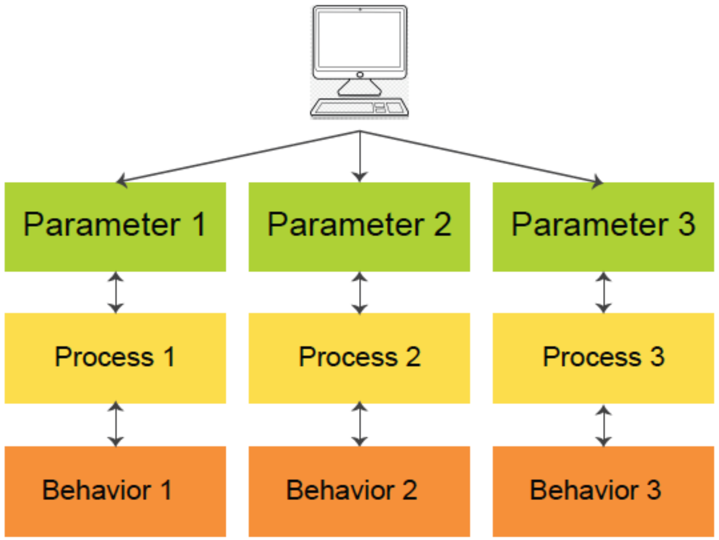
Interpreting reinforcement learning in the context of individual differences.

Anne Collins, UC Berkeley

CPC Zurich – September 2022

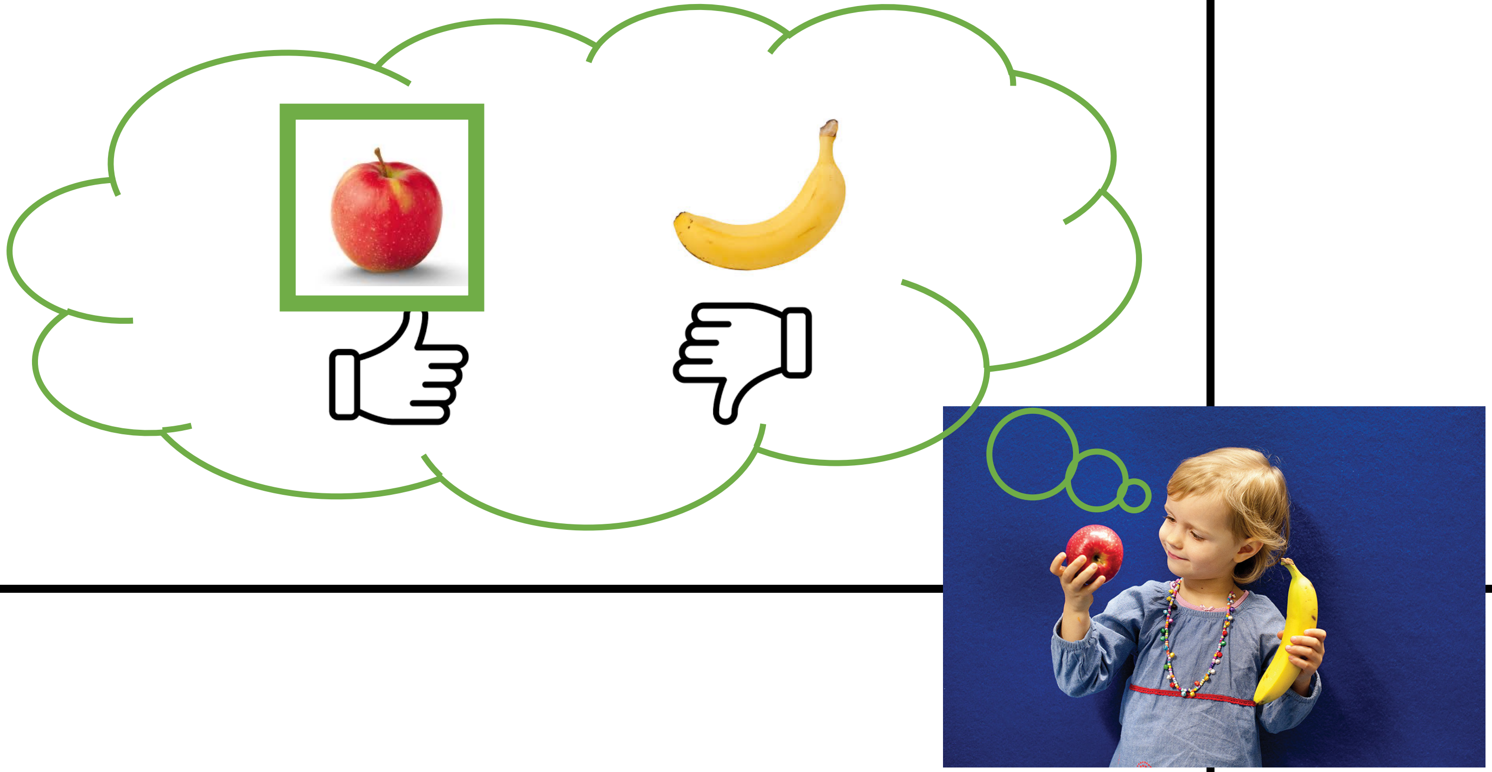
Computational psychiatry

- "carve cognition at its joints"
- Use computational modeling to reveal computational processes that support behavior
  - Match processes to neural substrates (via latent variables)
  - Match processes to individual differences (via meaningful parameters)



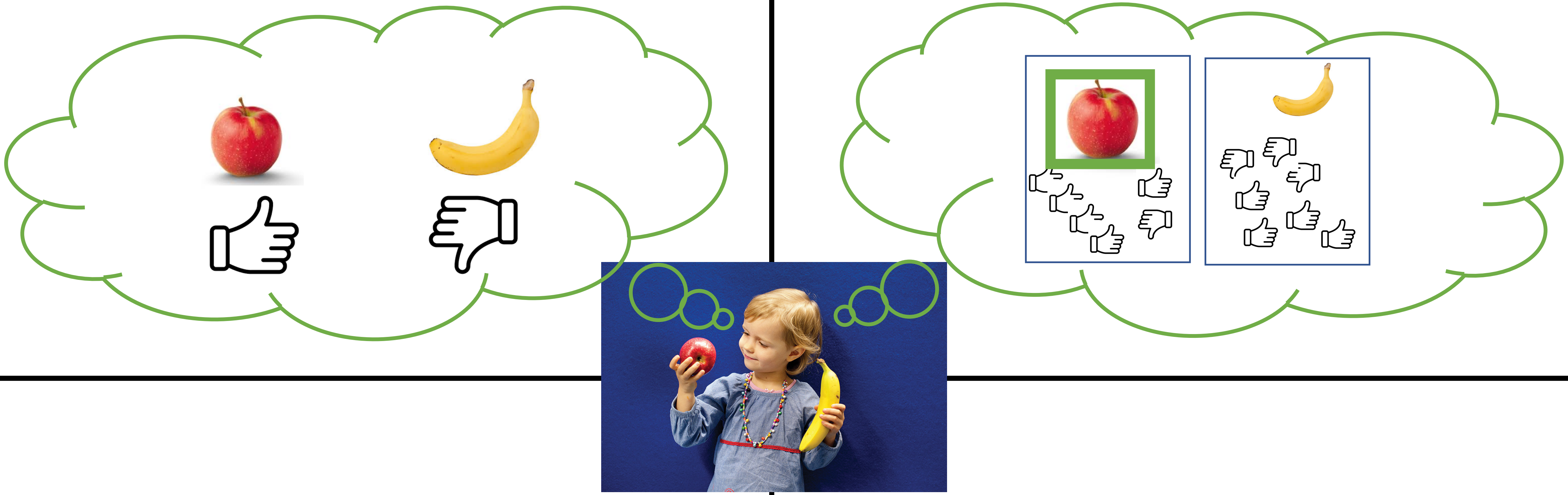
Computational modeling: Benefits and Risks

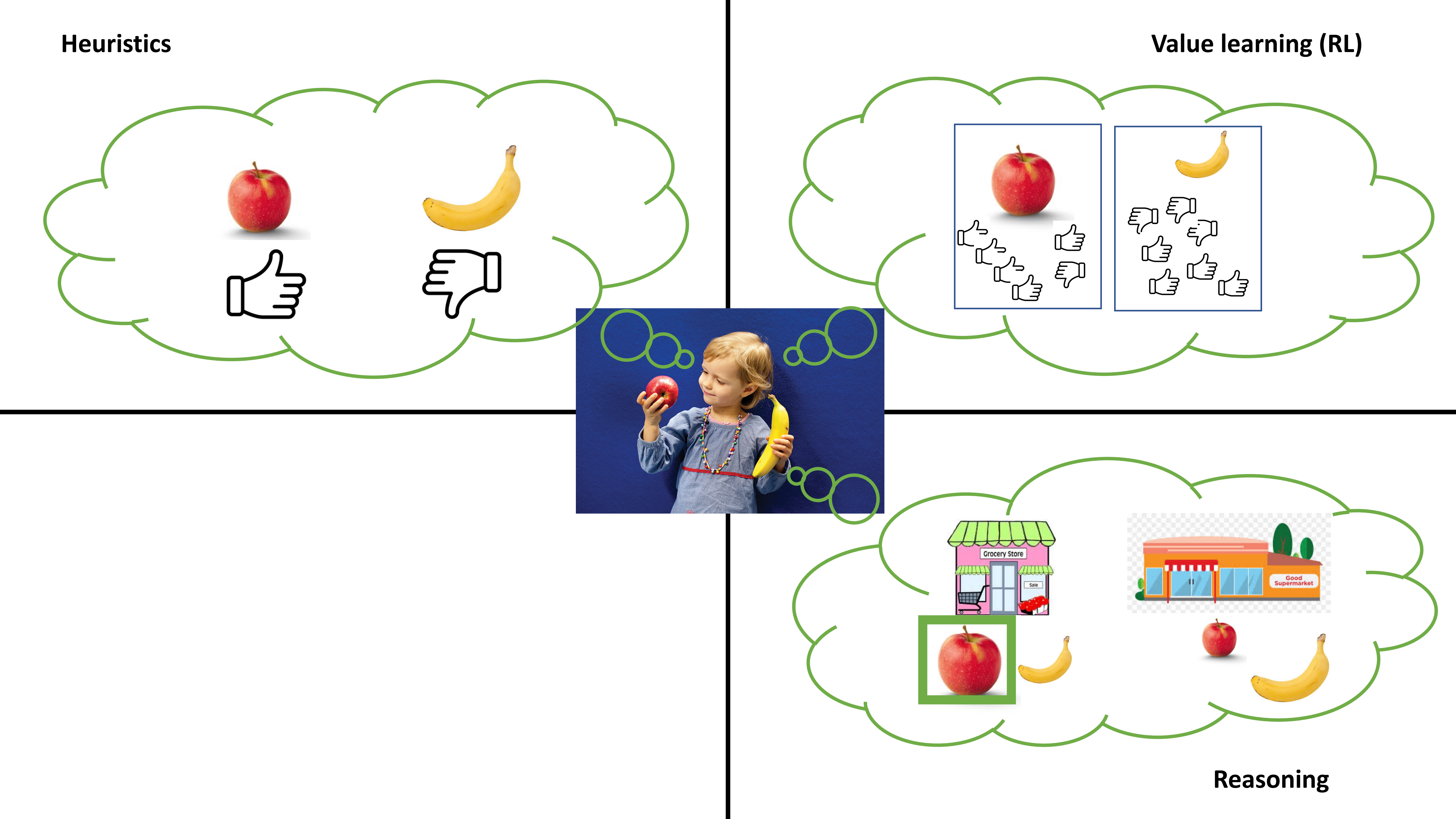
Heuristics

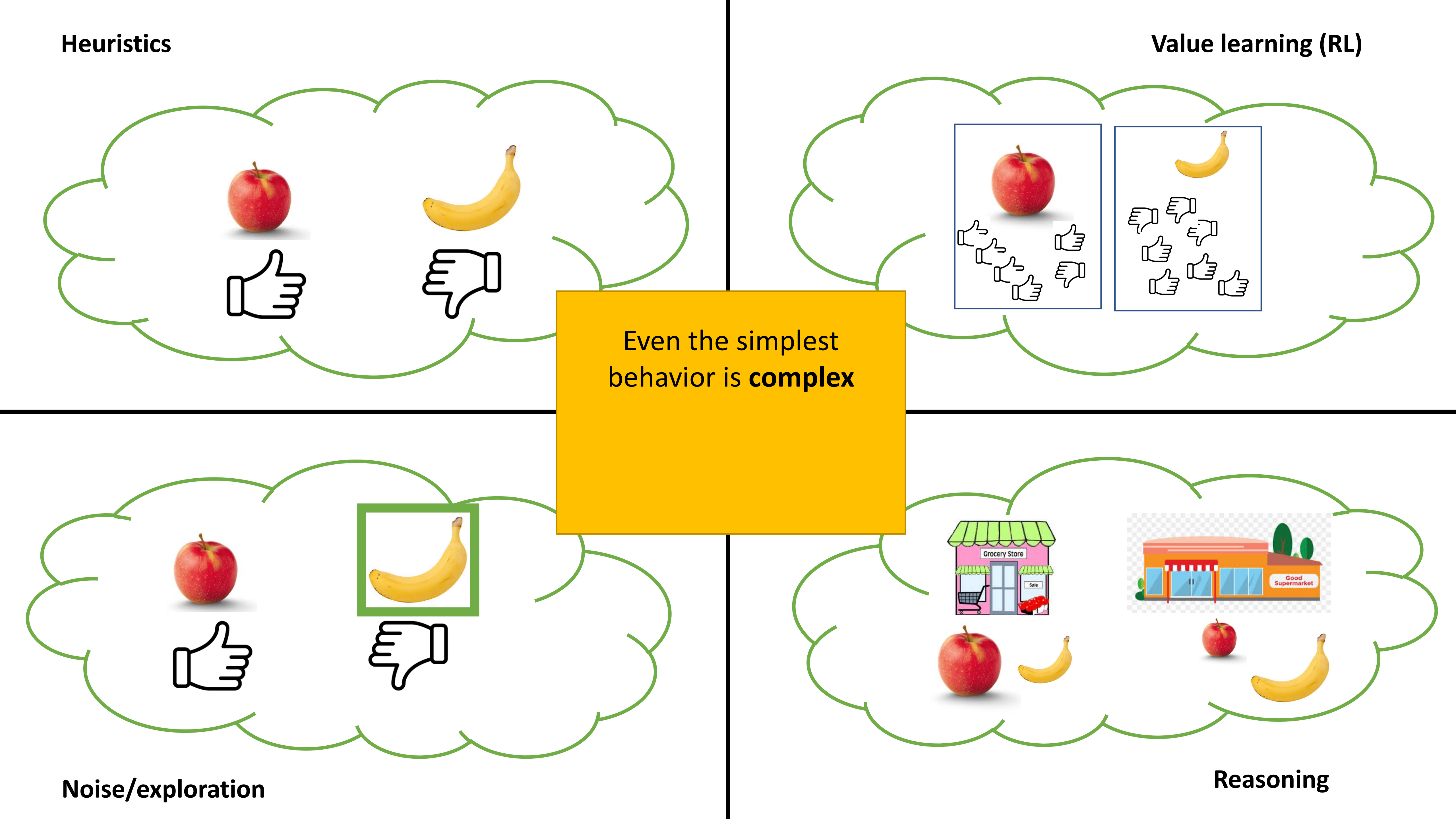


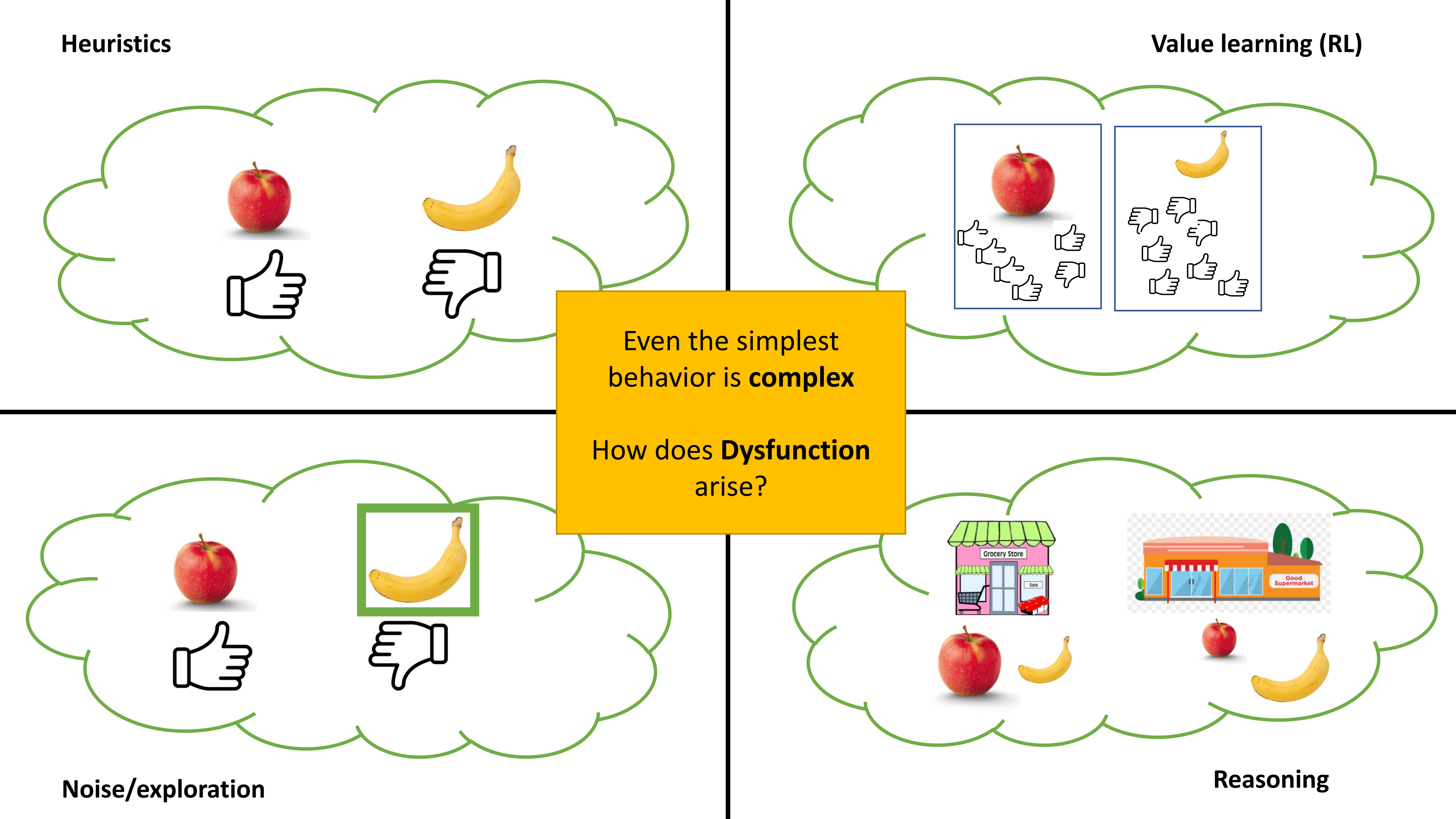
Heuristics

Value learning (RL)

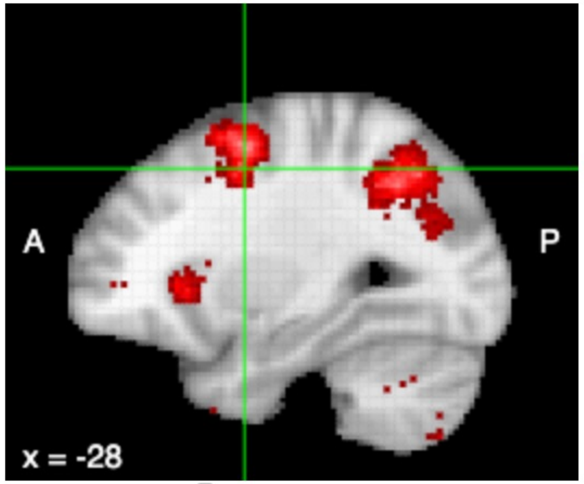



Heuristics
Value learning (RL)
Grocery Store
Sale
Good Supermarket
Reasoning

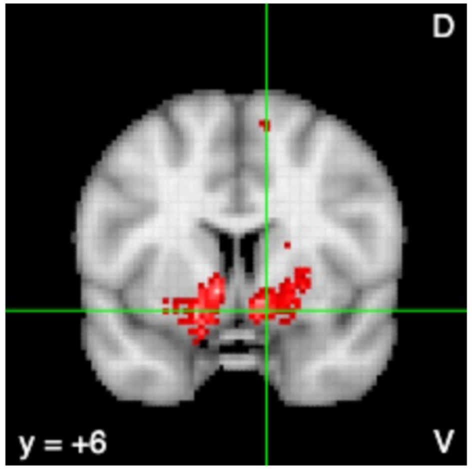

Heuristics
Value learning (RL)
Even the simplest behavior is **complex**
Grocery Store
Sale
Good Supermarket
Noise/exploration
Reasoning


Heuristics
Value learning (RL)
Even the simplest behavior is **complex**
How does **Dysfunction** arise?
Noise/exploration
Reasoning
Grocery Store
Sale
Good Supermarket

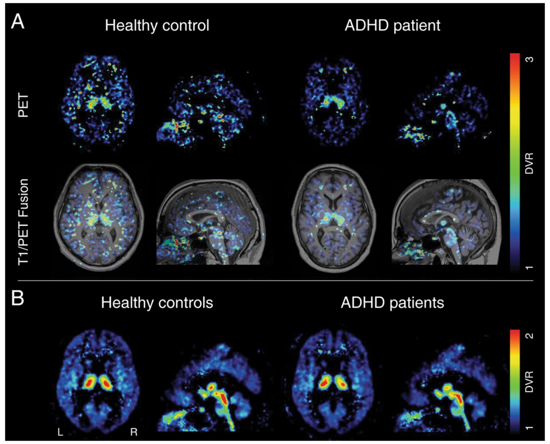
Heuristic – e.g. Working memory

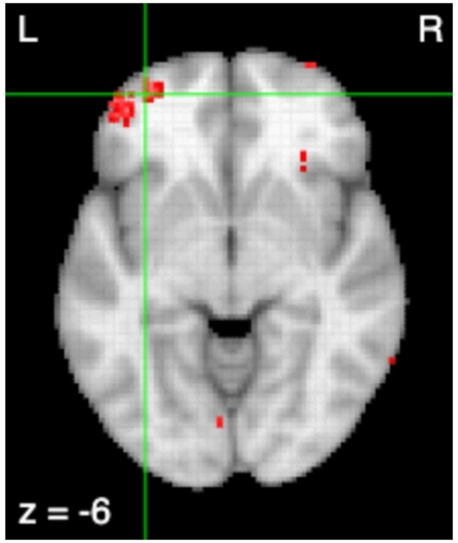


Value learning (RL)



How does the **brain** support
Complex behavior?
How does **dysfunction** arise?





Noise/exploration

Reasoning

Working memory

- Load effects
- Short term delay effects

Value learning (RL)

- Effects of cumulative reward
- Effects of positive/negative outcomes

Need to design **translatable** tasks that can **identify** the underlying processes.

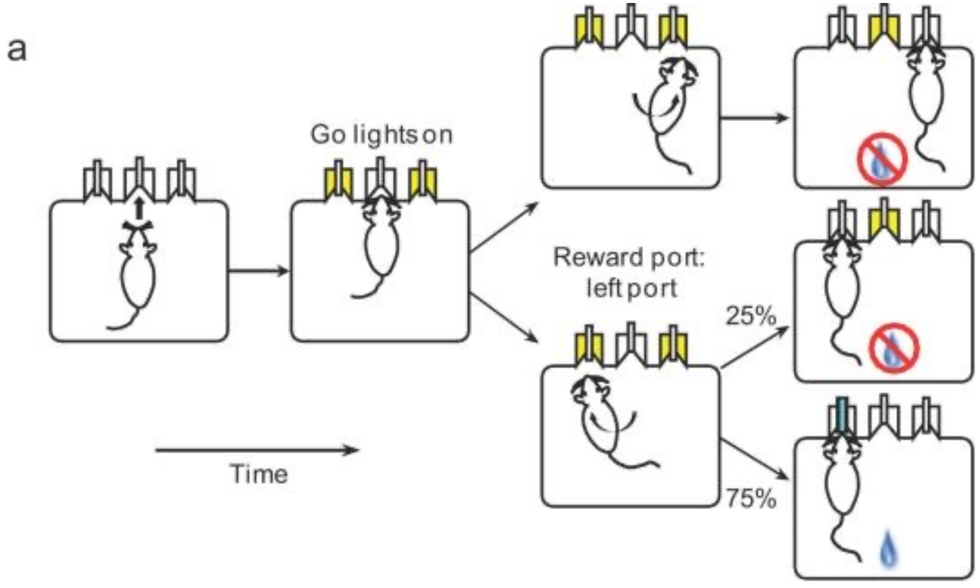
Choice variability

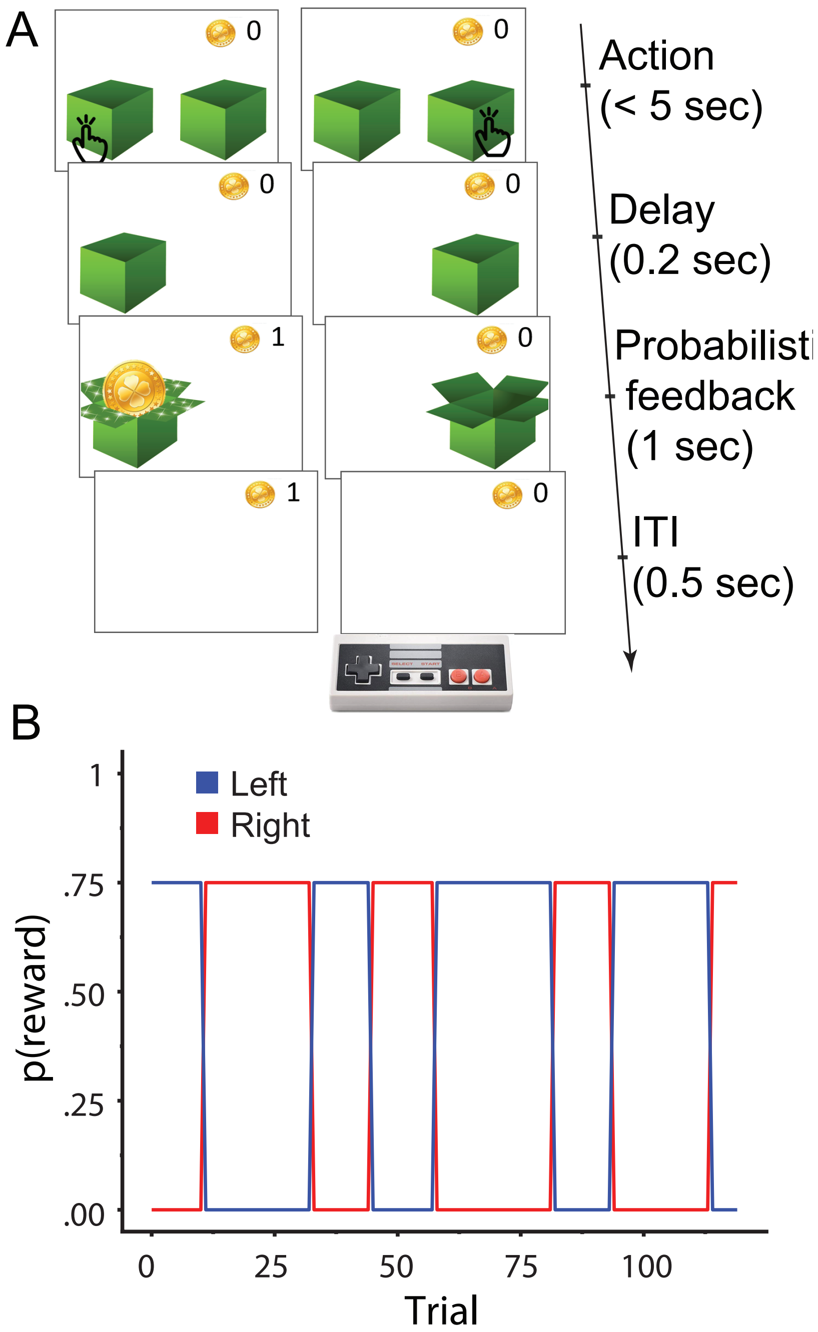
Noise/exploration

Fast inference in uncertain/ changing environments

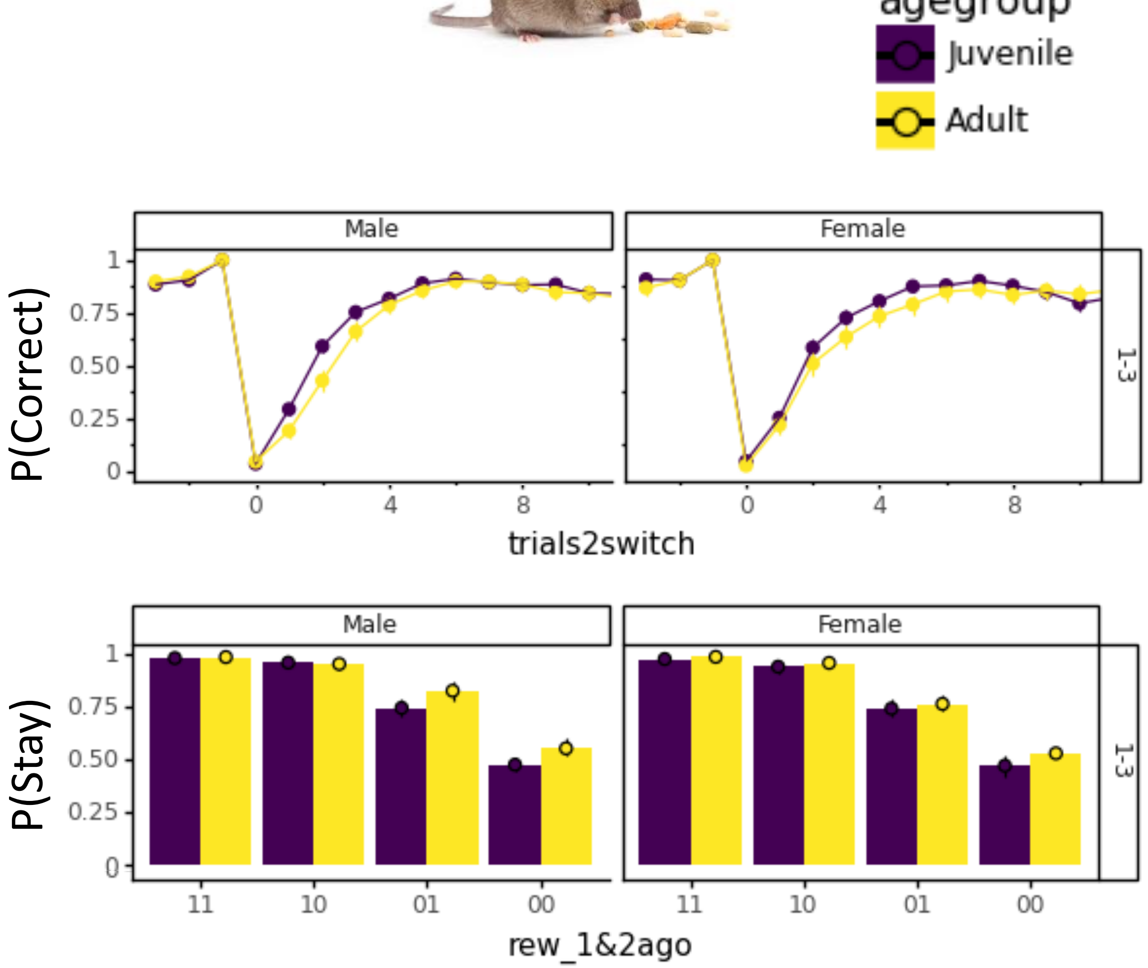
Reasoning

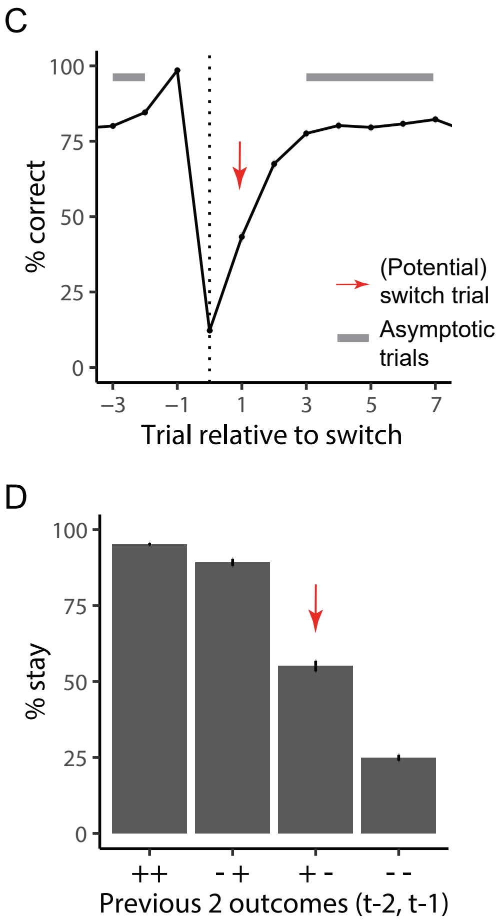
A task example:
Probabilistic reversal learning





Tai et al 2012
Eckstein et al, 2022





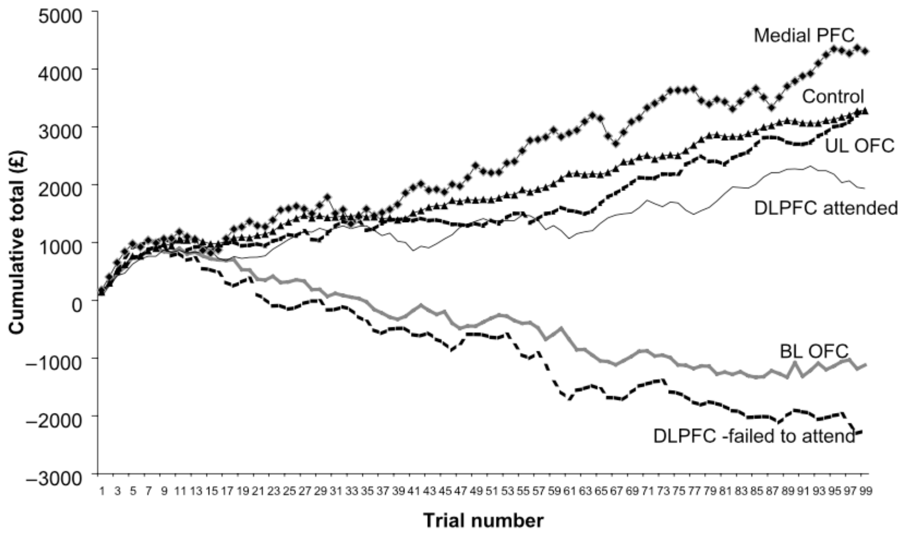
- Task learned very fast by mice and humans.
- Comparable mice and human data makes this task highly ***translatable.***
- Sensitive to many relevant individual differences (e.g. impulsivity, age, …)

Eckstein et al, in prep
(with Linda Wilbrecht)

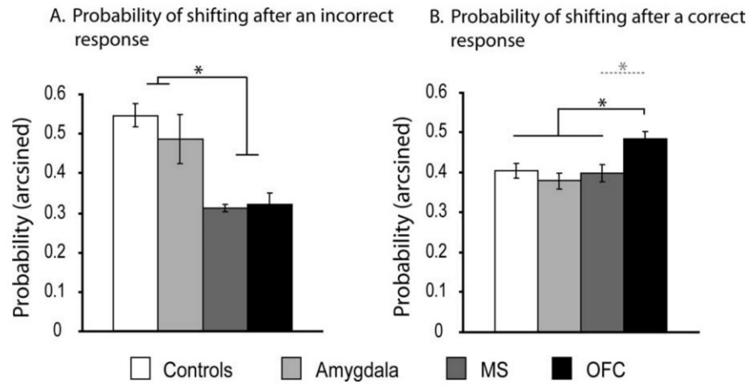
Brain

- Orbitofrontal cortex?
- DLPFC?
- Dopamine function?
- Striatum?

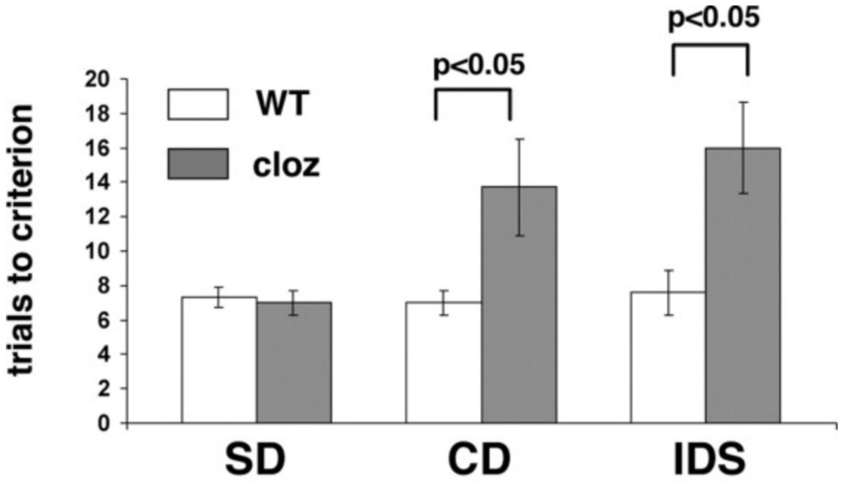
Hornak et al 2004

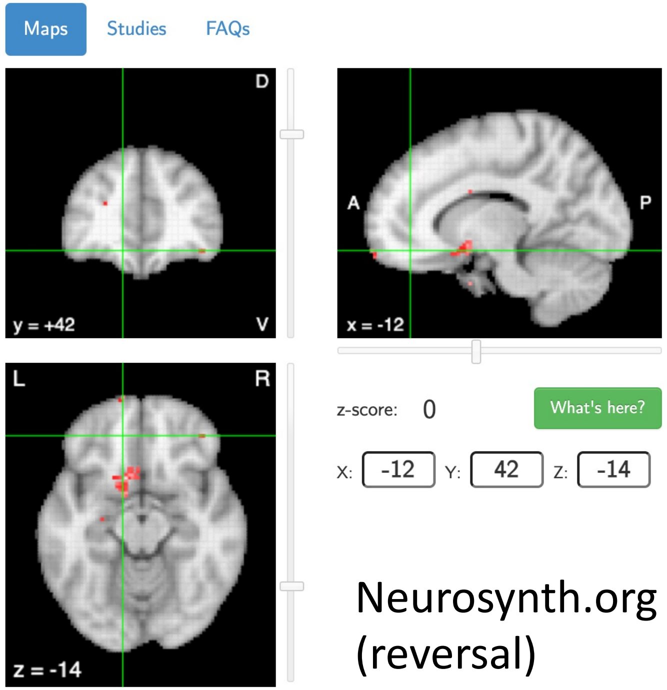


Clarke et al 2008



DeSteno & Schmauss 2009





Neurosynth.org (reversal)

Probing the cognitive mechanisms with computational modeling of behavior

Heuristic:

Win-Stay Lost-Shift + noise

Heuristic:

Switch after two negative outcomes + noise

More complex heuristics

Value learning:

Track and update the value of each option
Pick the best (+noise)

More complex value learning

(e.g. counterfactual learning, different positive/negative outcome valuation, etc…)

Reasoning/inference:

Track and update a belief about which box is “correct”, based on a model of the task
Pick the correct box (+noise)

More complex inference:

e.g. also infer the model of the task
…

Heuristic:

$P(\text{switch}_t \mid r_{t-1} = 1) = 1 - \varepsilon$
$P(\text{switch}_t \mid r_{t-1} = 0) = 1 - \varepsilon$

Heuristic:

Switch after two negative outcomes + noise

More complex heuristics

Value learning:

$Q(\text{side}_t) \leftarrow Q(\text{side}_t) + \alpha(r_t - Q(\text{side}_t))$

Choice ~ softmax(β Q)

More complex value learning

(e.g. counterfactual learning, different positive/negative outcome valuation, etc…)

Reasoning/inference:

$B_{t+1}(\text{side}_t) \sim B_t(\text{side}_t) * P(r_t \mid \text{side}_t)$

Choice ~ softmax(β B)

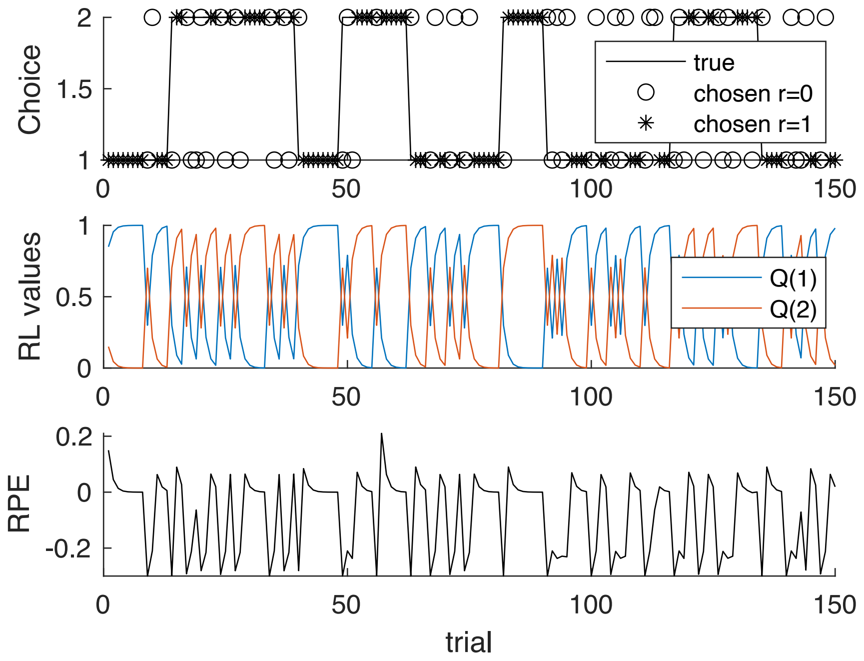
More complex inference:

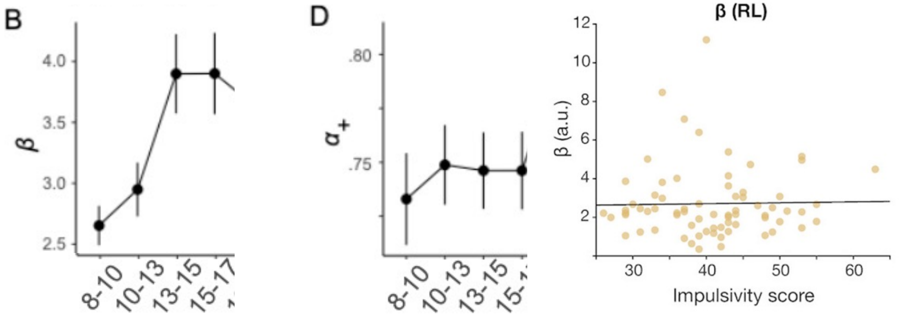
e.g. also infer the model of the task

…

Benefits of computational models

- Formal, quantitative predictions
- Latent variables as markers of processes:
  - E.g. do **reward prediction errors** relate to dopaminergic signaling in this task?
  - E.g. does the **belief** about the latent state relate to OFC signaling?
- Parameters as markers of tuning of the processes:
  - Effect of experimental conditions:
    - E.g. does learning rate change with administration of dopaminergic drug?
  - Effect of individual differences:
    - E.g. does exploration change with age/disease model group?

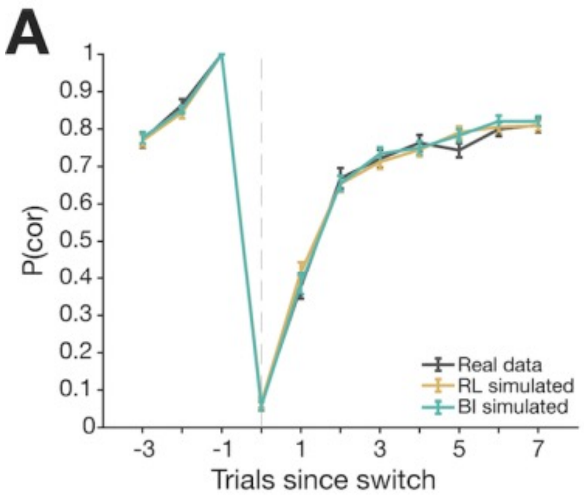


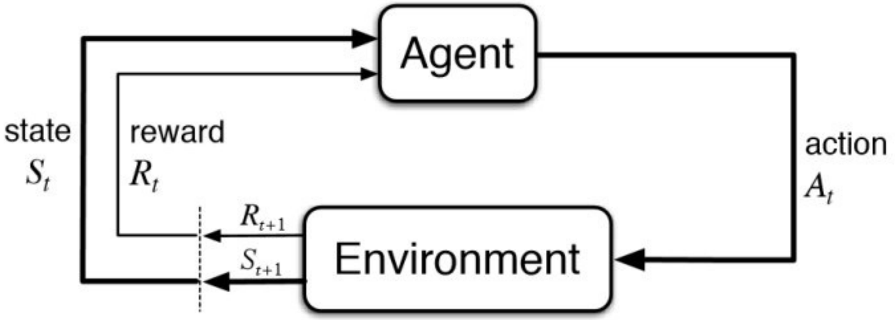


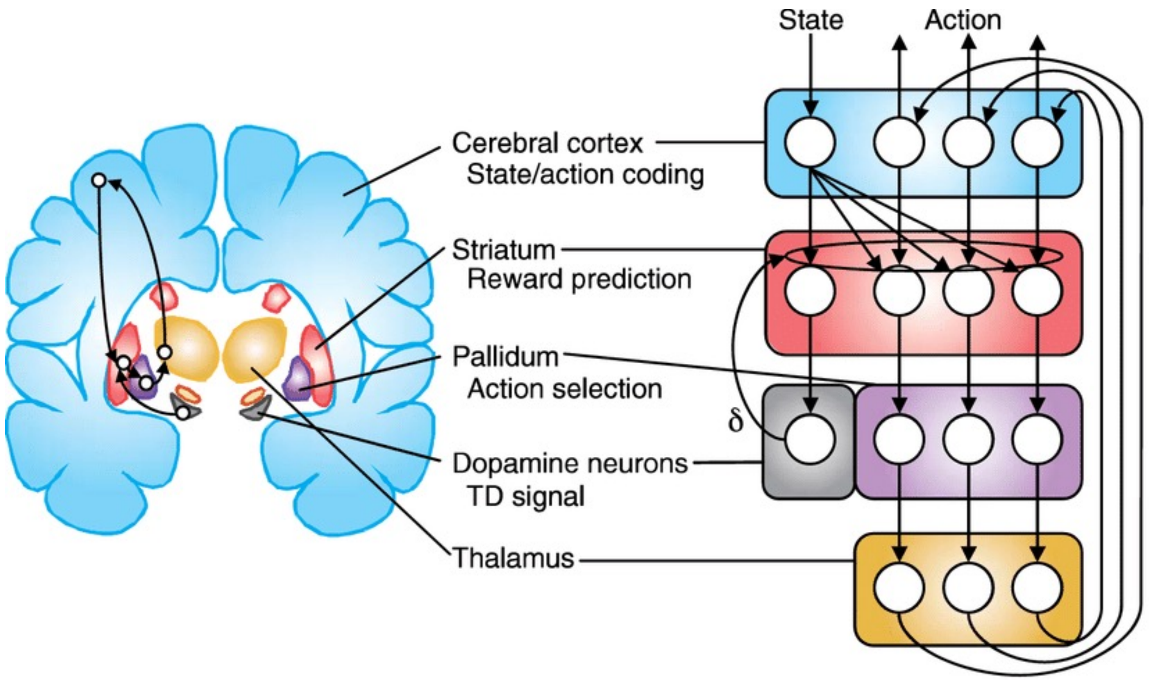
Eckstein et al, 2022; Zou et al 2022

Example of success – Reinforcement learning

- Dopamine encodes reward prediction error
- Striatum encodes values/policies, learned via dopamine-dependent plasticity.







Reinforcement learning (RL) models capture both behavior and a brain mechanism.

Sutton & Barto 1998, Doya et al 2008, Zou et al in prep

Caveats –
Risks of computational modeling

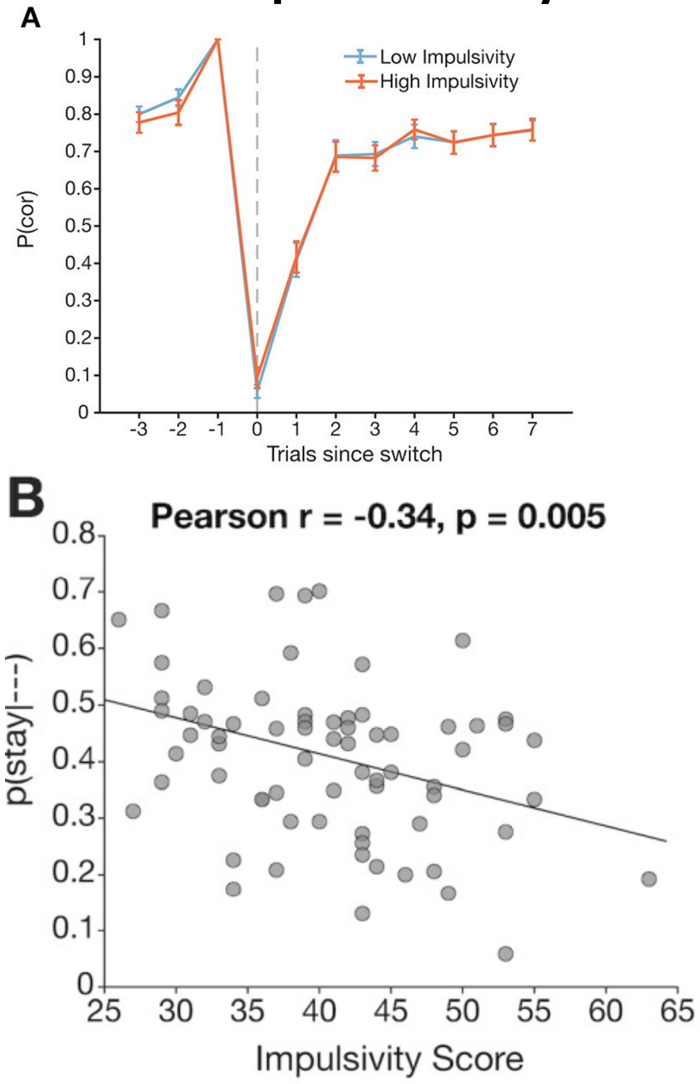
1. Identifying computations

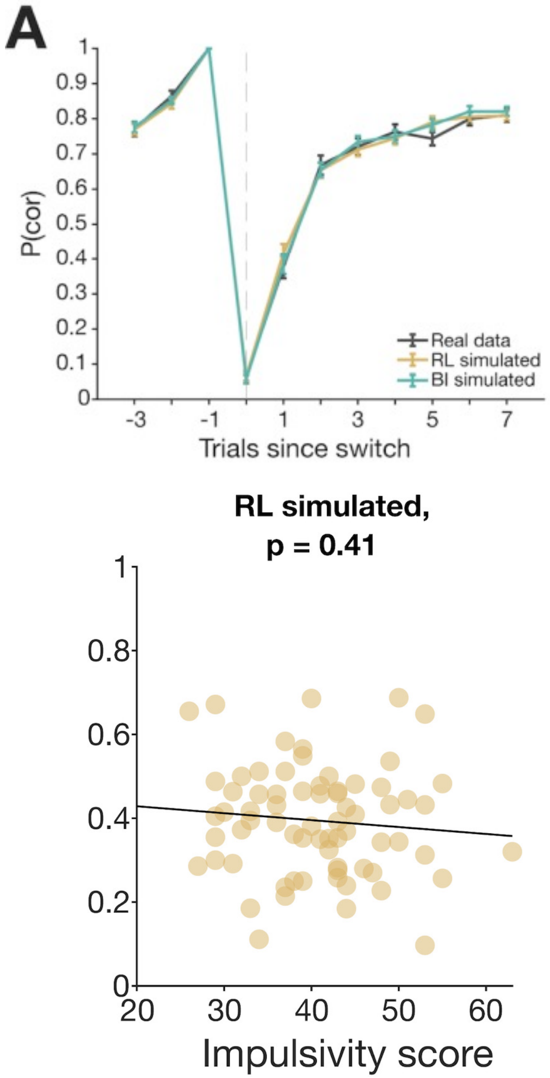
2. Mapping computations to mechanisms

1. Identifying computations
    a. **Even simple behavior is complex**

Caveat 1 – behavioral complexity

- Even simple behavior is complex
- Models need to trade off simplicity and expressivity
- Behavior may be more complex than can be identified in a given task with a given model



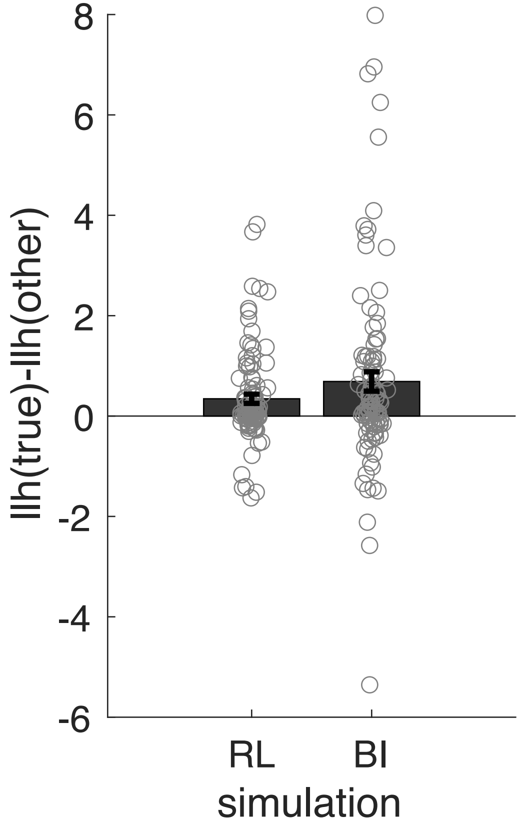


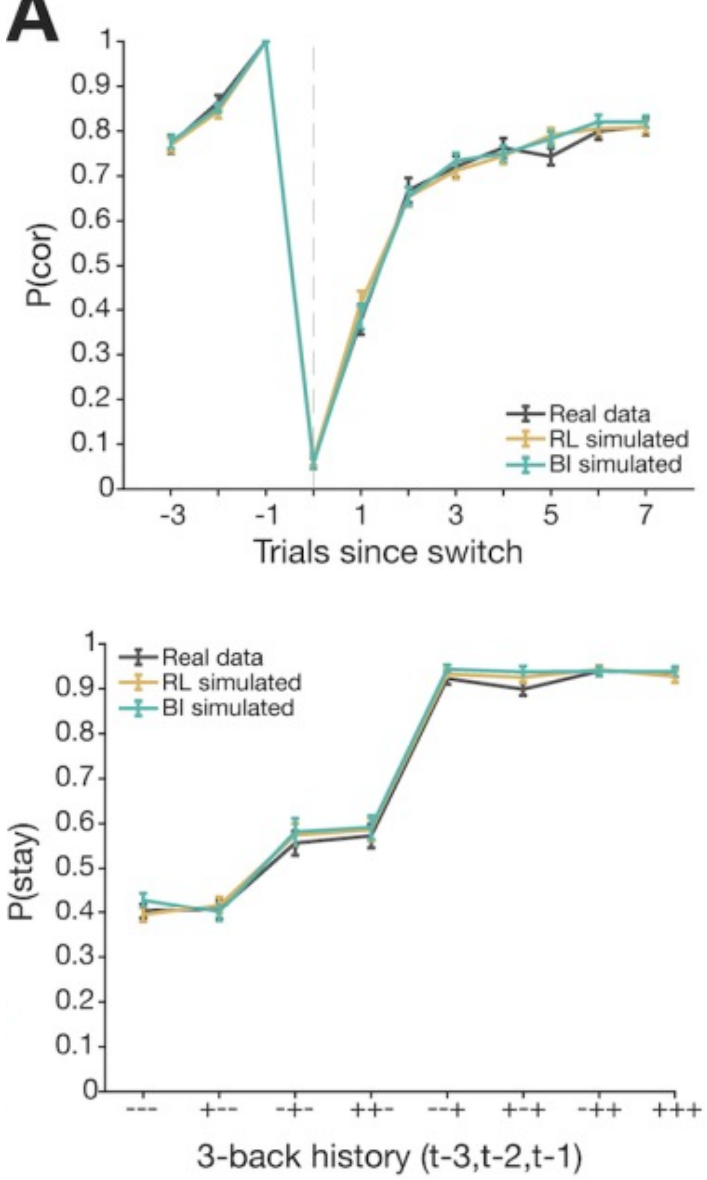
Zou et al, 2022

1. Identifying computations
    a. Even simple behavior is complex
    b. **Can we identify computations that support cognition from simple behavioral tasks?**

Caveat 2 – model identifiability

- Two very different models may capture behavior similarly well (while remaining technically identifiable)
- Which model should be used to probe the brain/individual differences?





Zou et al, 2022

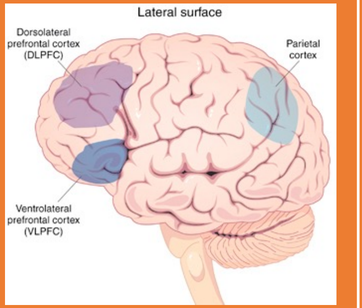
Caveat 2 – model identifiability

- Model vs. tasks
  - Model identifiability
  - Solution: design tasks with models in mind: some processes will not be identifiable in some tasks! Ensure qualitatively different predictions between competing theories.

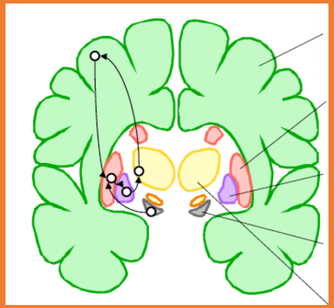
- An example - RLWM

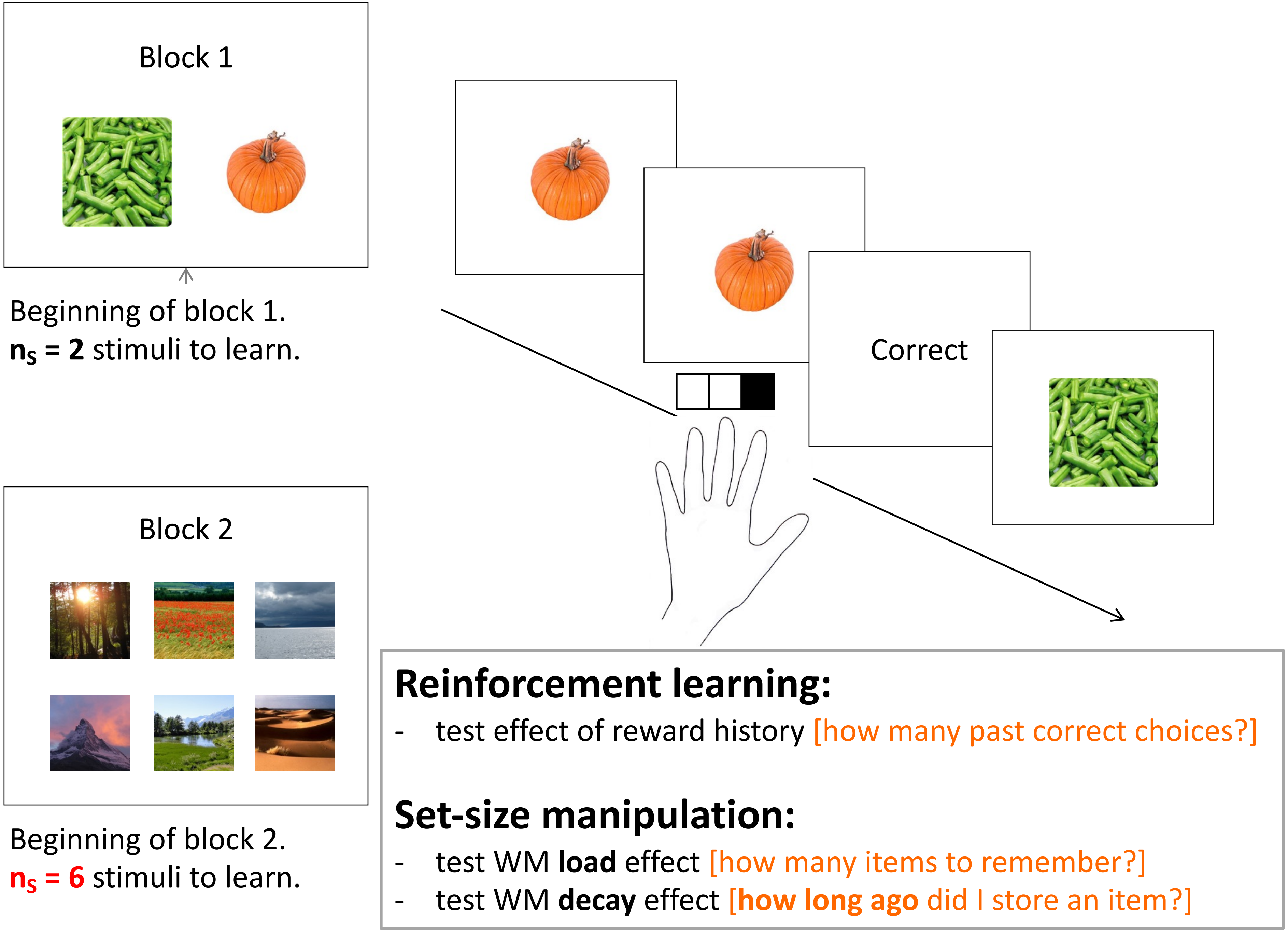
Two **redundant** components for learning: optimized for different trade-offs

WM



RL





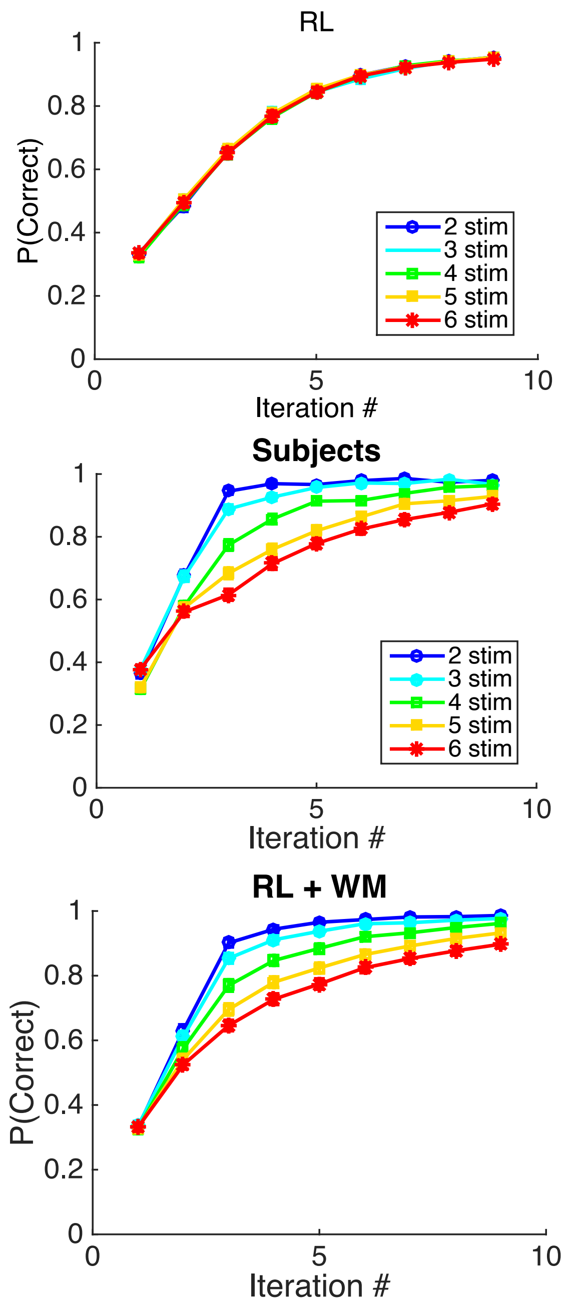
Block 1
Beginning of block 1.
n_S = 2 stimuli to learn.
Correct
Block 2
Beginning of block 2.
n_S = 6 stimuli to learn.
Reinforcement learning:
- test effect of reward history [how many past correct choices?]
Set-size manipulation:
- test WM load effect [how many items to remember?]
- test WM decay effect [how long ago did I store an item?]

Experimental results: n=78

Effect of set-size on learning performance is not accounted for by classic RL

WM
- Fast & flexible
- Forgetful
- Capacity limited

RL
- Slow & inflexible
- Effortless
- Broad and robust



Collins & Frank, 2012

1. Identifying computations
    a. Can we identify computations that support cognition from simple behavioral tasks?
    b. Even simple behavior is complex

2. **Mapping computations to mechanisms**
    a. **Overly expressive models**

Caveats 3 – computational modeling pitfalls

- RL is a good model of dopaminergic signaling
  - In a pavlovian learning task
  - (extended to other RL tasks)

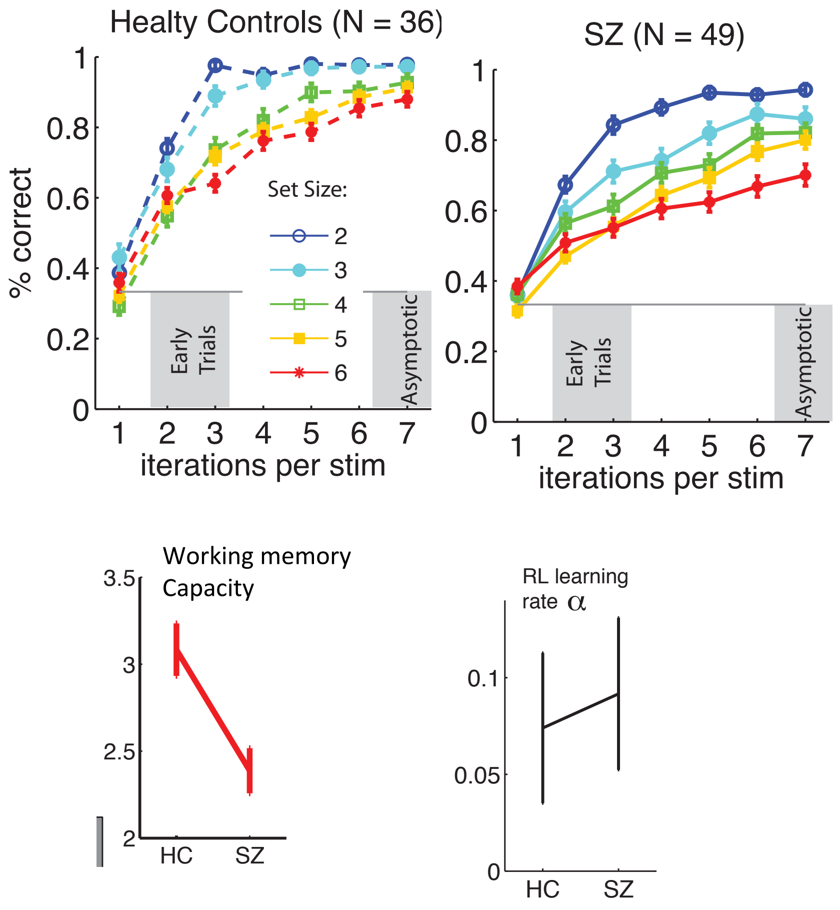
- RL captures behavior well in a probabilistic reversal task

- Should I conclude that RL variables in the probabilistic reversal task will be informative about the brain's RL system?
  - **Or is the RL family of model so expressive that it can capture both, but without actually relating them?**

Caveats 3 – overly expressive models

- Behavior that is well captured by RL models often relies on **non-RL processes** as well as RL processes
  - Working memory, episodic memory

- If this is not taken into account, the RL processes inferred by modeling behavior will be **contaminated.**



Collins et al, JoN, 2014

1. Identifying computations
   a. Can we identify computations that support cognition from simple behavioral tasks?
   b. Even simple behavior is complex

2. Mapping computations to mechanisms
   a. Overly expressive models
   b. **How generalizable are findings?**

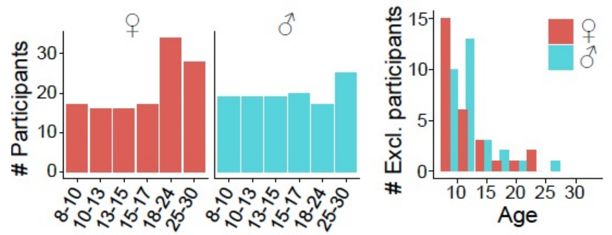
Caveats 4 – generalizability

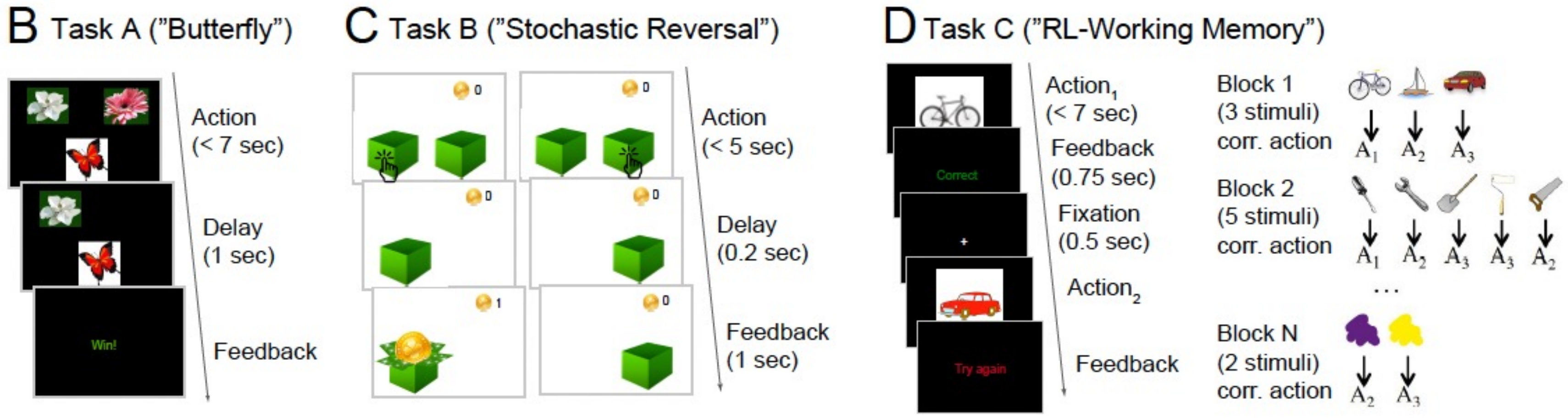
- Such statements are frequent:
  - “RL learning rates are impaired in population X”
  - “positive, but not negative, learning rates were correlated with symptom Y”

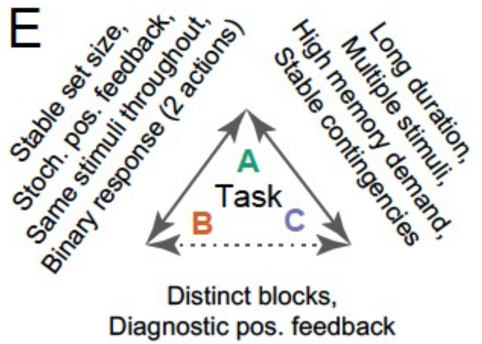
- Is this valid or too general?
- Will findings observed in one RL task and model generalize to a different (but similar) RL task and model?

Maria Eckstein, PhD
(now at DeepMind)

Comparing individual differences in three learning tasks





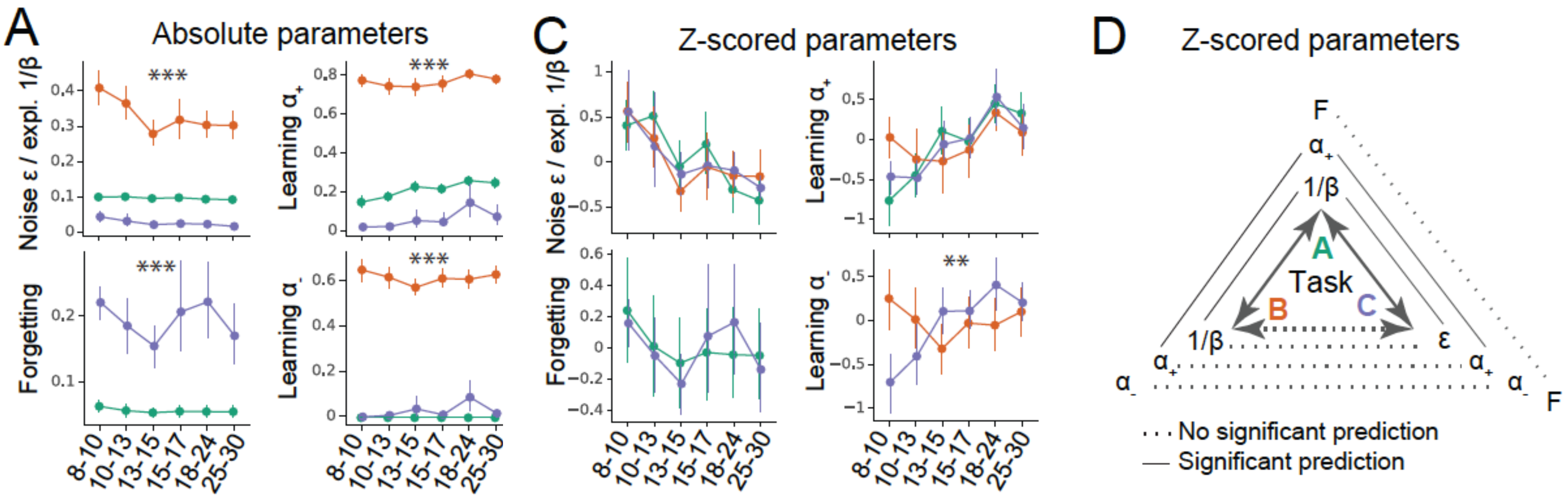


F

	$1/\beta$ and ε	α_+	α_-	Forget F	Persist. P	WM pars.
Model A	Y ($1/\beta$)	Y	Fixed at 0	Y	—	—
Model B	Y ($1/\beta$)	Y	Y	—	Y	—
Model C	Y (ε)	Y	Y	Y	—	Y

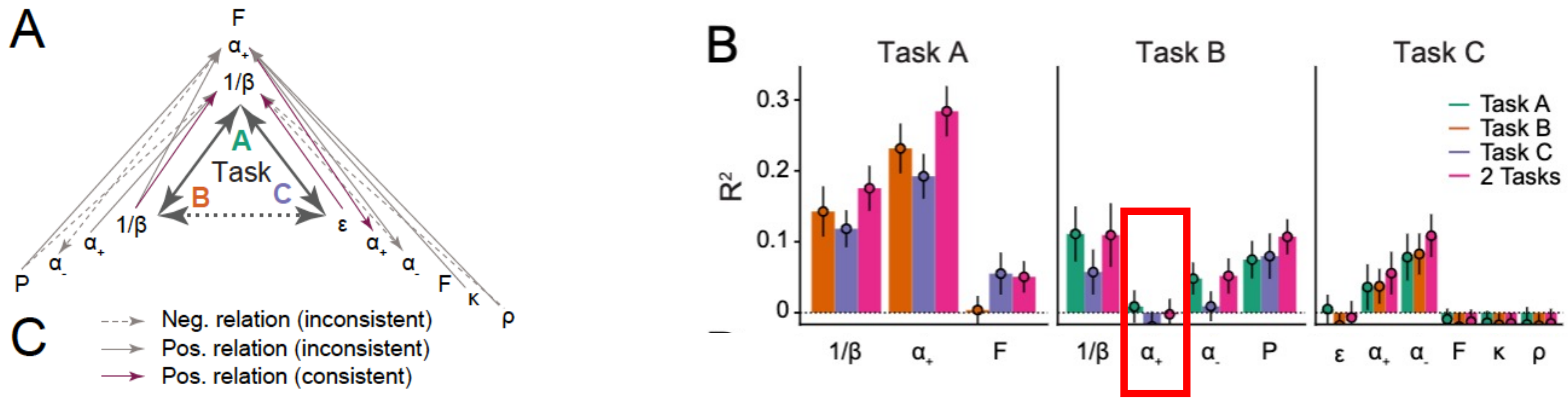
Eckstein et al, in press, eLife

Comparing individual differences in three learning tasks



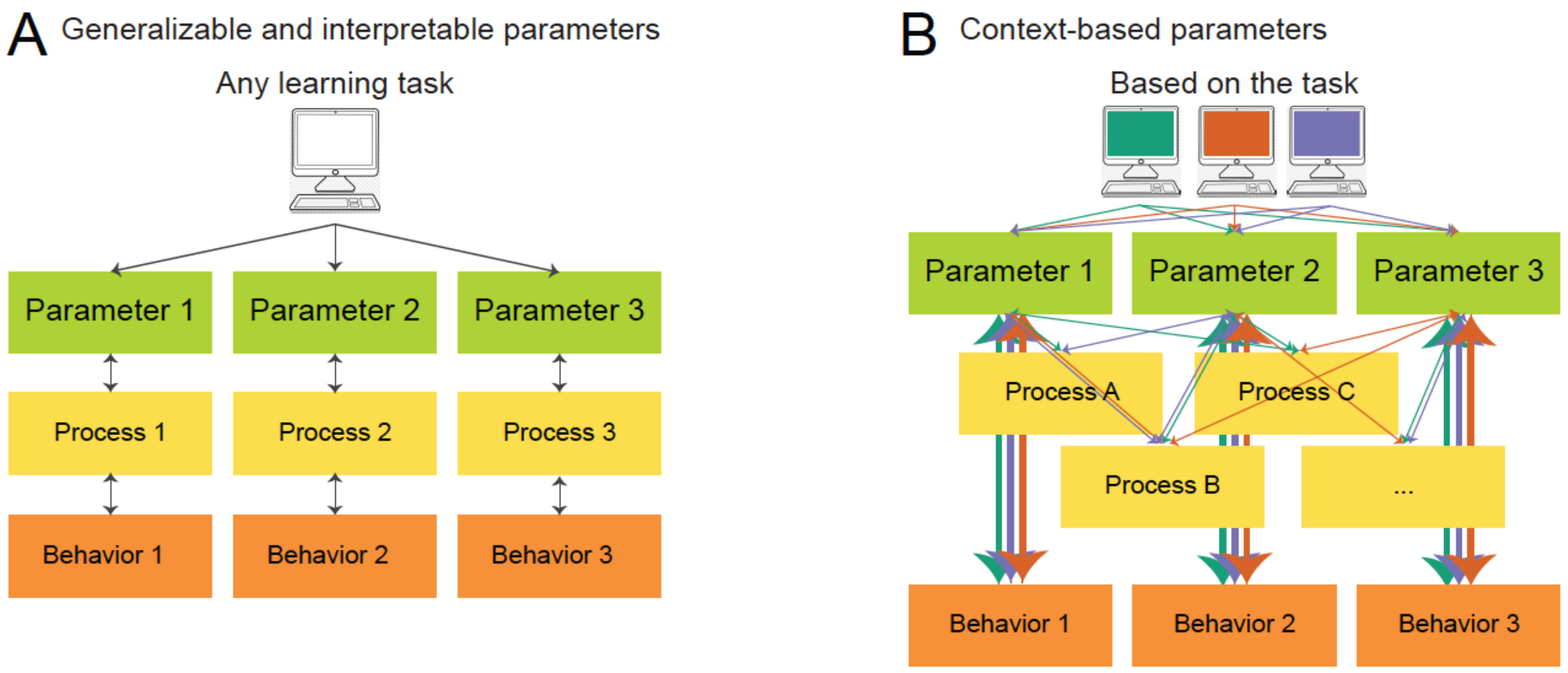
Eckstein et al, in press, eLife

Comparing individual differences in three learning tasks – low generalizability



Variants of the same learning task, all well fit by RL, can recruit different underlying processes, to the point that **inferred RL processes share no variance**.

Eckstein et al, in press, eLife


A
Generalizable and interpretable parameters
Any learning task
Parameter 1
Parameter 2
Parameter 3
Process 1
Process 2
Process 3
Behavior 1
Behavior 2
Behavior 3
B
Context-based parameters
Based on the task
Parameter 1
Parameter 2
Parameter 3
Process A
Process C
Process B
...
Behavior 1
Behavior 2
Behavior 3

Discussion – 1/2

- Computational modeling of behavior is an **essential tool** to probe the brain, individual differences, and dysfunction.
- It formalizes **quantitatively theories** about the cognitive processes that underly behavior.
- It provides access to **latent variables** to probe these cognitive processes as neural mechanisms.
- It provides **parameters** that capture how the processes are tuned and can be related to neural function.

Discussion - 2/2

- However, computational modeling is **not a silver bullet**.
- Computational modeling of behavior is only as powerful as the task it attempts to capture – **experimental design** remains essential.
- Computational modeling is **less generalizable** than often assumed (Eckstein & Collins 2021; Eckstein et al 2022) – being aware of the limits is important.
- Computational models can give **false confidence** that we are bridging from brain to behavior. Testing this bridge in different contexts is essential to not drawing false conclusions.

Thank you!

CCN lab (current and former)

CCN LAB

https://ccn.berkeley.edu
@Anne_on_tw
@ccnlab

Some collaborators :

Maria Eckstein

Linda Wilbrecht

...

Funding

- NSF SL-CN 1640885
- NIMH 119383-01
- NSF 2020844
- Helman Fellows Fund