# Neural Substrates for Conditioning

Bernard Balleine, UCLA

Folk Psychology describes three classes of behavior:

**REFLEX – VOLITION – HABIT**

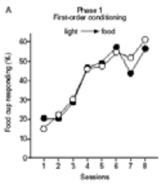
What learning processes contribute to the plasticity of these behavioral responses?

Can one theory of learning explain observations in all three categories of behavior?

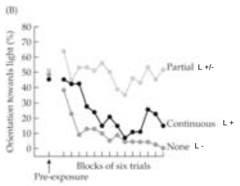
**Reflex** – Volition – Habit

## Conditioned reflexes

## Pavlovian conditioning



*Holland (2000)*



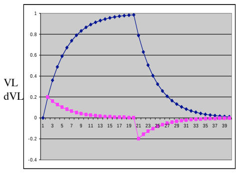
*Kaye & Pearce (1987)*

**US processing**
(e.g. Rescorla -Wagner)

$$\delta V = \alpha\beta(\lambda - \Sigma V)$$

$\alpha$ = stimulus intensity
$\beta$ = US intensity
$\lambda$ = US magnitude

Acquisition: Light - US: $\lambda$=1



For extinction: Light - Ø: $\lambda$ = 0

**CS processing** (associability)
(e.g. Pearce-Hall)

$$\delta V^n = \alpha S \lambda$$

S = CS intensity
$\lambda$ = US magnitude

$$\alpha^n = |\lambda^{n-1} - \Sigma V^{n-1}|$$

For extinction the omission of an expected US leads to the formation of CS-noUS association, $\overline{V}$, where:

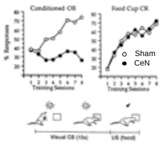
$$\delta \overline{V}_A = S_A \, . \, \alpha_A \, . \, \overline{\lambda}$$

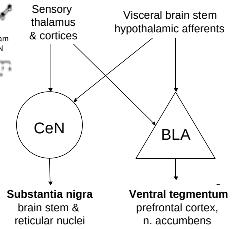
$$\overline{\lambda} = (\Sigma V - \Sigma \overline{V}) - \lambda$$

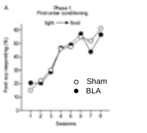
## Differential involvement of amygdala nuclei in US and CS processing

**CeN: CS processing**

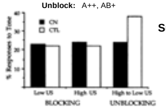
**BLA: US processing**



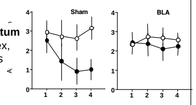




**Low US:** A+, AB+
**High US:** A++, AB++
**Unblock:** A++, AB+



| Pavlovian cond. | Contingency degradation |
|---|---|
| A – **O1**; B – O2 | A – **O1**; B – O2; **O1** |



*Ostlund & Balleine*

*(i)* **A-O1; B-O2** *(Ignoring context)*

*(ii)* **AC-O1; BC-O2; C-**

*(Predictive validity: the learning rule is sensitive to error and so reduces the predictive status of C)*

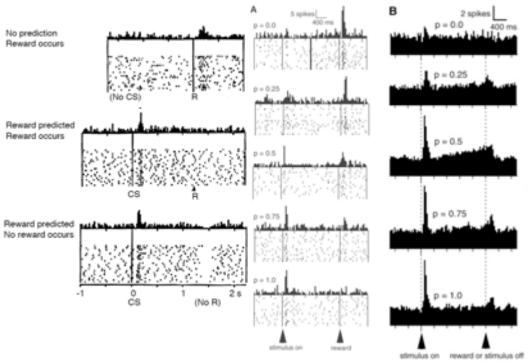
*(iii)* **AC-O1; BC-O2; C-O1**

*(C becomes a better predictor of O1 than A - it provides better information about the occurrence of O1*
*- sometimes called context blocking)*

**Evidence of both US and CS processing in midbrain dopamine neurons (Schultz)**

Probability of the US given the CS = 1

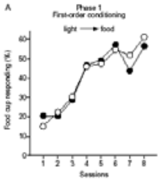
Probability of the US given the CS varied between 0 and 1



**But aren't Pavlovian conditioned responses in some sense goal-directed??**

Why do rats approach the location of food delivery during a stimulus paired with food?

*Perhaps rats behave in this way because they believe that these responses are required to get food.*



*Holland (2000)*

**Prediction:**

Rats should not acquire approach responses if the **response-outcome contingency** is reversed such that approach responding cancels access to food …

**$CR_{approach}$ → no food**

## Anticipatory approach

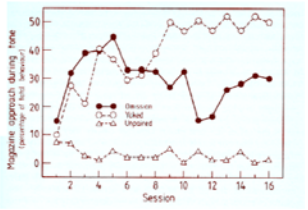
**Holland (1979):**

TONE (10-s) is paired with FOOD (pellets)

**Gp Omission:** The tone is paired with food EXCEPT if the rats approach the food source during the tone; then they get no food!
**Gp Yoked:** This group gets tone paired with food when the omission group does
**Gp Unpaired:** The tone is never paired with food delivery

Notice that rats in the omission condition are never given the information that approach is either NECESSARY OR SUFFICIENT to get food

Nevertheless, they acquire and maintain the approach response even though by doing so they lose over 30% of the available food …



**Reflex – Volition: Goal-directed action**

- From human action theory:

  **belief** (knowledge): 'action A → reward X'

  **desire** (reward): X

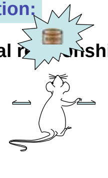
- Any action must satisfy 2 criteria to be called goal-directed:

  **Contingency criterion:** it must be sensitive to changes in the causal relation between action and outcome

  **Goal criterion:** it must be sensitive to changes in the value of the the goal

## Goal-directed action:

### 1. Action - goal r nship



In each second of the session:

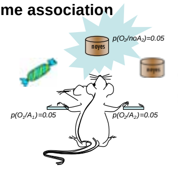
EARNED: p(**O1**/R1)=0.05
p(O2/R2)=0.05
FREE: p(**O1**/noR1 or noR2)=0.05

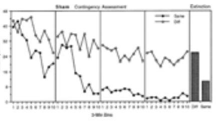
$\mathbf{V}_{\mathbf{R1\text{-}O1}}$ = p(**O1**/R1) - p(**O1**/noR1) which in this case = 0

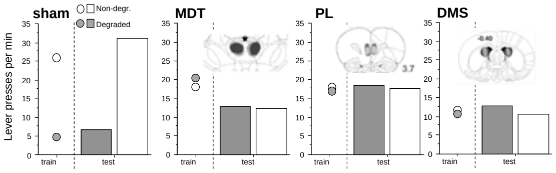
Or: $\Delta P = P(e+|c+) - P(e+|c-)$

## Neural bases of goal-directed action:

### 1. Action - outcome association

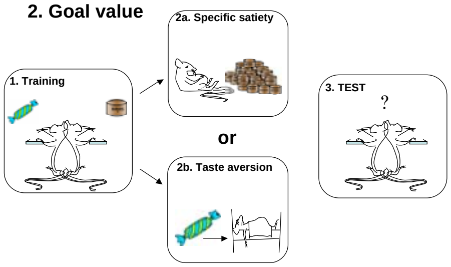






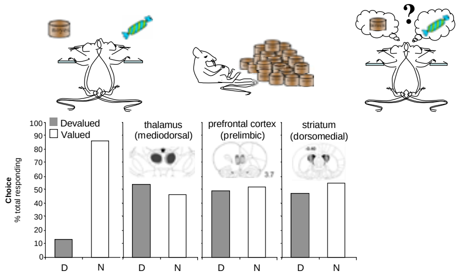
## Goal-directed action:

1. Action - goal relationship
2. Goal value



## Goal-directed action:

1. Action - goal relationship
2. Goal value

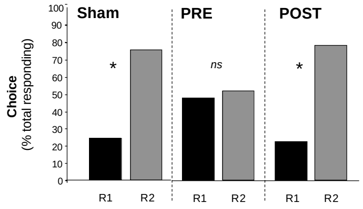


### Structures where lesions affect BOTH sensitivity to changes in action-outcome contingency and outcome devaluation:

| | AREA | conting. | deval. | Reference |
|---|---|---|---|---|
| * | PL | X | X | *Balleine & Dickinson, 1998, Corbit & Balleine, 2003, Ostlund & Balleine, 2005* |
| | OFC | X | - | *Ostlund & Balleine, in press* |
| * | DMS | X | X | *Yin, Ostlund, Knowlton & Balleine, 2005* |
| | DLS | - | - | *Yin, Knowlton & Balleine, 2004; 2006* |
| * | MDT | X | X | *Corbit, Muir & Balleine, 2003* |
| | ANT | - | - | *Corbit, Muir & Balleine, 2003* |
| | NACco | - | X | *Corbit, Muir & Balleine, 2001; Corbit & Balleine, in prep* |
| | NACsh | X | - | *Corbit, Muir & Balleine, 2001; Corbit & Balleine, in prep* |
| | HPC | - | - | *Corbit & Balleine, 2000; Corbit, Ostlund & Balleine, 2002* |
| | EC | X | - | *Corbit, Ostlund & Balleine, 2002* |

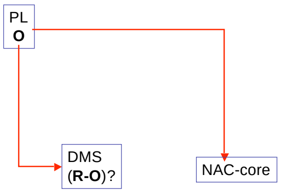
**PREDICTION: Post-training lesions of prelimbic cortex should be just as effective as pretraining lesions:**

**X** pre R1 → O1;R2 → O2 | **X** post 1 hr O1 then: R1 vs. R2

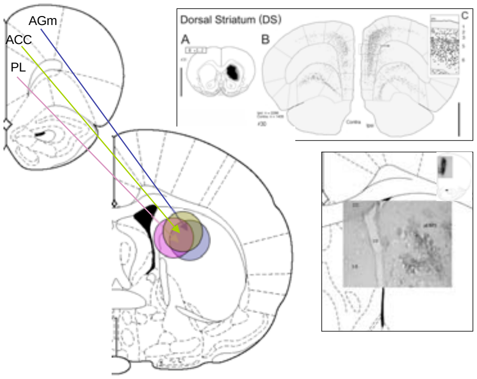


*(Sean Ostlund)*

## Two primary striatal efferents of prelimbic cortex:

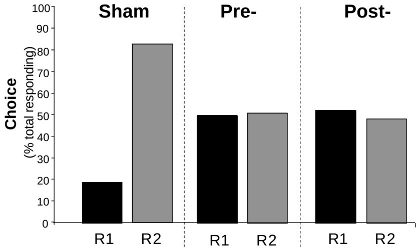


1. **Ventral striatum** - long regarded as a limbic-motor interface. BUT more closely related to factors affecting <u>performance rather than learning</u>
2. **Dorsomedial (associative) striatum:**



## Pre- vs. post-training lesions of the pDMS

X pre | R1 → O1; R2 → O2 | X post | 1 hr O1 then: R1 vs. R2



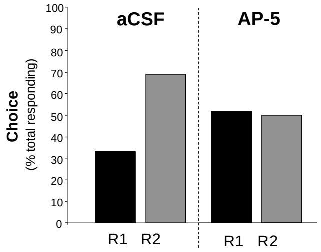
*(Henry Yin)*

## PREDICTION

Plasticity in dorso-medial striatum reportedly involves NMDA receptor-mediated long-term potentiation

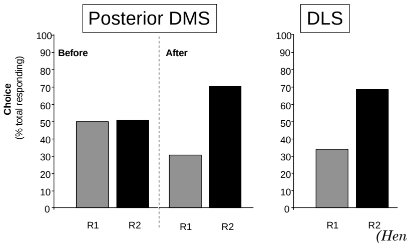
Infusion of AP-5, an NMDA antagonist, into the dorso-medial area during learning should be predicted to block the formation of the action-outcome association

| Pre-train | Training–1 session | Devaluation and test |
|---|---|---|
| R1 -> Oc<br>R2 -> Oc | **AP5 (0.5 µg/0.5 µL)**<br>R1 -> O1<br>R2 -> O2 | O1!: R1 vs. R2<br>(1 hr) |
| R1 -> Oc<br>R2 -> Oc | **aCSF (0.5 µL)**<br>R1 -> O1<br>R2 -> O2 | O1!: R1 vs. R2<br>(1 hr) |

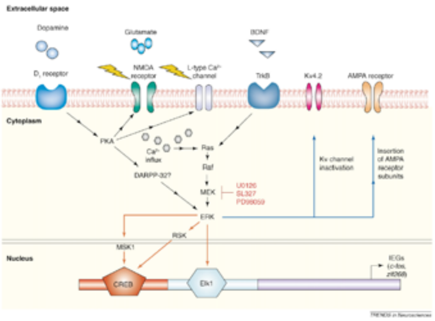


*(Henry Yin)*

| Pre-train | Training–1 session | Devaluation and test |
|---|---|---|
| R1 -> Oc<br>R2 -> Oc | **AP5 5-min before**<br>R1 -> O1<br>R2 -> O2 | O1!: R1 vs. R2<br>(1 hr) |
| R1 -> Oc<br>R2 -> Oc | **AP5 6-hr before**<br>R1 -> O1<br>R2 -> O2 | O1!: R1 vs. R2<br>(1 hr) |

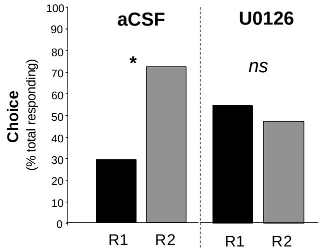


*(Henry Yin)*



*Lu et al, (2006) TINS*

| Pre-train | Training–1 session | Devaluation and test |
|---|---|---|
| **R1 -> Oc**<br>**R2 -> Oc** | **U0126 (1.0 µg/0.5 µL)**<br>**R1 -> O1**<br>**R2 -> O2** | **O1!: R1 vs. R2**<br>**(1 hr)** |
| **R1 -> Oc**<br>**R2 -> Oc** | **aCSF (0.5 µL)**<br>**R1 -> O1**<br>**R2 -> O2** | **O1!: R1 vs. R2**<br>**(1 hr)** |

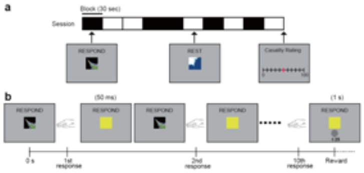


*(Mike Shiflett)*

**Instrumental conditioning, causality judgment and fMRI in humans**

with Saori Tanaka and John O'Doherty, *Caltech*

## Task paradigm



**RESPOND state**

- Subject can press button at any time
- Reinforcer = 25 ¢
- Low RR: after scheduled ratio
- Low VI: after scheduled interval
- Response cost = -1 ¢

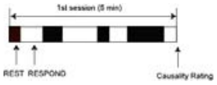
- Different figures indicate different schedules
- No instruction for the schedules

**REST state**

- No responses

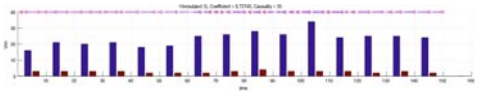
## Experimental Protocol

- RR with RI matched to the interval to reinforcer
  - RR-10
  - RI-(matched inter-reinforcer interval to RR-10)
- RI vs. RR matched to the response per reinforcer
  - RI-4
  - RR-(matched response rates to RI-4)
- Yoked-order design within subjects
- Randomized order of schedules
- Five "RESPOND" blocks and five "REST" blocks with pseudo-random order



## Correlation Coefficient

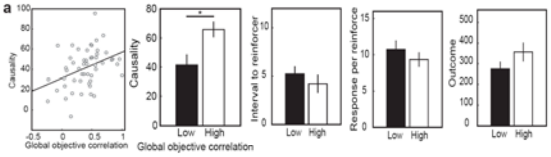
- Contingency between response rate and reinforcement rate during constant time bin may affect subject's causality
- Define correlation between response number per 10 sec bin and reinforce number per 10 sec bin as a measurement of contingency



Histogram of response and reinforce number per 10 sec bin
(blue: response, red: reinforcer)

## Behavioral Results

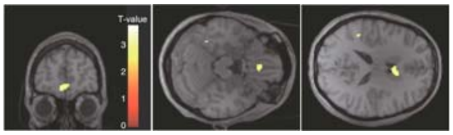
We calculated the correlation between response rate and outcome rate over 10 sec time bins during the various RR and RI components and examined various behavioral measures based on the component with the highest and with the lowest correlation coefficient for each subject.



## fMRI results

- High – Low correlation coefficient

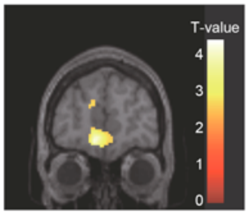
**Medial prefrontal cortex** **Medial orbital** **Caudate**



## fMRI results (cont)

High causality – Low causality

medial prefrontal cortex



Reflex – **Volition – Habit**

“A strictly voluntary act has to be guided by idea, perception, and volition, throughout its whole course. In an habitual action, mere sensation is a sufficient guide.”

*William James, 1890*

## Testing for habits

**Training conditions:**

Placing constraints on the rate of reward (e.g. by overtraining or the use of interval schedules) can cause the performance of goal-directed actions to become relatively inflexible; i.e. habitual:

- insensitive to changes in action-outcome contingency

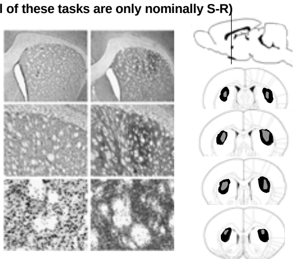
- insensitive to changes in goal value

## Testing for habits

**Lesions of dorsolateral striatum (DLS):**

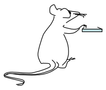
- McDonald & White (1993) win-stay
- Packard & McGaugh (1996) - place vs response
- Jog et al (1999); Barnes et al (2004) T-maze

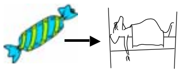
**(all of these tasks are only nominally S-R)**

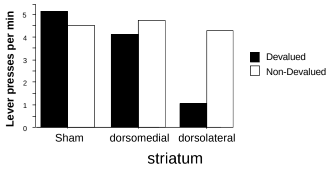


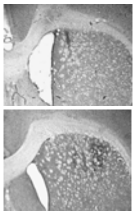
## Testing for habits

**Yin et al, 2004**









*(Henry Yin)*

**Habits are insensitive to changes in action-outcome contingency** (they are not governed by the same learning rule as goal-directed actions)

*Furthermore: inactivation of DLS increases sensitivity to omission*

**Phase 1: acquisition**

650 reinforced actions:
Lever press -> sucrose

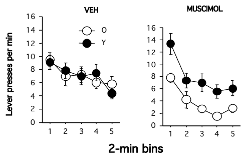
**Phase 2: omission**

Sucrose is delivered every 10s

Gp O: The next sucrose delivery is cancelled by lever pressing

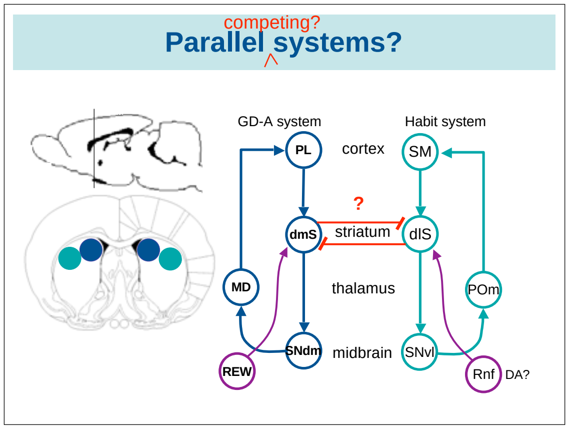
Gp Y: yoked to Gp OM and get same amount and pattern of suc delivered

Extinction test - off drug



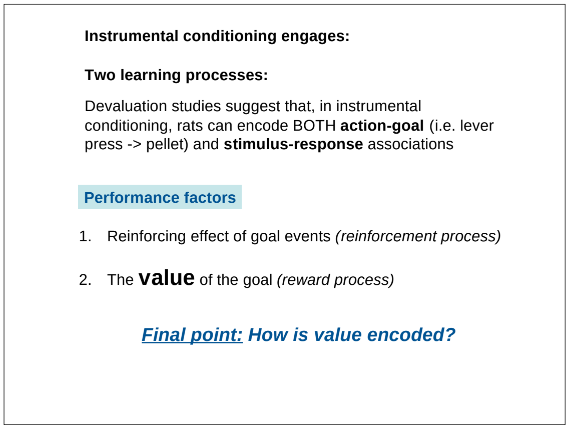
*(Yin et al, 2005)*

**Rats in Group O have to STOP lever pressing to get reward**



competing?

# Parallel systems?



**Instrumental conditioning engages:**

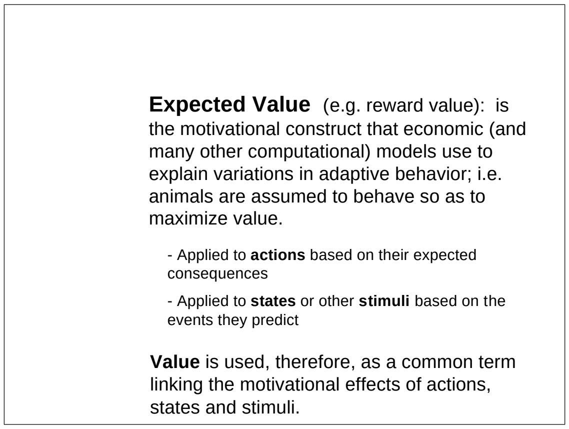
**Two learning processes:**

Devaluation studies suggest that, in instrumental conditioning, rats can encode BOTH **action-goal** (i.e. lever press -> pellet) and **stimulus-response** associations

**Performance factors**

1. Reinforcing effect of goal events *(reinforcement process)*
2. The **value** of the goal *(reward process)*

***Final point:* How is value encoded?**

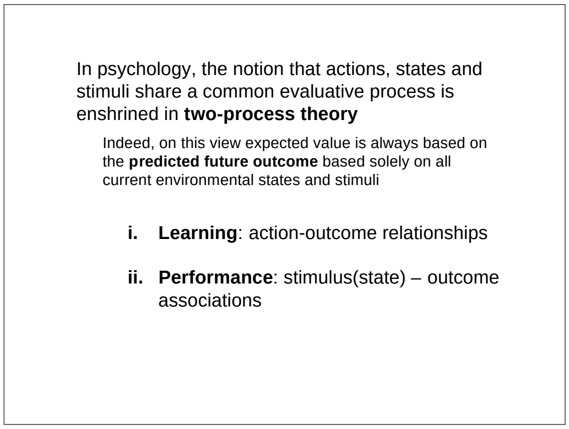


**Expected Value** (e.g. reward value): is the motivational construct that economic (and many other computational) models use to explain variations in adaptive behavior; i.e. animals are assumed to behave so as to maximize value.

- Applied to **actions** based on their expected consequences

- Applied to **states** or other **stimuli** based on the events they predict

**Value** is used, therefore, as a common term linking the motivational effects of actions, states and stimuli.

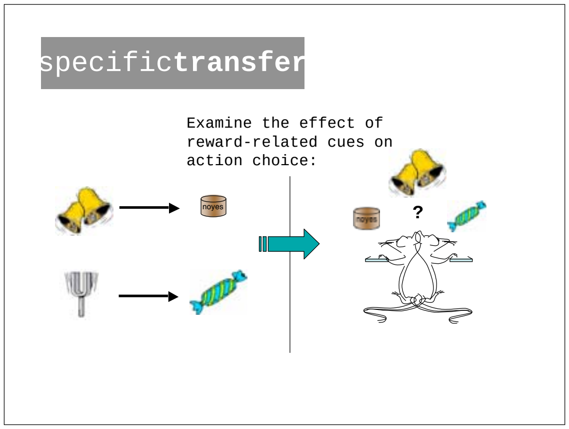


In psychology, the notion that actions, states and stimuli share a common evaluative process is enshrined in **two-process theory**

Indeed, on this view expected value is always based on the **predicted future outcome** based solely on all current environmental states and stimuli

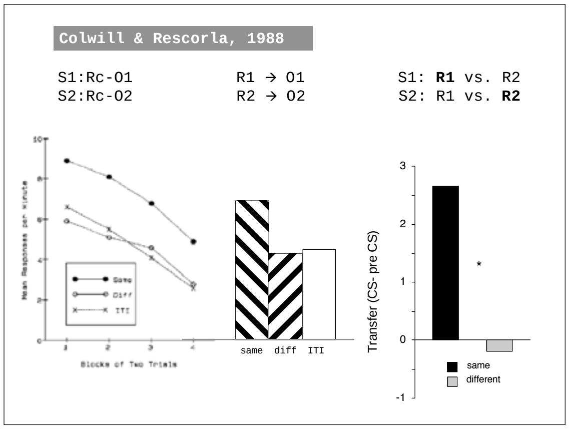
i. **Learning**: action-outcome relationships

ii. **Performance**: stimulus(state) – outcome associations



## specific**transfer**

Examine the effect of reward-related cues on action choice:



## Colwill & Rescorla, 1988

```
S1:Rc-O1          R1 → O1          S1: R1 vs. R2
S2:Rc-O2          R2 → O2          S2: R1 vs. R2
```

**Two process theory or theories?**

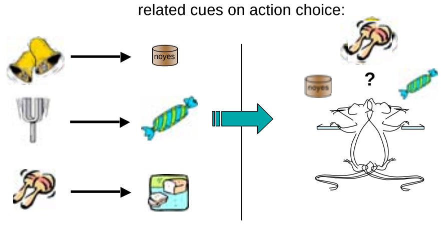
**There are two, two-process accounts**

**Motivational**: states/stimuli affect actions by causing changes in arousal or activation and so affect the relative ***vigor*** of responding

**Informational**: states/stimuli affect actions by retrieving or priming specific consequences in memory to influence action selection

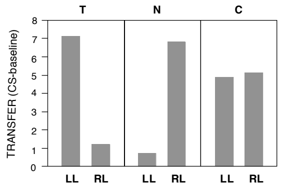
## **Transfer** motivation vs. information

Examine the effect of reward-related cues on action choice:

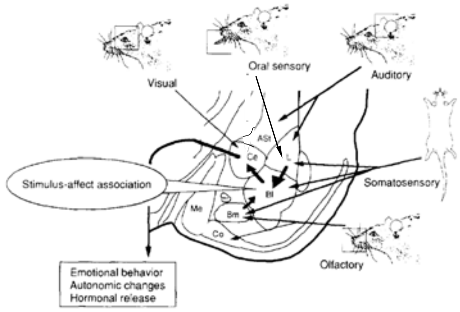


## Specific vs. general transfer: within-subjects assessment

| Pavlovian cond. | Instrumental cond. | Transfer test |
|---|---|---|
| T - pel; N - suc | LL → pel; RL → suc | T: **LL** vs. RL<br>N: LL vs. **RL**<br>**C: LL vs. RL** |
| **C-starch** | | |



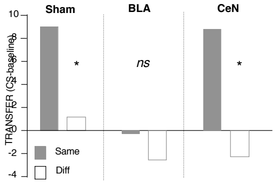
## **Amygdala:** sensory afferents



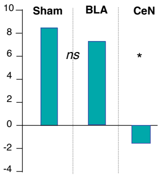
## **Specific vs. general transfer:** BLA and CeN lesions

| Pavlovian cond. | Instrumental cond. | Transfer test |
|---|---|---|
| T - pel; N - suc | LL → pel; RL → suc | T: **LL** vs. RL<br>N: LL vs. **RL**<br>**C: LL vs. RL** |
| **C-starch** | | |

**Specific transfer (T & N)**

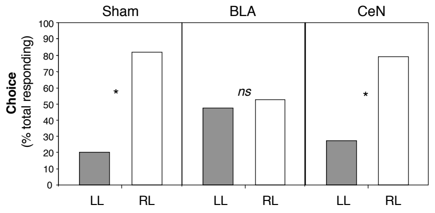


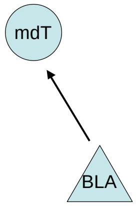
**General transfer (C)**



## BLA vs. CeN: Outcome devaluation

| Left lever → pellets;<br>Right lever → sucrose | 1 hr: pellets then: LL vs. RL |
|---|---|





## NMDA-induced lesions of mediodorsal thalamus:

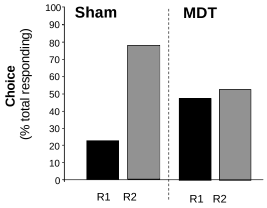
**Outcome devaluation**

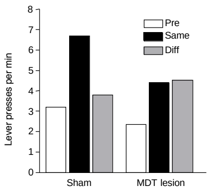
R1 → O1; R2 → O2

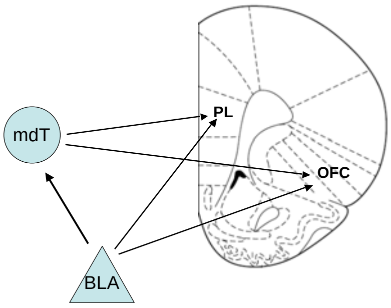
1 hr O1 then: R1 vs. R2

**Specific transfer**

| **Pavlovian cond** | **Instrumental cond** | **Transfer test** |
|---|---|---|
| T - pel; N - suc | LL → pel; RL → suc | T: **LL** vs. RL<br>N: LL vs. **RL** |

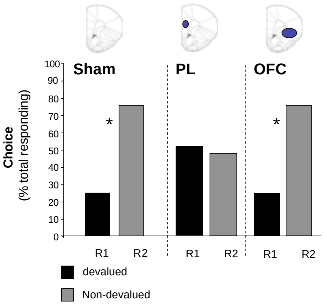






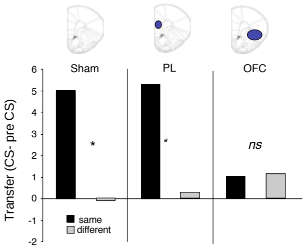
## NMDA-induced lesions of regions of prefrontal cortex

Lesions of orbitofrontal cortices have no effect on <u>outcome devaluation</u> **but PL lesions do:**



## NMDA-induced lesions of prelimbic cortex do not affect transfer BUT OFC lesions do!!

| **Pavlovian cond.** | **Instrumental cond.** | **Transfer test** |
|---|---|---|
| T - pel; N - suc | LL → pel; RL → suc | T: **LL** vs. RL<br>N: LL vs. **RL** |



This reflects a **double dissociation** within the prefrontal cortex in the influence of expected value based on prospective actions and expected value derived from environmental stimuli on action selection

| | PL lesion | OFC lesion |
|---|---|---|
| Outcome Devaluation (value of actions) | **x** | YES |
| Transfer (value of stimuli) | YES | **x** |

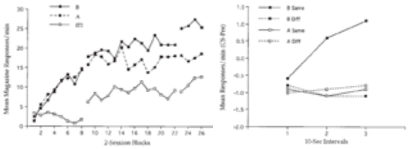
How are these distinct cortical values reconciled in action selection??

**Specific transfer** reflects the way information (such as advertising) can alter action selection

**Specific transfer** is sensitive to the predictive status of the CS: Indeed, degrading the predictive or causal status of the CS abolishes this transfer effect.

**Delamater, 1995**

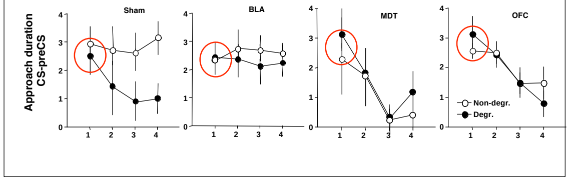
| Instrumental cond. | Pavlovian cond. | Contingency degradation | Specific transfer test |
|---|---|---|---|
| R1 – **O1**; R2 – O2 | A – **O1**; B – O2 | A – O1; B – O2; **O1** | A: **R1**, R2<br>B: R1, **R2** |



## Pavlovian contingency degradation

| Pavlovian cond. | Contingency degradation |
|---|---|
| S1 – **O1**; S2 – O2 | S1 – **O1**; S2 – O2; **O1** |

Each of the structures that abolish transfer also attenuate the rats' ability to calculate the relative validity of predictive cues.



There are two aspects of expected value that emerge from an assessment of the role of prefrontal cortex in action selection

The value of the expected outcome of an action *(strictly what is meant by reward value)*

The influence of the information provided by predictive cues.

**Action values** and **the 'value' of the information** provided by states and stimuli appear to be distinct forms of value.

## Reflex – Volition – Habit

So there appear to be three distinct forms of learning

Predictive learning
Goal-directed learning
Habit learning

and three distinct motivational processes that accompany each of these forms of learning and that independently influence decision processes

The value of predictive information
The reward value of goals
Reinforcing function of biologically significant events