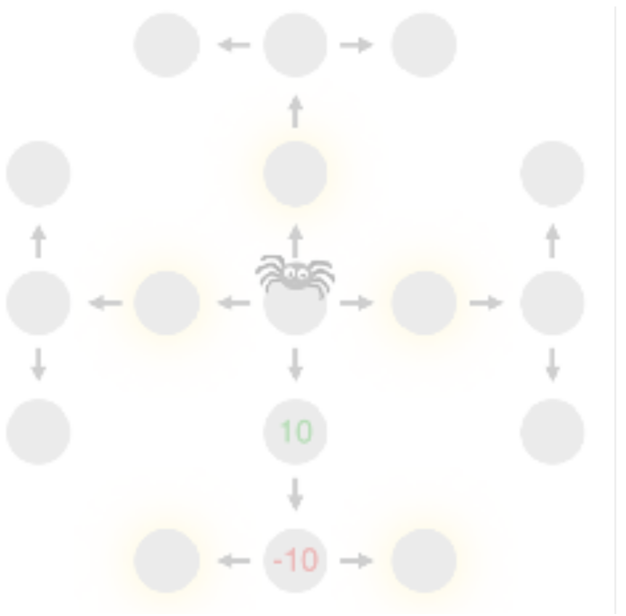


Multi-attribute decisions



Prizes	Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
A: 3 points	2		3	4	
B: 2 points	7				7
C: 2 points	7	4		2	
D: 21 points	7		8	6	
E: 2 points	9				6

Sequential decisions

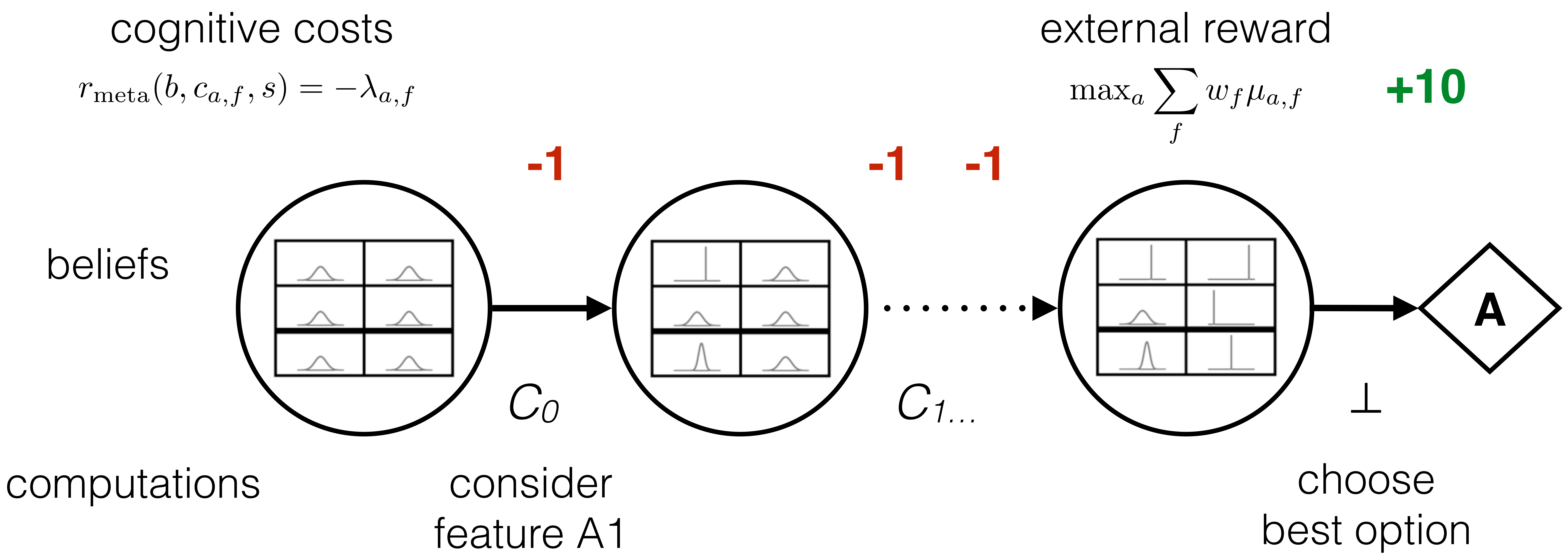


Model: Multi-attribute choice

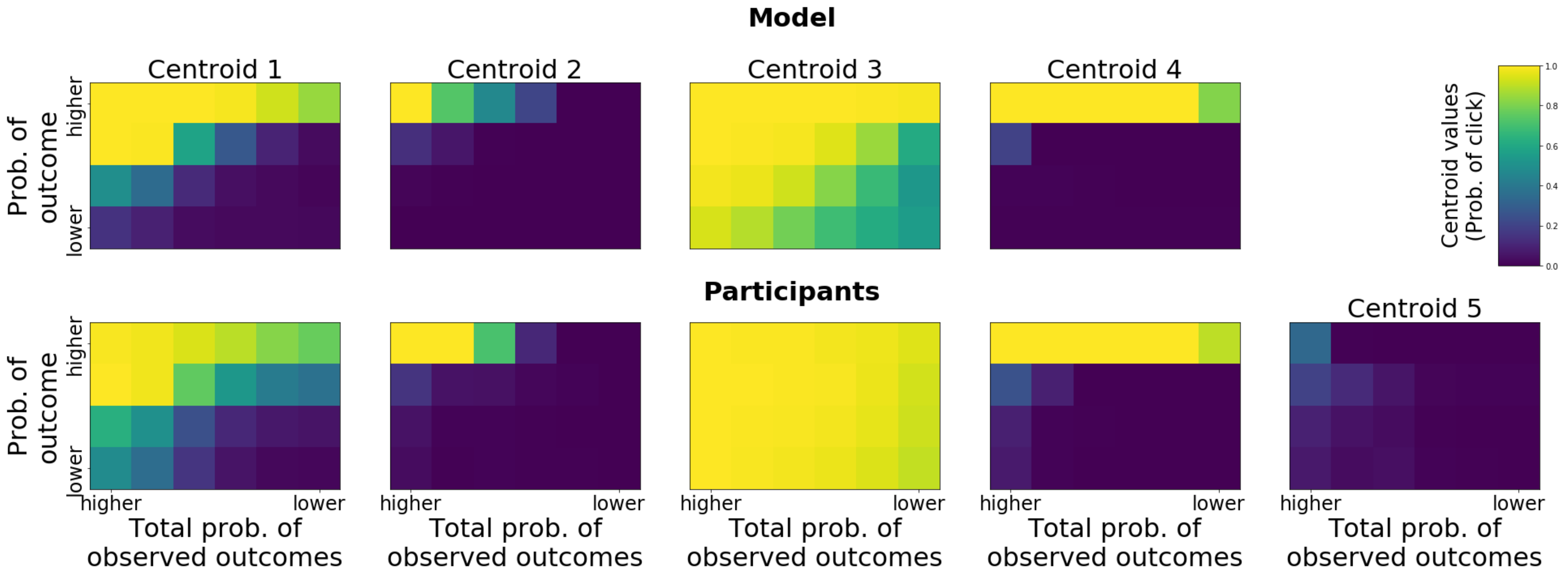
	option A	option B	
feature 1	$x_{A,1}$	$x_{B,1}$	w_1
feature 2	$x_{A,2}$	$x_{B,2}$	w_2
utility	r_A	r_B	$r_a = \sum_f w_f x_{a,f}$

Model: Belief updating

Model: Meta MDP



Results: Optimal decision heuristics



Krueger*, Callaway*, Lieder, Griffiths (in prep)

Results: Adaptation to the environment

Can we use this information to improve people's decisions?

Krueger*, Callaway*, Lieder, Griffiths (in prep)

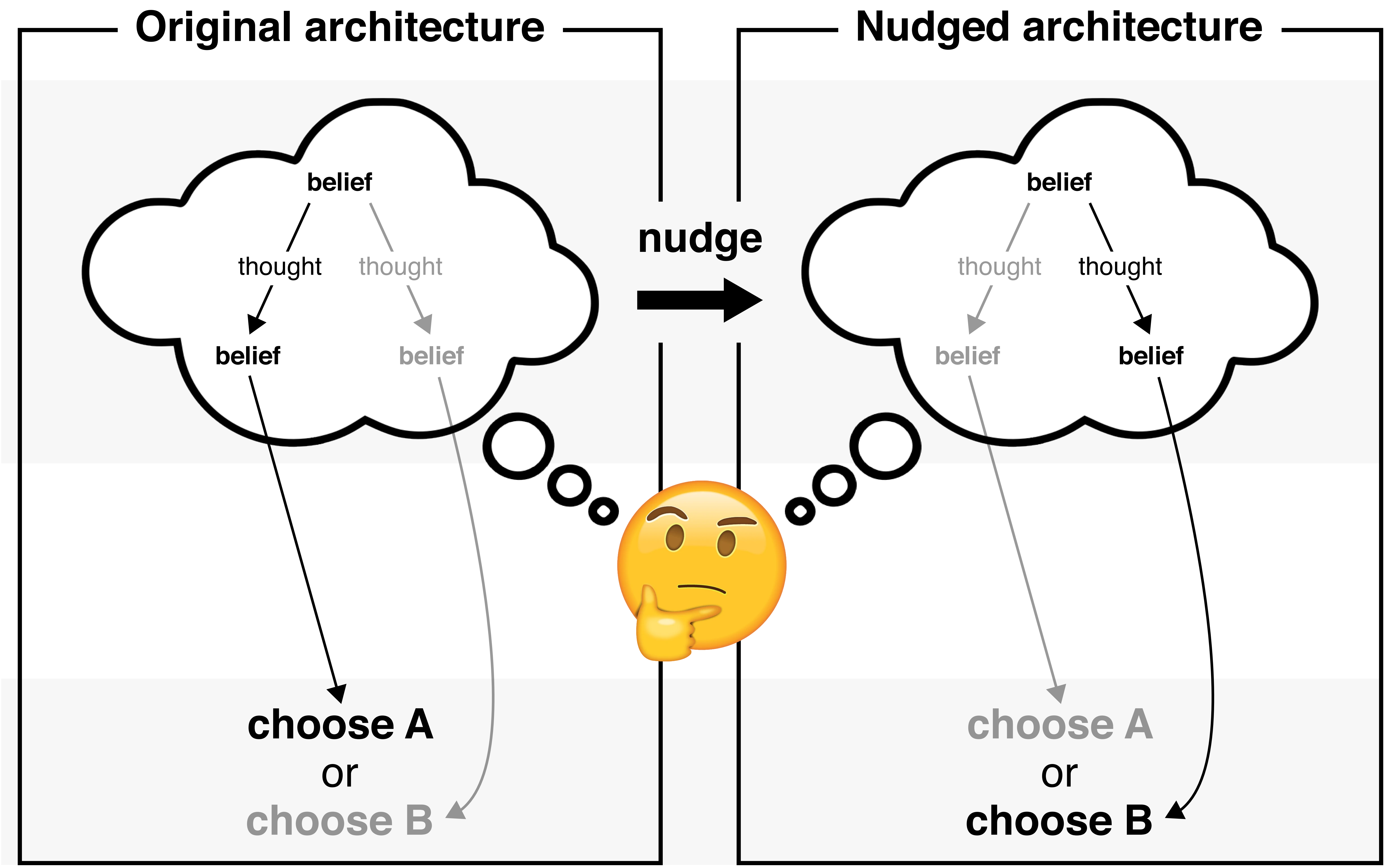
Application: Nudging

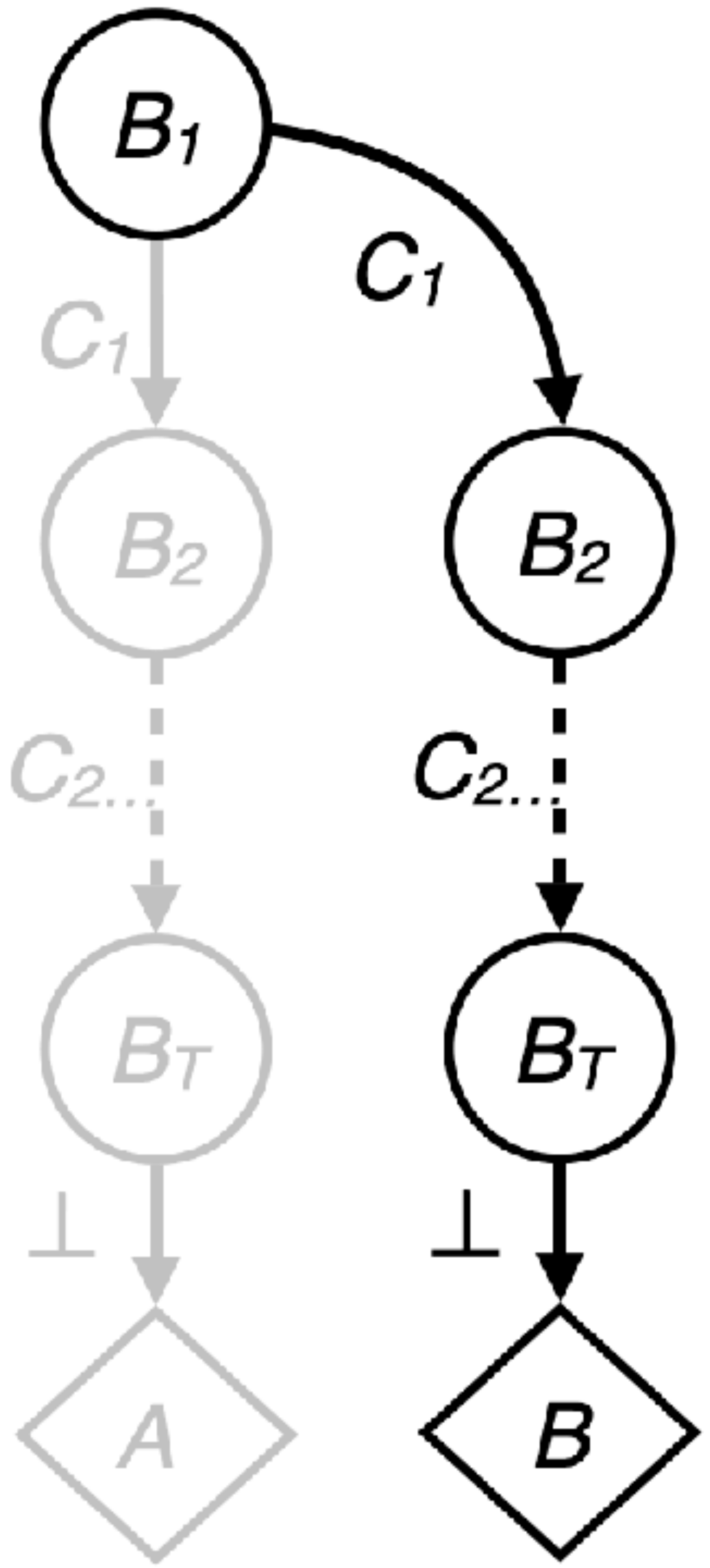
- Use findings from psychology to improve decisions by redesigning *choice architectures:* changing how choices are presented.
- Don't change economic incentives or restrict freedom of choice.

Examples

- Default options
- "Traffic light" labeling

Model: Nudging as modifying a meta MDP

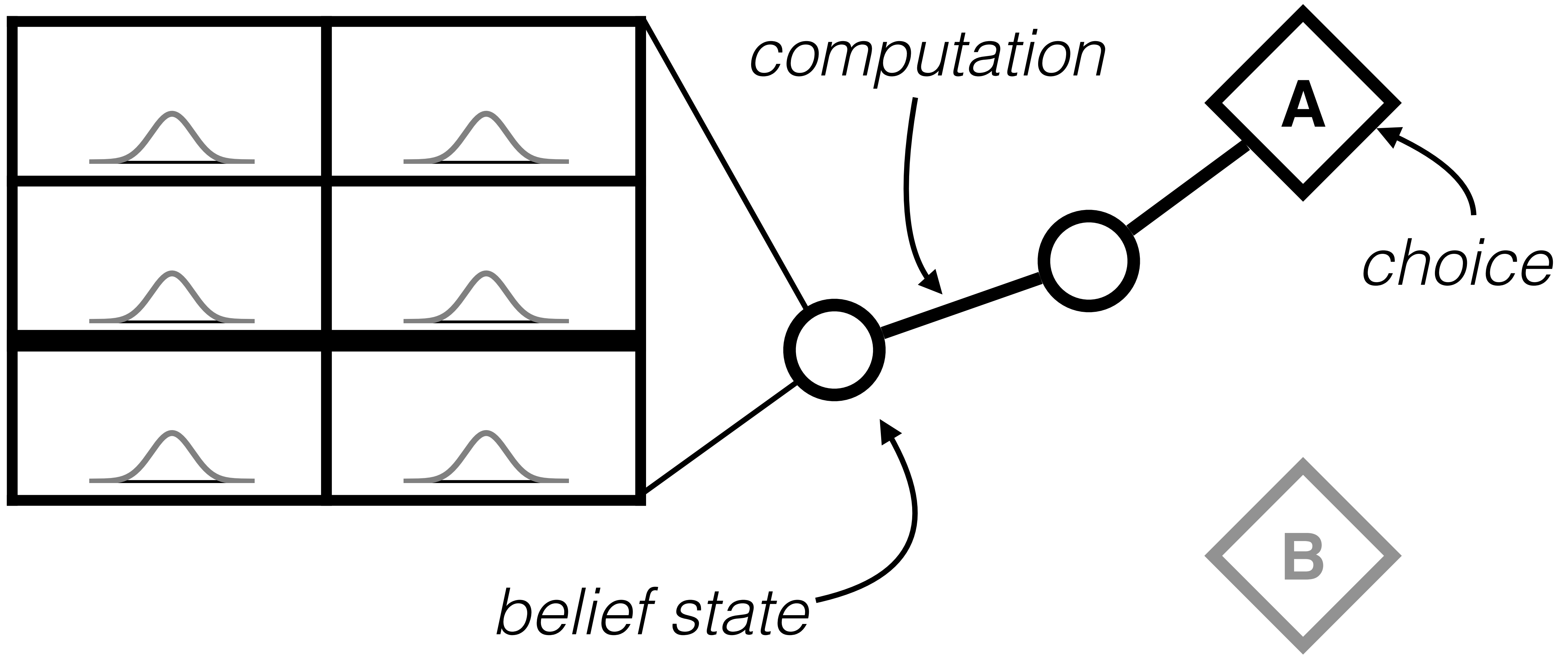




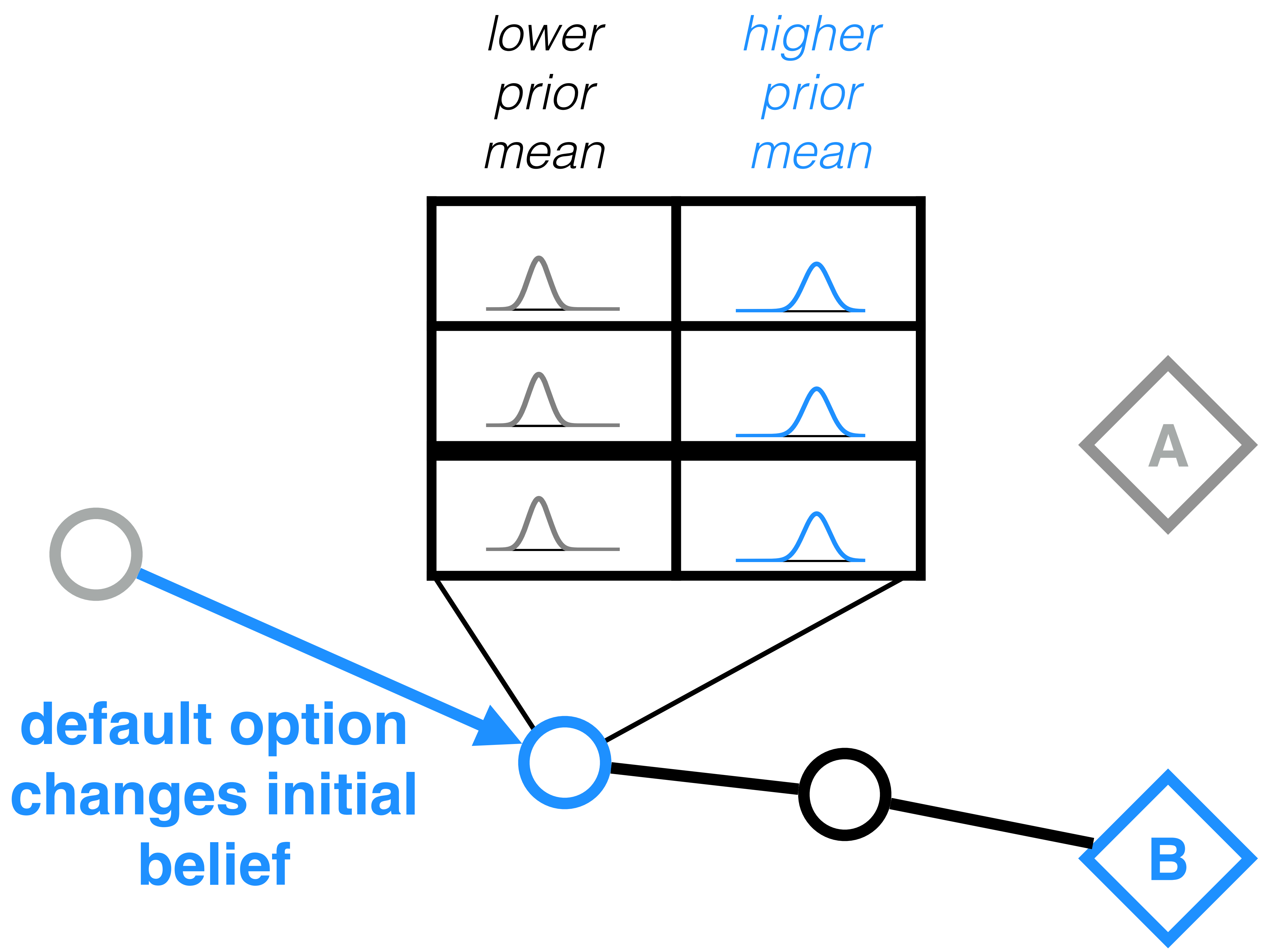
Callaway*, Hardy*, Griffiths (under review)

Model: Default options as recommendations

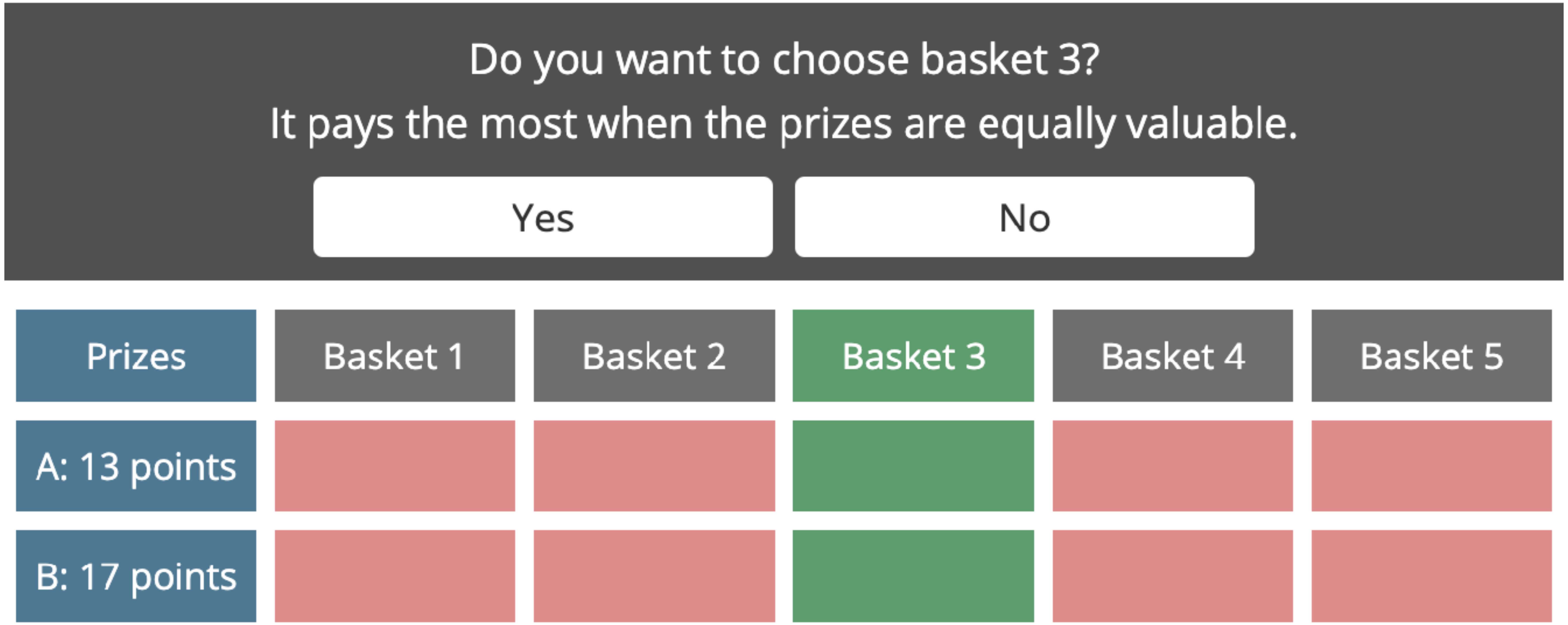
no default



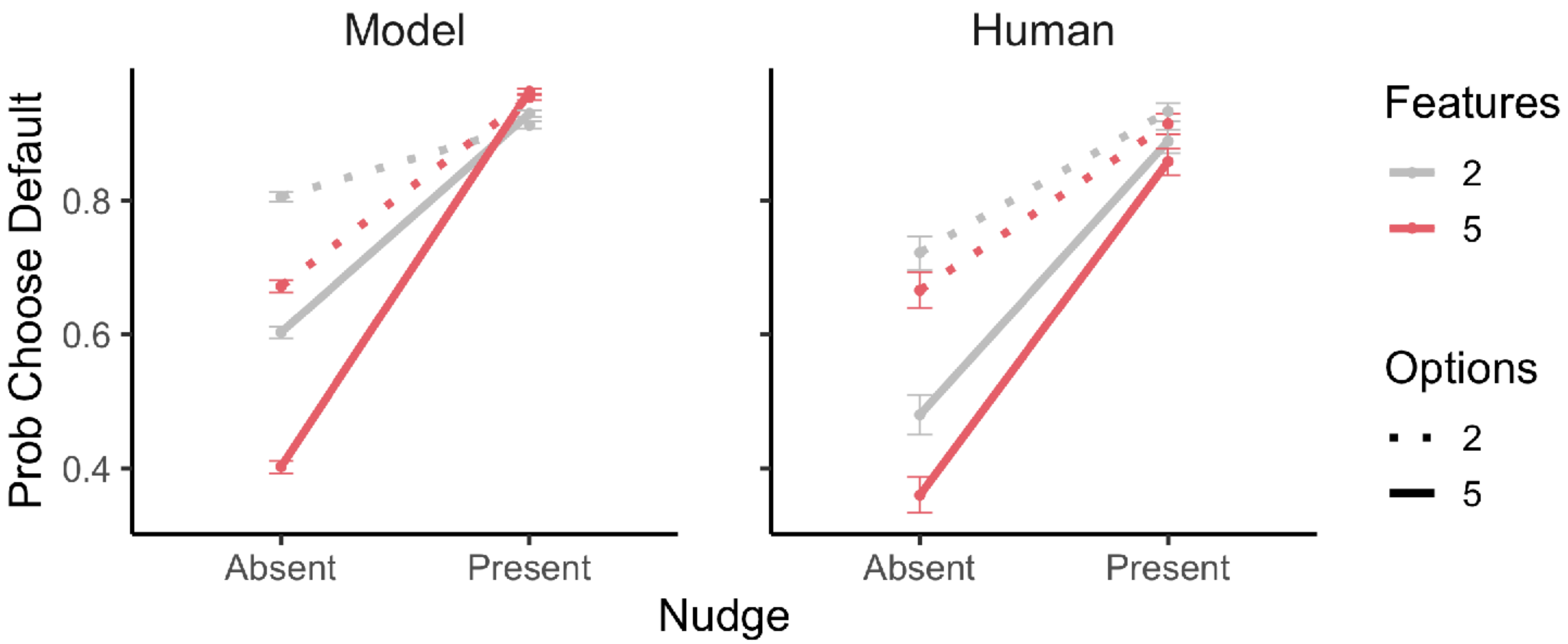
with a default



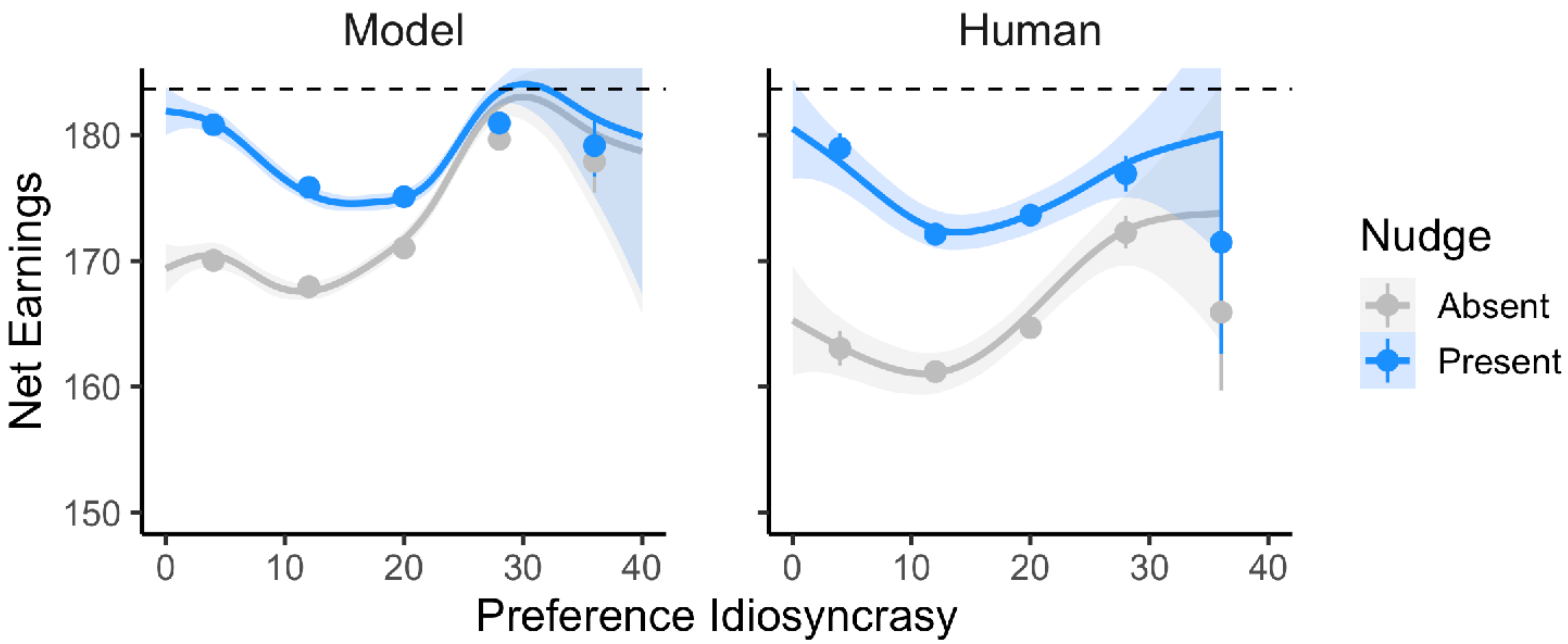
Experiment: Default options in Mouselab



Results: Defaults more effective on complex decisions



Results: Defaults more beneficial for typical preferences



Experiment: Traffic light labeling in Mouselab

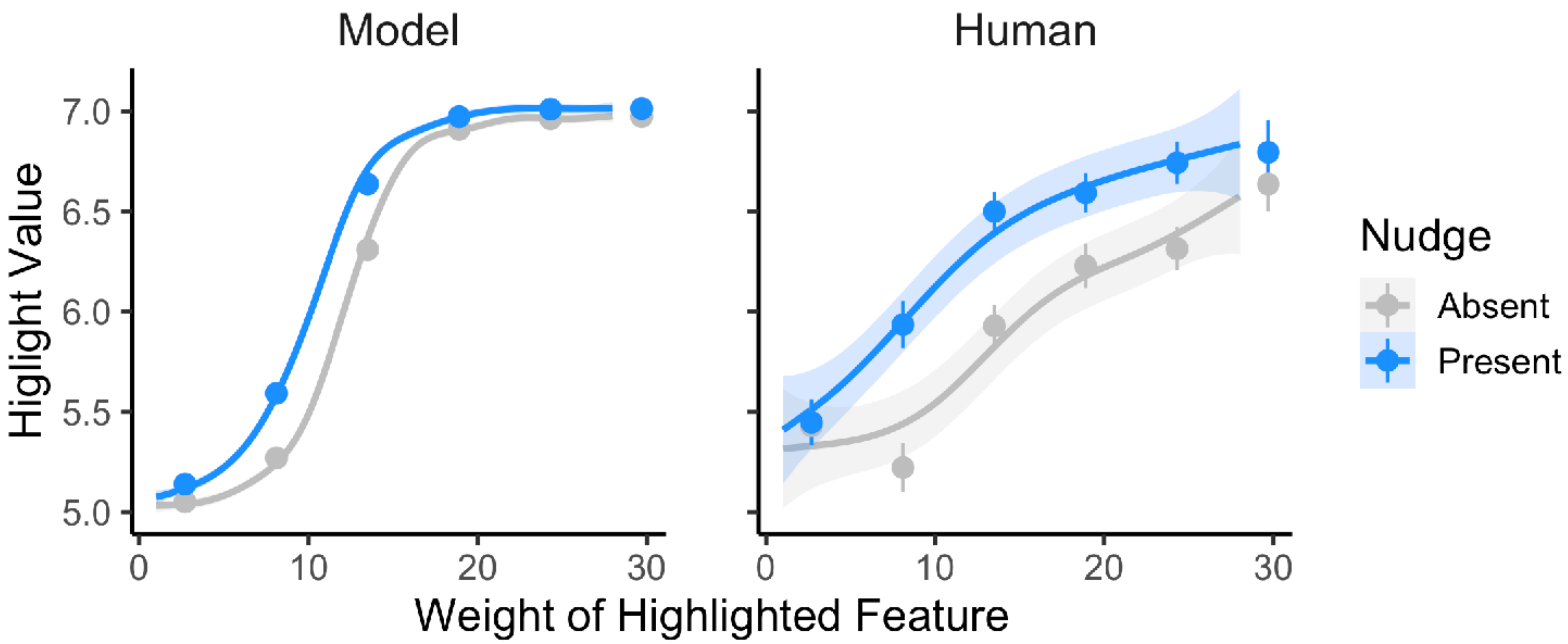
Clicks to reveal box

- 1
- 2
- 3

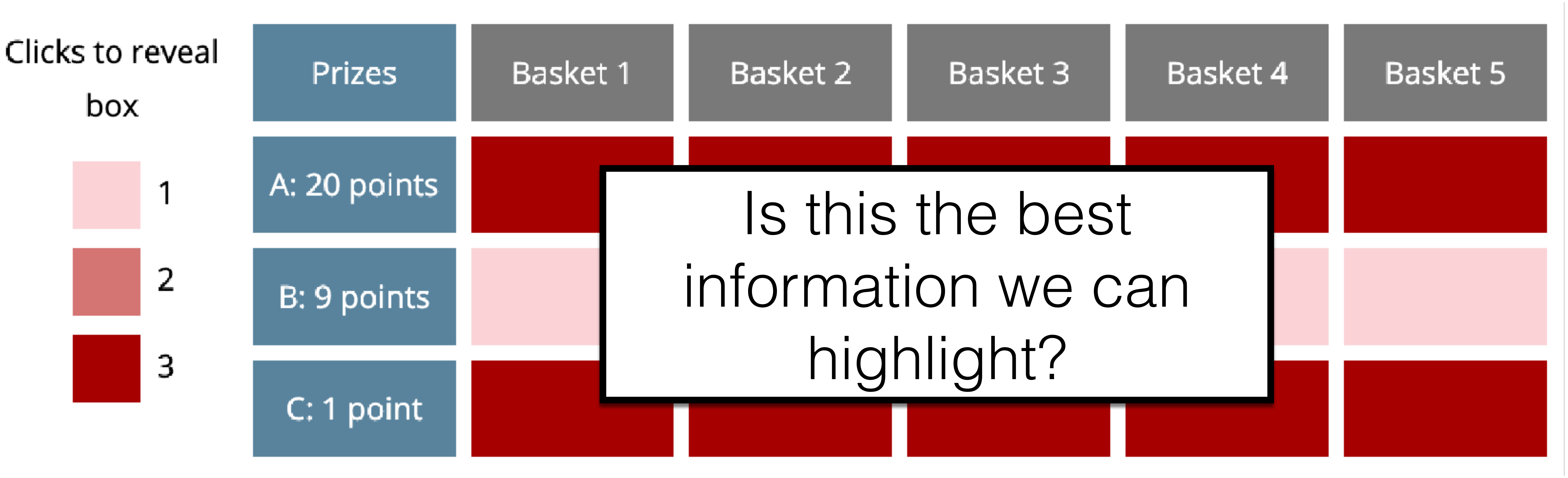
Prizes	Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
A: 20 points					
B: 9 points					
C: 1 point					

reduce cost of computations for one feature

Results: Most effective for moderate preferences



Experiment: Traffic light labeling in Mouselab



Model: Optimal nudging

Choice architect

Knows true feature values

Chooses modified meta MDP

Decision maker

Knows their preferences

Decides with modified meta MDP

Model: Optimal nudging

architect knowledge
(distribution over states)

optimal nudge
(modified meta MDP)

possible
world state

$$\tilde{M}^* = \underset{\tilde{M} \in \tilde{\mathcal{M}}}{\operatorname{argmax}} \, \mathrm{E}\left[g(\tilde{M}, s) \,\middle|\, s \sim b_{\mathrm{arch}} \right]$$

possible nudges

goal of nudge
(objective function)

Experiment: Optimal nudging

Prize values

Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
4	5	3	7	2
7	3	7	6	7
6	7	4	5	6
7	7	5	3	6
5	3	3	6	7

Original choice architecture

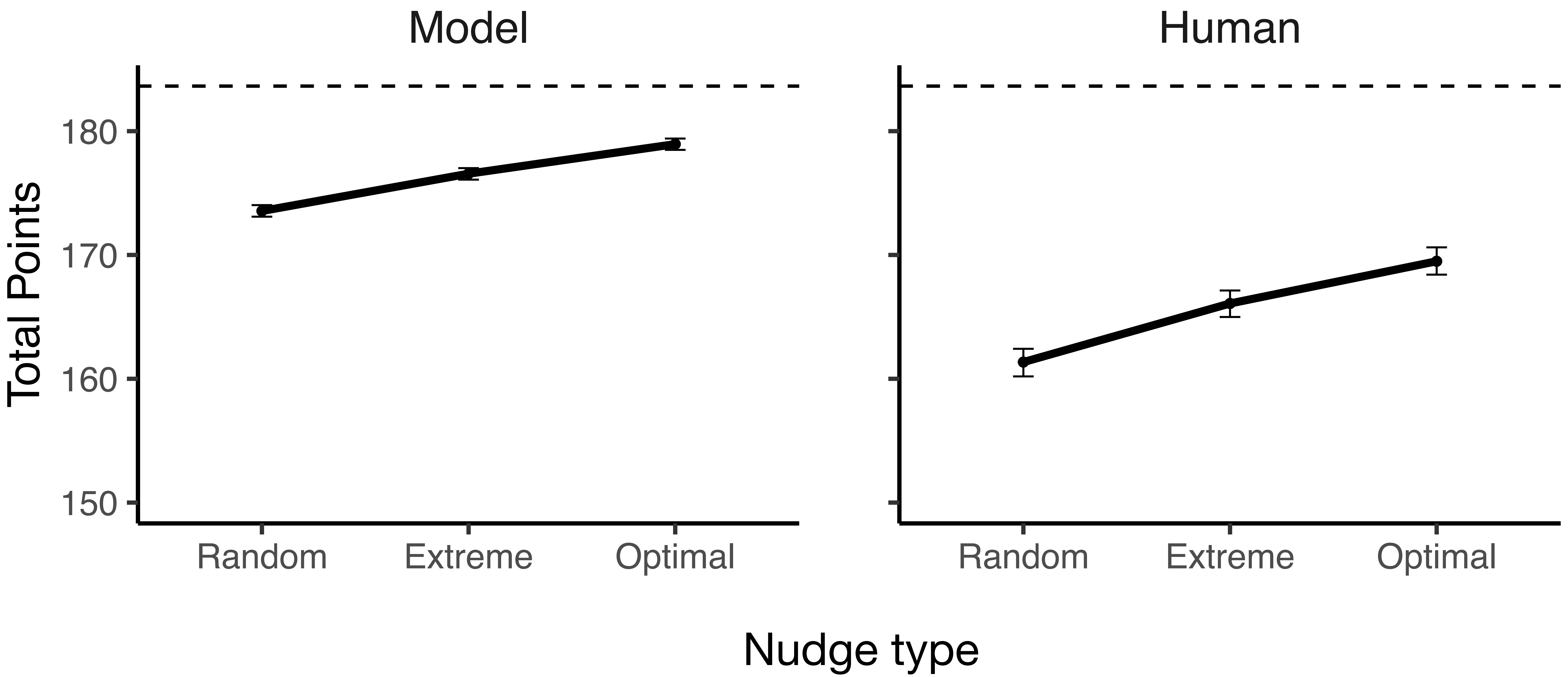
Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
		3		
				7
				7

optimal nudging

Optimal choice architecture

Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
		3		2
				7
6				
7				
				7

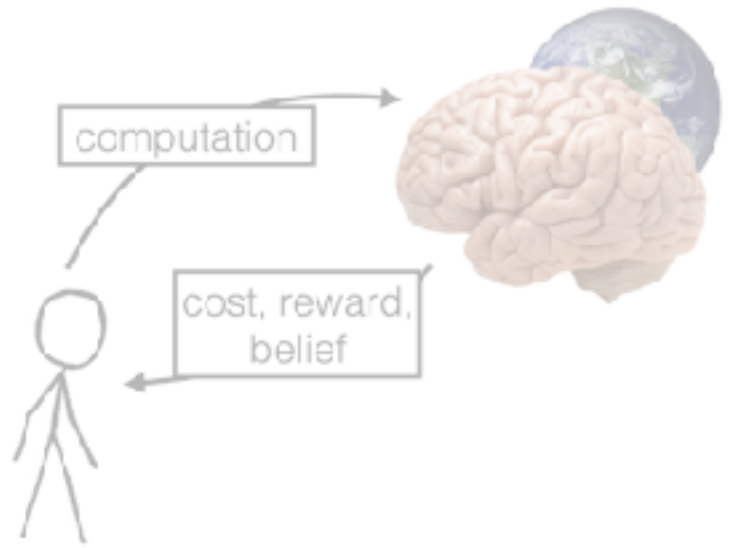
Results: Optimal nudges improve decisions



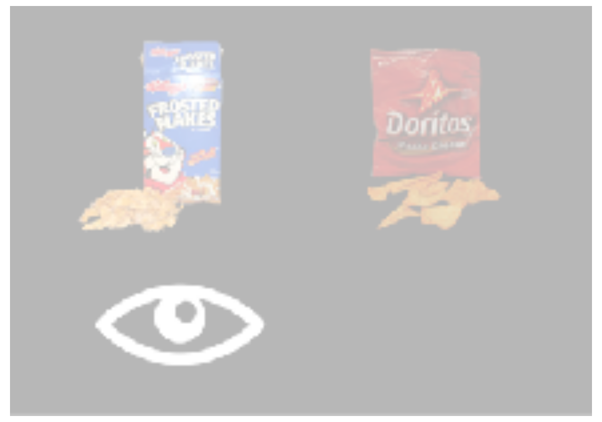
Summary: Predicting and nudging complex choices

- Multi-attribute decision problems can be modeled as a meta MDP where an agent sequentially considers features of each option.
- The optimal policy for that meta MDP depends on one's prior beliefs as well as the cost of considering different features.
- Modifying the meta MDP changes which features a rational agent considers, leading to predictable changes in behavior.
- This allows us to construct *optimal nudges*, changes to the meta MDP that maximized a desired outcome.

Metalevel MDPs



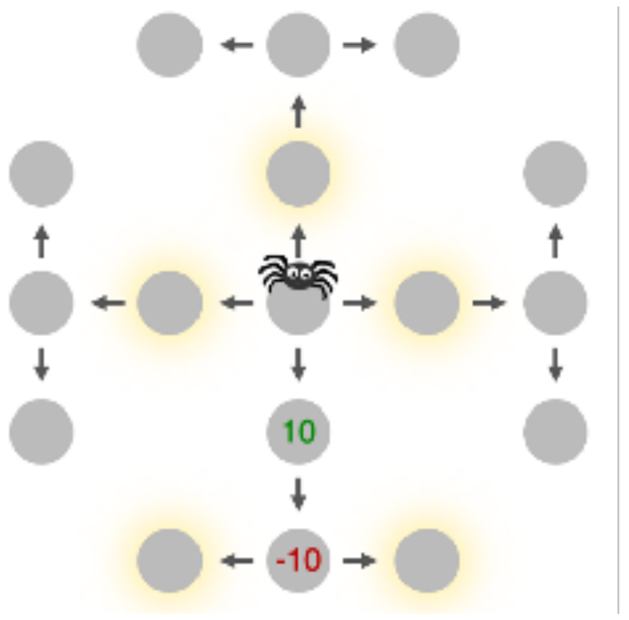
Simple decisions



Multi-attribute decisions

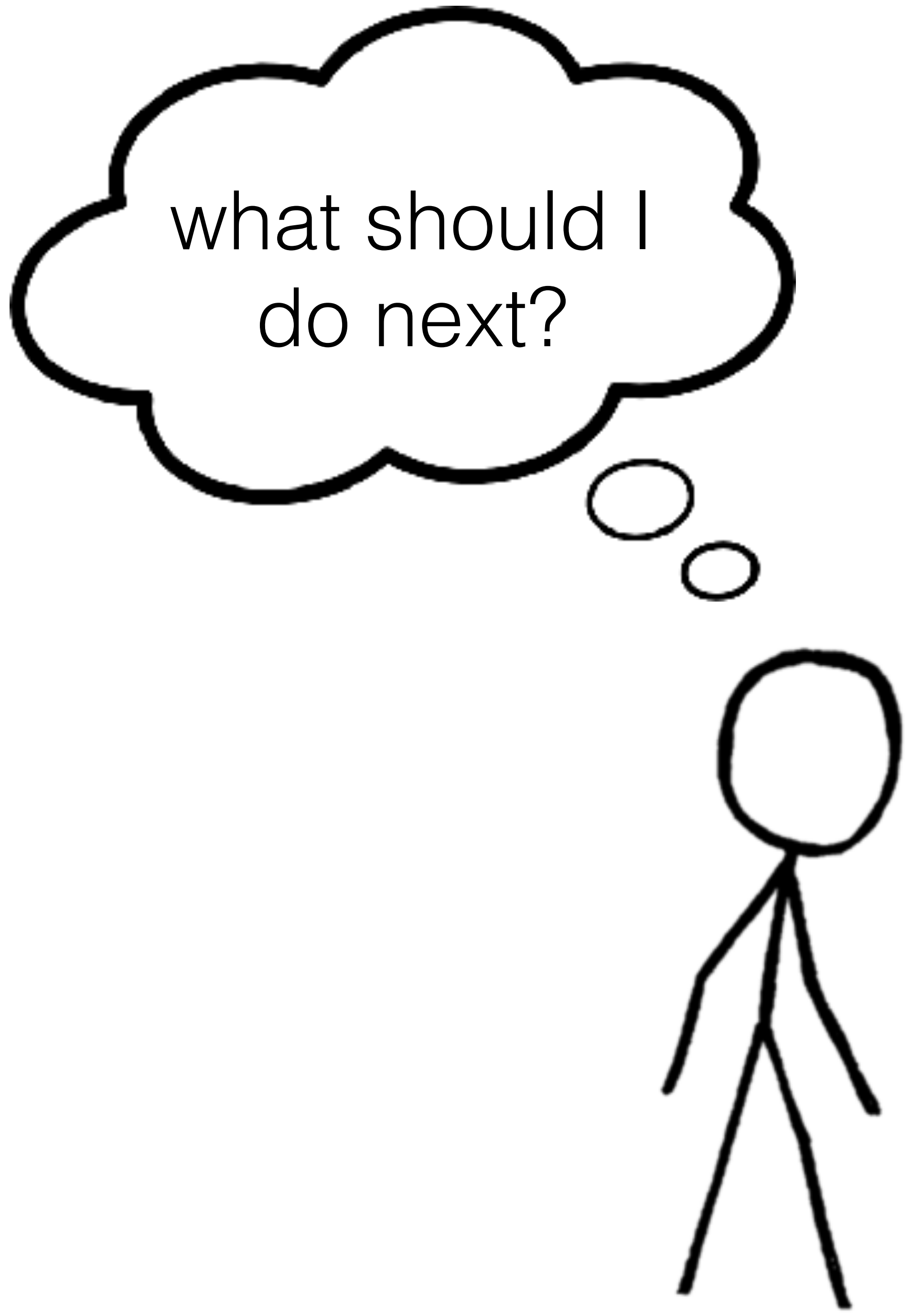
Prizes	Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
A: 3 points	2		3	4	
B: 2 points	7				7
C: 2 points	7	4		2	
D: 21 points	7		8	6	
E: 2 points	9				6

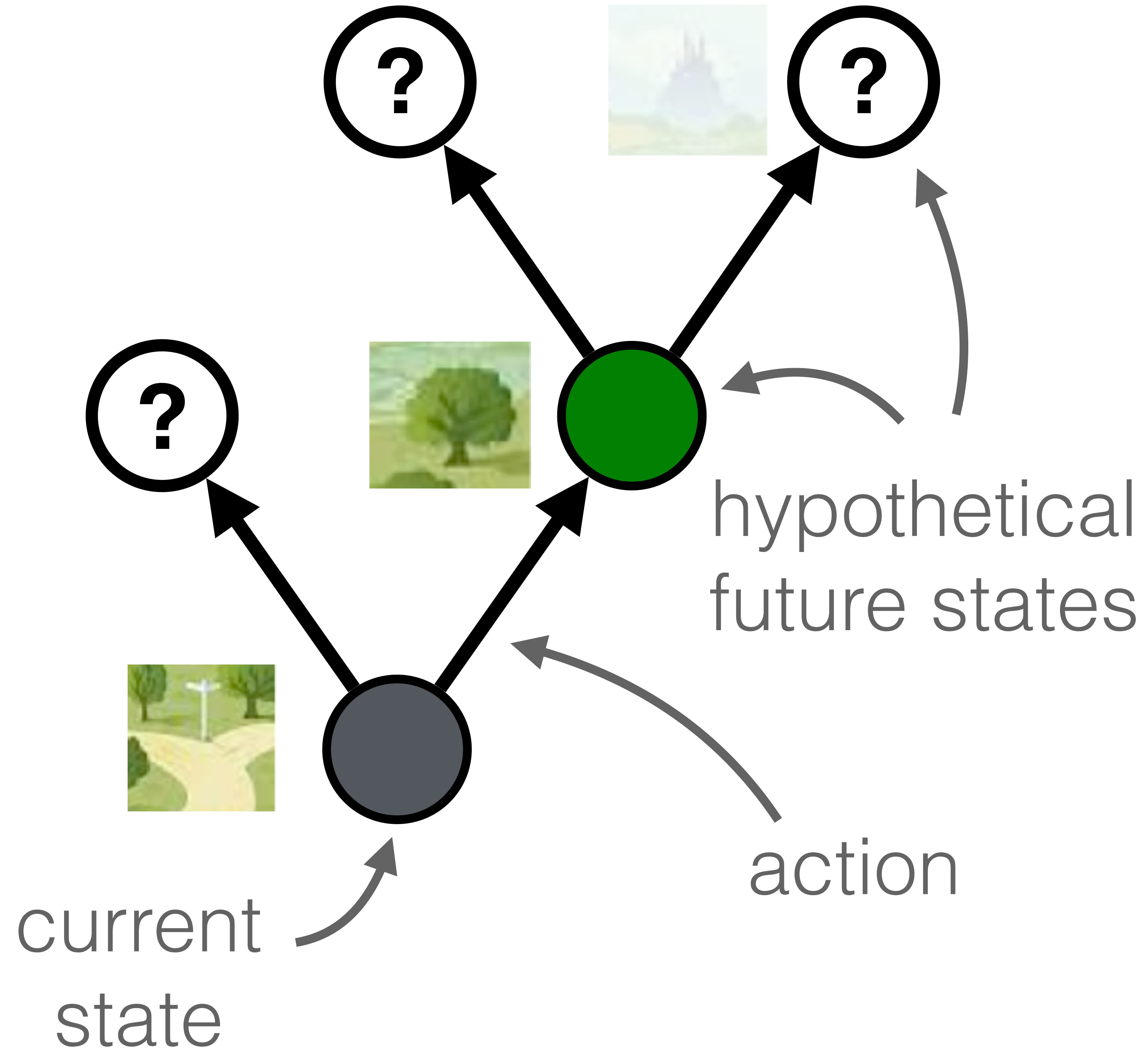
Sequential decisions



Background: Planning as decision-tree search

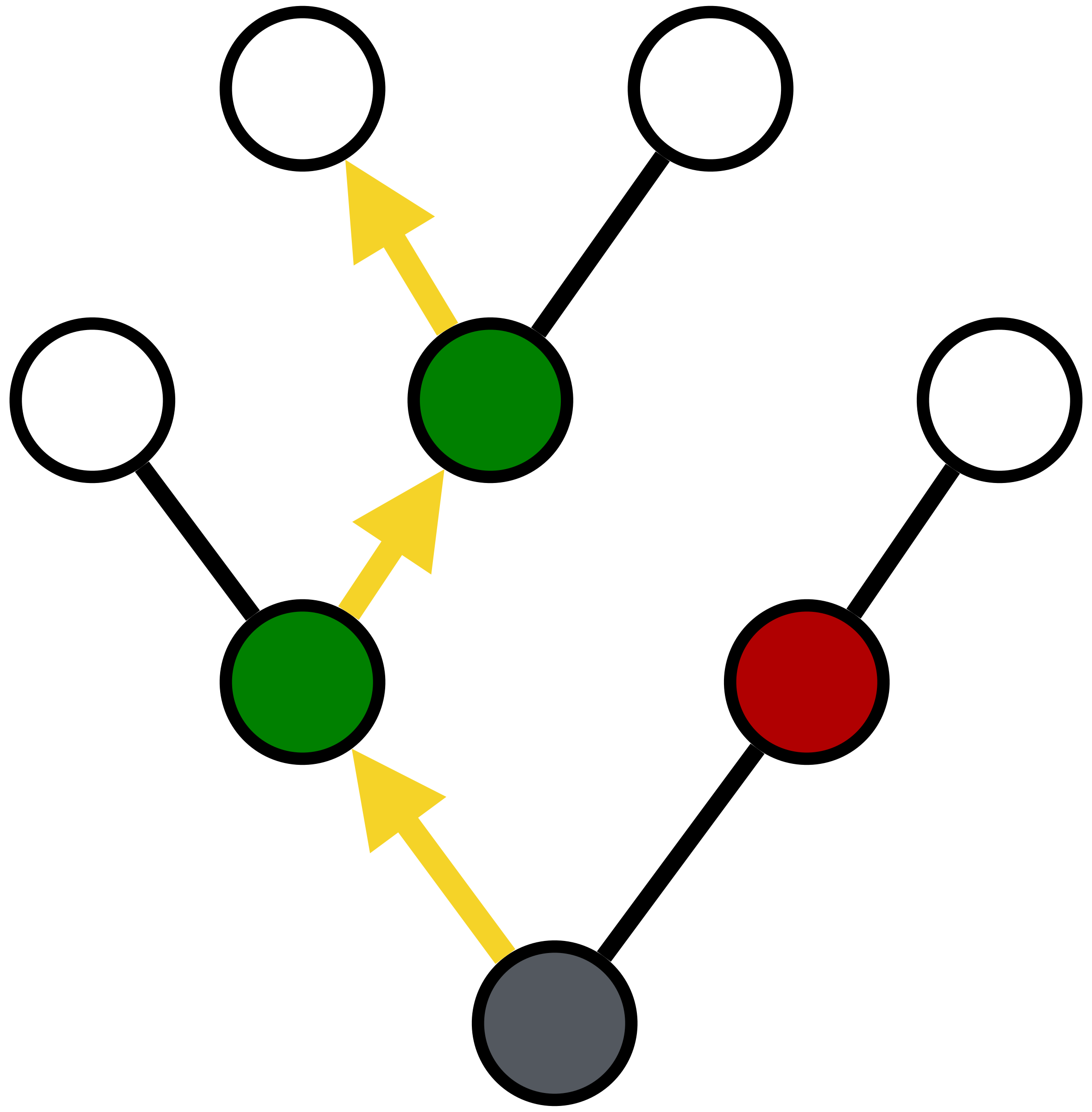
Which future state should you think about next?



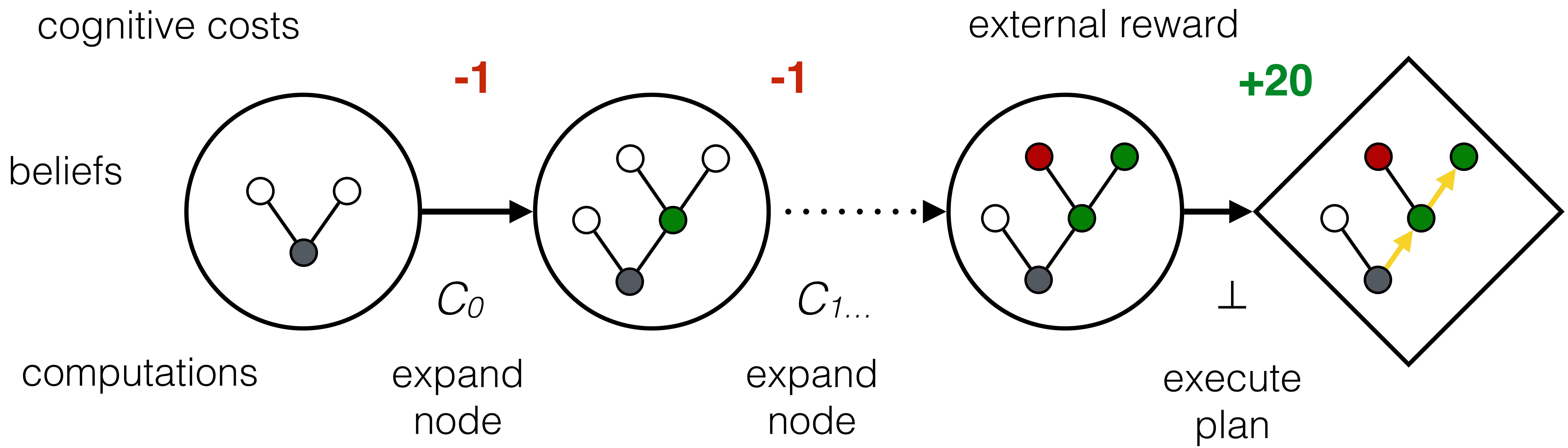


Callaway, van Opheusden, Gul, Das, Krueger, Griffiths, Lieder (2022)

Model: Decision-tree search



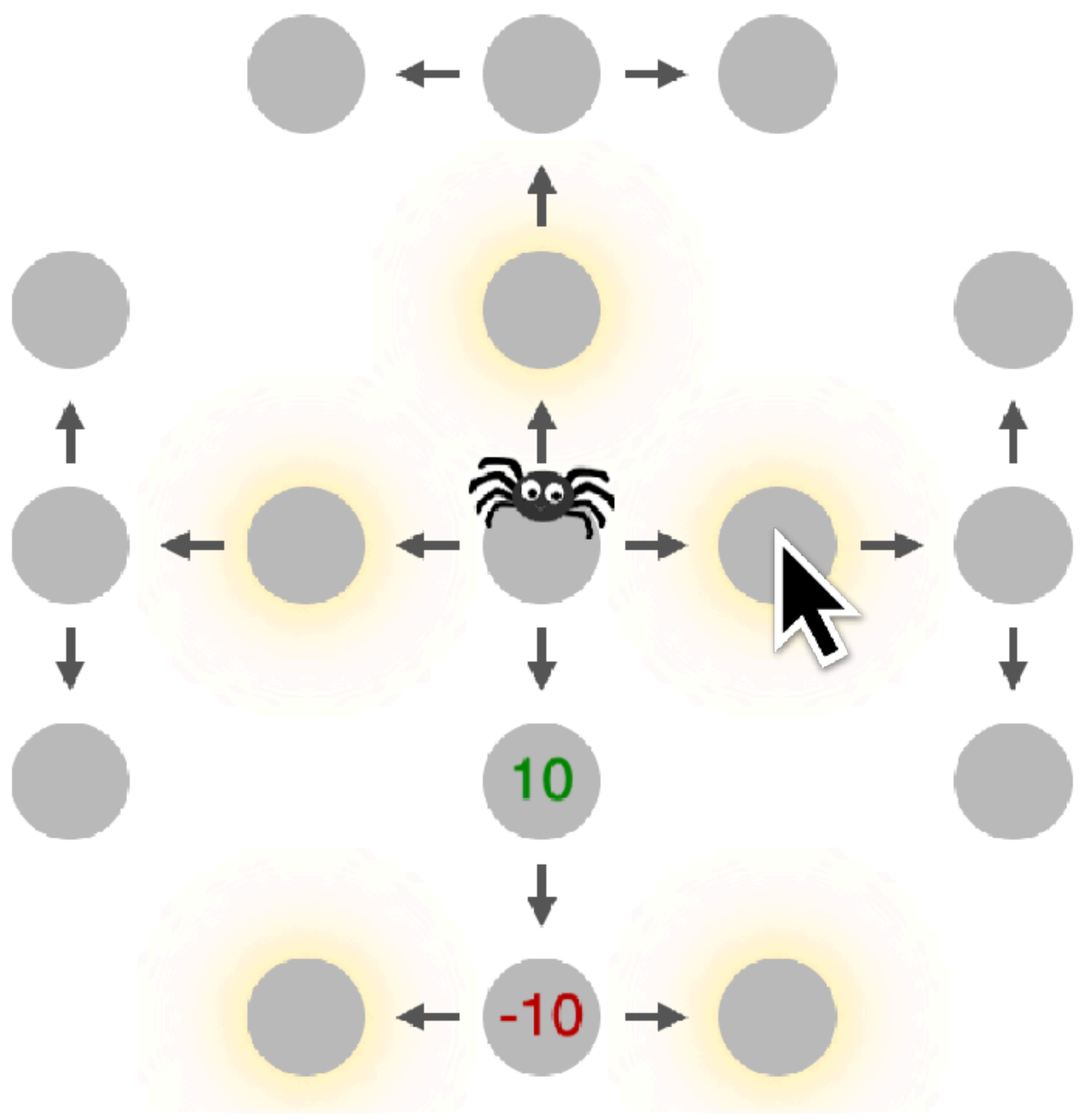
Model: Meta MDP



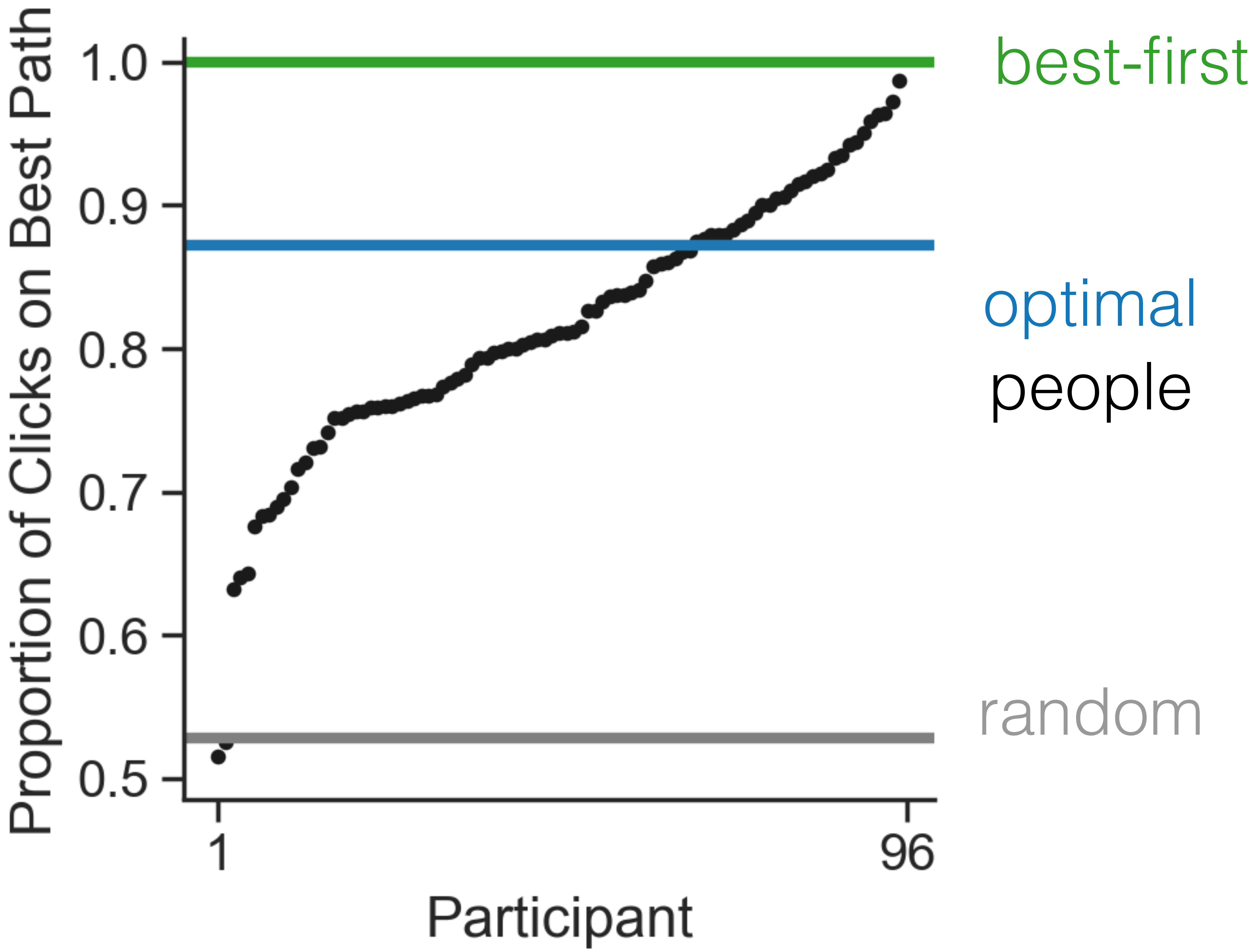
Experiment: Mouselab-MDP

- Route-planning problem: maximize total reward over three steps.
- Rewards are initially occluded, revealed by clicking.
- Extends the Mouselab paradigm to planning problems.

  Payne et al. (1988)

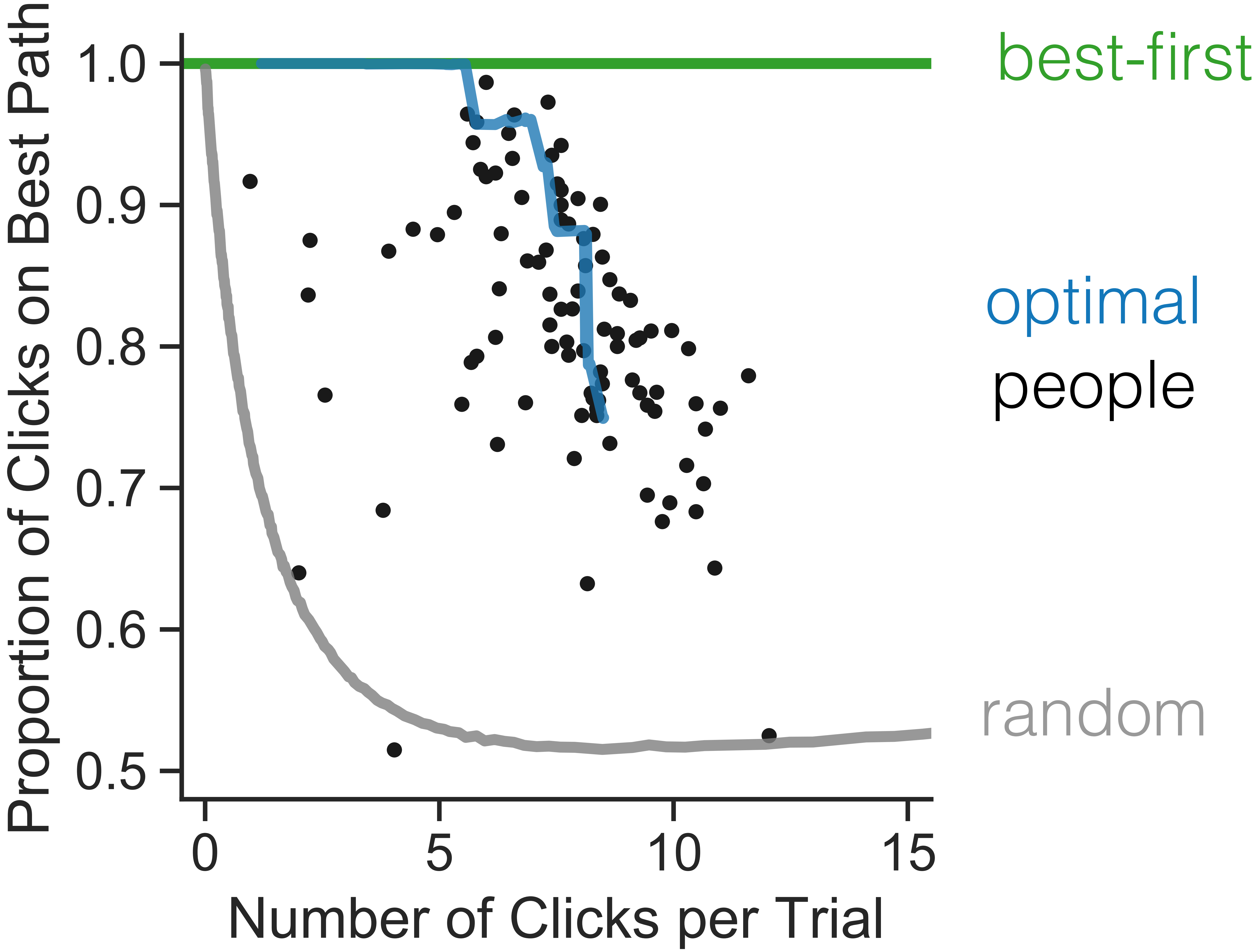


Results: Best-first search is optimal

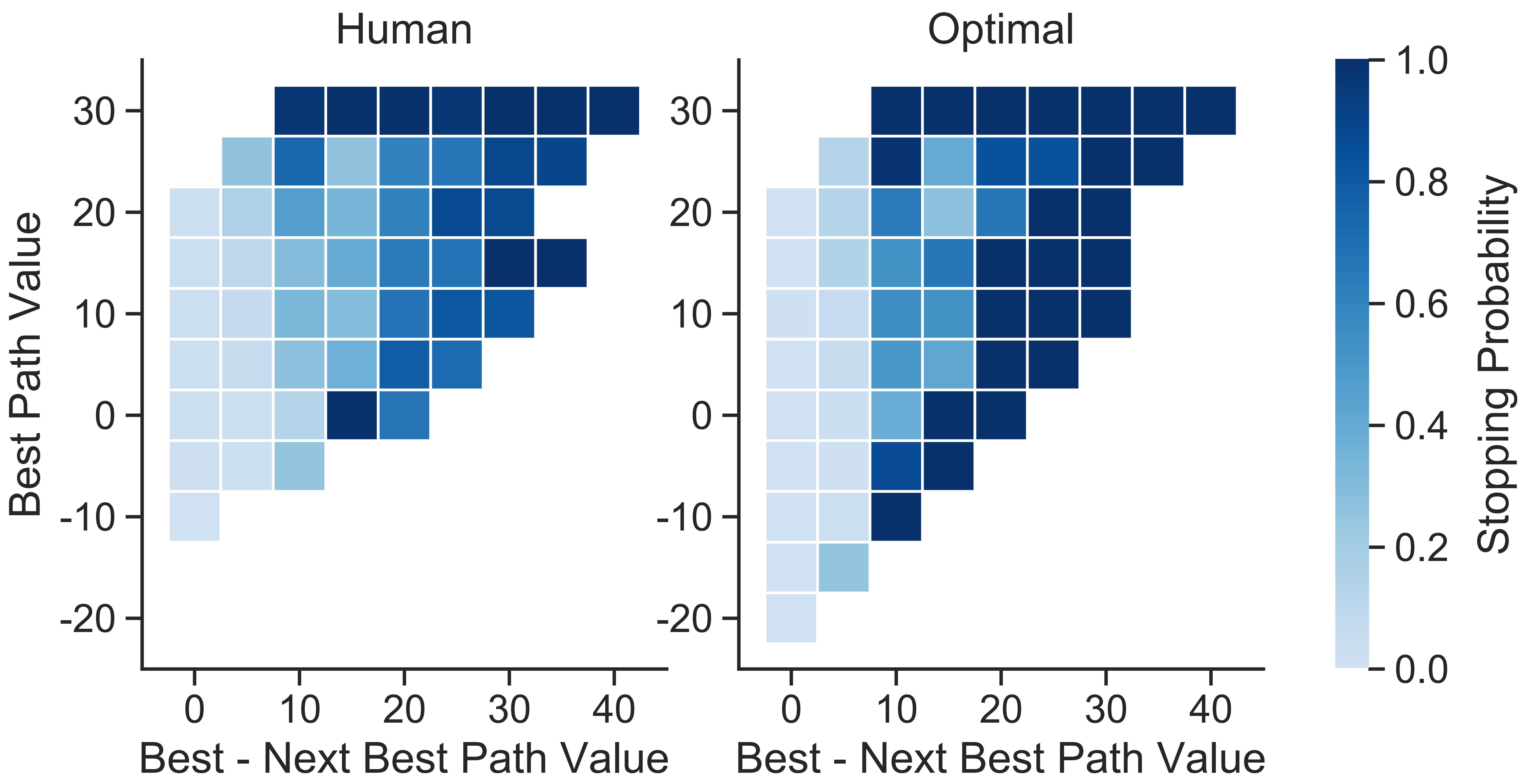


(depending on the cost)

Results: Best-first search is optimal



Results: Relative and absolute stopping rule

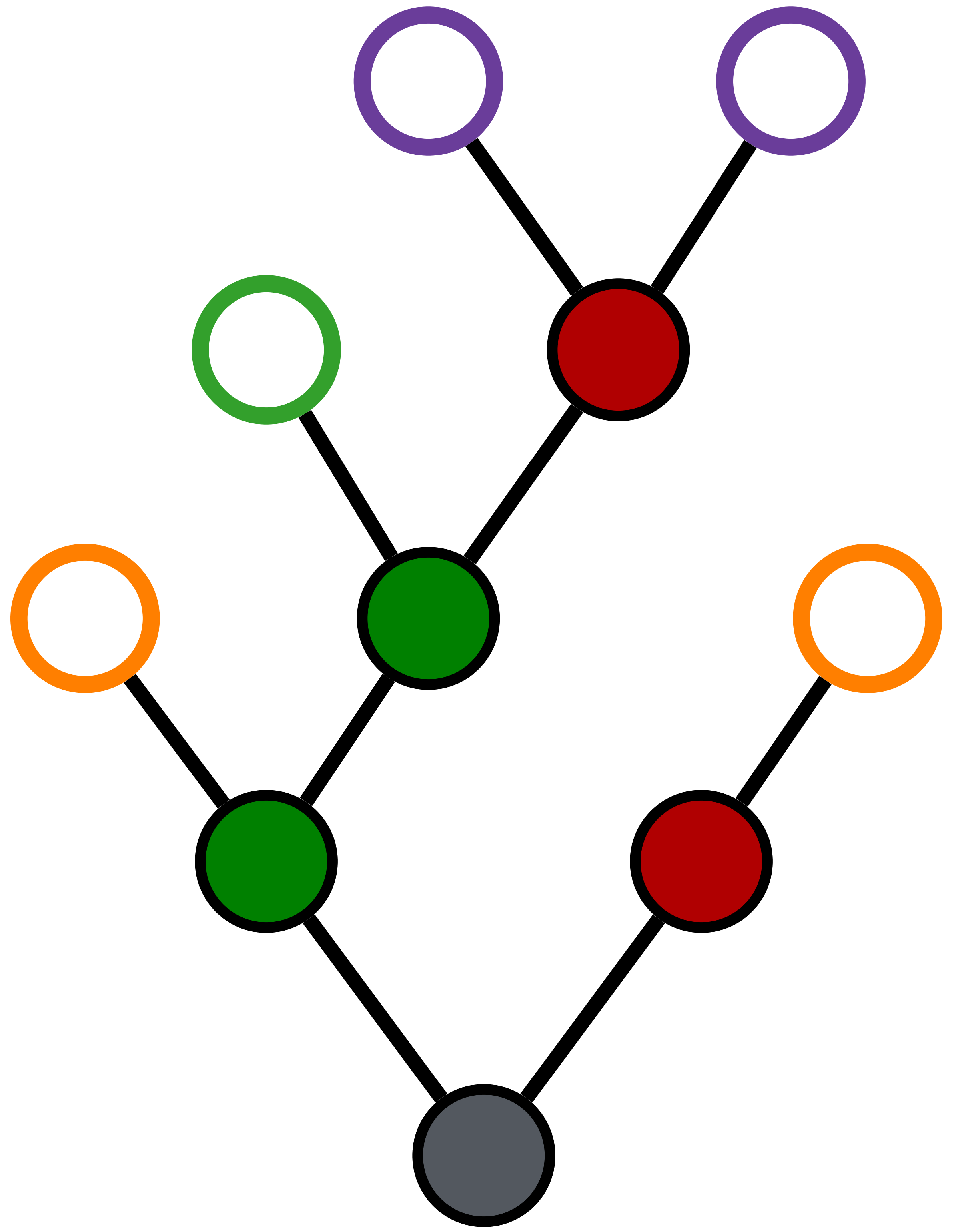


Model: Alternative search strategies

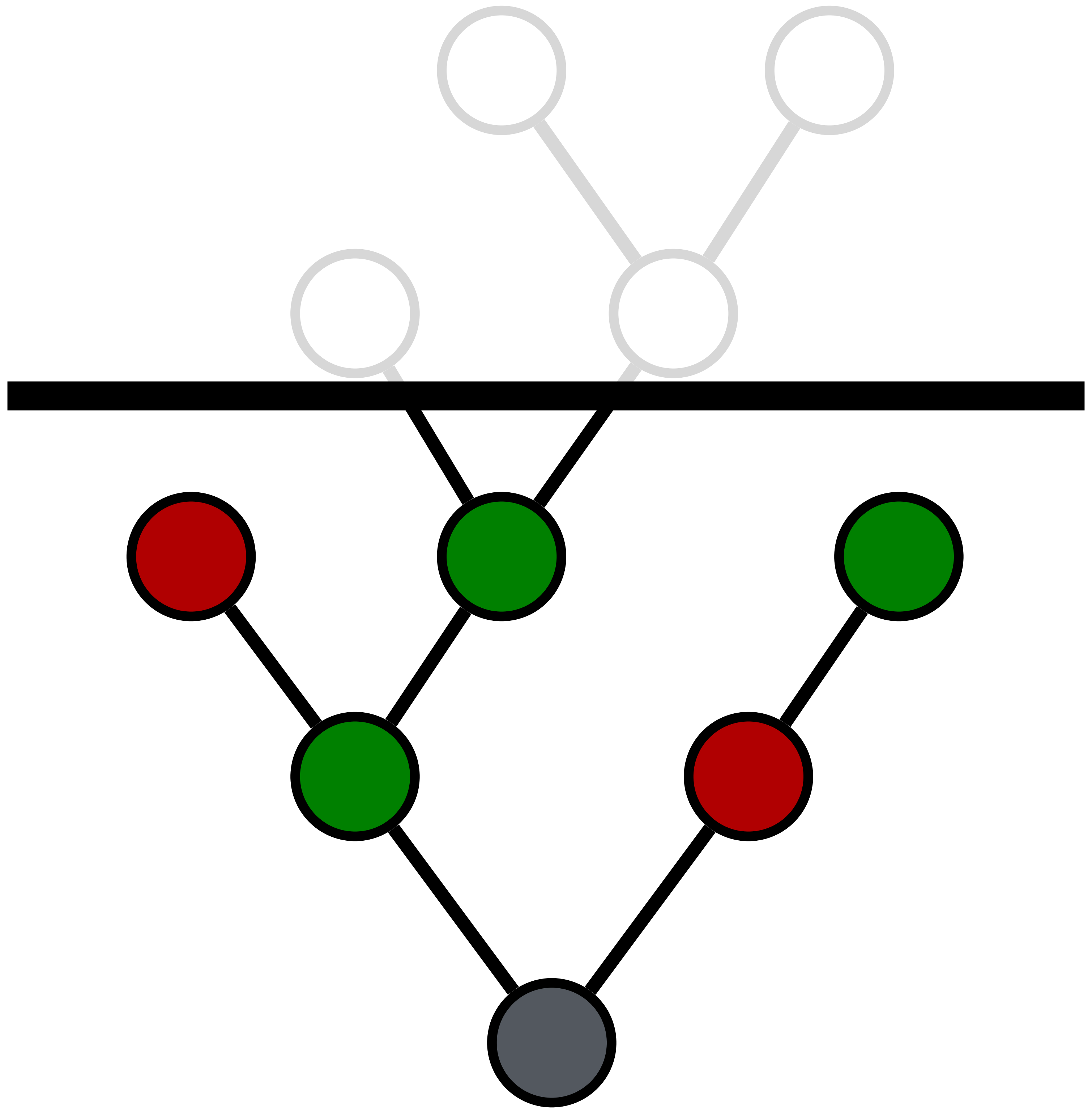
Best-First search expands nodes on high value paths

Depth-First search expands nodes that are far from the root

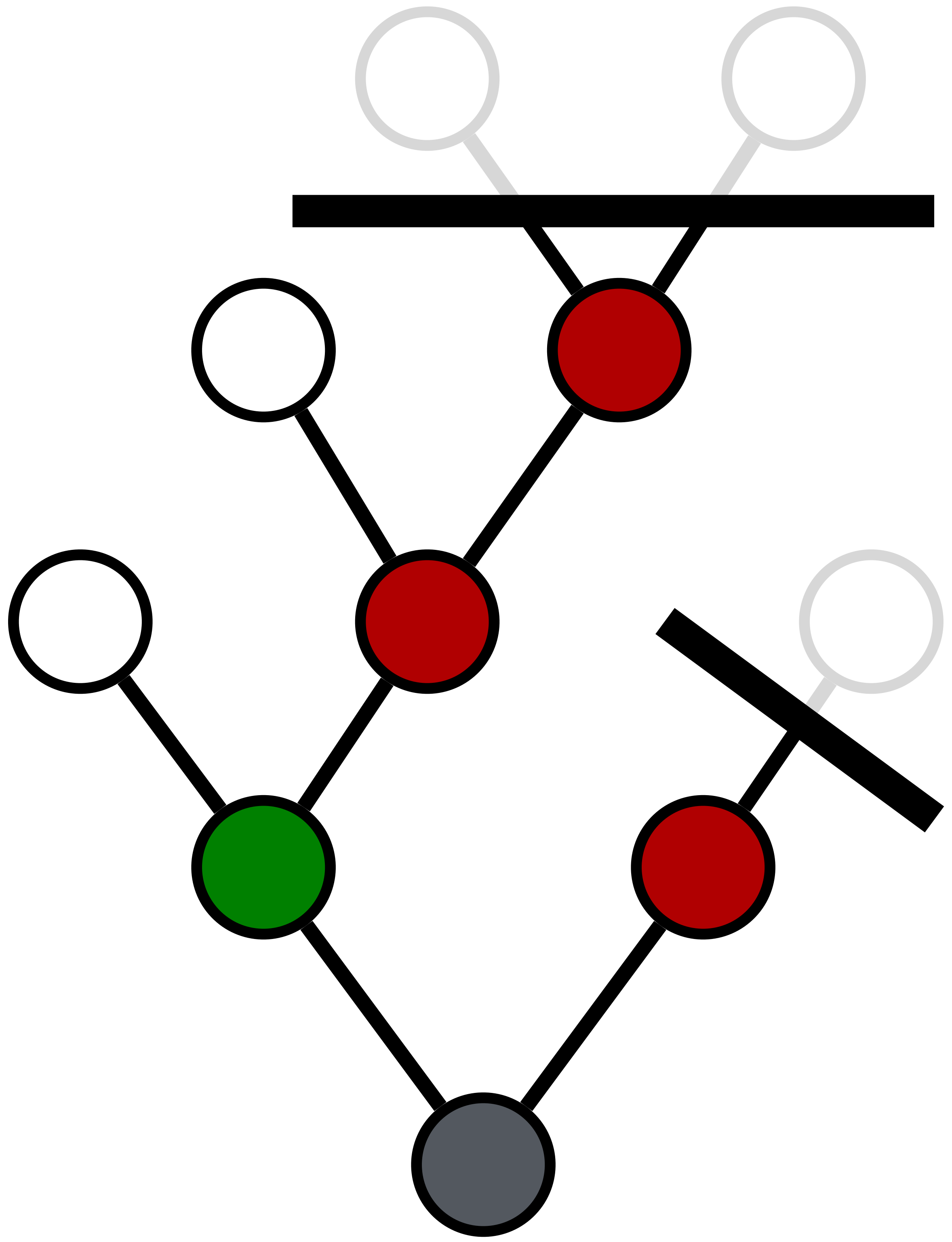
Breadth-First search expands nodes that are close to the root



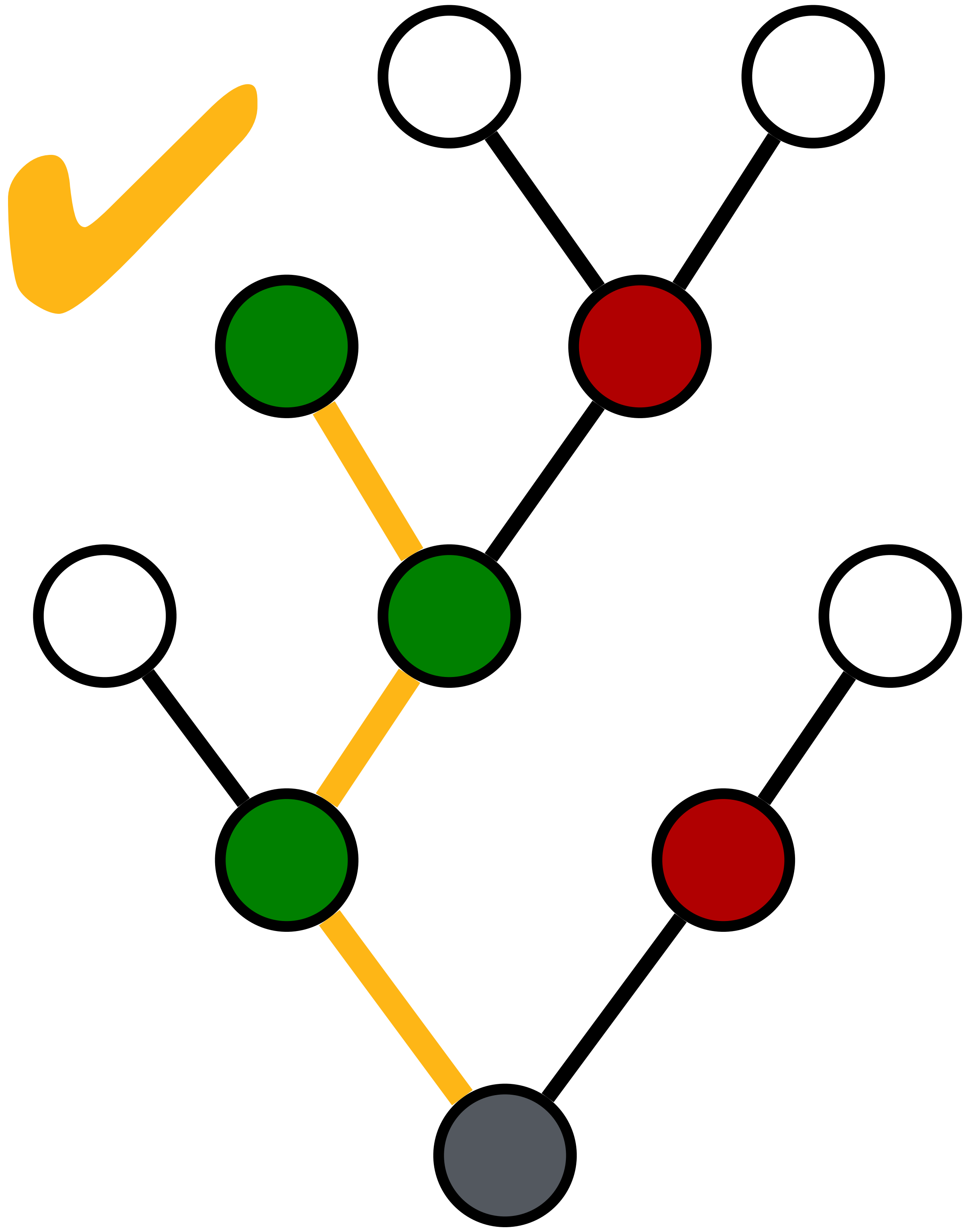
Model: Heuristics



depth limit

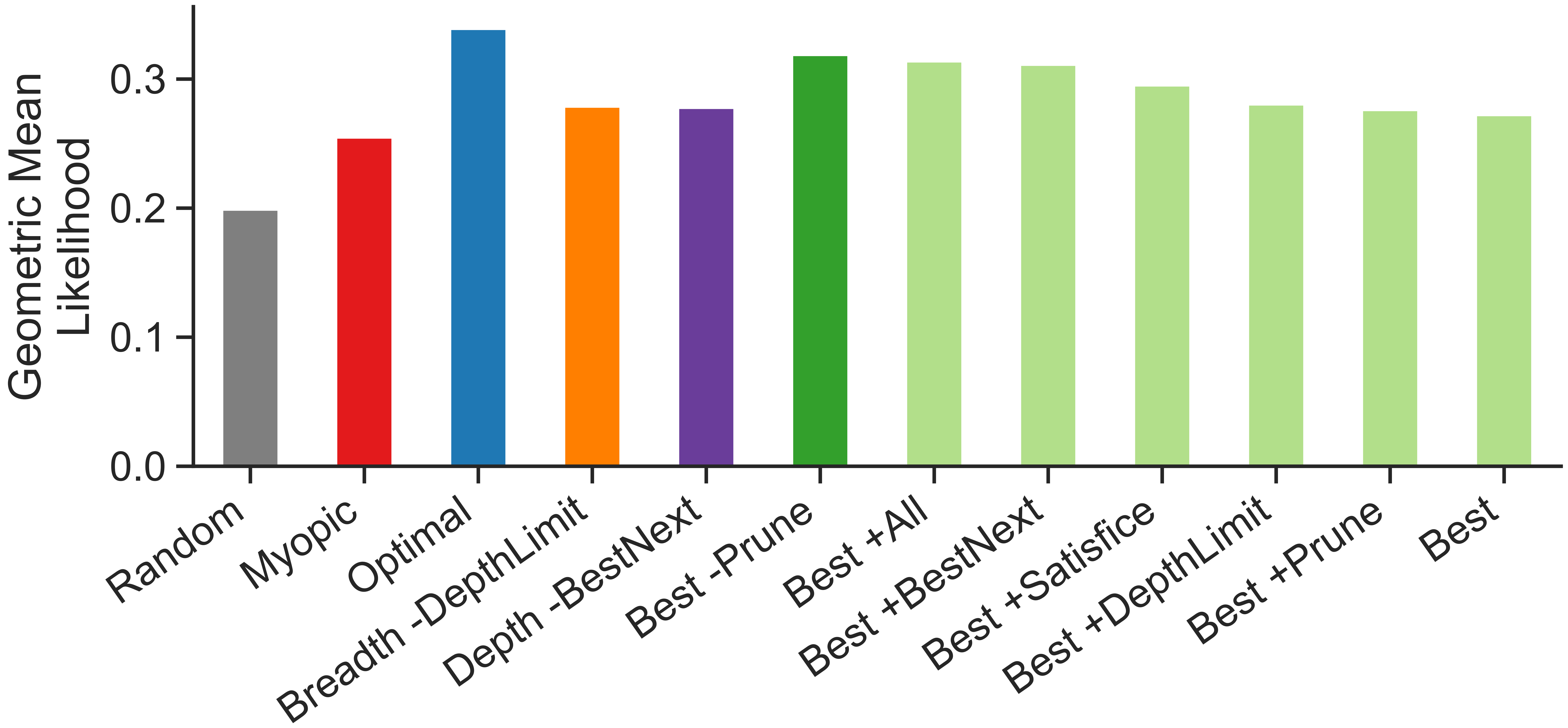


pruning

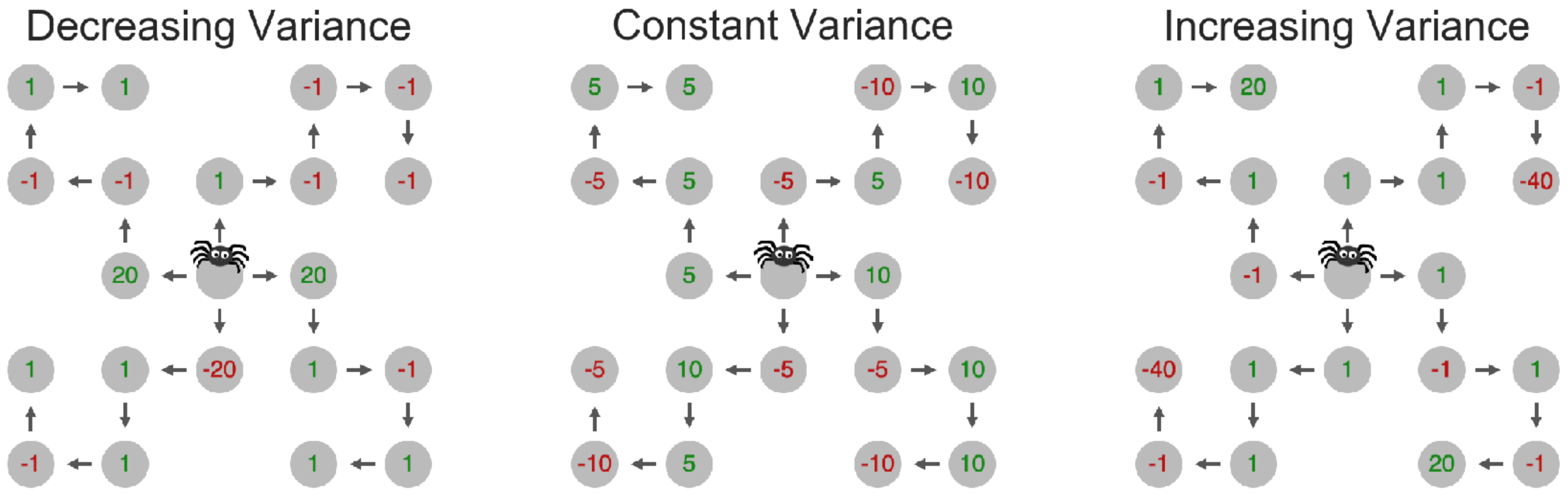


satisficing

Results: Model comparison

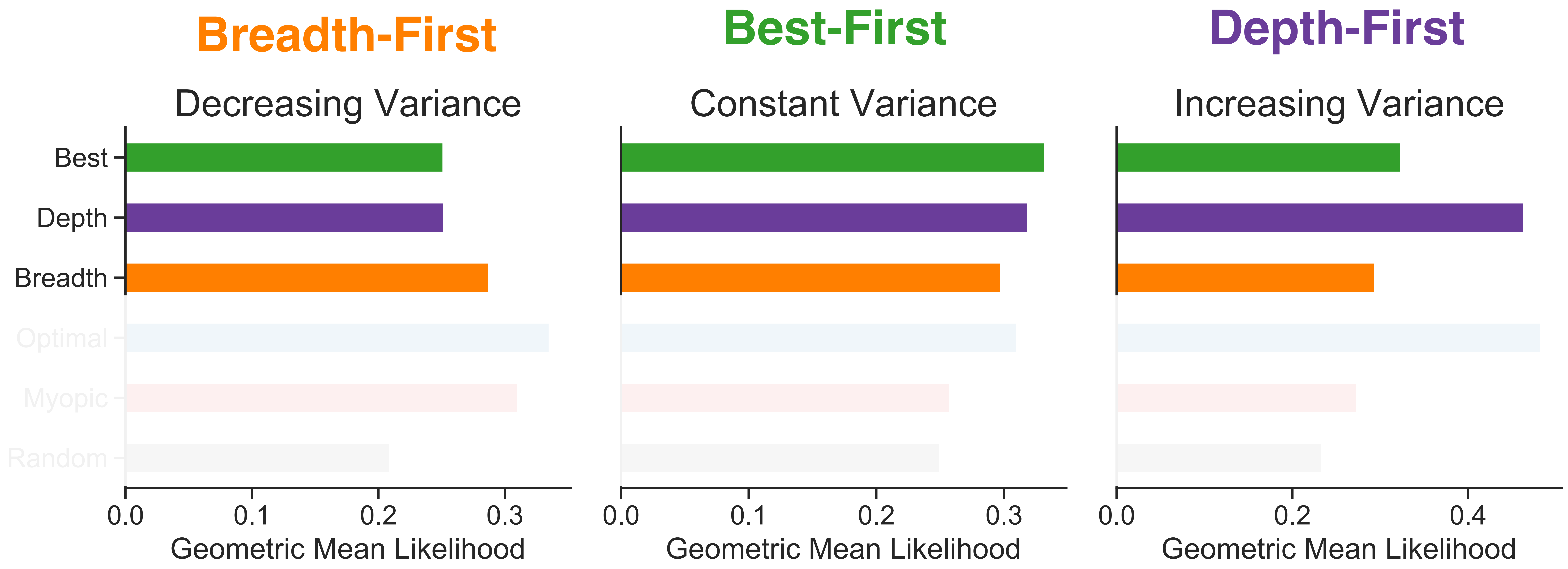


Experiment: Adapting to the environment



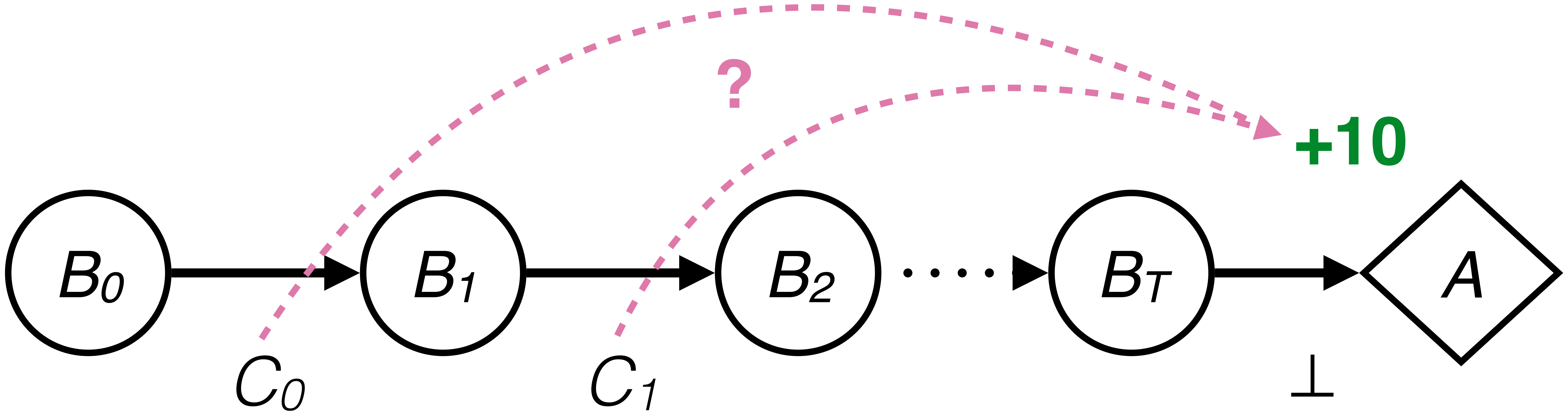
Breadth-First **Best-First** **Depth-First**

Results: Adapting to the environment



Application: Teaching efficient planning strategies

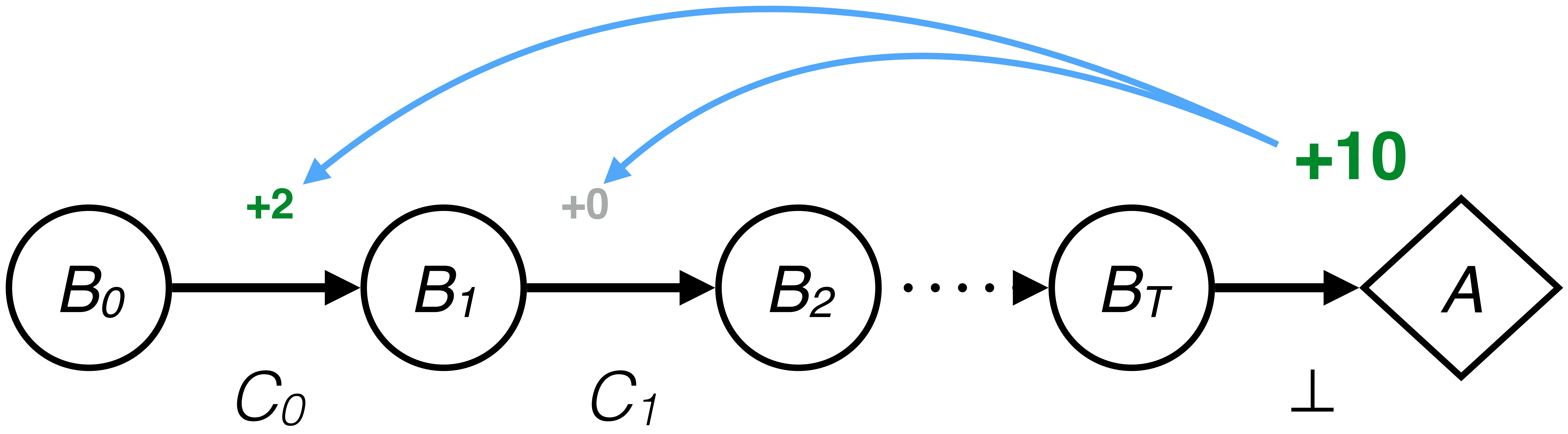
- **Challenge:** Learning strategies is hard because of the *temporal credit assignment problem*: which computations contributed to making a good decision?



Callaway, Jain, van Opheusden, Das, Iwama, Gul, Krueger, Becker, Griffiths, Lieder (2022)

Application: Teaching efficient planning strategies

- **Solution:** Use *reward shaping* to make the long-term consequences of thinking immediately salient.



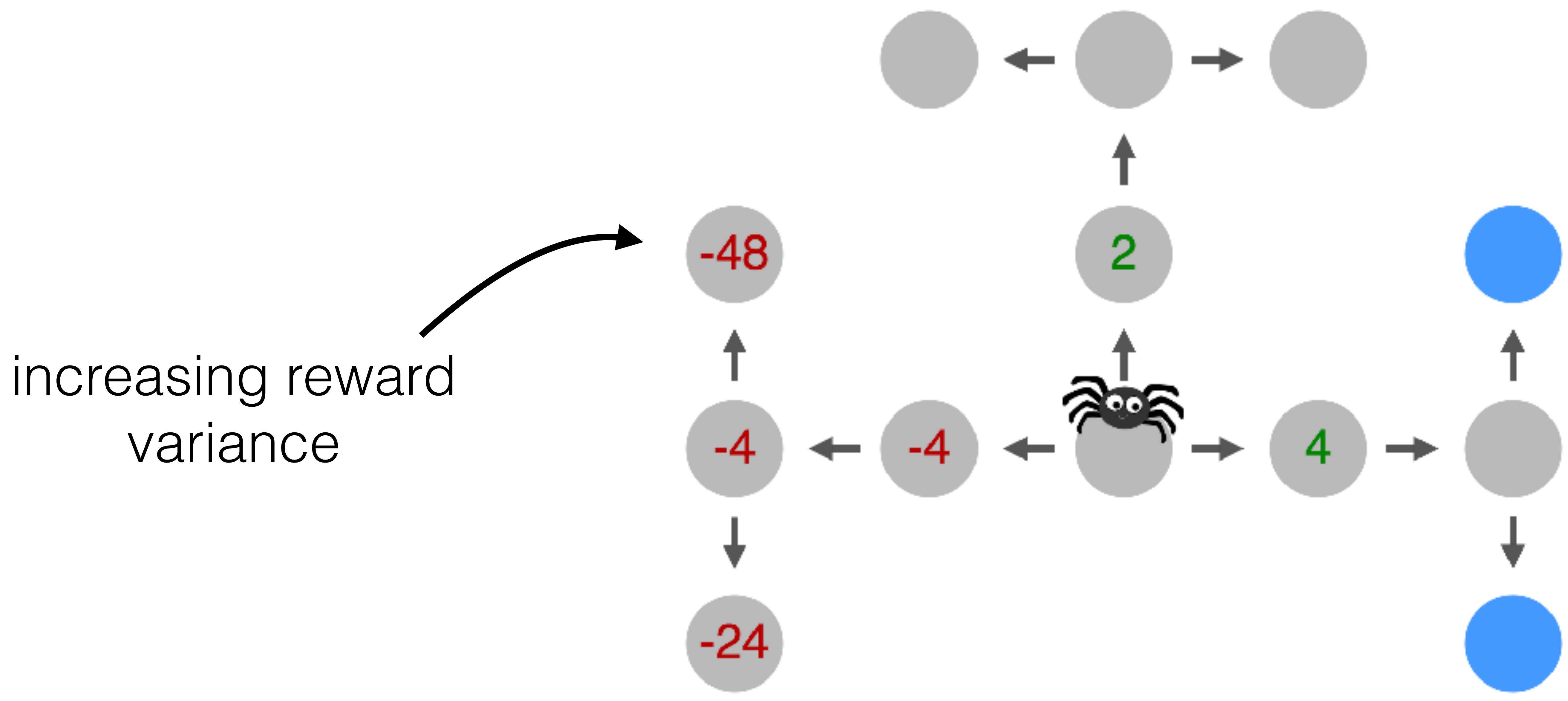
$$\text{loss}(b, c) = \max_{c'} Q_{\text{meta}}(b, c') - Q_{\text{meta}}(b, c)$$

Callaway, Jain, van Opheusden, Das, Iwama, Gul, Krueger, Becker, Griffiths, Lieder (2022)

Experiment: Teaching backward planning

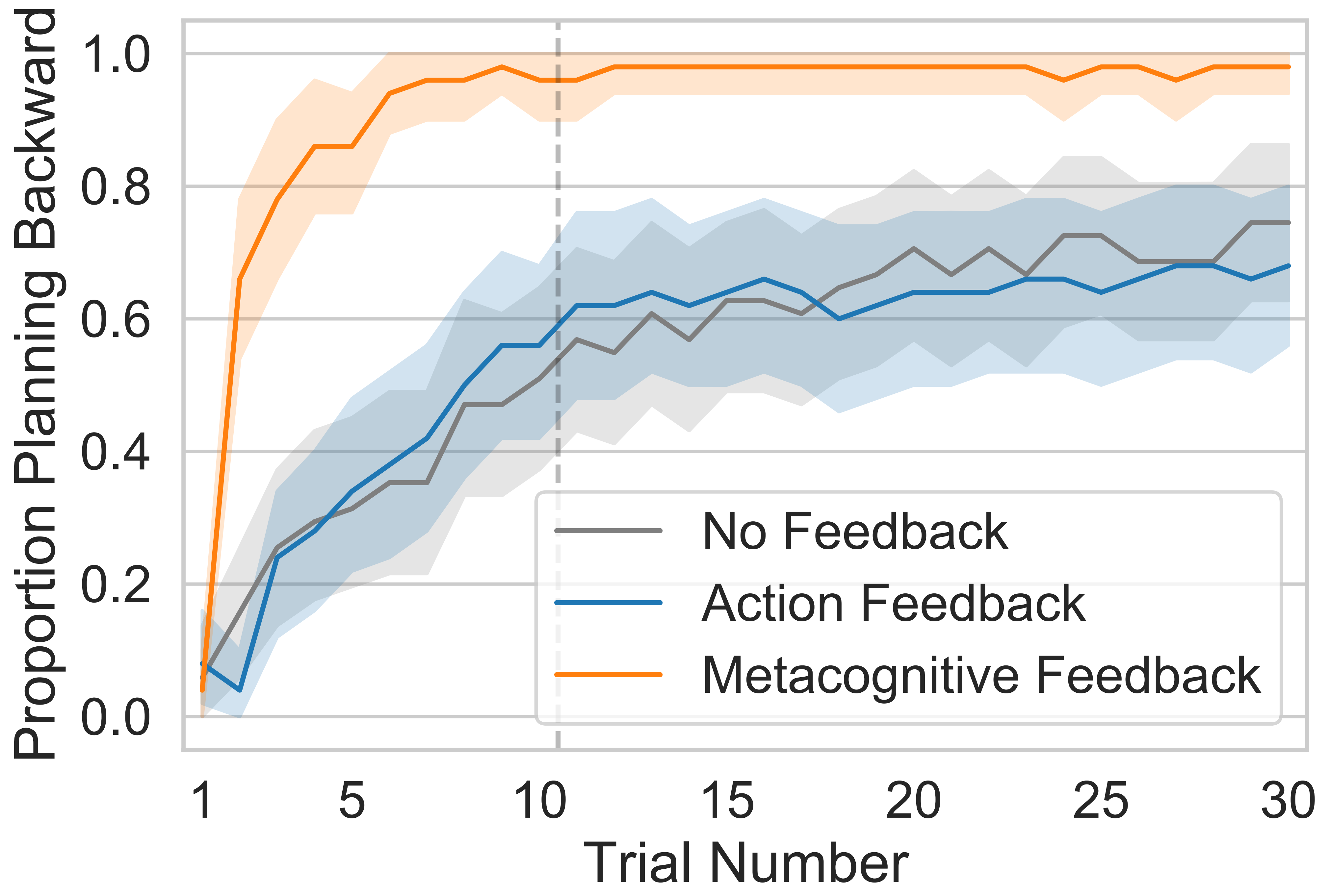
You should have inspected one of the highlighted nodes.

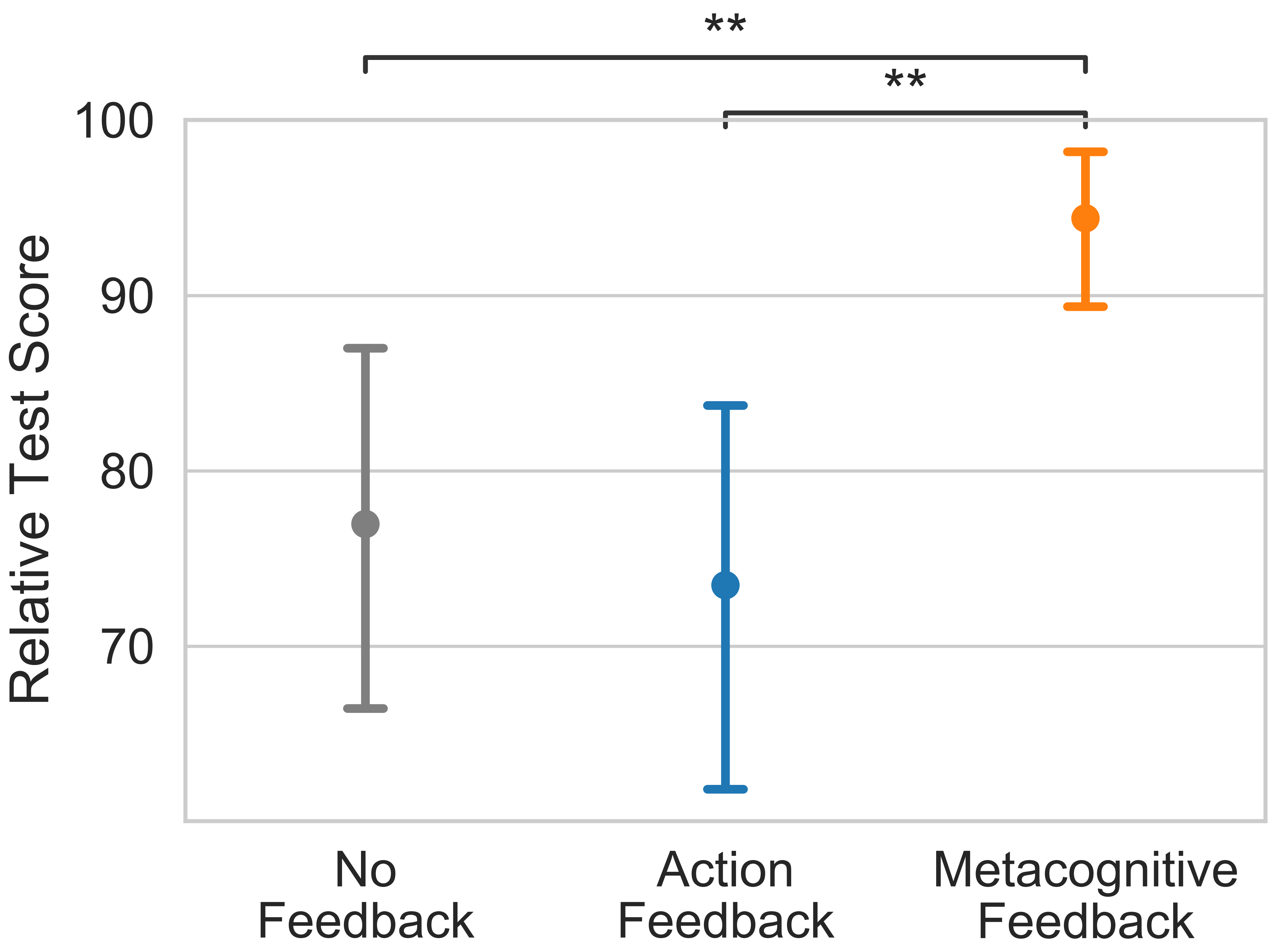
Please wait 7 seconds.



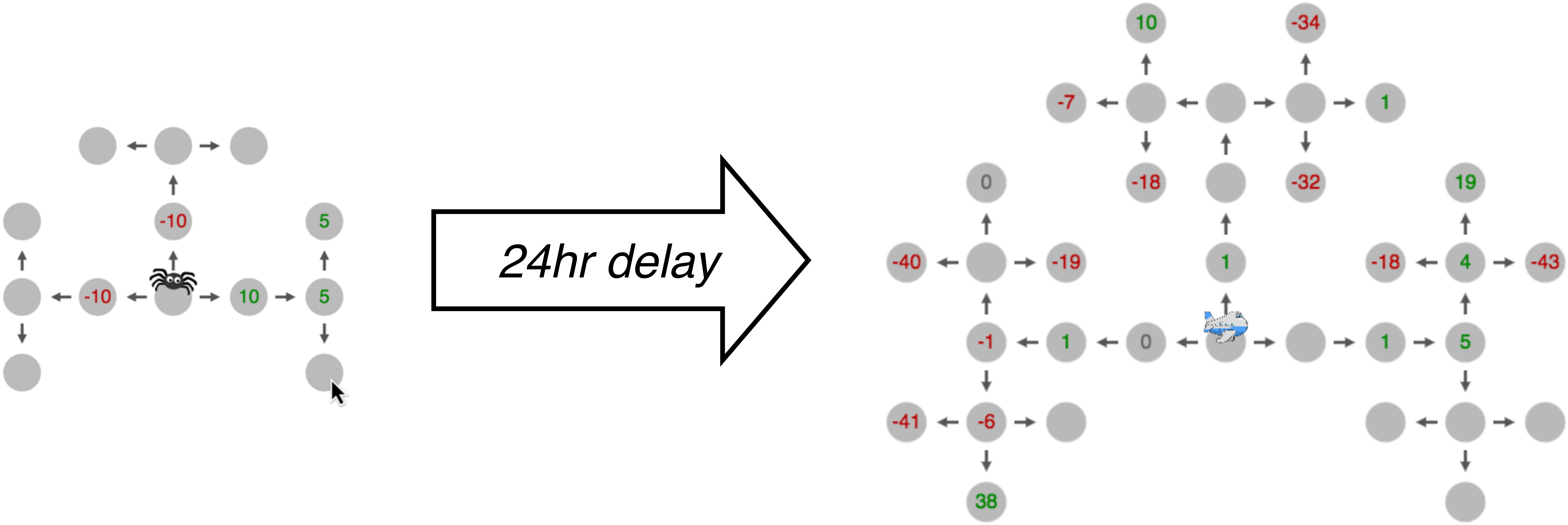
Callaway, Jain, van Opheusden, Das, Iwama, Gul, Krueger, Becker, Griffiths, Lieder (2022)

Results: Metacognitive feedback accelerates learning

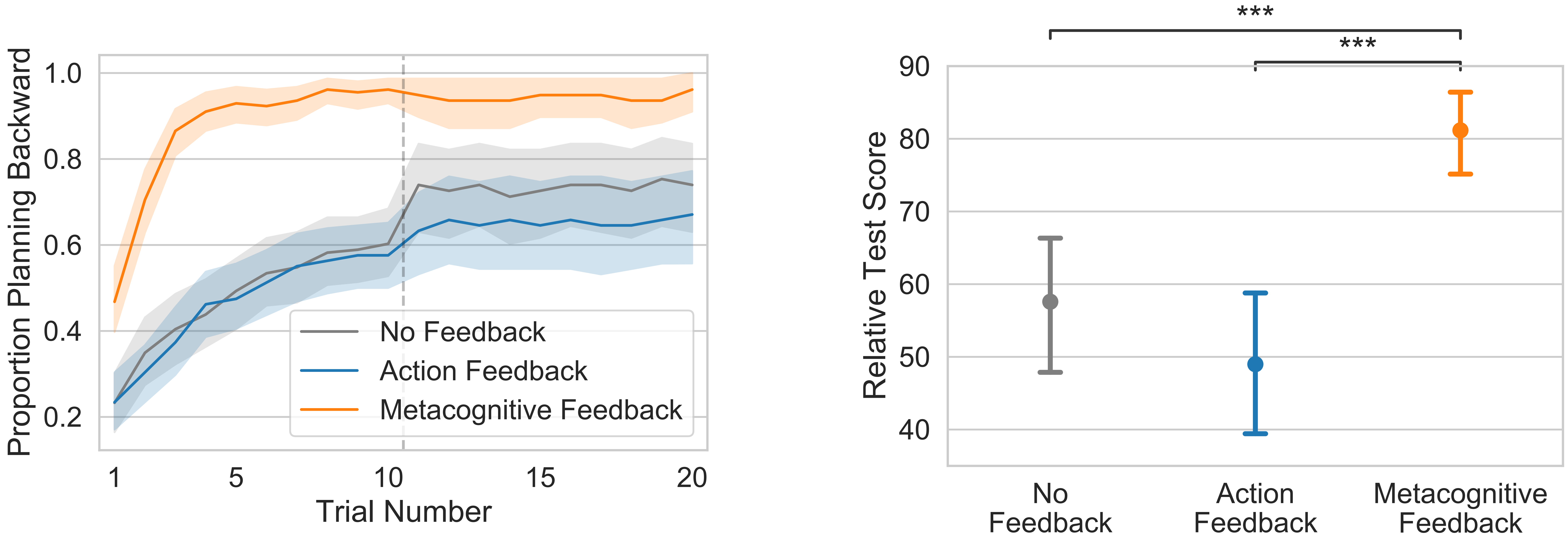




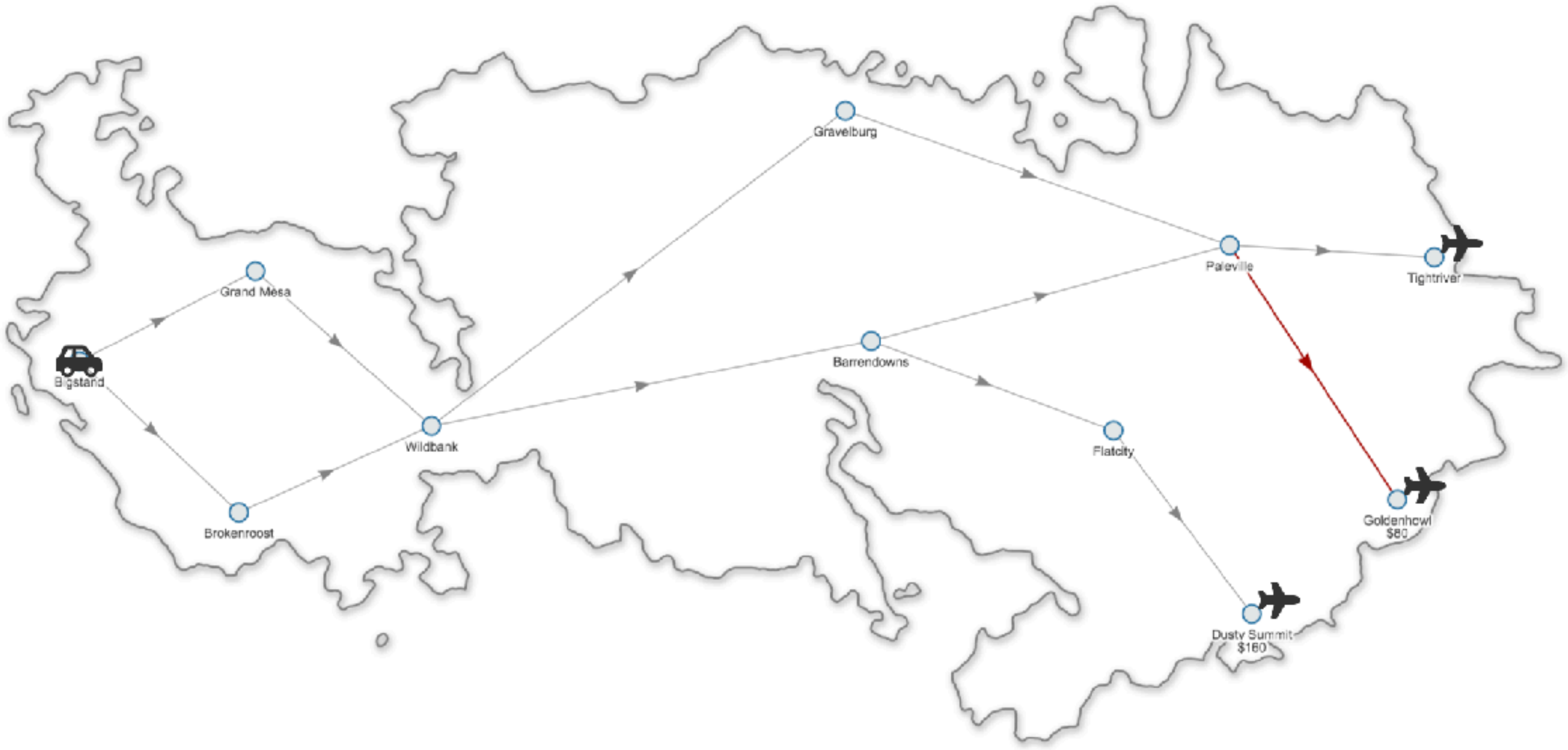
Experiment: Transfer and retention



Results: Strategy retained & applied on bigger problem



Experiment: Far transfer

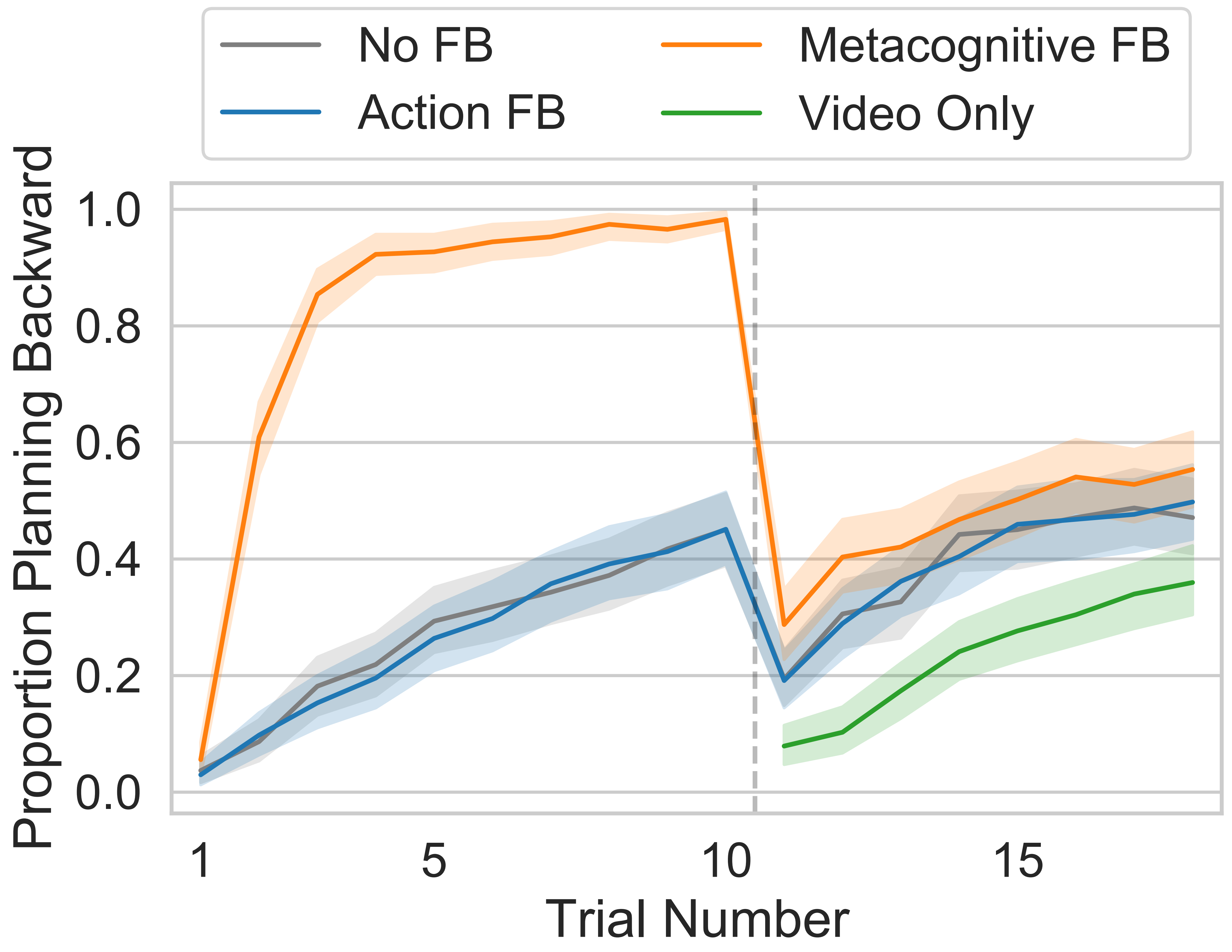


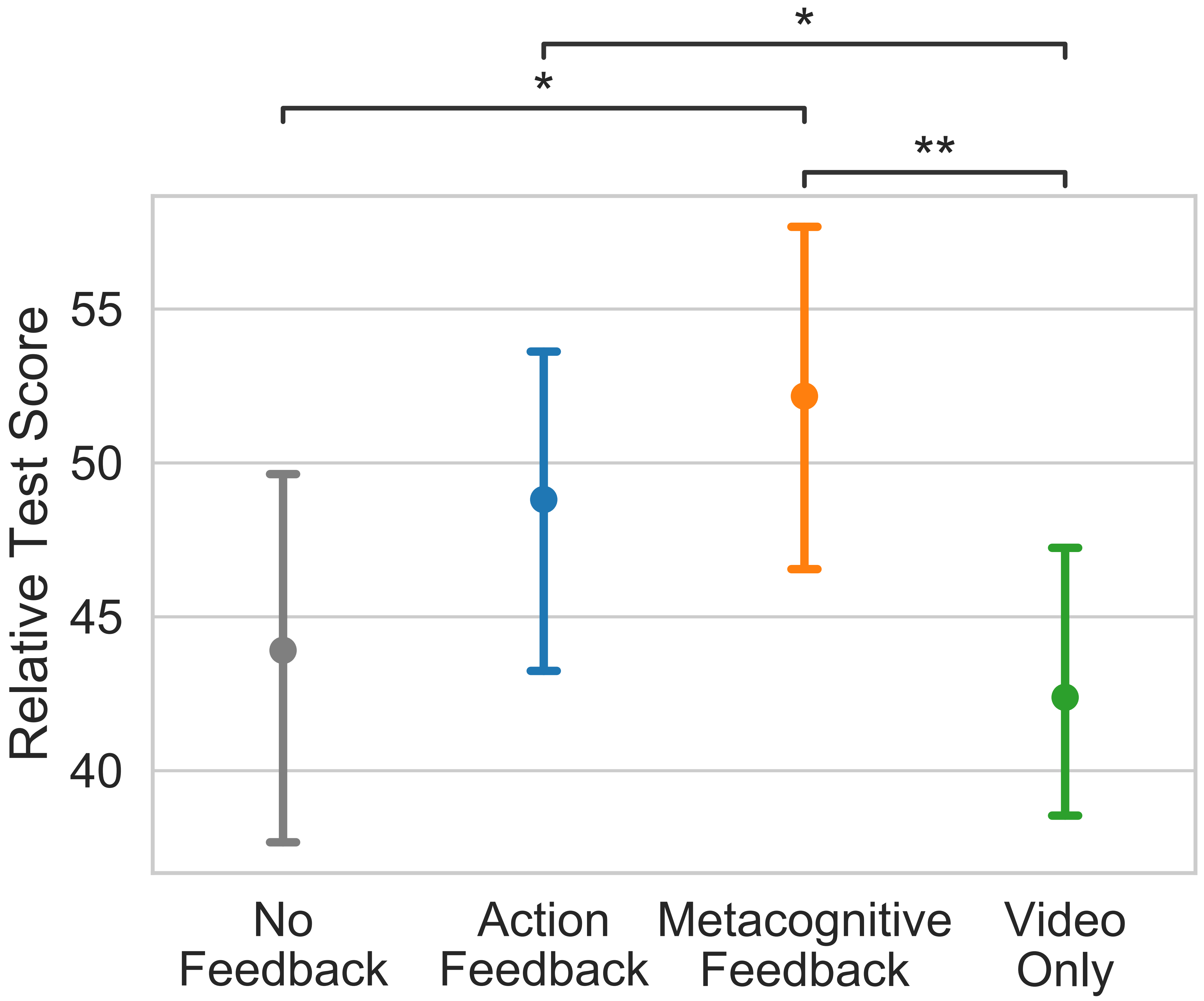
Barrendowns | Reveal | Submit

Looking up Barrendowns

Time cost: $82

Results: Weak transfer to new problems

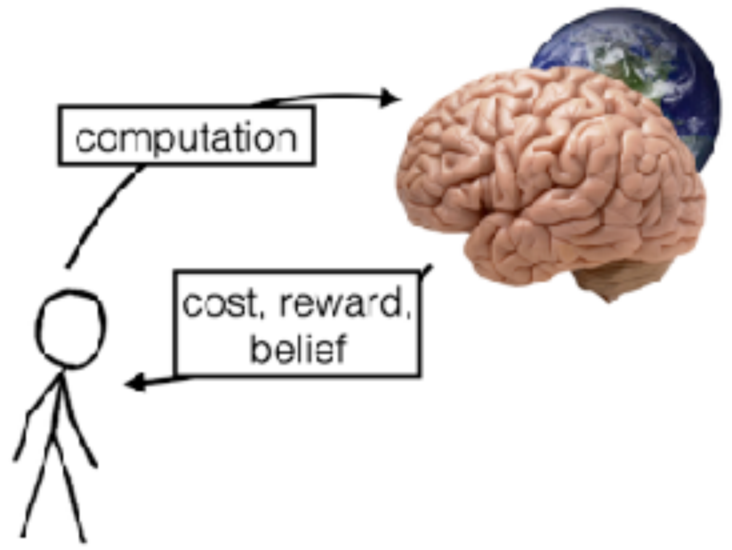




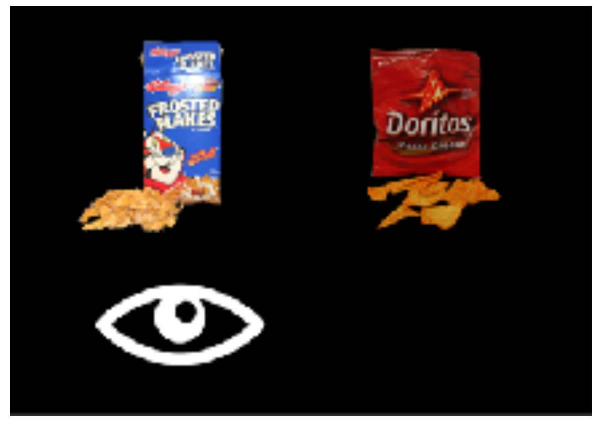
Summary: Discovering and teaching optimal planning strategies

- Planning can be modeled as a meta MDP where an agent decides which hypothetical future action to evaluate next.
- Human planning algorithms are more adaptive than previously proposed heuristic models can account for.
- We can help people learn even more efficient strategies using *reward shaping*, rewarding good thoughts immediately.
    - ➡ But transfer to new contexts presents a challenge.

Metalevel MDPs



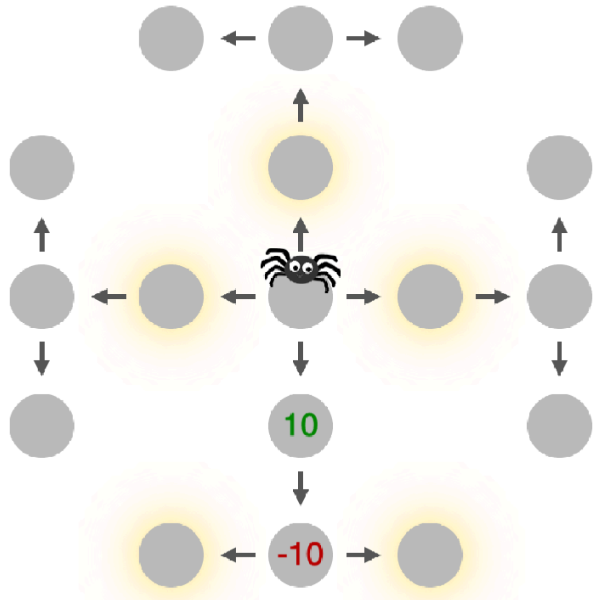
Simple decisions



Multi-attribute decisions

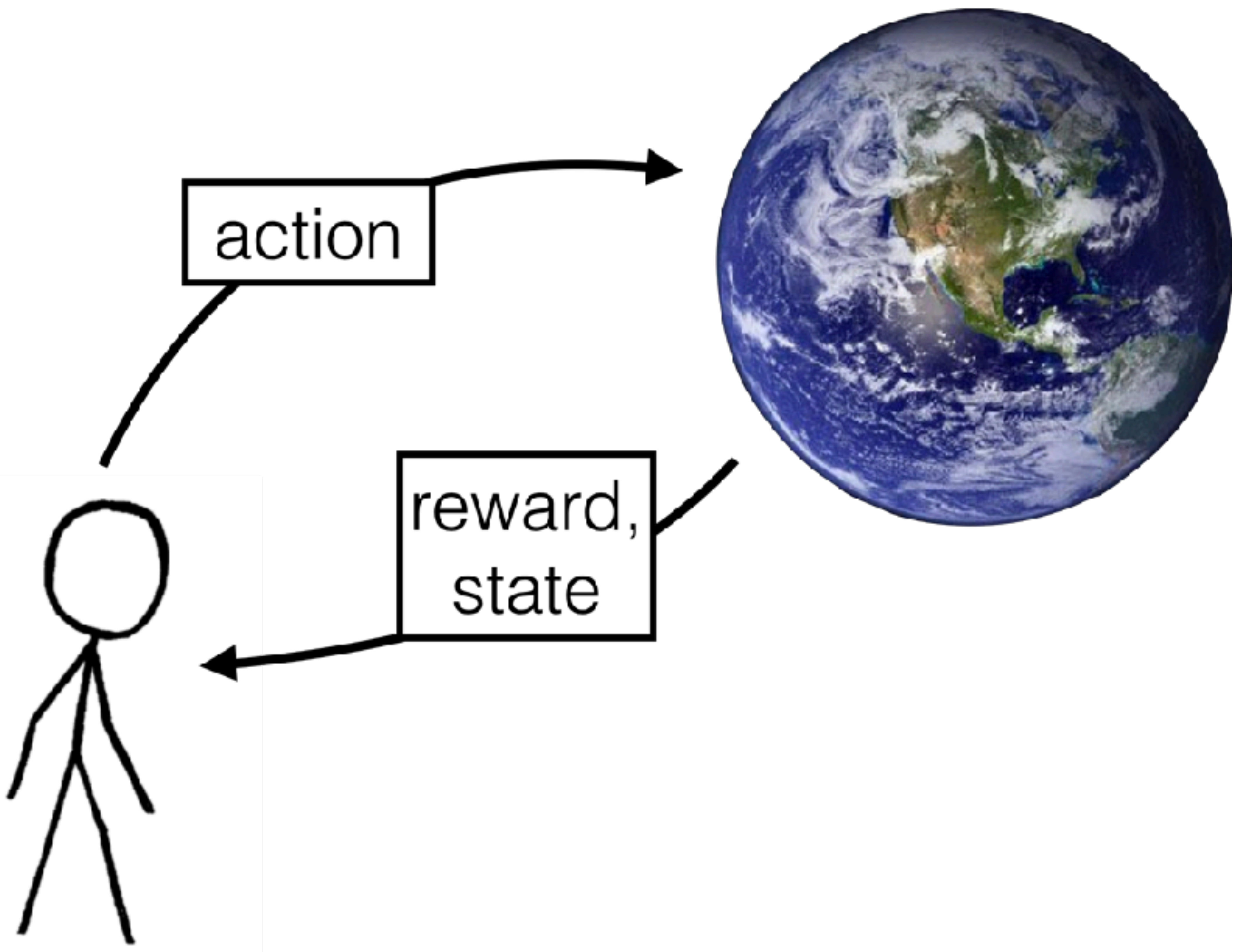
Prizes	Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
A: 3 points	2		3	4	
B: 2 points	7				7
C: 2 points	7	4		2	
D: 21 points	7		8	6	
E: 2 points	9				6

Sequential decisions

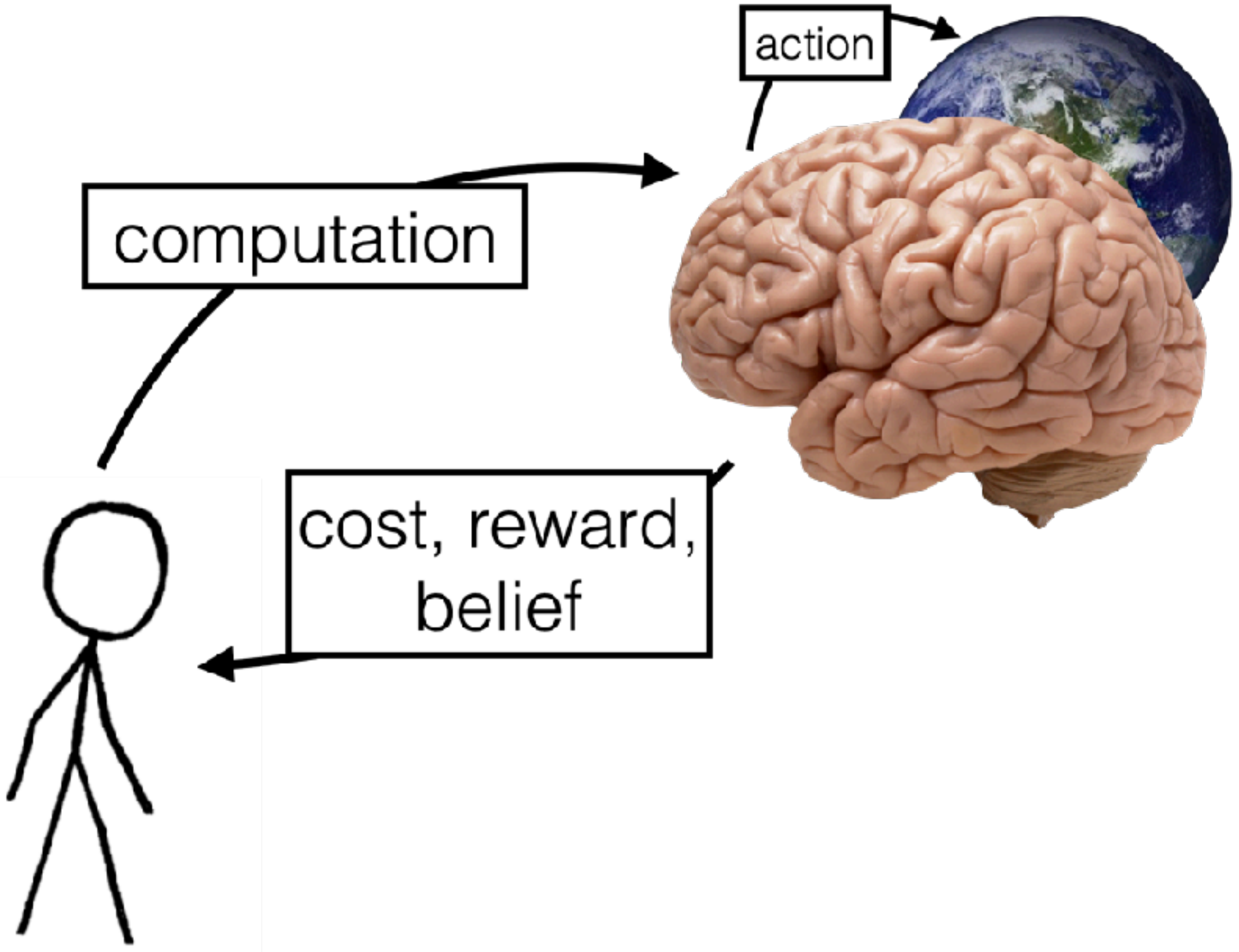


Conclusion: Making decisions in the world and the mind

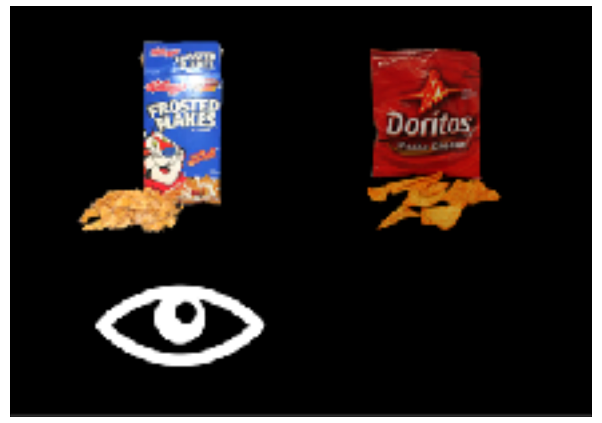
MDP



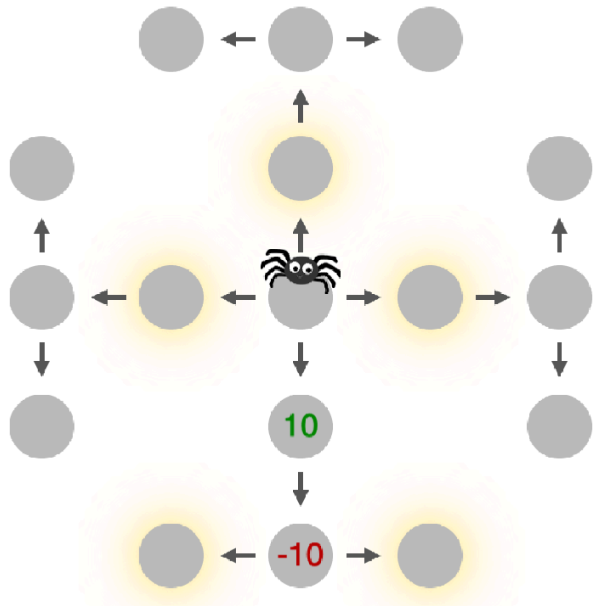
meta MDP

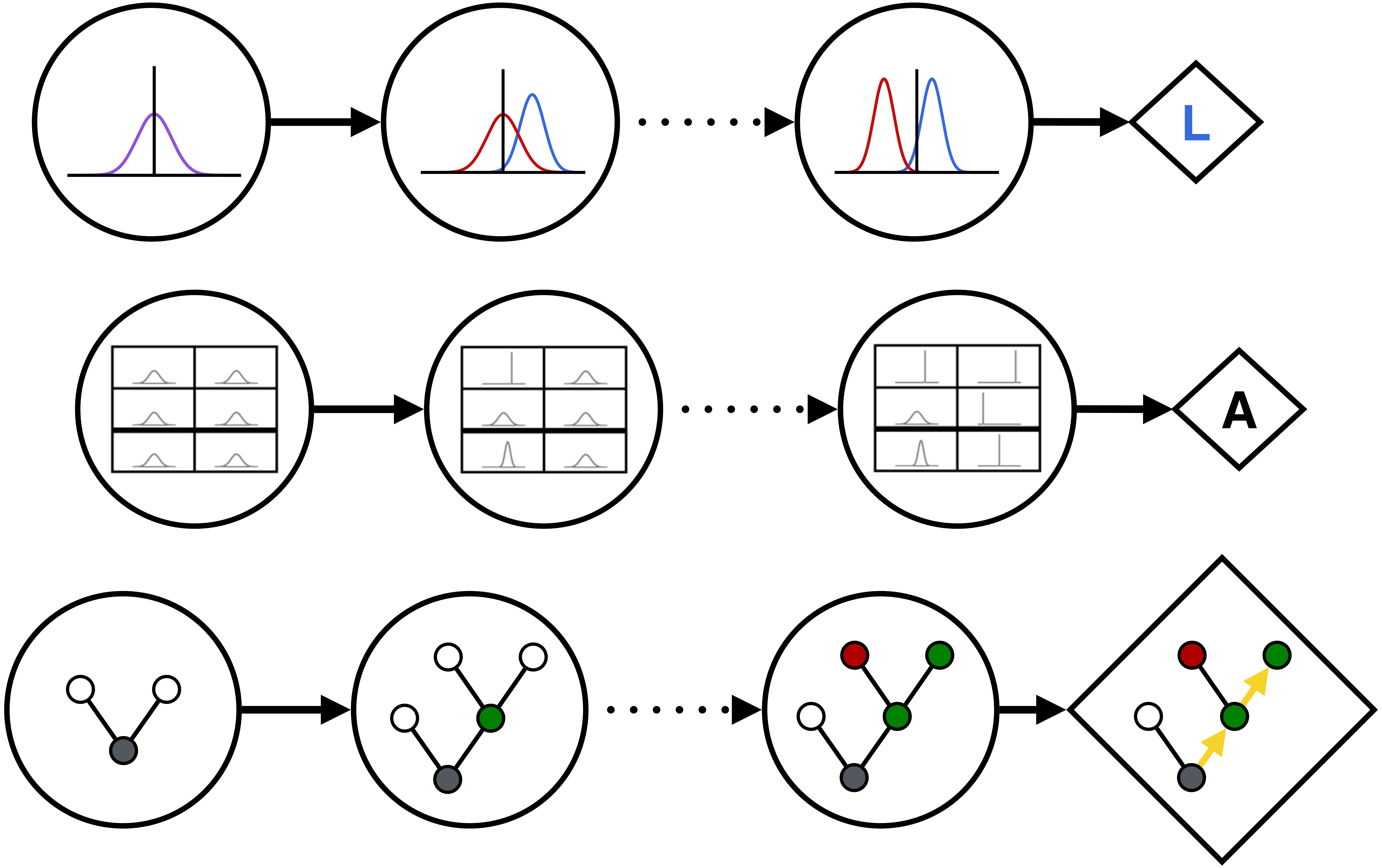


Conclusion: A general framework for resource-rationality

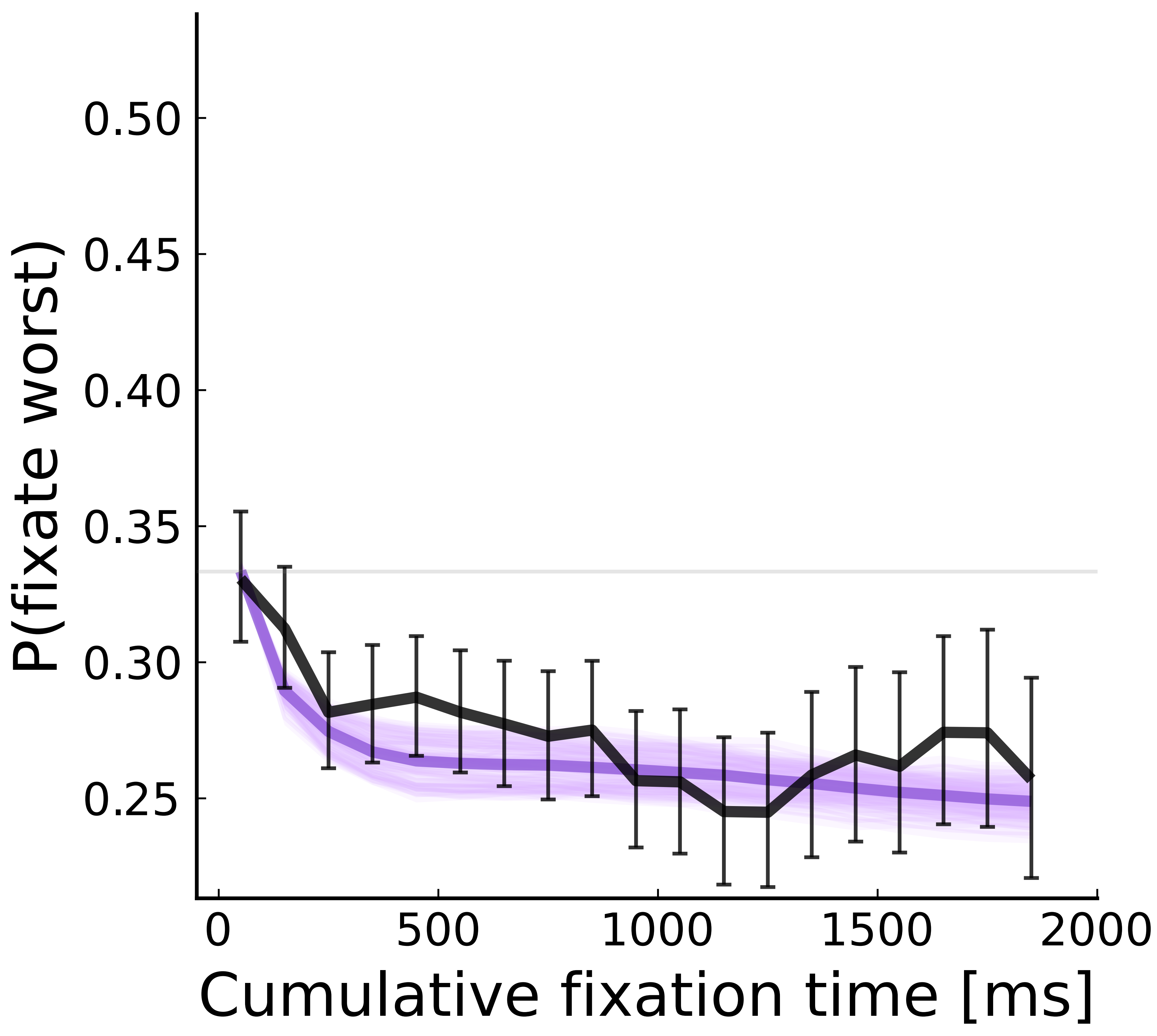


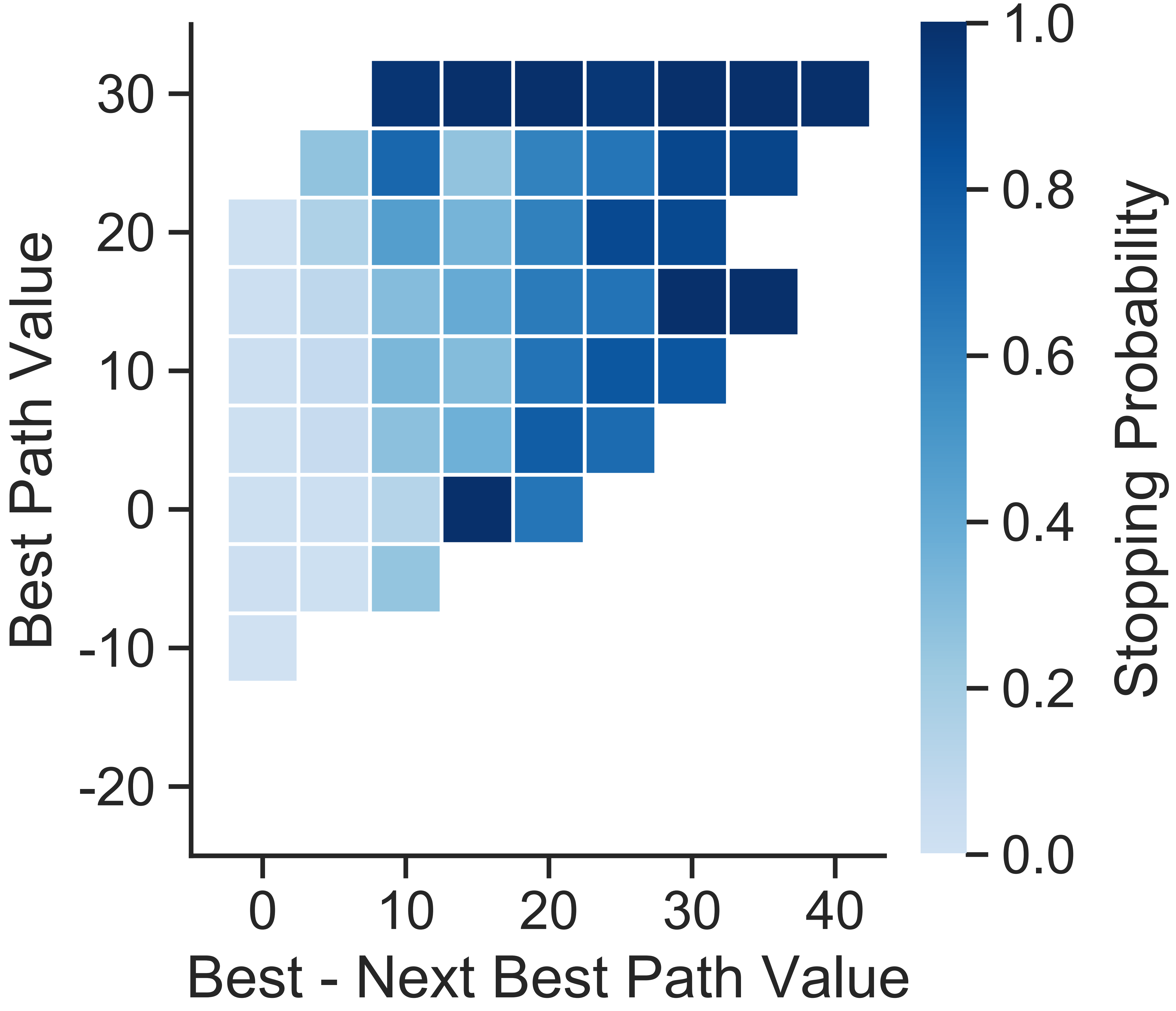
Prizes	Basket 1	Basket 2	Basket 3	Basket 4	Basket 5
A: 3 points	2		3	4	
B: 2 points	7				7
C: 2 points	7	4		2	
D: 21 points	7		8	6	
E: 2 points	9				6



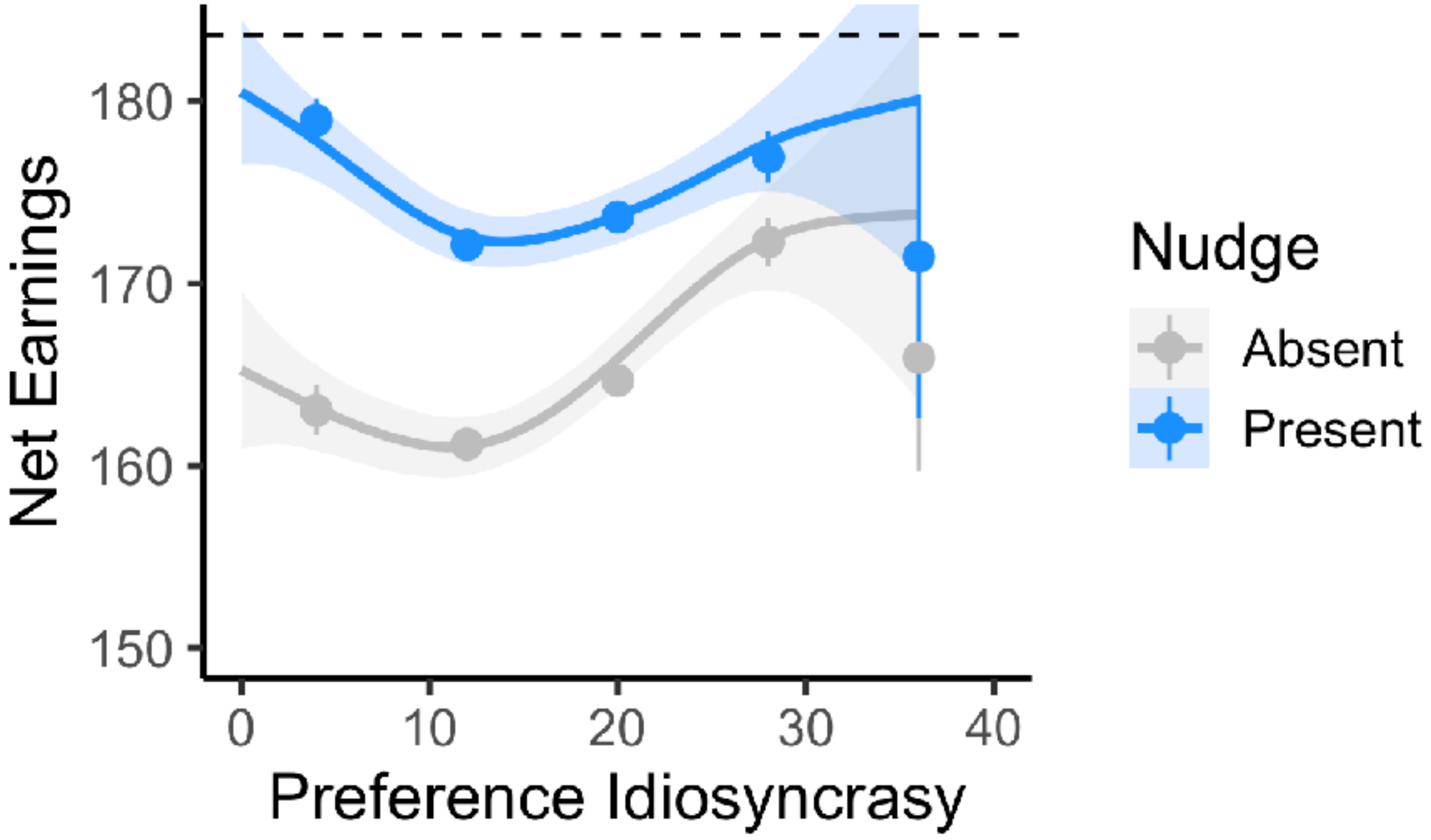


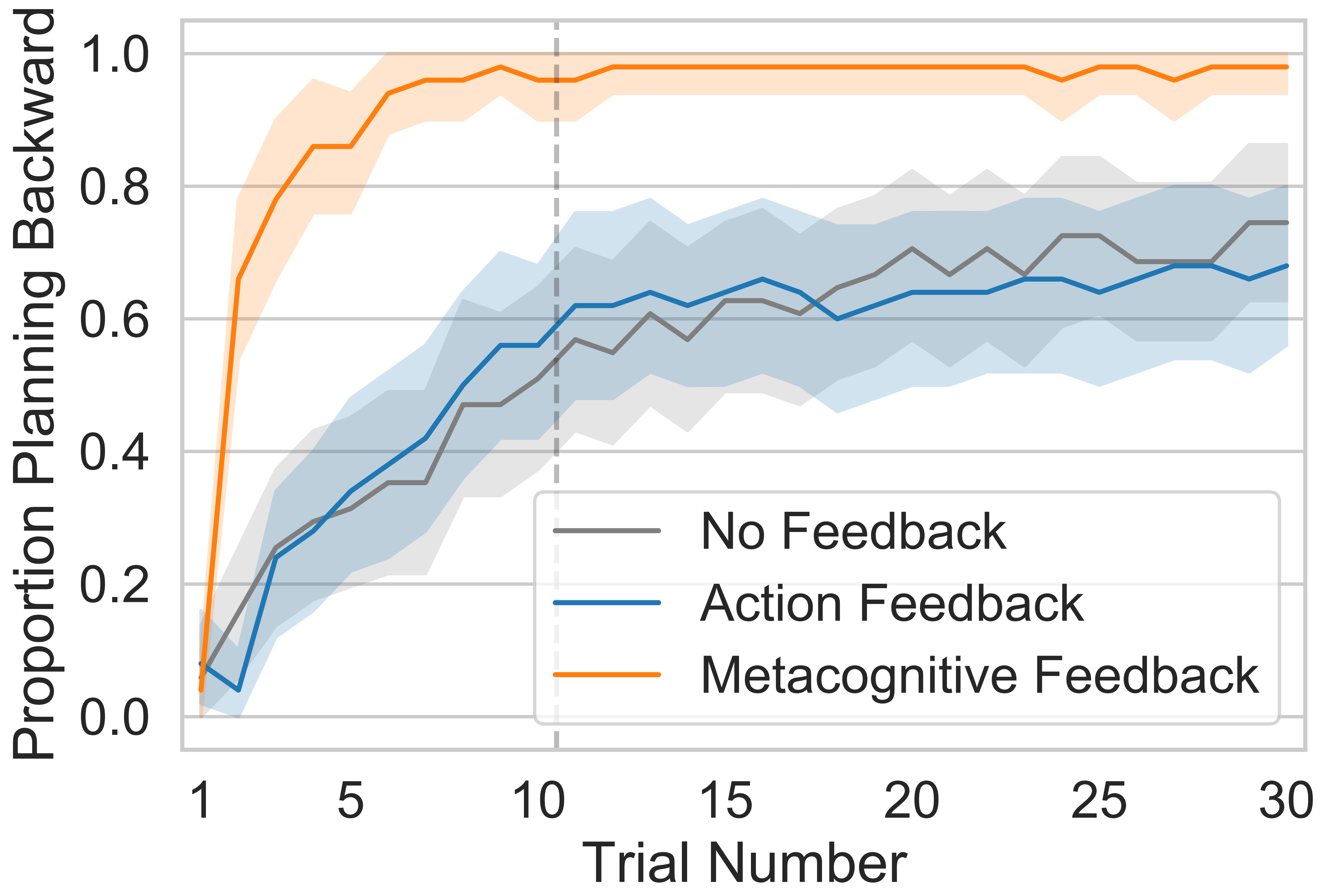
Conclusion: Explaining *how* people make decisions





Conclusion: And helping them make *better* decisions





Thanks!

Antonio Rangel

Matt Hardy

Tom Griffiths

Paul Krueger

Falk Lieder

Bas van Opheusden