COSMOS SUMMER SCHOOL

Computational summer school on modeling social and collective behavior
Tokyo Sept 29th - Oct 3rd, 2025

Tutorial 2: Models of social and individual learning

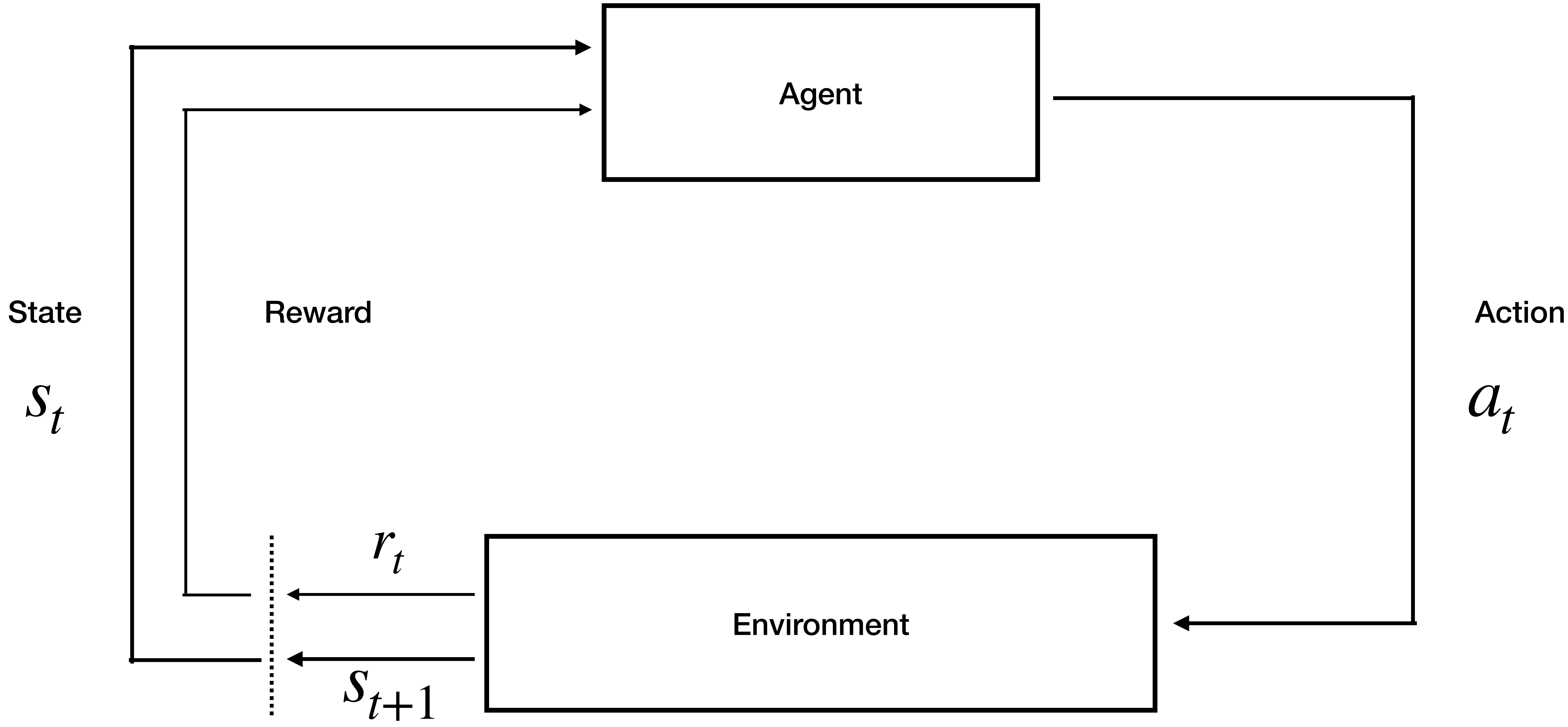
Charley Wu & Wataru Toyoka

Sept 30th

Goals of Tutorial 2:

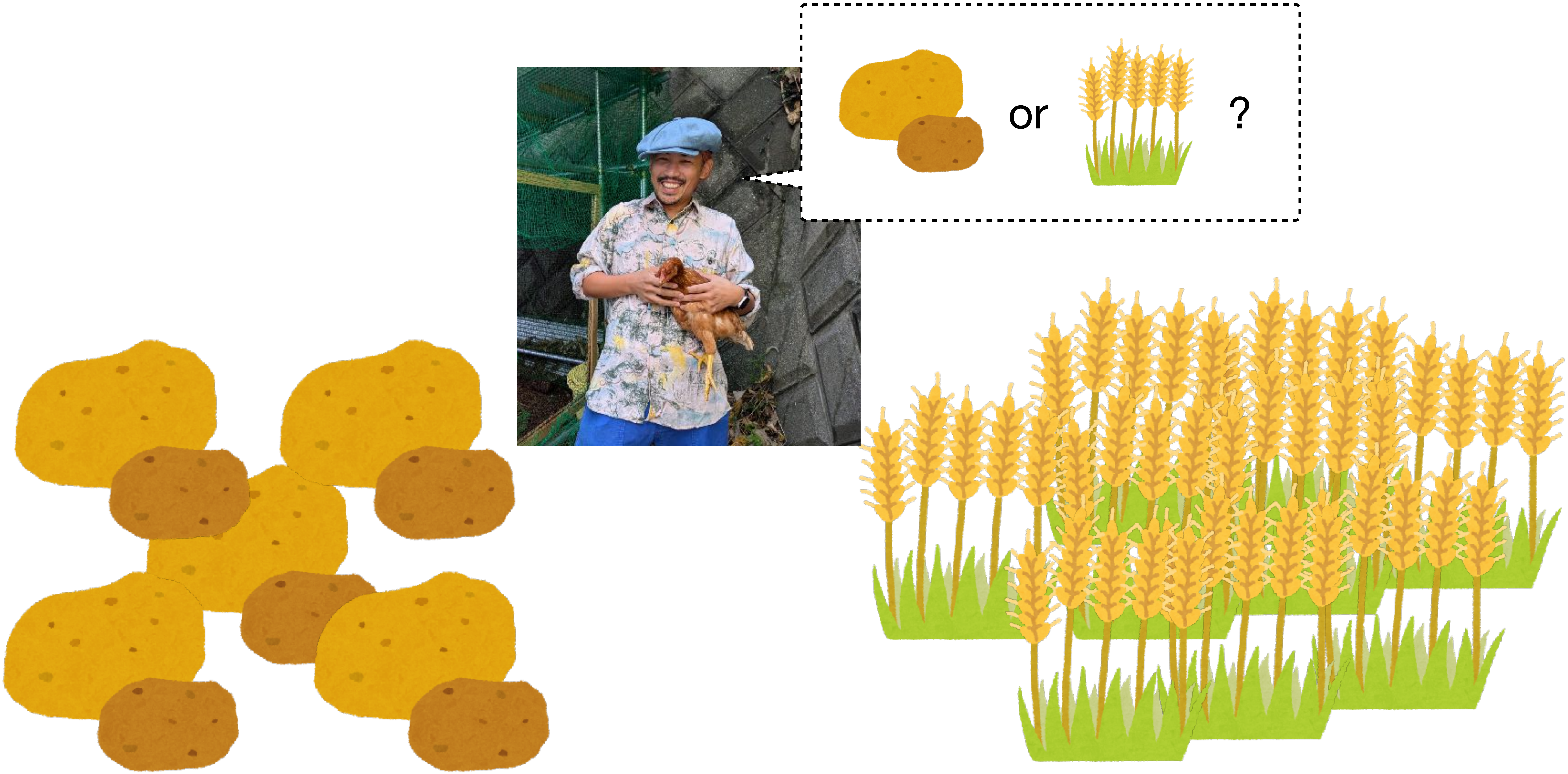
- **Brisk introduction to asocial RL**
  - Simulating data
  - Maximum likelihood estimation (MLE) of model parameters
  - Predicting choices
- **Social learning models**
  - Imitating actions
  - Combining asocial and social learning
  - Social learning hierarchy (from imitation to Theory of Mind)
- **Scaling up to more complex problems**

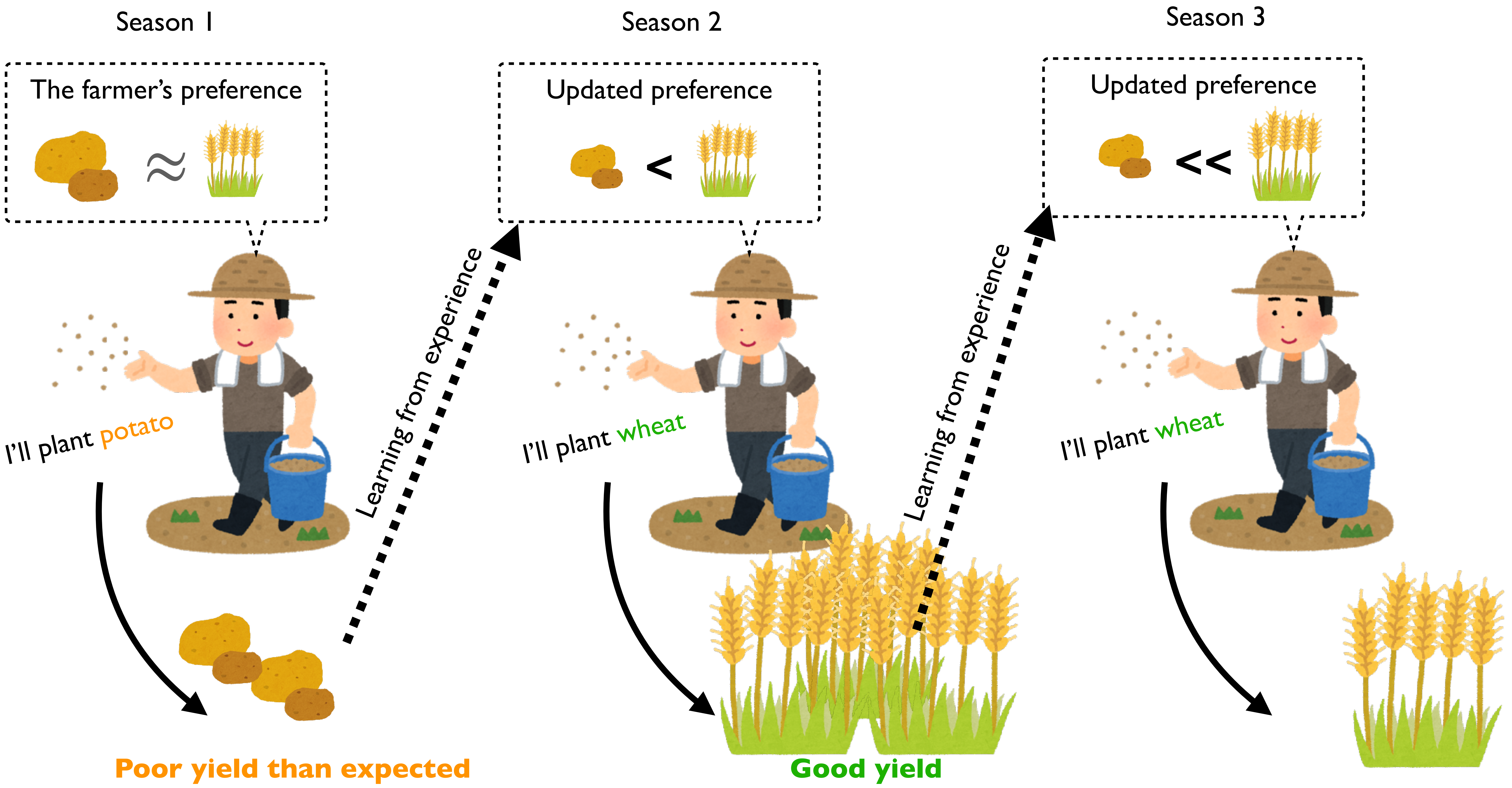
Reinforcement Learning (RL)



Sutton & Barto (1998)

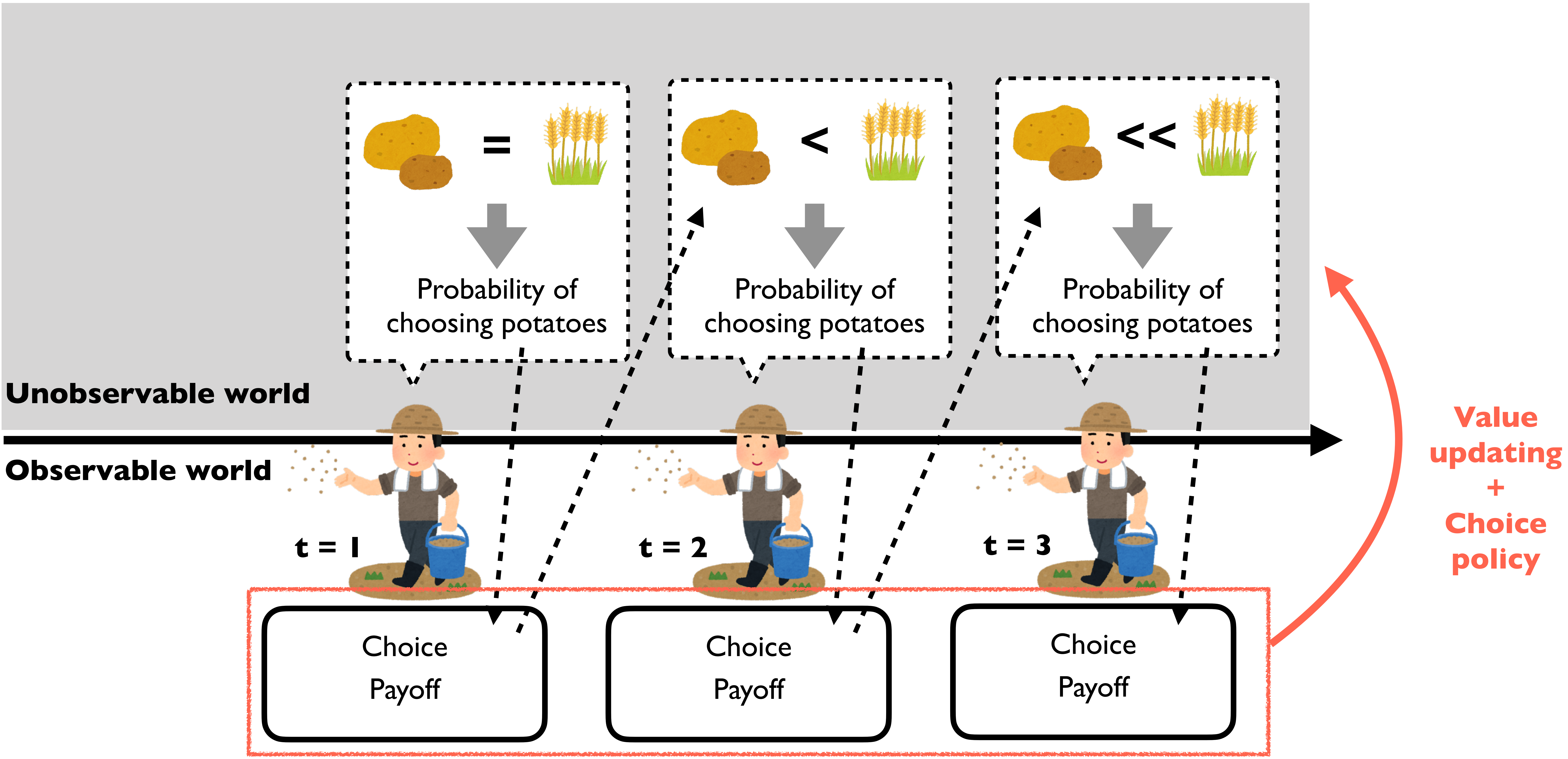
A multi-armed bandit task



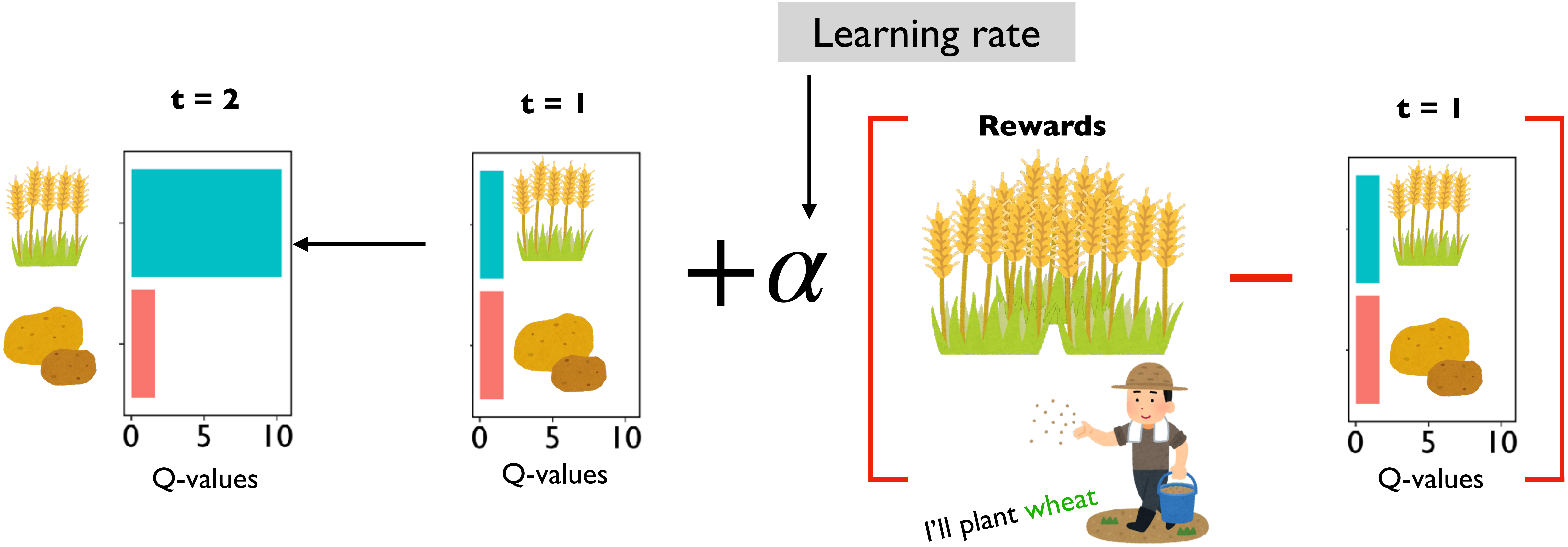


Season 1
The farmer's preference
≈
Season 2
Updated preference
<
Season 3
Updated preference
<<
I'll plant potato
Learning from experience
I'll plant wheat
Learning from experience
I'll plant wheat
Poor yield than expected
Good yield

RL process is not directly observable. It must be statistically inferred.



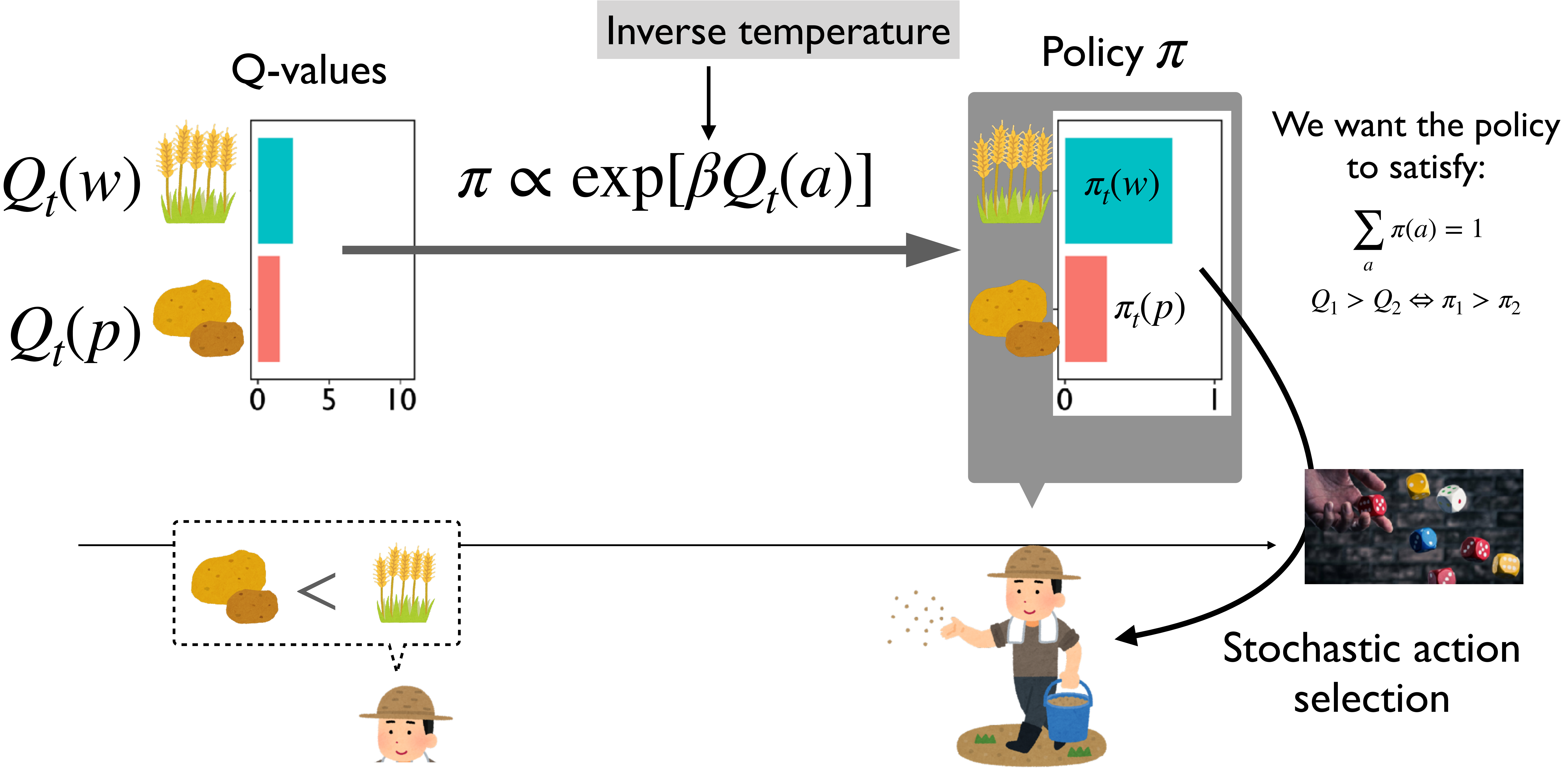
Value updating:
Q-Learning

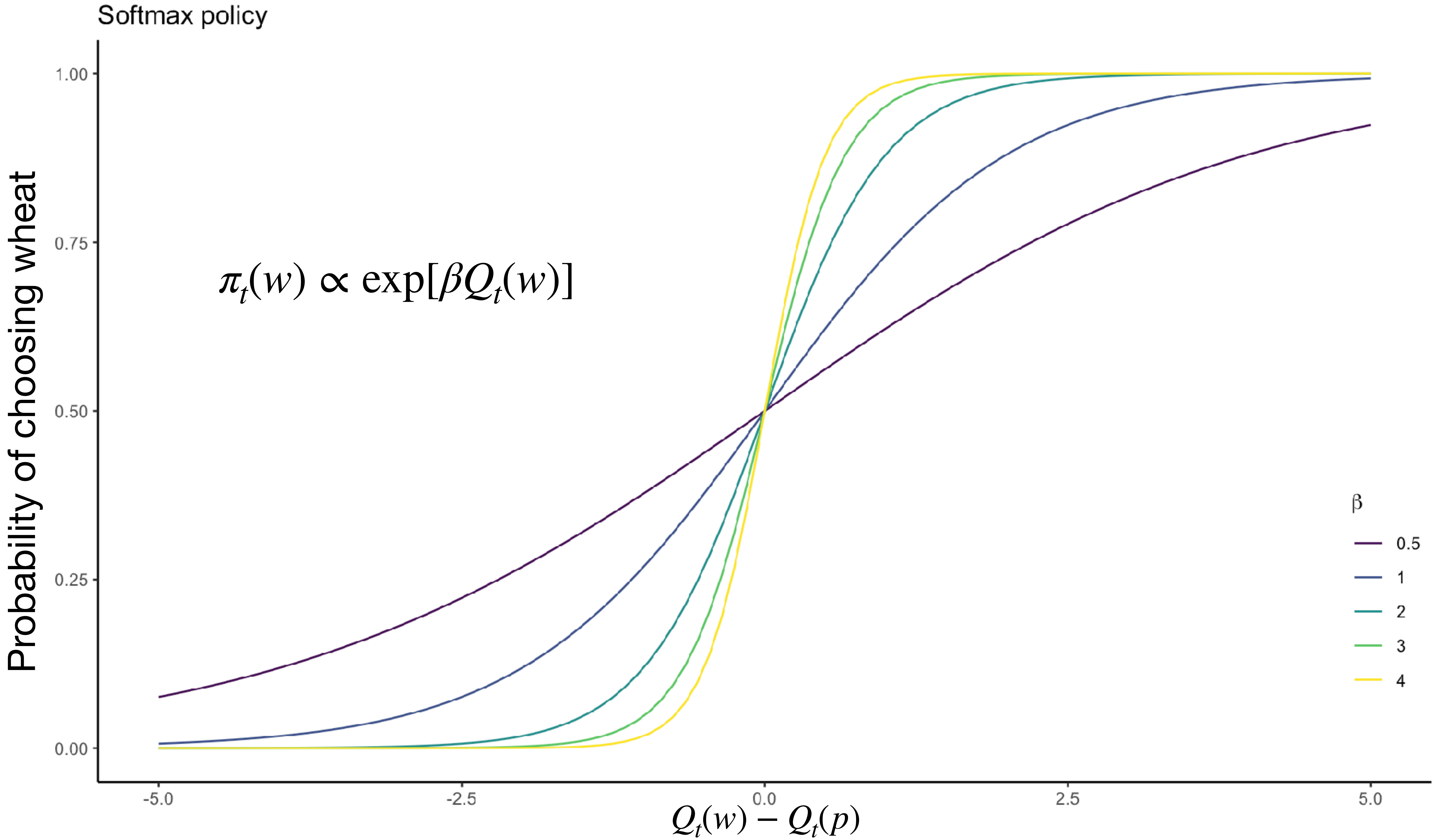


reward prediction error (RPE)

$$Q_{t+1}(a) \leftarrow Q_t(a) + \alpha\,[r_t(a) - Q_t(a)]$$

Choice policy: Softmax

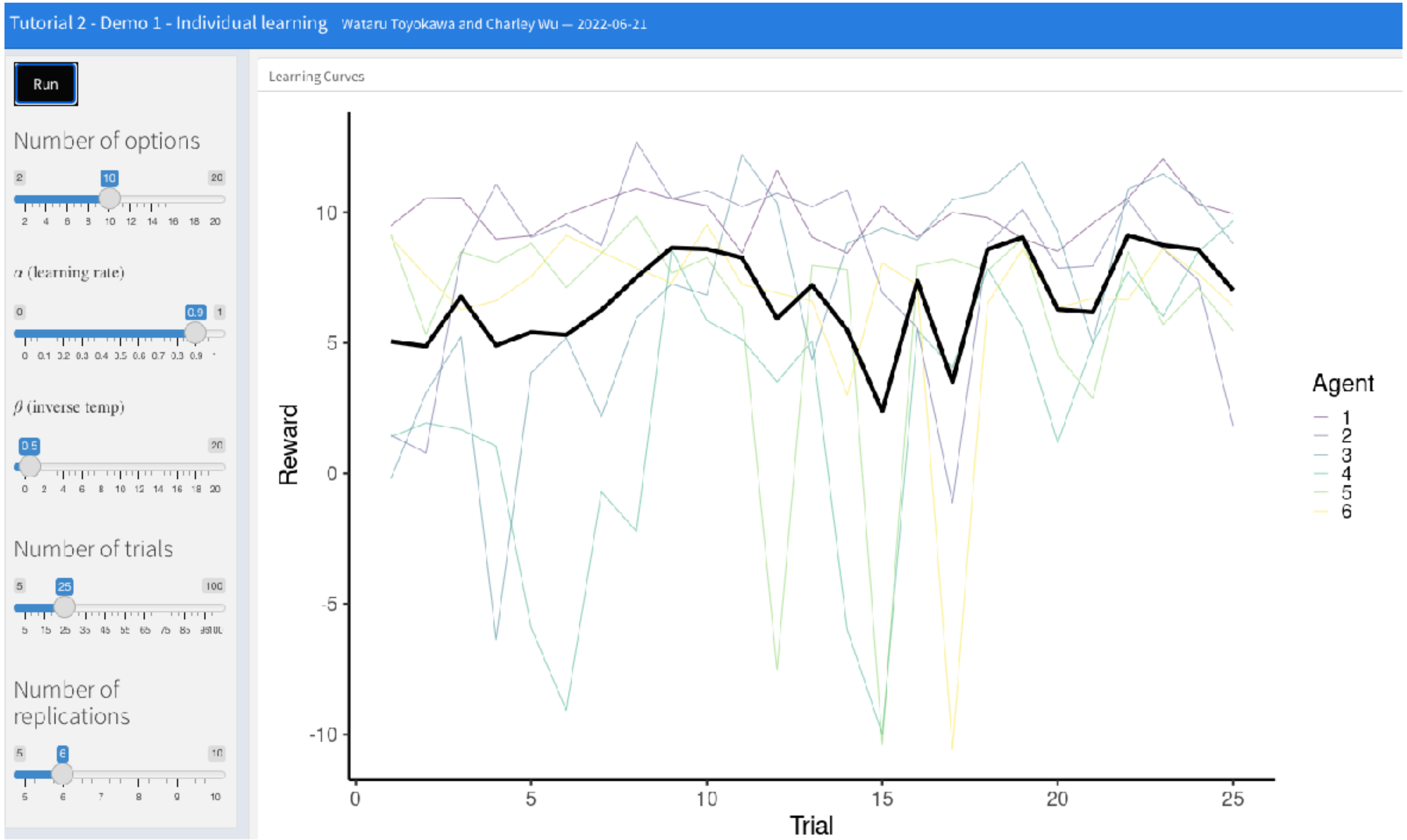



Softmax policy
Probability of choosing wheat
1.00
0.75
0.50
0.25
0.00
-5.0
-2.5
0.0
2.5
5.0
$Q_t(w) - Q_t(p)$
$\pi_t(w) \propto \exp[\beta Q_t(w)]$
β
0.5
1
2
3
4

Notebook

https://cosmossummerschool.github.io/notebooks/tutorial-2-models-of-learning.html

Demo 1: Tweaking individual learning parameters



Which learning parameters (α, β) typically produce the best results?

Likelihood function

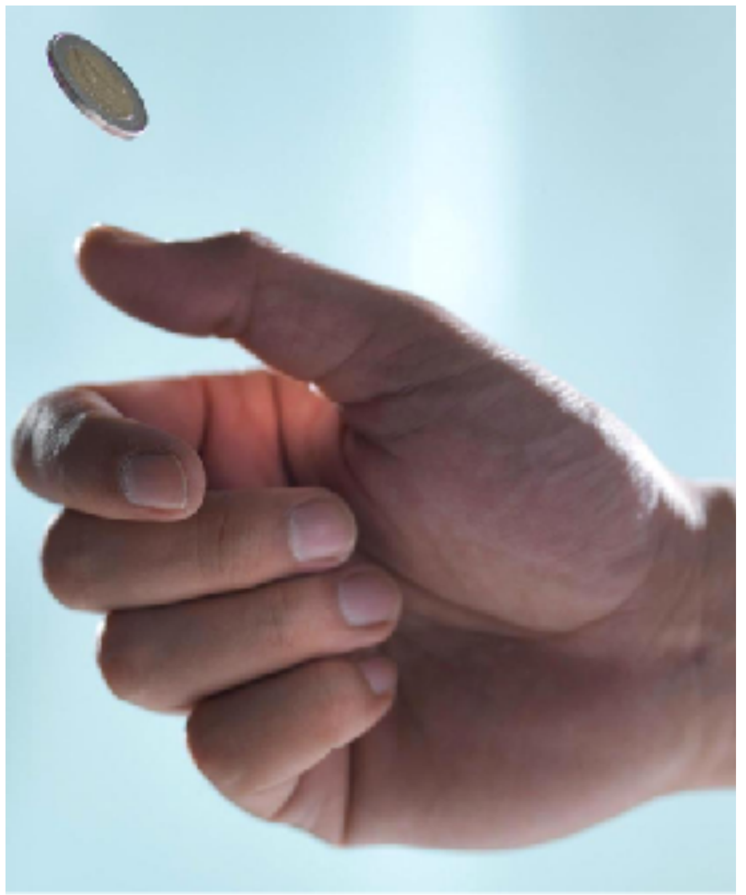
Beyond only simulating data, we also want to use models to fit experimental data.

To do so, we first need to define a **Likelihood Function:**

$$P(D\,|\,\theta)$$

describing the probability the observed data D was generated by model parameters θ

Coin Flip Model



Observed Data:

$D = \{$  $\}$

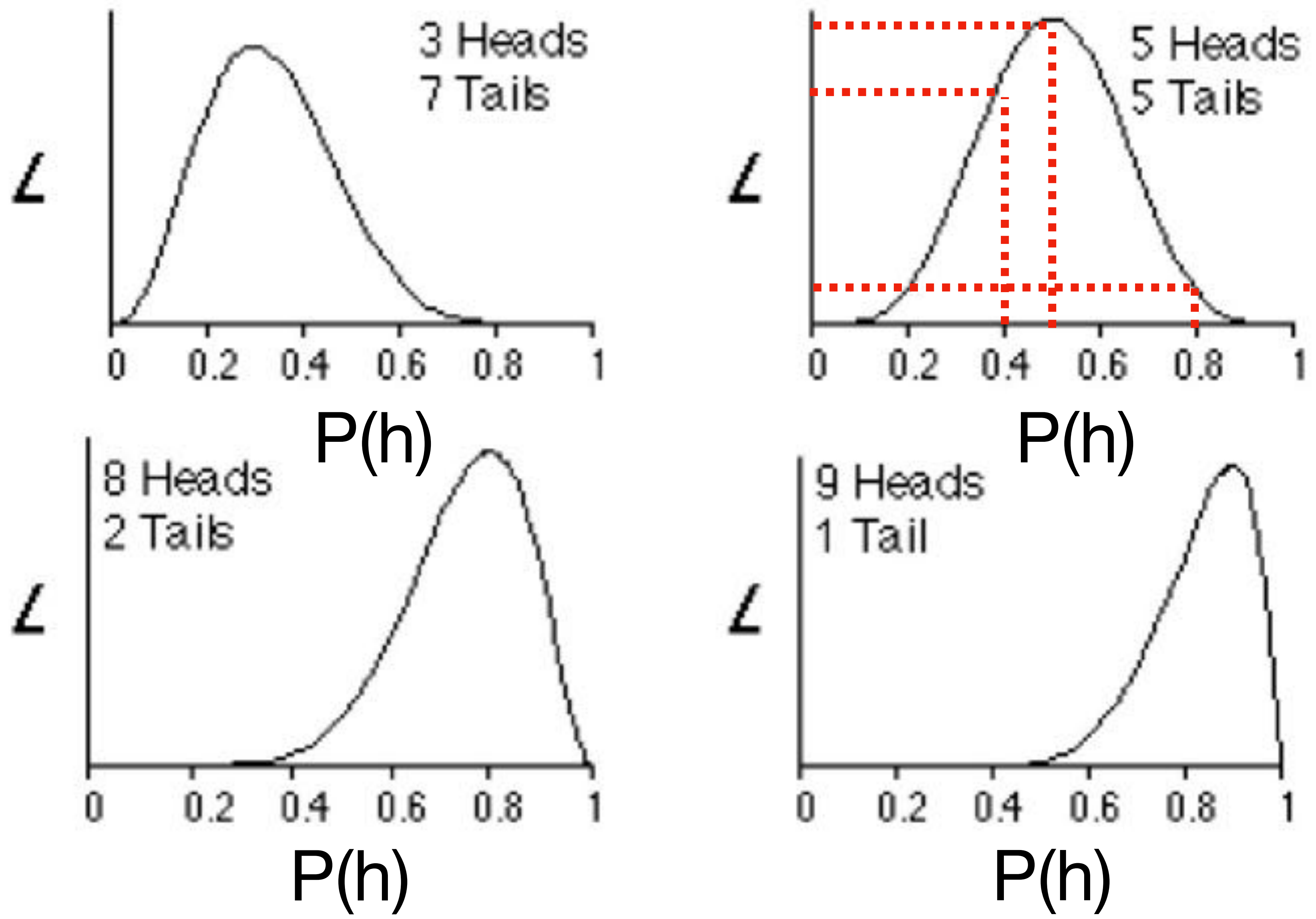
Model:

flips

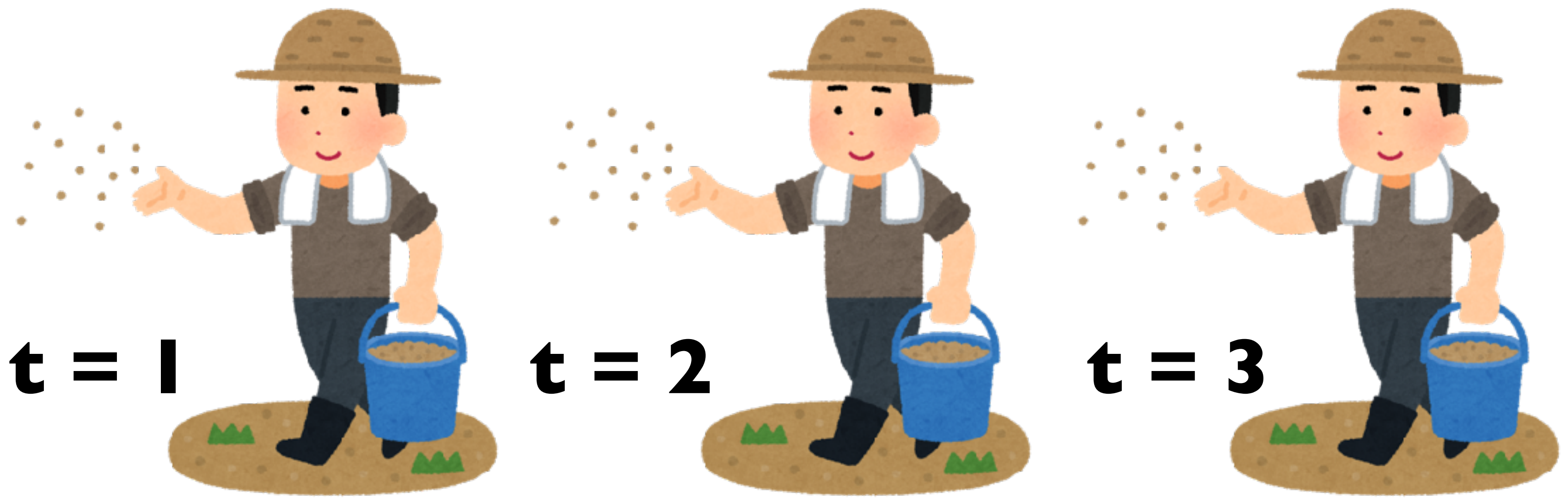
$D \sim$ **Binomial**(n, θ)

$\theta = P(h)$

P(heads)



Log Likelihoods



For multiple data points, we need to describe the **joint likelihood** over all observations:

$$P(D\,|\,\theta) = \prod_t P(d_t\,|\,\theta)$$

This is much easier using logarithms, since we can replace multiplication with summation in log space to compute the **log likelihood**

$$\log P(D\,|\,\theta) = \sum_t \log P(d_t\,|\,\theta)$$

Since probabilities are always <1, the log likelihood will always be negative. Thus, it's more convenient to express model fit using the **negative log likelihood** (nLL) by inverting the sign:

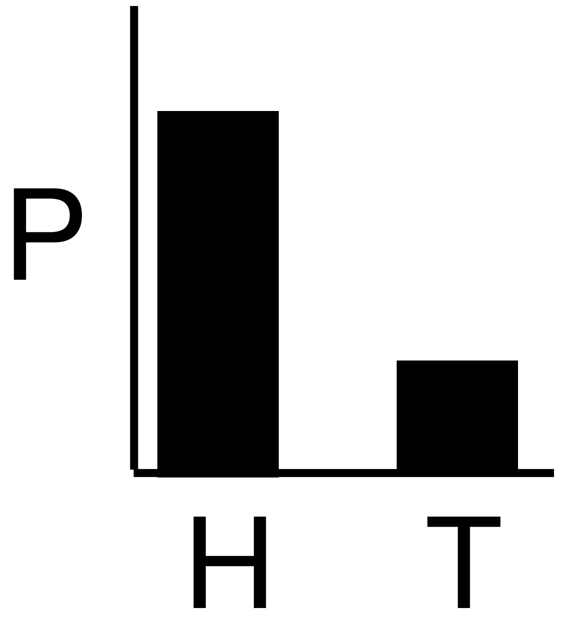
$$nLL = -\log P(D\,|\,\theta)$$

The nLL expresses the amount of error or loss (aka 'log loss') and will always be greater than zero. *Smaller values thus describe better model fits.*

*Natural logs are often used by default (also written as $\ln$) rather than base 10 logarithms $\log_{10}$

Likelihoods as Goodness of Fit

	Measure	Formula	Heads	Tails
	Likelihood	P(D|θ)	80%	20%
	Log likelihood	log P(D|θ)	–0.22	–1.61
aka Log Loss	Negative Log Likelihood (nLL)	– log P(D|θ)	0.22	1.61
Used in BIC/AIC	Deviance	–2 log P(D|θ)	0.44	3.22



Notebook

From model simulation to likelihood functions

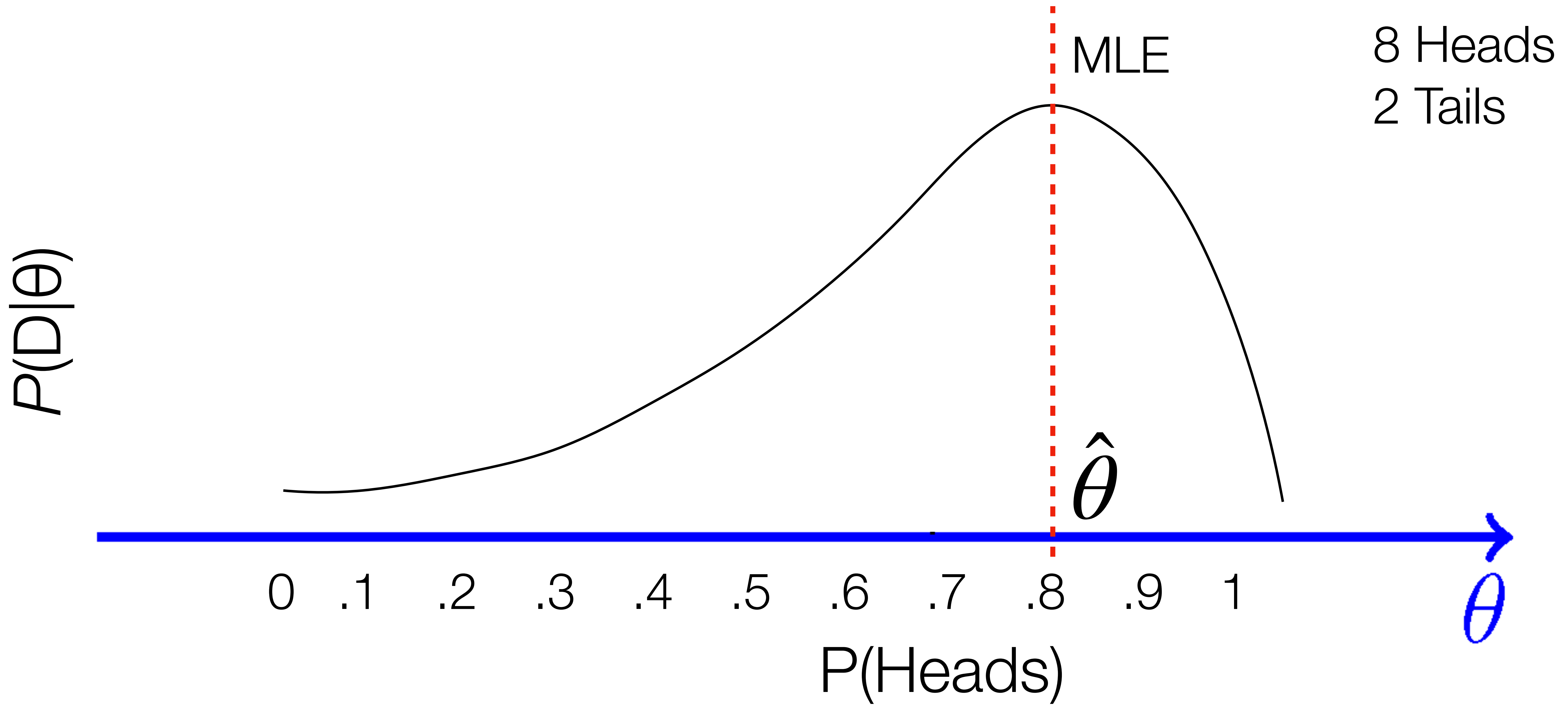
In practice, we can use code very similar to our model simulations to create a likelihood function

```r
likelihood <- function(params, data){
   nLL <- 0 #initialize negative log likelihood
   for (d in data){ #loop through data
     #make predictions
     predictions <- model(params)
     #define true outcome
     observedAction <- d
     #Update nLL
     nLL <- nLL -log(predictions[observedAction])
   }
   return(nLL)
}
```

```python
def likelihood(params, data, model):
    nLL = 0  # Initialize negative log-likelihood
    # Loop through each observed action
    for observed_action in data:
        # Make predictions
        predictions = model(params)
        # Pobability of the observed action
        prob = predictions[observed_action]
        # Update the negative log-likelihood
        nLL -= np.log(prob)
    return nLL
```

Fitting Models with Maximum Likelihood Estimation (MLE)

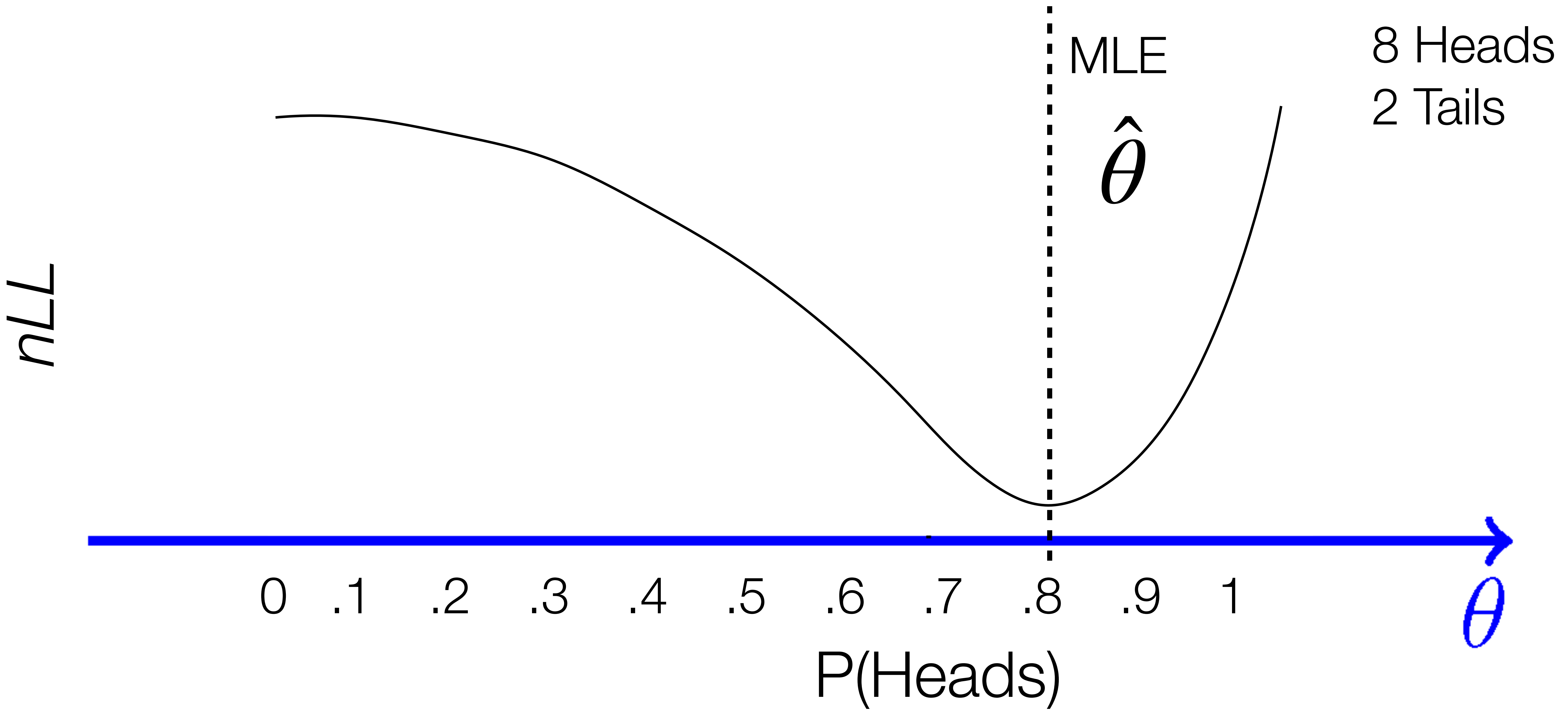
Use the likelihood function to find the parameters $\hat{\theta}$ where $P(D \mid \theta)$ is largest



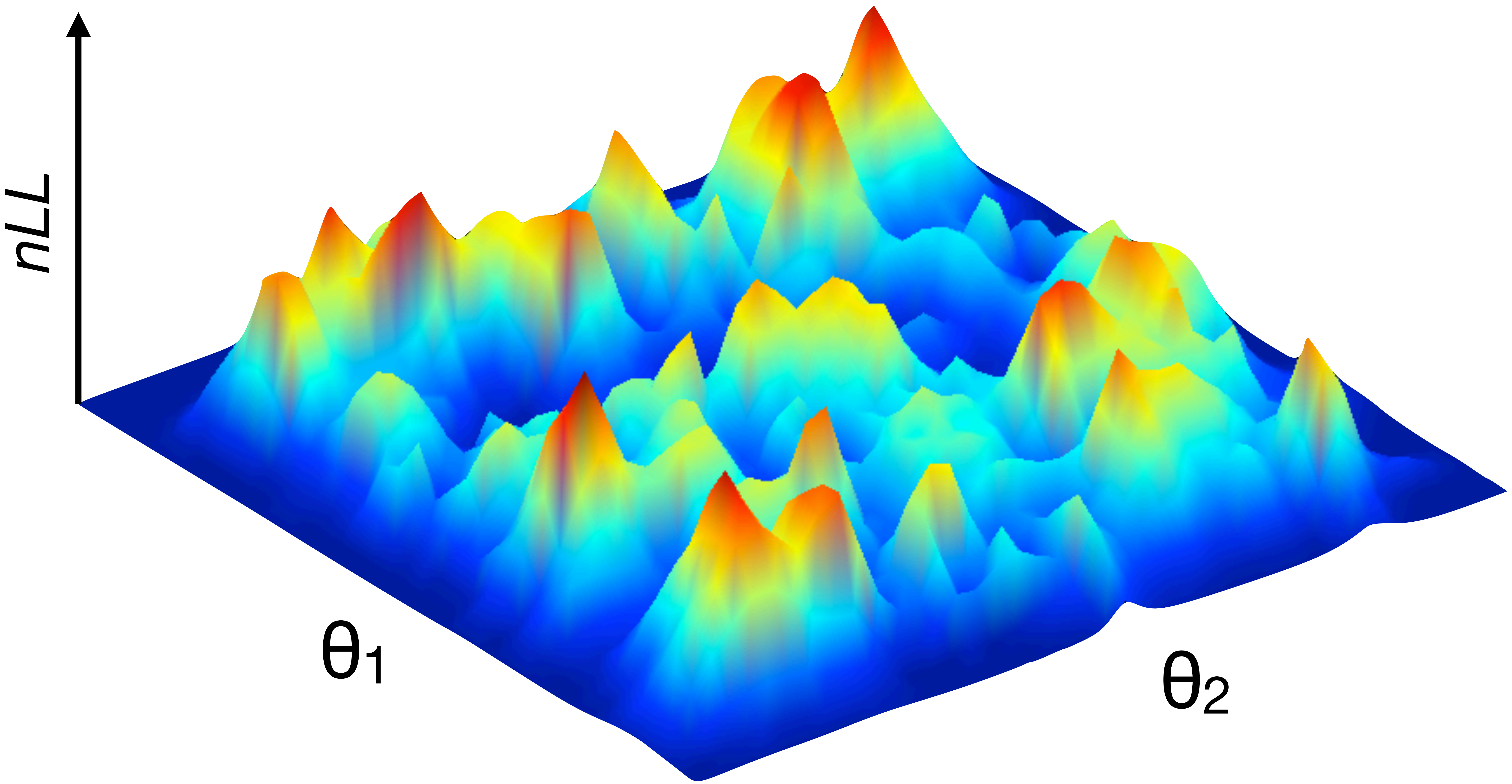
Fitting Models with Maximum Likelihood Estimation (MLE)

Use the likelihood function to find the parameters $\hat{\theta}$ where $P(D \mid \theta)$ is largest

…. or where nLL is lowest



Computing the MLE



Optimization function

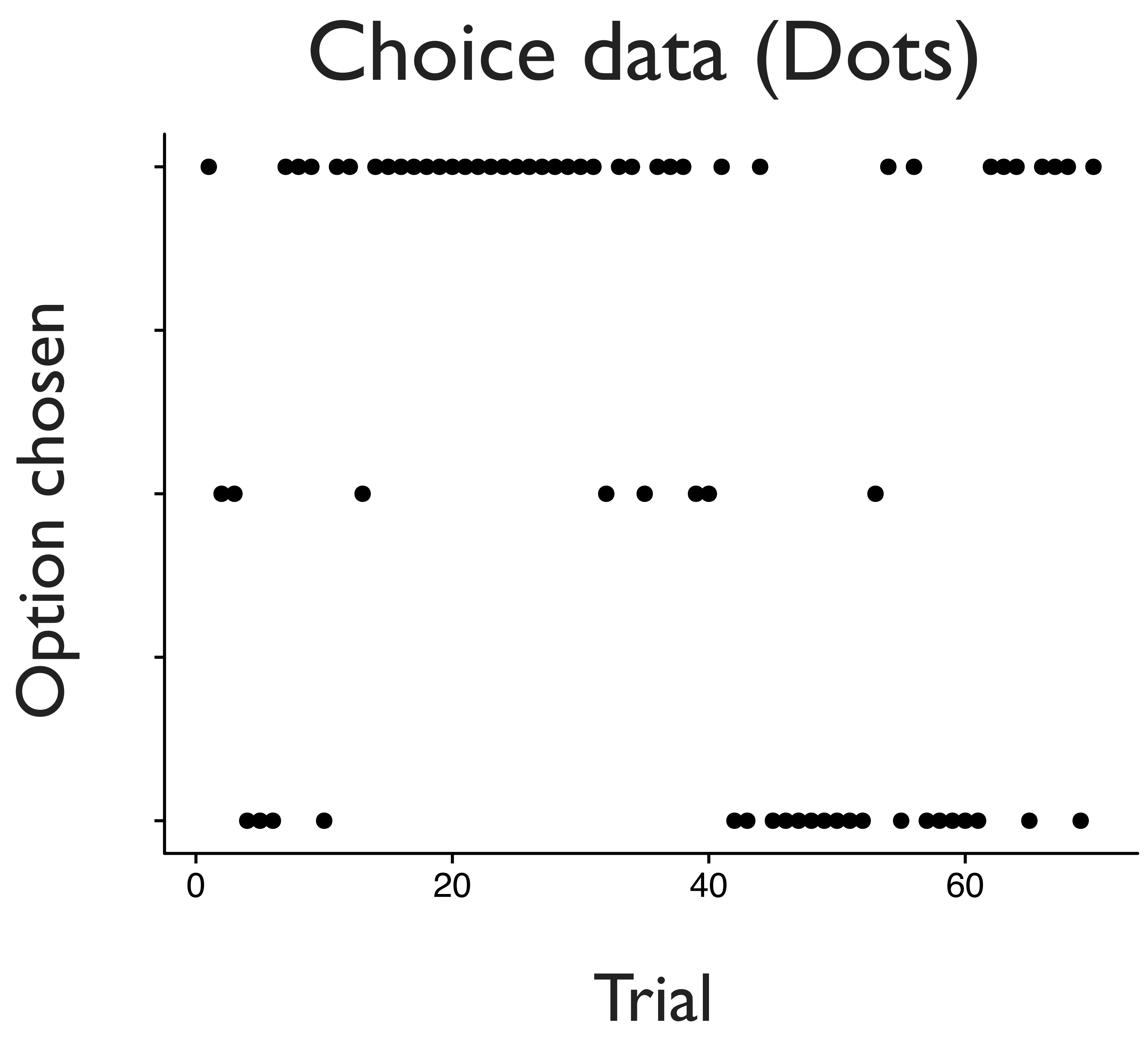
```
likelihood <-
function(params
, data)
```

Minimize nLL

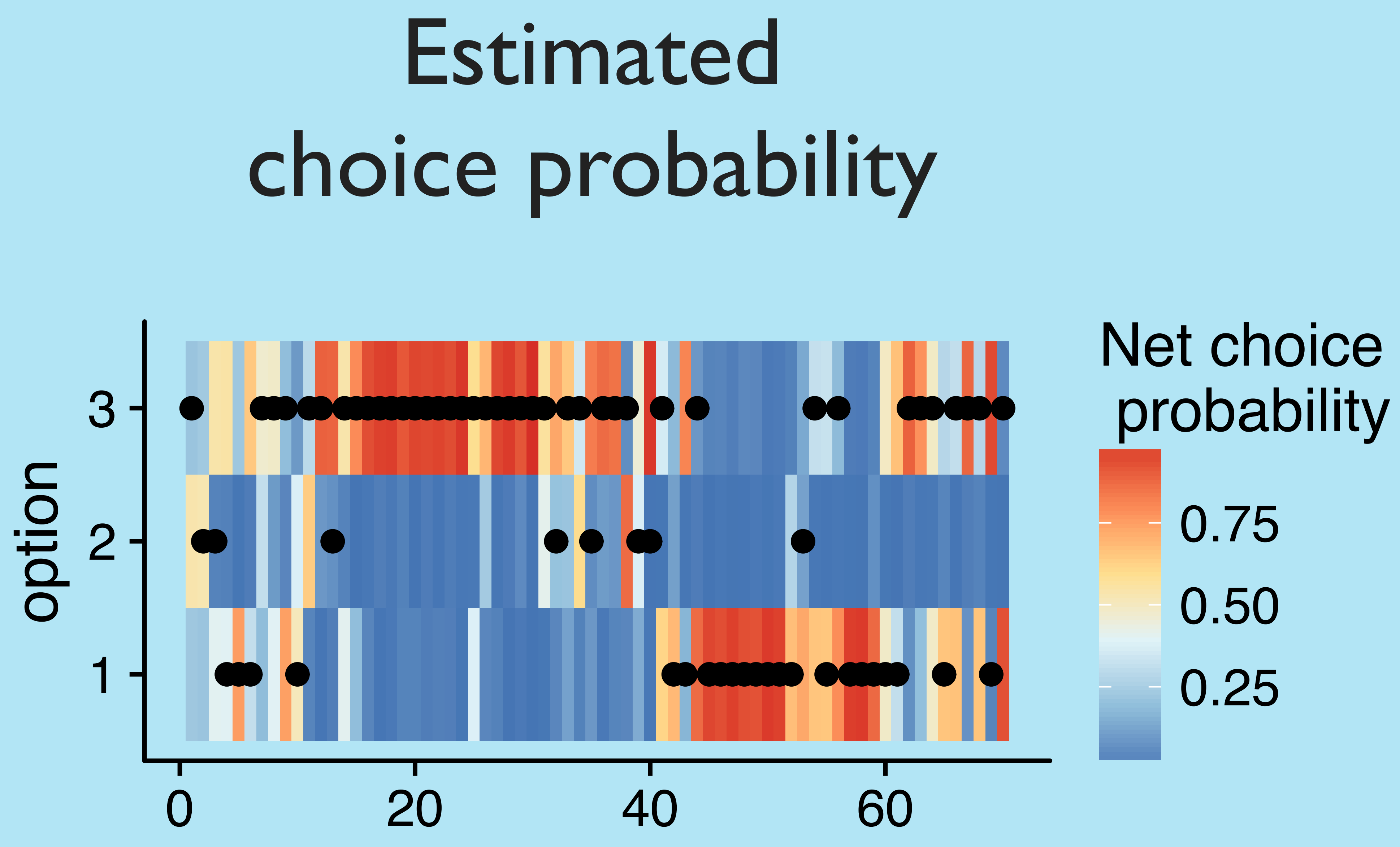
Types of optimization algorithms

- Gradient descent
- Simplex methods
- Differential evolution

MLE for a RL model



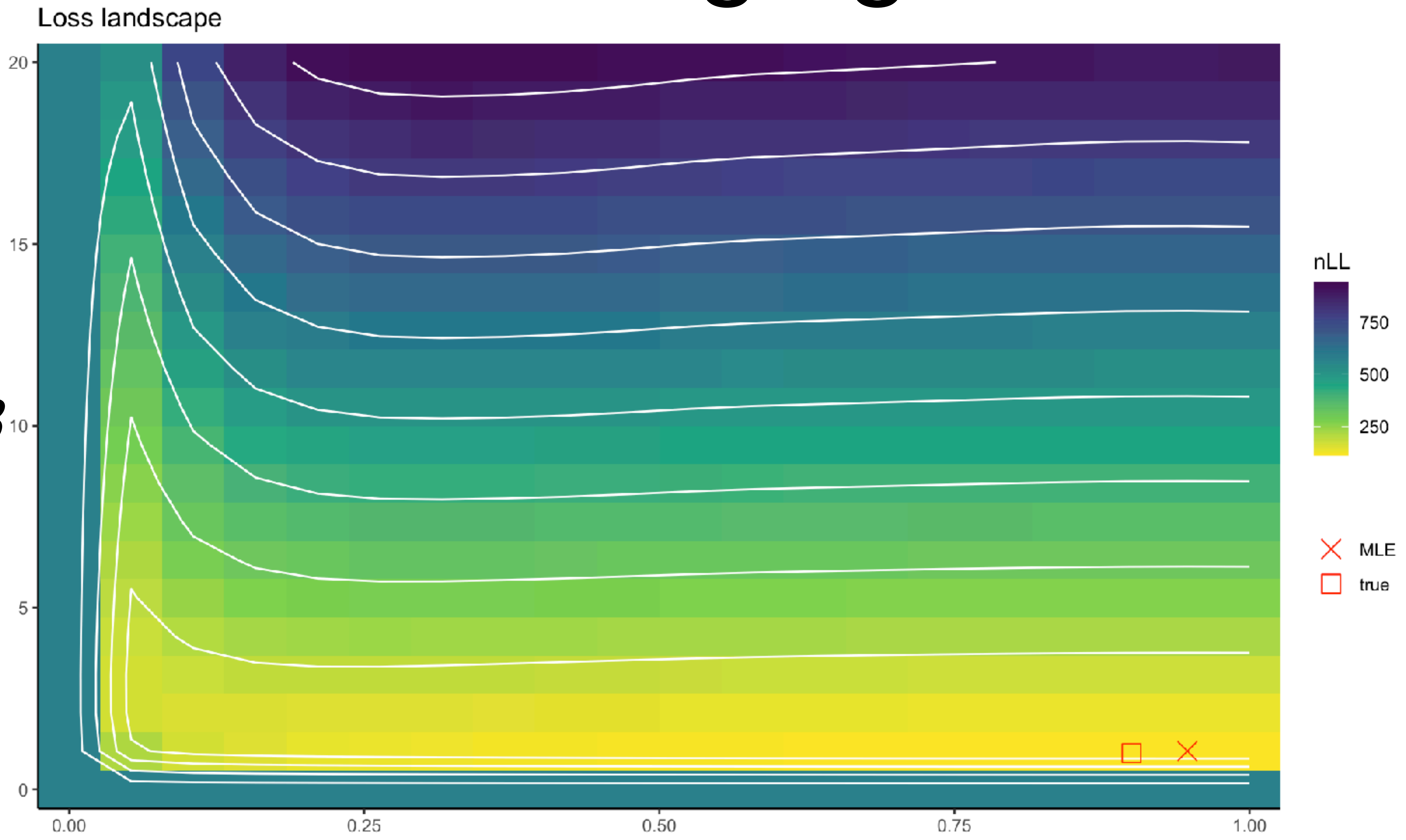
Search for
parameters that
maximize likelihood



Maximum Likelihood Estimate (MLE)

Notebook

Q-learning agent

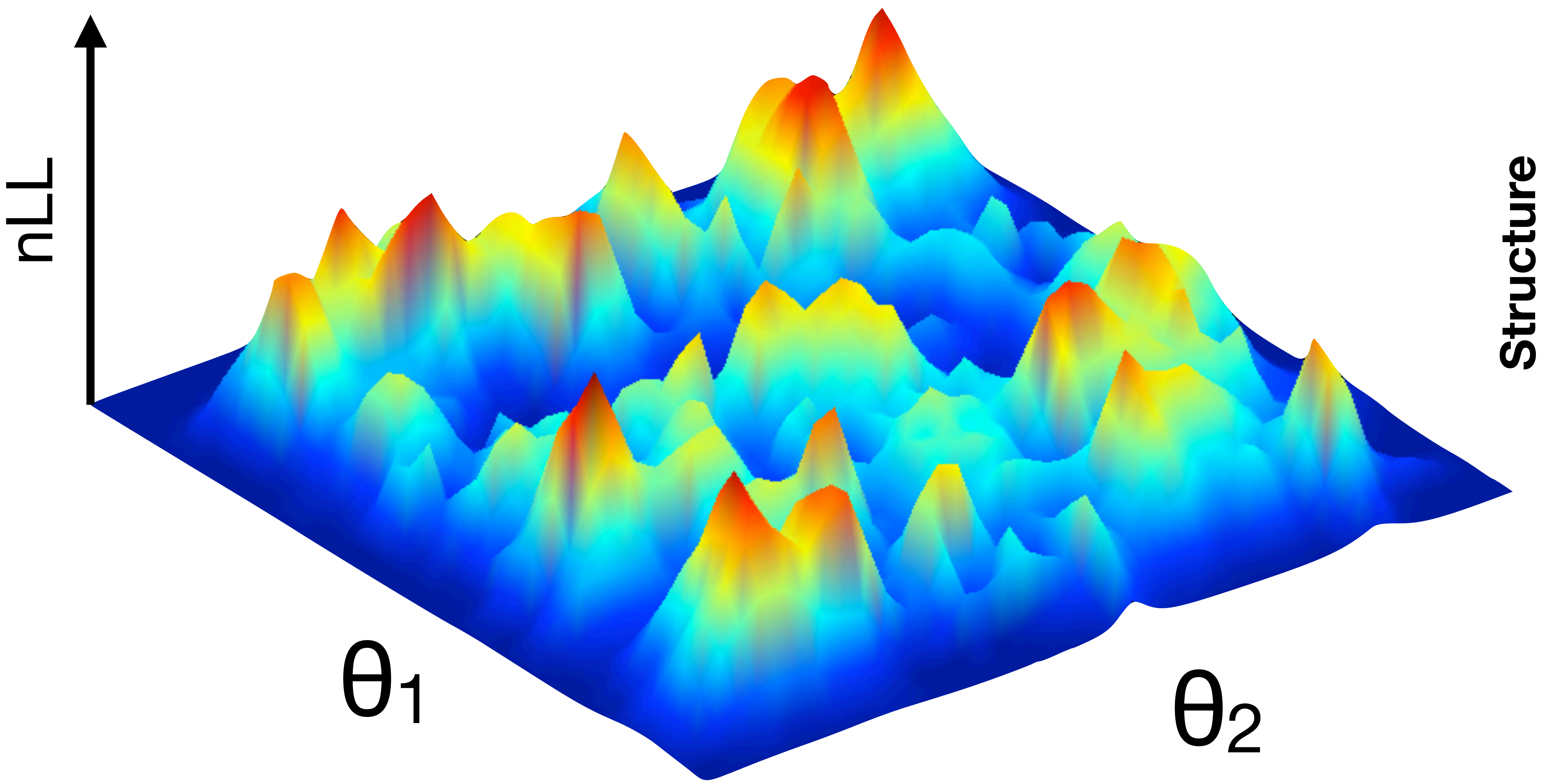


β

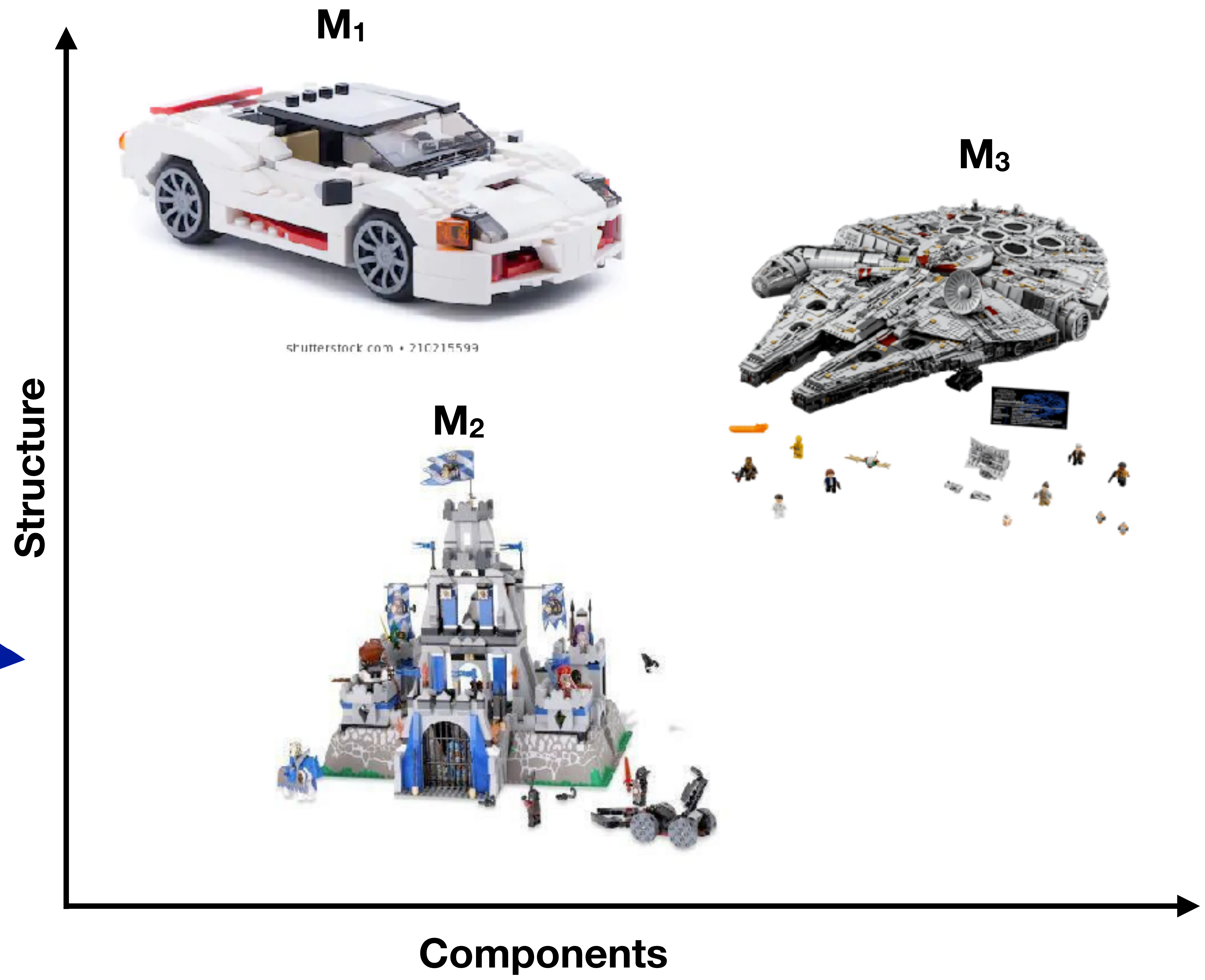
α

Parameter Space and Model Space

Parameter Space

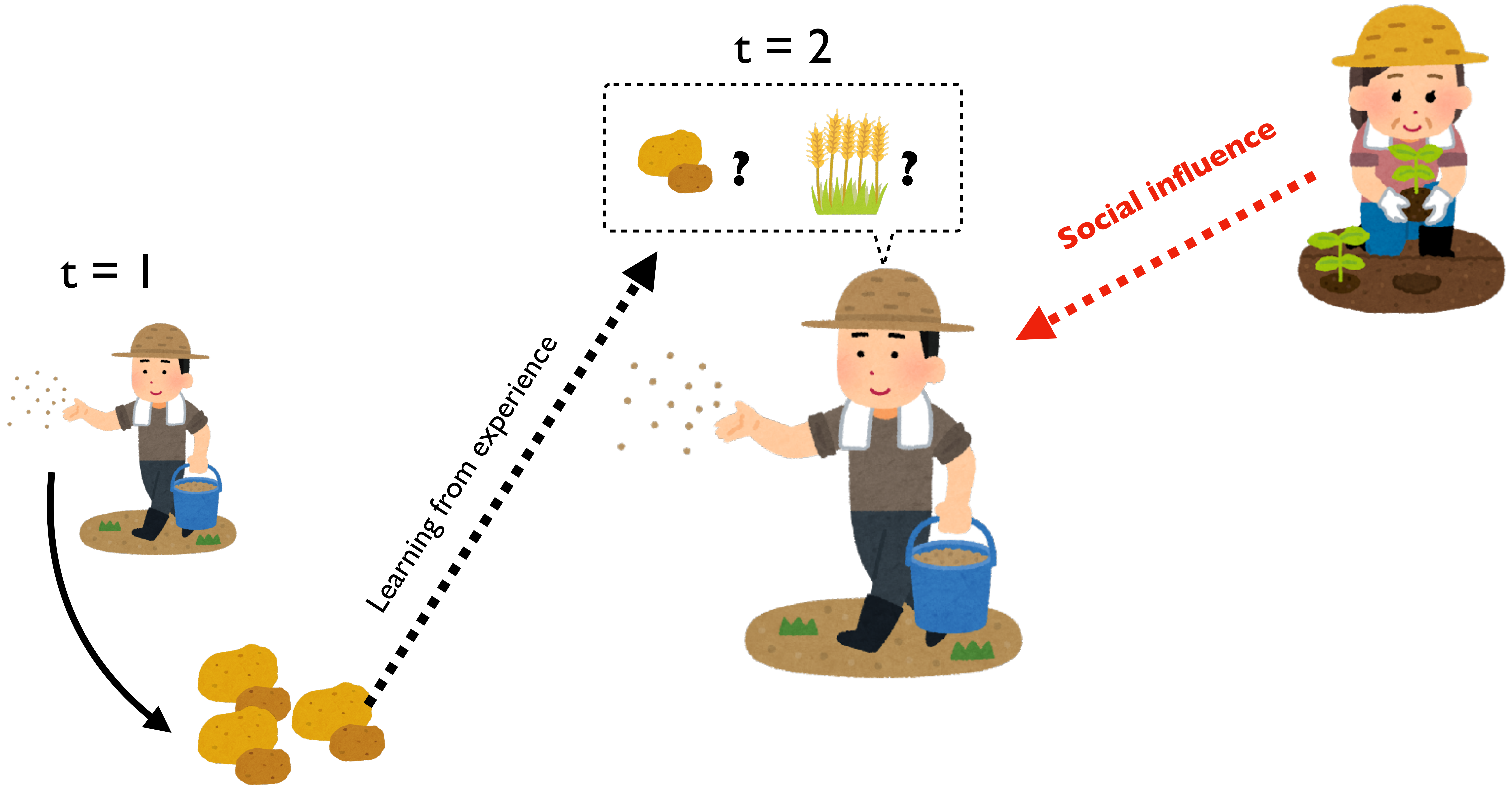


Model Space



Learning from social information

Learning from social information



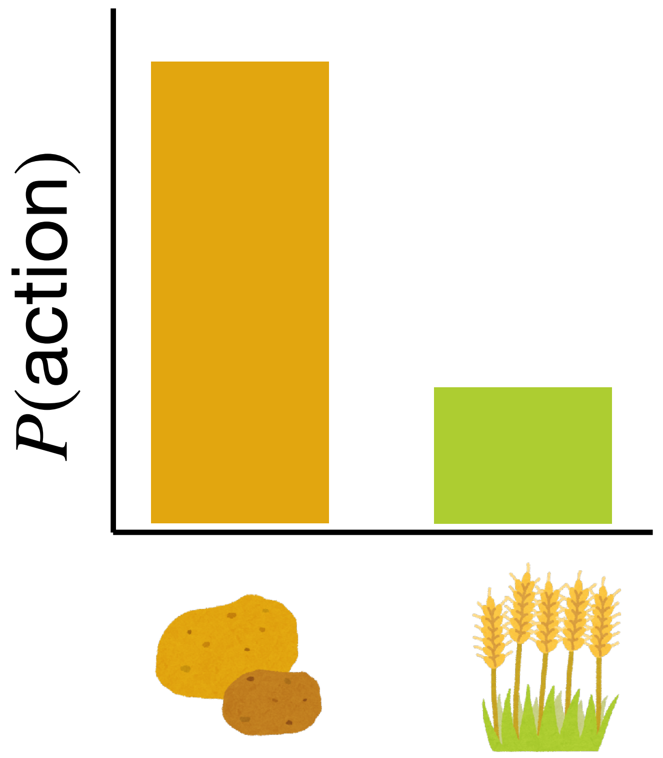
Imitating actions

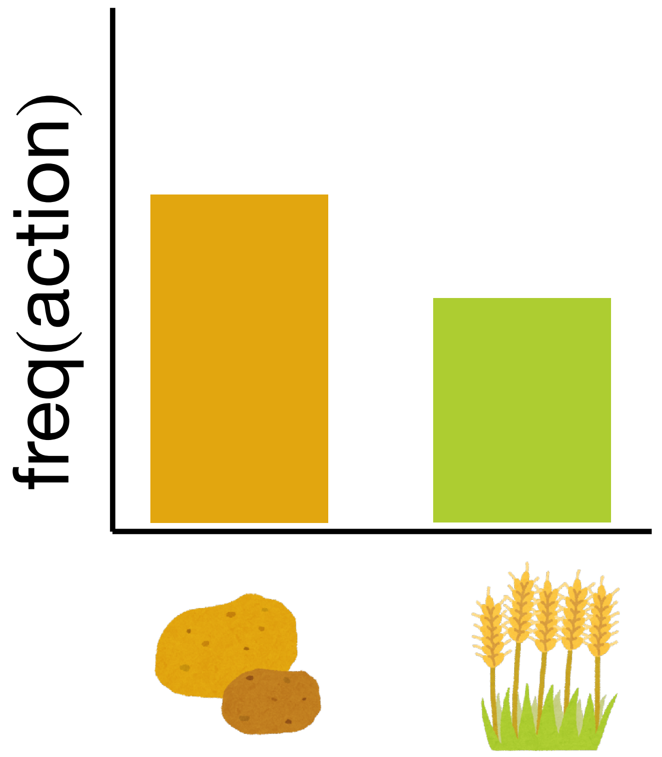
Frequency-dependent copying (FDC)

Probability of choosing option a $\propto$ frequency of other agents performing the same action

$$\pi_{\text{FDC}}(a) \propto \frac{f(a)^{\theta}}{\sum f(k)^{\theta}}$$

conformity exponent

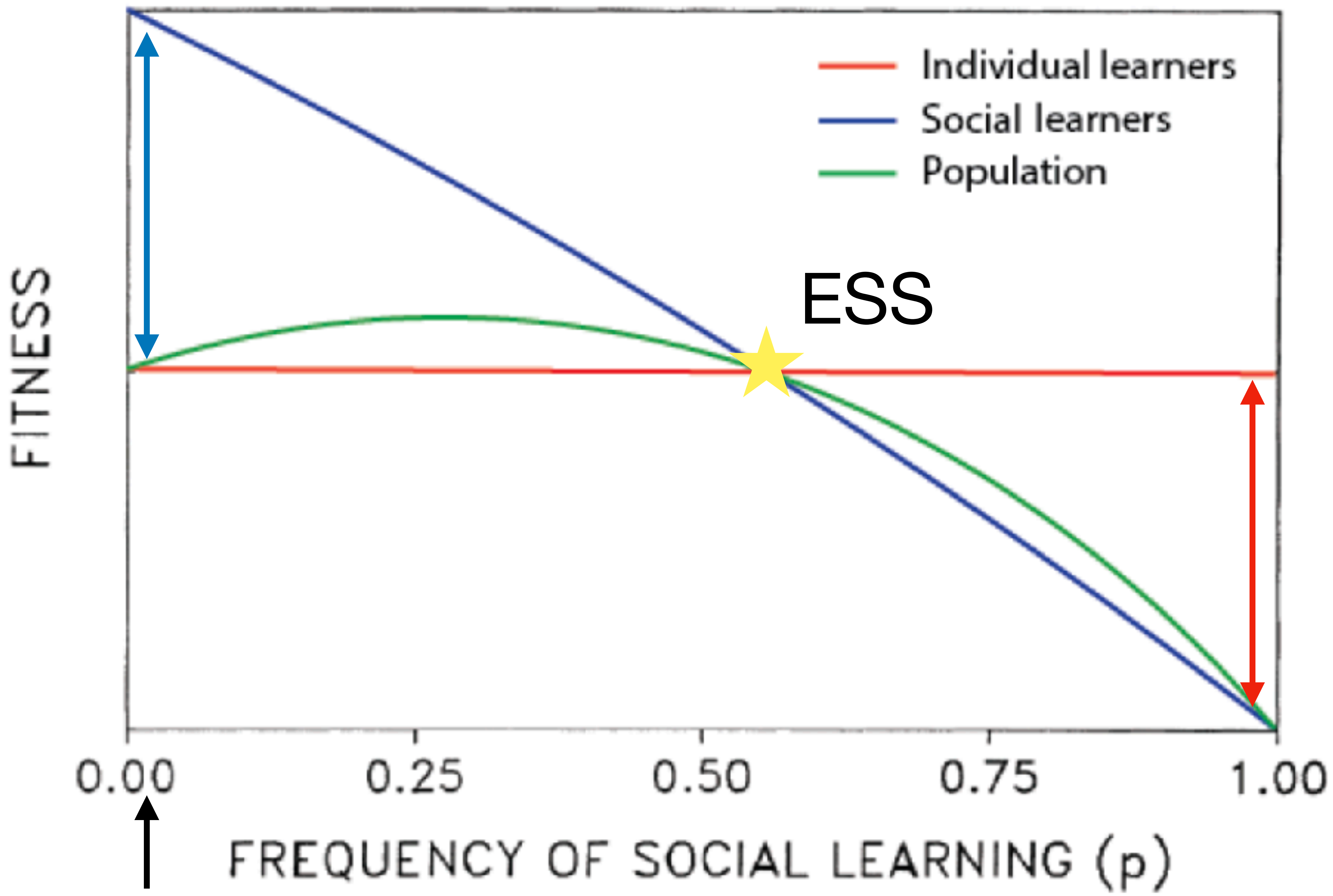




The wisdom and madness of crowds

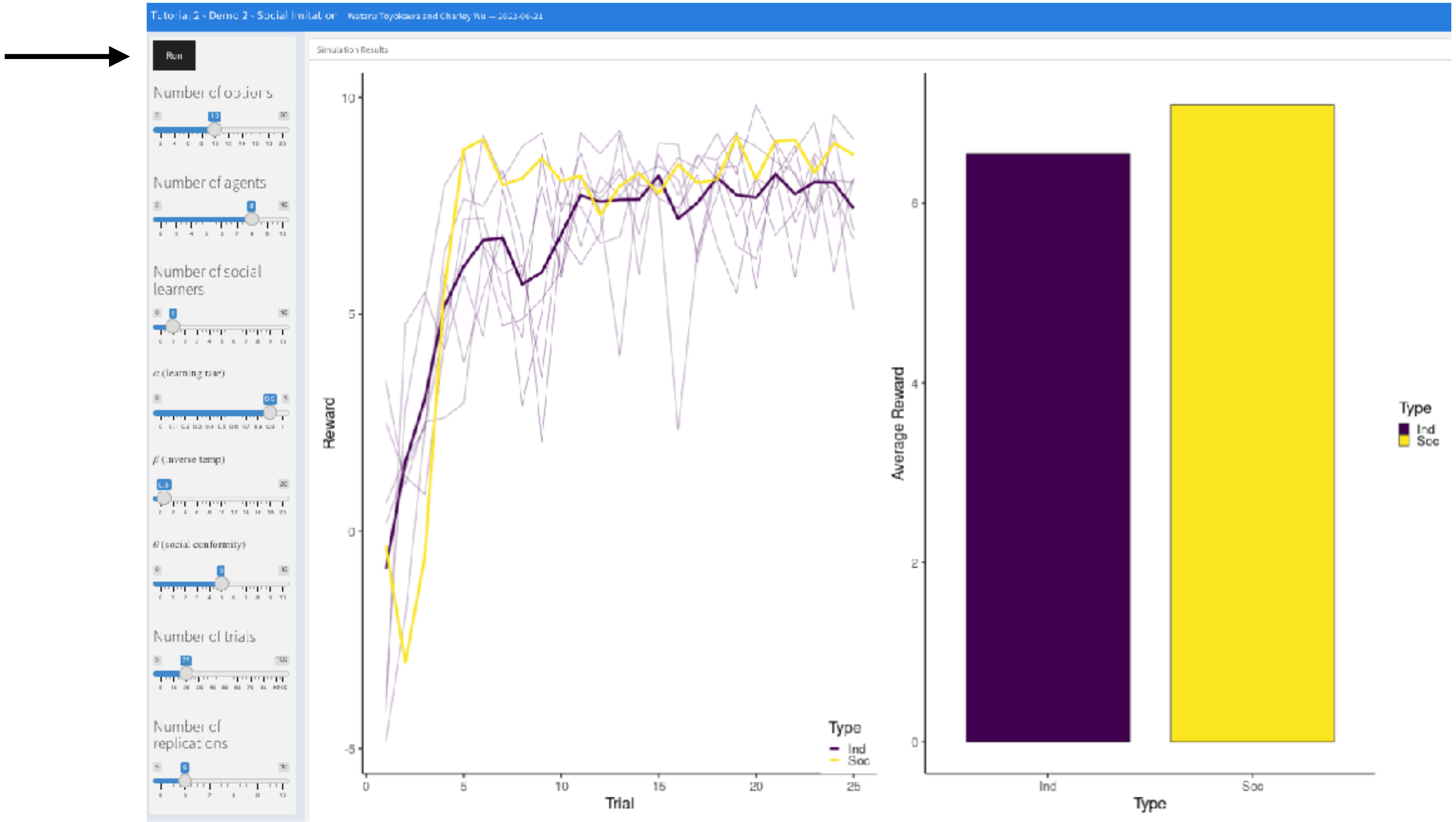
- Social learning via imitation has *frequency-dependent fitness*
  - High social learning fitness when *p* is small
  - Collapse in social learning fitness when *p* is large
- The best strategy depends on what other people do, creating a dynamic strategy selection problem
- Evolutionarily stable strategy (ESS) is an intermediate mixture

Rogers' (1988) paradox



Notebook https://cosmossummerschool.github.io/notebooks/tutorial-2-models-of-learning.html

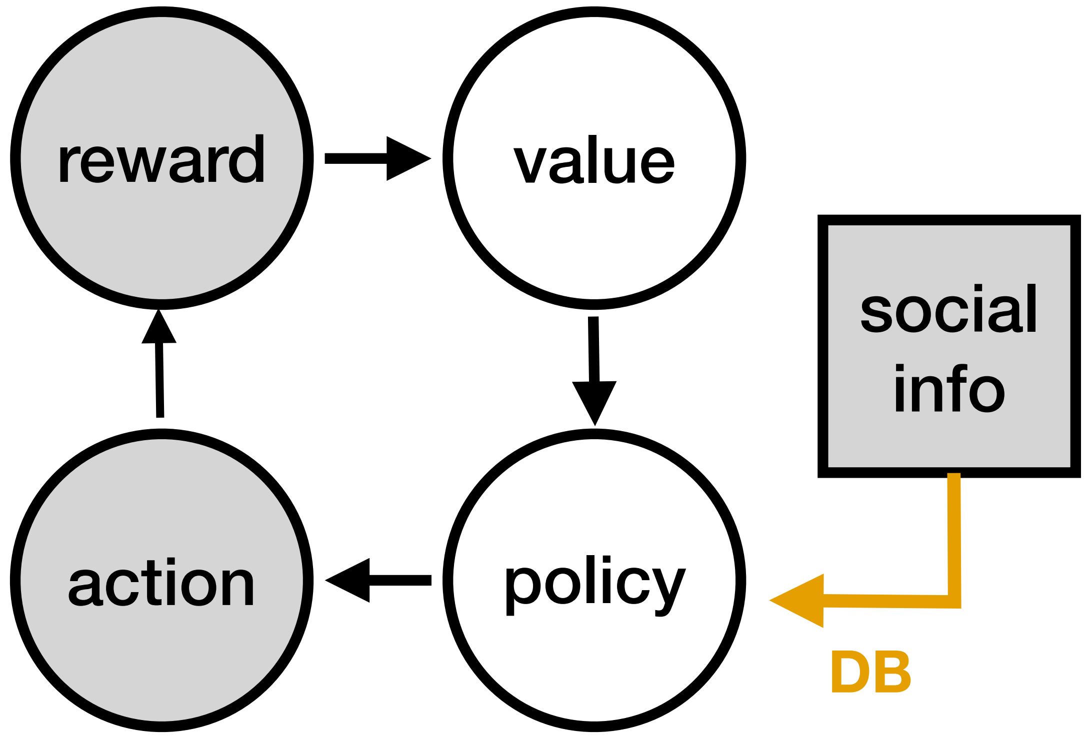
Demo 2: Imitation and Rogers' paradox



How do different ratios of individual vs. social learners change the performance of each agent type?

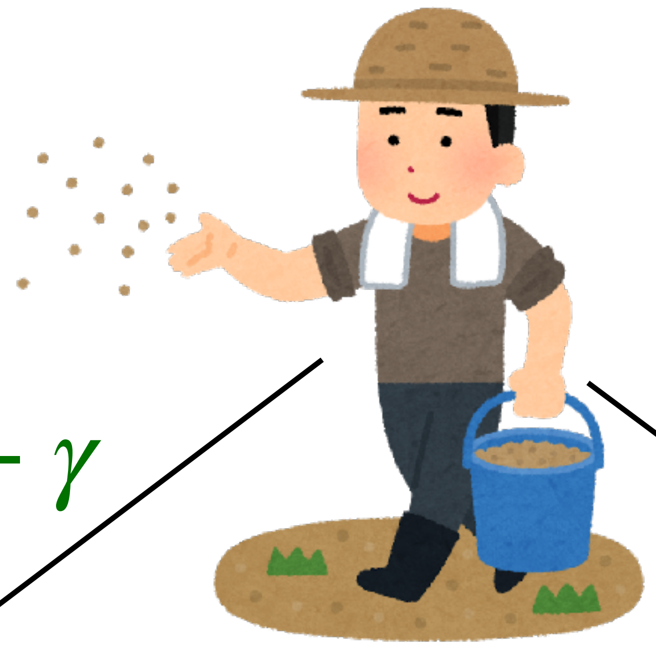
Combining imitation and value-learning

Decision-biasing (DB)



Mixture policy individual social

$$\pi_{DB} = (1-\gamma)\pi_{\text{Ind}} + \gamma\pi_{\text{Soc}}$$

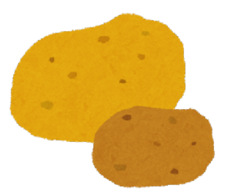
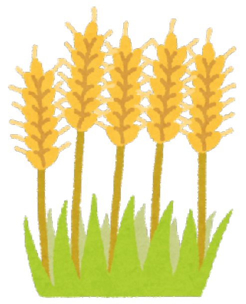


$1-\gamma$ γ

π_{Ind}

Q-learning + softmax

Q-values

 < 

$\exp[\beta Q_t(a)]$

π_{Soc}

Frequency-dependent copying

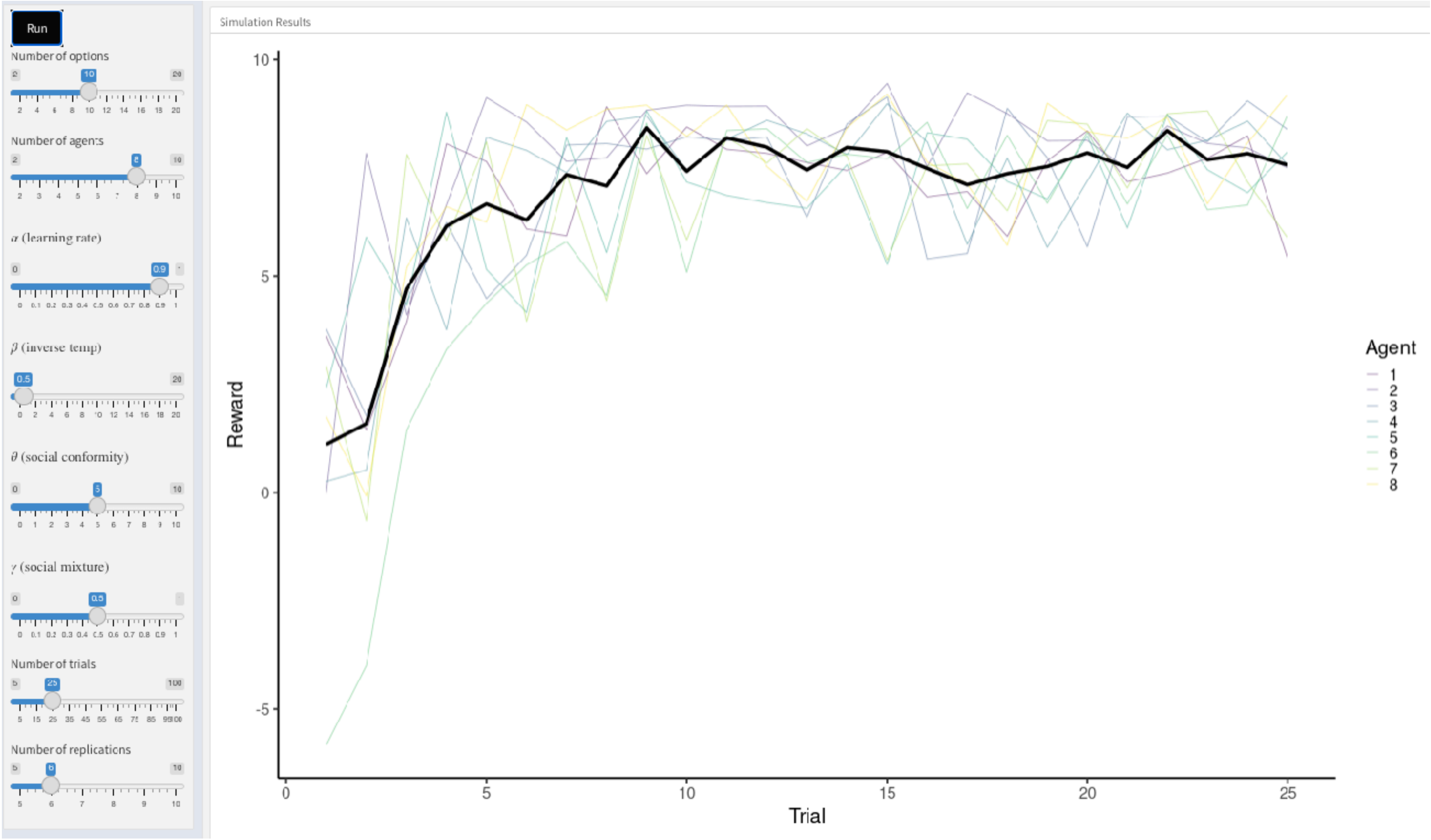
 > 

$$\frac{f(i)^\theta}{\sum_k f(k)^\theta}$$

Toyokawa et al., (NHB 2019)

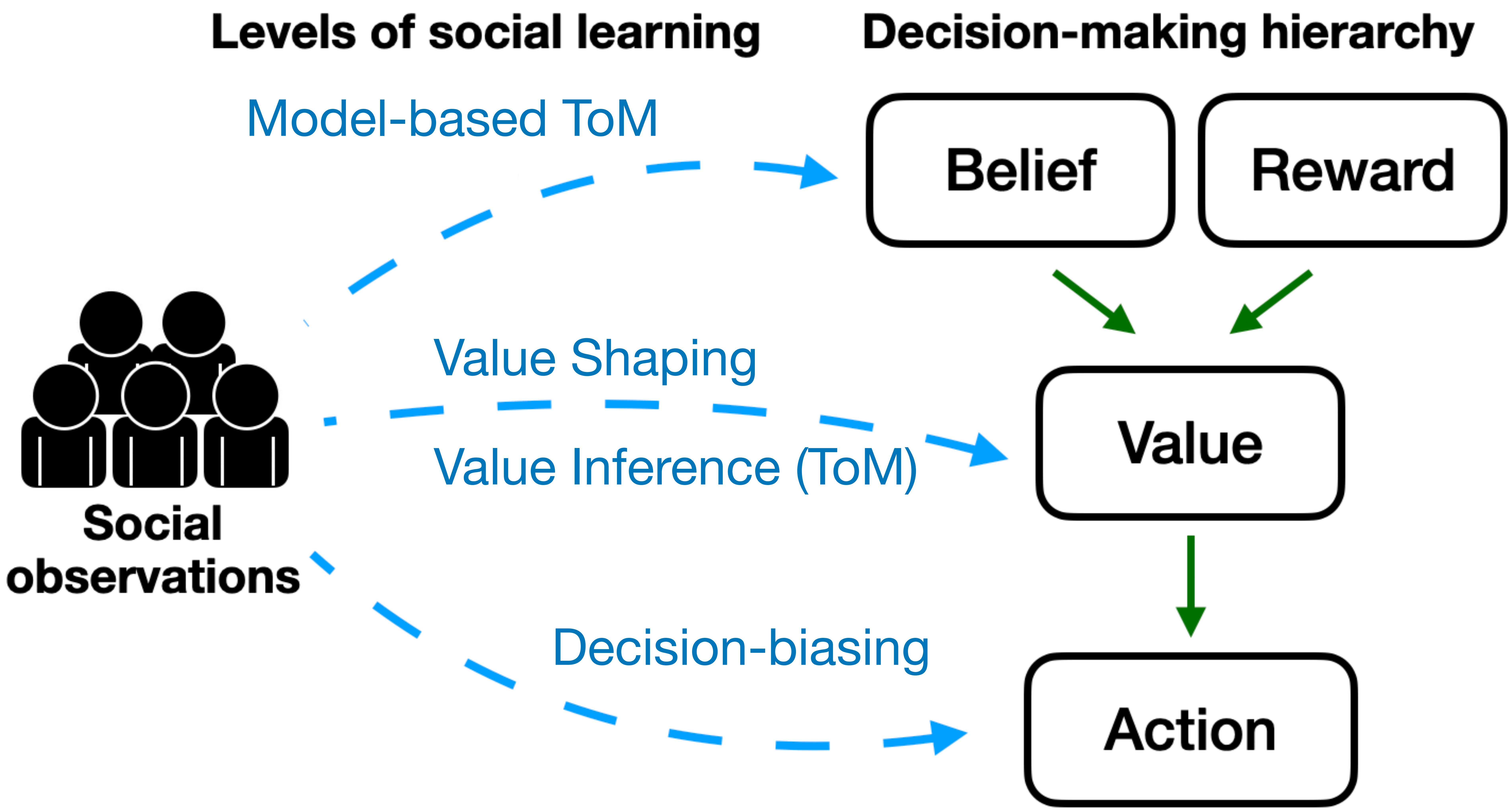
Notebook https://cosmossummerschool.github.io/notebooks/tutorial-2-models-of-learning.html

Demo 3: Decision-biasing



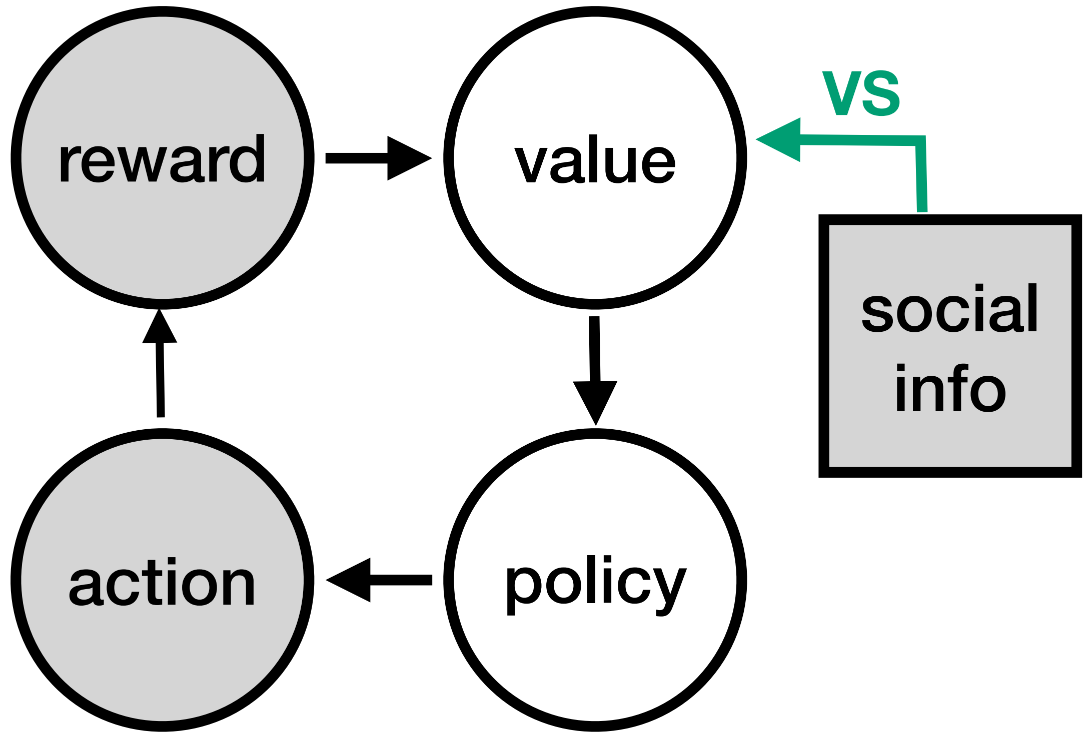
Which values of γ (social mixture) and θ (conformity exponent) typically produce the best results?

Social influence at different levels of learning



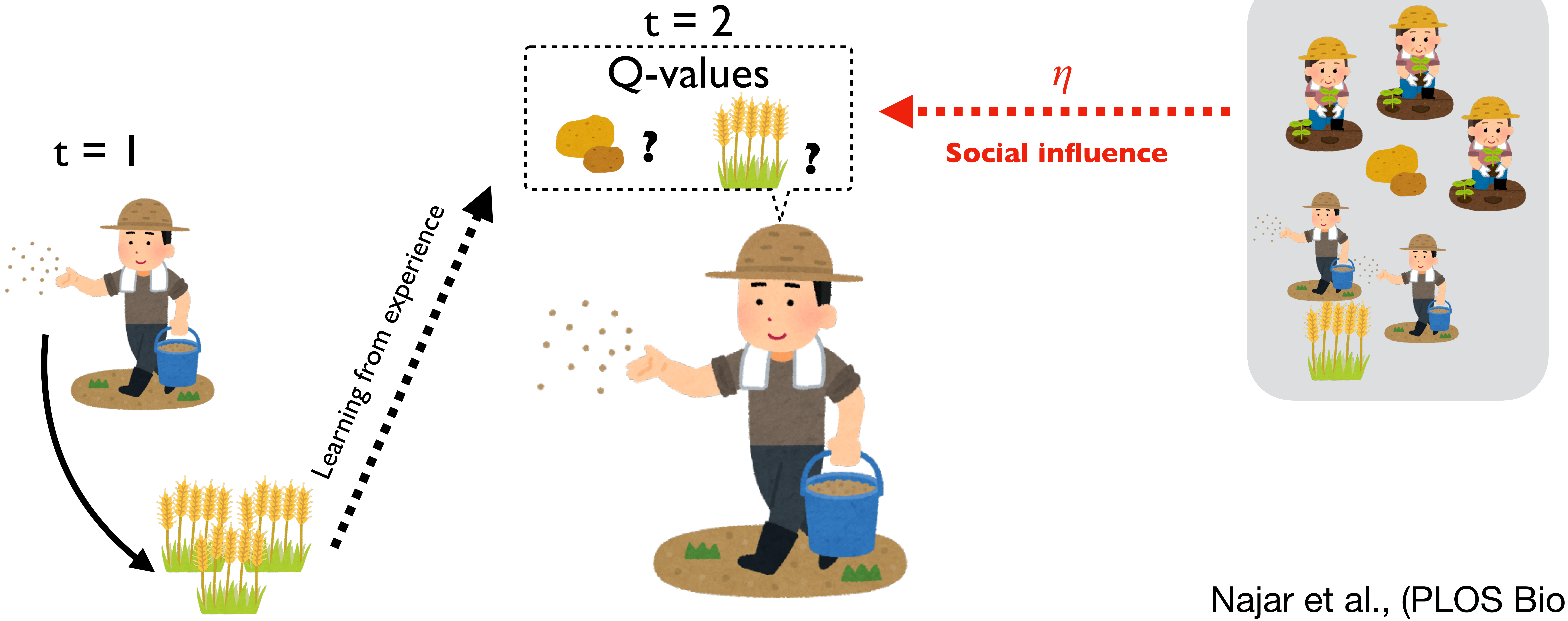
Wu, Vélez, & Cushman (2022)

Value-shaping (VS)



value bonus

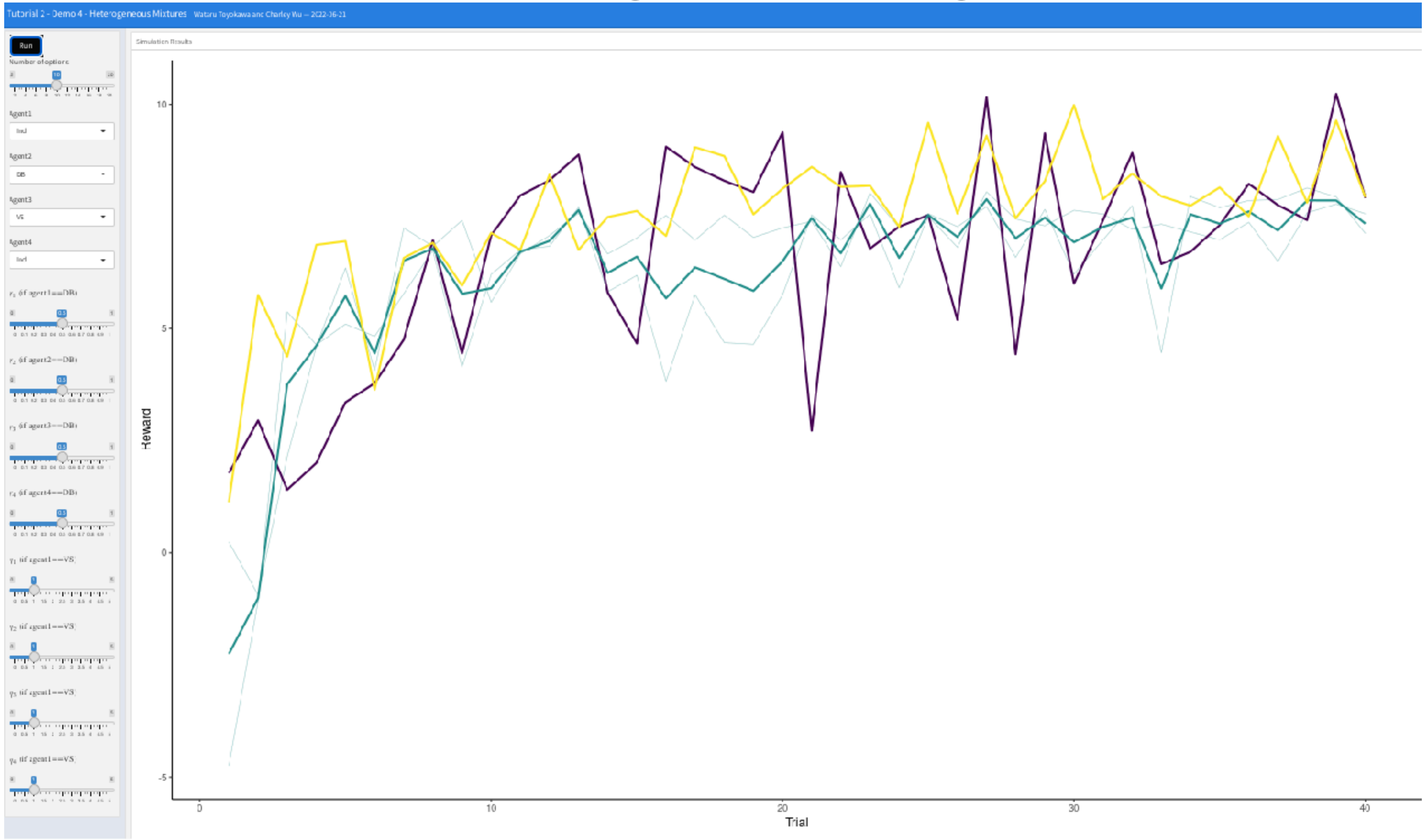
$$Q(a) \leftarrow Q(a) + \eta f(a)$$



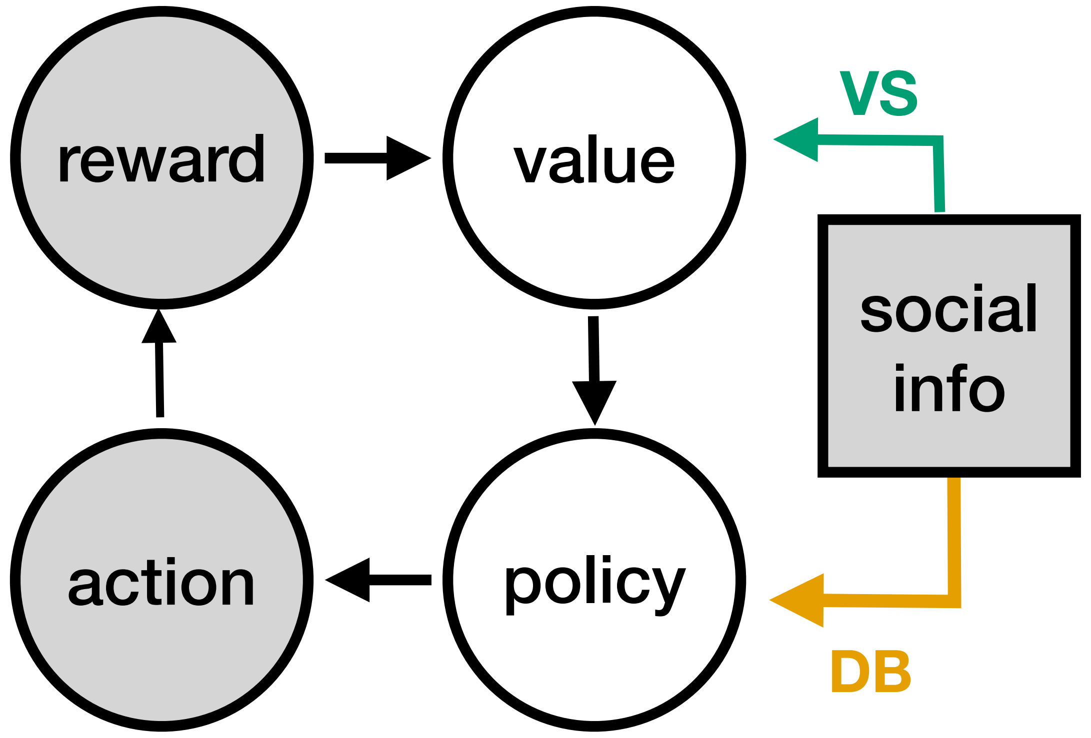
Najar et al., (PLOS Bio 2020)

Notebook https://cosmos-konstanz.github.io/notebooks/tutorial-2-models-of-learning.html#value-shaping

Demo 4: Heterogeneous groups

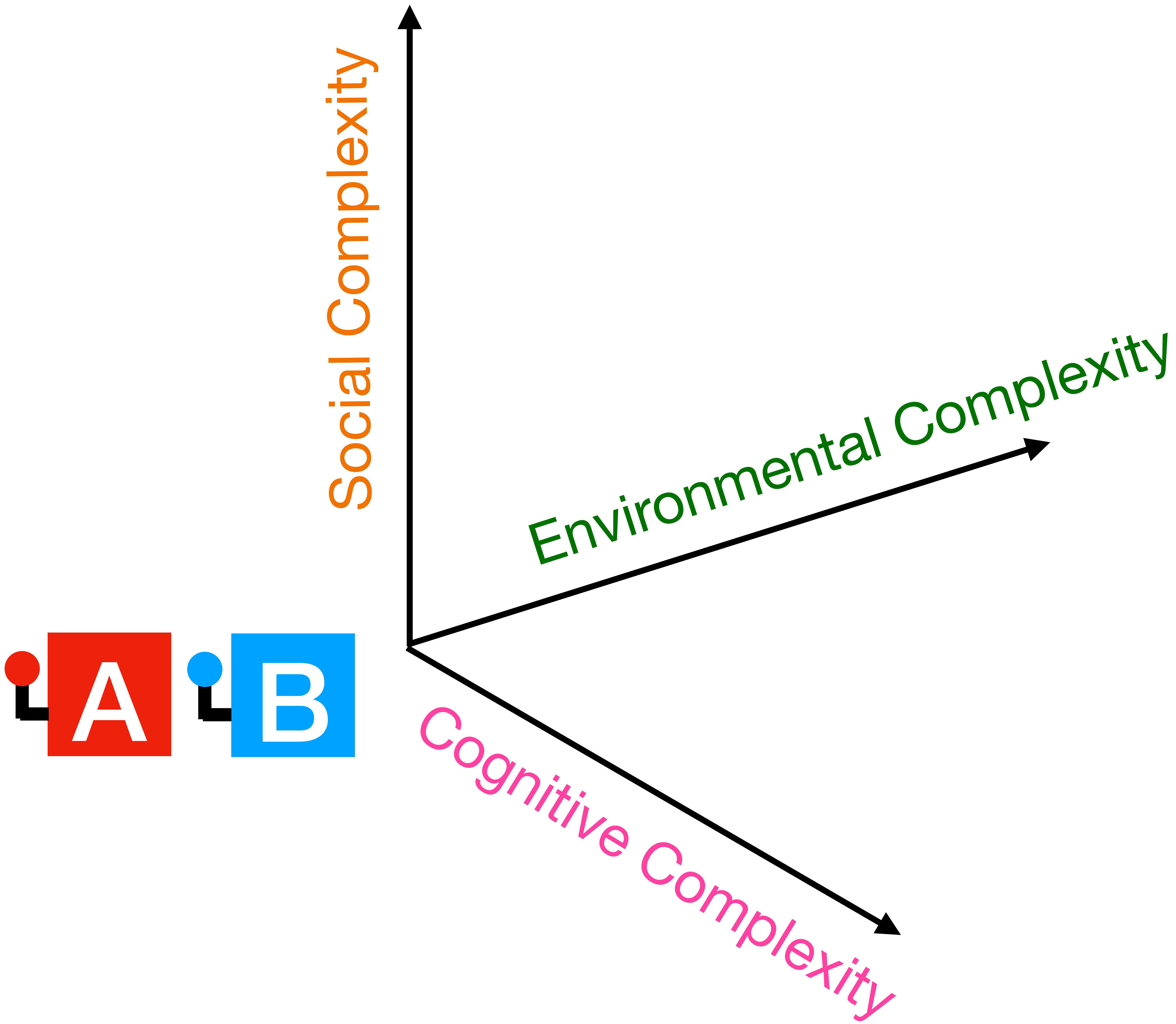


Ind vs **DB** vs **VS**



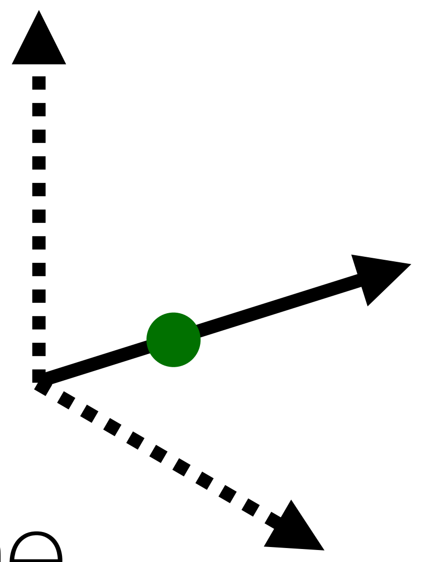
Which strategy performs better than others? Is it robust to different group compositions?

Scaling up to more diverse social learning tasks

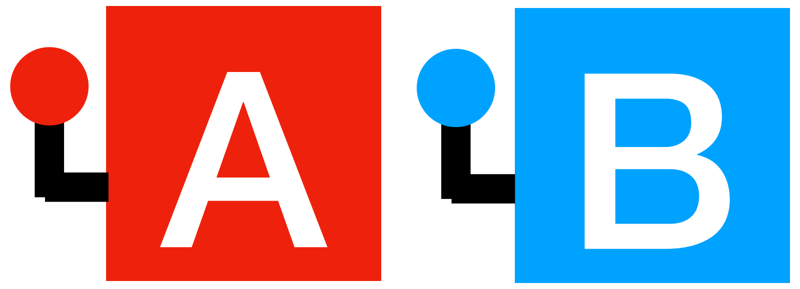


- Environmental complexity
  - Reward structure
  - Temporal dynamics
- Social complexity
  - Network structures
  - Incentives
- Cognitive complexity
  - imitation vs. emulation (ToM)
  - Pedagogy vs. observational learning

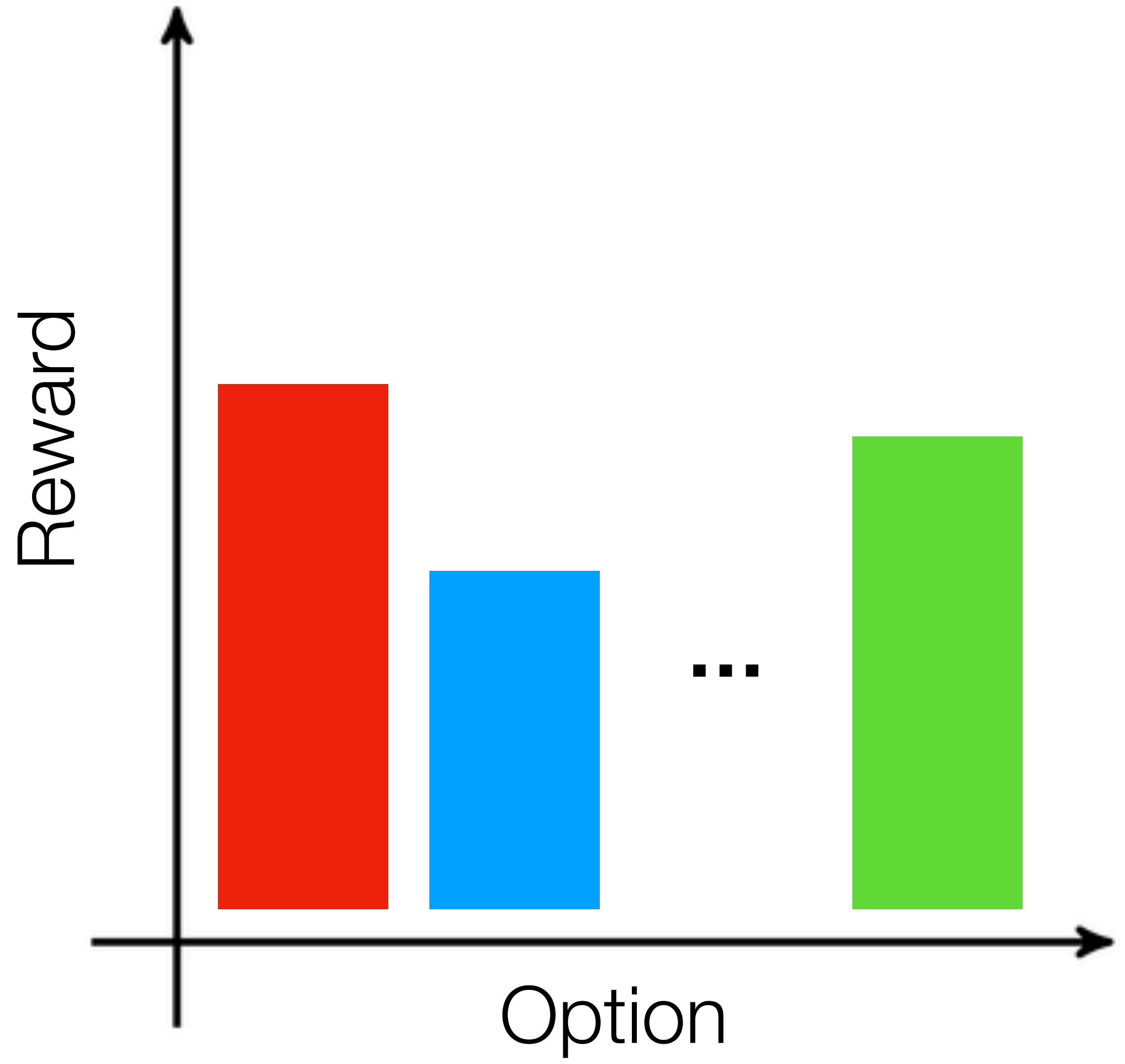
Environmental Complexity: Structured rewards



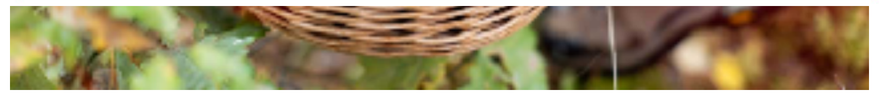
- Rather than assuming each action has independent rewards (**Tabular RL**), we can leverage the structure of the environment to *generalize* from familiar to new situations (**Function Approximation**)
- *Gaussian Process* (GP) regression is a versatile framework for generalization using Bayesian function approximation
  - Stimuli in similar locations, with similar features, or with similar network connections are expected to yield similar rewards



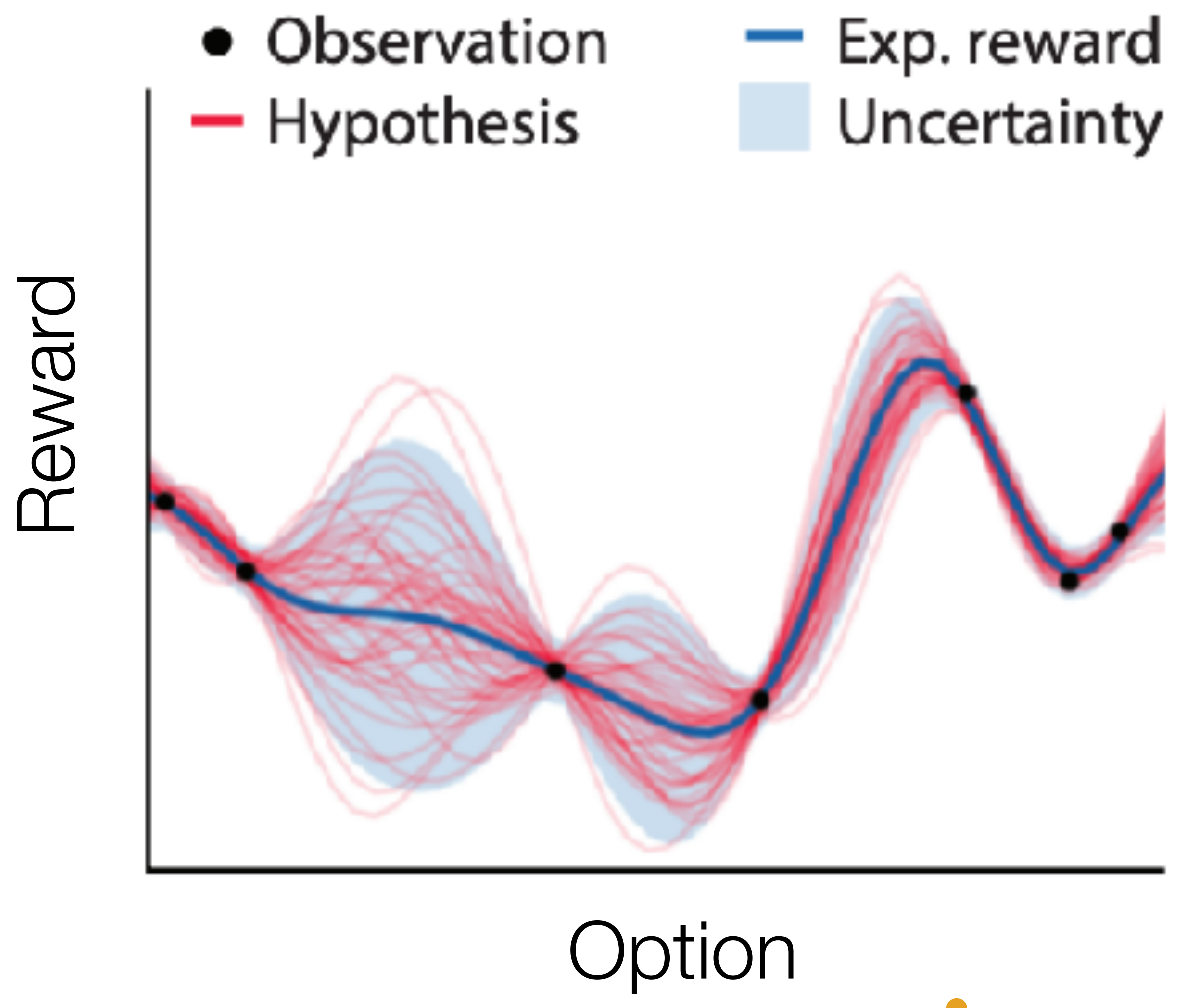
Tabular RL

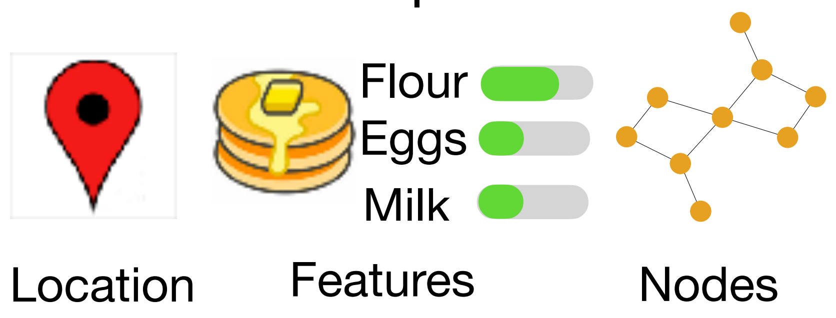


7	5	10	22	32	32	28	24	22	26	33
6	11	19	29	38	41	42	40	37	36	40
22	27	30	35	43	50	53	53	51	49	46
45	44	38	36	40	46	47	49	54	55	48
61	55	46	40	37	32	27	31	44	52	44
62	59	57	54	44	27	14	17	33	46	45
53	59	68	71	59	36	17	15	28	45	51
46	57	71	77	67	47	26	18	27	45	56
45	56	65	67	60	46	29	20	27	42	55
51	57	58	53	47	40	30	23	28	40	49
60	62	58	47	39	38	35	31	35	41	46



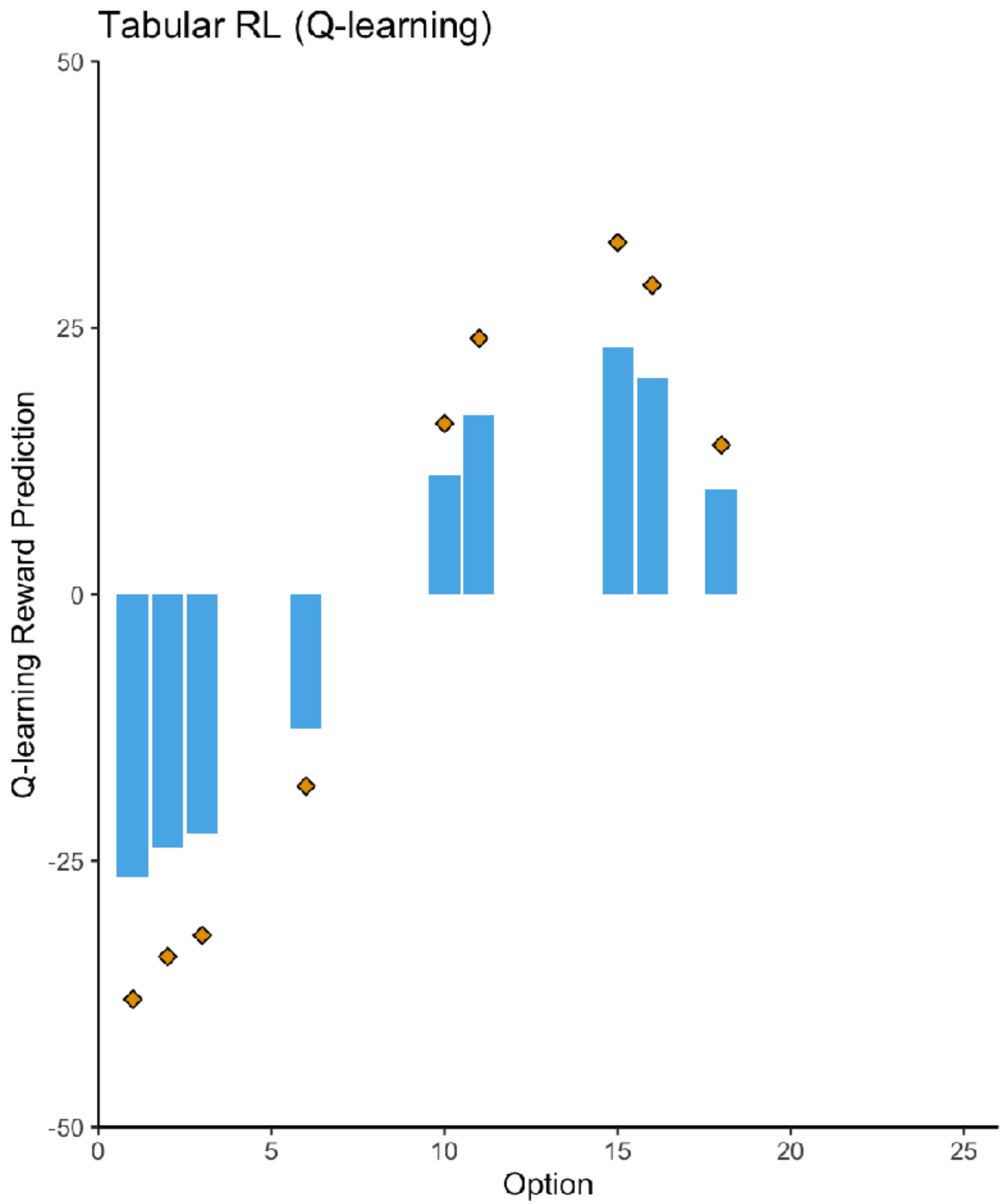
Function Approximation

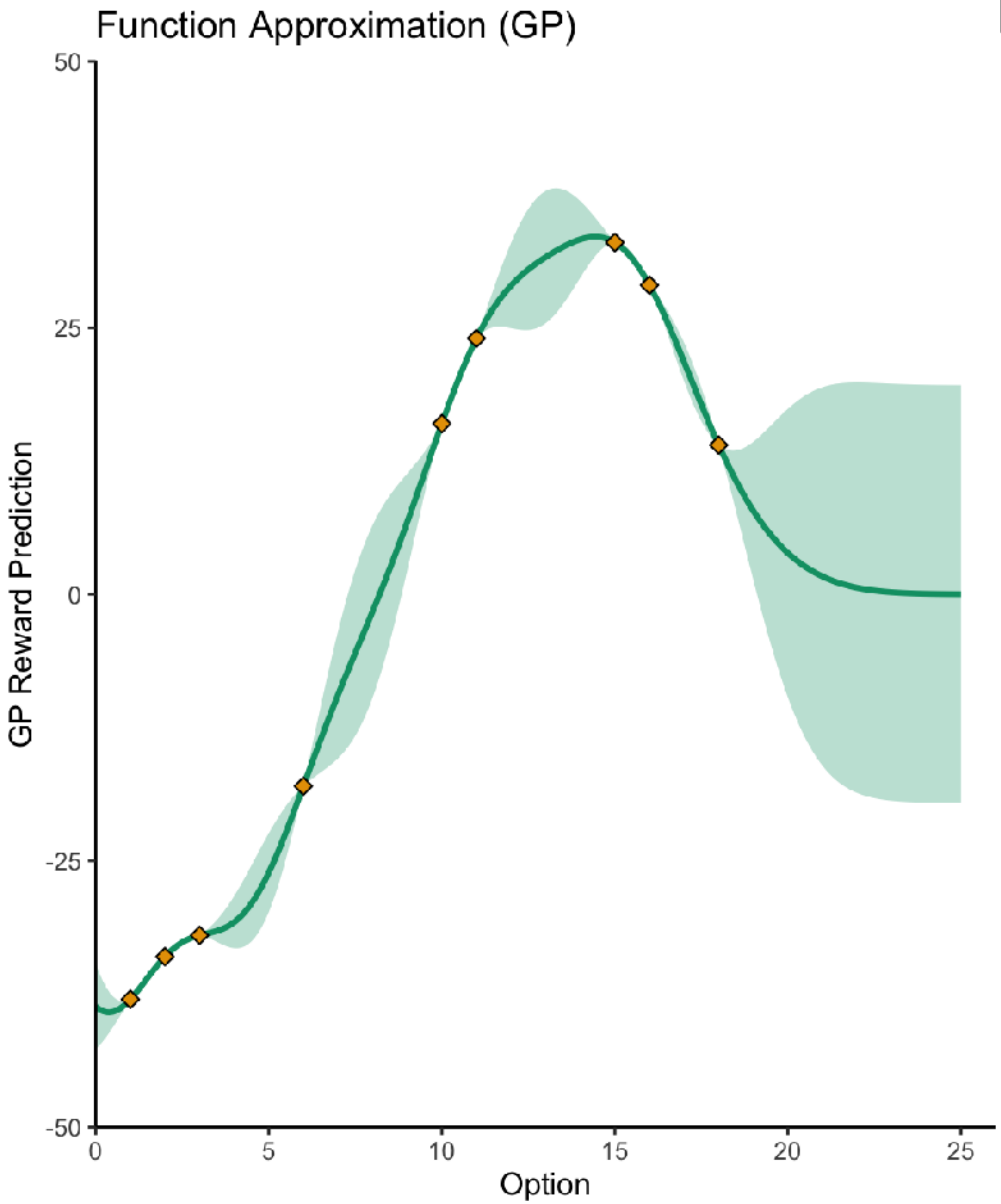




Wu et al., (*AnnRevsPsy* 2025); Wu et al., (*NHB* 2018); Wu et al., (*PLOS CB* 2020); Wu et al., (*CBB* 2021)

Notebook https://cosmossummerschool.github.io/notebooks/tutorial-2-models-of-learning.html

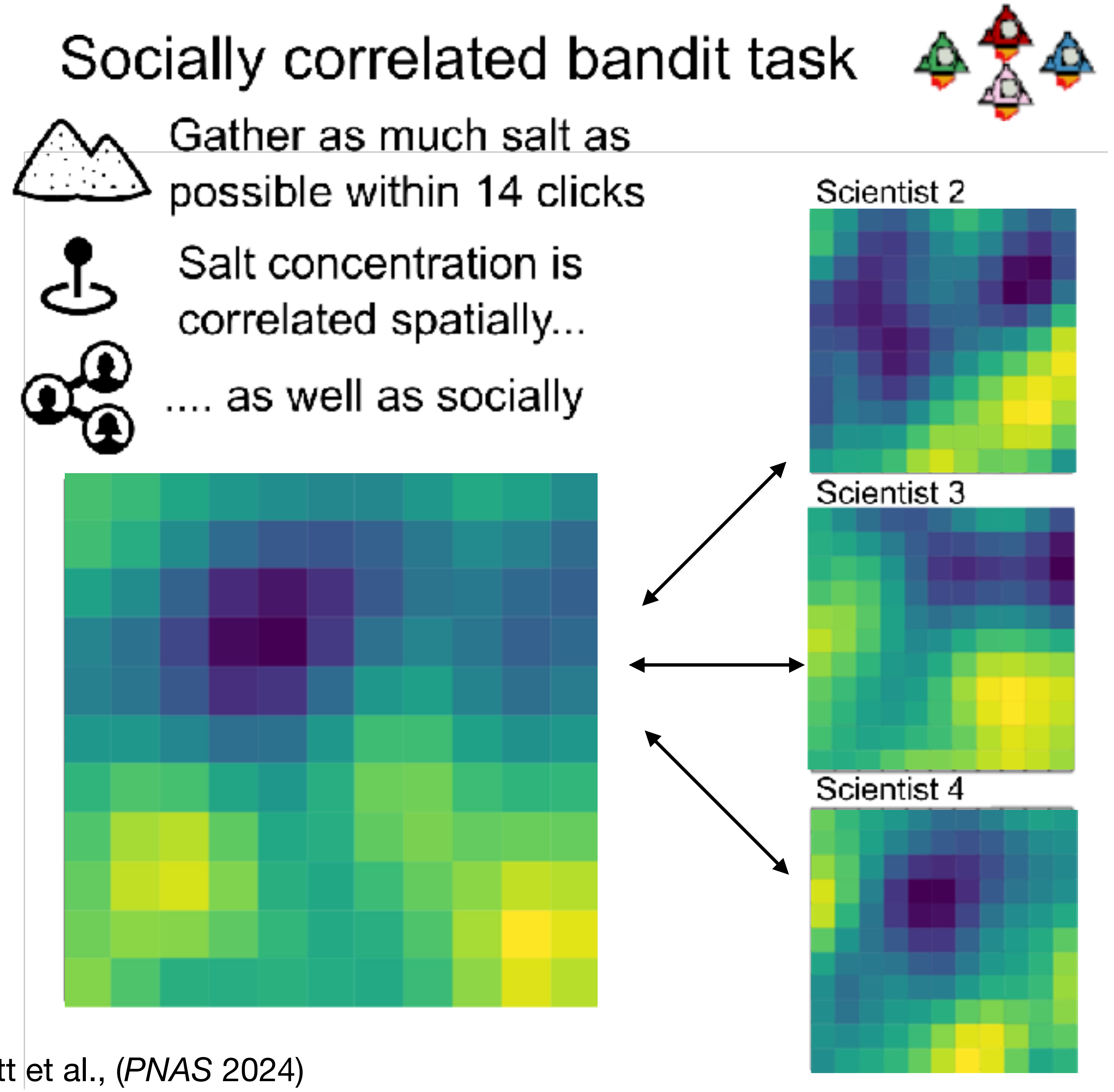


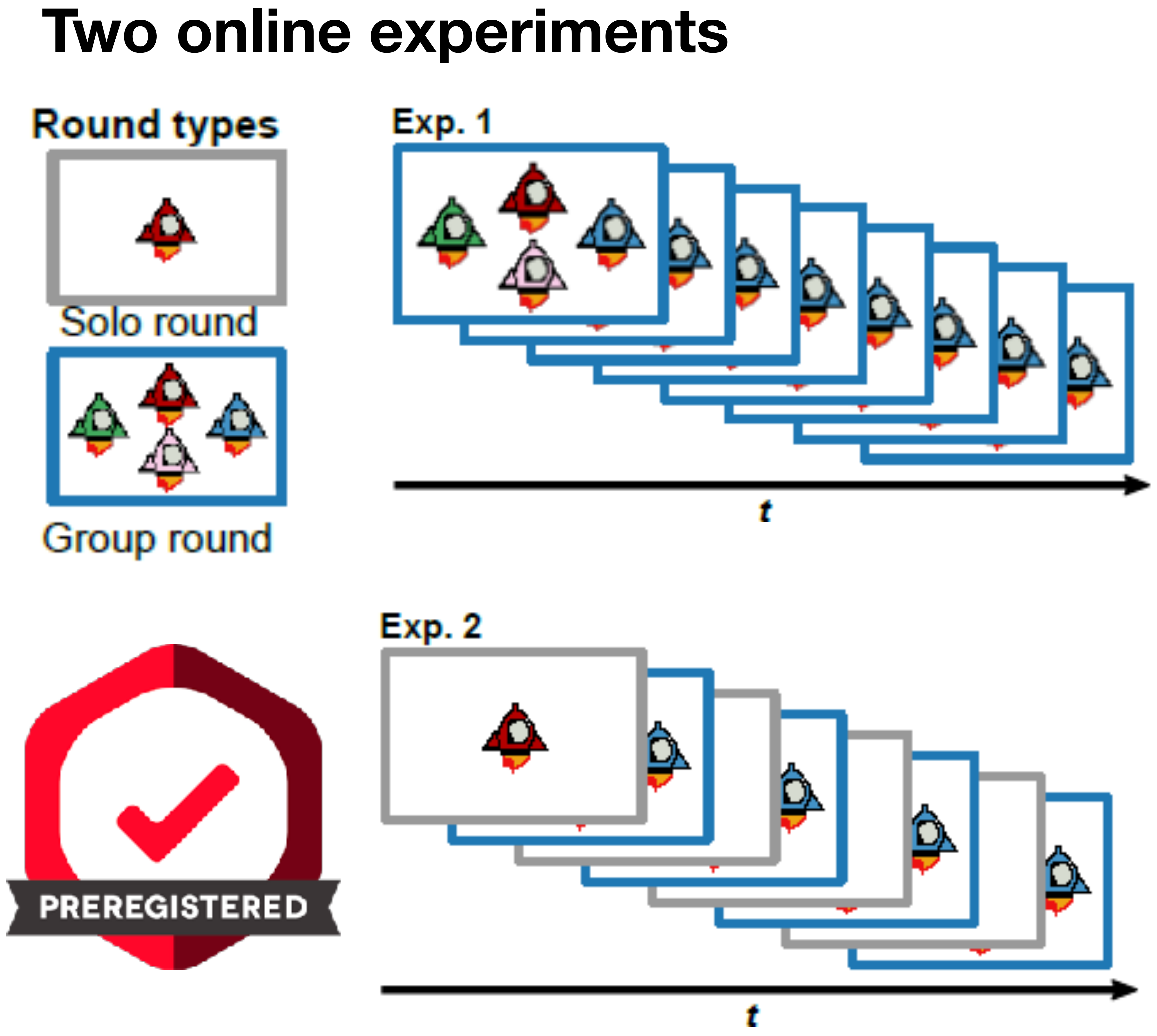


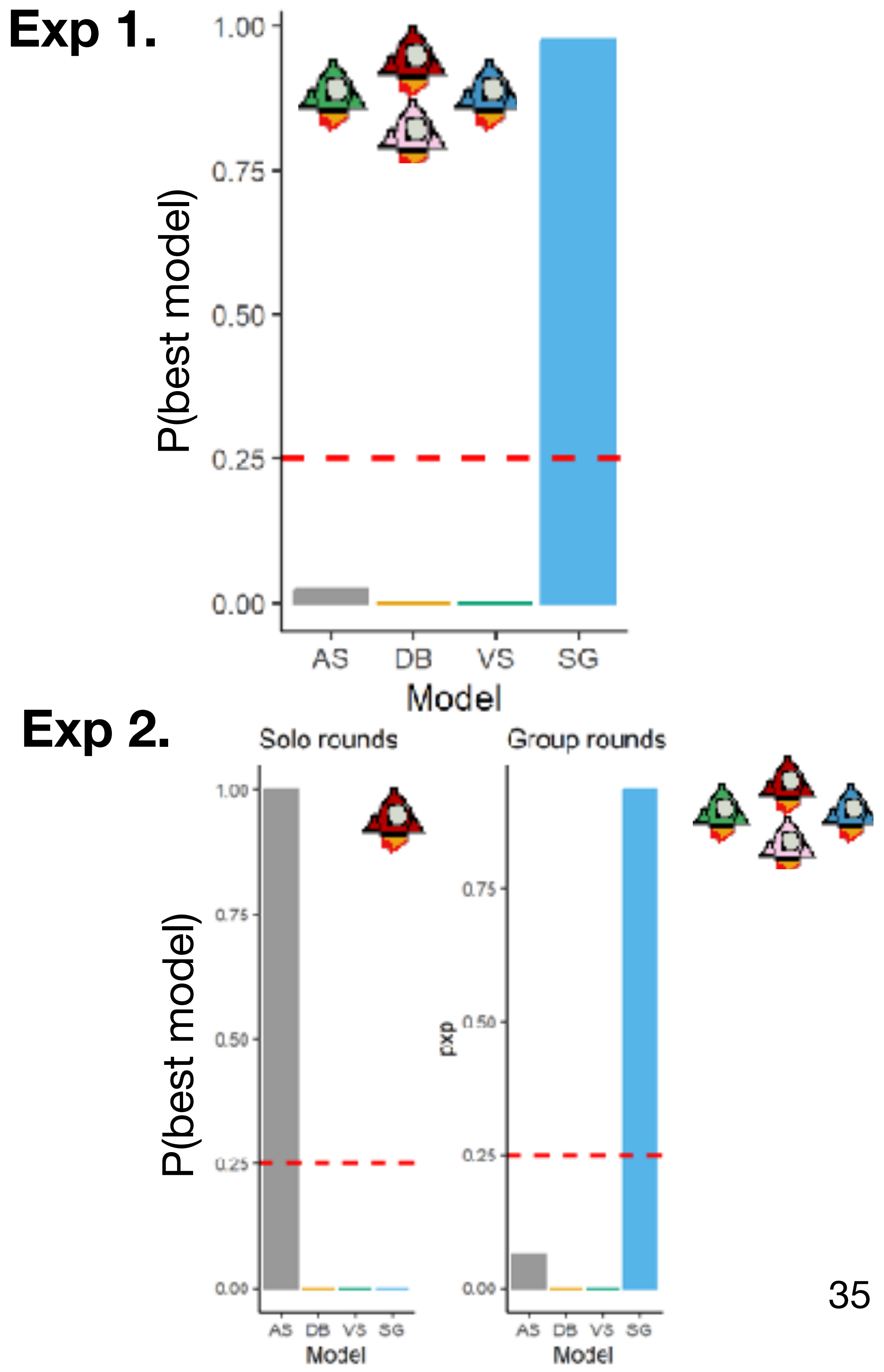
Environmental Complexity: Structured rewards

Alexandra Witt

- Social generalization
  - Social info from people with different preferences/goals should be "taken with a grain of salt" rather than used verbatim

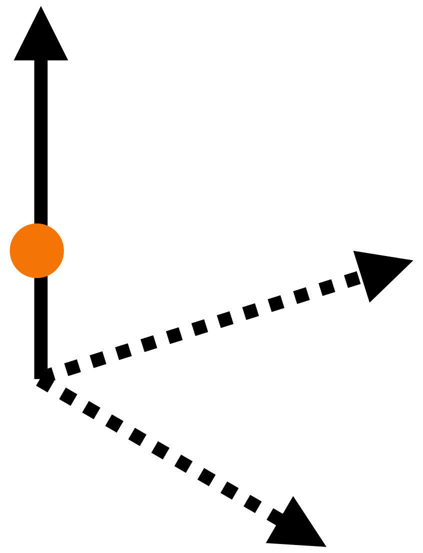






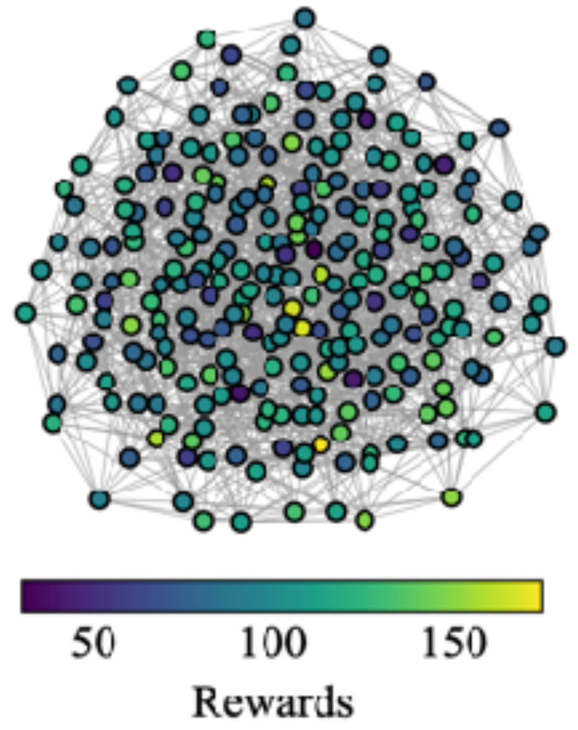
Witt et al., (*PNAS* 2024)

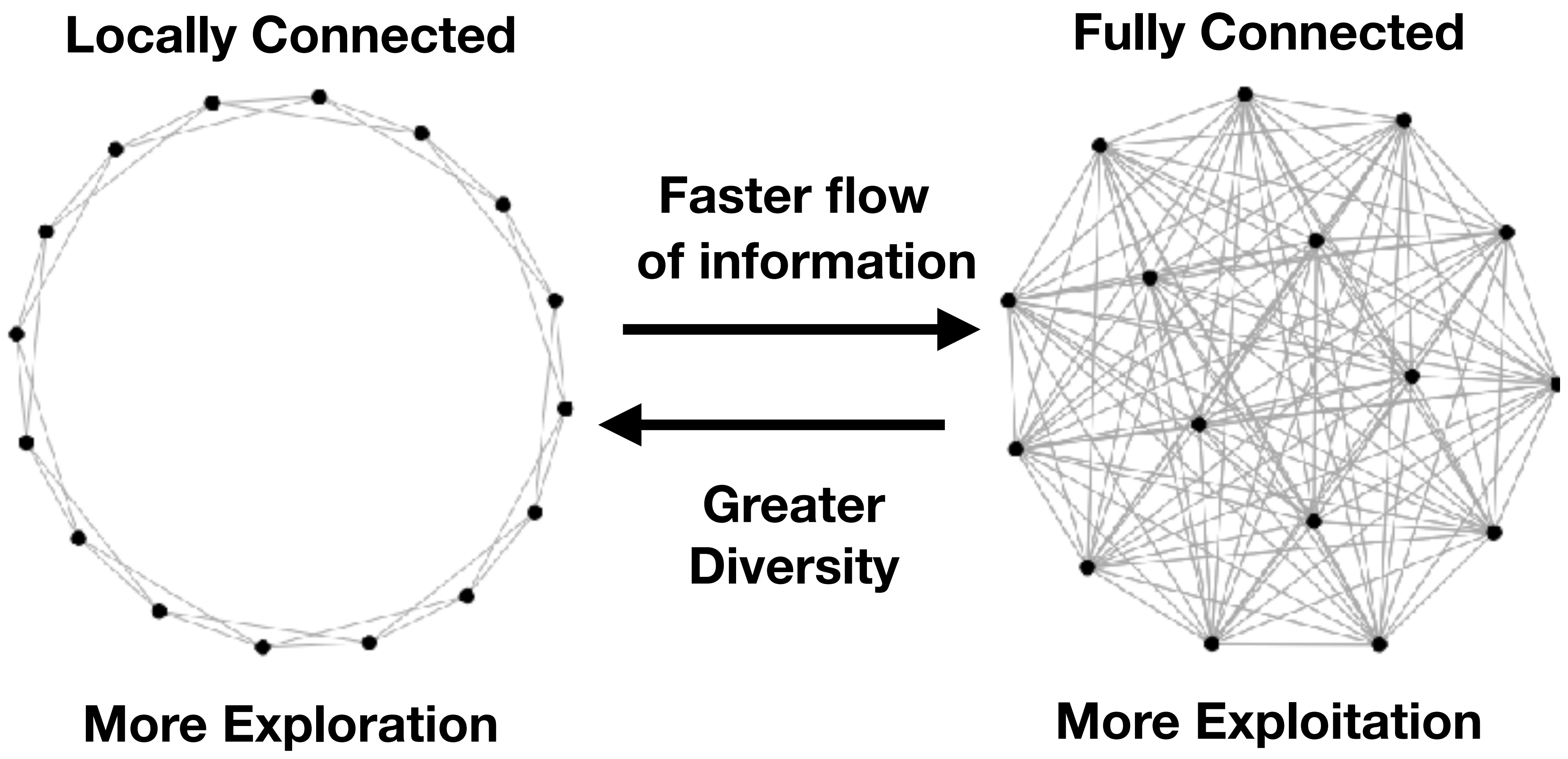
Social Complexity: Network structures



- Network structures shape explore-exploit trade-off at the group level
  - Locally connected networks maintain greater diversity, and facilitate more **exploration** at the group level
  - Fully connected networks lead to rapid convergence and faster **exploitation**
- Optimal balance depends on the environment and task

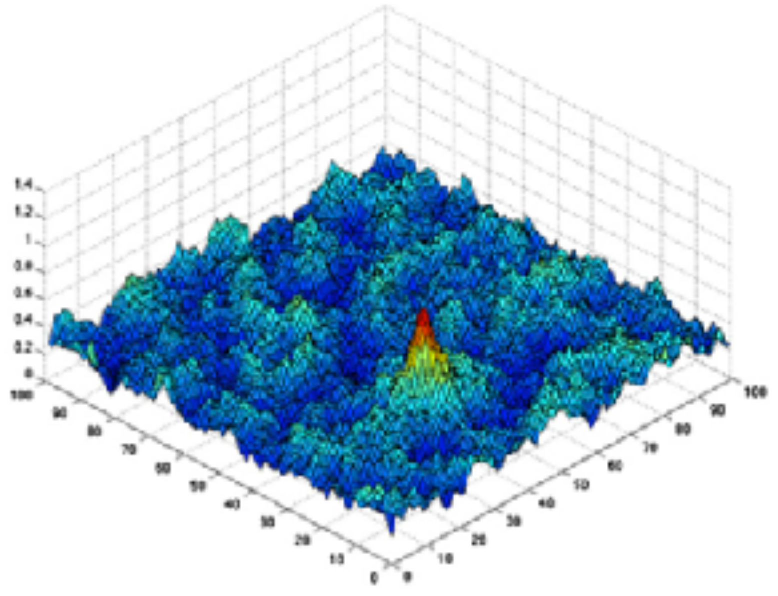
NK environments
(Lazer & Friedman, 2007; Szőcs et al., 2025)



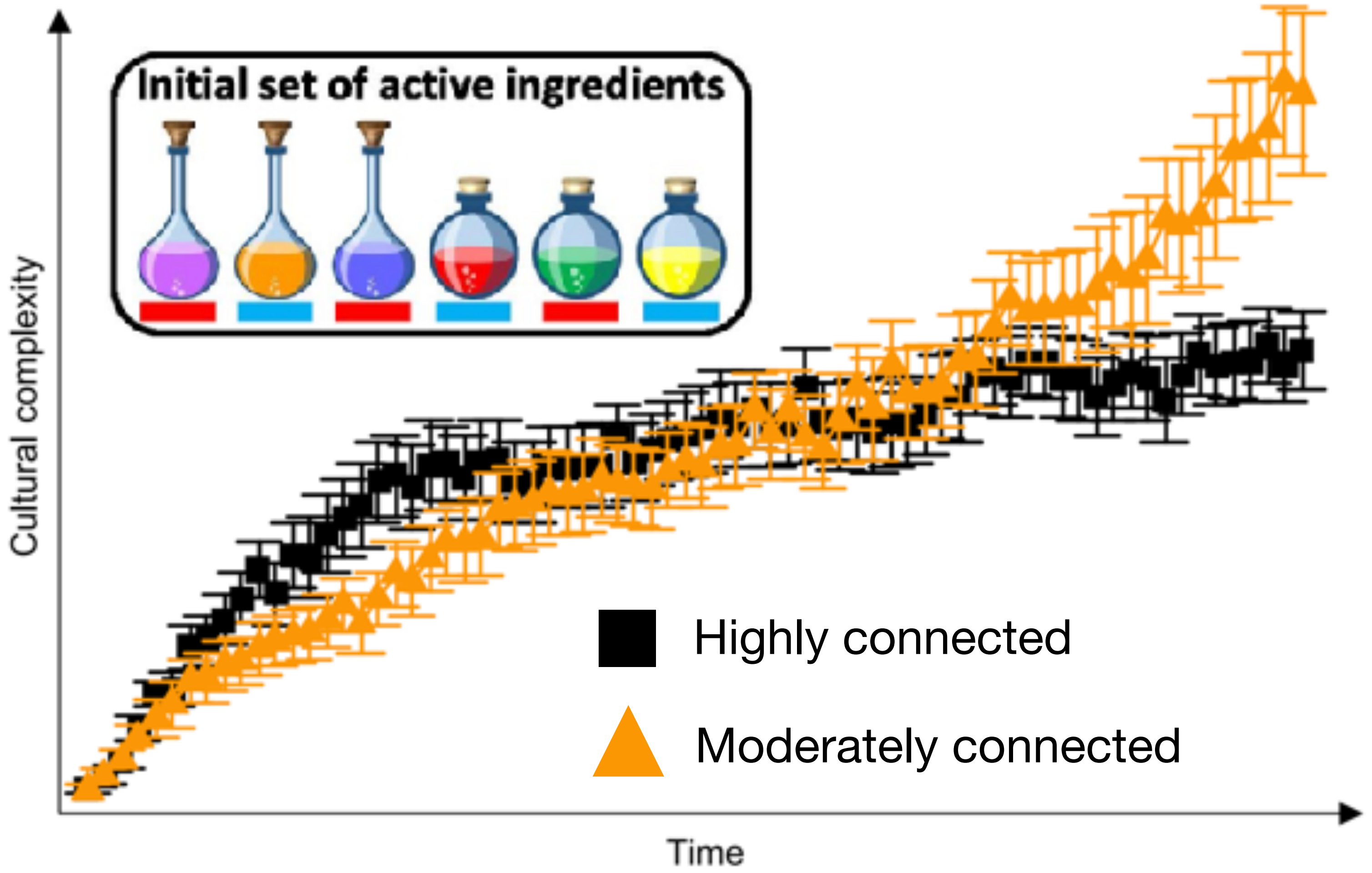


Barkoczi, Analytis & Wu (2016); Xi & Toyokawa (2025)

Rugged landscapes
(Mason & Watts, 2012)



Network structure influences cultural complexity

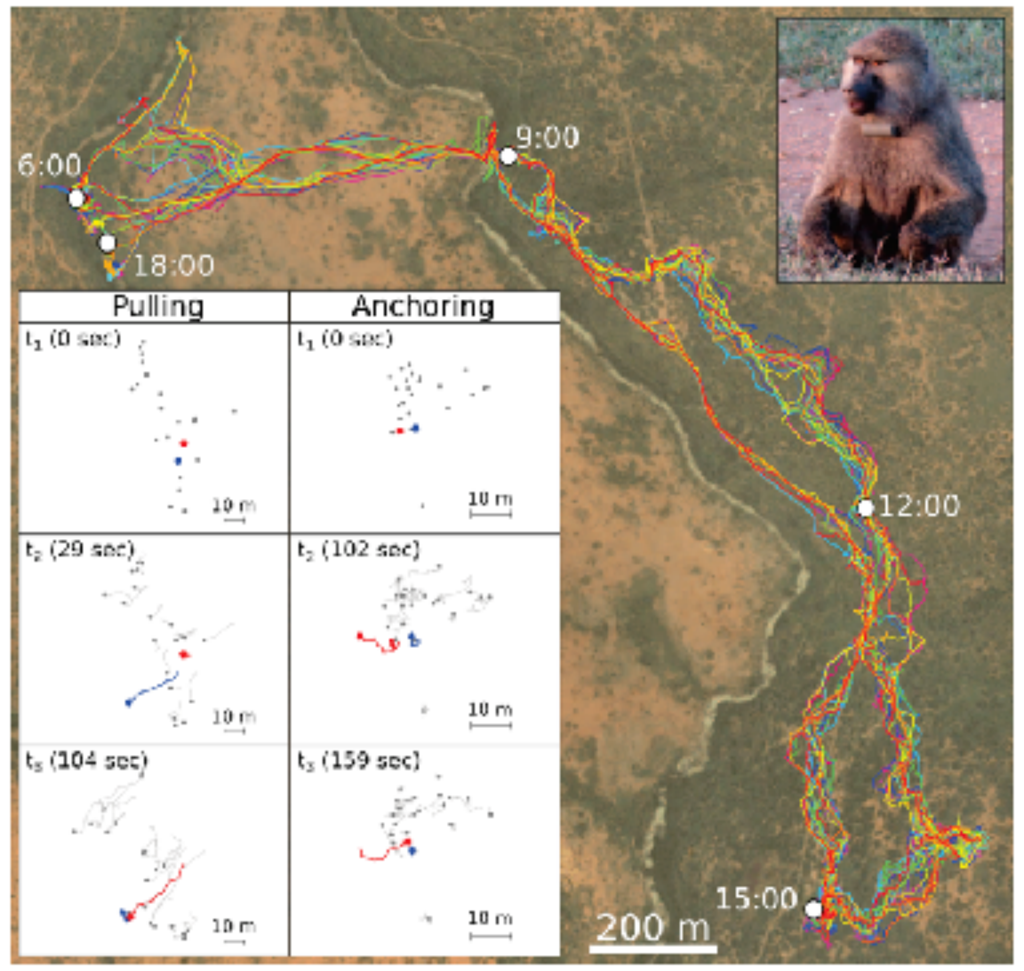


Derex & Boyd (2016); Derex & Mesoudi (2020)

Social Complexity: Network structures

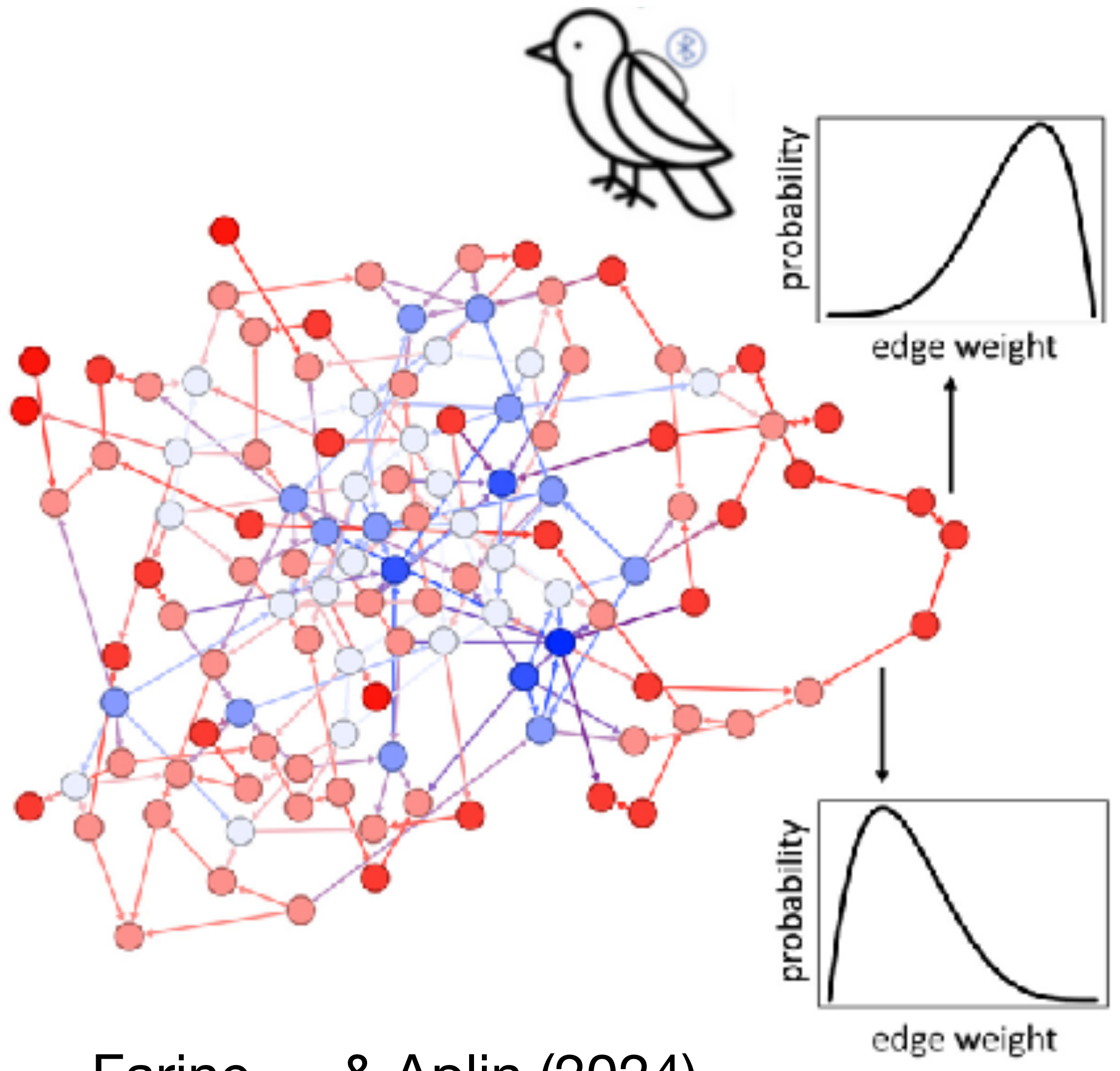
- Unlike in lab experiments, naturalistic social interactions are not explicitly structured and dynamically over through a variety of social factors
    - Thus, open challenges for inference of dynamic social interactions using GPS or bluetooth trackers, camera-based post estimation, or from field of view data

GPS collared Baboons



Strandburg-Peshkin et al. (2015)

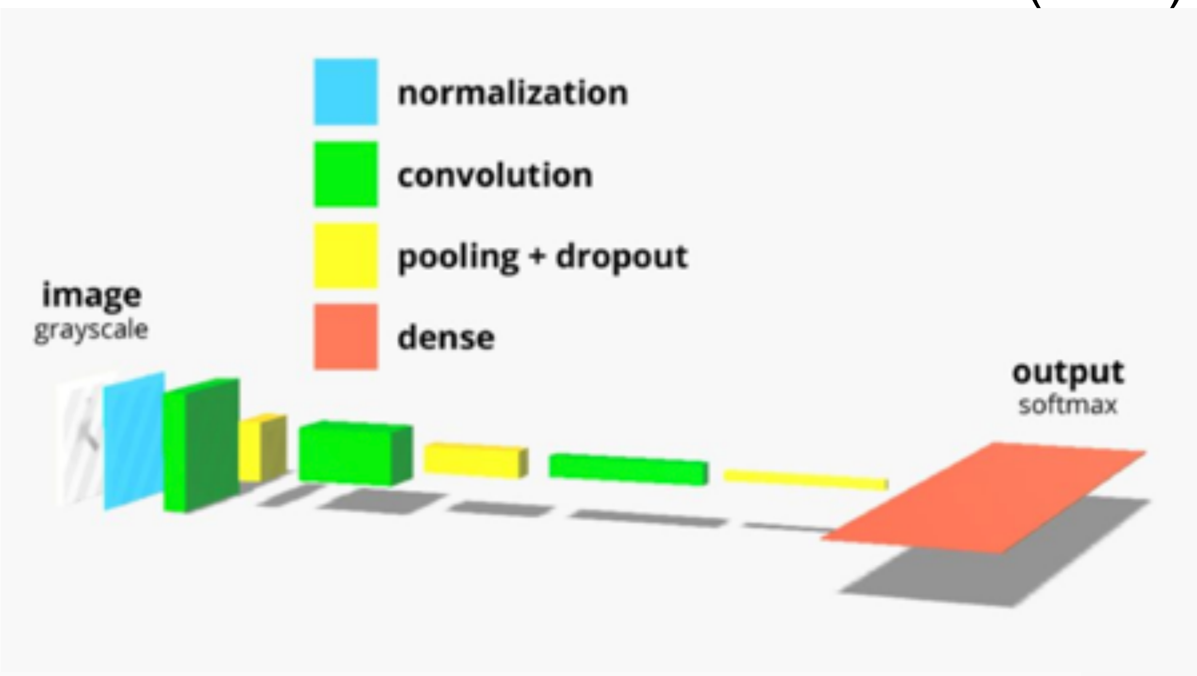
Bluetooth tracking using Apple's "Find My" network

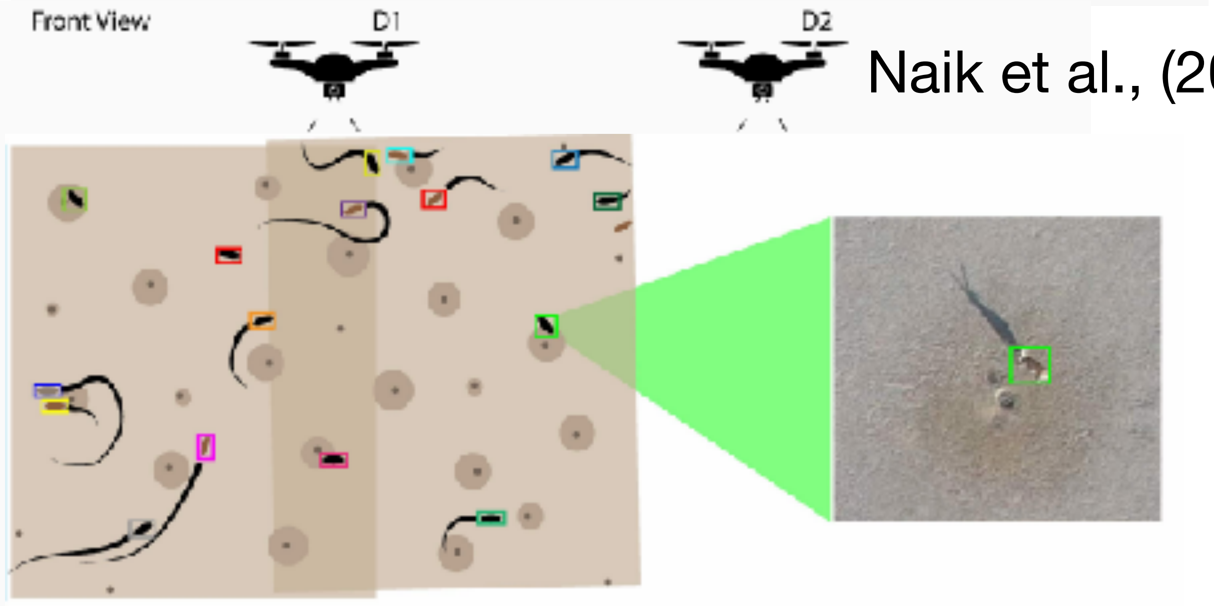


Farine … & Aplin (2024)

TRex: Camera-based pose estimation

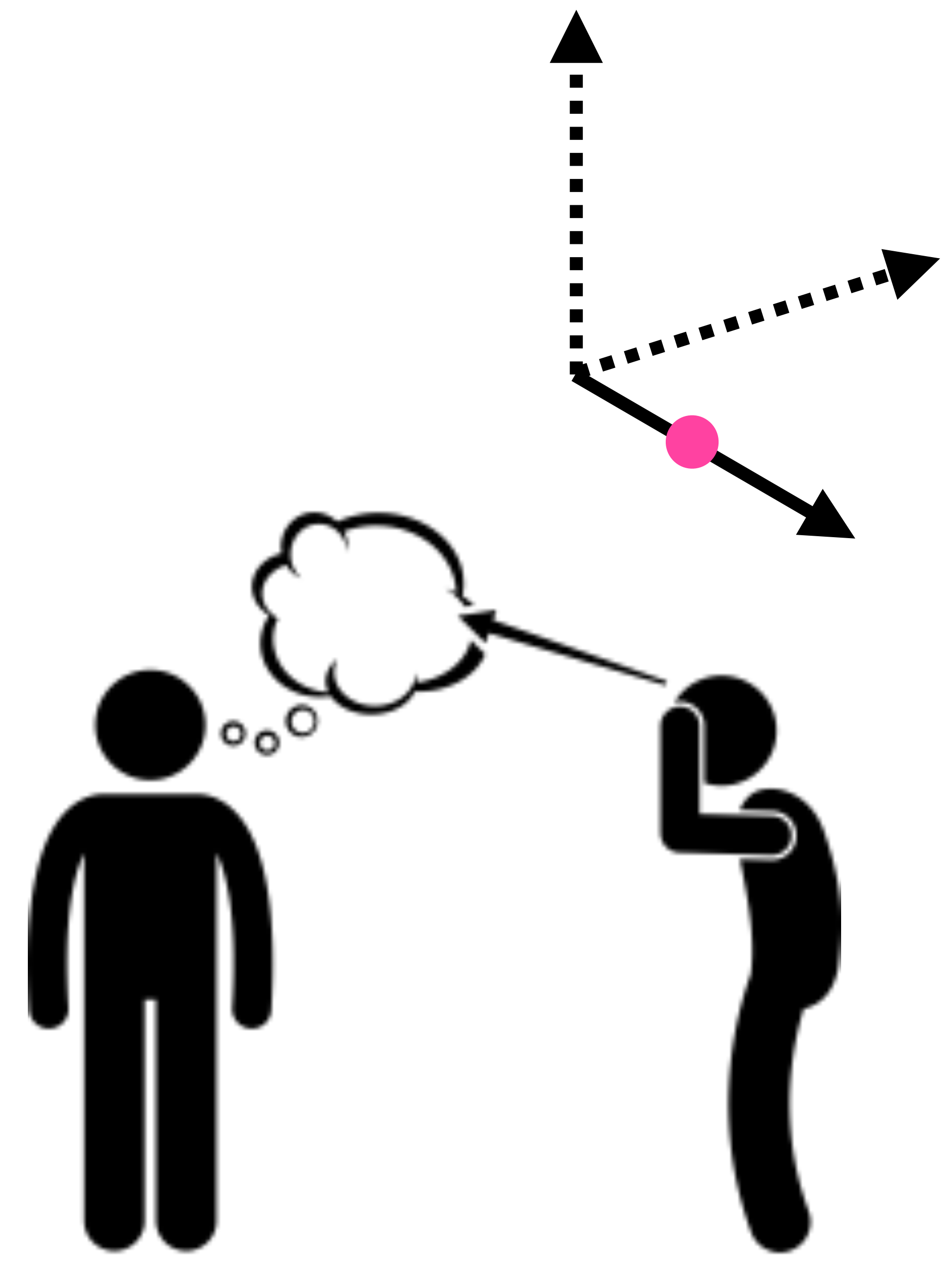
Walter & Couzin (2021)



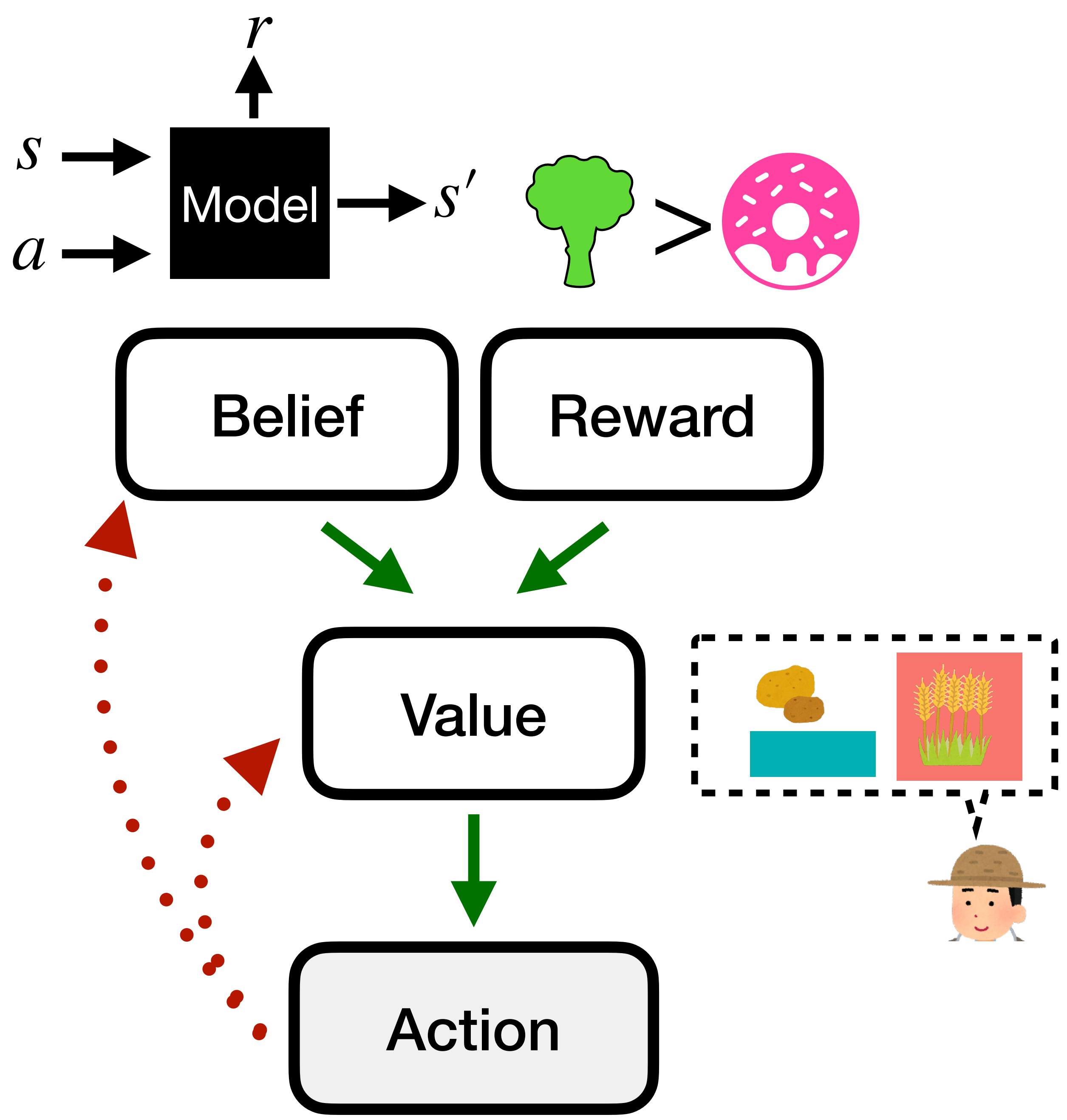


Naik et al., (2024)

Cognitive Complexity: ToM and Pedagogy



- So far, we have only described very simple social learning mechanisms…
  - Yet a key aspect of human social learning is our ability to "unpack" observed actions into imputed mental states
- This is known as *Theory of Mind* (ToM) inference and is often modeled using *Inverse reinforcement learning* (IRL)
- We can specify it at (atleast) two different levels:
  - **Value Inference**: Inferring **V**alues from **A**ctions: $P(V|A) \propto P(A|V)P(V)$
  - **Model-based Inference**: Inferring **B**eliefs about the structure of the world and intrinsic **R**eward specifications $P(B,R|A) \propto P(A|B,R)P(B,R)$

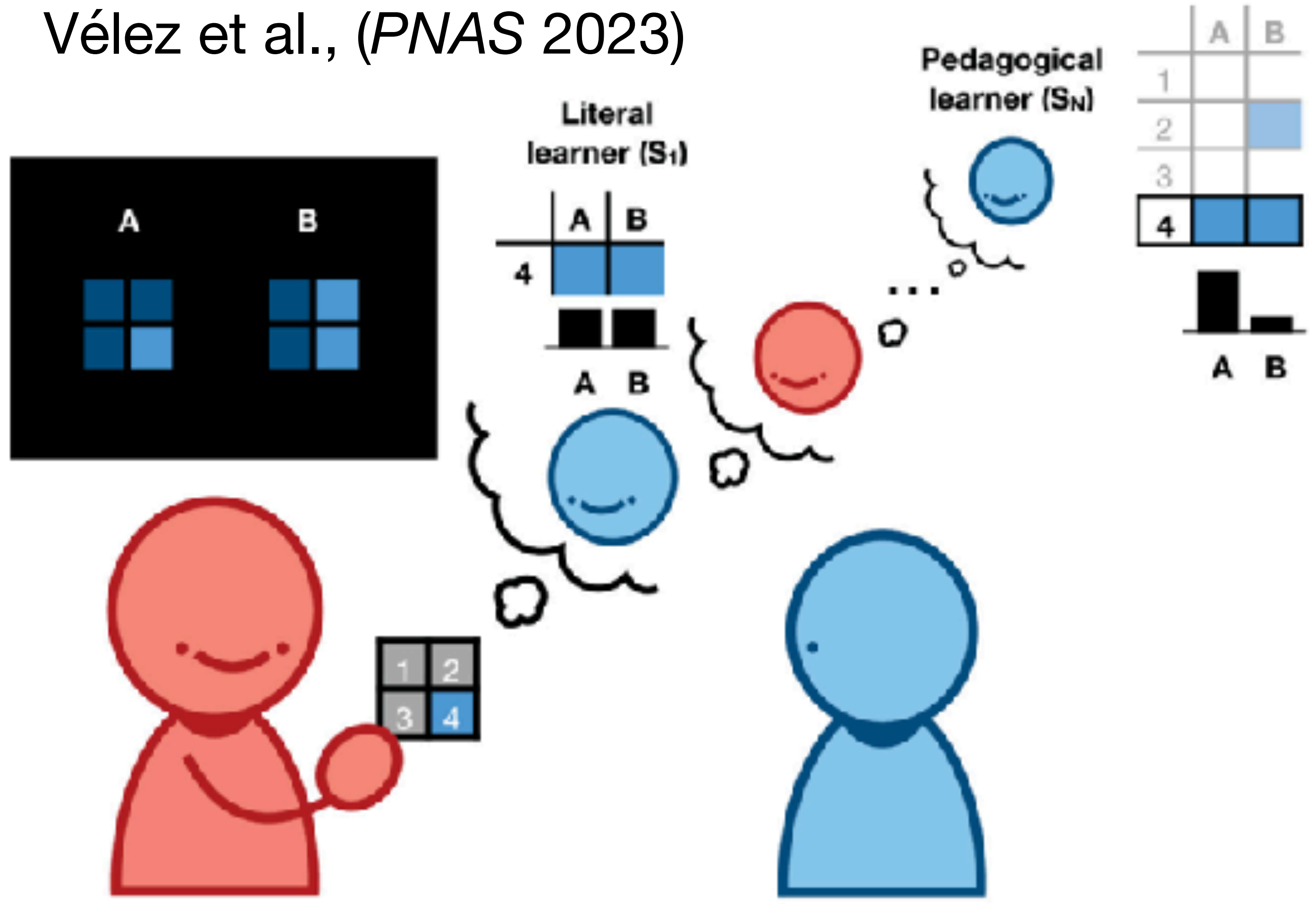


Wu, Vélez, & Cushman (2022)

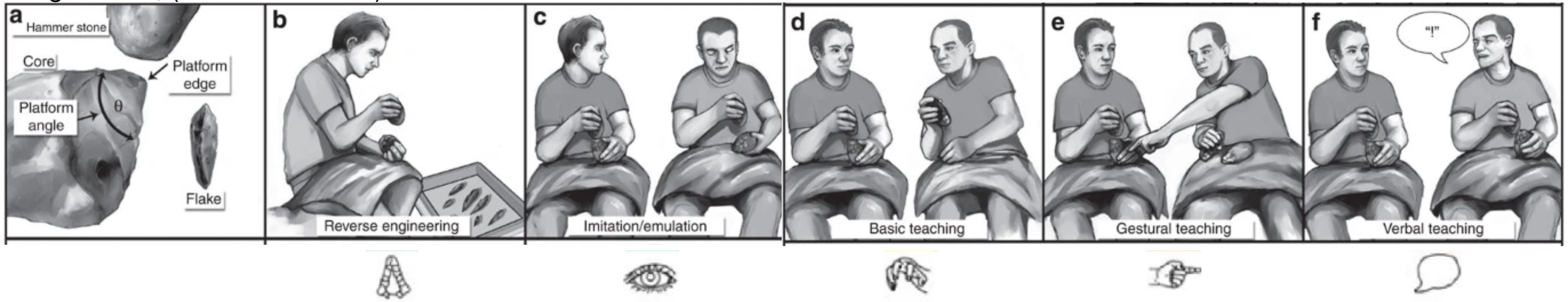
Cognitive Complexity: ToM and Pedagogy

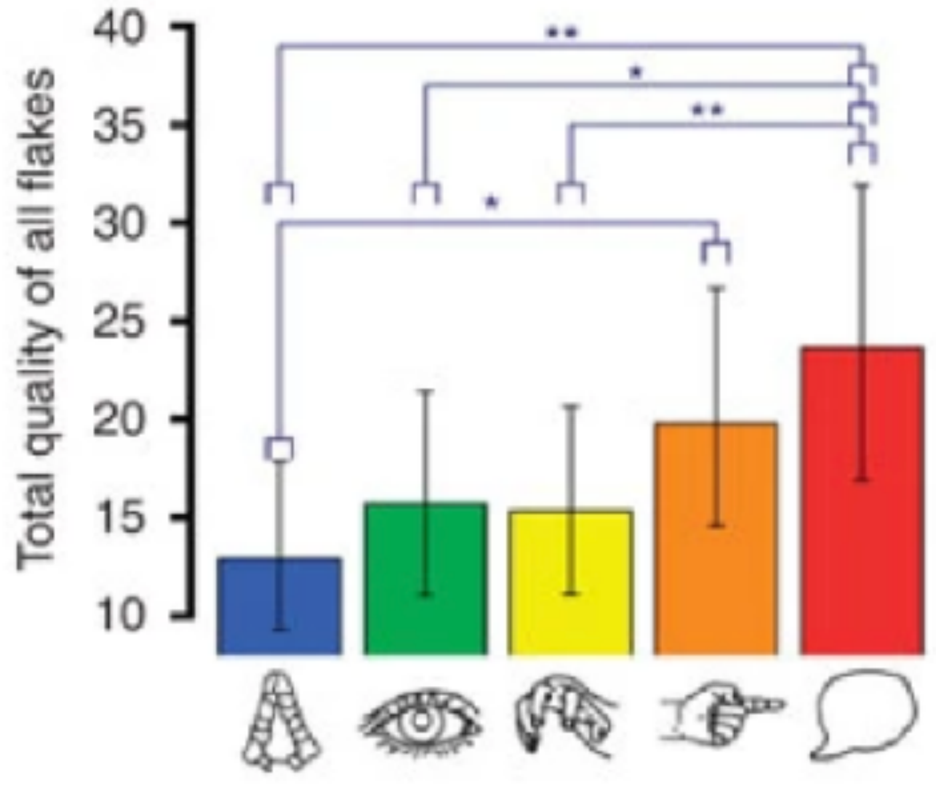
- ToM can be instrumentally useful…
    - Inferring values and beliefs can be more flexible than choice imitation to different skills, preferences, and goals
- … but it also plays an important role in pedagogy
    - Effective teaching requires inferring what the learner knows and doesn't know, in order to select the most informative information
- Pedagogy plays an important role in cultural transmission

Vélez et al., (*PNAS* 2023)



Morgan et al., (*NatComms* 2015)





Putting it all together: Cultural evolution

~ 60 minutes real time

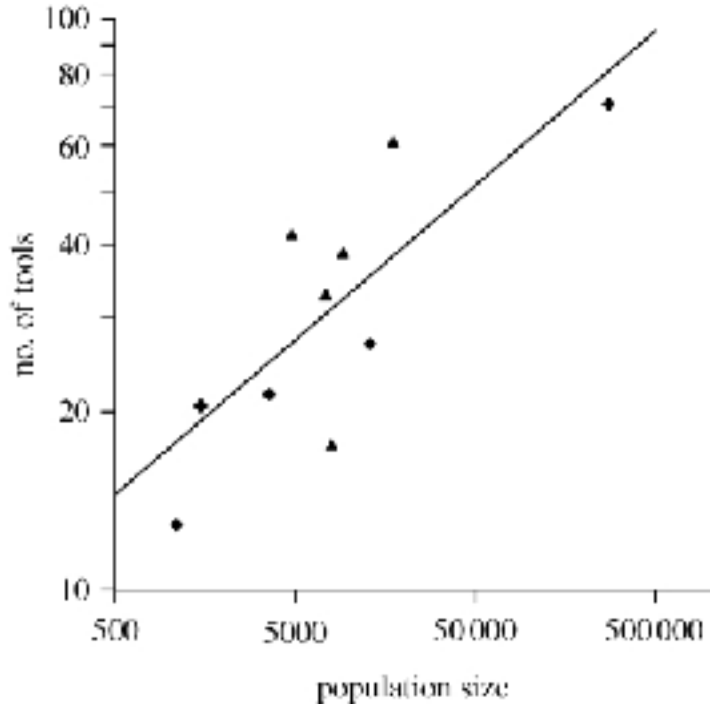
Born to a random player

Dependent on online player community to survive until adulthood

Incremental improvements lead to largescale technological progress over successive generations

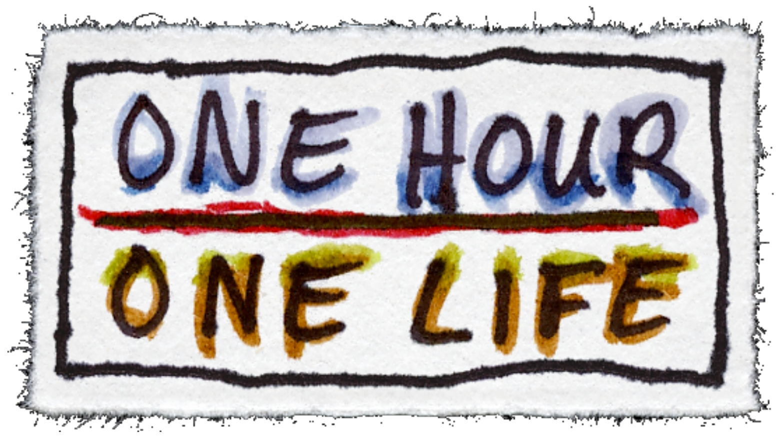
OHOL Dataset

Field studies



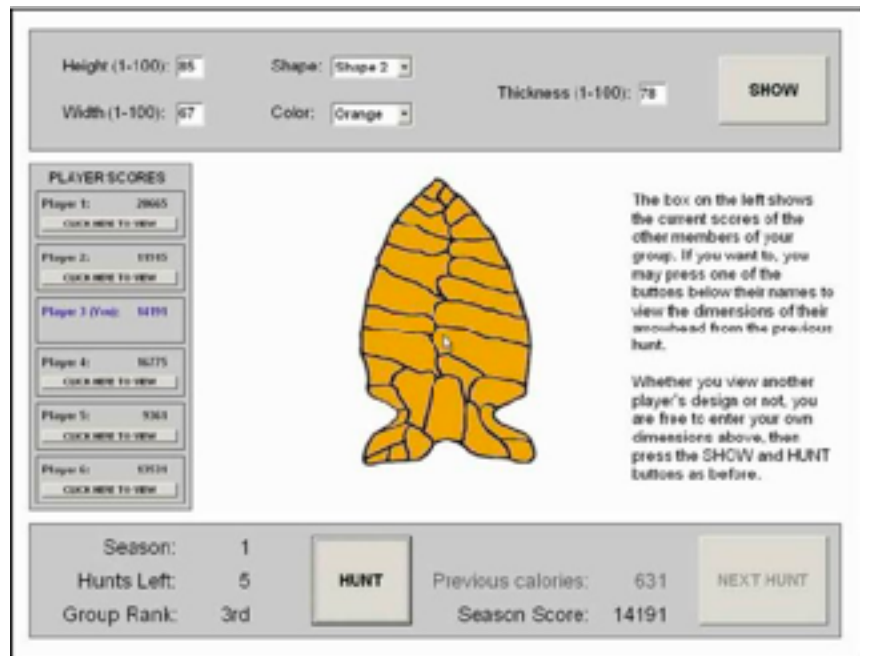
Kline & Boyd (2010)

Online data



Vélez et al., (2024)

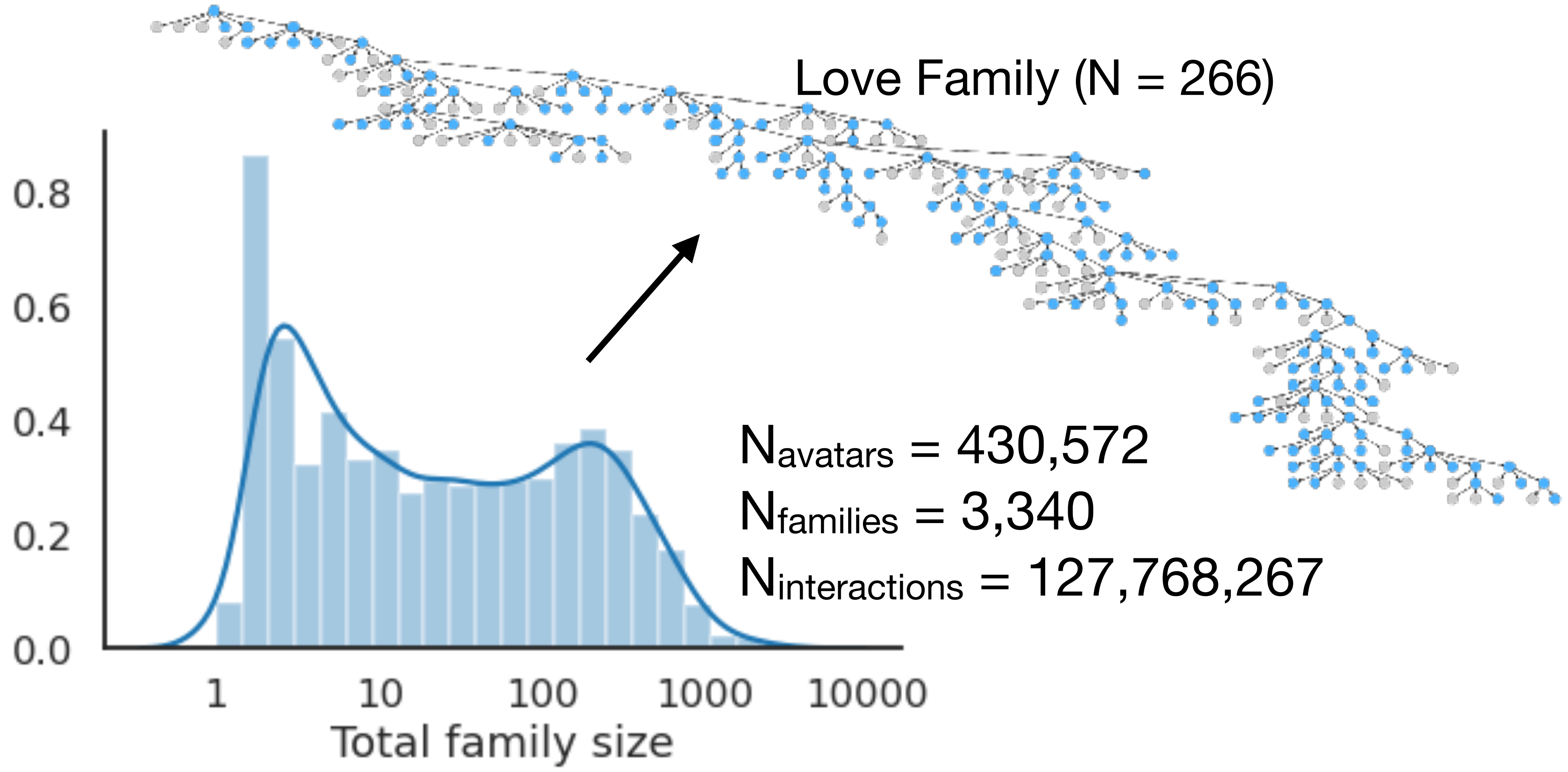
Lab studies



Mesoudi & O'Brien (2008)

Richness of communities

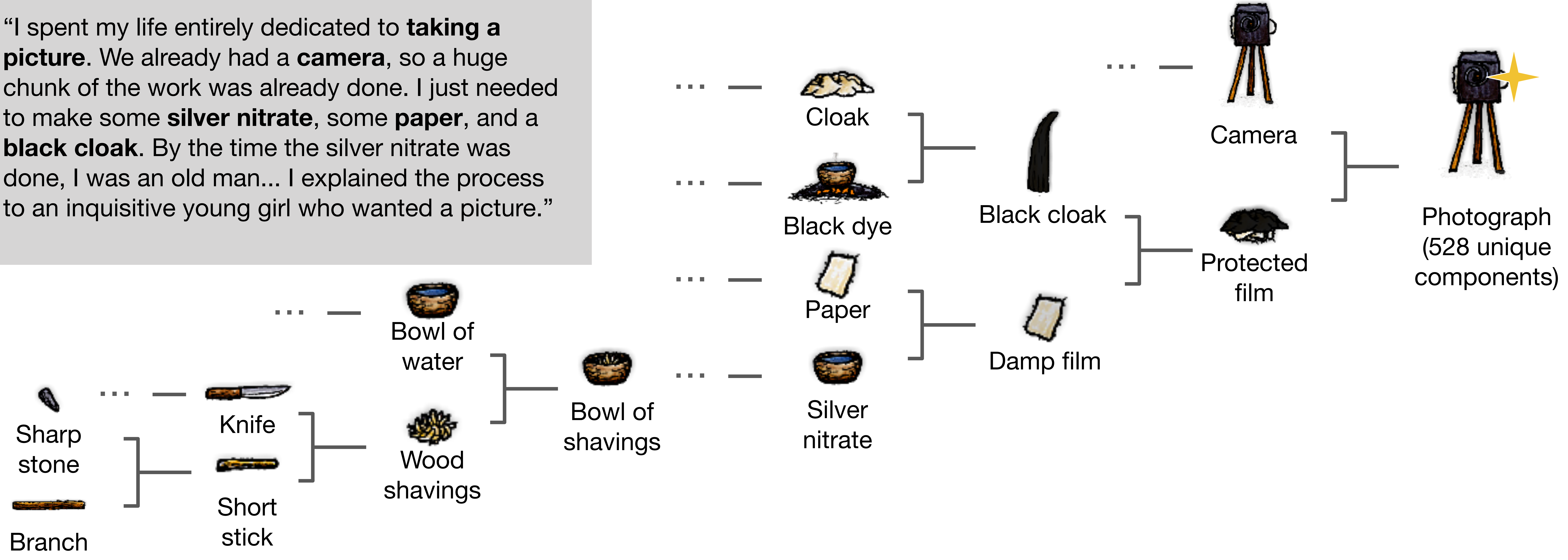
Completeness of data



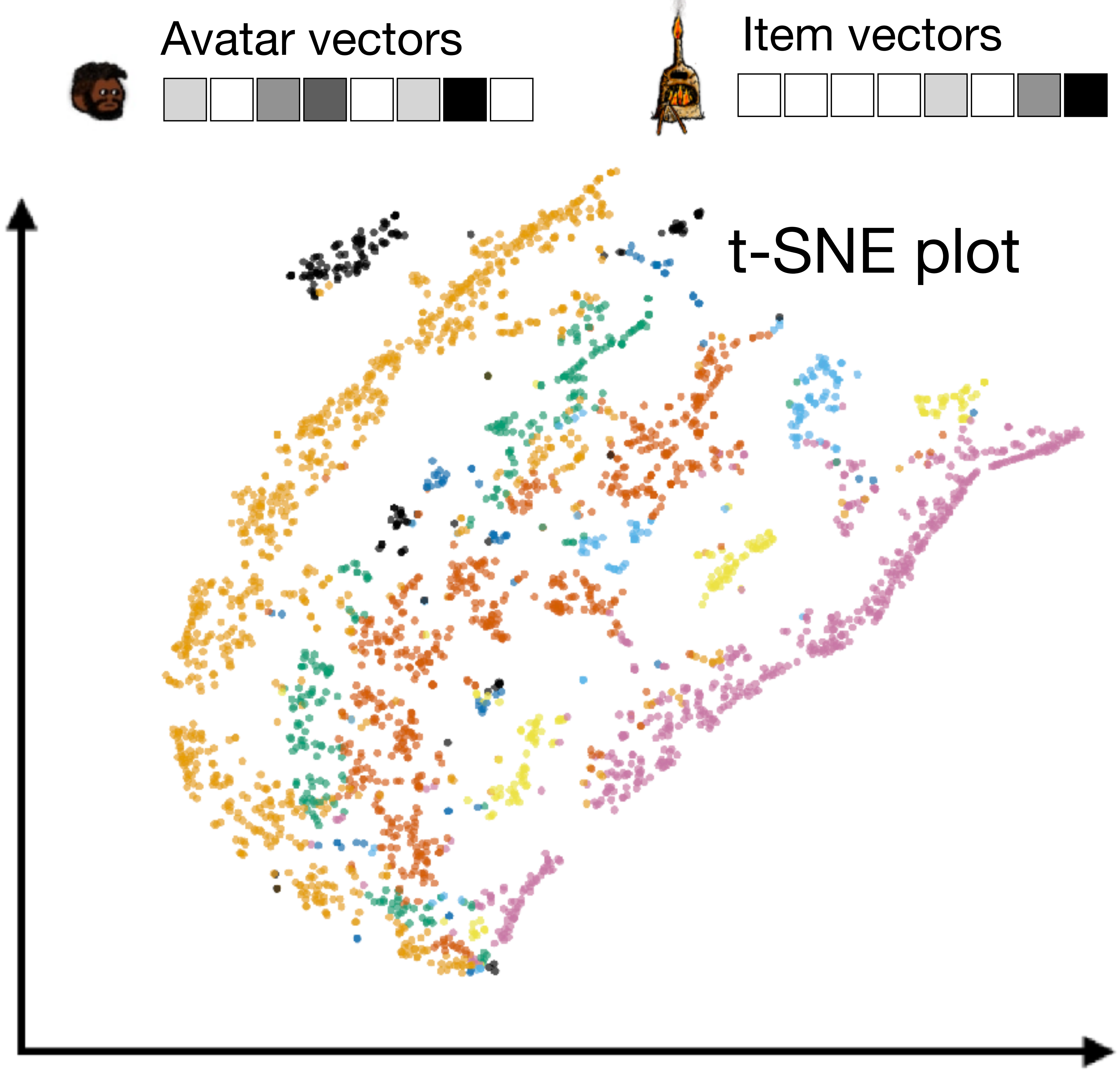
$N_{avatars}$ = 430,572
$N_{families}$ = 3,340
$N_{interactions}$ = 127,768,267

Technological development in a microcosm

"I spent my life entirely dedicated to **taking a picture**. We already had a **camera**, so a huge chunk of the work was already done. I just needed to make some **silver nitrate**, some **paper**, and a **black cloak**. By the time the silver nitrate was done, I was an old man... I explained the process to an inquisitive young girl who wanted a picture."

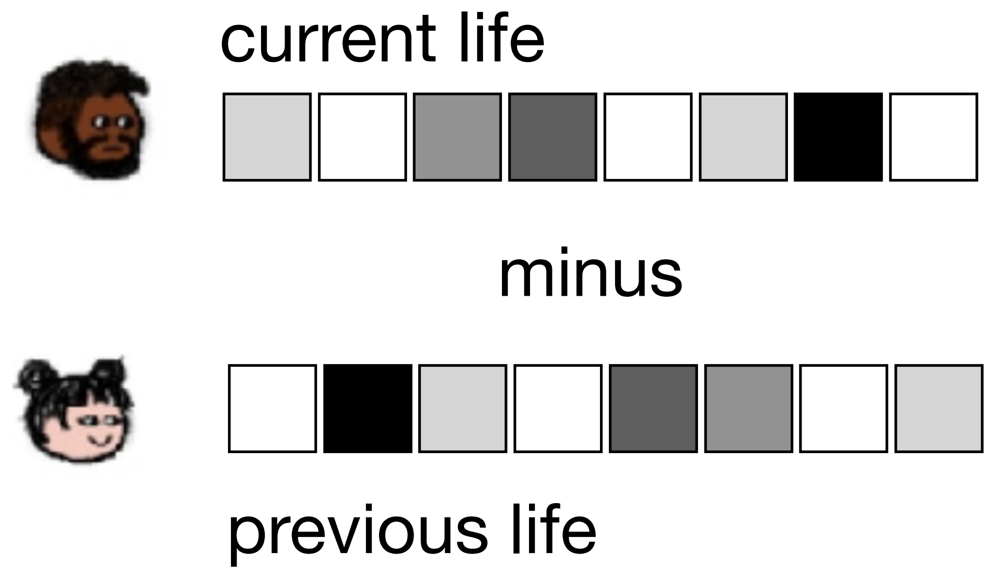


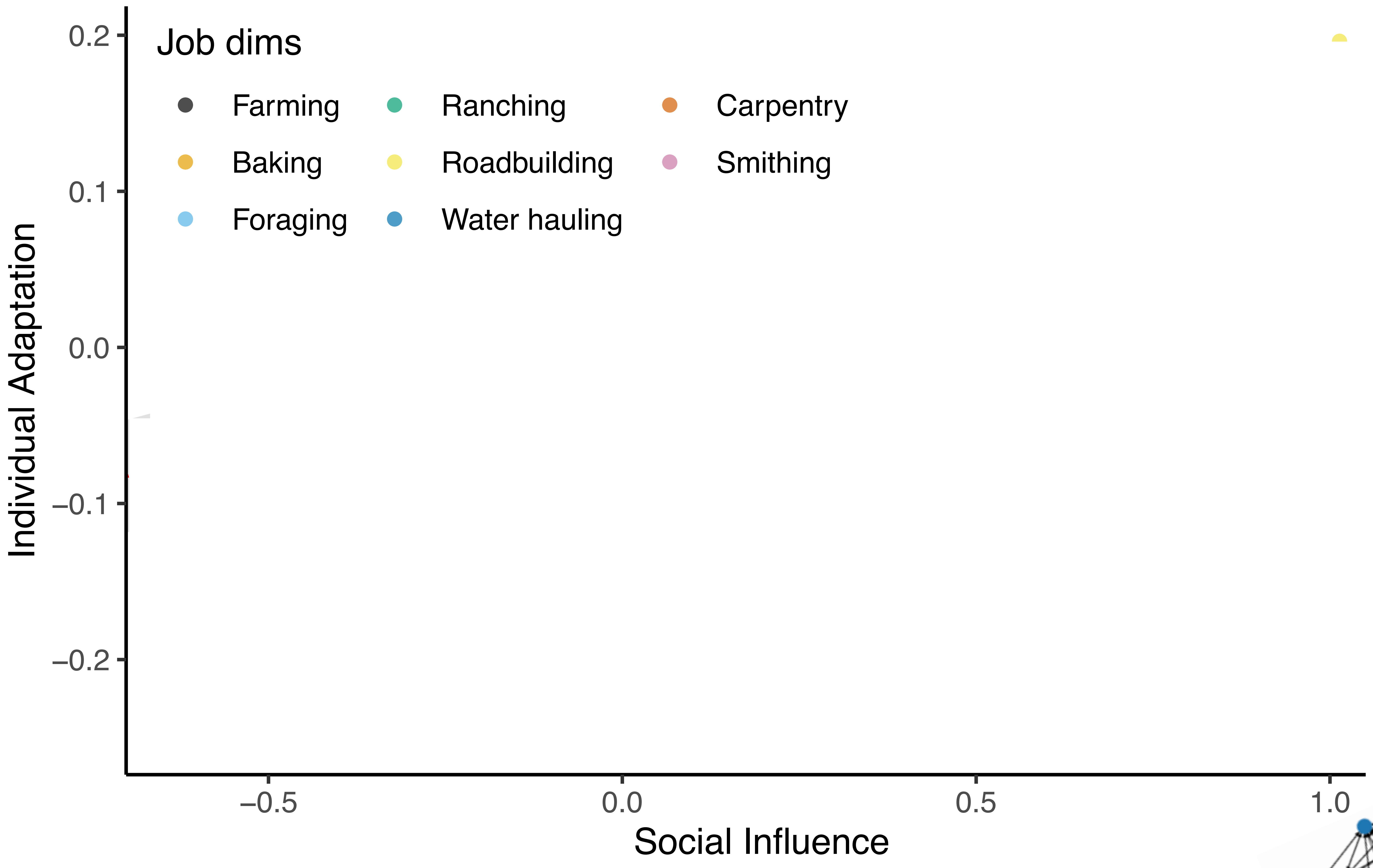
What jobs do people perform in their community?

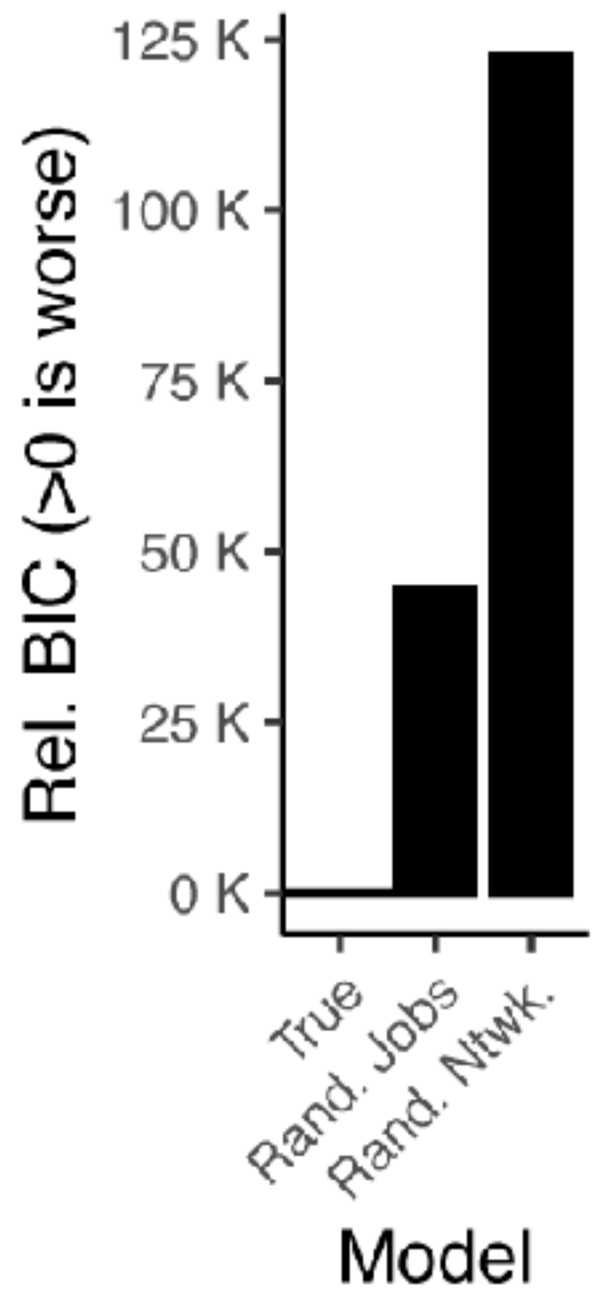


Dimension	Representative items (and factor loadings)		
Farming	Fertile soil pile (12.6)	Carrot pile (9.1)	Steel hoe (3.5)
Baking	Pie crust (8.5)	Hot adobe oven (8.1)	Bowl of flour (3.6)
Foraging	Basket (22.8)	Wild gooseberry bush (5.6)	Dead rabbit (1.7)
Ranching	Fleece (4.0)	Shorn sheep (3.4)	Ball of thread (3.7)
Roadbuilding	Floor stakes (14.8)	Stone road (5.2)	Pine floor (4.2)
Water hauling	Bucket of water (41.3)	Canada goose pond (14.5)	Dry gooseberry bush (4.4)
Carpentry	Horse-drawn cart (73.8)	Fence (2.3)	Yew bow (0.8)
Smithing	Firing forge (15.8)	Iron ore pile (11.0)	Stack of steel ingots (7.4)

Modeling cultural transmission of expertise







$$\sum \Big(\text{freq}(\;)\text{job}(\;) + \ldots + \text{freq}(\;)\text{job}(\;)\Big)$$

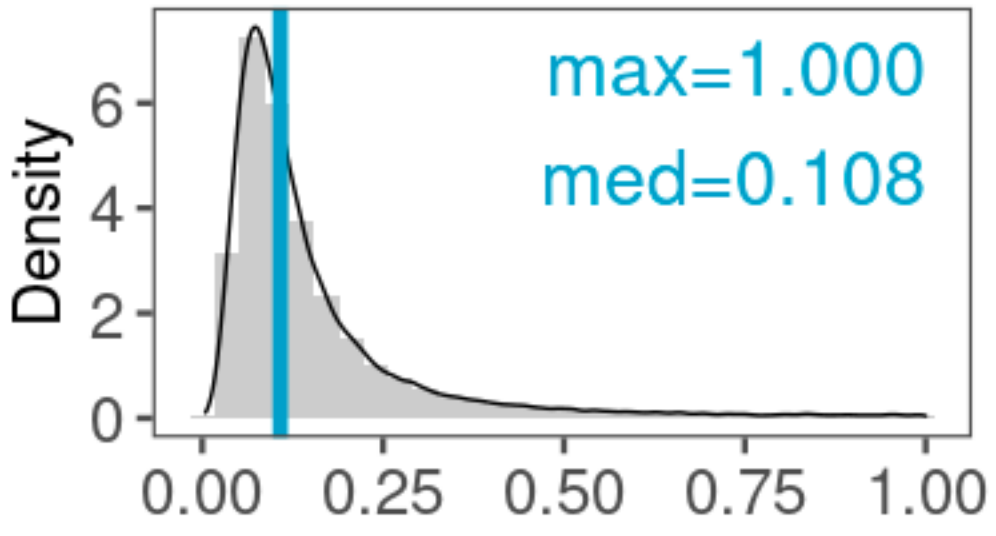
How do community characteristics affect technological development?

Specialization

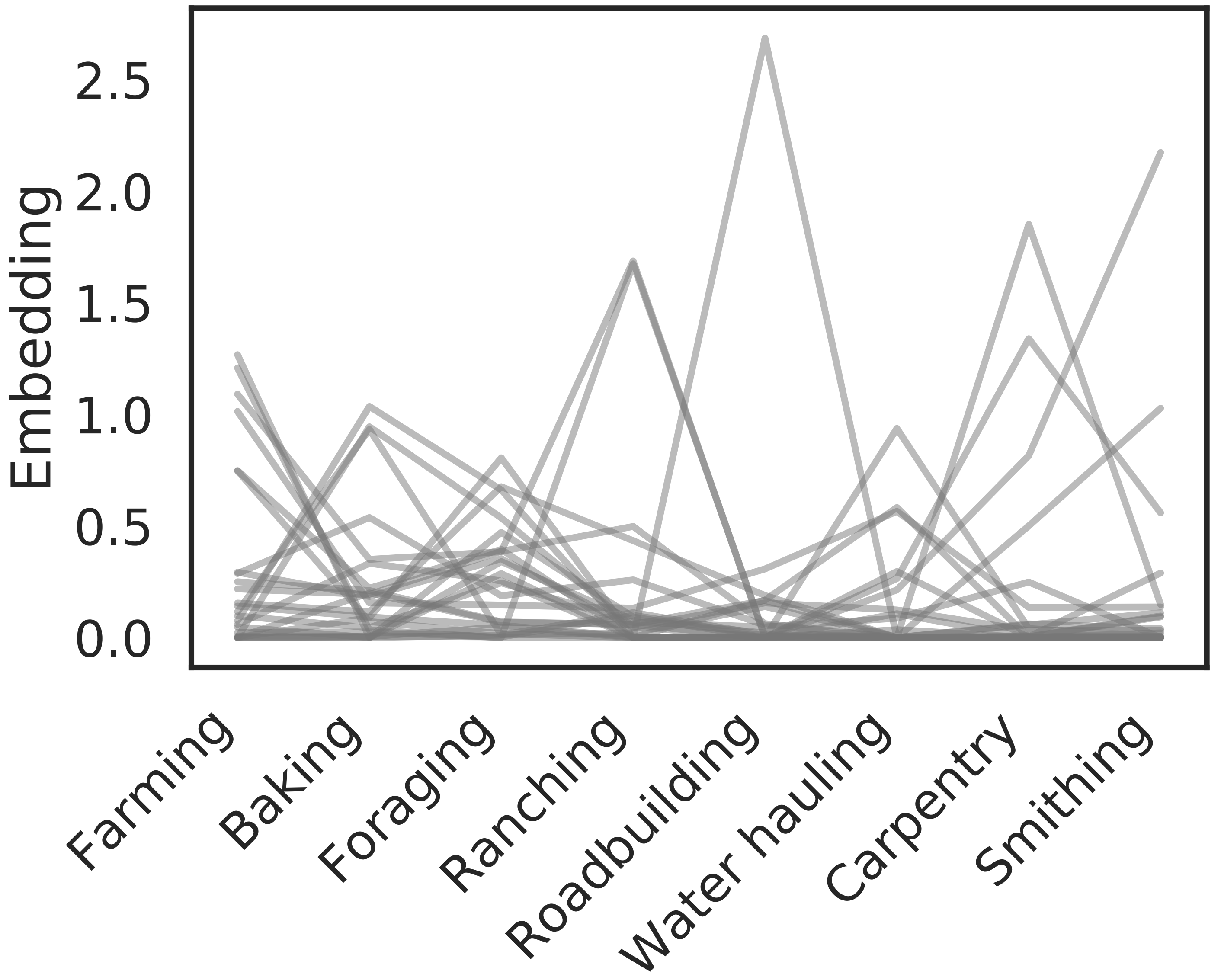
Herfindal-Hirschman Index (HHI)

$$HHI = \sum_{i=1}^{n} \boxed{s_i^2}$$

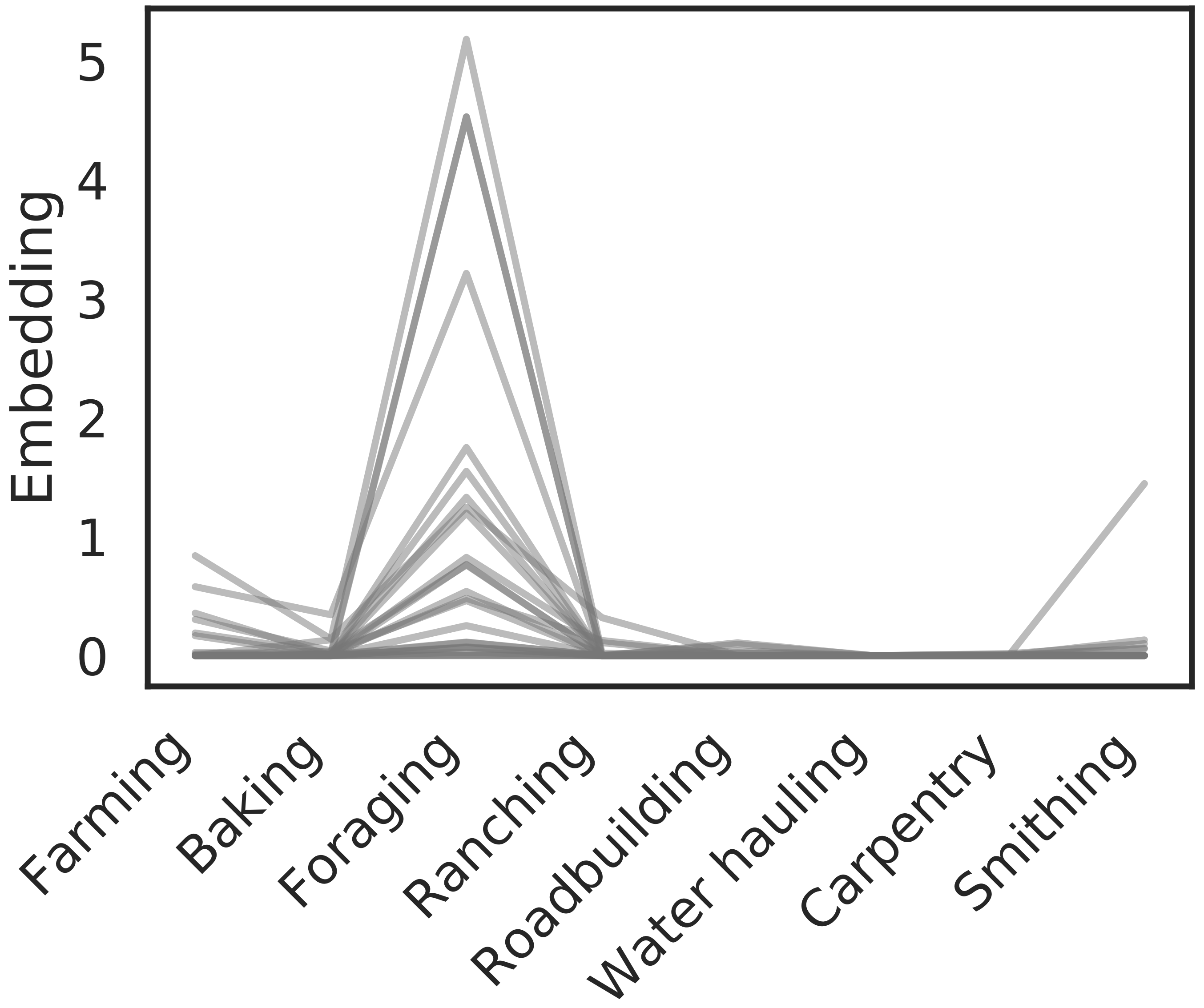
Share of community activity in job *i*



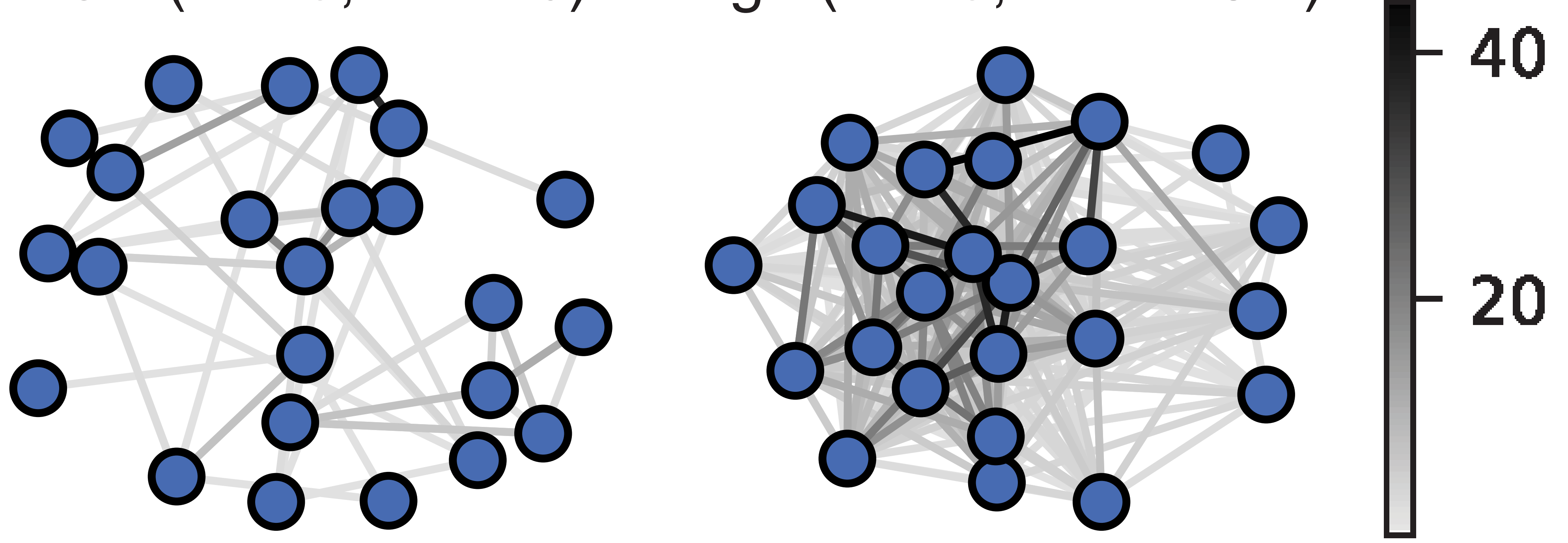
Low specialization
ESPERANZA family
HHI = 0.009



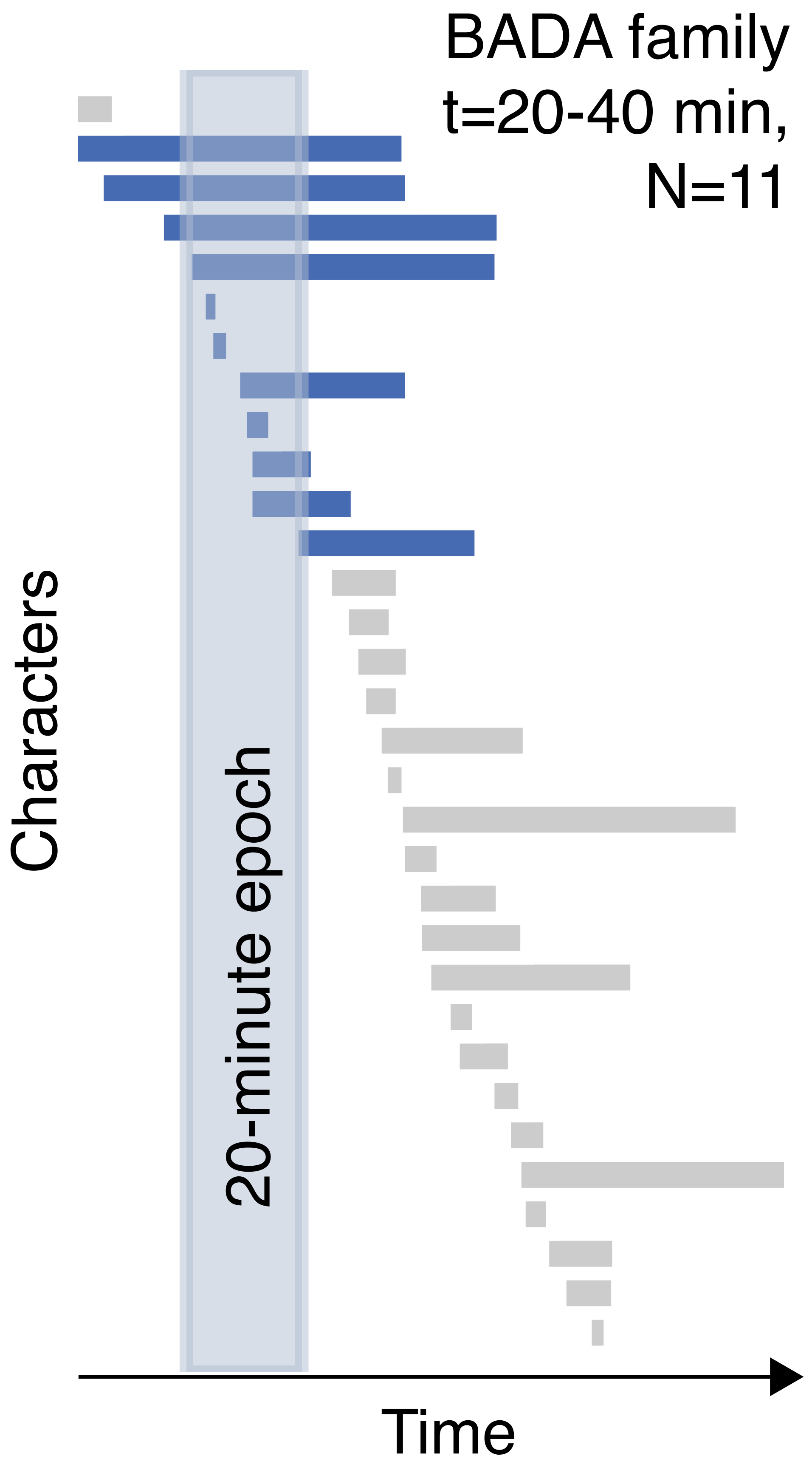
High specialization
MIA family
HHI = 0.7

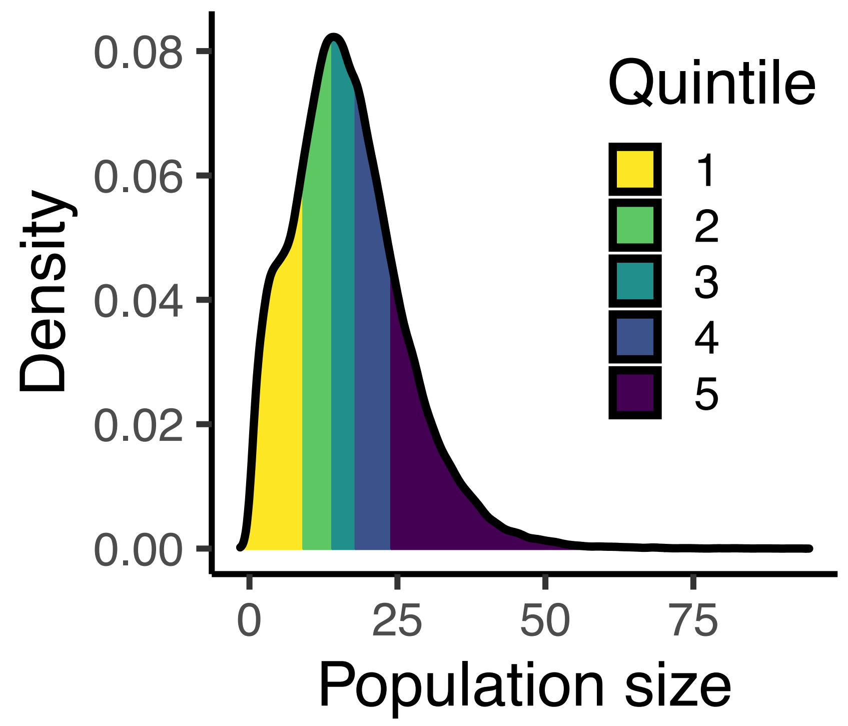


Interactivity

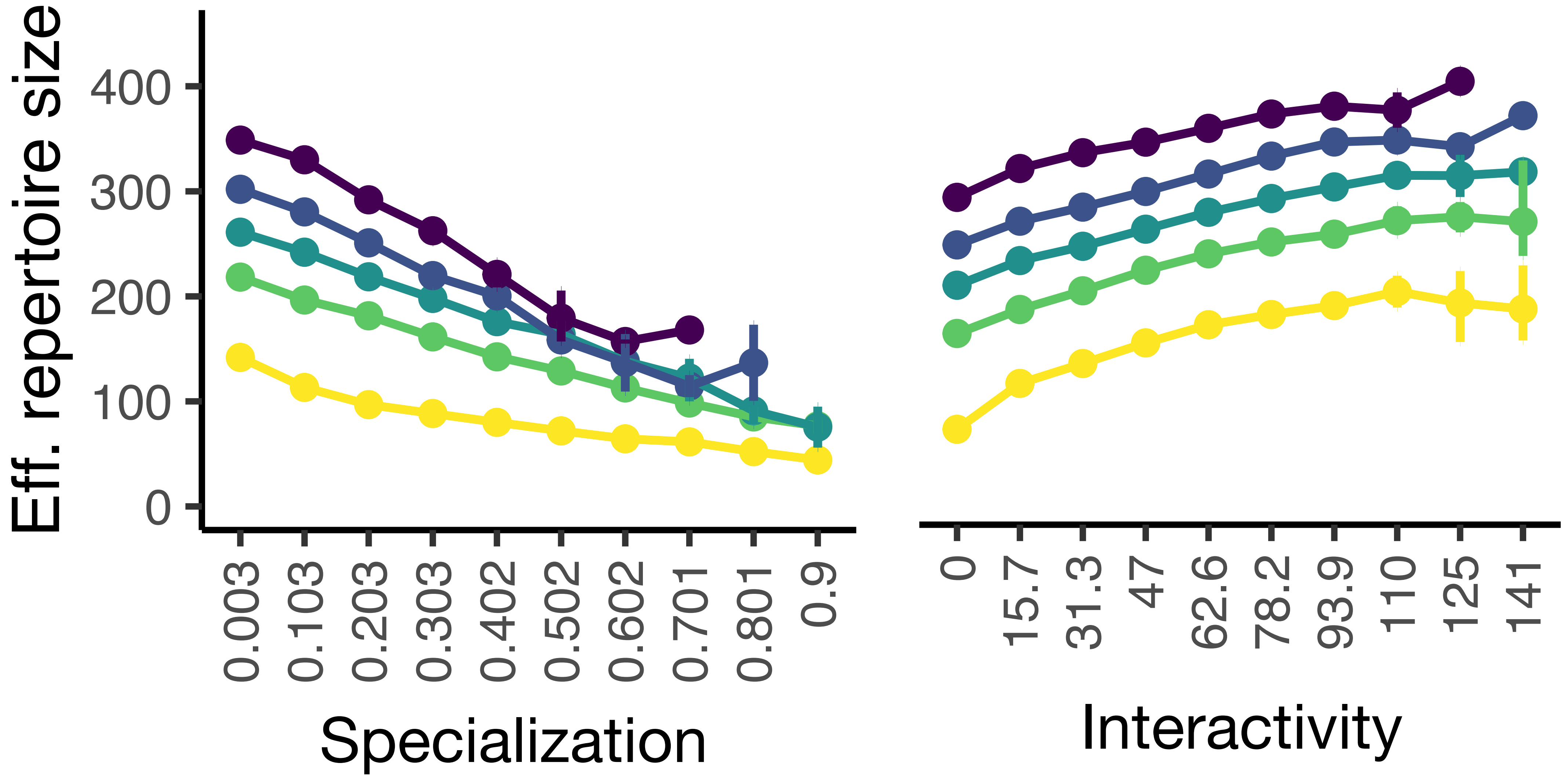


How do community characteristics affect technological development?

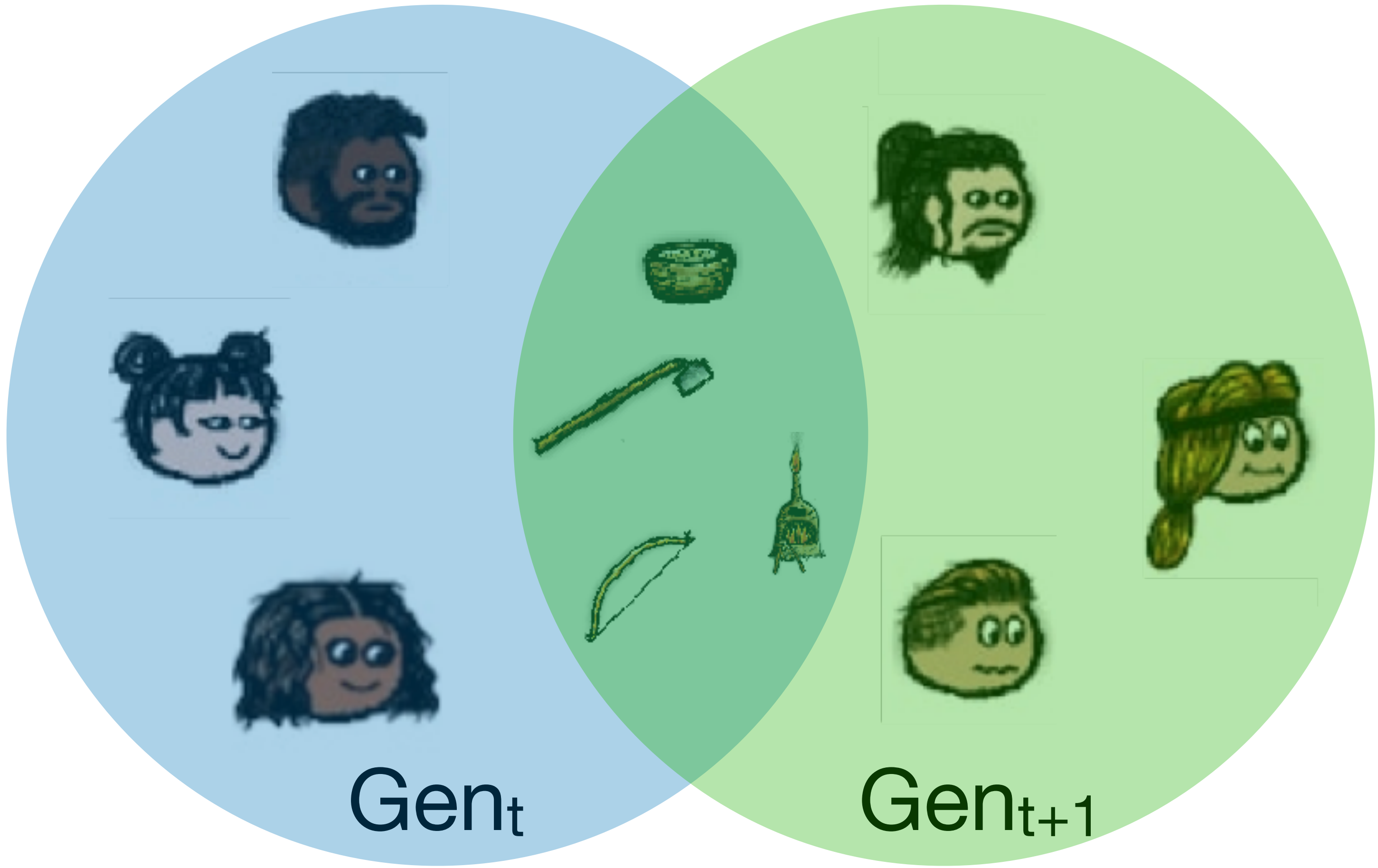


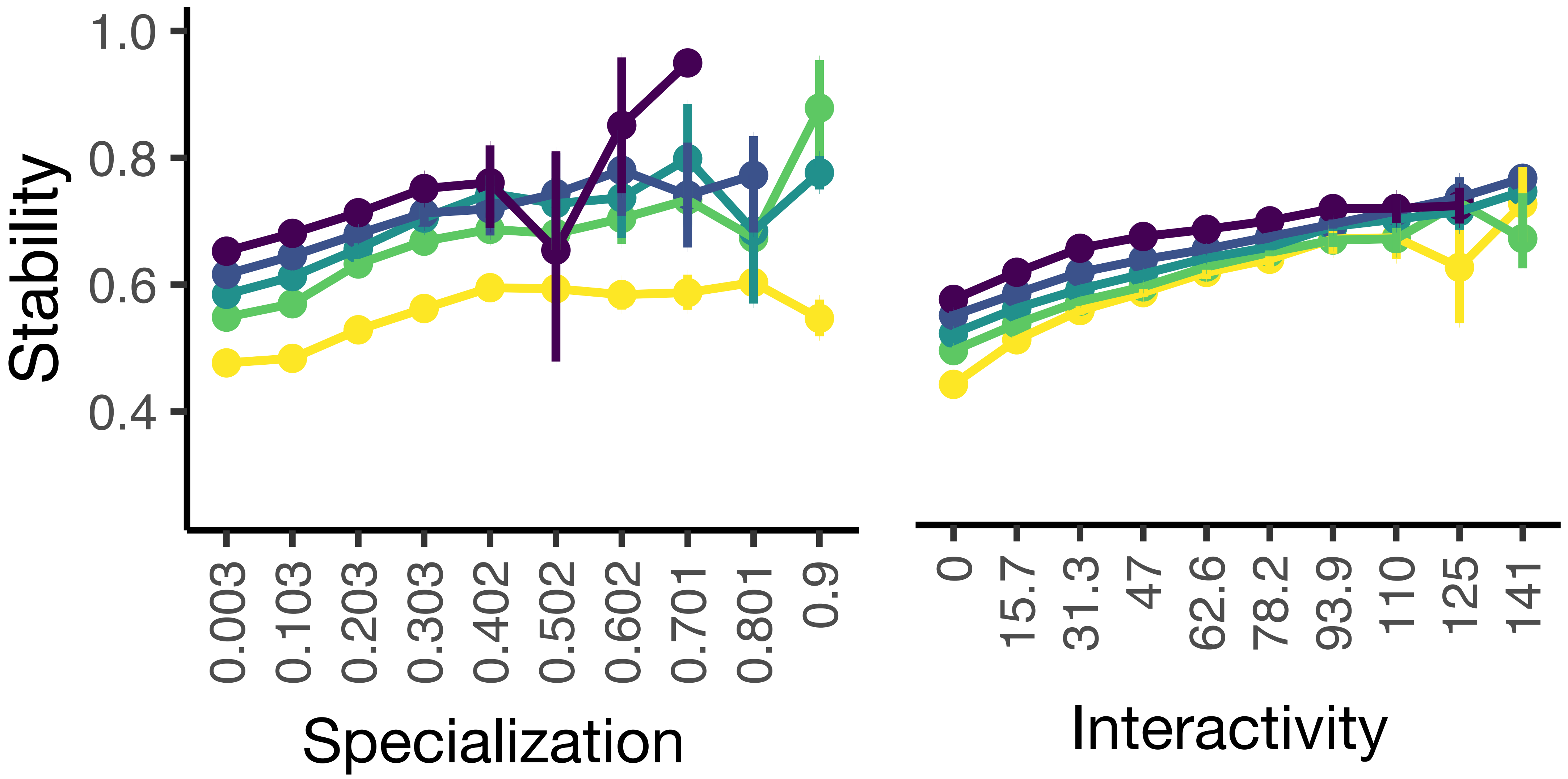


Effective Repertoire



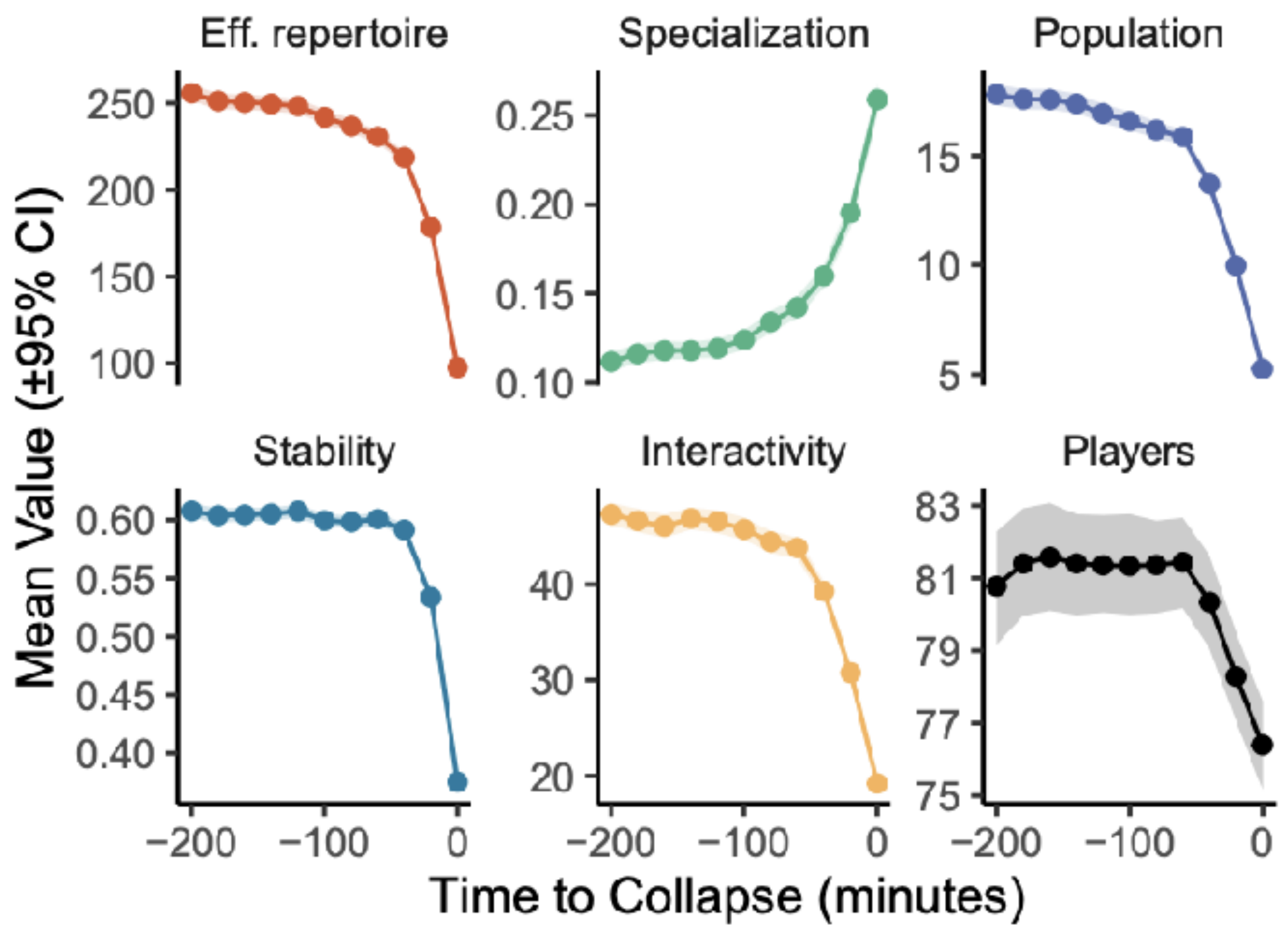
Stability

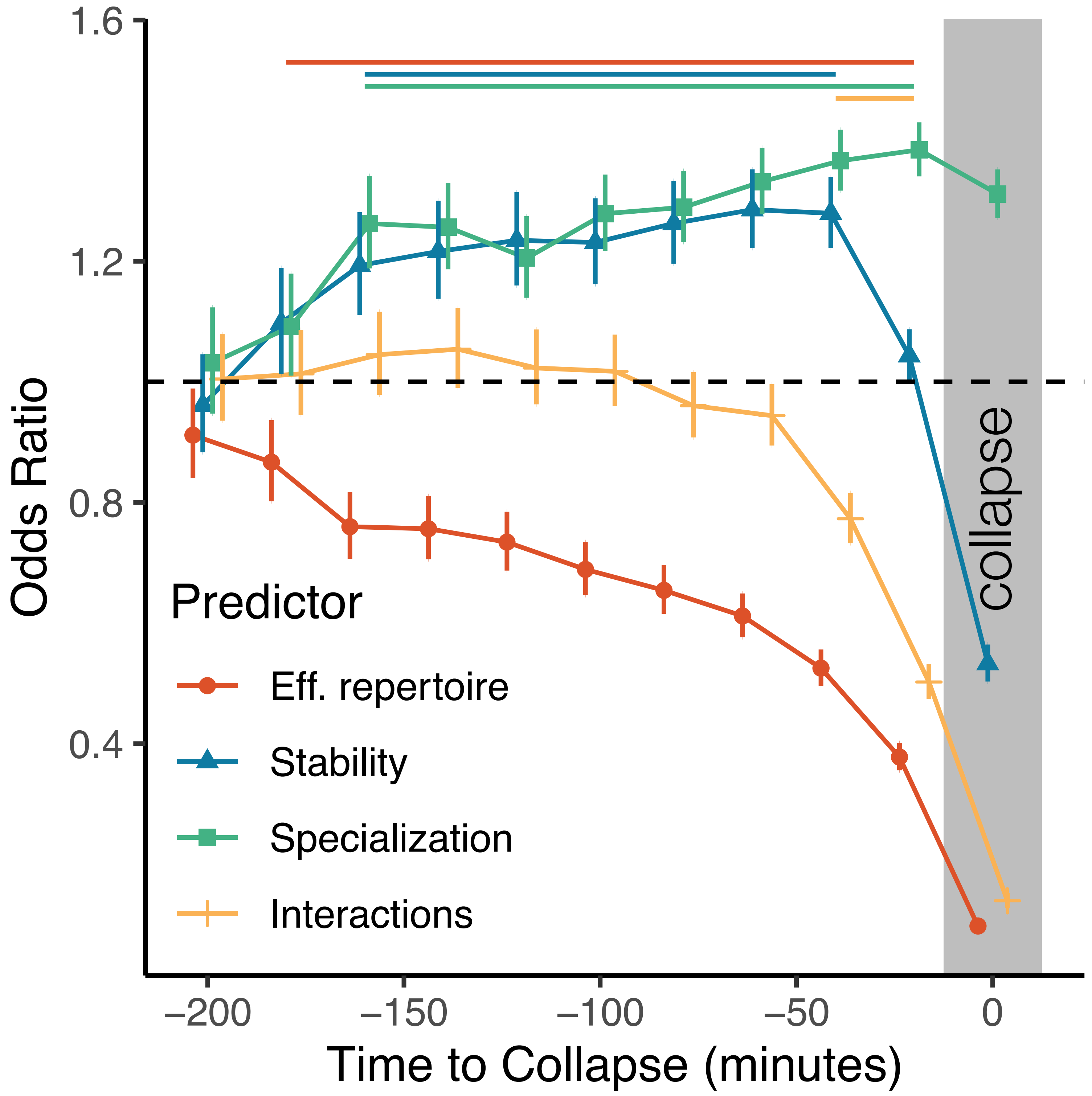




Predicting collapse

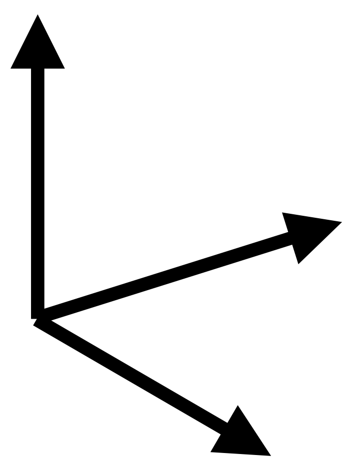
- Beyond predicting growth and stability, can we predict when communities will collapse? (pop_size —> 0)
- **Top right:** raw variables aligned at time of collapse
- **Bottom right**: temporal cluster analysis predicting collapse at different temporal offsets
  - Lines at the top indicate significant clusters
  - A decline in effective reportoirs is an early warning signal of collapse
  - But over-specialization is just as predictive only 1 epoch later
- **Take-home: loss of diversity predicts collapse!**





Summary and open challenges

- Social learning deploys a wide range of tools:
  - **imitation**: directly copy observed behaviors
  - **value-shaping:** add a heuristic bonus to observed behaviors
  - **ToM Inference**: inferring hidden value representations or hidden beliefs about the world
  - + extensions across multiple dimensions of complexity



- Yet for each mechanism we can describe verbally, we can also define a computational model that makes more precise commitments to the mechanisms of behavior
- Through experimentation and modeling, we can iteratively tweak and refine our understanding of social learning.