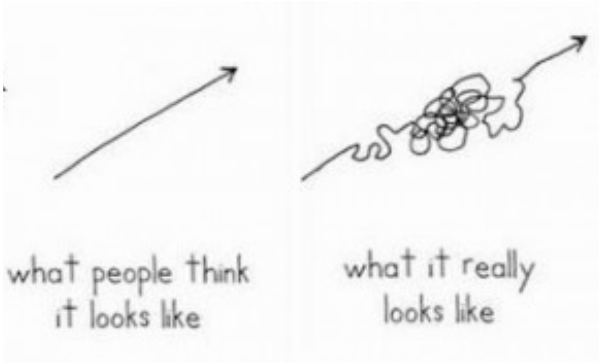
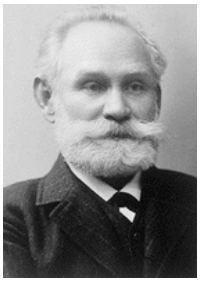
Classical Conditioning III: Temporal difference learning

PSY/NEU338: Animal learning and decision making:
Psychological, computational and neural perspectives

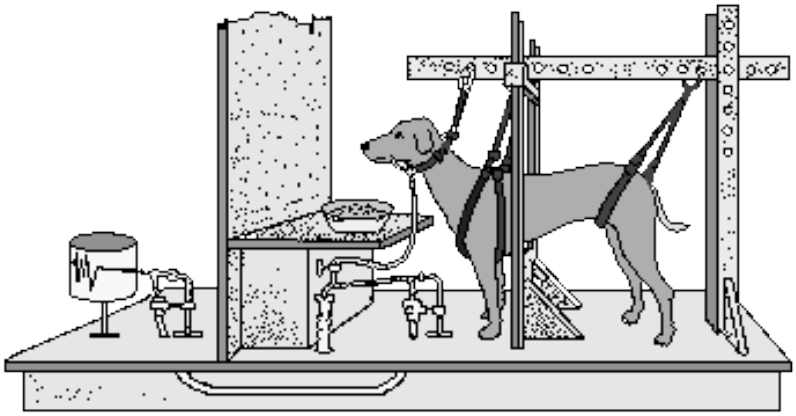
understanding a new concept

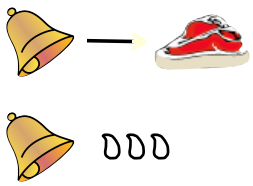


recap: animals learn predictions



Ivan Pavlov





= Unconditional Stimulus

= Conditional Stimulus

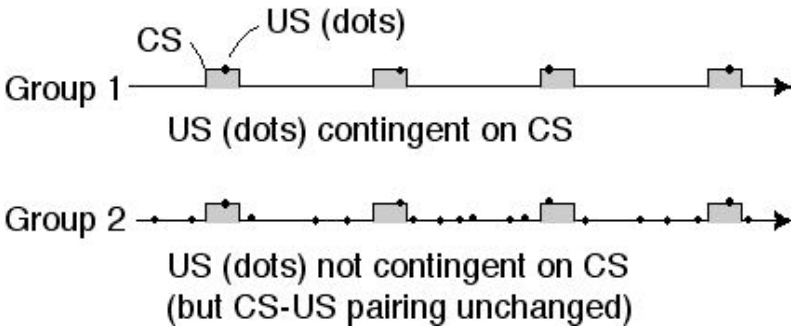
= Unconditional Response (reflex);
Conditional Response (reflex)

what makes conditioning Pavlovian?

procedurally: Pavlovian/classical conditioning is a learning situation in which the reinforcer *does not depend* on the animal's response

from the animal's point of view: the conditioned response is *unaviodable*, like a *reflex*, not utilitarian or flexible; direct result of a prediction

But... 1) Rescorla's control condition



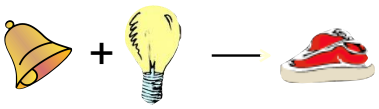
temporal contiguity is not enough - need contingency

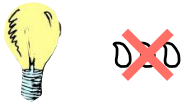
P(food | light) > P(food | no light)

But... 2) Kamin's blocking

Phase I

Phase II





contingency is also not enough.. need surprise

P(food | noise+light) ≠ P(food | noise alone)

Rescorla & Wagner (1972)

The idea: error-driven learning

Change in value is proportional to the difference between actual and predicted outcome

$$\Delta V(CS_i) = \eta[R_{US} - \sum_{j \in \text{trial}} V(CS_j)]$$

Two assumptions/hypotheses:

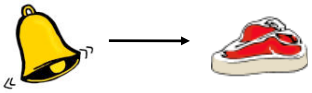
(1) learning is driven by error (formalize notion of surprise)

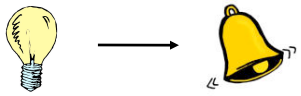
(2) summations of predictors is linear

what does the theory explain?

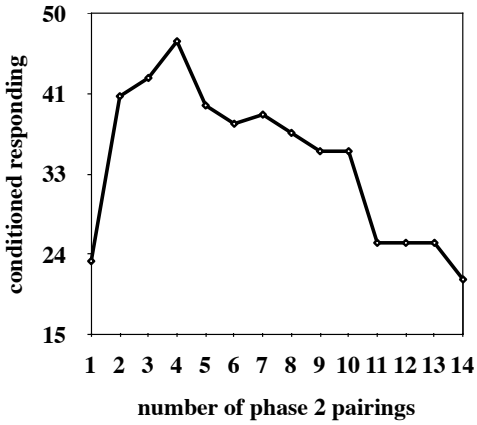
	R-W	
acquisition	√	
extinction	√̸	
blocking	√	
overshadowing	√	
temporal relationships	X	
overexpectation	√	

but: second-order conditioning

phase 1: 

phase 2: 

test:  ?



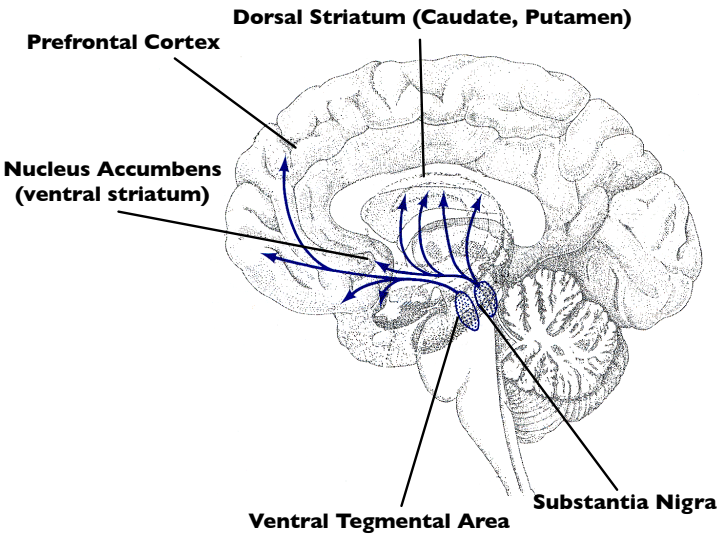
what do you think will happen?
what would Rescorla-Wagner learning predict here?

animals learn that a predictor of a predictor is also a predictor!
⇒ not interested solely in predicting immediate reinforcement..

Challenge:

- Can you modify the R-W learning rule to account for second order conditioning?
- Group work:
  - what is the fundamental problem here?
  - ideas how to solve it?

hint: dopamine



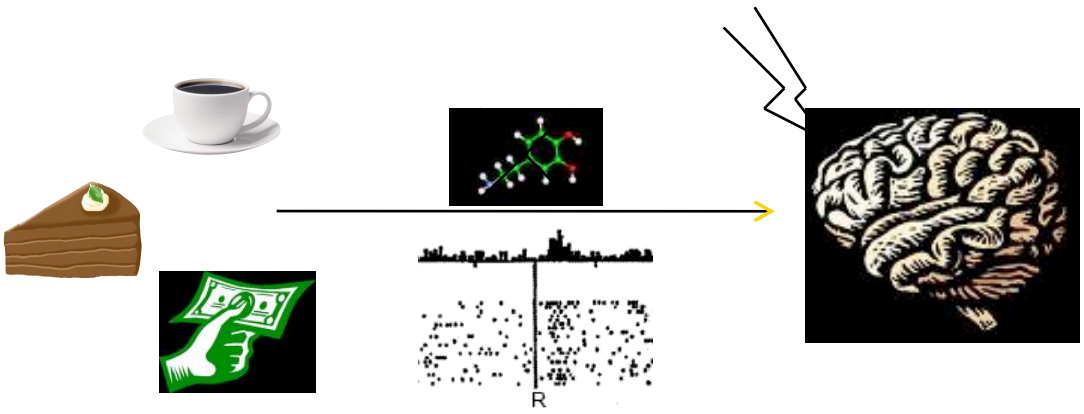
Parkinson's Disease
→ Motor control / initiation?

Drug addiction, gambling,
Natural rewards
→ Reward pathway?
→ Learning?

Also involved in:
- Working memory
- Novel situations
- ADHD
- Schizophrenia
- …

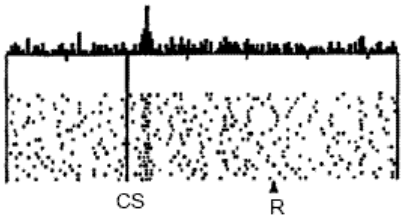
the anhedonia hypothesis (Wise, '80s)

- Anhedonia = inability to experience positive emotional states derived from obtaining a desired or biologically significant stimulus
- Neuroleptics (dopamine antagonists) cause anhedonia
- Dopamine areas are a target of intra-cranial self stimulation (ICSS)
- Dopamine is important for reward-mediated conditioning

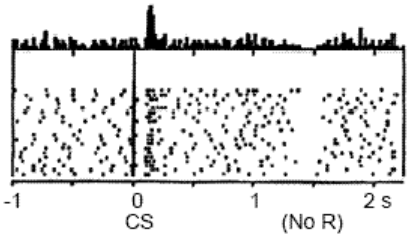


but...

predictable reward



omitted reward



(Schultz *et al.* '90s)

- behavioral puzzle: second order conditioning
- neural puzzle: dopamine responds to reward predicting stimuli instead of to rewards
- think about it…

understanding the brain: where do we start?!

David Marr (1945-1980) proposed three levels of analysis:

1. the problem (Computational Level)
2. the strategy (Algorithmic Level)
3. how its actually done by networks of neurons (Implementational Level)

lets start over, this time from the top...

The problem: optimal prediction of future reinforcement

$$V_t = E\left[\sum_{i=t+1}^{\infty} r_i\right]$$

want to predict expected sum of future reinforcement

$$V_t = E\left[\sum_{i=t+1}^{\infty} \gamma^{i-t-1} r_i\right]$$

want to predict expected sum of *discounted* future reinf. ($0<\gamma<1$)

$$V_t = E\left[\sum_{i=t+1}^{T} r_i\right]$$

want to predict expected sum of future reinforcement in a trial/episode

lets start over, this time from the top...

The problem: optimal prediction of future reinforcement

$$\begin{aligned} V_t &= E\left[r_{t+1} + r_{t+2} + \ldots + r_T\right] \\ &= E\left[r_{t+1}\right] + E\left[r_{t+2} + \ldots + r_T\right] \\ &= E\left[r_{t+1}\right] + V_{t+1} \end{aligned}$$

(note: *t* indexes time within a trial)

$$V_t = E\left[\sum_{i=t+1}^{T} r_i\right]$$

want to predict expected sum of future reinforcement in a trial/episode

lets start over, this time from the top...

The problem: optimal prediction of future reinforcement

$$\begin{aligned} V_t &= E\left[r_{t+1} + r_{t+2} + \ldots + r_T\right] \\ &= E\left[r_{t+1}\right] + E\left[r_{t+2} + \ldots + r_T\right] \\ &= E\left[r_{t+1}\right] + V_{t+1} \end{aligned}$$

(note: *t* indexes time within a trial)

Think football…
What would be a sensible learning rule here?
How is this different from Rescorla-Wagner?