ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

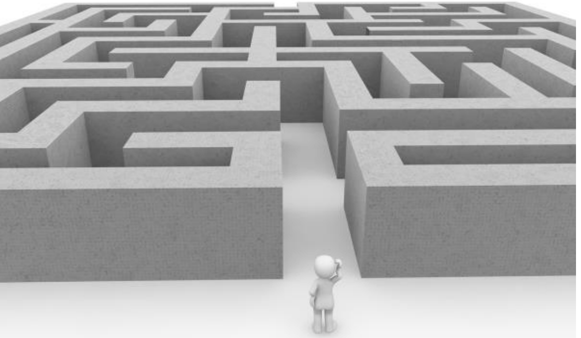
Systems for instrumental learning: habitual and goal-directed

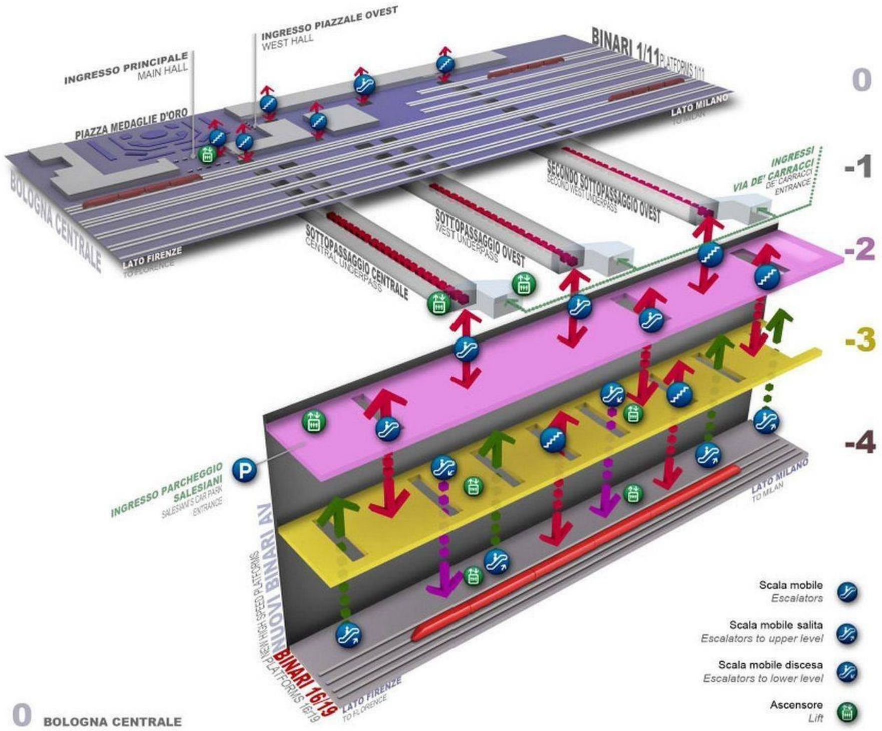
Cognition and Neuroscience
Academic year 2023/2024

Francesca Starita

francesca.starita2@unibo.it

Uscite Exits Via de' Carracci Quartiere Bolognina
P.zza Medaglie d'Oro
Binari Platforms 1-11 1-7 Ovest 1-4 Est 16-19
P1





0 **BOLOGNA CENTRALE**

-1 **PIANO SOTTOPASSAGGI**
UNDERPASS LEVEL

-2 **PIANO KISS&RIDE**
KISS & RIDE LEVEL

-3 **PIANO NUOVA HALL ALTA VELOCITÀ**
NEW HIGH SPEED TRAINS HALL

-4 **PIANO NUOVI BINARI ALTA VELOCITÀ (16/19)**
NEW HIGH SPEED PLATFORMS LEVEL

Scala mobile
Escalators

Scala mobile salita
Escalators to upper level

Scala mobile discesa
Escalators to lower level

Ascensore
Lift

Scala fissa
Stairs

Parcheggio
Car Park

Percorsi in ingresso/uscita nuovi binari AV (16/19)
New high speed platforms (16/19) entrance/exit passageways 🟥🟥🟥

Percorsi ingresso nuovi binari AV (16/19)
Passageways to new high speed platforms (16/19) 🟪🟪🟪

Percorsi di uscita da nuovi binari AV (16/19)
Exit passageways from new high speed platforms (16/19) 🟩🟩🟩

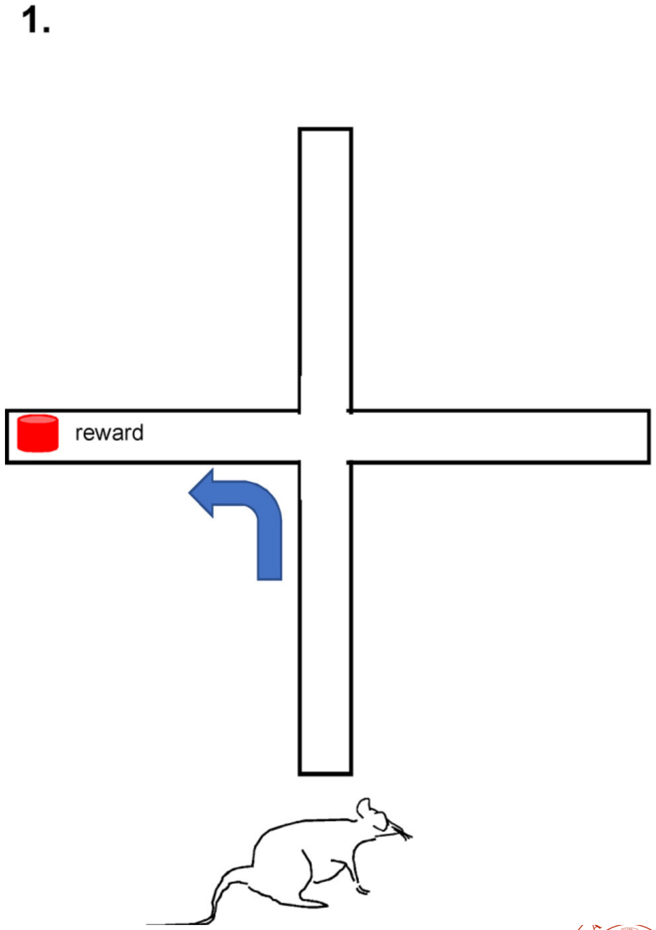
Different learning strategies: place/map vs response strategy

In a plus or T maze, animals start in a particular location (e.g. a south arm) and from there they must enter a different arm, e.g. the west arm, to find a food reward.

In such a situation, the animals could use **one of two strategies** to solve the task.

1. They could form a **cognitive map** and learn that **reward is located in the west location** and so travel there to find it.
2. They could learn that a particular sequence of motor responses, ultimately **turning left, leads to reward**.

HOW WOULD YOU TEST WHICH STRATEGY WAS USED?



Corbit, Laura H. "Understanding the balance between goal-directed and habitual behavioral control." *Current opinion in behavioral sciences* 20 (2018): 161-168.

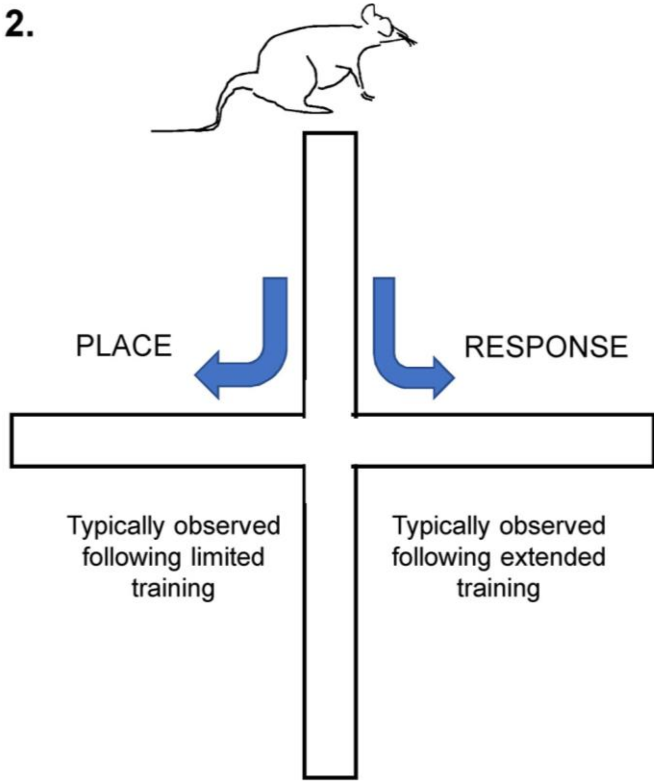
Different learning strategies: place/map vs response strategy

To find out what strategy was used, the animal is placed in a novel location, e.g. the north arm, and allowed to choose between arms from there.

1. If they navigate to the west arm --> they have learned to solve the maze using a **cognitive map/place strategy**.
2. If they turn left --> they have learned to solve the maze using a **response strategy** (they perform the same response that led to reward in the past).

Early studies found that **rats learned more readily about places** (i.e. place strategy) than about the particular response sequence (i.e. response strategy).

Nonetheless, given extensive training, rats come to rely on a response strategy.


2.
PLACE
RESPONSE
Typically observed following limited training
Typically observed following extended training

Corbit, Laura H. "Understanding the balance between goal-directed and habitual behavioral control." *Current opinion in behavioral sciences* 20 (2018): 161-168.

Multiple systems contribute to learning and controlling behavior in animals

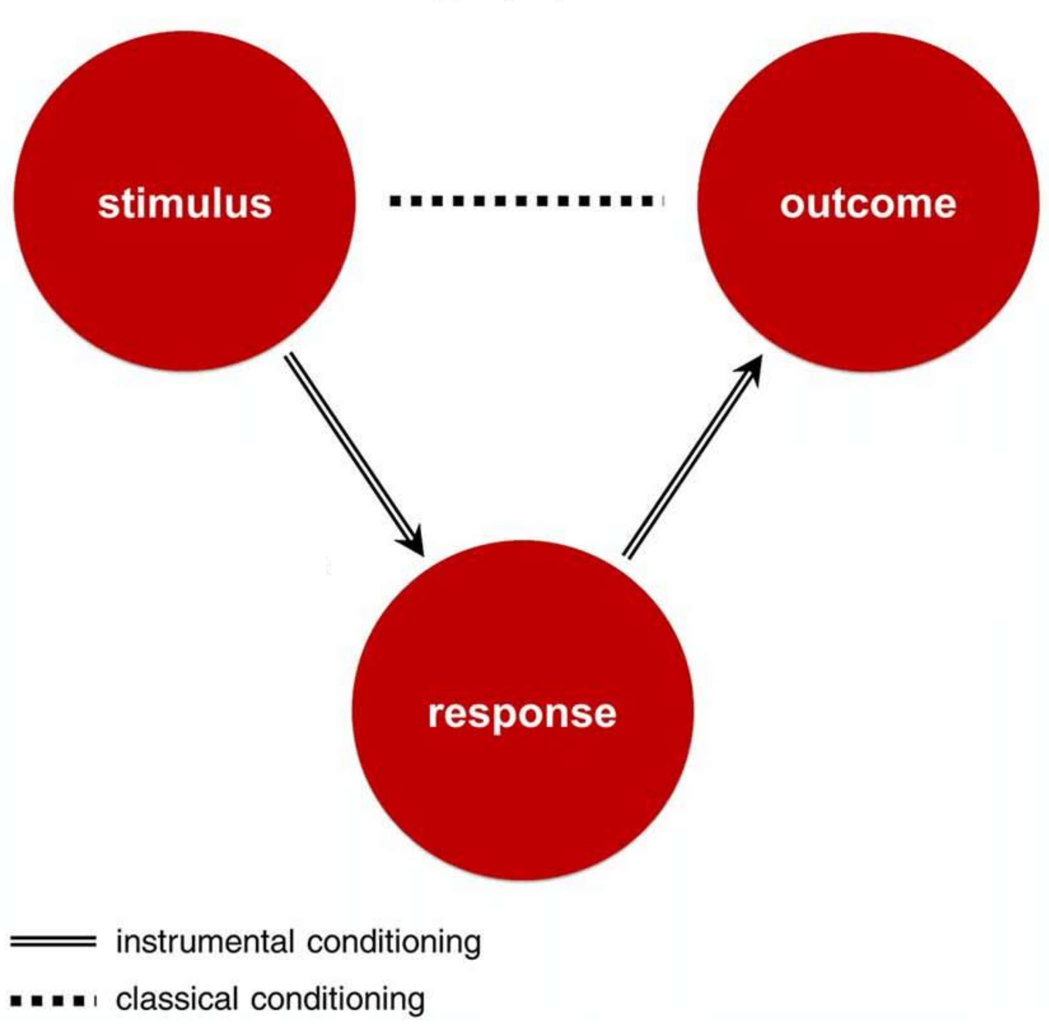
Three learning systems enable organisms to draw on previous experience to **make predictions** about the world and to **select behaviors** appropriate to those predictions:

1. a **Pavlovian system** that learns to predict biologically significant events so as to trigger appropriate responses;

Instrumental system that comprises

2. a **habitual system** that learns to repeat previously successful actions;
3. a **goal-directed system** that evaluates actions on the basis of their specific anticipated consequences.

Predictions are for control



Instrumental learning (or operant conditioning) involves associating an action with an outcome

Thorndike's Law of effect

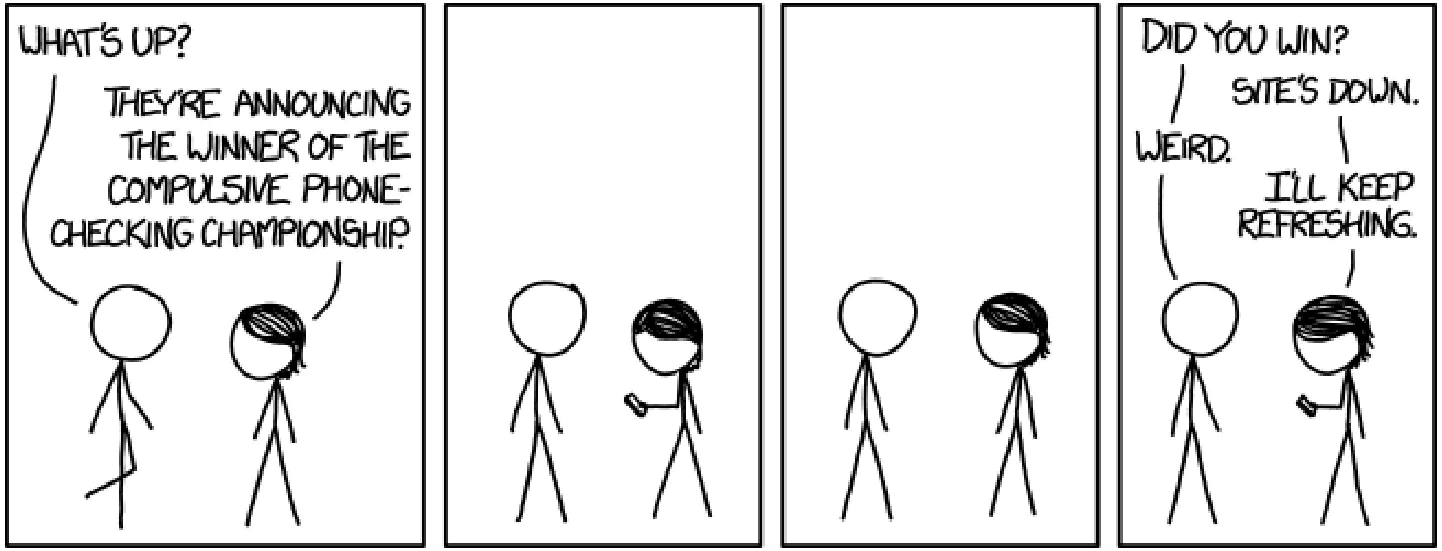
"Of several **responses** made to the same situation, those which are accompanied or closely **followed by satisfaction** to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they **will be more likely to recur**; those which are accompanied or closely **followed by discomfort** to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they **will be less likely to occur**. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond." (Thorndike, 1911)

We act to produce outcomes that are desirable or to avoid those that are harmful or aversive

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

But how do we select (decide) which is the appropriate action to take?

- Are we flexible in the actions we take?
- Do we choose the action to take with the goal in mind or do we automatically select actions based on previous (rewarded) experiences?
- Are our choices directed by the goal we want to achieve or are they automatically/habitually triggered based on our past (rewarded) experiences?


WHAT'S UP?
THEY'RE ANNOUNCING THE WINNER OF THE COMPULSIVE PHONE-CHECKING CHAMPIONSHIP.
DID YOU WIN?
SITE'S DOWN.
WEIRD.
I'LL KEEP REFRESHING.

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

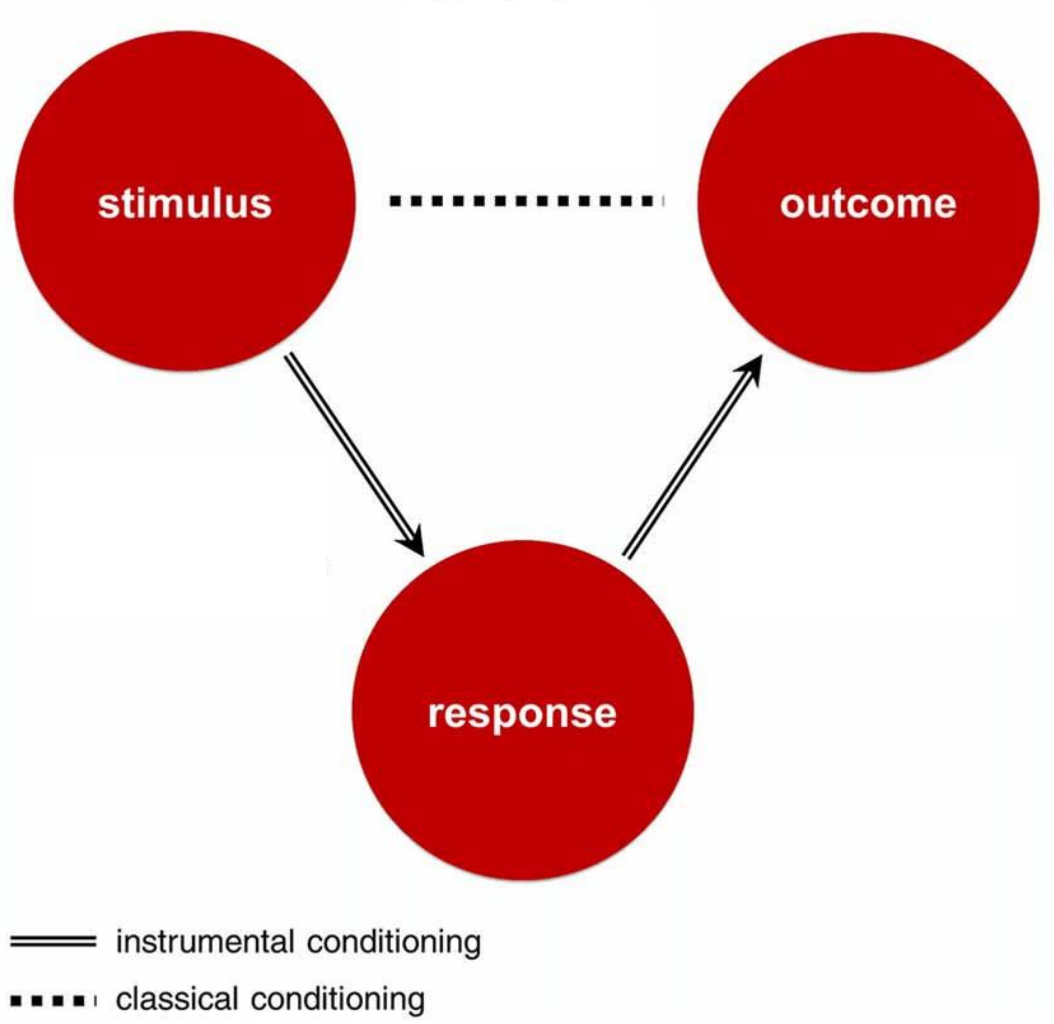
Multiple systems contribute to learning and controlling behavior in animals

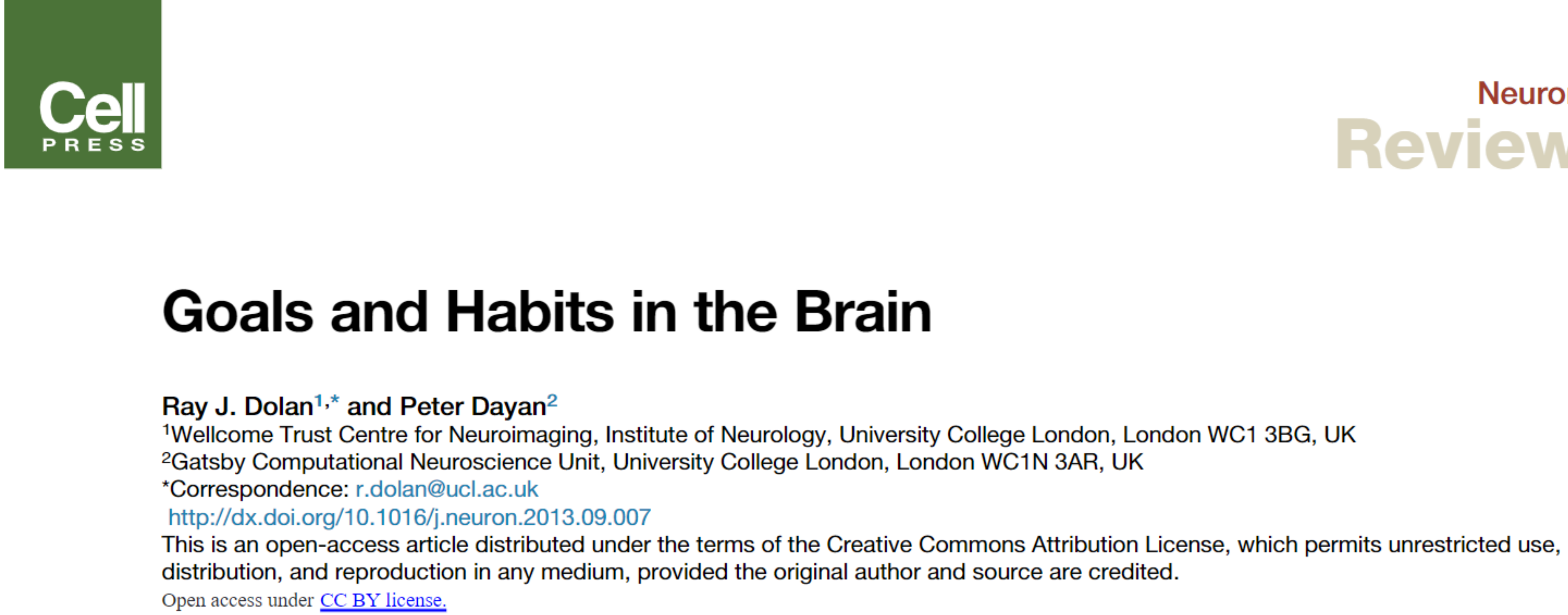
- Are we flexible in the actions we take?
- Do we choose the action to take with the goal in mind or do we automatically select actions based on previous (rewarded) experiences?
- Are our choices directed by the goal we want to achieve or are they automatically/habitually triggered based on our past (rewarded) experiences?

Instrumental system that comprises

2. a **habitual system** that learns to repeat previously successful actions;
3. a **goal-directed system** that evaluates actions on the basis of their specific anticipated consequences.

Predictions are for control





Neuron
Review

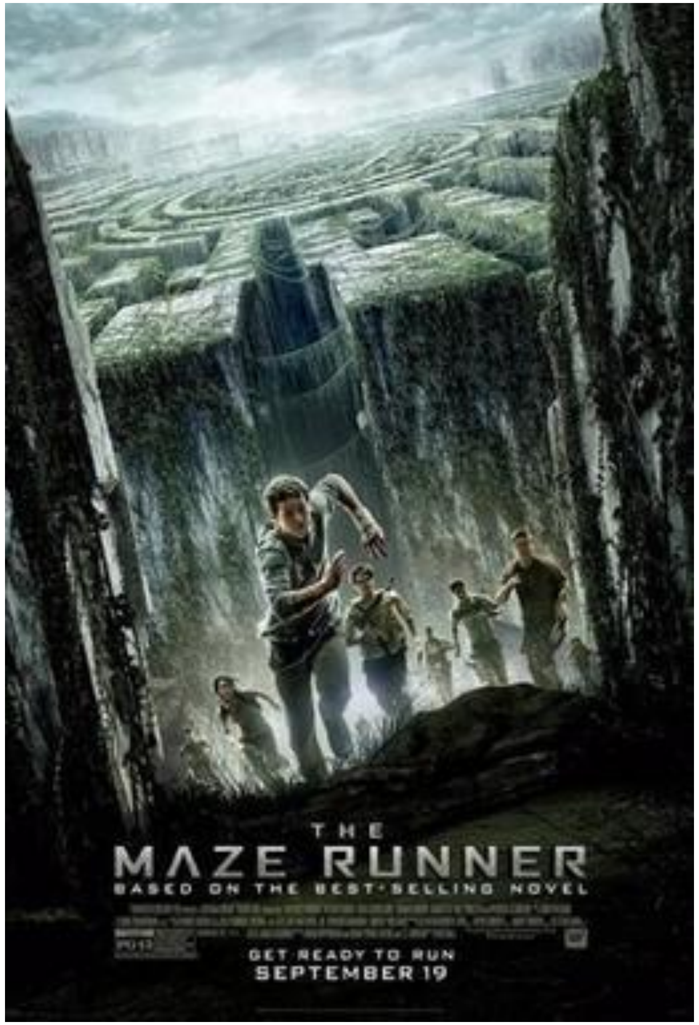
Goals and Habits in the Brain

Ray J. Dolan[1,*] and Peter Dayan[2]
[1]Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1 3BG, UK
[2]Gatsby Computational Neuroscience Unit, University College London, London WC1N 3AR, UK
*Correspondence: r.dolan@ucl.ac.uk

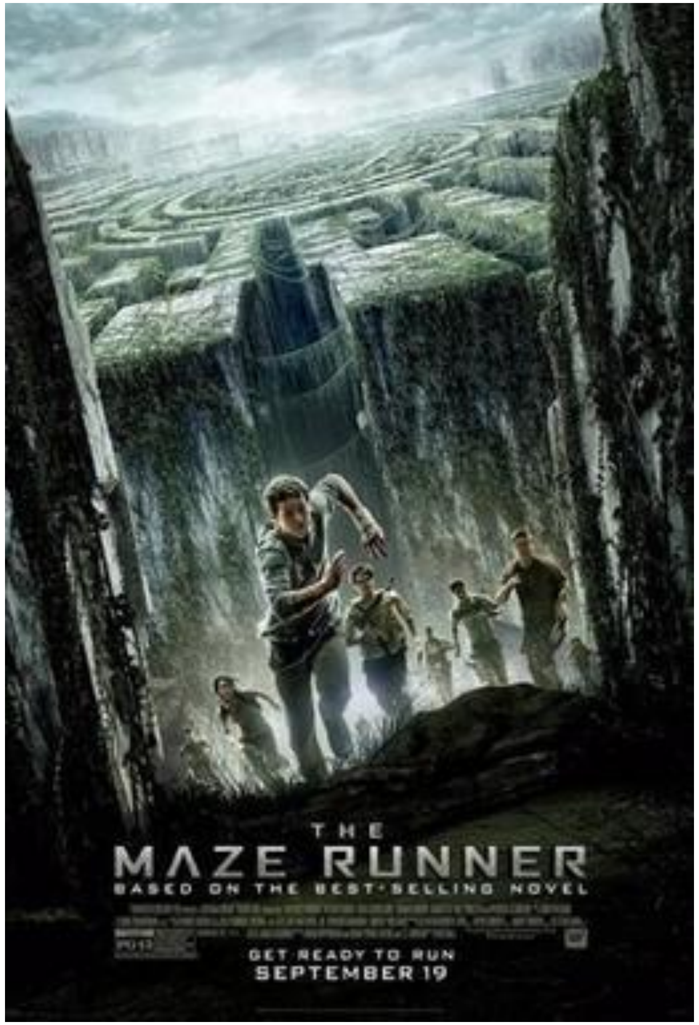
http://dx.doi.org/10.1016/j.neuron.2013.09.007

- **Generation 0**: cognitive maps vs stimulus-response [experimental psychology]
- **Generation 1**: goal-directed vs habitual actions [experimental psychology]
- **Generation 2**: goal-directed vs habitual actions in the human brain [cognitive neuroscience]
- **Generation 3**: model-based vs model-free computational analyses [computational neuroscience]

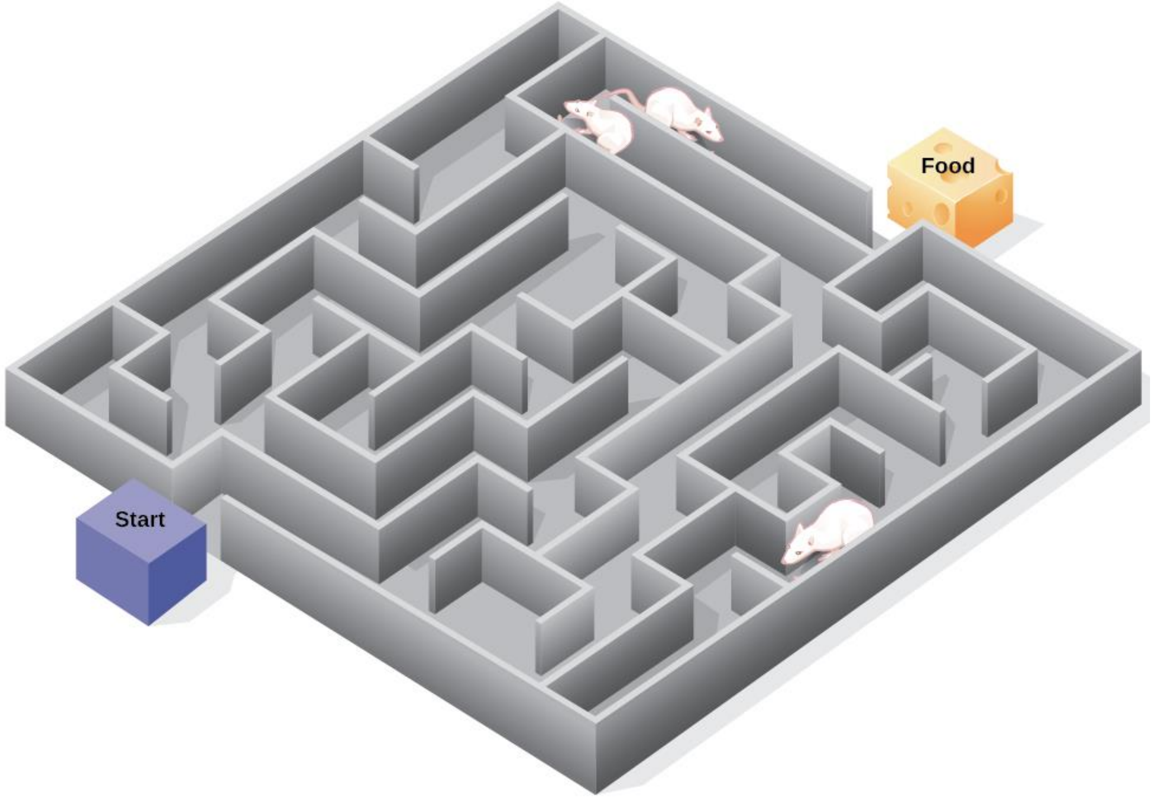
Generation 0: cognitive maps vs stimulus-response



Generation 0: cognitive maps vs stimulus-response



How does the animal learn to solve the maze?

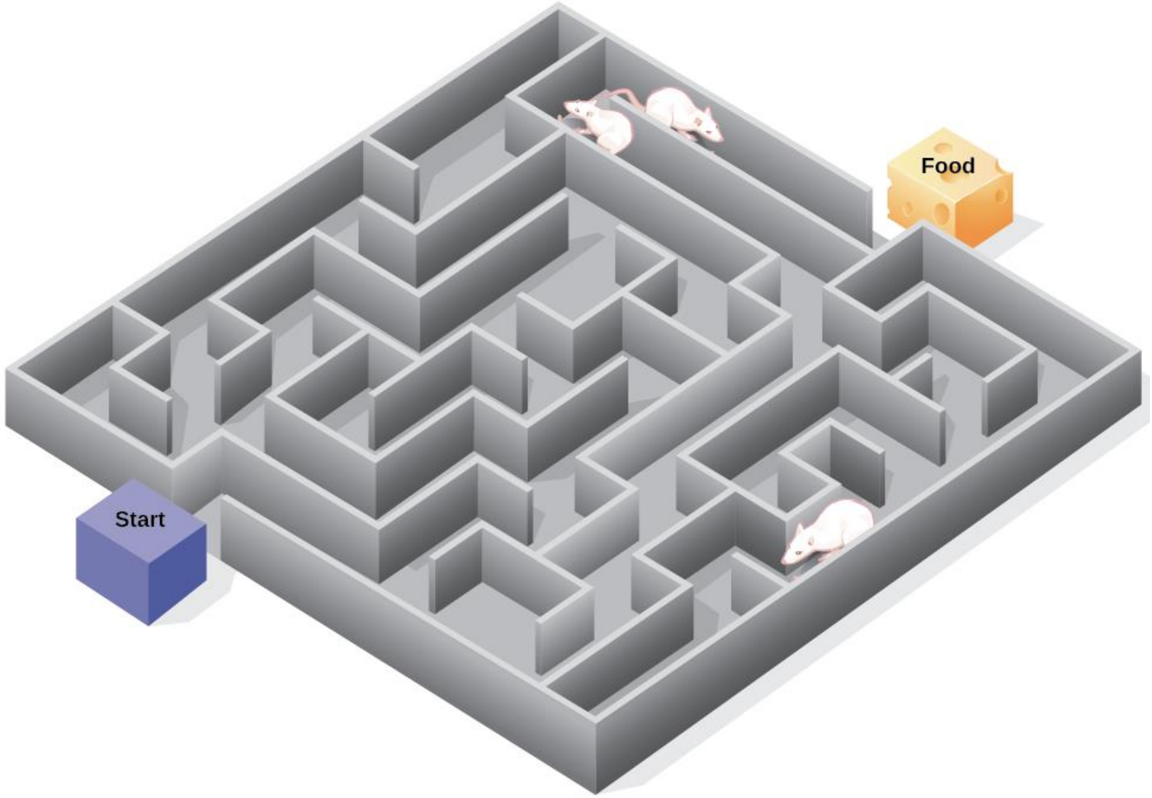


Generation 0: cognitive maps vs stimulus-response

Stimulus-response (S-R) theories

- the bedrock of psychology in the first half of the 20th century
- Solving the maze is a matter of **individual stimulus-response one-to-one connections**
- Learning depends on strengthening of some connections and weakening of others
- the animal helplessly responds to a succession of external and internal stimuli that callout the actions to take (e.g. turnings) and the like that follows

How does the animal learn to solve the maze?



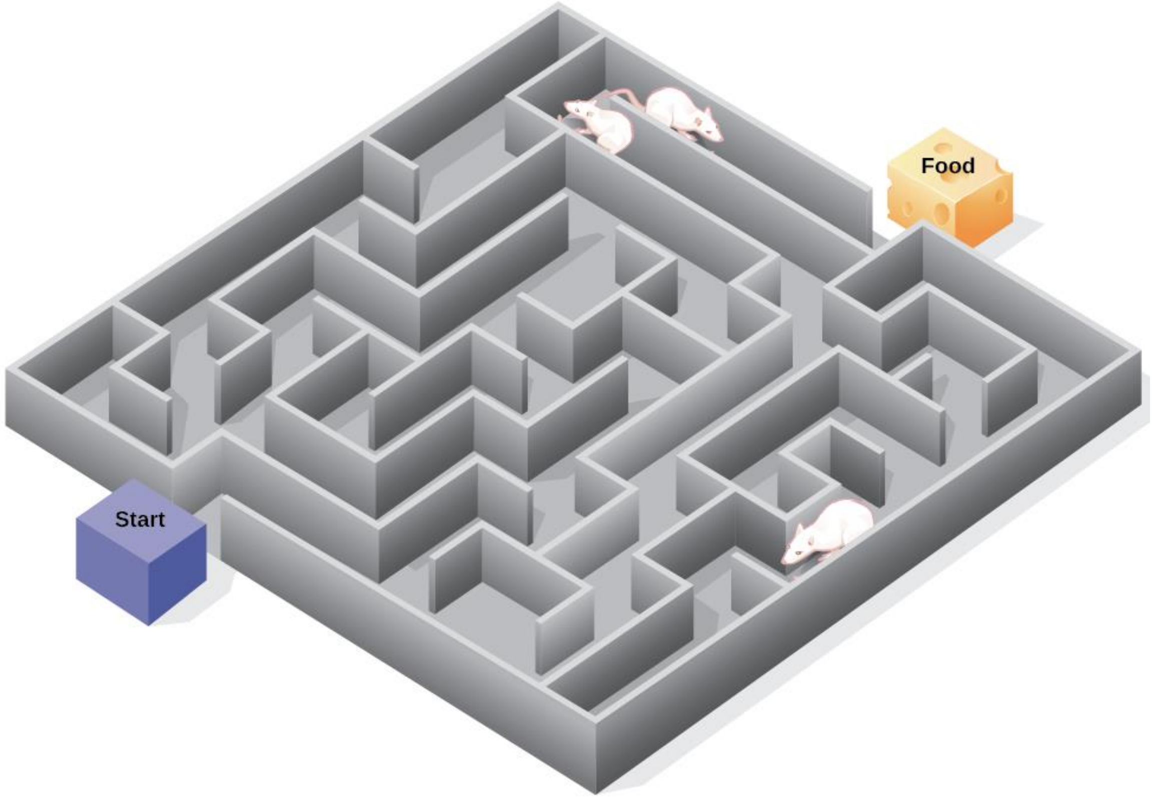
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

Generation 0: cognitive maps vs stimulus-response

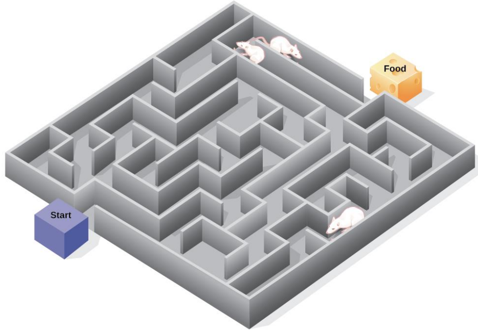
Field theories

- Solving the maze is a matter of **creating a mental/cognitive map that includes multiple sets of connections**
- The mental map then guides what responses the animal will perform
- The mental map acts as a representational template that enables an animal to find the best possible action at a particular state

How does the animal learn to solve the maze?



Generation 0: cognitive maps vs stimulus-response



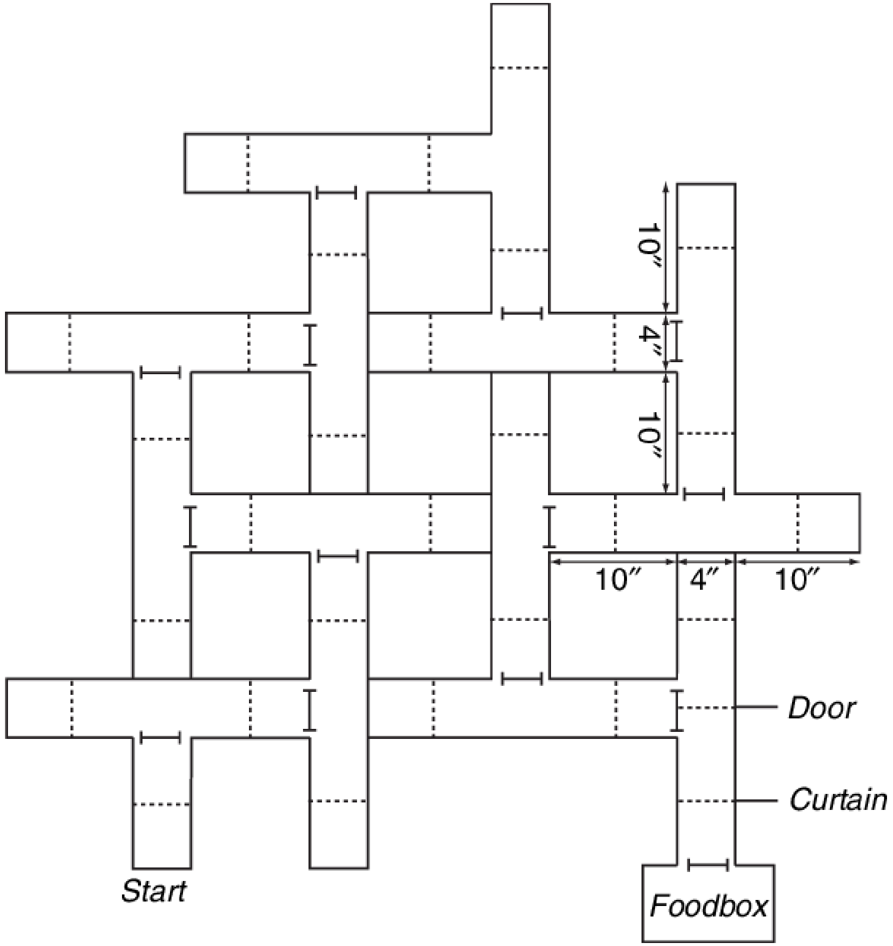
Tolman's maze

The maze had lots of doors and curtains to make it difficult for the rats to master.

Doors swung both directions, which prevented the rat from seeing most of the junctions as it approached. This forced the rat to go through the door to discover what was on the other side.

Curtains hung down and prevented the rat from getting a long distance perspective and it also meant that they could not see a wall at the end of a wrong turn until they had already made a choice and moved in that direction.

The rat was always in a small area, unable to see beyond the next door or curtain, so learning the maze was a formidable task.



Maze used by Tolman and Honzik (1930) to study latent learning in rats. From Tolman EC (1948) Cognitive maps in rats and men. Psychol. Rev. 55: 189-208

https://courses.lumenlearning.com/wmopen-psychology/chapter/psychology-in-real-life-latent-learning/

Generation 0: cognitive maps vs stimulus-response

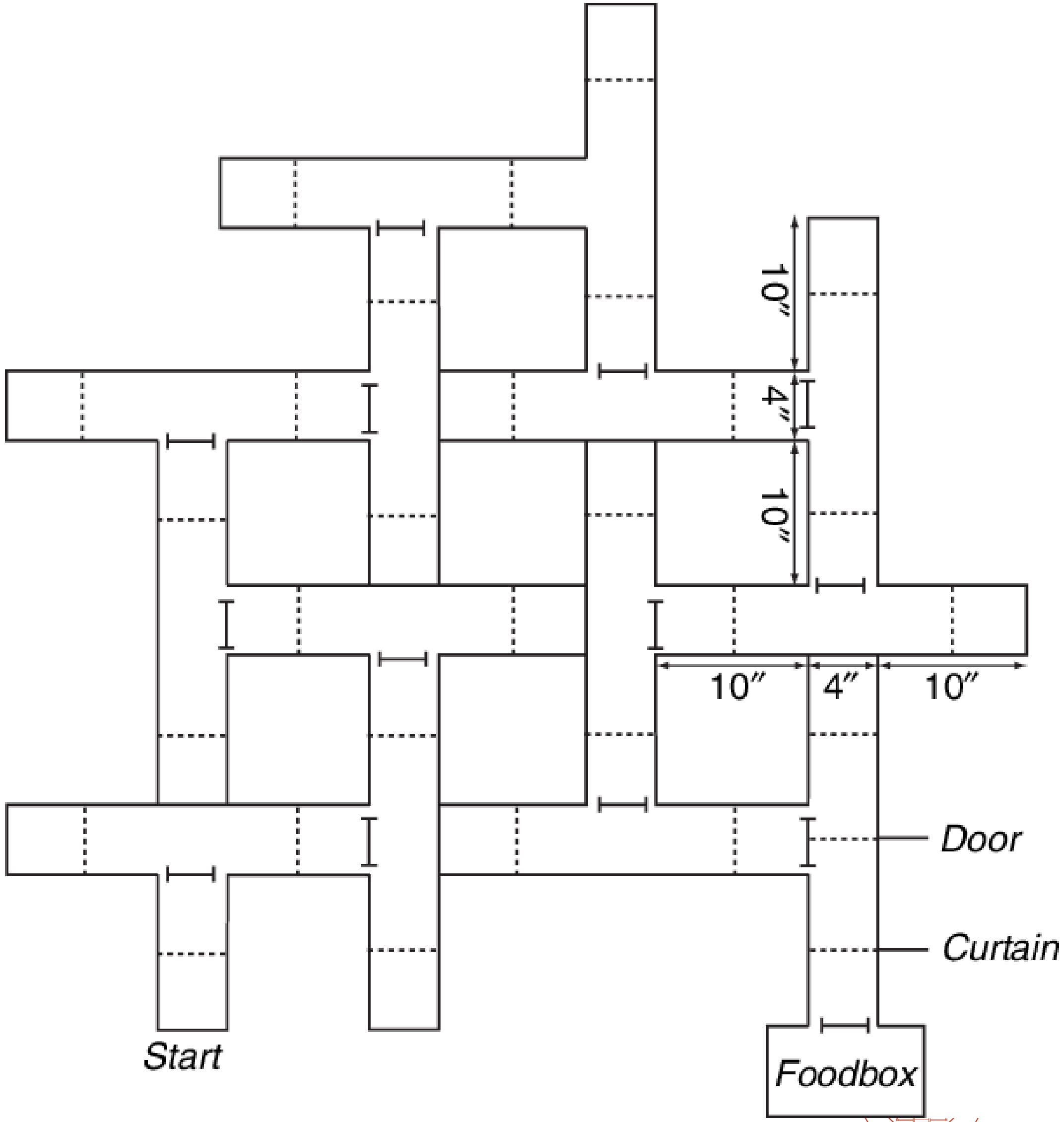
Experiment

Hungry rats have to find their way out through a maze

Group 1: no reward for solving the maze

Group 2: food reward for solving the maze

Which group completed the maze faster?



Maze used by Tolman and Honzik (1930) to study latent learning in rats. From Tolman EC (1948) Cognitive maps in rats and men. Psychol. Rev. 55: 189-208

Generation 0: cognitive maps vs stimulus-response

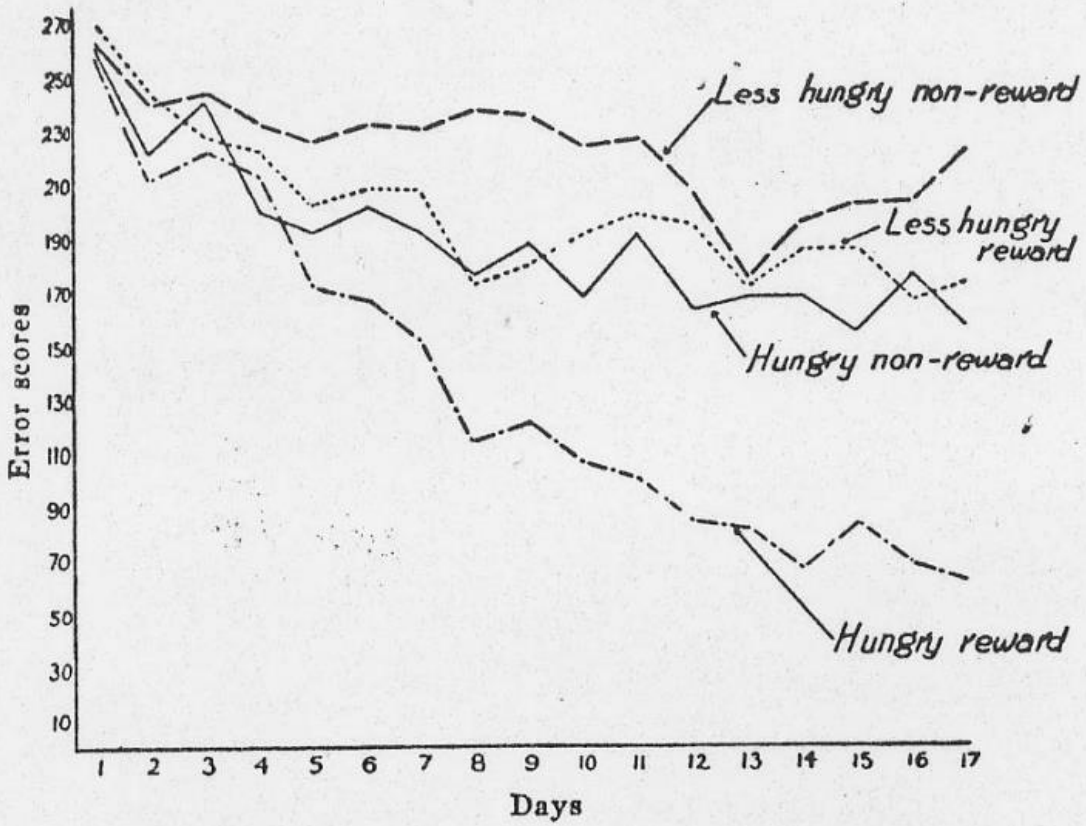
Experiment

Hungry rats have to find their way out through a maze

Group 1: no reward for solving the maze

Group 2: food reward for solving the maze

Which group completed the maze faster?



Error curves for four groups, 36 rats.

Fig. 3

(From E. C. Tolman and C. H. Honzik, Degrees of hunger, reward and non-reward, and maze learning in rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, No. 16, p. 246. A maze identical with the alley maze shown in Fig. 1 was used.)

Generation 0: cognitive maps vs stimulus-response

Experiment

Hungry rats have to find their way out through a maze

Group 1: no reward for solving the maze

Group 2: food reward for solving the maze

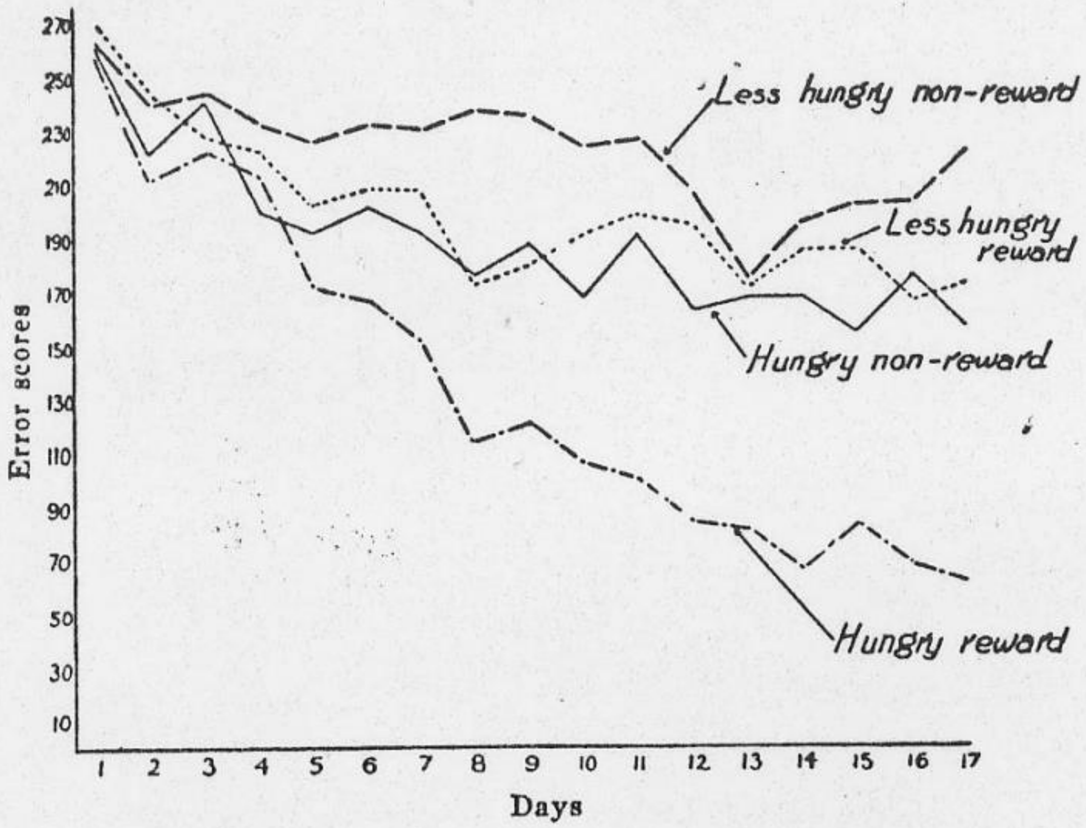
Which group completed the maze faster?

Group 2

Reward & motivation are crucial to learn

But...

Have the other groups learnt anything at all?



Error curves for four groups, 36 rats.

FIG. 3

(From E. C. Tolman and C. H. Honzik, Degrees of hunger, reward and non-reward, and maze learning in rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, No. 16, p. 246. A maze identical with the alley maze shown in Fig. 1 was used.)

Generation 0: cognitive maps vs stimulus-response

Experiment

Hungry rats have to find their way out through a maze

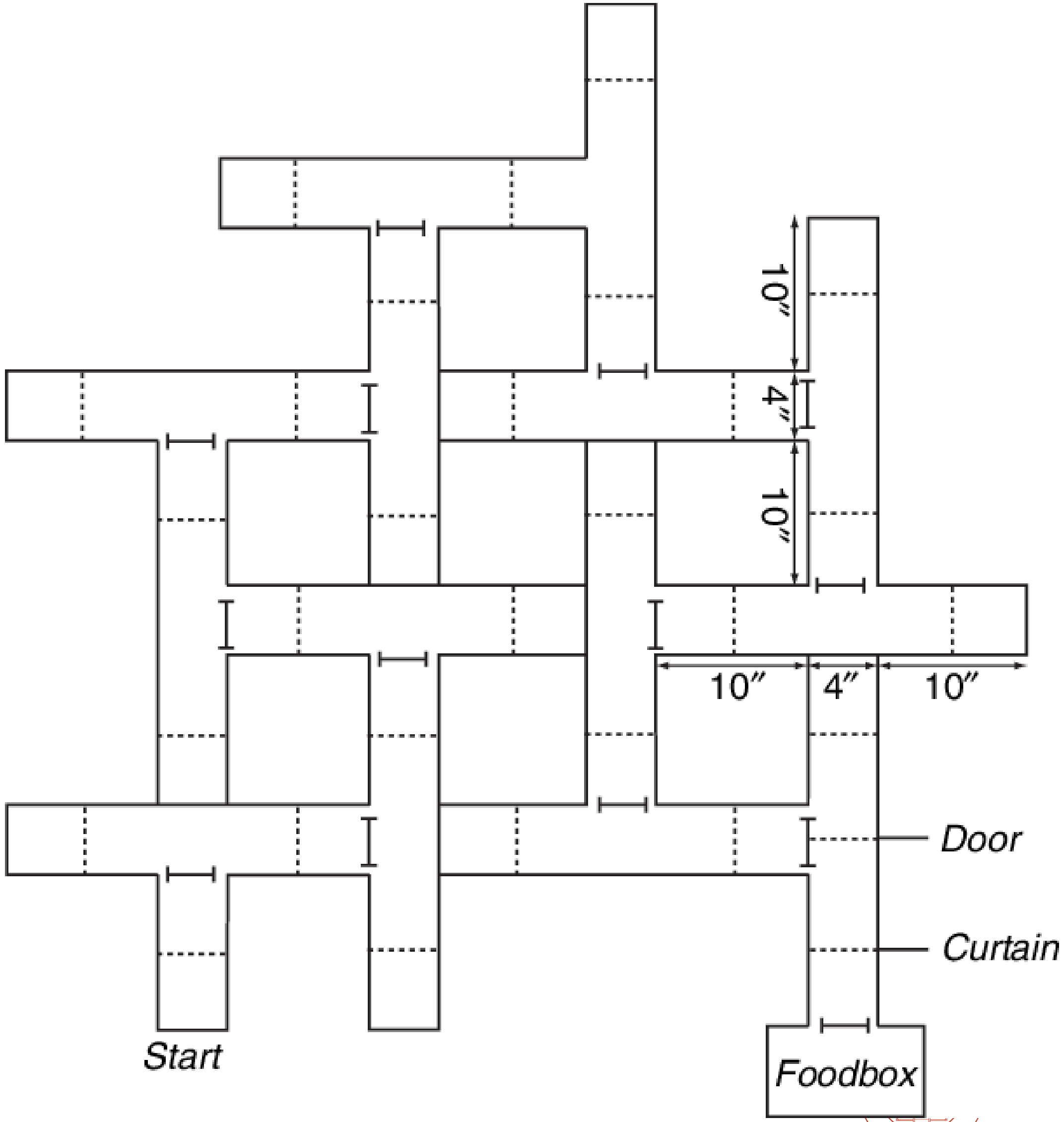
Group 1: no reward for solving the maze

Group 2: food reward for solving the maze

Group3: food reward for solving the maze provided only at day 11

What do you think happens to performance?

Note: the S-R theories argues that no learning occurs when there is no reward



Maze used by Tolman and Honzik (1930) to study latent learning in rats. From Tolman EC (1948) Cognitive maps in rats and men. Psychol. Rev. 55: 189-208

Generation 0: cognitive maps vs stimulus-response

Experiment

Hungry rats have to find their way out through a maze

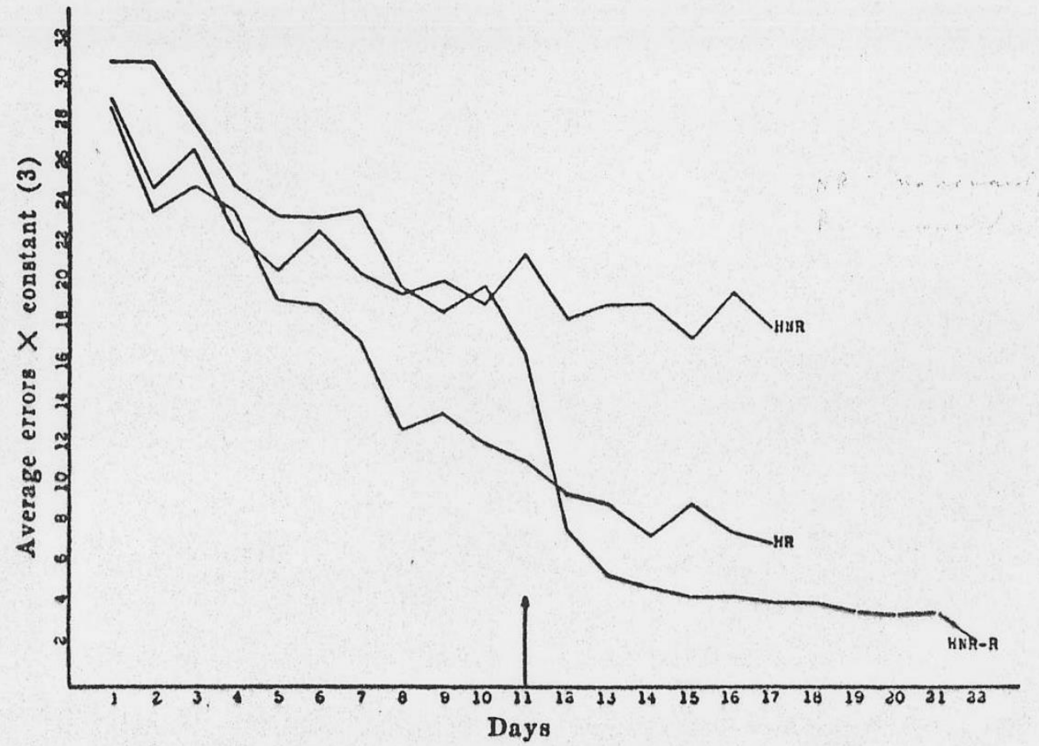
Group 1: no reward for solving the maze

Group 2: food reward for solving the maze

Group3: food reward for solving the maze provided only at day 11

What do you think happens to performance?

As soon as the rats in group 3 was given the food, they were able to find their way through the maze quickly, just as quickly as the comparison group, which had been rewarded with food all along



Error curves for HR, HNR, and HNR-R

FIG. 6

(From E. C. Tolman and C. H. Honzik, Introduction and removal of reward, and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1930, 4, No. 19, p. 267.)

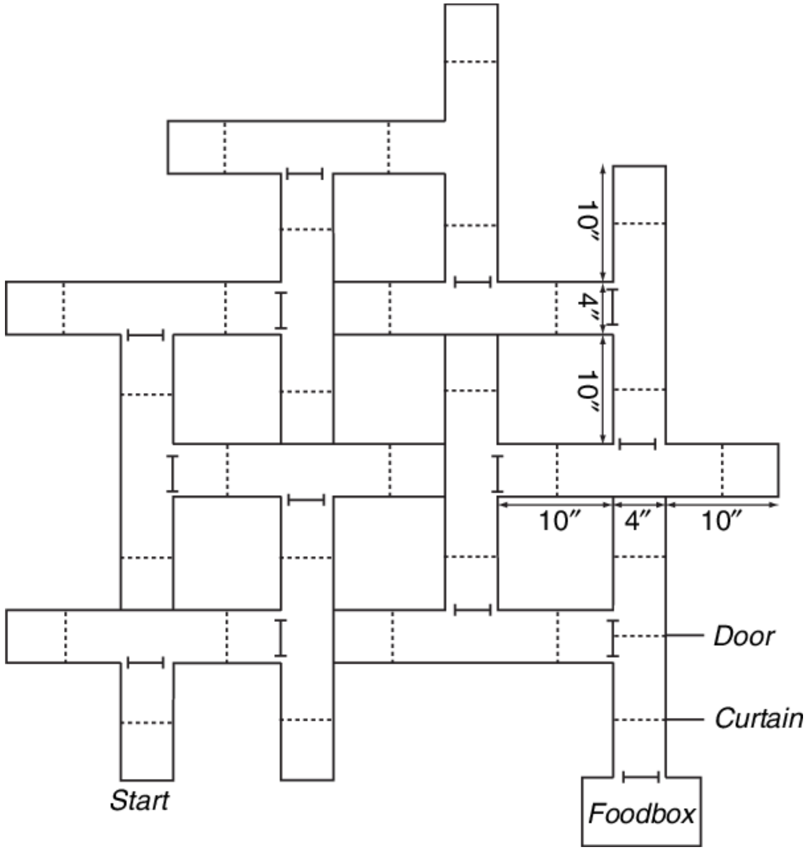
Latent learning & cognitive maps

Latent learning

- Learning that is not shown behaviorally until there is sufficient motivation
- It occurs without any obvious reinforcement of the behavior or associations that are learned

Cognitive map

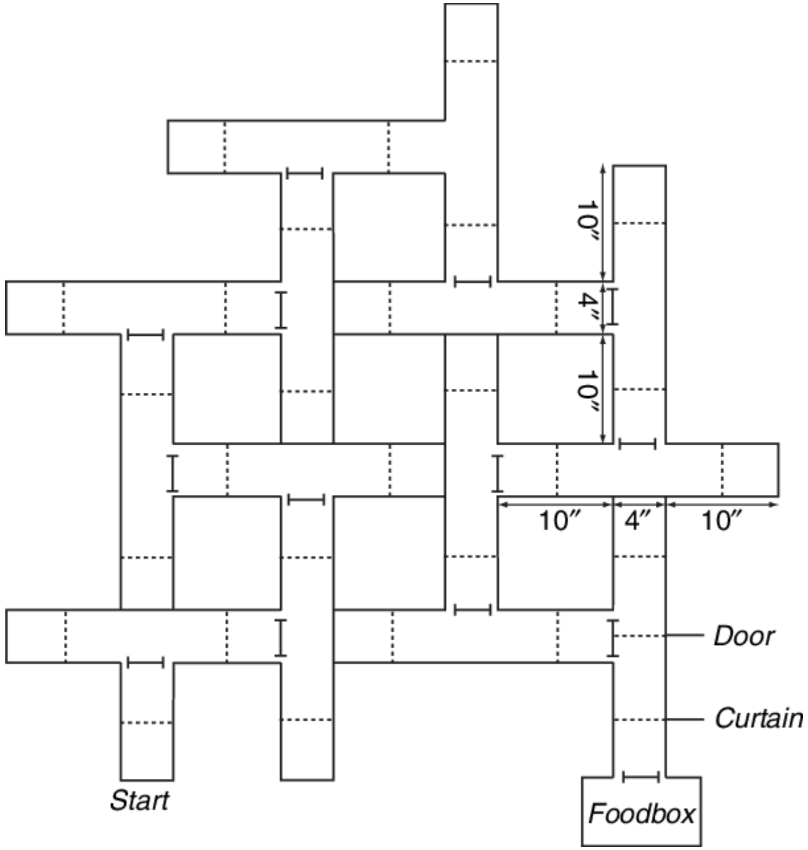
- Rats behaved as if they were responding to a mental representation of the overall layout of the maze rather than blindly exploring different parts of the maze through trial and error
- Mental representation of the space field that can guide what actions should be performed at any stage to achieve a particular goal



Maze used by Tolman and Honzik (1930) to study latent learning in rats. From Tolman EC (1948) Cognitive maps in rats and men. Psychol. Rev. 55: 189-208

Generation 0: Implications for the field

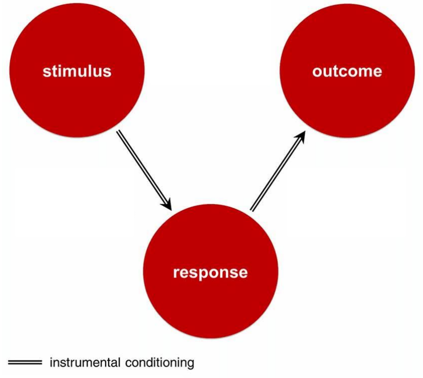
- Challenged the constraints of behaviorism, which stated that processes must be directly observable and that learning was the direct consequence of conditioning to stimuli
- Challenged the prevailing stimulus-response (S–R) view of learning and behavior, which corresponds to the simplest model-free way of learning policies
- Conditioning in volves more than the simple formation of associations between sets of stimuli or between responses and reinforcers. It includes learning and representing other facets of the total behavioral context



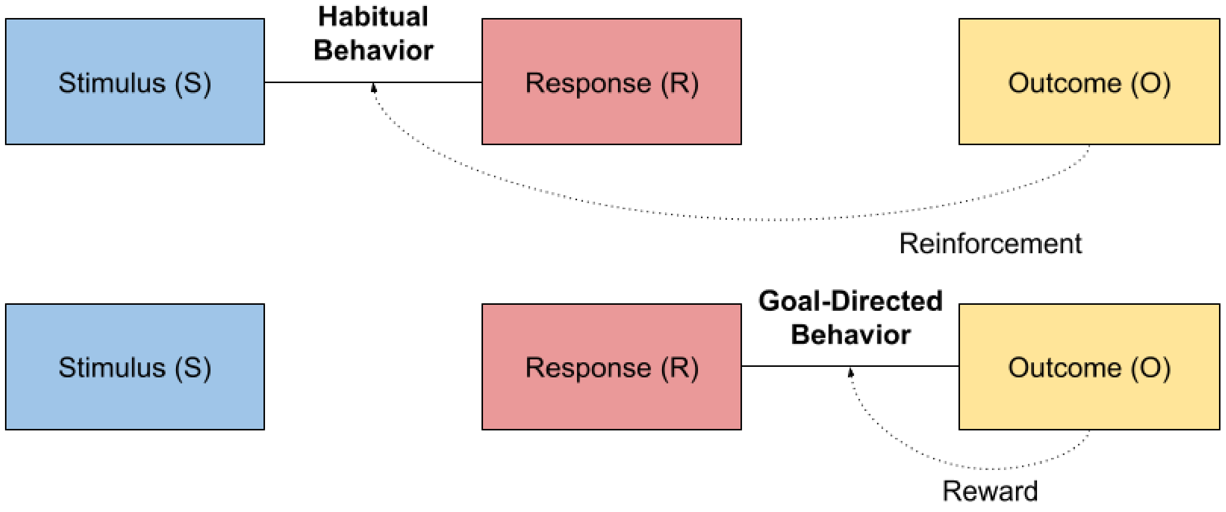
Maze used by Tolman and Honzik (1930) to study latent learning in rats. From Tolman EC (1948) Cognitive maps in rats and men. Psychol. Rev. 55: 189-208

Generation 0 studies established a dichotomy between decision behavior controlled by a cognitive map and by S-R associations

Generation 1: Goal-Directed vs Habitual actions

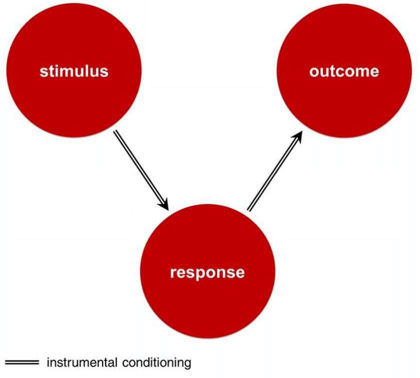


- Operationalized the use of cognitive maps for learning to choose appropriate actions (I.e. that maximize rewards/minimize punishments) in nonspatial domains
- Termed this as **goal-directed behavior/actions**
- Contrasted it with **habitual behavior/actions**
- Focused on animal studies to identify the neural bases of the two types of behaviors



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

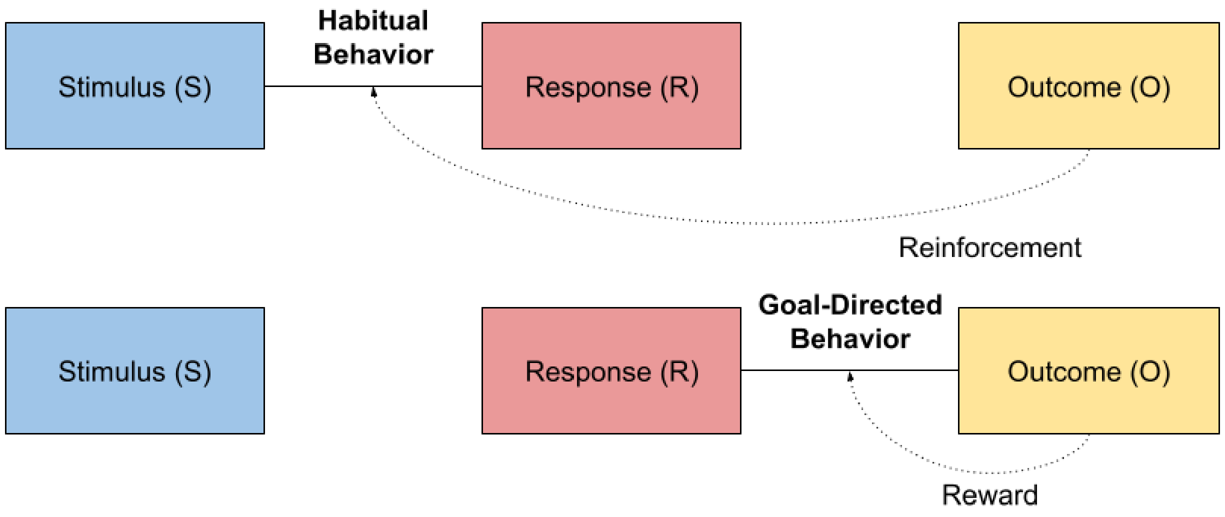
Goal-directed behavior/actions



-The action is made because we think that they will lead to outcomes that we desire

-Two criteria make an action goal-directed

1. There must be **knowledge of the relationship between an action** (or sequence of actions) and its **consequences** --> response-outcome or R-O control
2. The **outcome should be motivationally relevant** or desirable at the moment of choice/action

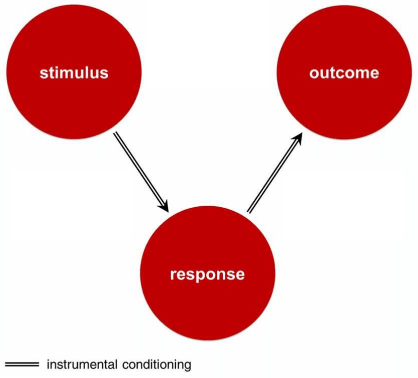


Goal-directed behavior:

- Involves active deliberation
- Has high computational cost
- Shows adaptive flexibility to changing of environmental contingencies (e.g. the behavior stops if no reward follows the action)

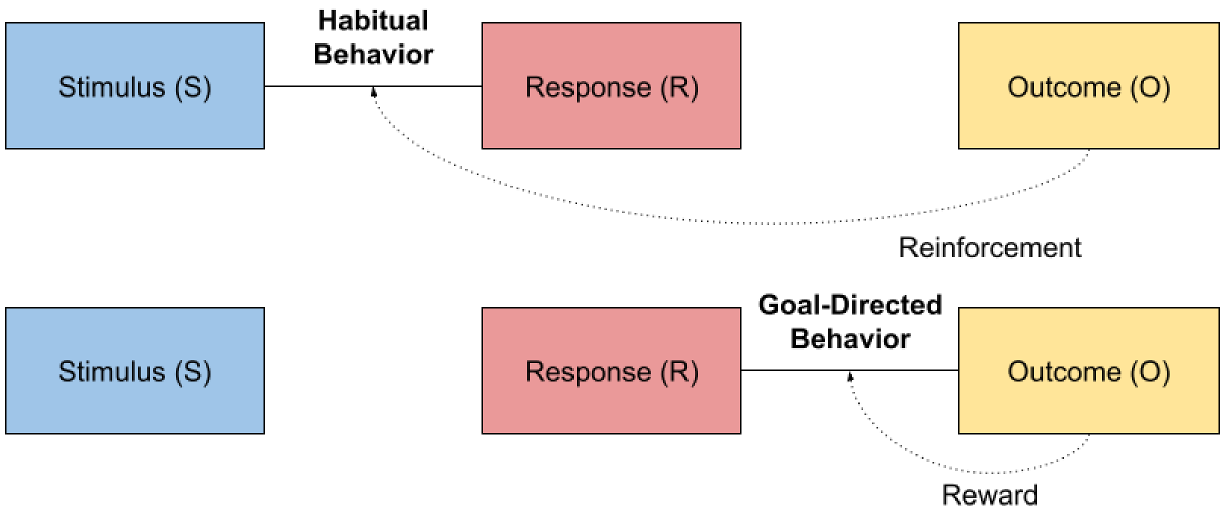
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

Habitual behavior/actions



The action is

- made automatically, just because it has been rewarded in the past
- not influenced by the current value of the outcome it leads to
- continues to be enacted even when the outcome is undesired



Habitual behavior:

- Automatic (no active deliberation)
- Has low computational cost
- Is inflexible to changing of environmental contingencies (e.g. the behavior does not stop even if no reward follows the action)

Habitual behavior/actions

The action is

- made automatically, just because it has been rewarded in the past
- not influenced by the current value of the outcome it leads to
- continues to be enacted even when the outcome is undesired

Habitual behavior:

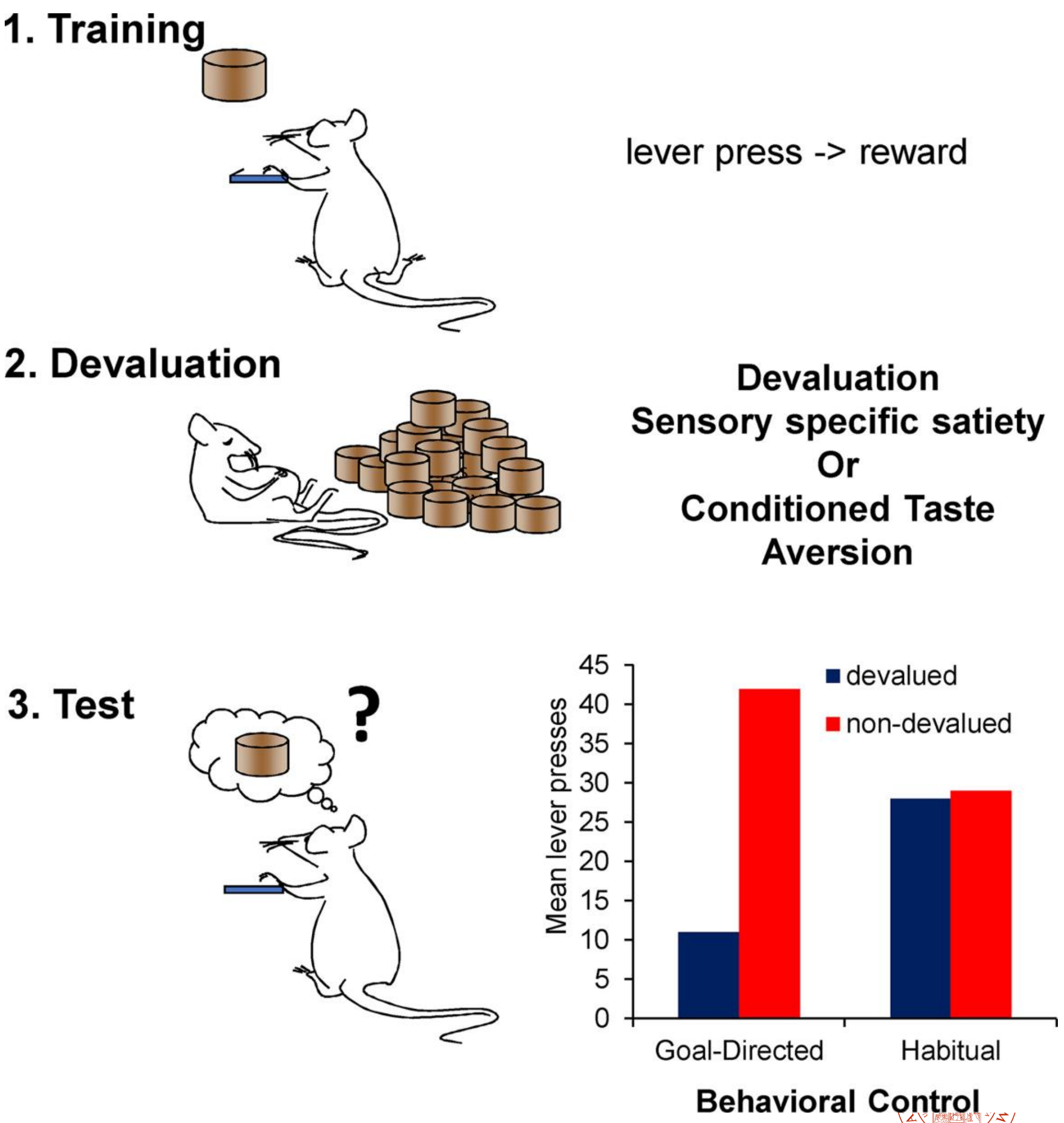
- Automatic (no active deliberation)
- Has low computational cost
- Is inflexible to changing of environmental contingencies (e.g. the behavior does not stop even if no reward follows the action)

IT'S NOT THAT SIMPLE

If you repeat the same behavior regularly without getting the results you desire, you're crazy. Choose new habits. Get sane, and get results.
www.ninaamir.com
NINA AMIR
INSPIRATION TO Creation COACH

Testing if a behavior is goal-directed (vs habitual)

1. training session: the animal undergoes instrumental learning (learns that some actions will lead to rewards)
2. Post-training manipulation
   1. reinforcer **devaluation**
   2. contingency degradation
3. Testing session: the animal repeats the actions learned during instrumental training under extinction
   - If the action associated to the devalued reinforcer is performed less, then the behavior is goal-directed
   - if not, it's habitual



1. Training

lever press -> reward

2. Devaluation

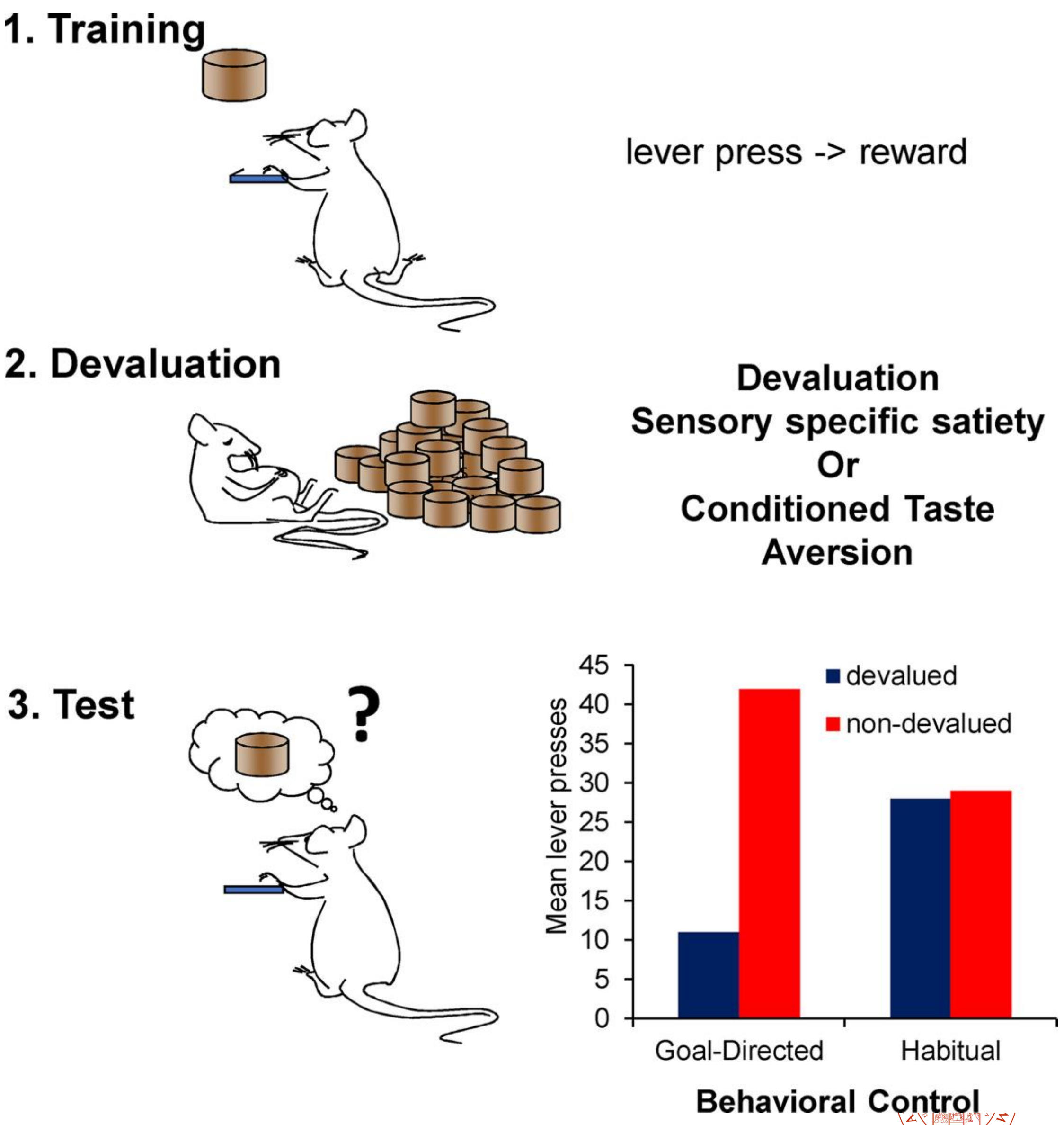
Devaluation
Sensory specific satiety
Or
Conditioned Taste Aversion

3. Test

Corbit, Laura H. "Understanding the balance between goal-directed and habitual behavioral control." *Current opinion in behavioral sciences* 20 (2018): 161-168.

Testing if a behavior is habitual (vs goal-directed)

1. Extensive training session: **overtraining**
   - the animal undergoes instrumental learning (learns that some actions will lead to rewards)
   - This time the training is extensive
2. Post-training manipulation
   1. reinforcer **devaluation**
   2. contingency degradation
3. Testing session: the animal repeats the actions learned during instrumental training under extinction
   - If the action associated to the devalued reinforcer is performed less, then the behavior is goal-directed
   - if not, it's habitual



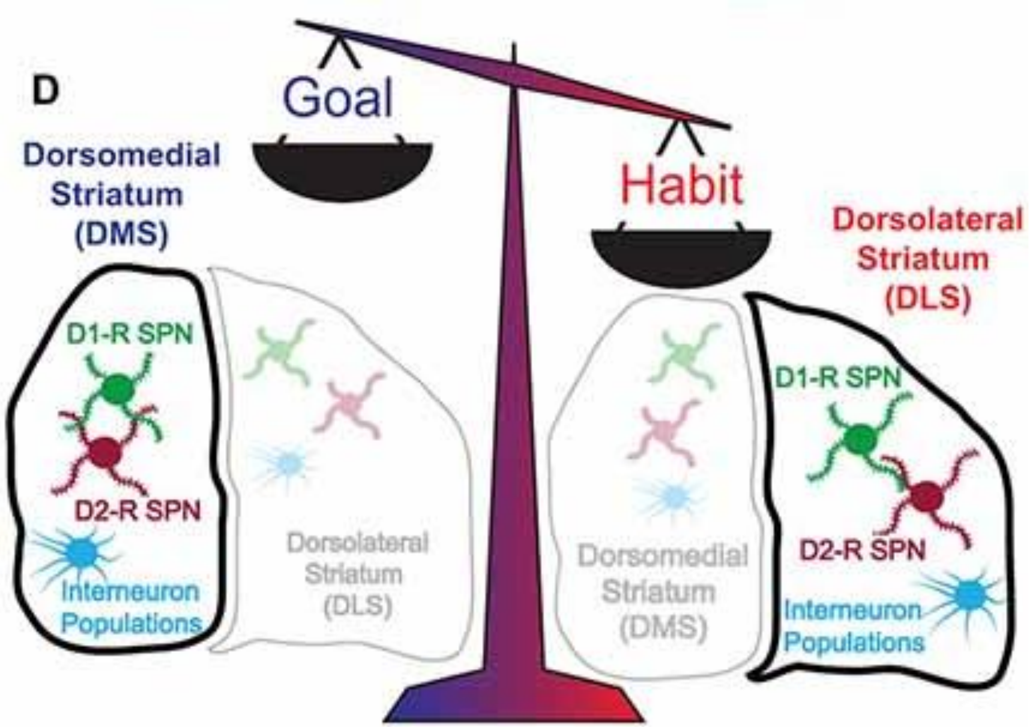
Dissociation of goal-directed vs habitual behavior in the striatum

Dorsomedial striatum

supports goal-directed behavior

Dorsolateral striatum

supports habitual behavior



D
Goal
Habit
Dorsomedial Striatum (DMS)
Dorsolateral Striatum (DLS)
D1-R SPN
D2-R SPN
Interneuron Populations
Dorsolateral Striatum (DLS)
Dorsomedial Striatum (DMS)
D1-R SPN
D2-R SPN
Interneuron Populations

Lipton, David M., Ben J. Gonzales, and Ami Citri. "Dorsal striatal circuits for habits, compulsions and addictions." *Frontiers in systems neuroscience* (2019): 28.

The dopaminergic pathways

1. Nigrostriatal pathway

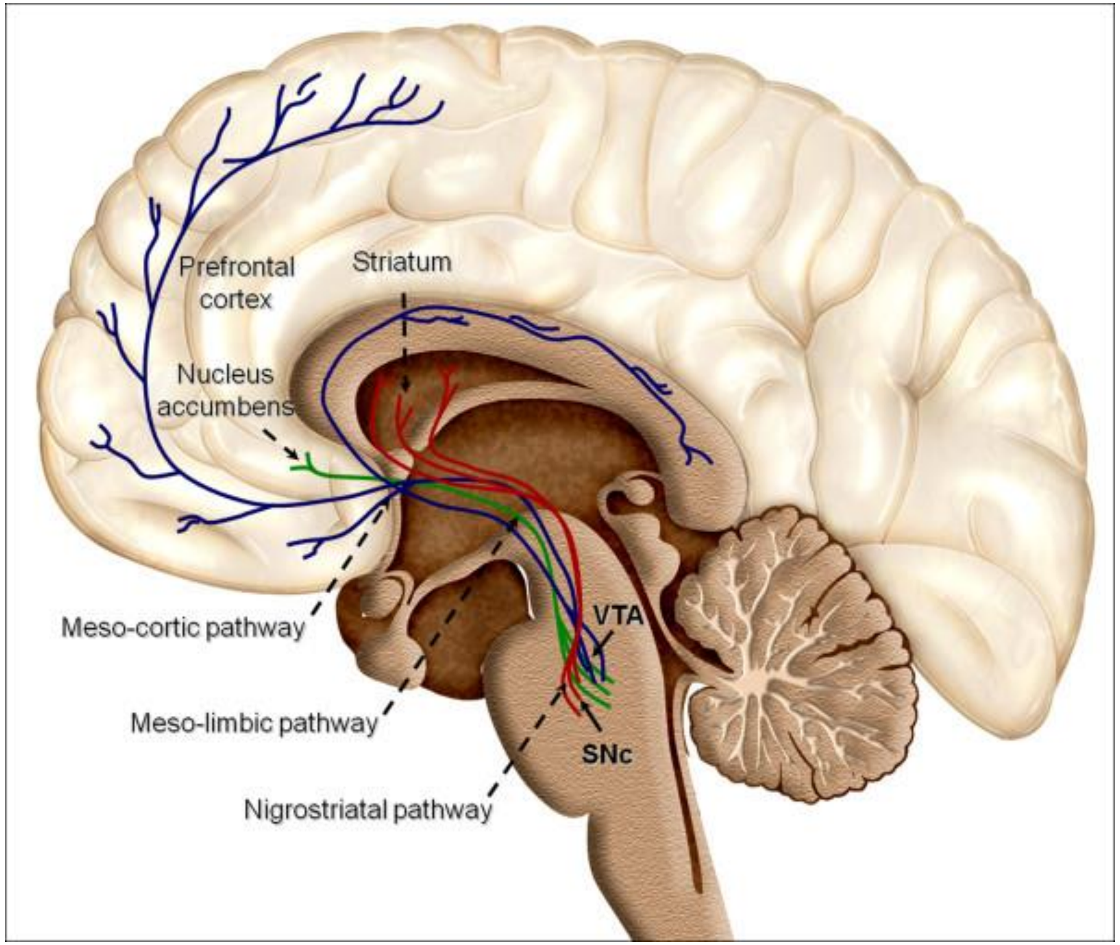
- originates in the substantia nigra pars compacta (SNc)
- projects primarily to the **caudate–putamen** (dorsal striatum in rodents)
- It is critical in the production of **movement** as part of the basal ganglia motor loop

2. Mesolimbic pathway

- originates in the VTA
- projects to the **nucleus accumbens**, septum, amygdala and hippocampus

3. Mesocortical pathway

- Originates in the VTA
- projects to the medial prefrontal, cingulate, **orbitofrontal** and perirhinal cortex

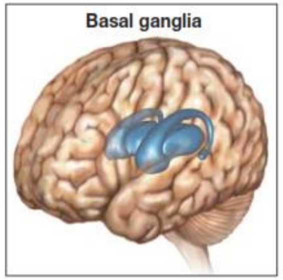


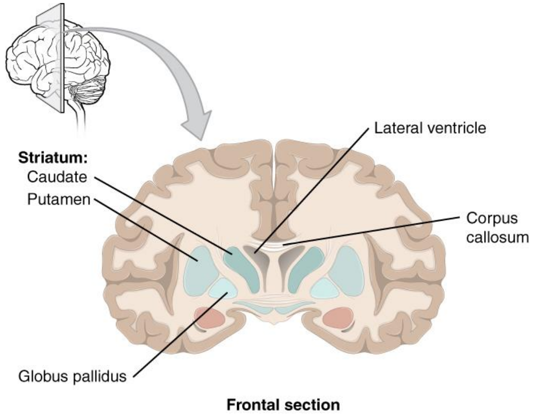
https://www.brainfacts.org/3d-brain#intro=false&focus=Brain

Striatum: linking motivation-action

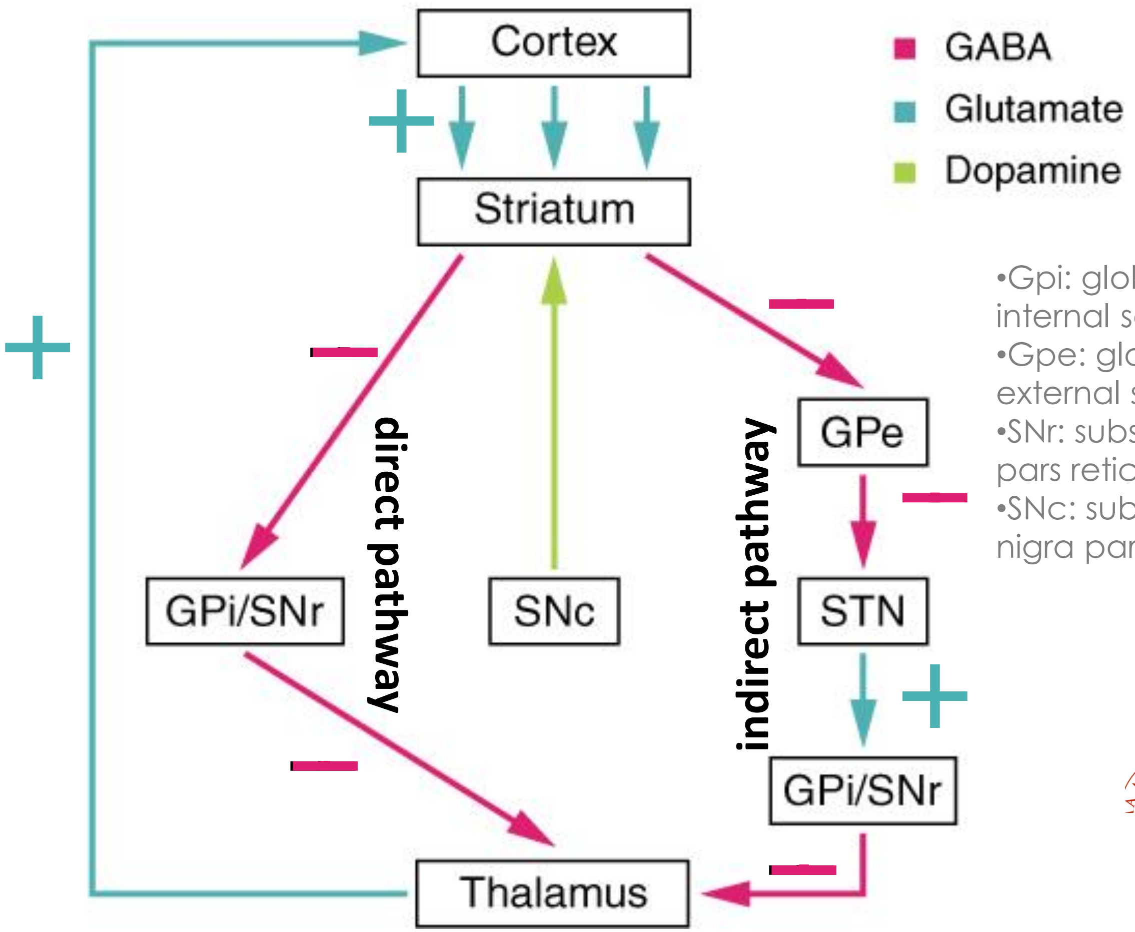
The striatum may be the interface where reward influences action

- The basal ganglia are involved in the selection of actions
- Rewards may influence which actions are selected
    - by affecting plasticity in the striatum, so as to reinforce rewarded actions and make them more likely to recur





Basal nuclei



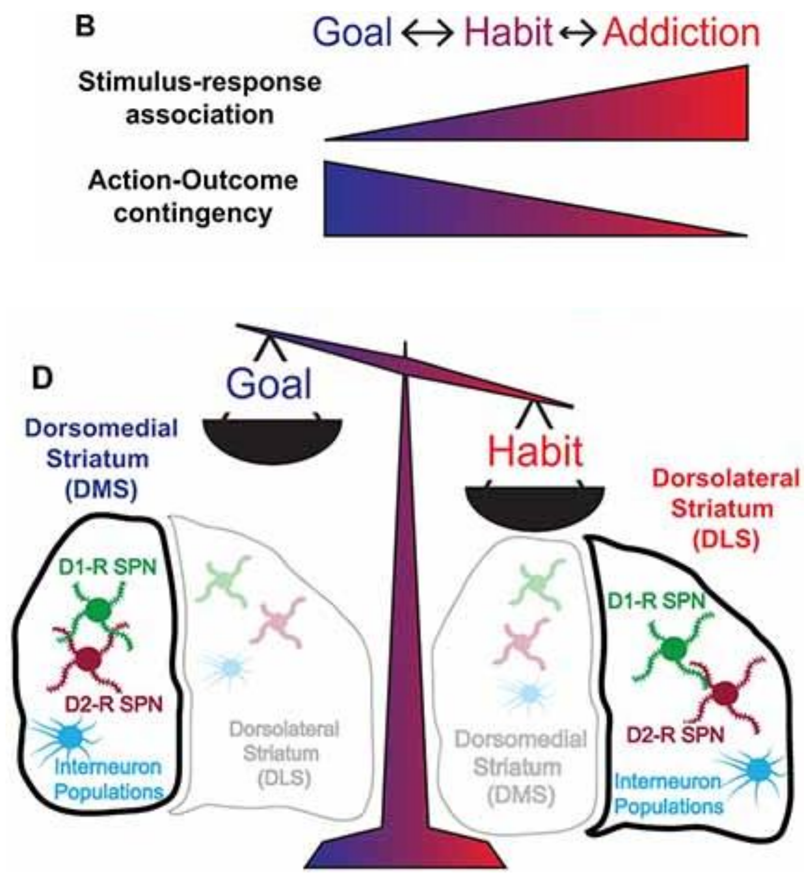
•Gpi: globus pallidus internal segment
•Gpe: globus pallidus external segment
•SNr: substantia nigra pars reticulata
•SNc: substantia nigra pars compacta

From goal-directed to habitual behavior: a continuum

- **Generation 1 results** show that the need for **overtraining to make a behavior habitual** implies that behavior is initially goal directed but then becomes habitual over the course of experience

- In the brain, there is a dynamic **inter-dependency** between goal-directed and habitual systems, which may **act simultaneously and competitively**

- If habit and goal-directed processes act concurrently, we may wonder what are the factors that influence the integration and competition between the two systems


B
Goal ↔ Habit ↔ Addiction
Stimulus-response association
Action-Outcome contingency
D
Goal
Habit
Dorsomedial Striatum (DMS)
D1-R SPN
D2-R SPN
Interneuron Populations
Dorsolateral Striatum (DLS)
Dorsomedial Striatum (DMS)
Dorsolateral Striatum (DLS)
D1-R SPN
D2-R SPN
Interneuron Populations

Lipton, David M., Ben J. Gonzales, and Ami Citri. "Dorsal striatal circuits for habits, compulsions and addictions." *Frontiers in systems neuroscience* (2019): 28.

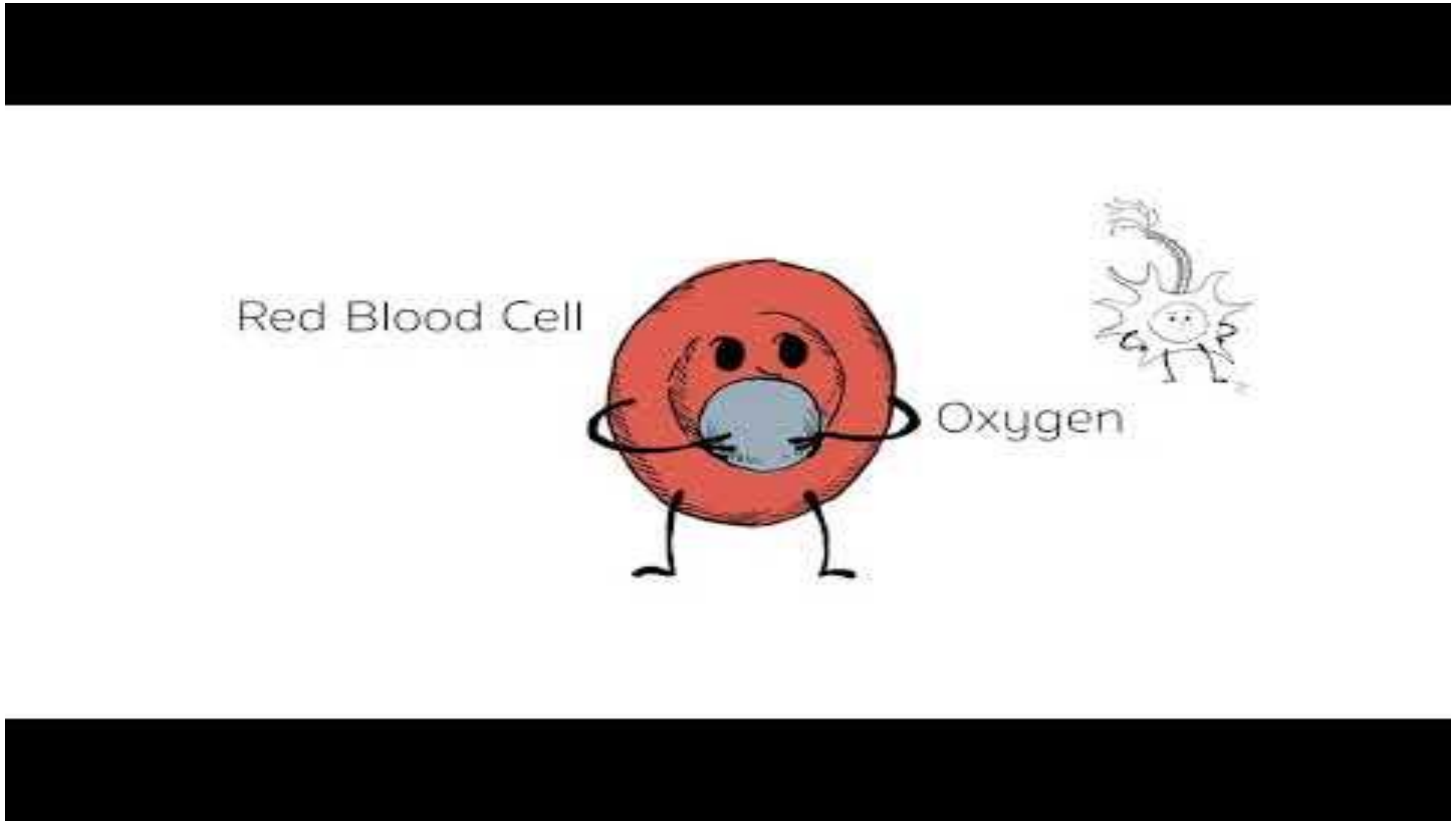
Generation 2: Actions and Habits in the Human Brain

- Successful animal paradigms were adapted for human experiments

- Use of fMRI in order to investigate the neural bases of
  - Goal-directed actions
  - Habitual actions

Functional Magnetic Resonance Imaging (fMRI)

- Measures the ratio of oxygenated to deoxygenated hemoglobin
  - this value is referred to as the blood oxygen level–dependent (**BOLD**) signal
  - Indirect measure of neuronal activity
- Correlational evidence
- High spatial resolution but low temporal resolution
  - Appropriate to know where things happen but not when things happen

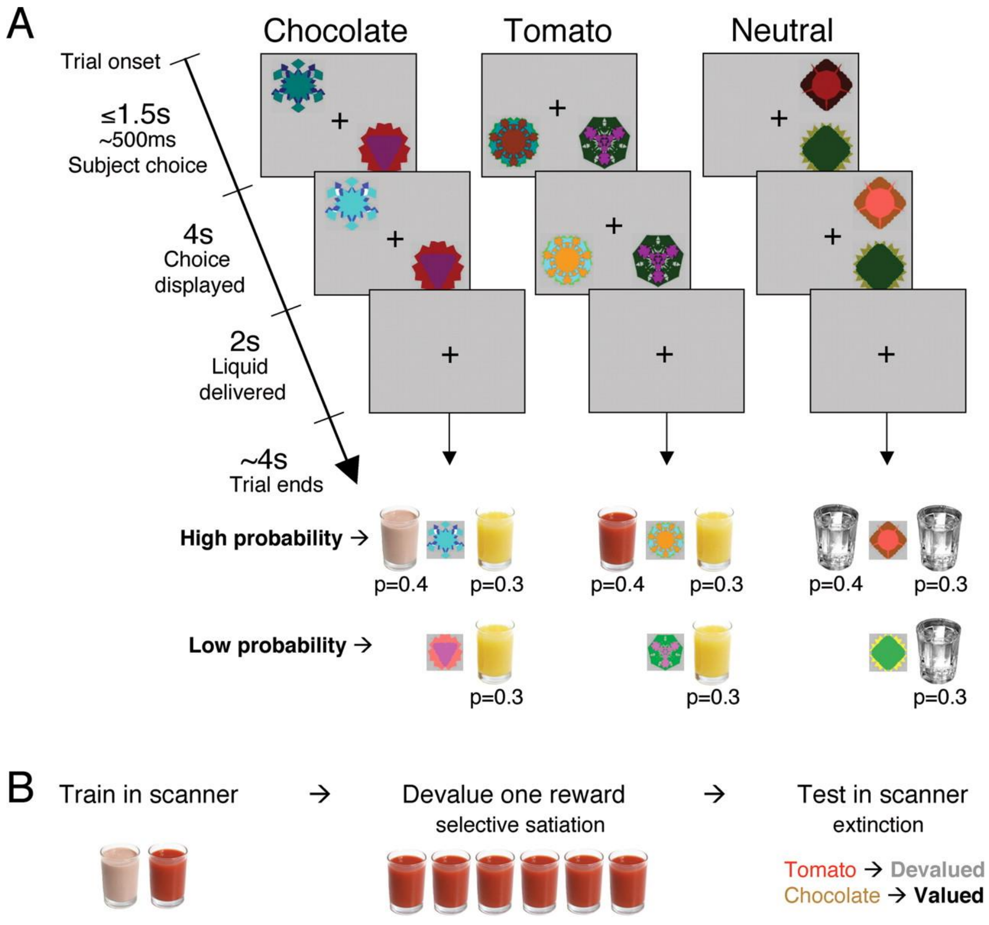


https://youtu.be/4UOeBM5BwdY

Neural substrates of the goal-directed behavior in humans

Method:

- human subjects were trained on a task in which two different actions resulted in two distinct food reward outcomes (I.e. instrumental conditioning)
- One of the outcomes was then devalued(by feeding subjects that food to satiety, i.e., until they would consume no more of it)
- the values of other foods not eaten remained high
- After devaluation participants performed the instrumental actions (choice of stimuli) under extinction



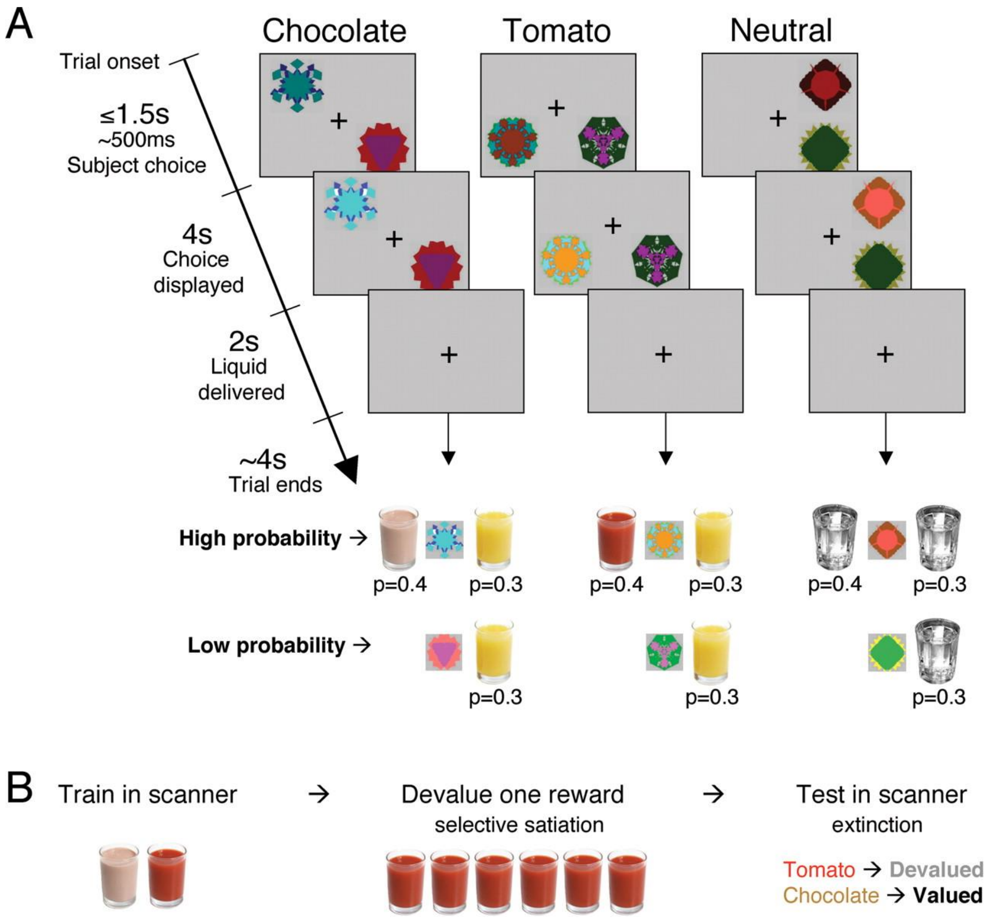
Valentin, V.V., Dickinson, A., and O'Doherty, J.P. (2007). Determining the neu-ral substrates of goal-directed learning in the human brain. J. Neurosci.27,4019–4026.

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

Neural substrates of the goal-directed behavior in humans

Method:

- fMRI was recorded at train and test to examine brain areas responding during action selection
  - looking for areas that showed sensitivity to the change in value of the associated outcomes
  - such area(s) would be candidate regions for implementing goal-directed behavior in humans

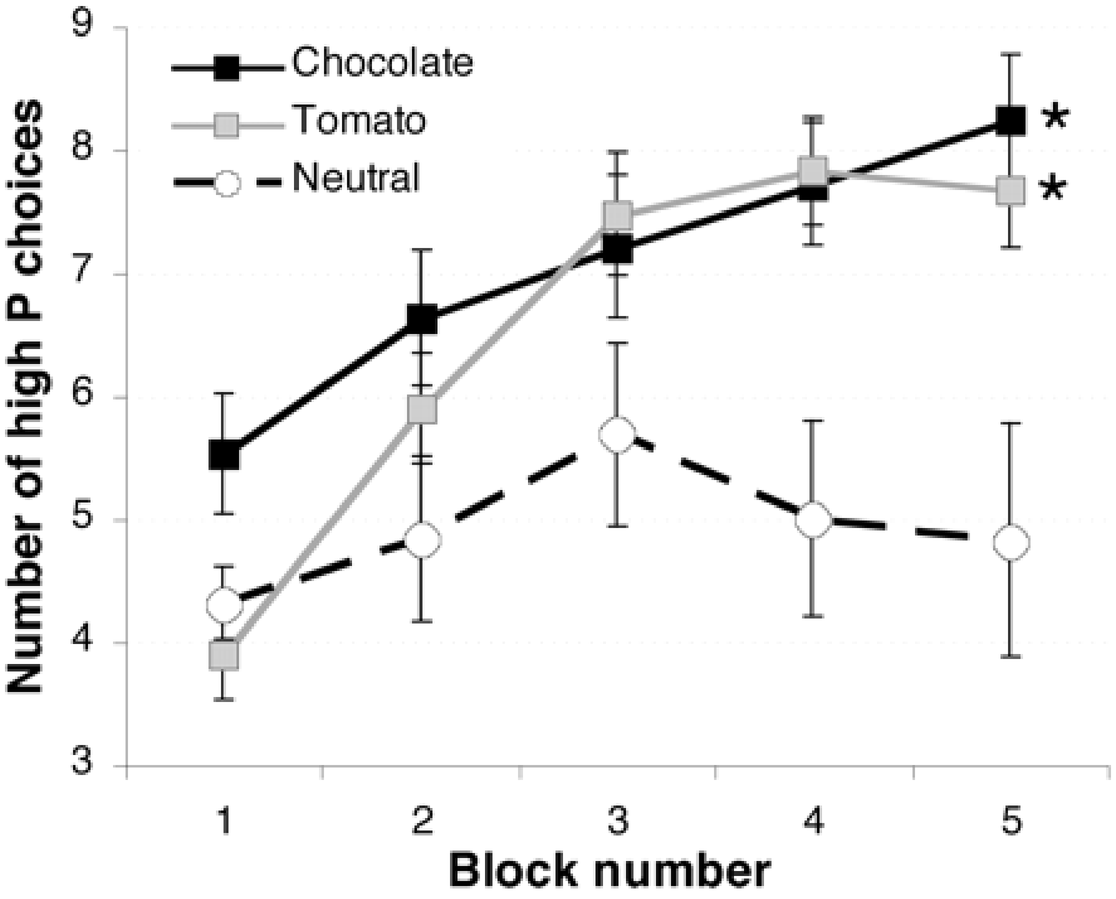


Valentin, V.V., Dickinson, A., and O'Doherty, J.P. (2007). Determining the neu-ral substrates of goal-directed learning in the human brain. J. Neurosci.27,4019–4026.

Neural substrates of the goal-directed behavior in humans

Results: behavioral

Learning curves. Total number of high-probability action choices over five 10-trial blocks shown averaged across 19 subjects during training. Over the course of training, **subjects increasingly favored the high-probability actions associated with tomato juice or chocolate milk over their low-probability counterparts**, but this was not the case for the neutral condition where subjects were indifferent between the high- and low-probability actions (*p < 0.0005, one-tailed). Error bars indicate SEM.



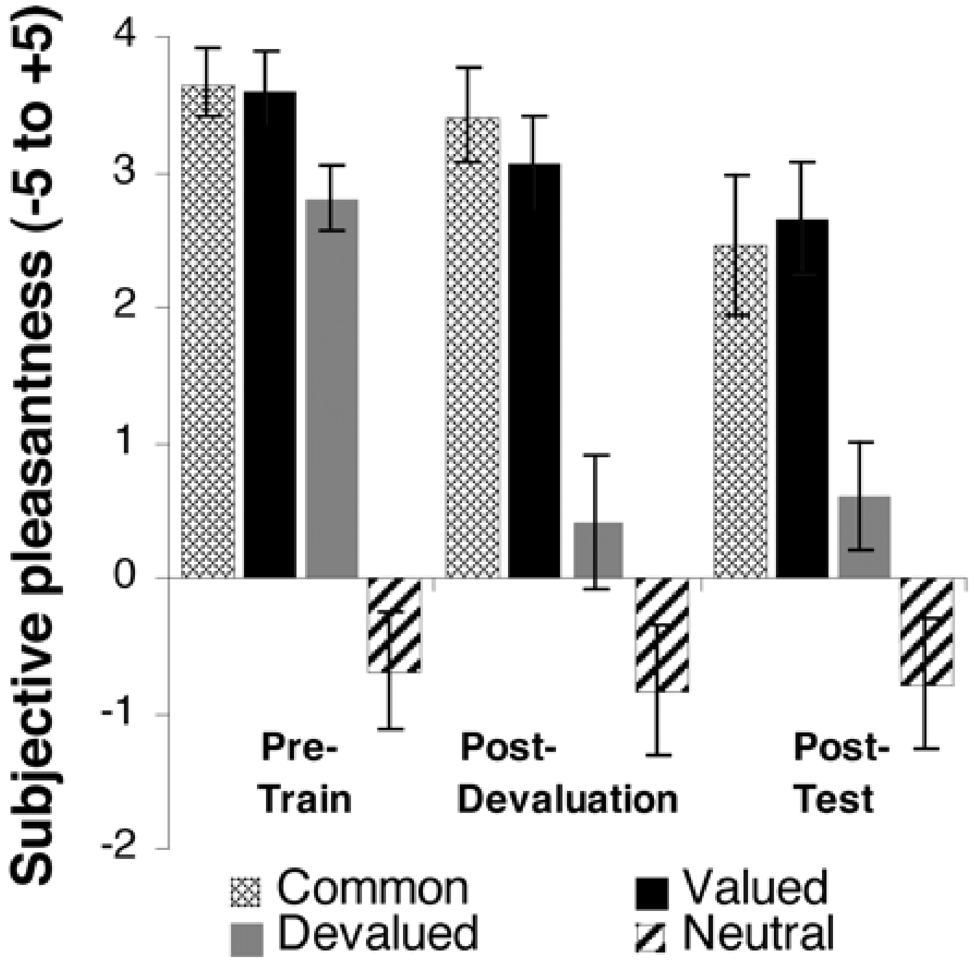
Valentin, V.V., Dickinson, A., and O’Doherty, J.P. (2007). Determining the neu-ral substrates of goal-directed learning in the human brain. J. Neurosci.27,4019–4026.

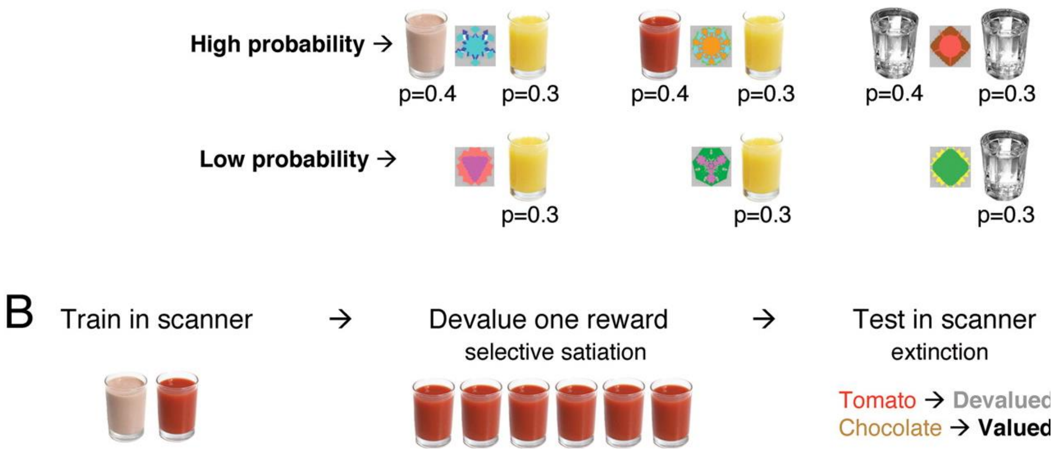
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

Neural substrates of the goal-directed behavior in humans

Results: behavioral

Subjective pleasantness ratings on a scale of −5 (very unpleasant) to +5 (very pleasant) before training, after devaluation, and after test. The rating for the food eaten (devalued) significantly decreased compared with the food not eaten (valued) after the selective devaluation procedure (interaction at $p < 0.01$). Error bars indicate SEM.



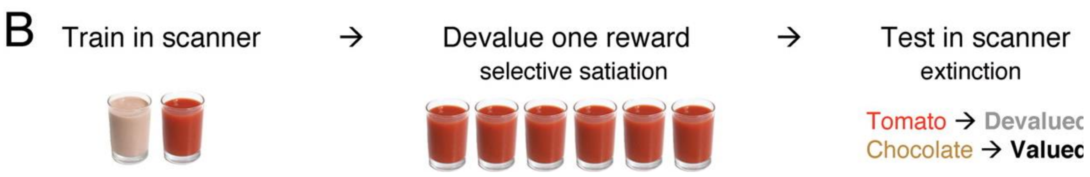


Valentin, V.V., Dickinson, A., and O'Doherty, J.P. (2007). Determining the neu-ral substrates of goal-directed learning in the human brain. J. Neurosci.27,4019–4026.

Neural substrates of the goal-directed behavior in humans

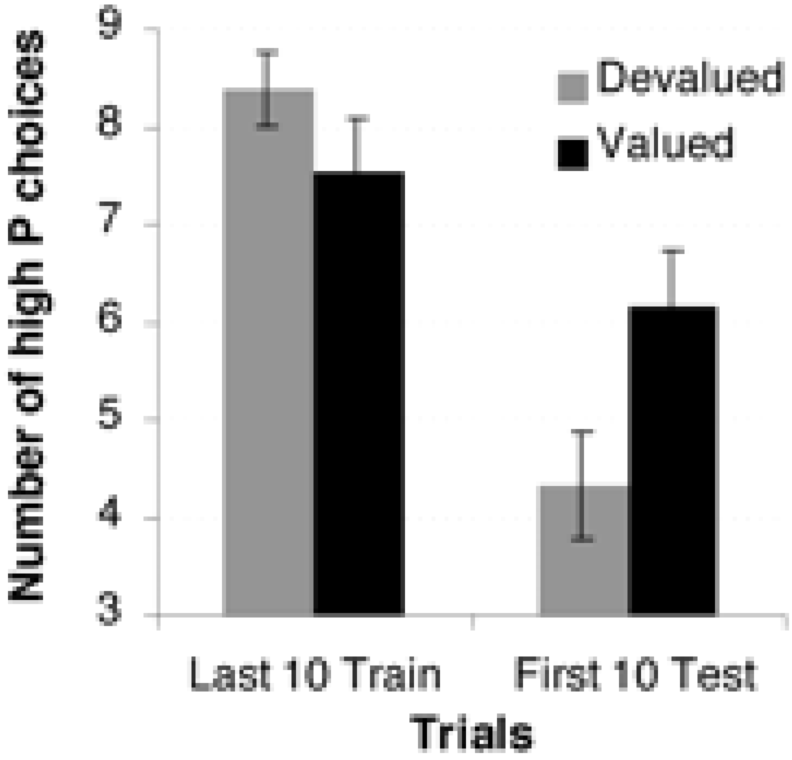
Results: behavioral

At testing, after devaluation, subjects reduced their choices of the high-probability action associated with the devalued food significantly more than that of the valued food (interaction with $p < 0.01$). Error bars indicate SEM.



B Train in scanner → Devalue one reward selective satiation → Test in scanner extinction

Tomato → Devaluec
Chocolate → Valuec



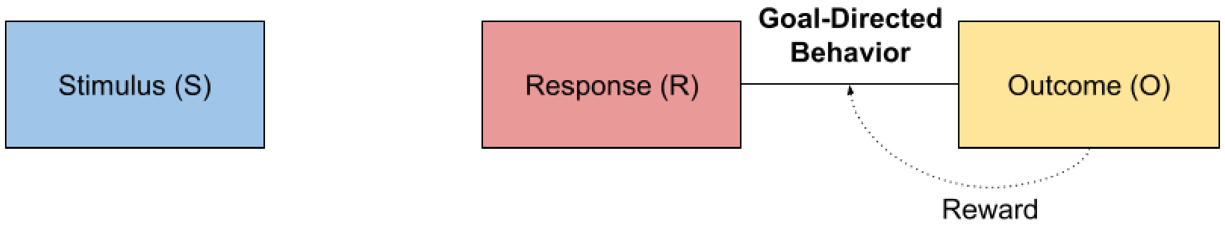
Valentin, V.V., Dickinson, A., and O’Doherty, J.P. (2007). Determining the neu-ral substrates of goal-directed learning in the human brain. J. Neurosci.27,4019–4026.

Goal-directed behavior/actions

The action is made because we think that they will lead to outcomes that we desire

Two criteria make an action goal-directed

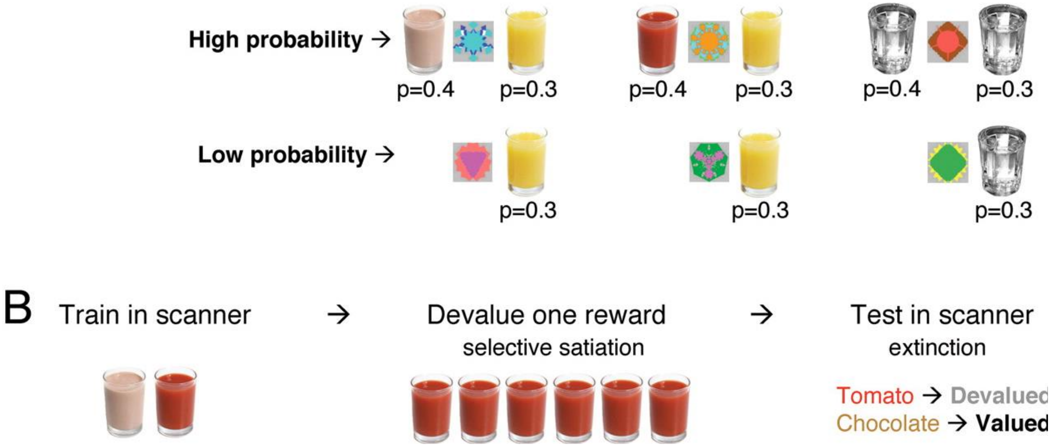
1. There must be **knowledge of the relationship between an action** (or sequence of actions) and its **consequences** --> response-outcome or R-O control
2. The **outcome should be motivationally relevant** or desirable at the moment of choice/action

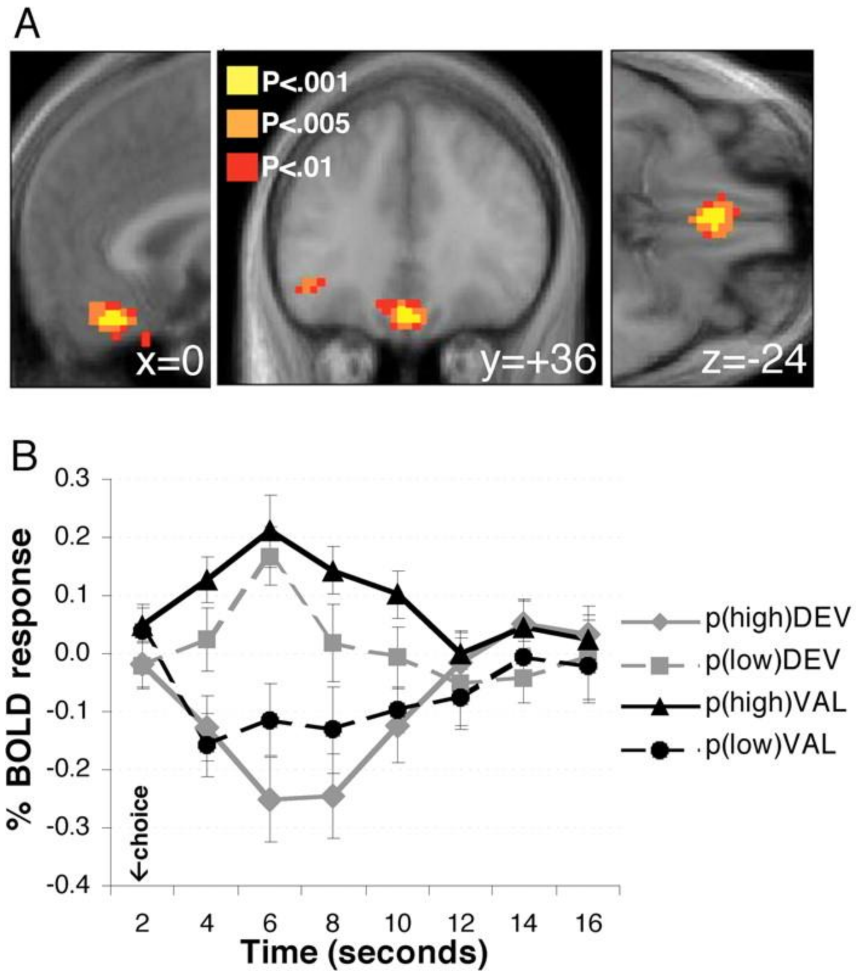


ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

Neural substrates of the goal-directed behavior in humans

Are there brain areas that respond differently between the still motivationally relevant outcome (I.e. valued) and the devalued one?



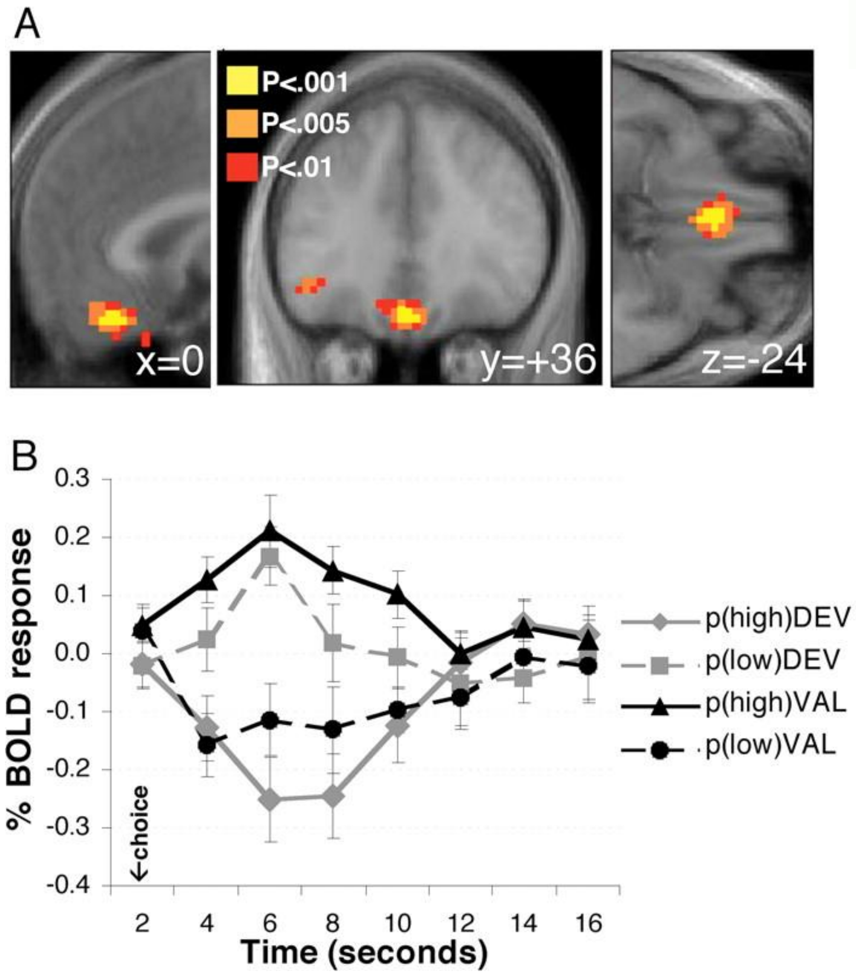


Valentin, V.V., Dickinson, A., and O’Doherty, J.P. (2007). Determining the neu-ral substrates of goal-directed learning in the human brain. J. Neurosci.27,4019–4026.

Neural substrates of the goal-directed behavior in humans

Results: neural

A, A region of the **medial OFC** showing a significant modulation in its activity during instrumental action selection as a function of the value of the associated outcome (mOFC; x = −3, *y* = 36, z = −24 mm; *Z* = 3.29; *p* < 0.001). ***B***, Time-course plots derived from the peak voxel (from each individual subject) in the mOFC during trials in which subjects chose each one of the four different actions (choice of the high- vs low-probability action in either the valued or devalued conditions).



Valentin, V.V., Dickinson, A., and O'Doherty, J.P. (2007). Determining the neu-ral substrates of goal-directed learning in the human brain. J. Neurosci.27,4019–4026.

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

The dopaminergic pathways

1. **Nigrostriatal pathway**

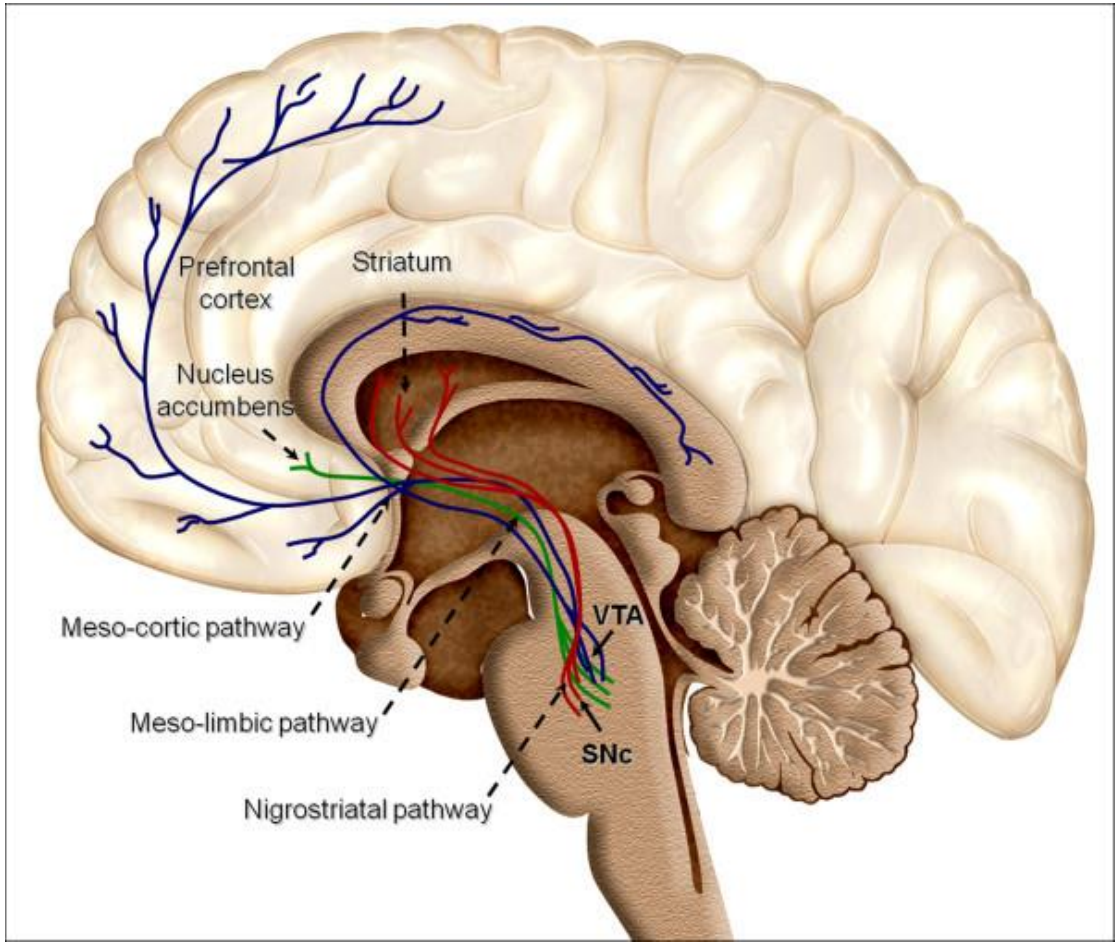
- originates in the substantia nigra pars compacta (SNc)
- projects primarily to the **caudate–putamen** (dorsal striatum in rodents)
- It is critical in the production of **movement** as part of the basal ganglia motor loop

2. **Mesolimbic pathway**

- originates in the VTA
- projects to the **nucleus accumbens**, septum, amygdala and hippocampus

3. **Mesocortical pathway**

- Originates in the VTA
- projects to the medial prefrontal, cingulate, **orbitofrontal** and perirhinal cortex



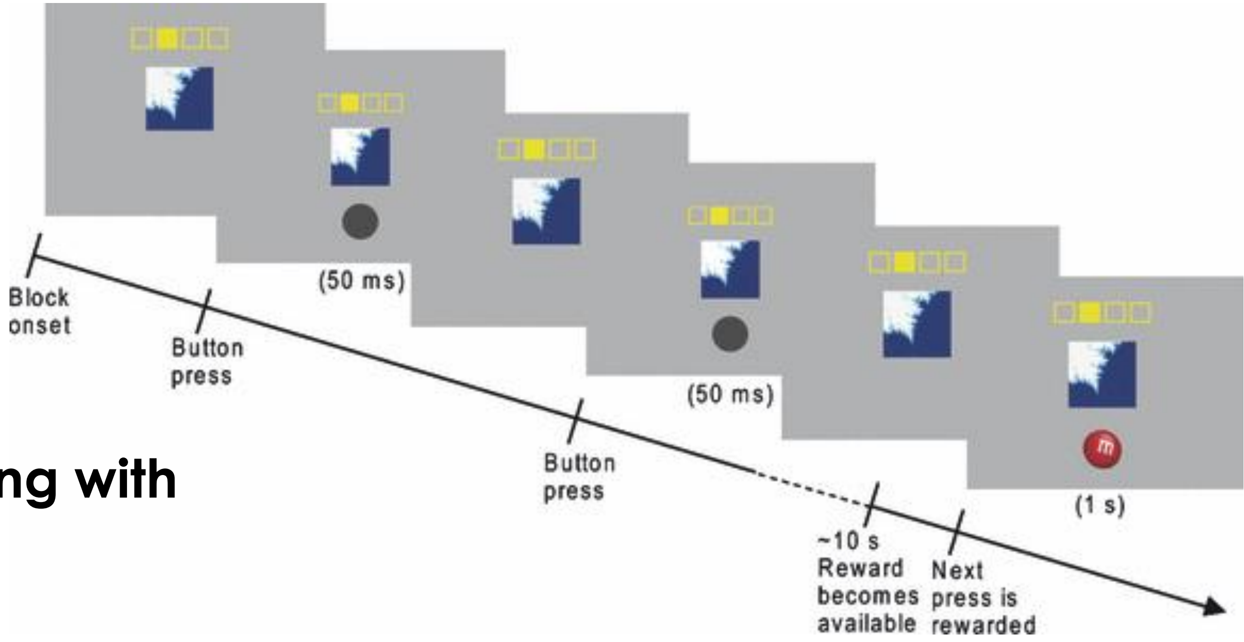
https://www.brainfacts.org/3d-brain#intro=false&focus=Brain

Neural substrates of habitual behavior in humans

Method:

Group 1: extensive training (6 times more than group2)

Group 2: little training



A fractal image was shown on the screen, along with a schematic indicating which button to press. Participants were instructed to press the indicated button as often as they liked; after each button press either a **gray circle** briefly appeared (50ms), indicating no reward, or a **picture of an M&M or Frito** appeared (1000ms), indicating a food reward corresponding to the picture. Only presses of the indicated button led to the display of the gray circle or food picture. Rewards were delivered on a variable interval 10-s schedule

Tricomi, E.M., Balleine, B.W., and O'Doherty, J.P. (2009). A specific role forposterior dorsolateral striatum in human habit learning. Eur. J. Neurosci.29,2225–2232

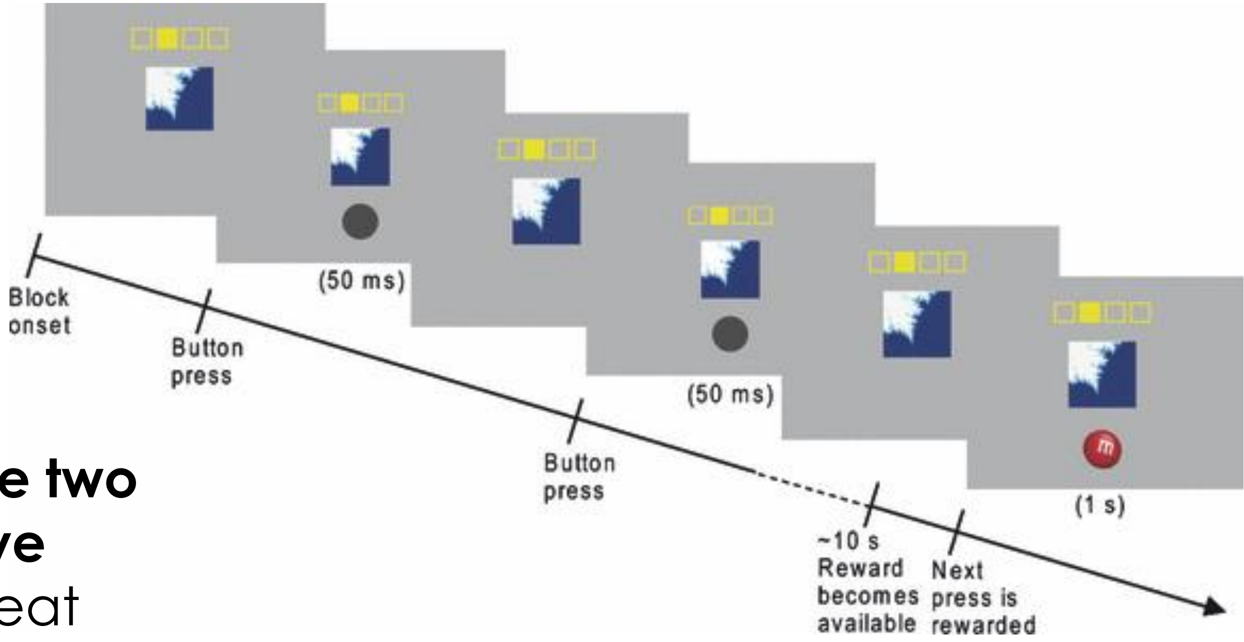
Neural substrates of habitual behavior in humans

Method:

Group 1: extensive training (6 times more than group2)

Group 2: little training

Following the final session of training, **one of the two food outcomes was devalued through selective satiation**, in which participants were asked to eat that food until it was no longer pleasant to them.



Tricomi, E.M., Balleine, B.W., and O’Doherty, J.P. (2009). A specific role forposterior dorsolateral striatum in human habit learning. Eur. J. Neurosci.29,2225–2232

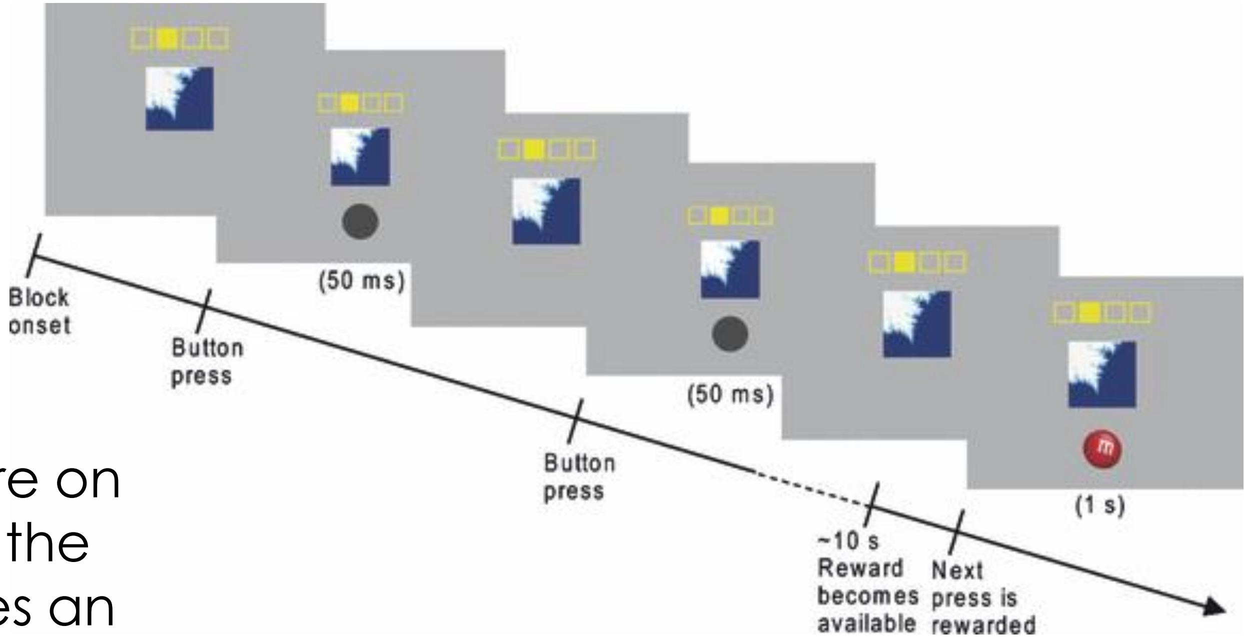
Neural substrates of habitual behavior in humans

Method:

Group 1: extensive training (6 times more than group2)

Group 2: little training

To test the **effects of the devaluation** procedure on behavior, participants were placed back into the scanner for a 3-min **extinction test**. This provides an explicit test for the presence of habitual behavior. If behavior remains goal-directed, participants should respond less for the food they no longer find pleasant than for the still valued food; however, if behavior has become habitual, the fractal cues should elicit responding irrespective of the outcome value.



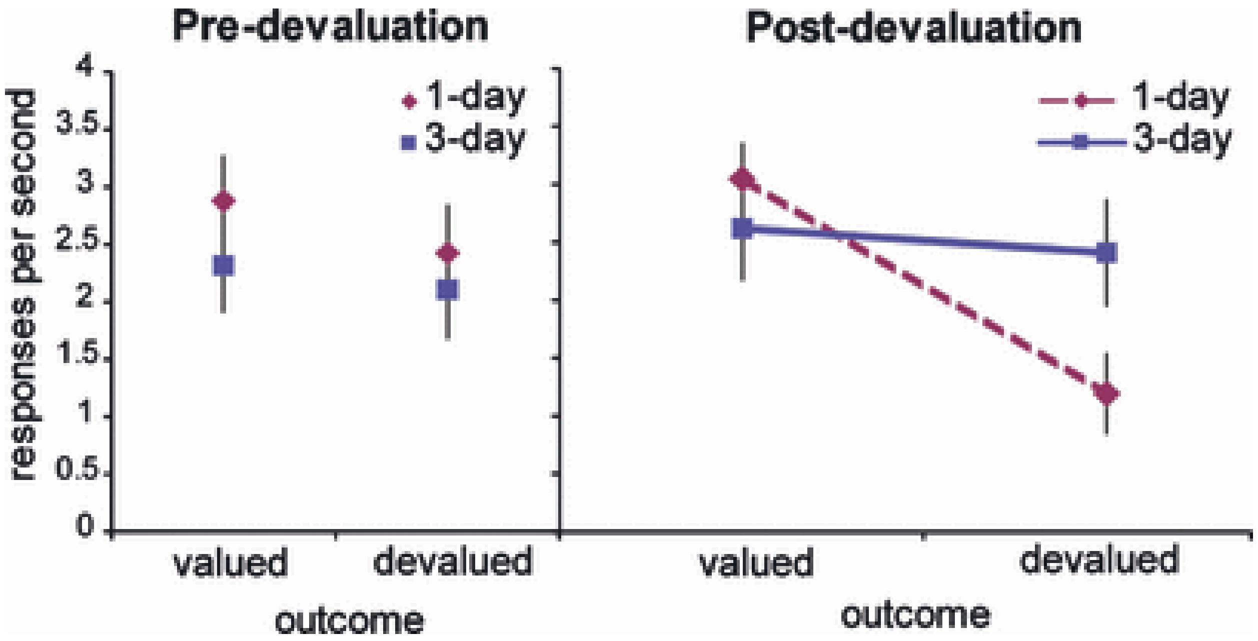
Tricomi, E.M., Balleine, B.W., and O'Doherty, J.P. (2009). A specific role forposterior dorsolateral striatum in human habit learning. Eur. J. Neurosci.29,2225–2232

Neural substrates of habitual behavior in humans

Results: behavioral

During the last session of training, **prior to the devaluation procedure** (left), there were no significant differences in response rates between groups or when responding for the two food rewards (one which will be devalued through selective satiation and one which will not).

During the test **following the devaluation procedure**, response rates for the still-valued outcome remained high, as did response rates for the devalued outcome for the 3-day group. In contrast, response rates for **the 1-day group for the devalued outcome were reduced**.

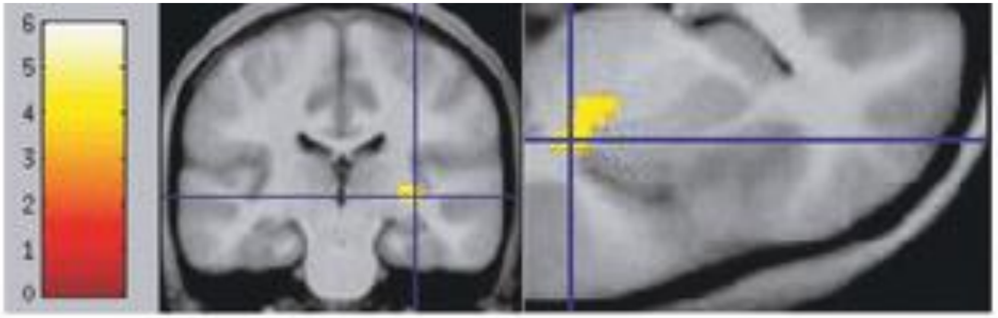


Tricomi, E.M., Balleine, B.W., and O’Doherty, J.P. (2009). A specific role forposterior dorsolateral striatum in human habit learning. Eur. J. Neurosci.29,2225–2232

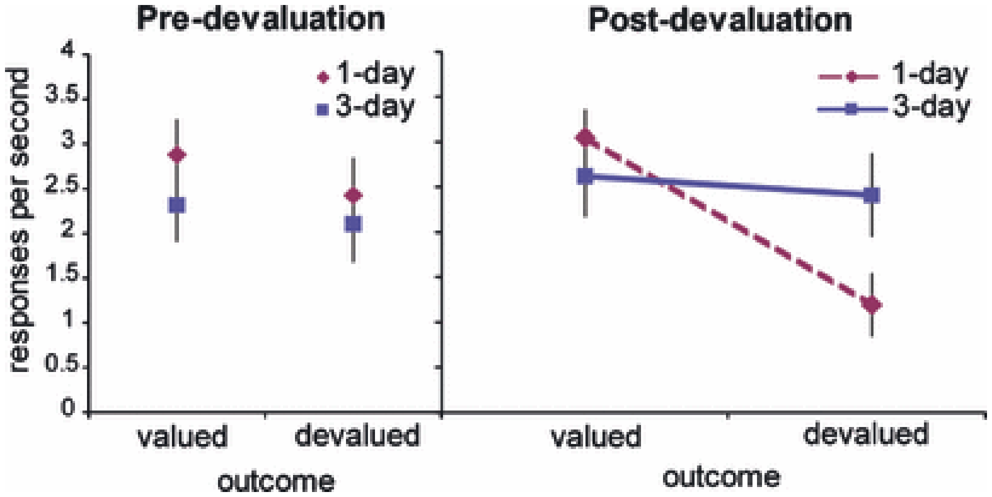
Neural substrates of habitual behavior in humans

Results: neural

Our within-subjects analysis of the last two sessions of training versus the first two sessions **in the 3-day group** revealed several significant voxel clusters, including a region within the **dorsolateral striatum** (DLS), in the right posterior putamen–globus pallidus



Right Putamen



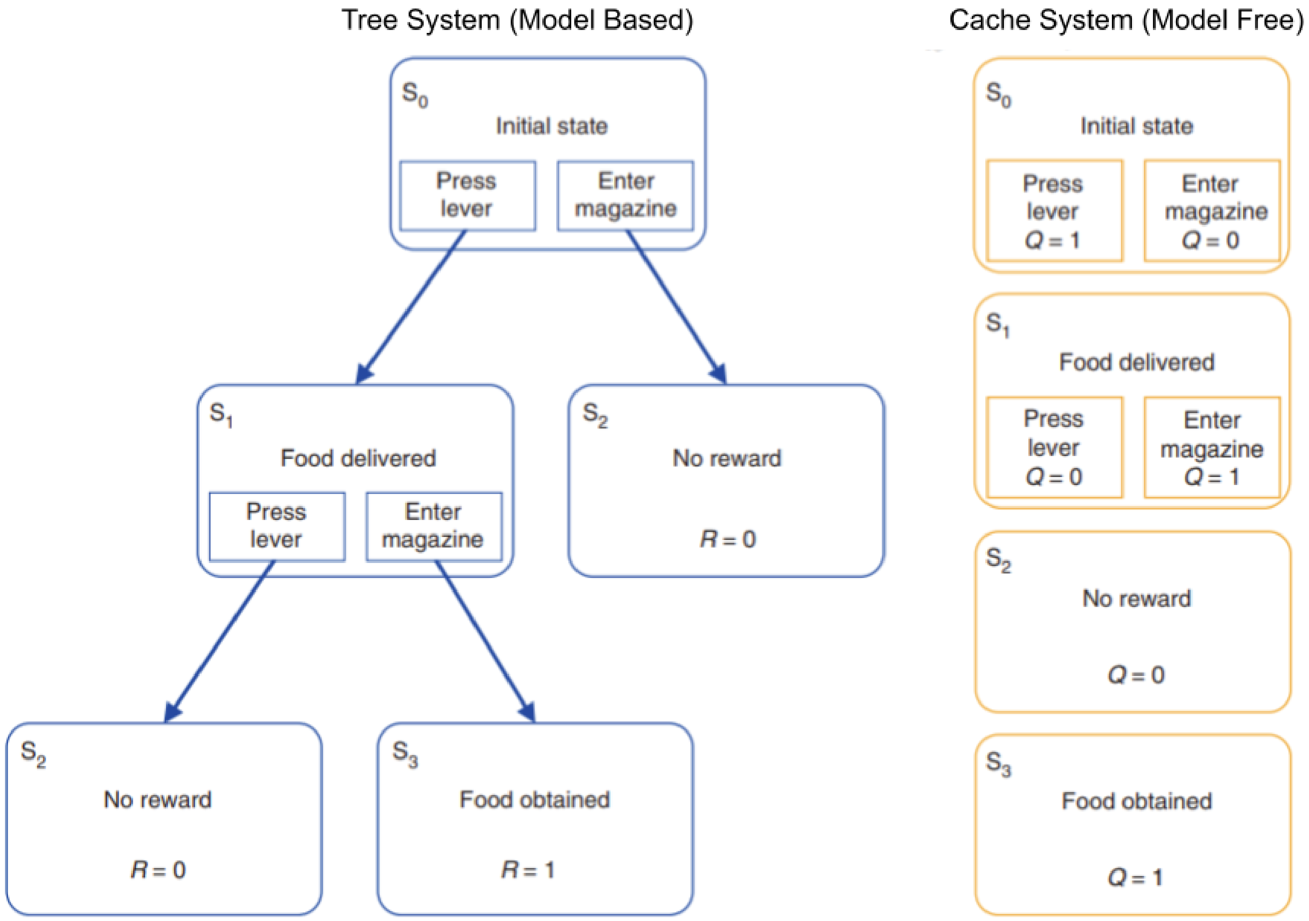
Tricomi, E.M., Balleine, B.W., and O’Doherty, J.P. (2009). A specific role forposterior dorsolateral striatum in human habit learning. Eur. J. Neurosci.29,2225–2232

Generation 3: model-based vs model-free computational analyses

Computational formalization of

- Goal-directed actions --> model-based
- Habitual actions --> model-free
- Their interaction

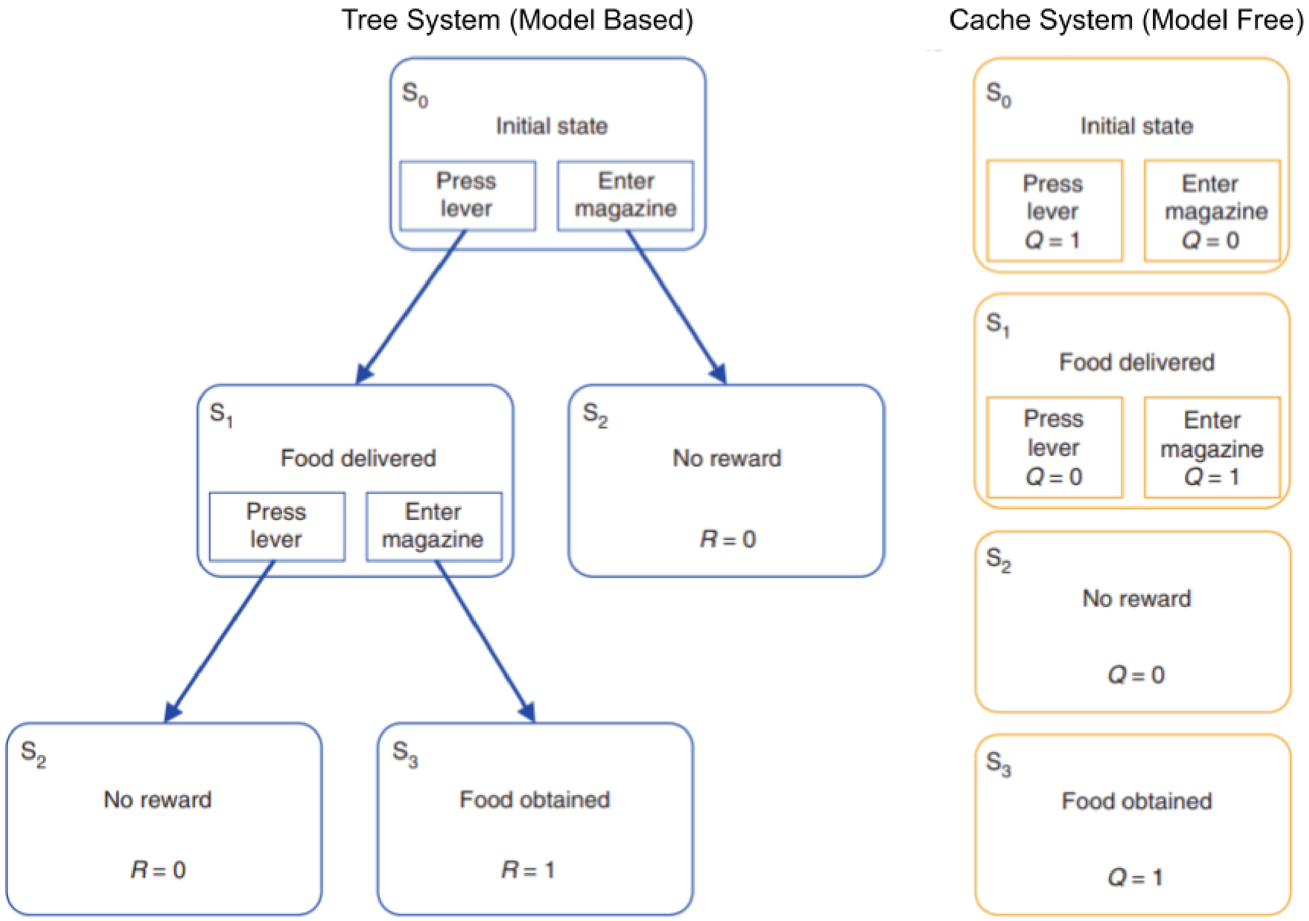
Model means anything an agent can use to predict how its environment will respond to its actions in terms of state transitions and rewards



Daw et al (2005)

Generation 3: model-based vs model-free computational analyses

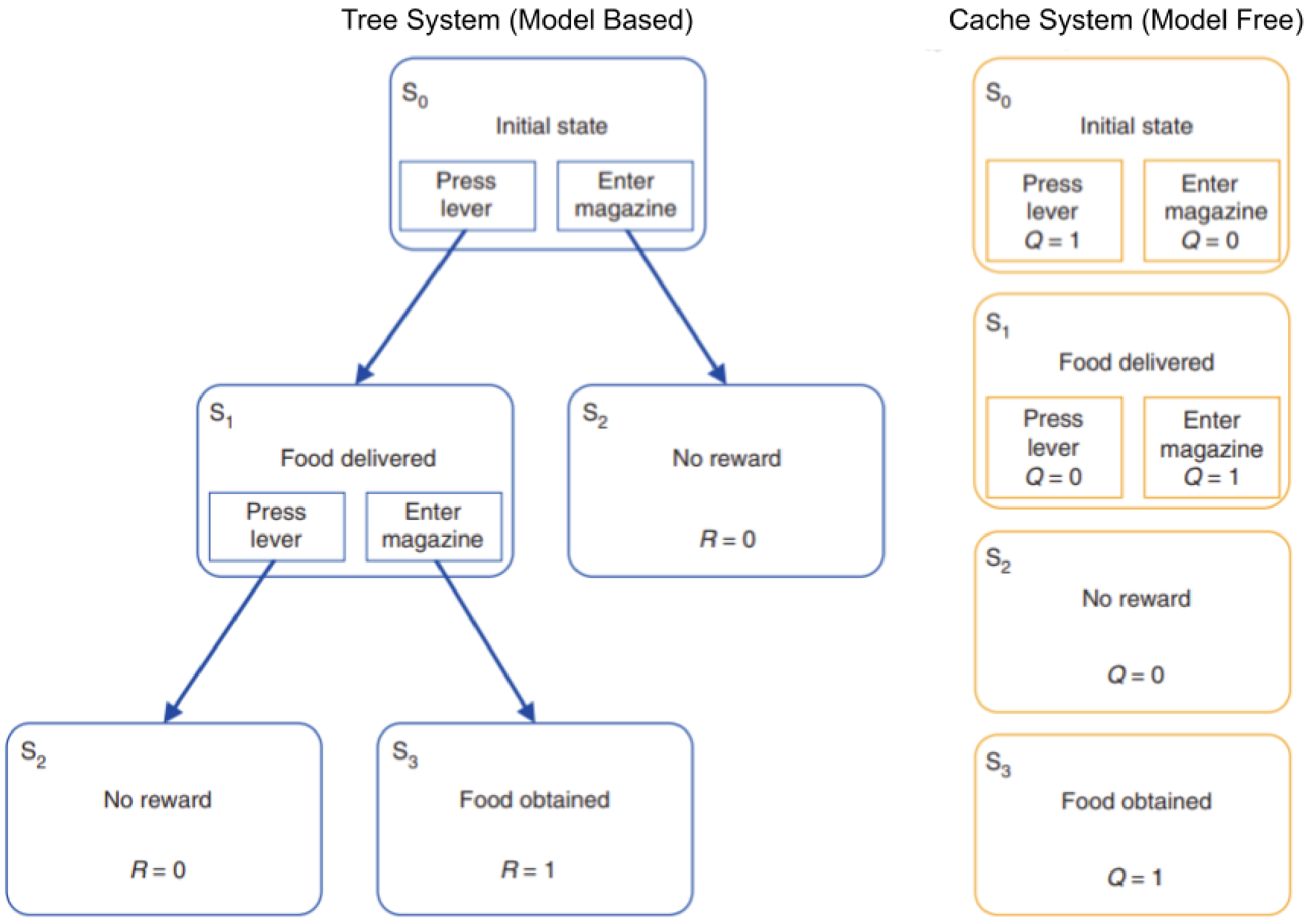
- Goal-directed actions --> model-based
  - A model-based algorithm selects actions by using a model to predict the consequences of possible courses of action in terms of future states and the reward signals expected to arise from those states
- Habitual actions --> model-free
  - A model-free algorithm selects actions relying on stored action values for all the state–action pairs obtained over many learning trials



Daw et al (2005)

Generation 3: model-based vs model-free computational analyses

- Goal-directed actions --> model-based
  - When the environment of a model-based agent changes the way it reacts to the agent’s actions, the agent can update the value (policy) of future states without the need to move to them.
- Habitual actions --> model-free
  - When the environment of a model-free agent changes the way it reacts to the agent’s actions, the agent has to move to that state, act from it, possibly many times, and experience the consequences of its actions.

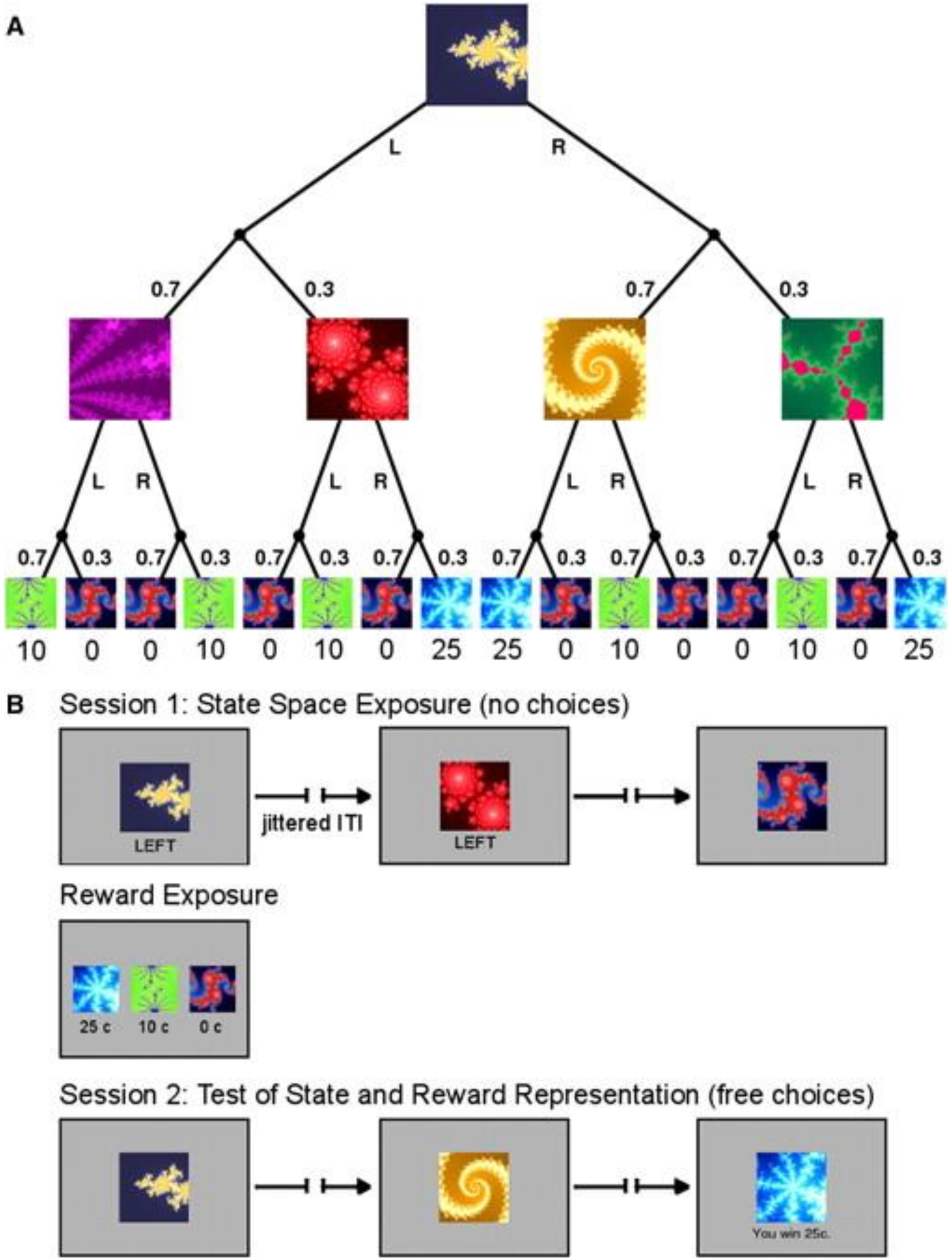


Daw et al (2005)

Latent learning in humans?

Method

(A) The experimental task was a **sequential two-choice Markov decision task** in which all decision states are represented by fractal images. The task design follows that of a binary decision tree. Each trial begins in the same state. Subjects can choose between a left (L) or right (R) button press. With a certain probability (0.7/0.3) they reach one of two subsequent states in which they can choose again between a left or right action. Finally, they reach one of three outcome states associated with different monetary rewards (0¢, 10¢, and 25¢).



Gläscher J, Daw N, Dayan P, O'Doherty J. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron.* 2010;66:585–595

Latent learning in humans?

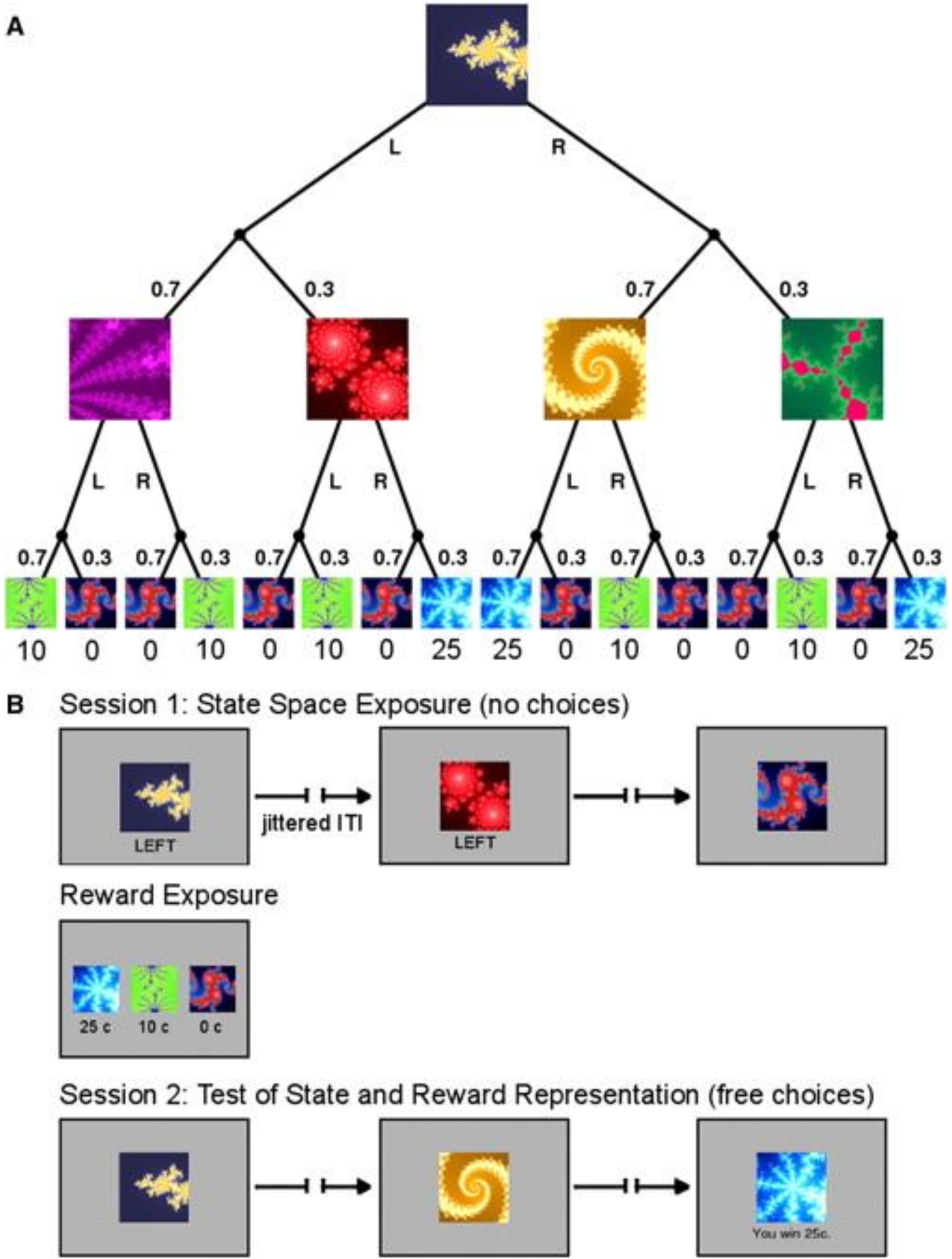
Method

(B) The experiment proceeded in two fMRI scanning sessions of 80 trials each.

In the **first session**, subject **choices were fixed** and presented to them below the fractal image. However, subjects could still learn the transition probabilities.

Between scanning **sessions** subjects were presented with the **reward schedule** that maps the outcome states to the monetary payoffs. This mapping was rehearsed in a short choice task.

Finally, in the **second scanning session**, subjects were **free to choose** left or right actions in each state. In addition, they also received the payoffs in the outcome states.



Gläscher J, Daw N, Dayan P, O'Doherty J. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron.* 2010;66:585–595

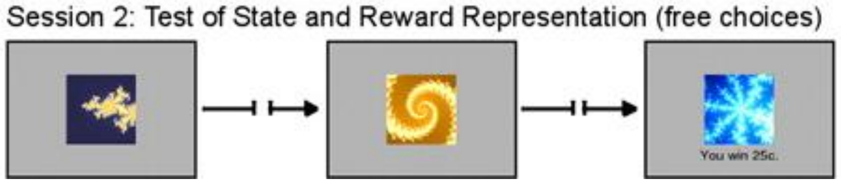
Latent learning in humans?

Results: Behavioral from free-choice session

Test if participants were able to make optimal choices by combining the knowledge they acquired about state transitions (session 1) and reward contingencies (between sessions).

Any successful learning would be possible with model-based, but not model-free, learning

Can you tell why?


Session 2: Test of State and Reward Representation (free choices)
You win 25c.

Gläscher J, Daw N, Dayan P, O'Doherty J. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron.* 2010;66:585–595

Latent learning in humans?

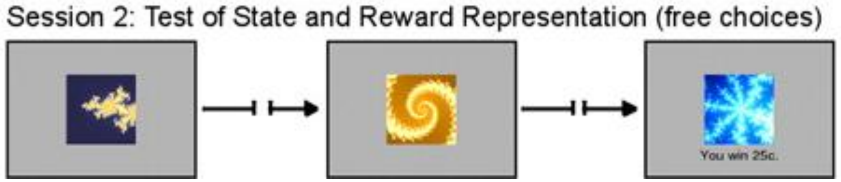
Results: Behavioral from free-choice session

Test if participants were able to make optimal choices by combining the knowledge they acquired about state transitions (session 1) and reward contingencies (between sessions).

Any successful learning would be possible with model-based, but not model-free, learning

Can you tell why?

Model-free learning focuses exclusively on predicting rewards, so it learns only if rewards are given.


Session 2: Test of State and Reward Representation (free choices)
You win 25c.

In state 1, at the first trial, of all 18 subjects:

- 13 made the optimal choice
- 5 made the wrong choice in state

indicating that their choice of behavior **cannot be** explained by model-free learning theory.

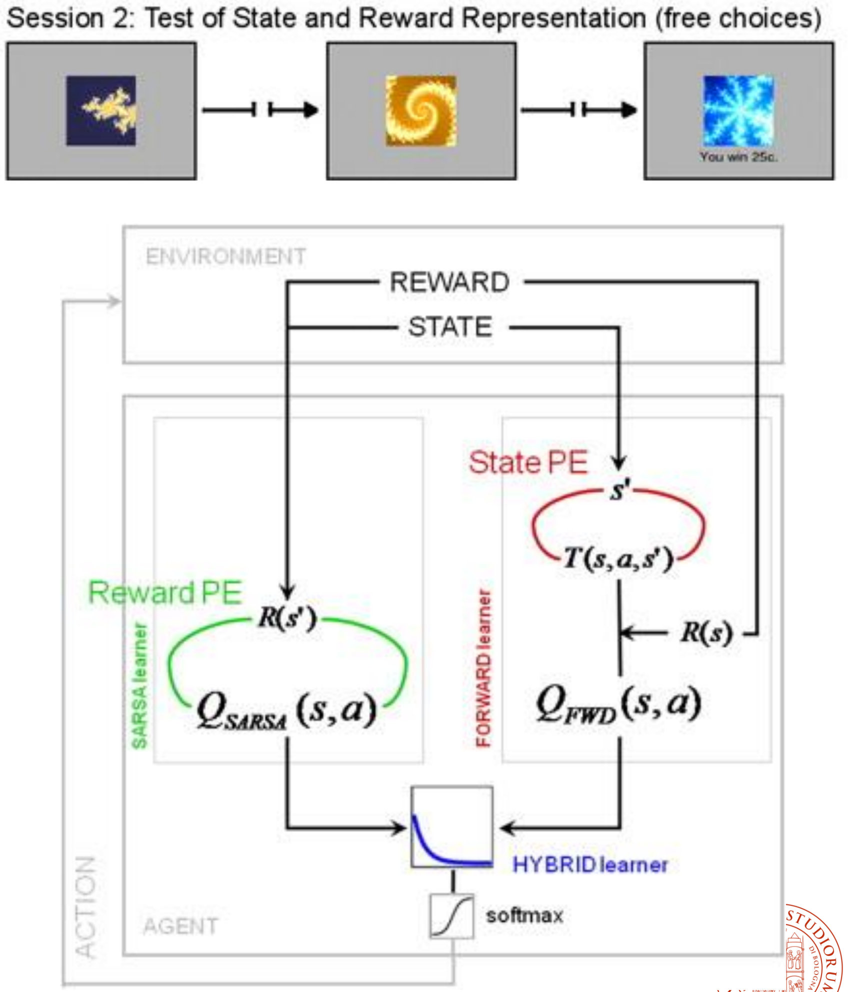
Gläscher J, Daw N, Dayan P, O'Doherty J. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron.* 2010;66:585–595

Latent learning in humans?

Results: Computational from free-choice session

Choice behavior during the entire session was best explained by hybrid model that integrates both

- Reward PE: model-free (similar TD model)
- State PE: model-based



Gläscher J, Daw N, Dayan P, O'Doherty J. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron.* 2010;66:585–595

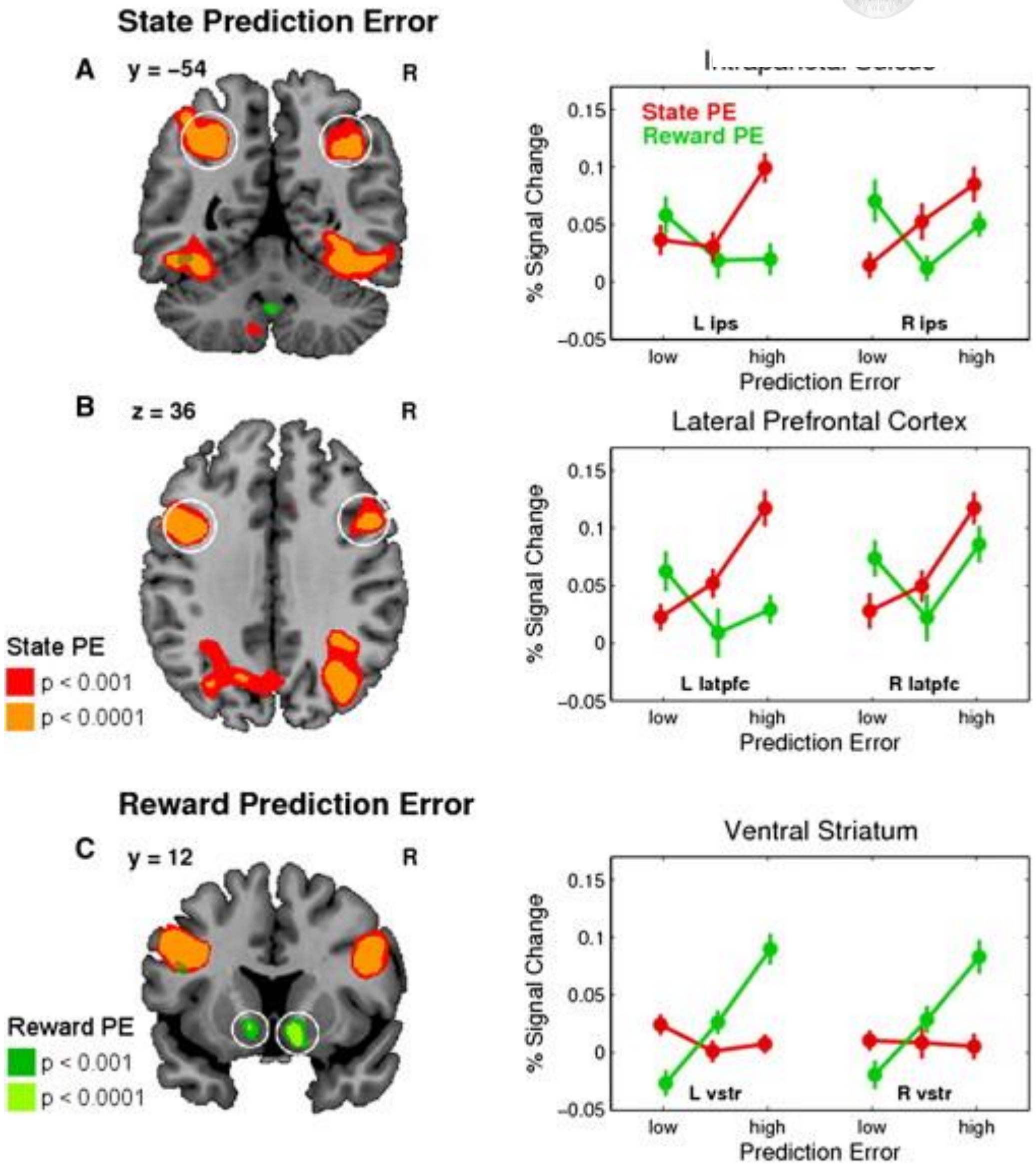
Neural Signatures of Reward PE and State PE

Results: Neural

Parameters estimated from computational models were used to find activations that correlated with SPE & RPE.

(A and B) Significant effect for **SPE** bilaterally in the intraparietal sulcus (ips) and **lateral prefrontal cortex** (lpfc).

(C) Significant effects for **RPE** in the **ventral striatum** (vstr).



Gläscher J, Daw N, Dayan P, O'Doherty J. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron.* 2010;66:585–595

Let's try this:

On a computer (sorry, not a phone) go to

https://nivlab.github.io/jspsych-demos/tasks/two-step/experiment.html

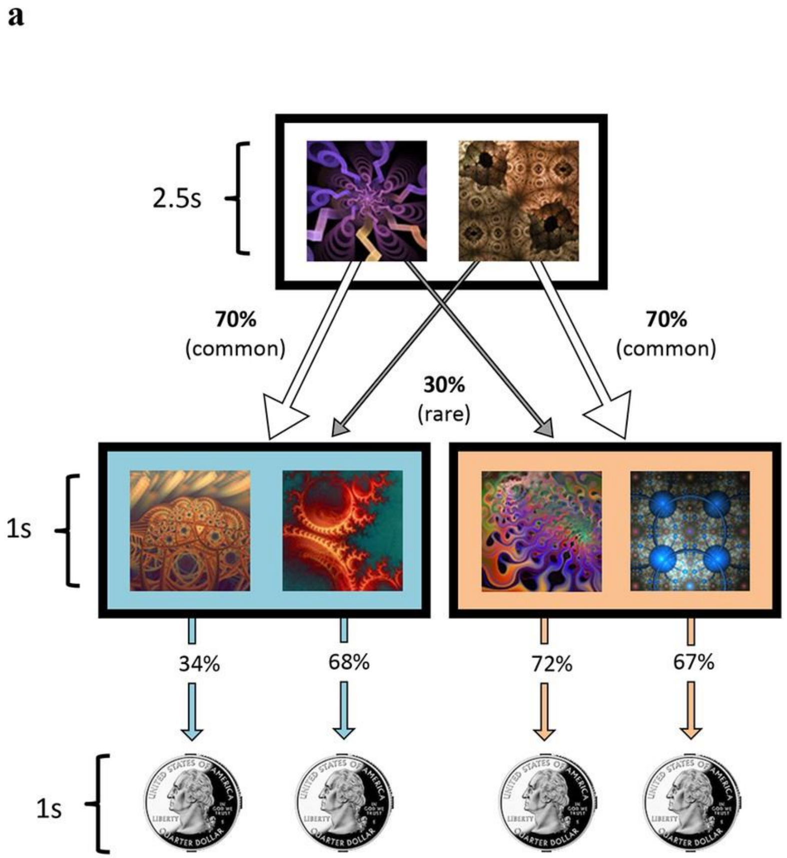
and play through instructions & game for a while

https://nivlab.github.io/jspsych-demos/

Sequential two-choice Markov decision tasks

Developed to

- discern the influence of model-free vs model-based controller on behavior
- to determine whether neural signals are correlated with predictions and prediction errors specific to each controller



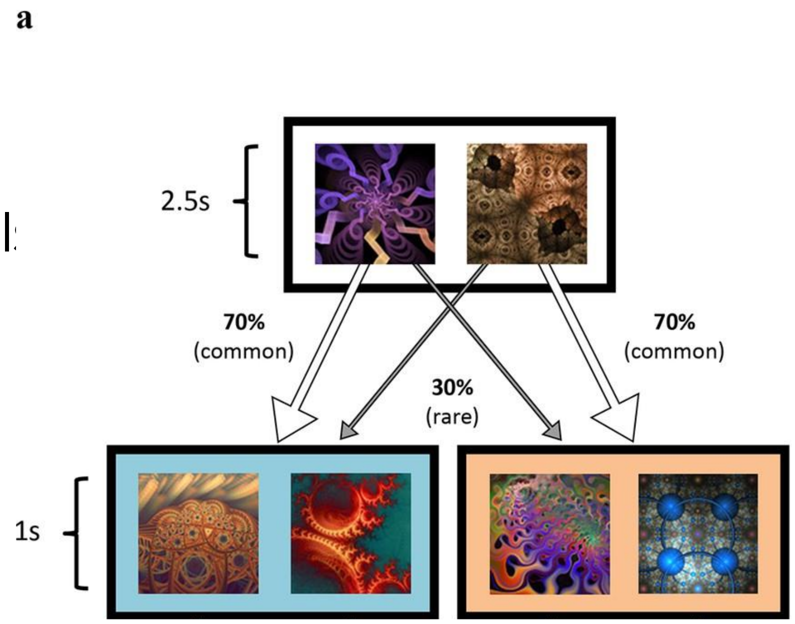
https://doi.org/10.7554/eLife.11305.003

Sequential two-choice Markov decision tasks

"Your task is to maximize the reward"

(**a**) Subjects chose between two fractal which probabilistically determined whether they would transition to the orange or blue second stage state.

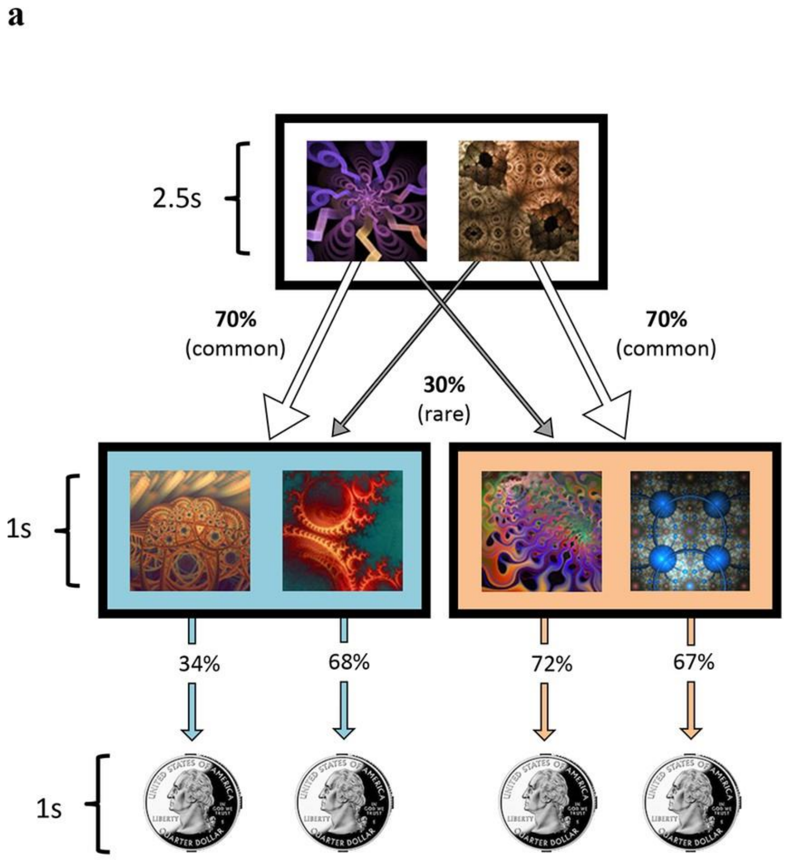
Action at the first state is associated with one likely and one unlikely transition. For example, the fractal on the left had a 70% chance of leading to the blue second stage state (‘common’ transition) and a 30% chance of leading to the orange state (‘rare’ transition). **These transition probabilities were fixed and could be learned over time.**



https://doi.org/10.7554/eLife.11305.003

Sequential two-choice Markov decision tasks

(**a**) In the second stage state, subjects chose between two fractals, each of which was associated with a distinct probability of being rewarded with a 25 cents coin.



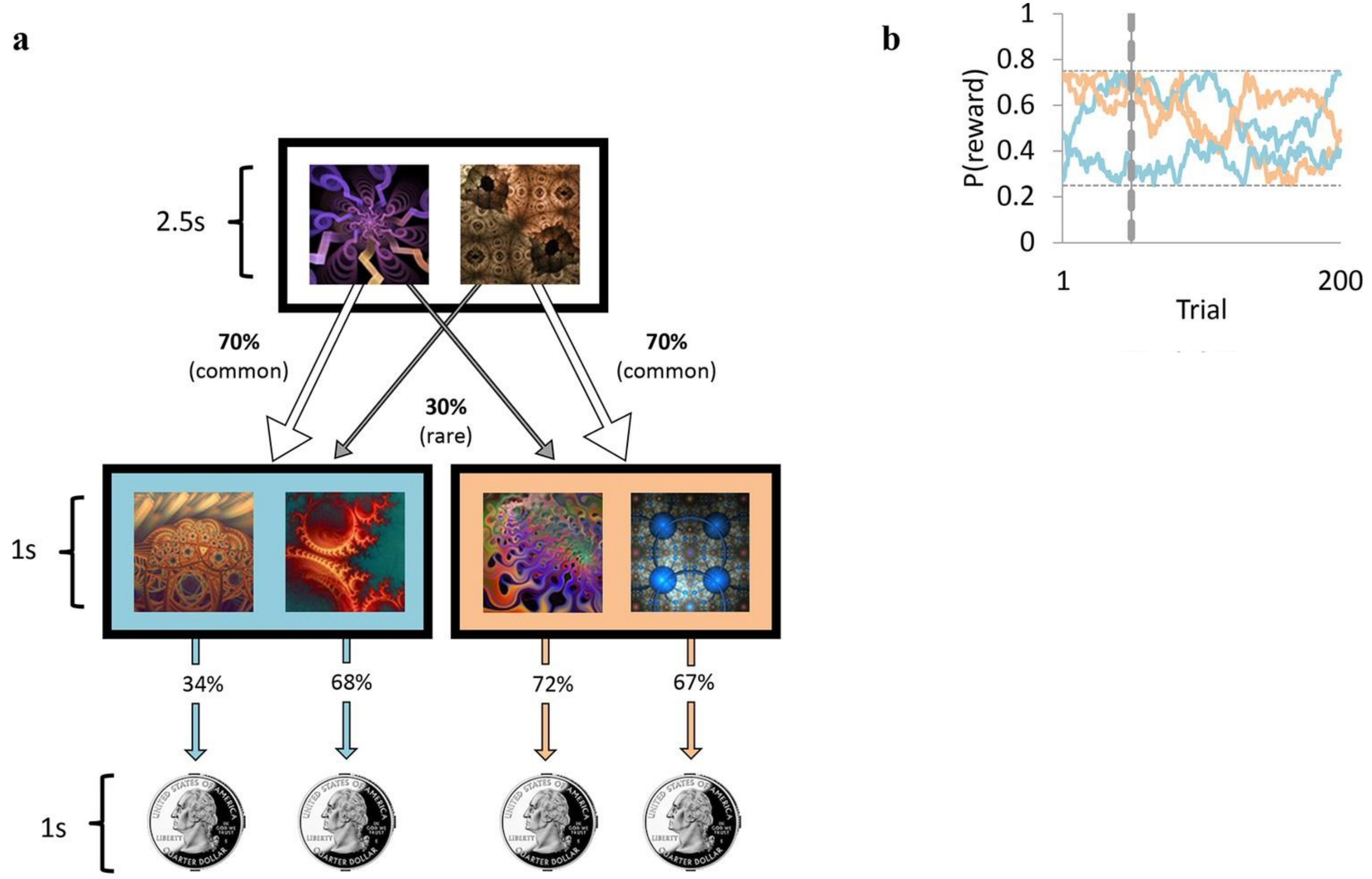
https://doi.org/10.7554/eLife.11305.003

Sequential two-choice Markov decision tasks

(**b**) Drifting reward probabilities determined by Gaussian Random Walks for 200 trials with grey horizontal lines indicating boundaries at 0.25 and 0.75.

To incentivize subjects to continue learning throughout the task, the chances of p**ay off associated with the four second-stage options were changed** slowly and independently.

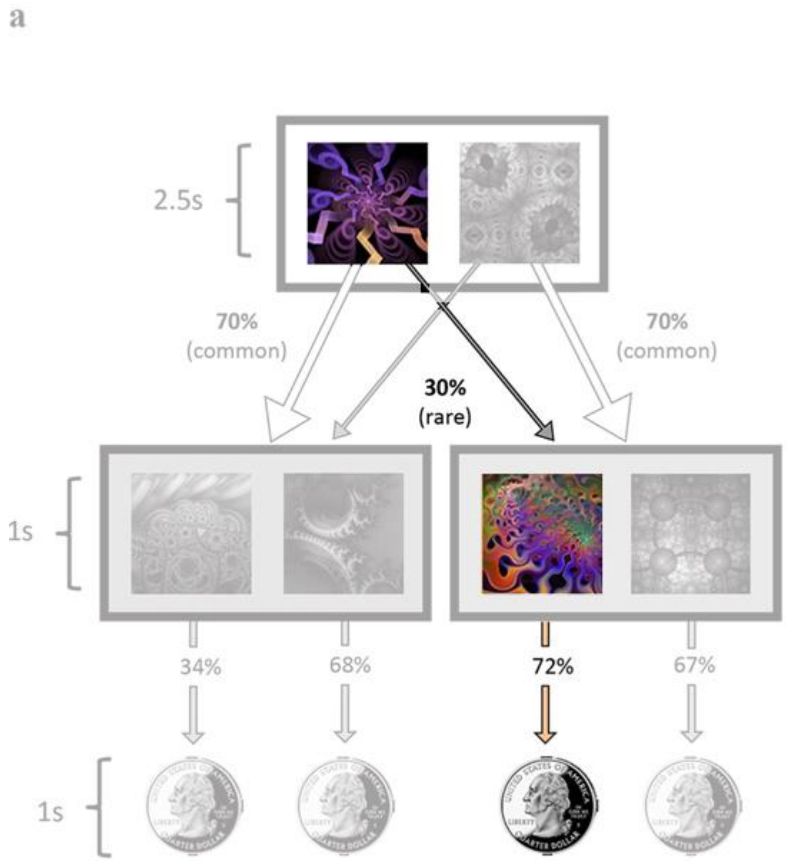
The chance of winning is almost stochastic.



https://doi.org/10.7554/eLife.11305.003

Sequential two-choice Markov decision tasks

How does bottom-stage outcome affect top-stage choices?



https://doi.org/10.7554/eLife.11305.003

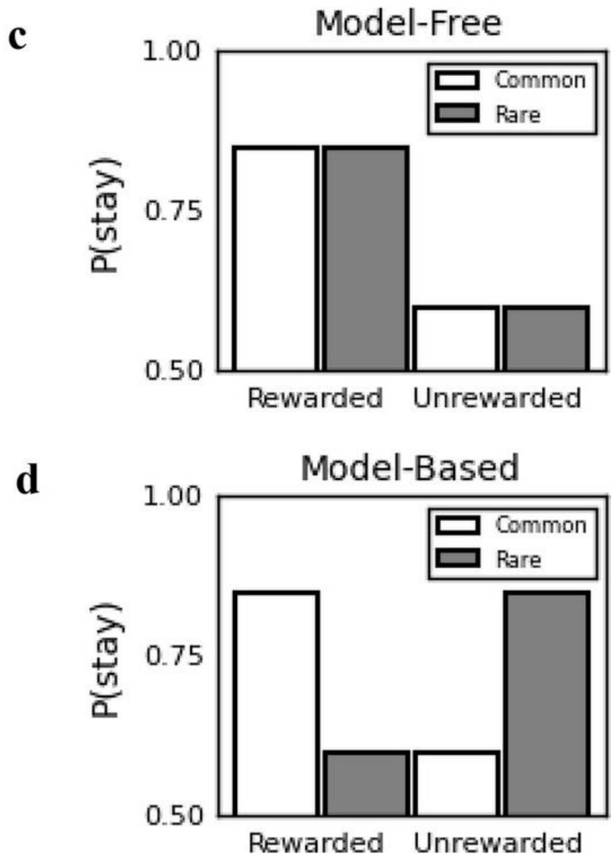
Sequential two-choice Markov decision tasks

https://doi.org/10.7554/eLife.11305.003

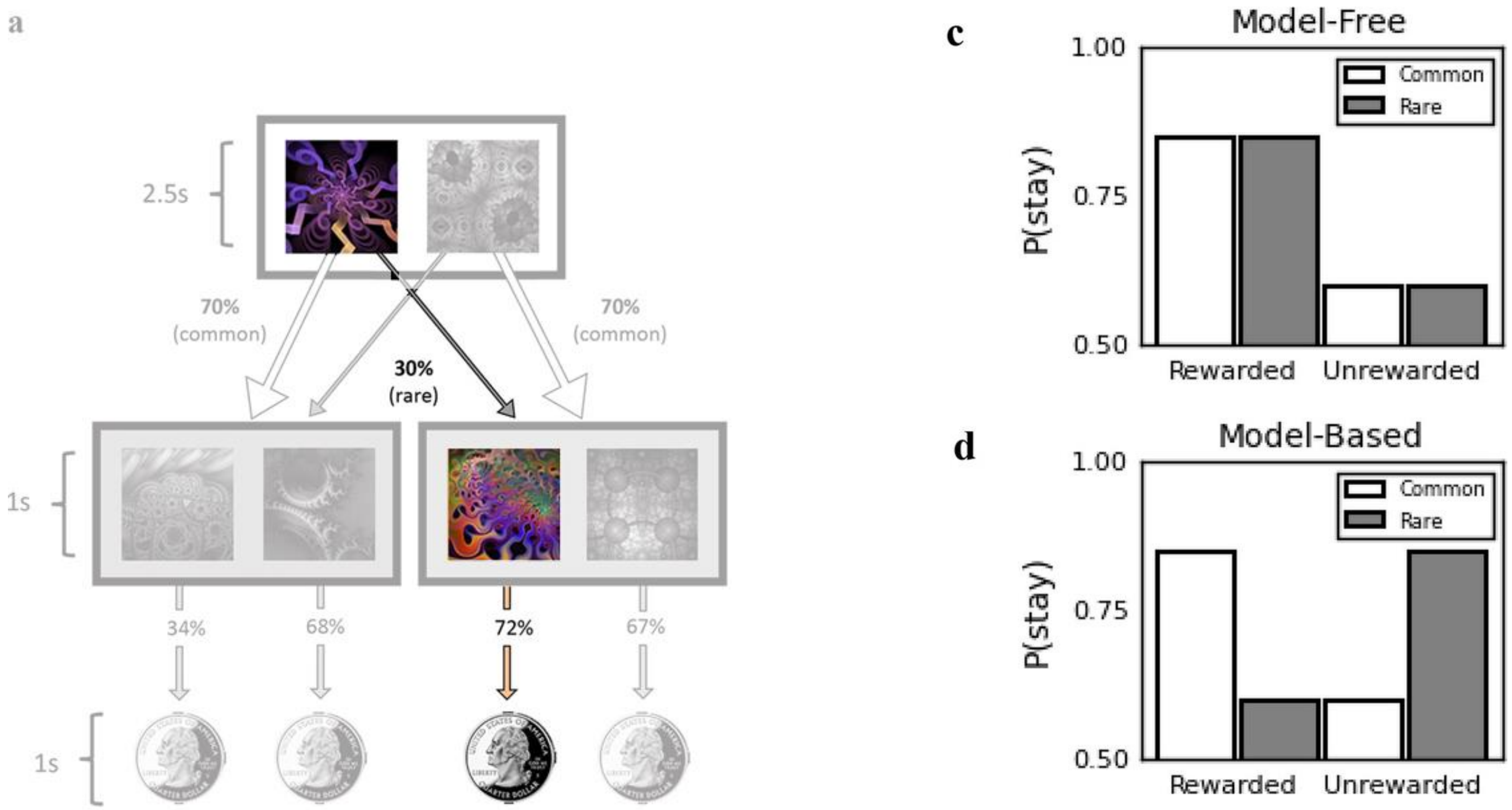
How does bottom-stage outcome affect top-stage choices?

Model-free and model-based agents differ in the action selected after a **rare transition**.

- Model-free agent
  - ignores transition structure
  - prefers to repeat actions that lead to reward, irrespective of the likelihood of that first transition.
- Model-based agent
  - respects transition structure
  - can ascribe rewards following a rare transition to an alternative (non-selected) action—which, despite not predicting reward on the current trial, will be more likely to lead to reward on future trials.
  - model-based strategy predicts a crossover interaction between the two factors (reward and transition), because a rare transition inverts the effect of the subsequent reward.



Sequential two-choice Markov decision tasks

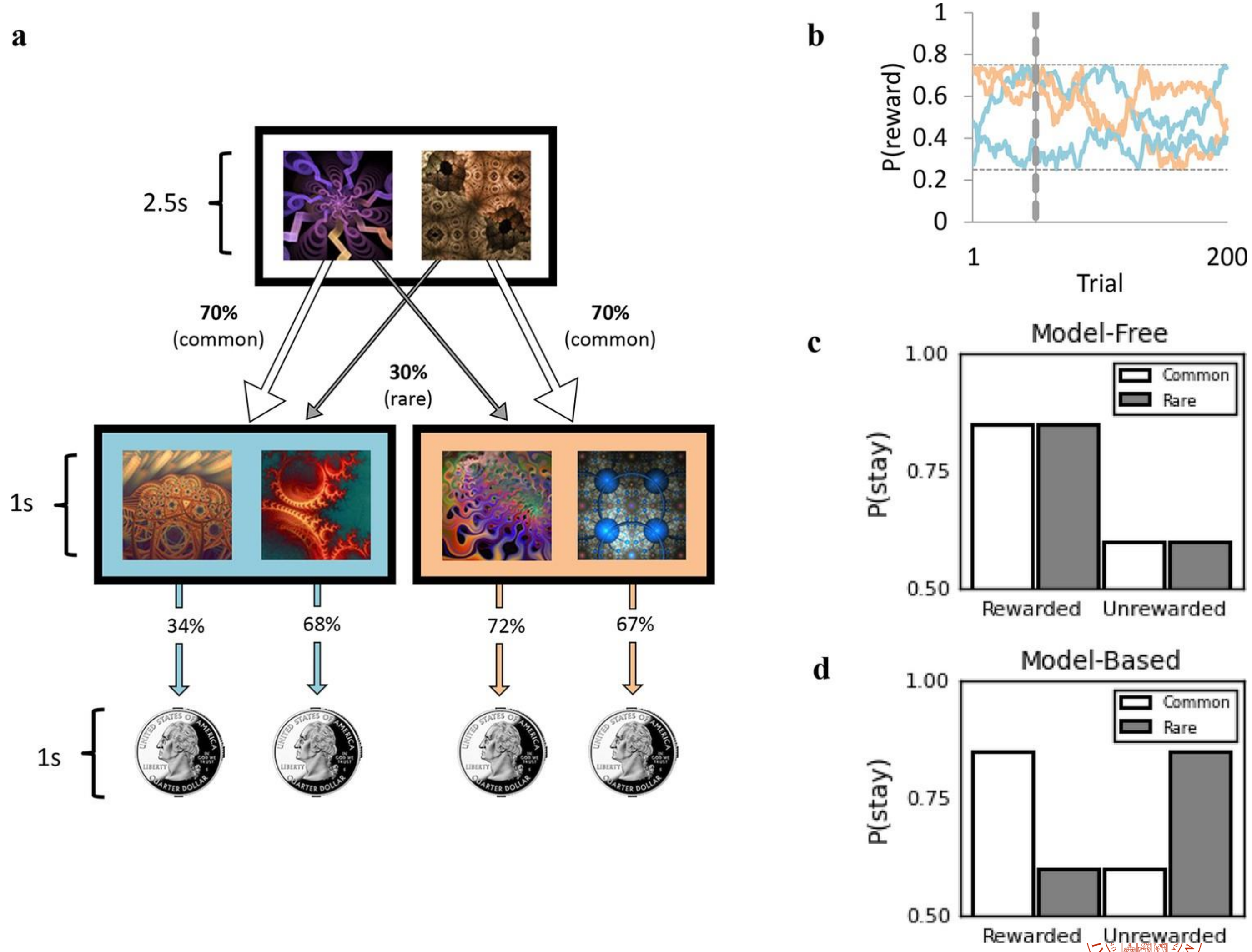


https://doi.org/10.7554/eLife.11305.003

Sequential two-choice Markov decision tasks

(**c**) Schematic representing the performance of a purely ‘model-free’ learner, who only exhibits sensitivity to whether or not the previous trial was rewarded vs. unrewarded, and does not modify their behavior in light of the transition that preceded reward.

(**d**) Schematic representing the performance of a purely ‘model-based’ learner, who is more likely to repeat an action (i.e. ‘stay’) following a rewarded trial, only if the transition was common. If the transition to that rewarded state was rare, they are more likely to switch on the next trial.

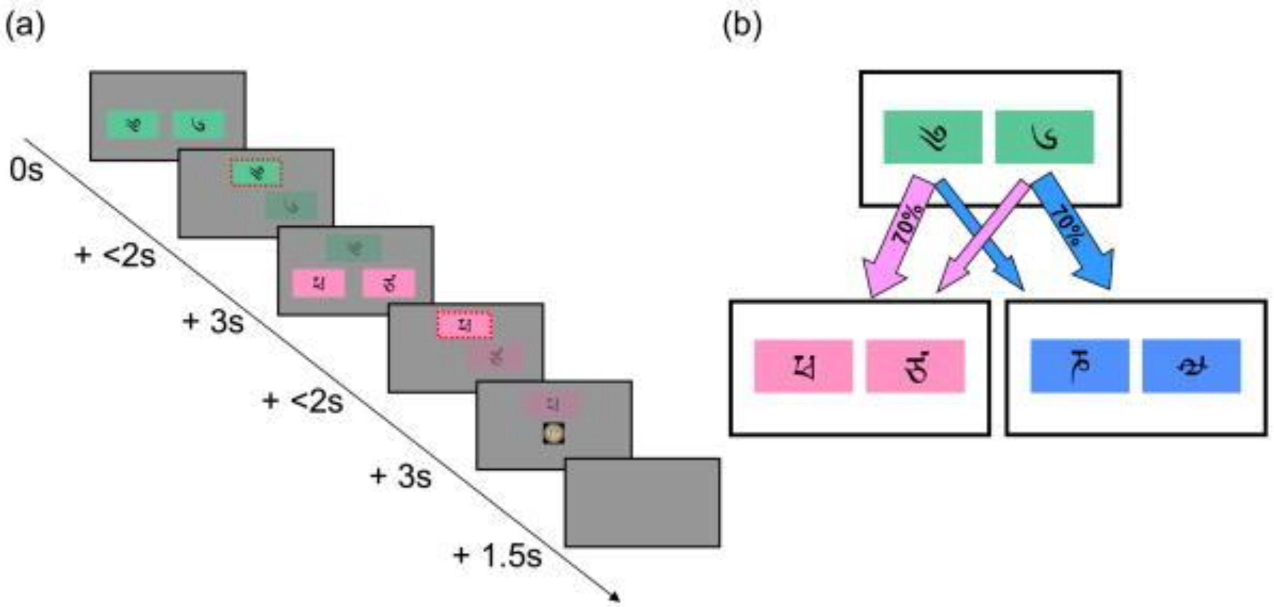


https://doi.org/10.7554/eLife.11305.003

Detecting simultaneous correlates of model-free and model-based systems

Method

(A) Timeline of events in trial. A first-stage choice between two options (green boxes) leads to a second-stage choice(here, between two pink options), which is reinforced with money.(B) State transition structure. Each first-stage choice is predominantly associated with one or the other of the second-stage states, and leads there 70% of the time.



To incentivize subjects to continue learning throughout the task, the chances of pay off associated with the four second-stage options are changed slowly and independently

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. *Neuron*. 2011;69(6):1204-1215. doi:10.1016/j.neuron.2011.02.027

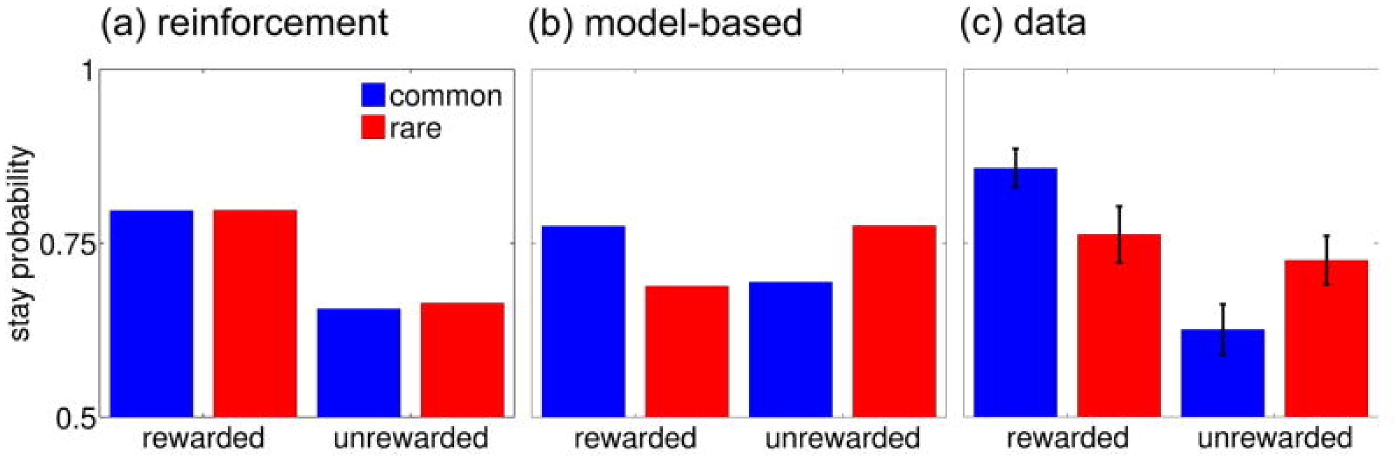
Detecting simultaneous correlates of model-free and model-based systems

Results: Analysis of choice behavior.

(a) Simple reinforcement predicts that a first-stage choice resulting in reward is more likely to be repeated on the subsequent trial, regardless of whether that reward occurred after a common or rare transition.

(b) Model-based prospective evaluation instead predicts that a rare transition should affect the value of the other first-stage option, leading to a predicted interaction between the factors of reward and transition probability.

(c) Actual stay proportions, averaged across subjects, display hallmarks of both strategies. Error bars: 1 SEM.



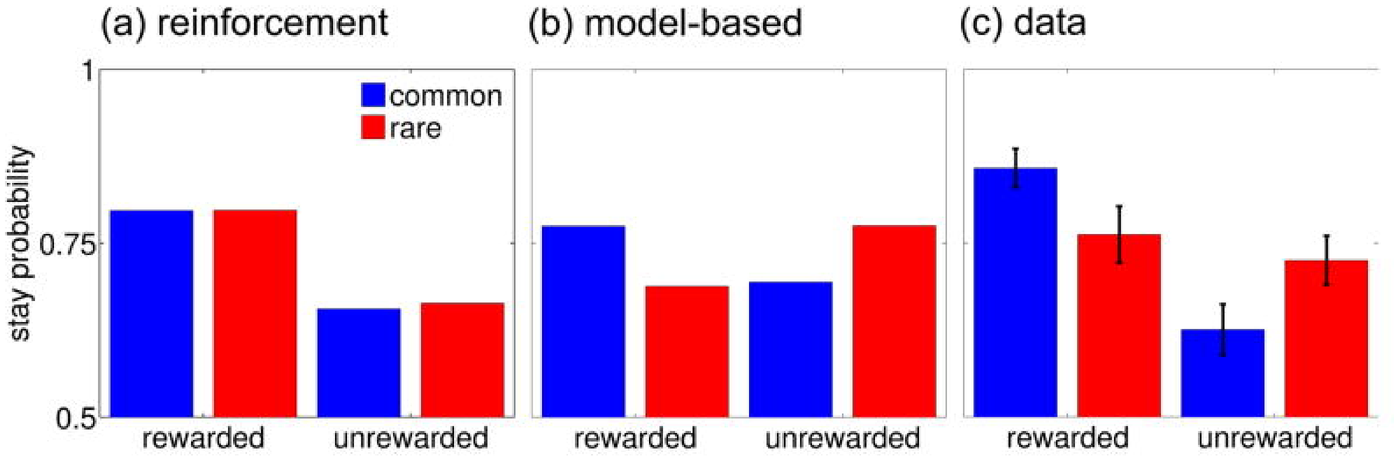
Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. *Neuron*. 2011;69(6):1204-1215. doi:10.1016/j.neuron.2011.02.027

Detecting simultaneous correlates of model-free and model-based systems

Results: Computational

Choice behavior during was best explained by hybrid model that integrates both

- Reward PE: model-free (similar TD model)
- State PE: model-based



Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. *Neuron*. 2011;69(6):1204-1215. doi:10.1016/j.neuron.2011.02.027

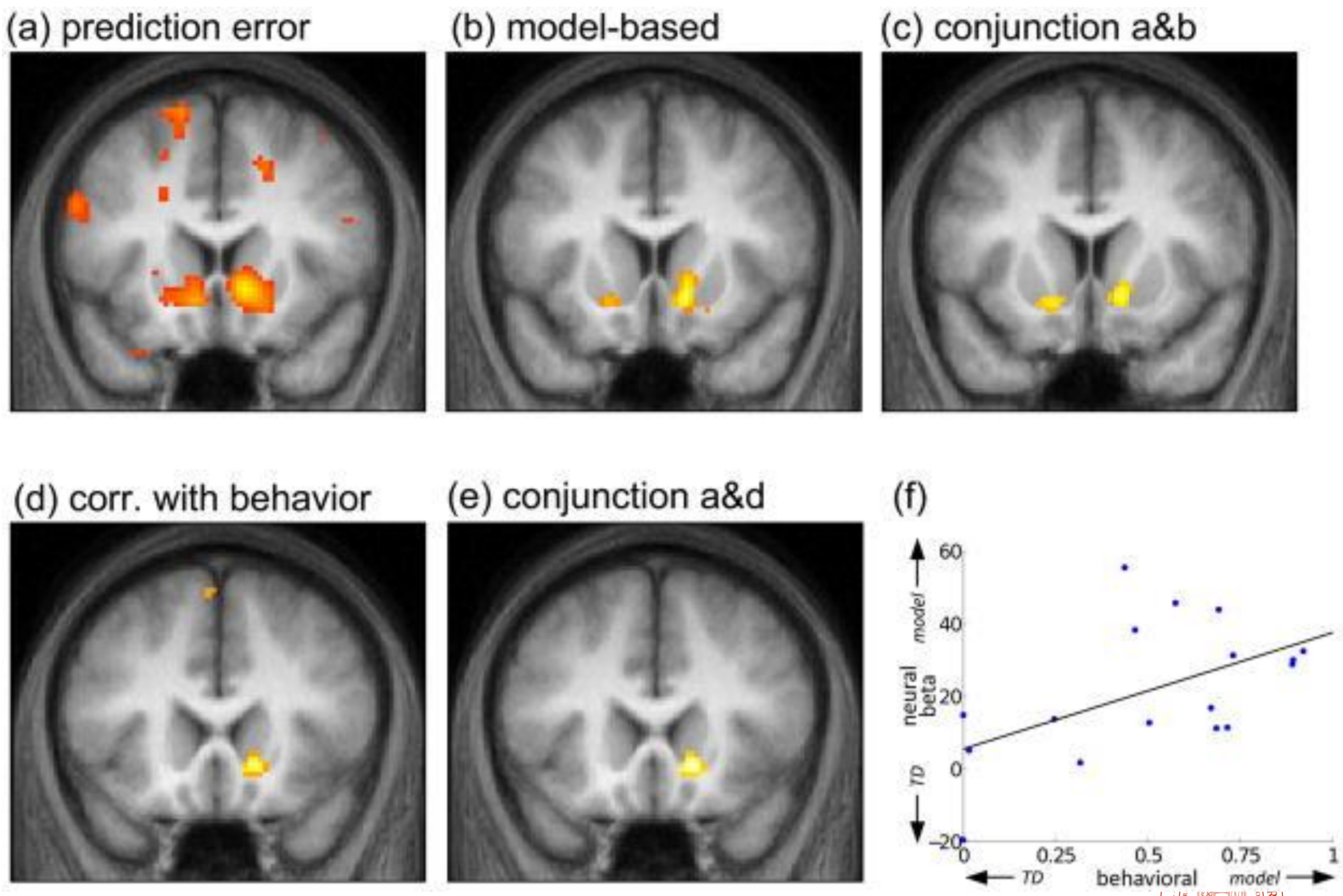
Detecting simultaneous correlates of model-free and model-based systems

Results: Neural

Parameters estimated from computational models were used to find activations that correlated with SPE/model-based & RPE/model-free

Activity in striatum occurred both for model-free and model-based prediction error.

This activity correlated with the extent to which that subject's behavior was model based.

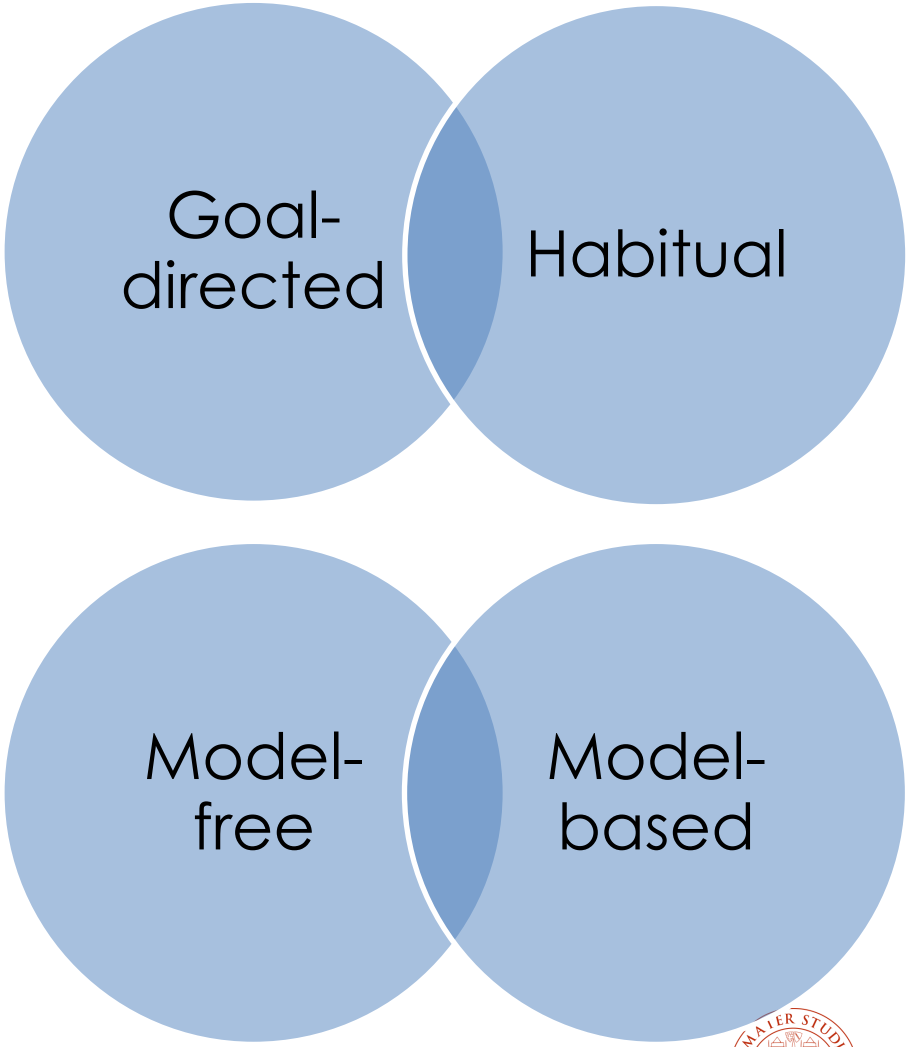


Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. *Neuron.* 2011;69(6):1204-1215. doi:10.1016/j.neuron.2011.02.027

Goal-directed/model-based habitual/model-free behavior are integrated

- **Generation 3 results** challenge the notion of a separate model-based vs model-free learner and suggest a more **integrated computational and neural architecture** for high-level human decision-making

- In the brain, there is a dynamic **inter-dependency** between goal-directed/model-based and habitual/model-free systems, which may **act simultaneously and competitively**



Recommended readings

- Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

- Daw, N. D., & O'Doherty, J. P. (2014). Multiple systems for value learning. In Neuroeconomics (Chapter 21, pp. 393-410). Academic Press.