Reinforcement and aversive Learning
Behavior, The basal-ganglia, The amygdala

Rony Paz

Reinforcement learning

Supervised learning –

all knowing teacher, detailed feedback

Reinforcement learning –

scalar (correct/incorrect) feedback

Unsupervised learning –

self organization

The law of effect

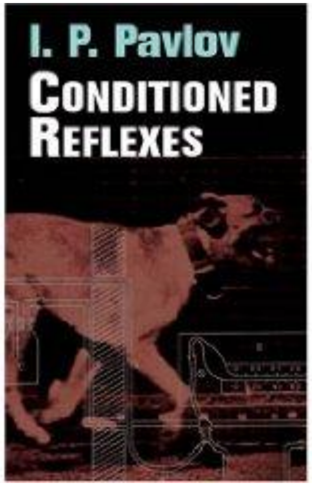
"The Law of Effect is that: Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur"

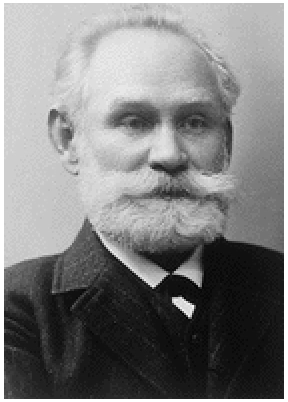
Edward Lee Thorndike (1911)

Classical conditioning (Pavlov, 1927)

The Elements:

- **US**: Unconditioned stimulus
- **UR**: Unconditioned response
- **NS**: Neutral stimulus
- **CS**: Conditioned stimulus
- **CR**: Conditioned response





And in Neurons

Donald Hebb

- Hebbian plasticity (1949):

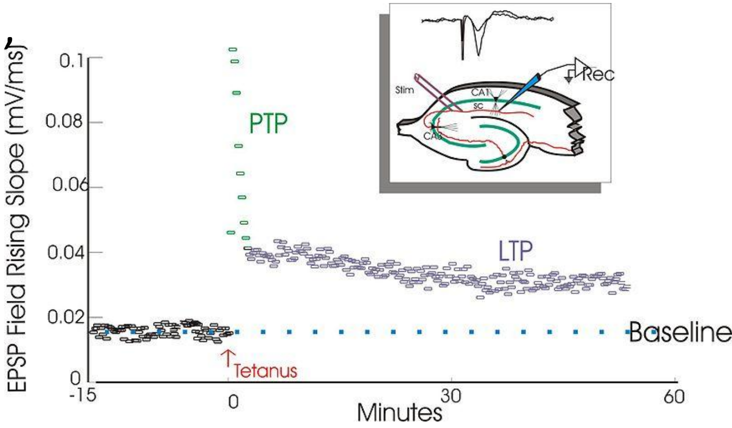
The general idea is an old one, that any two cells or systems of cells that are repeatedly active at the same time will tend to become 'associated', so that activity in one facilitates activity in the other

When one cell repeatedly assists in firing another, the axon of the first cell develops synaptic knobs (or enlarges them if they already exist) in contact with the soma of the second cell

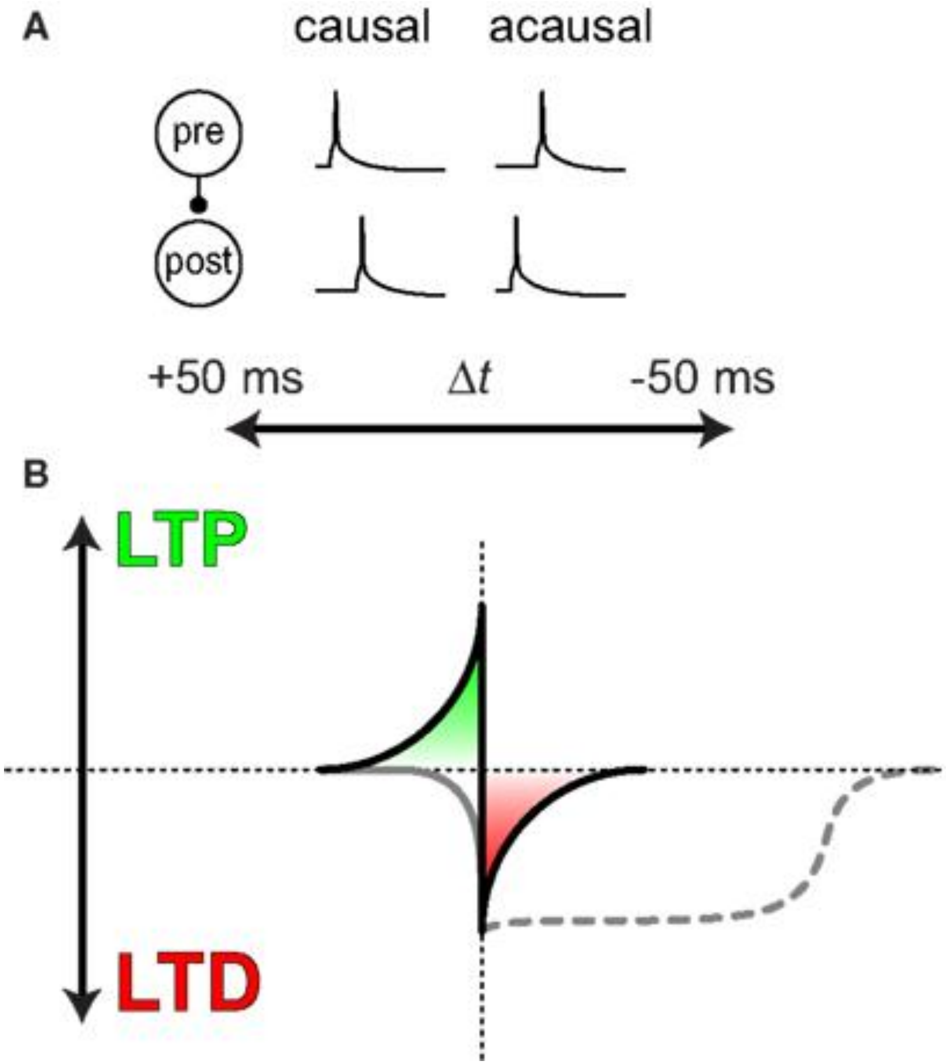
Long-term-potentiation (LTP)

- Lomo, Bliss, Andersen, 1966, Hippocampus.
- Induced artificially by tetanic stimulation
- Long-lasting enhancement in signal transmission between two neurons that results from stimulating them synchronously.
- Increase in synaptic strength
- A cellular mechanism for learning and memory.
- Requires protein synthesis

- **Hebbian LTP** requires simultaneous pre- and postsynaptic depolarization for its induction (“fire together – wire together”)
  - Specificity: to synapse
  - Associativity: when a ‘weak’ pathway is not enough, simultaneous strong input will associate both
  - Coopertaivity: weak stimulation of many that converge



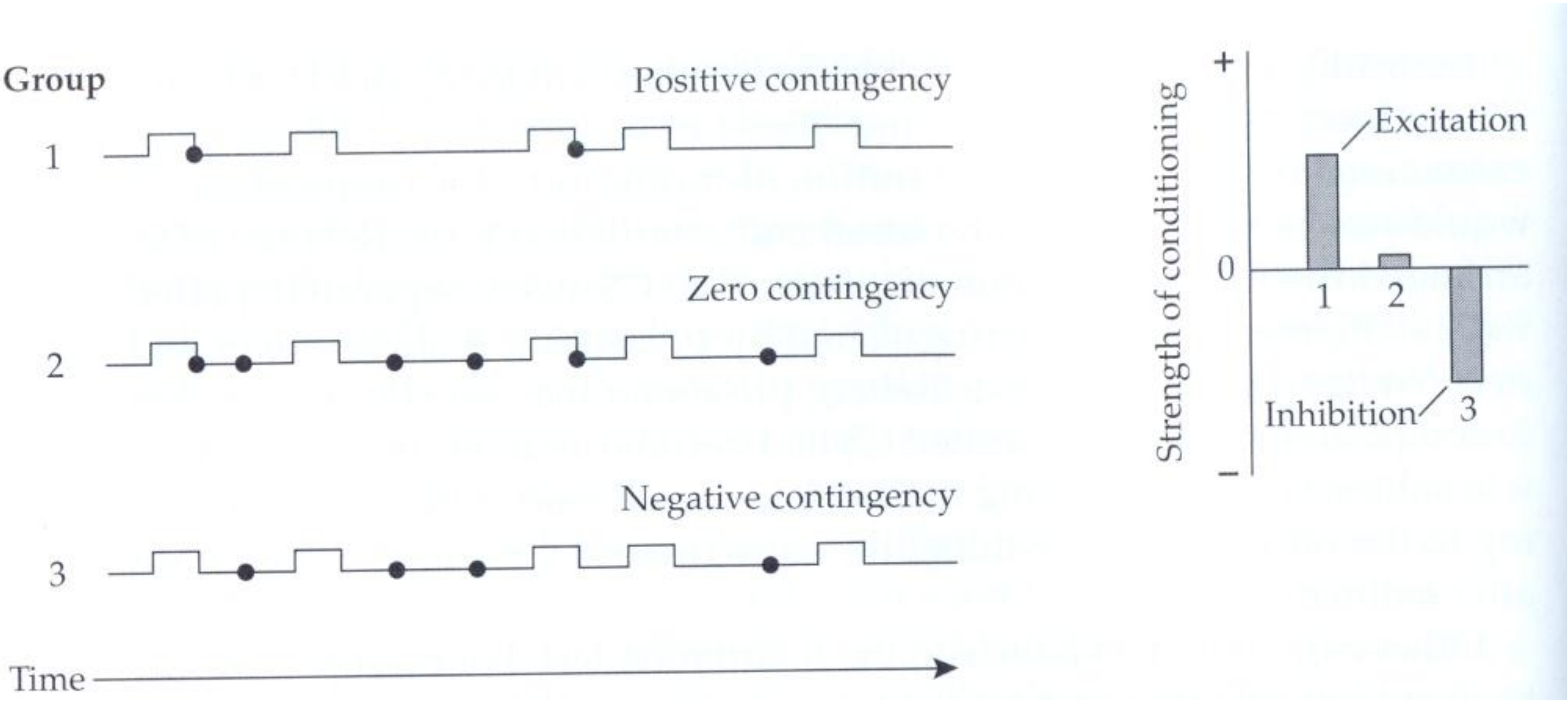
Spike-timing-dependent-plasticity (STDP)



So far…

- Coincidence (co-occurrence) model can explain learning and associations.

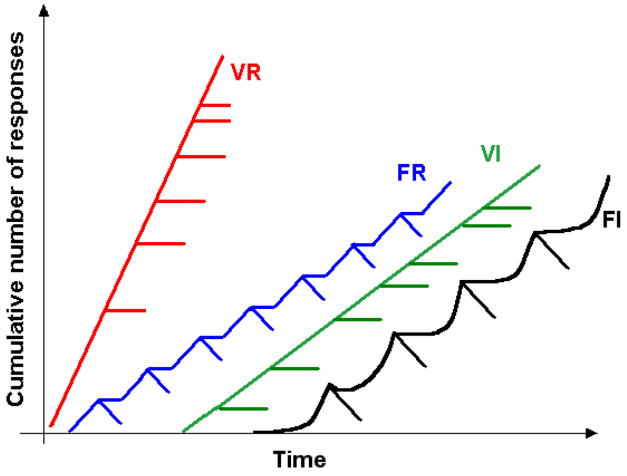
More than just co-occurrence: Reliable prediction - contingency



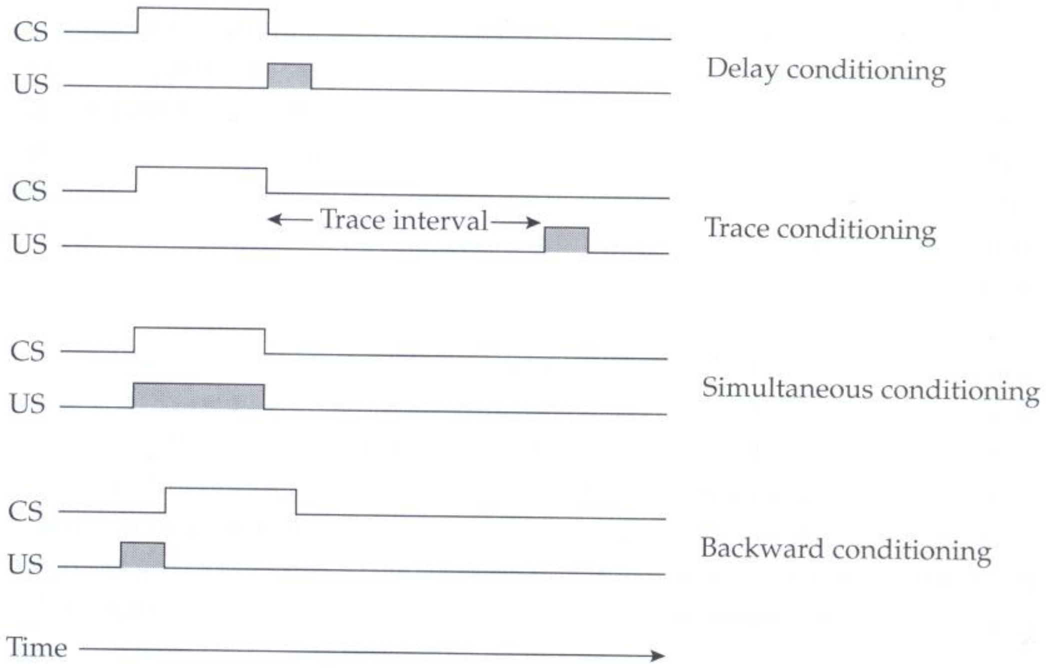
Schedules of reinforcement

Lead to different kinds of learning rates:

- Fixed interval
- Variable interval
- Fixed ratio
- Variable ratio



Conditioning strength

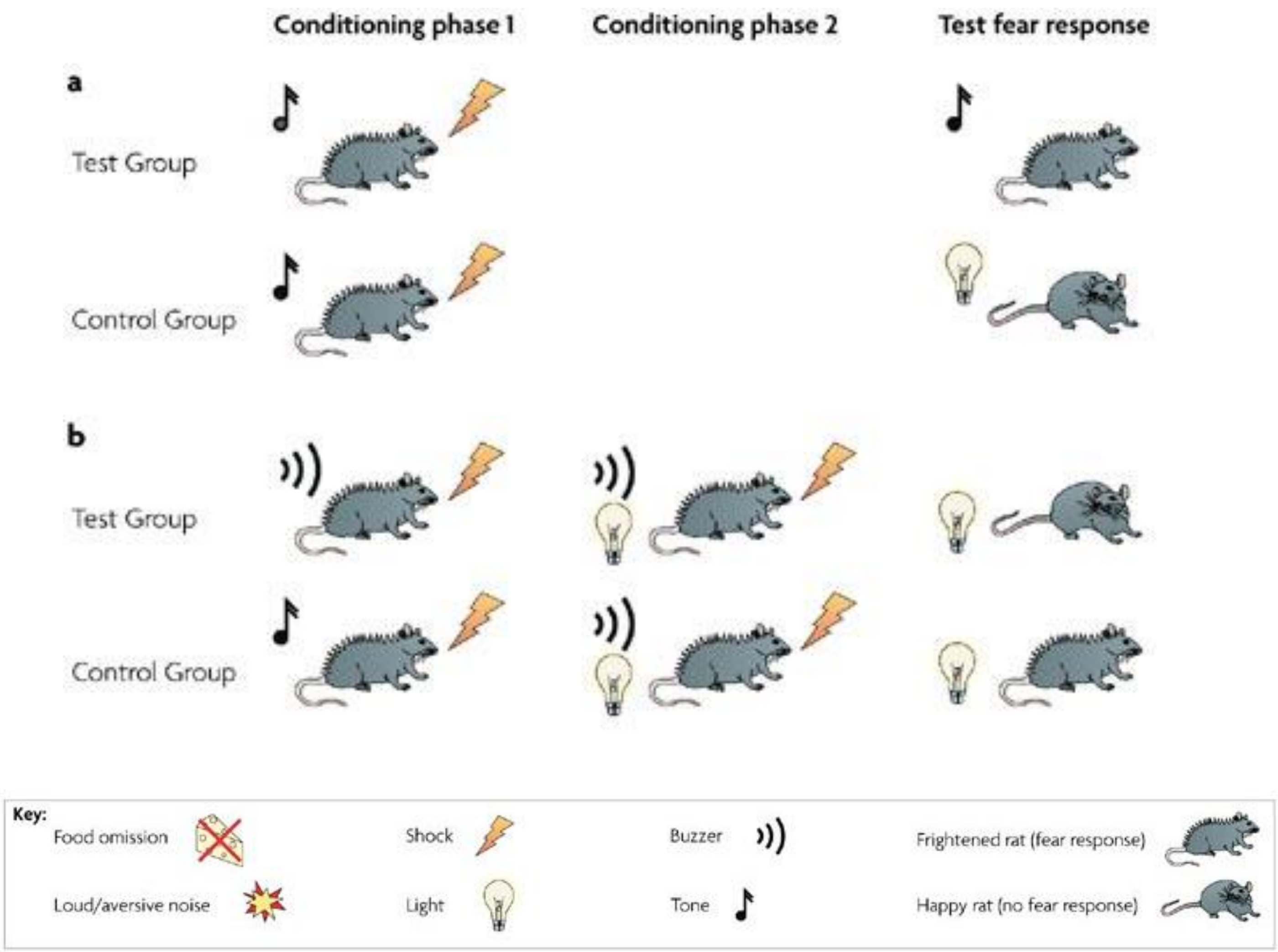


- Backward < simultaneous < trace < delay
- In trace: short interval > long interval
- In delay: short CS > long CS
- Salience of the CS
- Strength of the US
- Spaced trials is better than massed trials (the ratio between inter-trial-interval and the CS)

Properties of classical conditioning

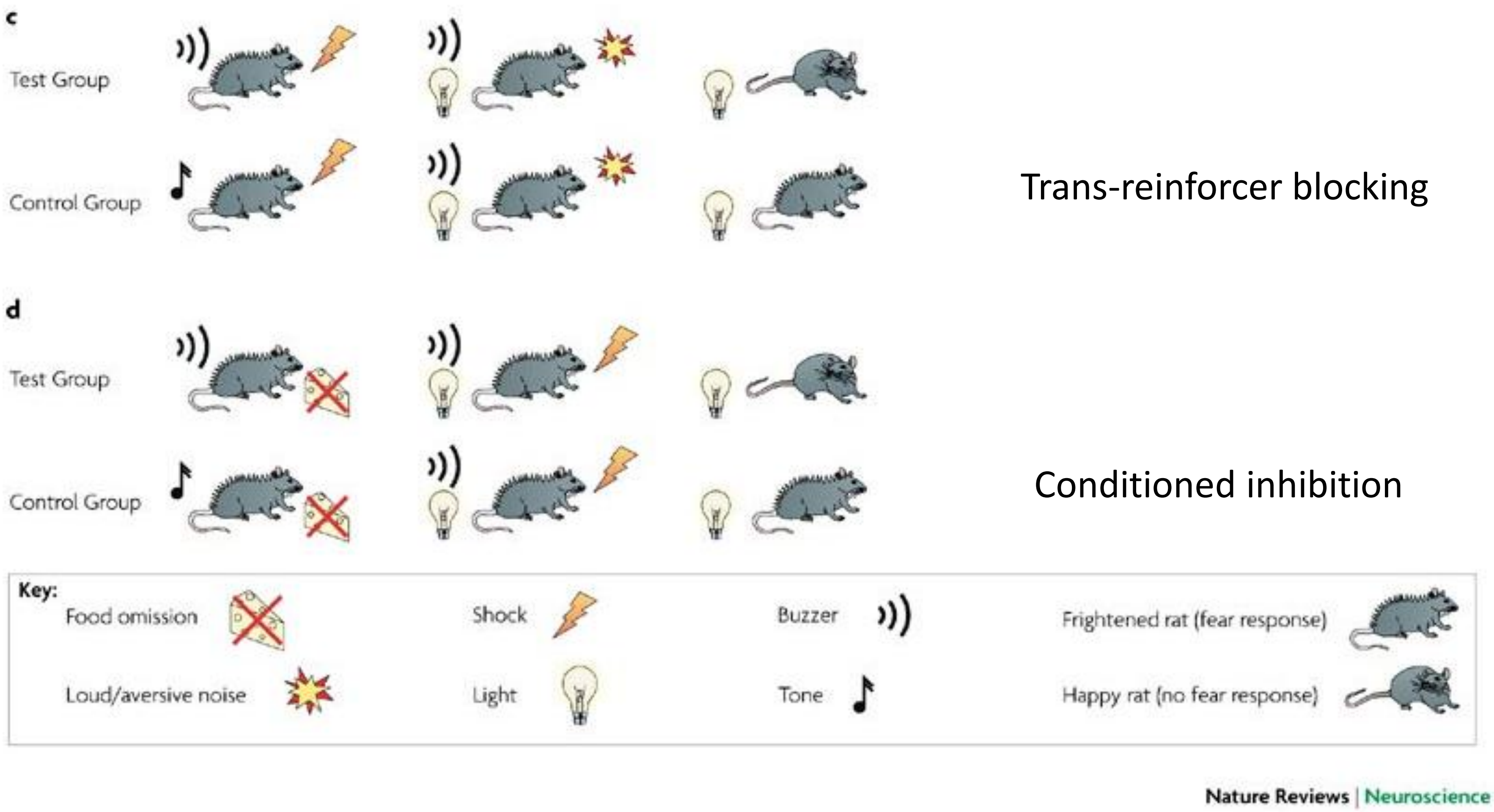
- **Acquisition**.
- **Partial Reinforcement**
- **Generalization** (little albert, watson&rayner, 1920)
- **Interstimulus Interval (ISI) effects**.
- **Intertrial Interval (ITI) effects**.

Blocking (Kamin, 1968)



Nature Reviews | Neuroscience

Conditioned inhibition and more



Trans-reinforcer blocking

Conditioned inhibition

Suggests common brain mechanisms

Relative validity (wagner 1968)

- **Experimental Group**
- 10 x Tone and Light followed by food
- 10 x Click and Light followed by nothing causing extinction
- **Control Group**
- 5 trials of Tone and Light followed by food
- 5 trials of Tone and Light followed by nothing causing extinction
- 5 trials of Click and Light followed by food
- 5 trials of Click and Light followed by nothing causing extinction
- Total experience of the light is the same for both groups as both have 10 light food pairings and 10 light no food pairings yet the animals in the experimental group associated less with the light.
- In simple terms it is attending more to a stimulus that constantly predicts the outcome and attending less to a poor predictor

- *Learning occurs not because two events co-occur, but because that co-occurrence is UNPREDICTED*

Rescorla-Wagner rule (1972)

Learning to predict reward R given stimulus U=1

Goal: Form a prediction of the reward V of the form:

$V=\omega U$

And learn to change ω :

$\Delta \omega = \varepsilon(R-V)U$

After learning of consistent pairing: $\omega=R$

Where:
U=CS availability (0,1);
V=reward prediction:
R=reward availability (0,1) :
ω = weight of the connection between U and V
ε = learning rate
R-V = prediction error

Blocking

- Given U1, U2 and R, after U1 has been learnt:
- $\omega 1 = R$
- $V = \omega 1 U1 + \omega 2 U2$

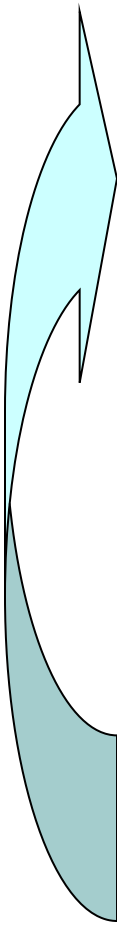
- Prediction error: $R - V = 0$

  And no learning occurs for $\omega 2$

But: two main problems

- Temporal credit assignment (or who is to blame?)
  - Rewards are delayed, and come after many actions and states has occur.
  - We need to propagate the rewards back…
- Exploration / exploitation tradeoff
  - Trust one set of reasonably good cards, and the ace might hide in the other

TD learning

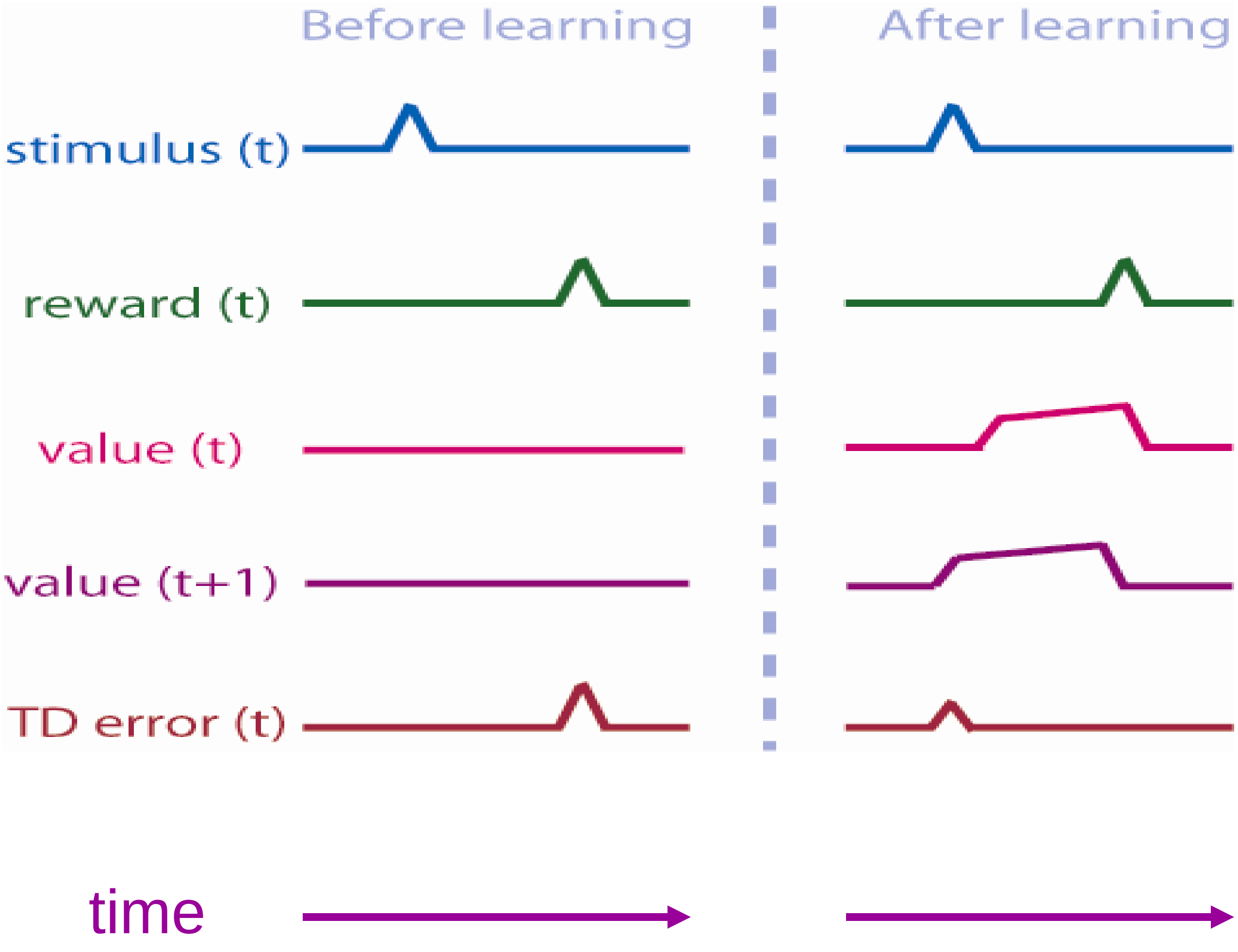


1. Estimate value of current state $(V_t{=}r_t{+}\ \gamma' r_{t+1}{}^{+\cdots})$: (discounted) sum of expected rewards
2. Measure ‘truer’ value of current state: reward at present state + estimated value of next state $(r_t{+}\ \gamma V_{t+1})$
3. TD error $\delta_t = r_t + \gamma V_{t+1} - V_t$
4. Use TD error to improve 1 $(V_t^{k+1}{=}V_t^{k}{+}\eta\ \delta_t)$

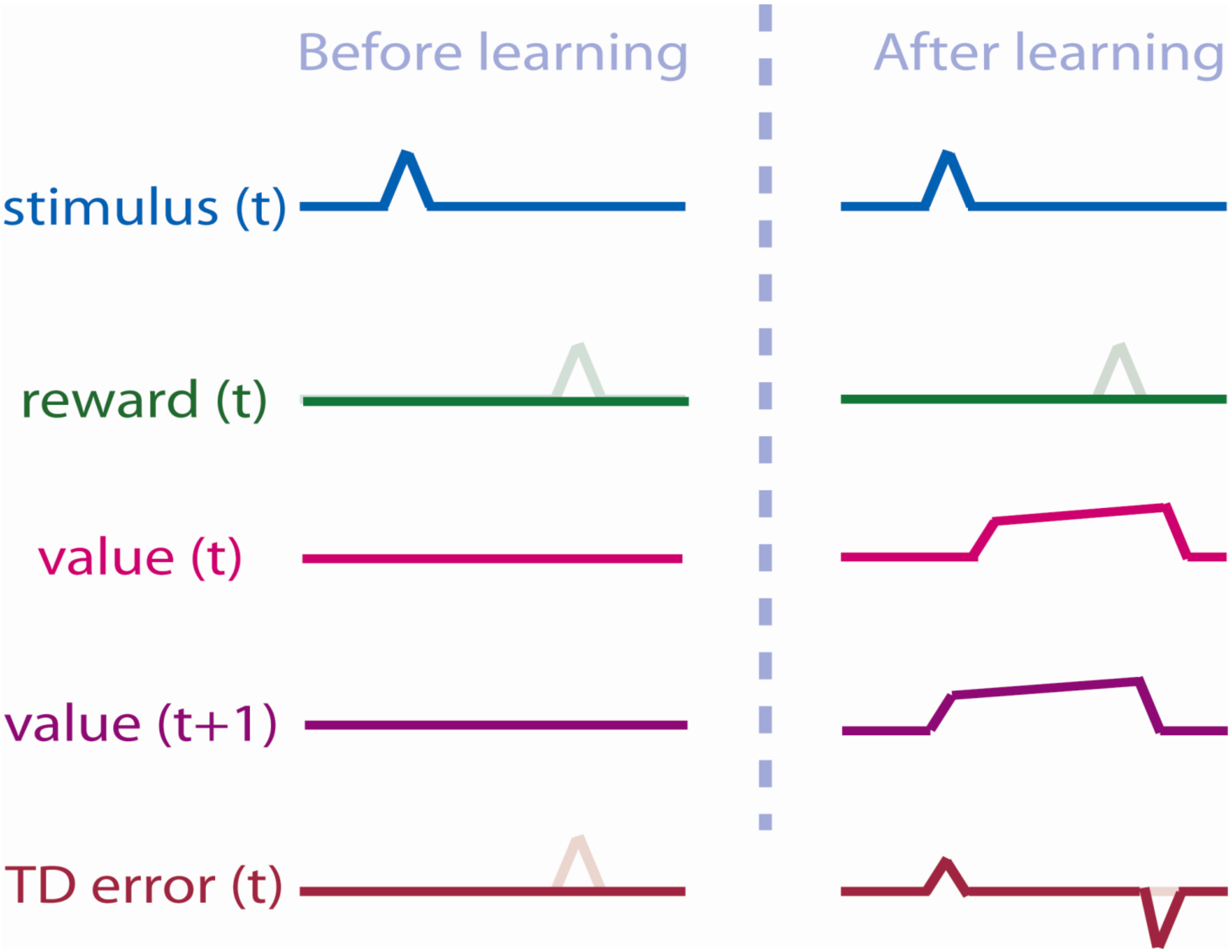
where: $V_{t\ =\ value}$ *of the state reached at time t in iteration k*

r_t *= reward given at time t;* η *= learning rate,* δ *= prediction error*

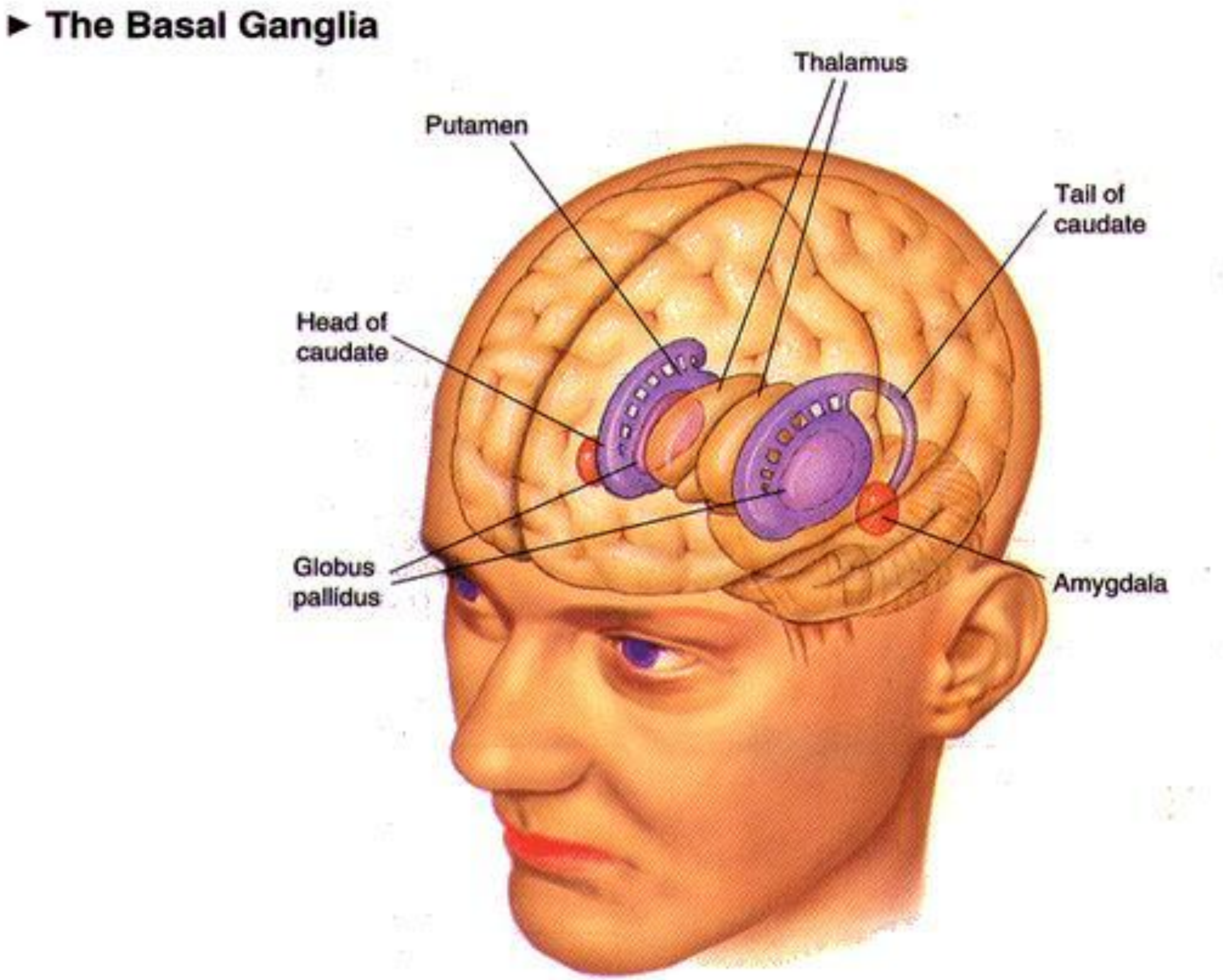
TD error: $\delta_t = r_t + \gamma V_{t+1} - V_t$



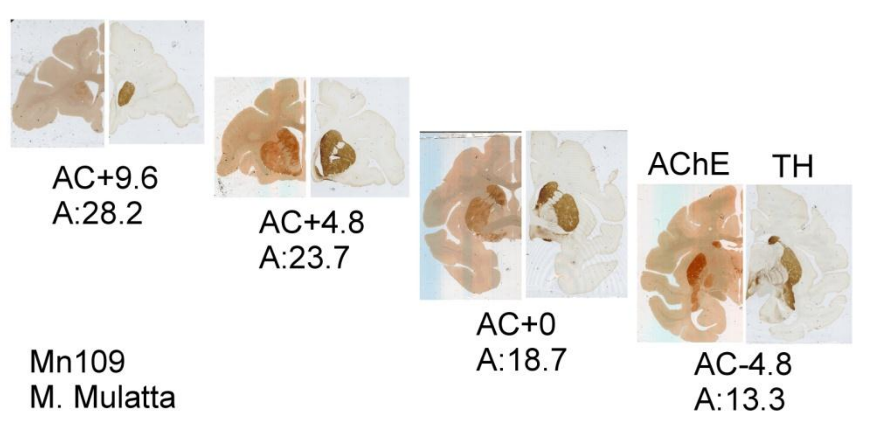
Reward omission

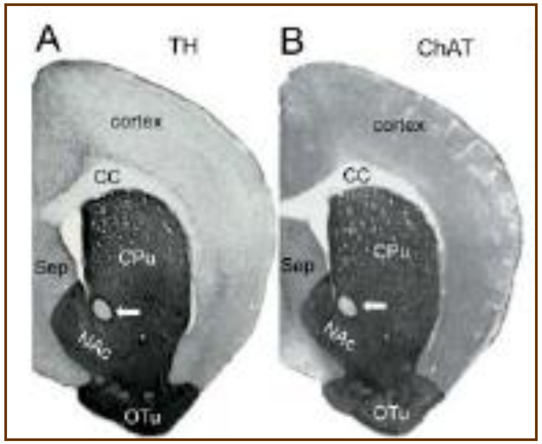


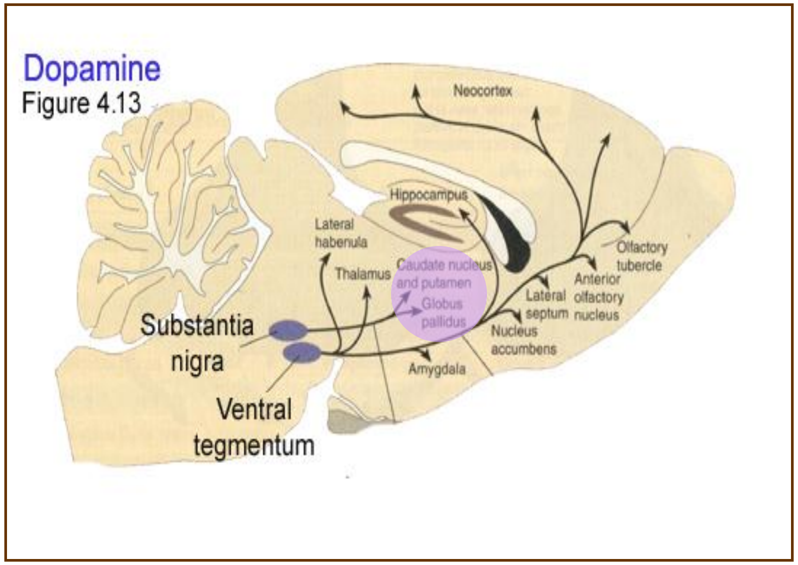
The basal ganglia

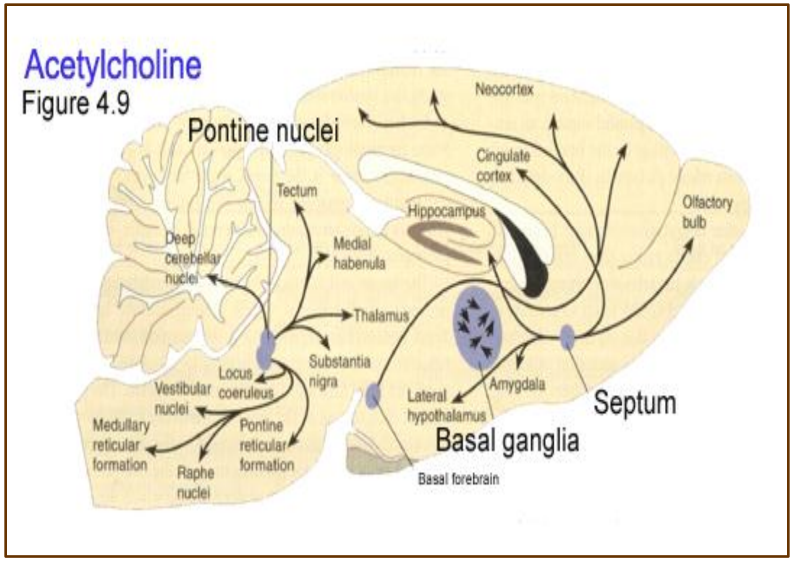


Dopamine and acetylcholine meet in the striatum

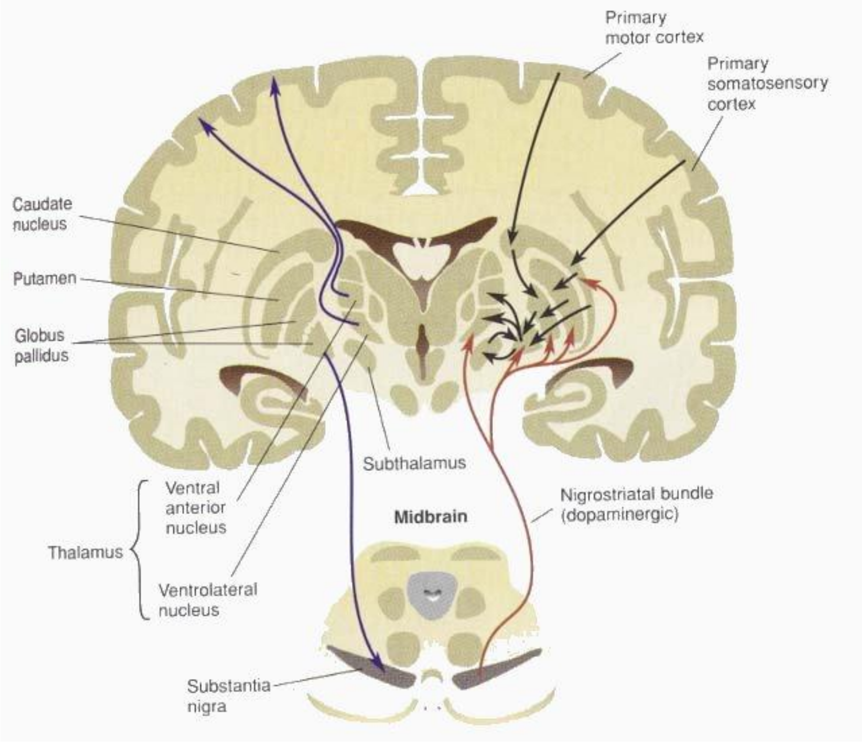


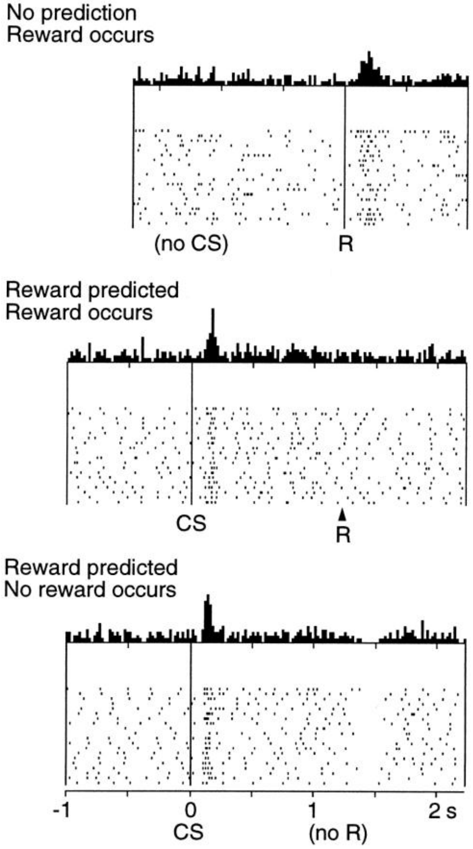


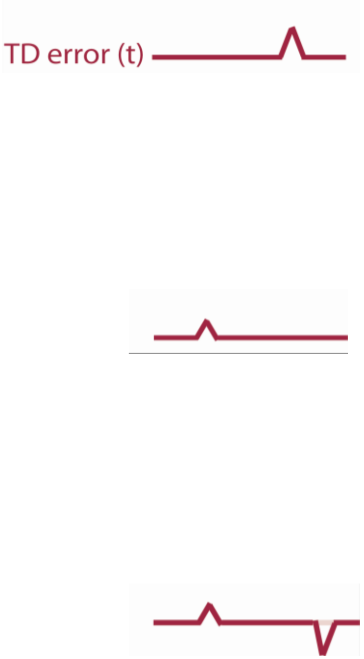




Dopamine match surprise signal

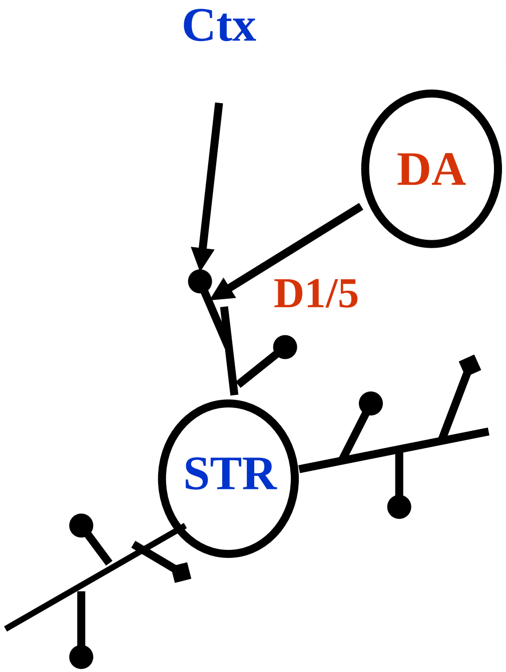


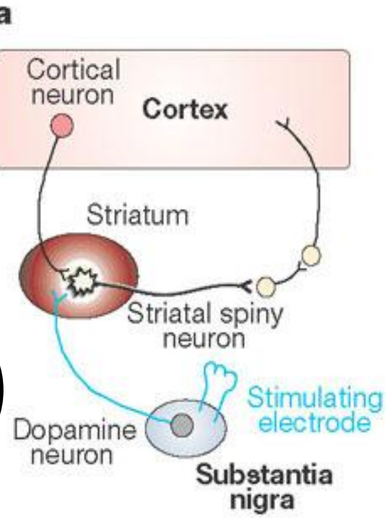


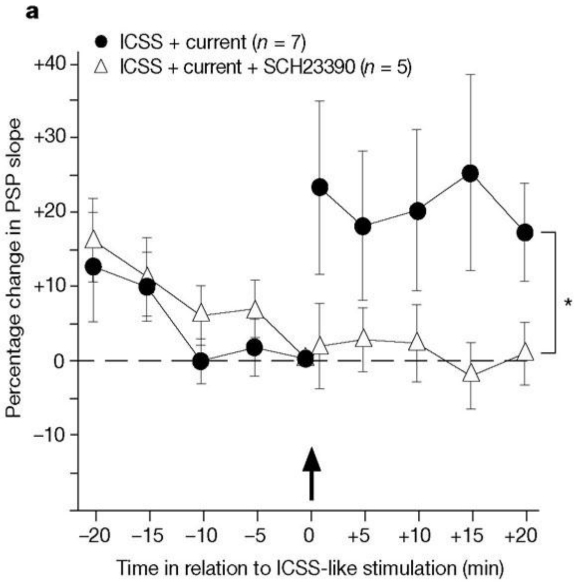


Schultz et al,
JNS 13:
900-913 ,1993

LTP in cortico-striatal synapses

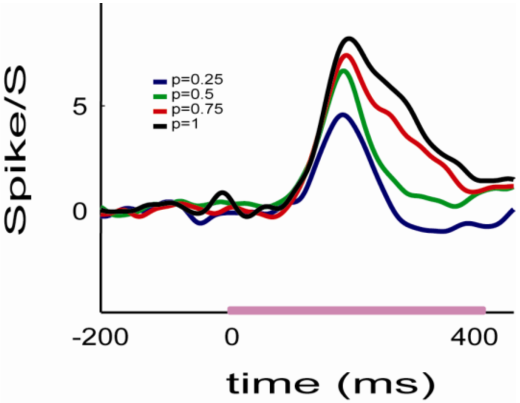


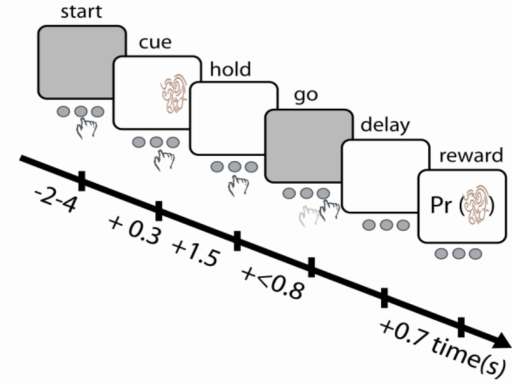


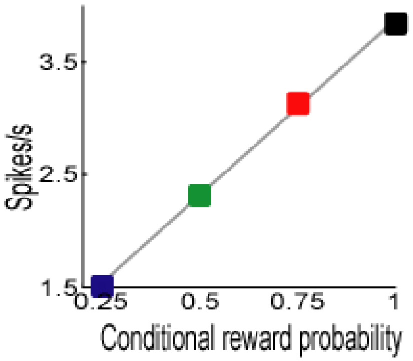


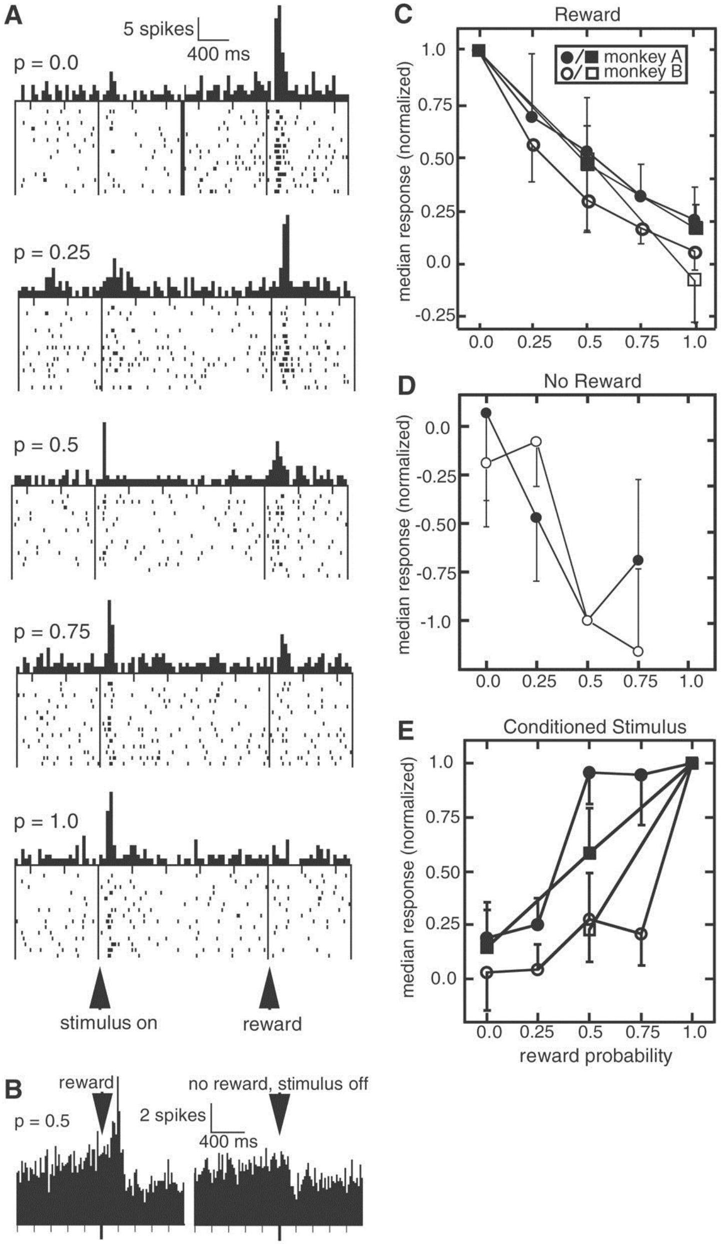
Reynolds et al, A cellular mechanism of reward-related learning Nature 413, 67 - 70 (2001)

Dopamine reflects probability of cue giving reward



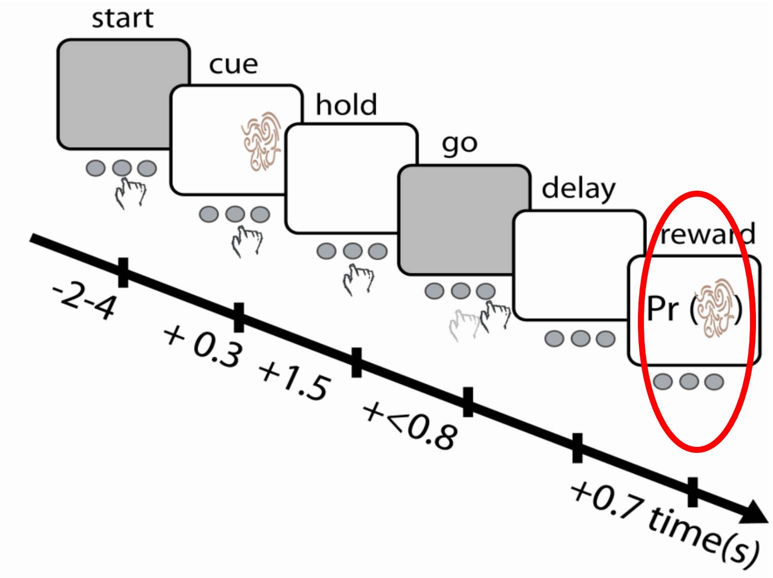


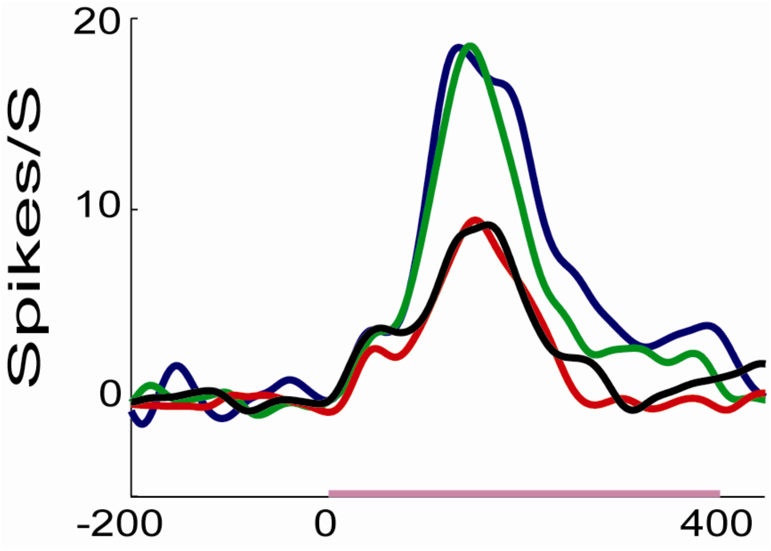


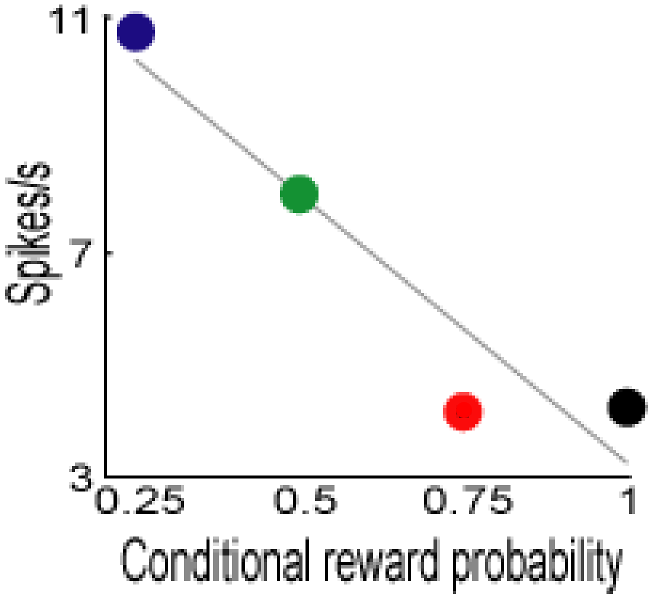


Morris et al, Neuron, 2004
Fiorillo et al, Science, 2003

And inversely to the reward:





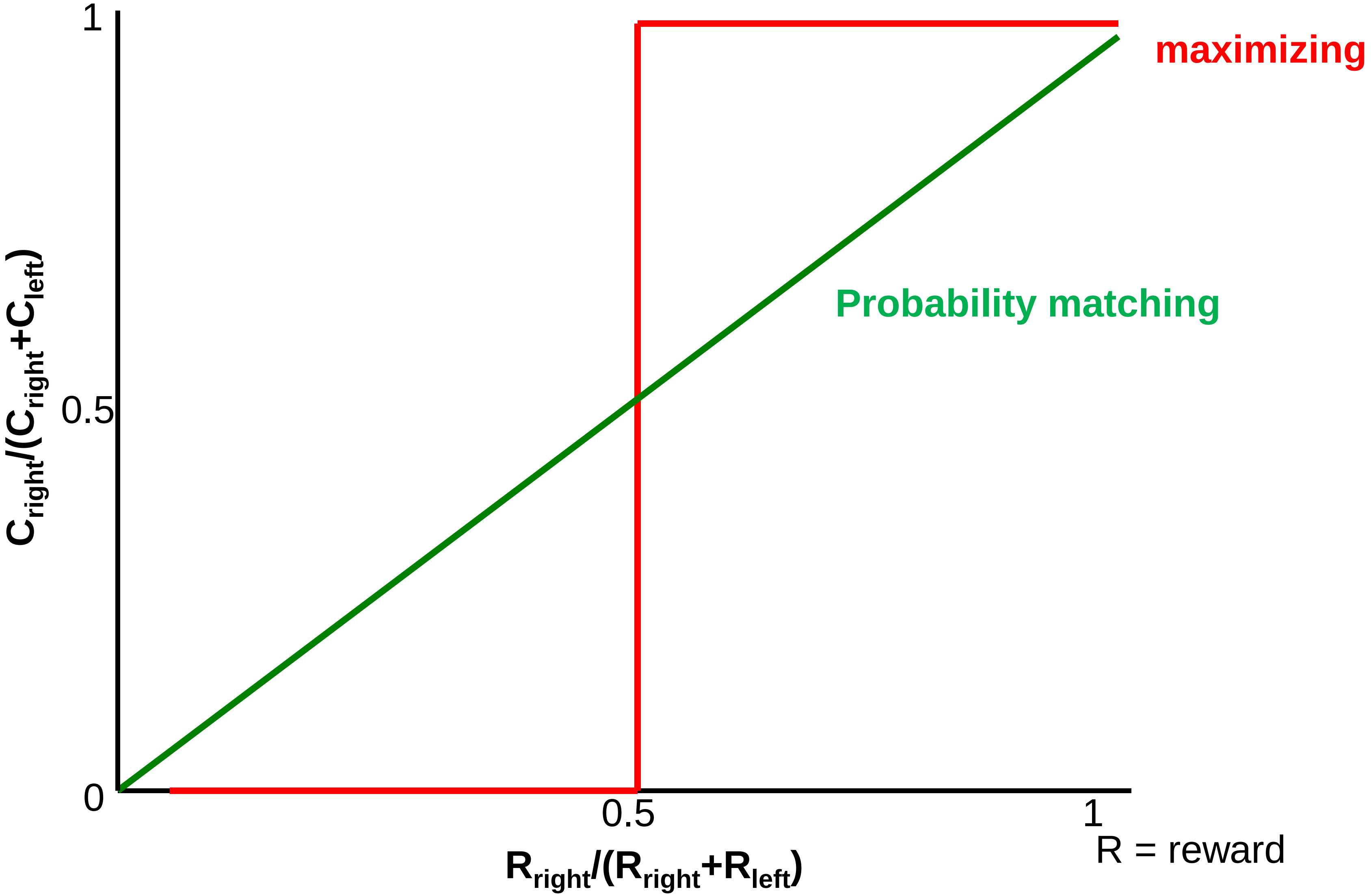


Dopamine responses

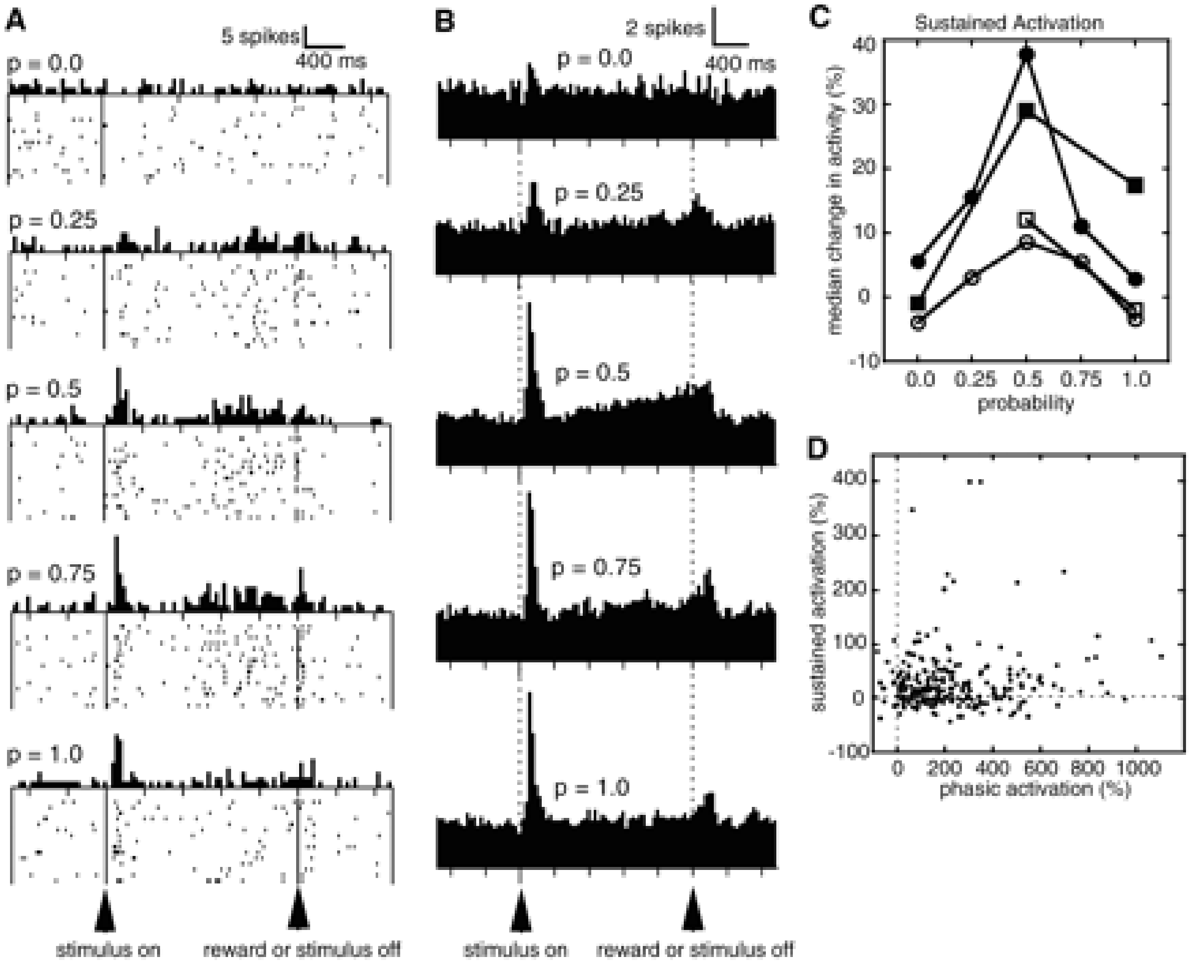
- Responses to visual cue are correlated with future reward probability
- Responses to reward are inversely correlated with reward probability
- Dopamine neurons provide an accurate surprise signal (but only in the positive domain)

- What about actions?

Exploration-exploitation: decision behavior



Uncertainty signal in dopamine neurons

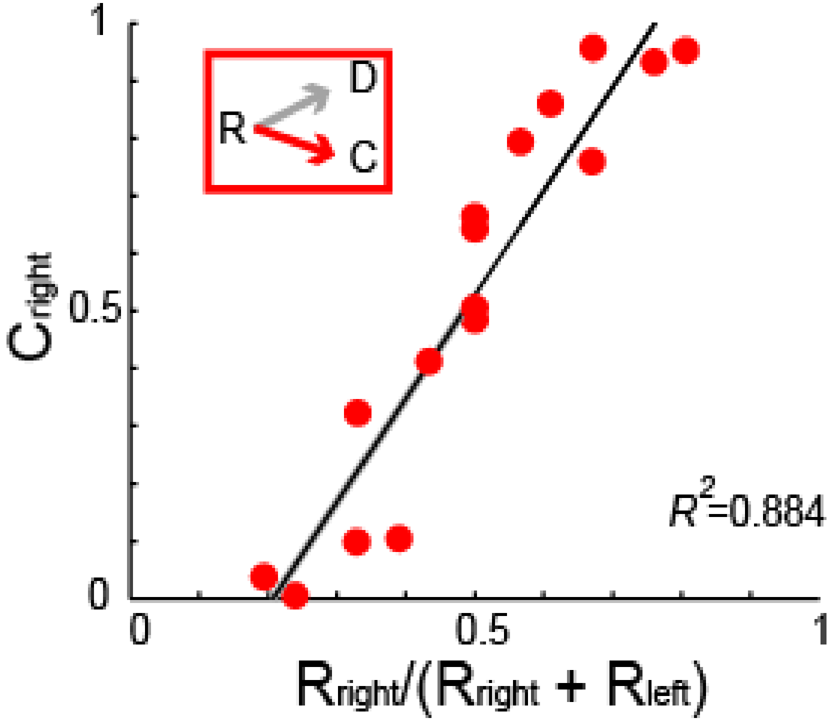


Risk taking?

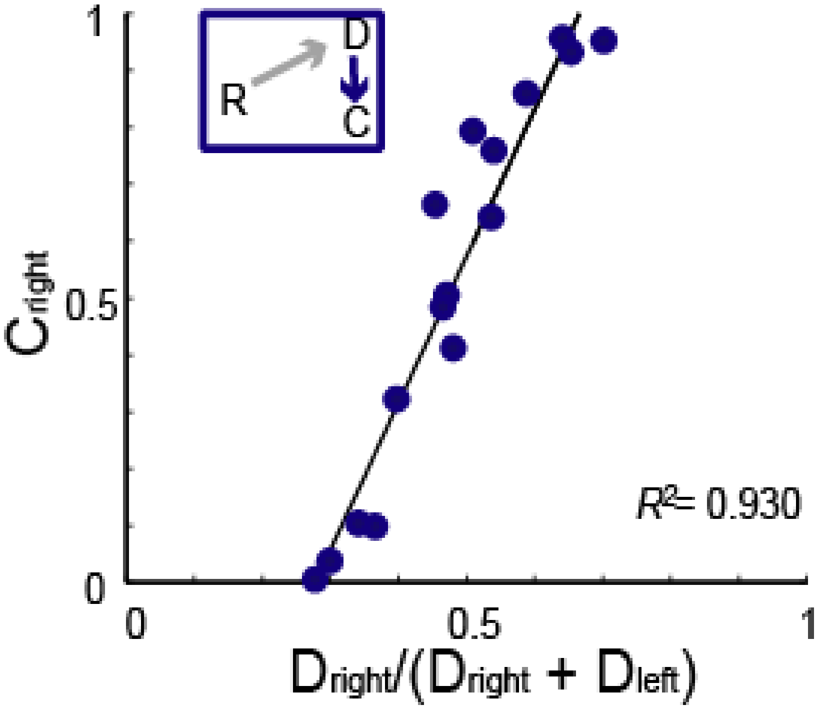
Fiorillo, Science, 2003

Probability matching in monkeys

Behavior



Dopamine



Morris, Nat. Neurosci. 2006

Fear thou not – the negative domain

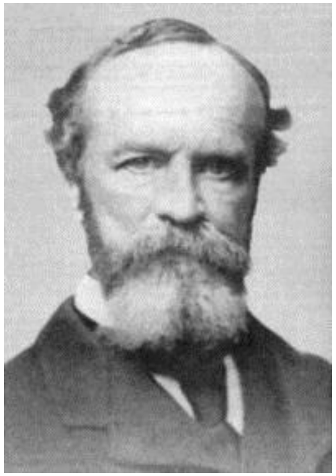
- What is a “reward”?
- Learning is motivated by threats to survival
- Threats are reinforcers
- Fear is a prime motivator

	Decreases behavior	Increases behavior
Presented	Positive punishment	Positive reinforcer
Taken away	Negative punishment	Negative reinforcer

Taking drugs?

More fun,
less withdrawal

What are emotions?



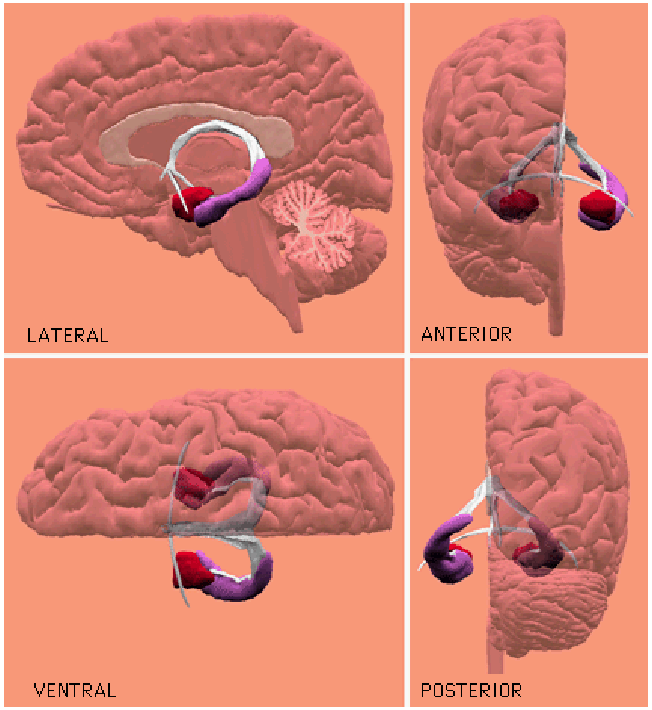
William James
1842-1910

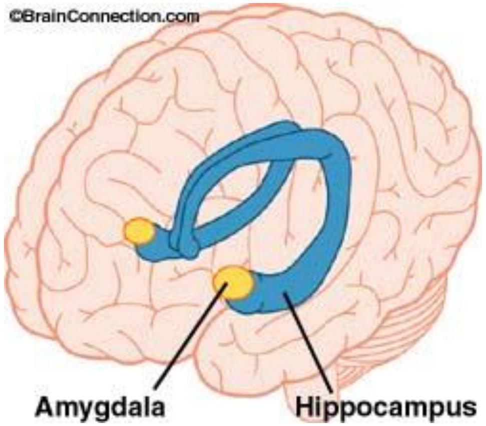
Do we run from a bear because we are afraid, or are we afraid because we run?

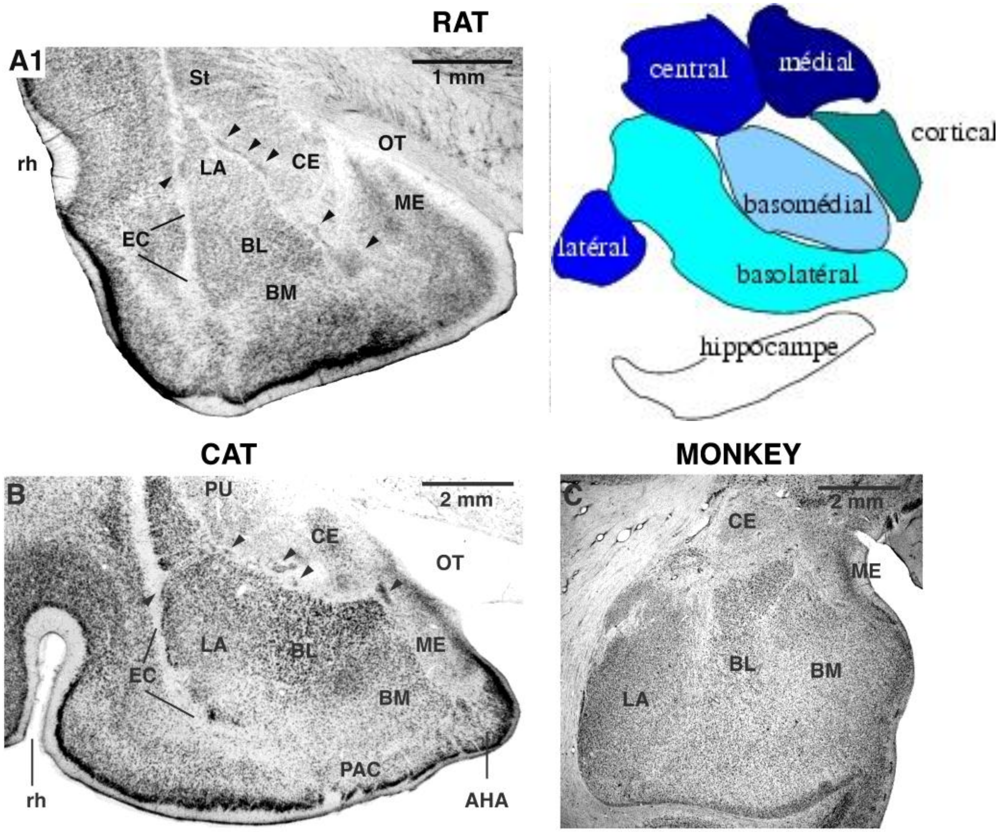
James proposed that the obvious answer, that we run because we are afraid, was **wrong**, and instead argued that we are afraid because we run.

Perception=>bodily changes=>feeling

The amygdala

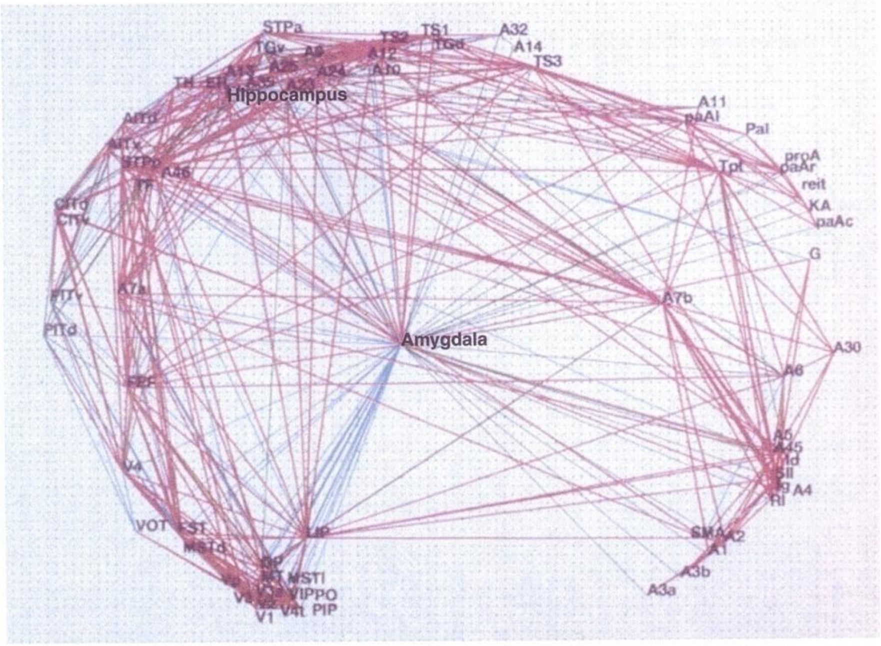


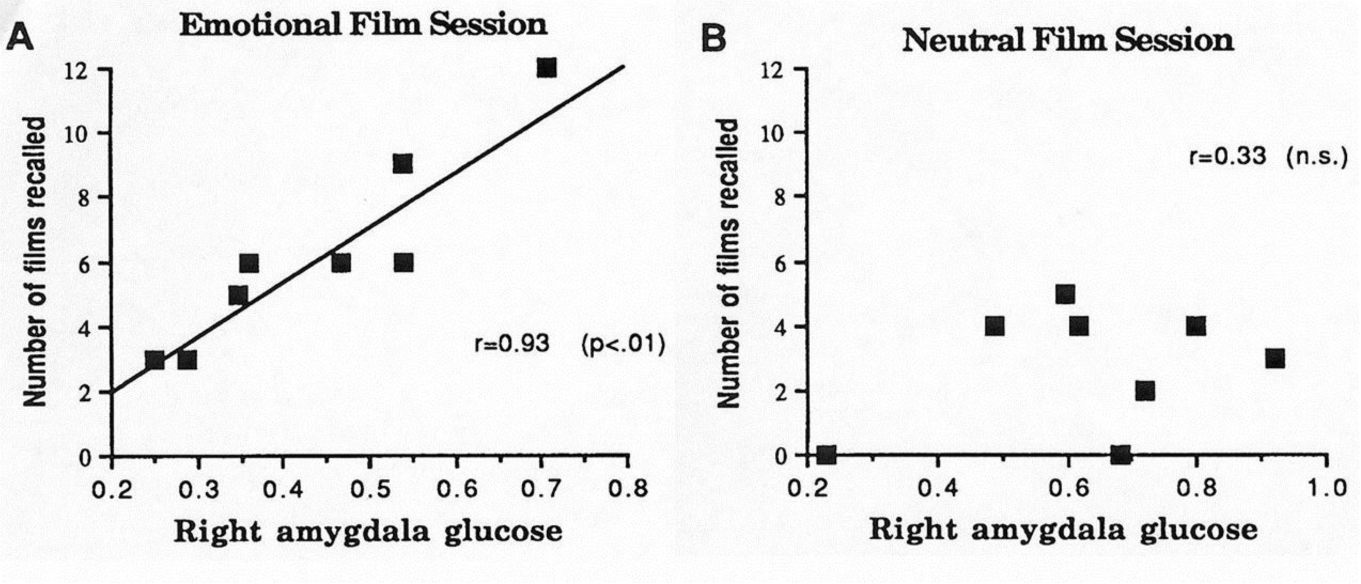




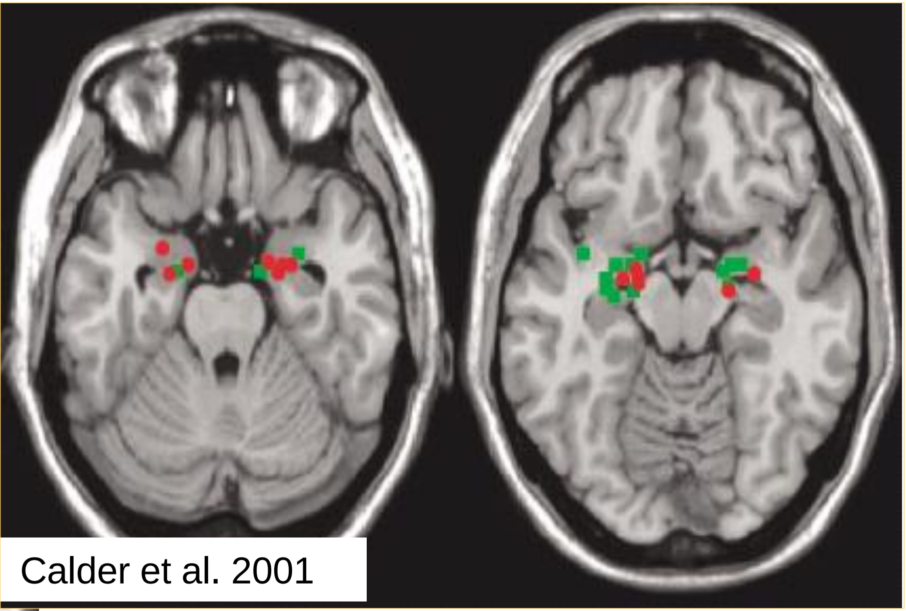
Amygdala and its basolateral complex (BLA)

- BLA evolution parallels that of the prefrontal cortex
- BLA cell types reminiscent of cortex
- Cortical projections are much more extensive in primates
- Most cortical projections of the amygdala originate from BLA (none from CEA)





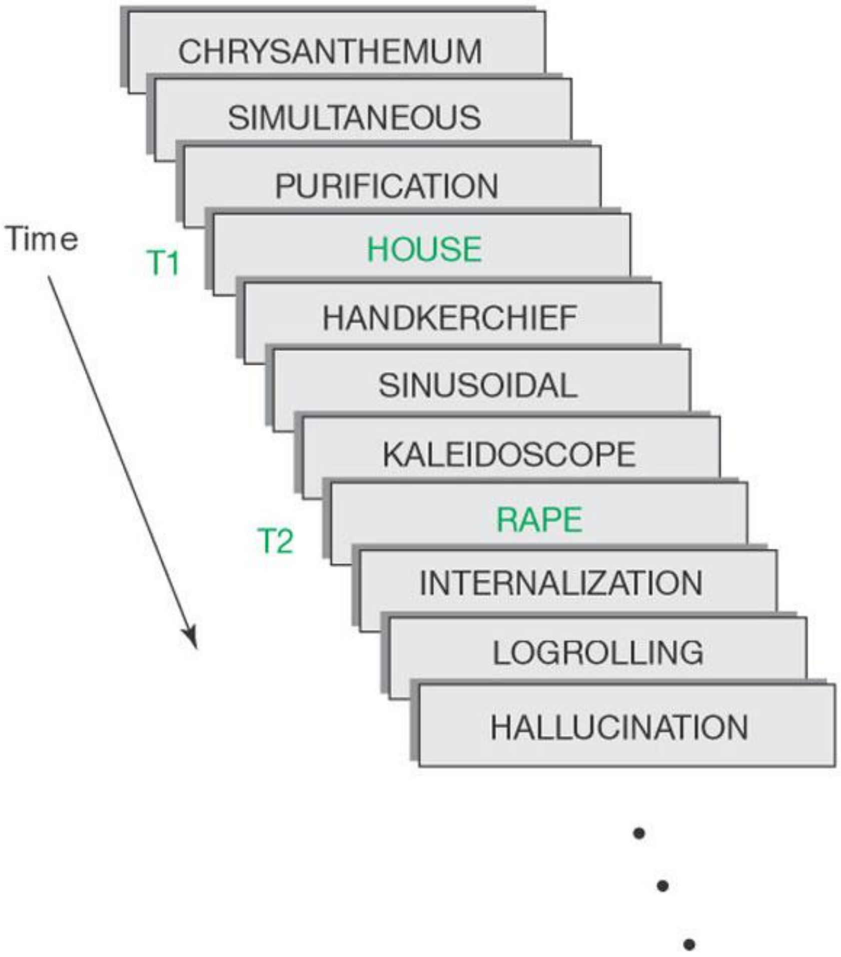
Cahill et. al. PNAS, 1996

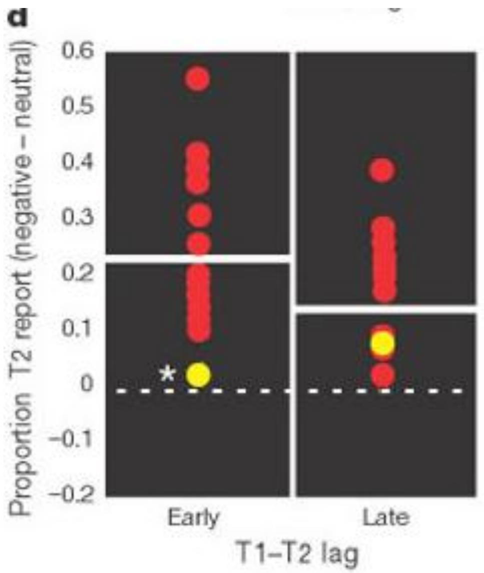


Calder et al. 2001

Red – תגובה לפנים מפוחדות (8)
Green – תגובה להתנית פחד (6)

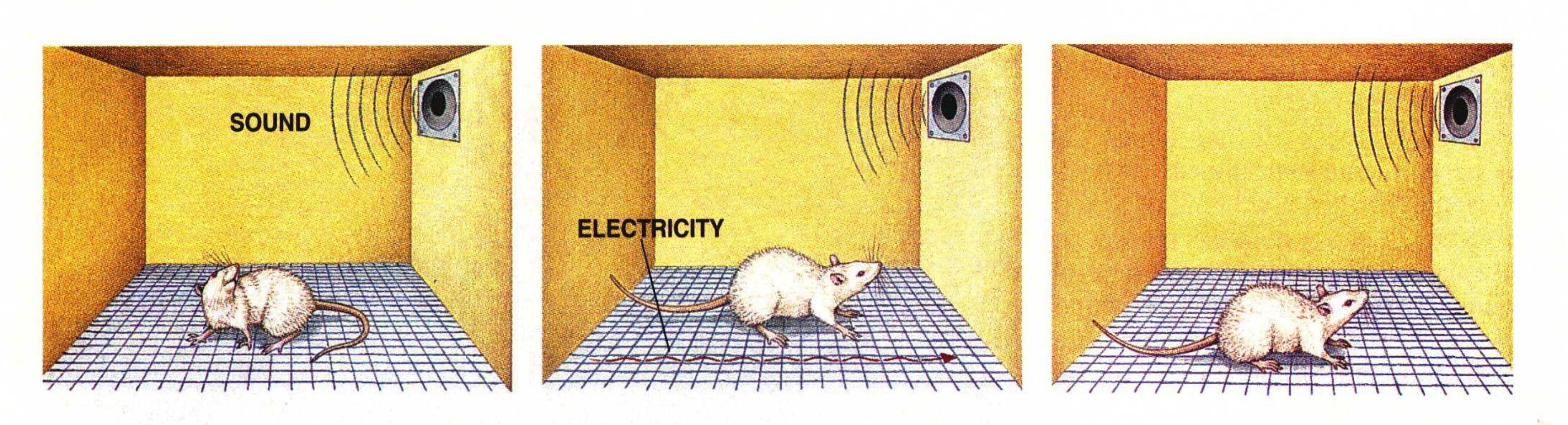
Emotional affect on "Attentional blink" is reduced with amygdala damage

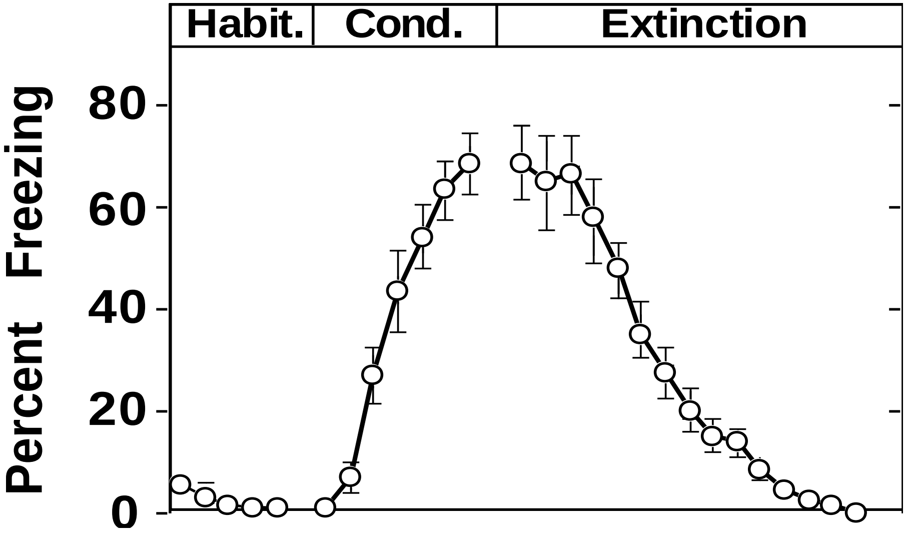




Anderson, Nature, 2001

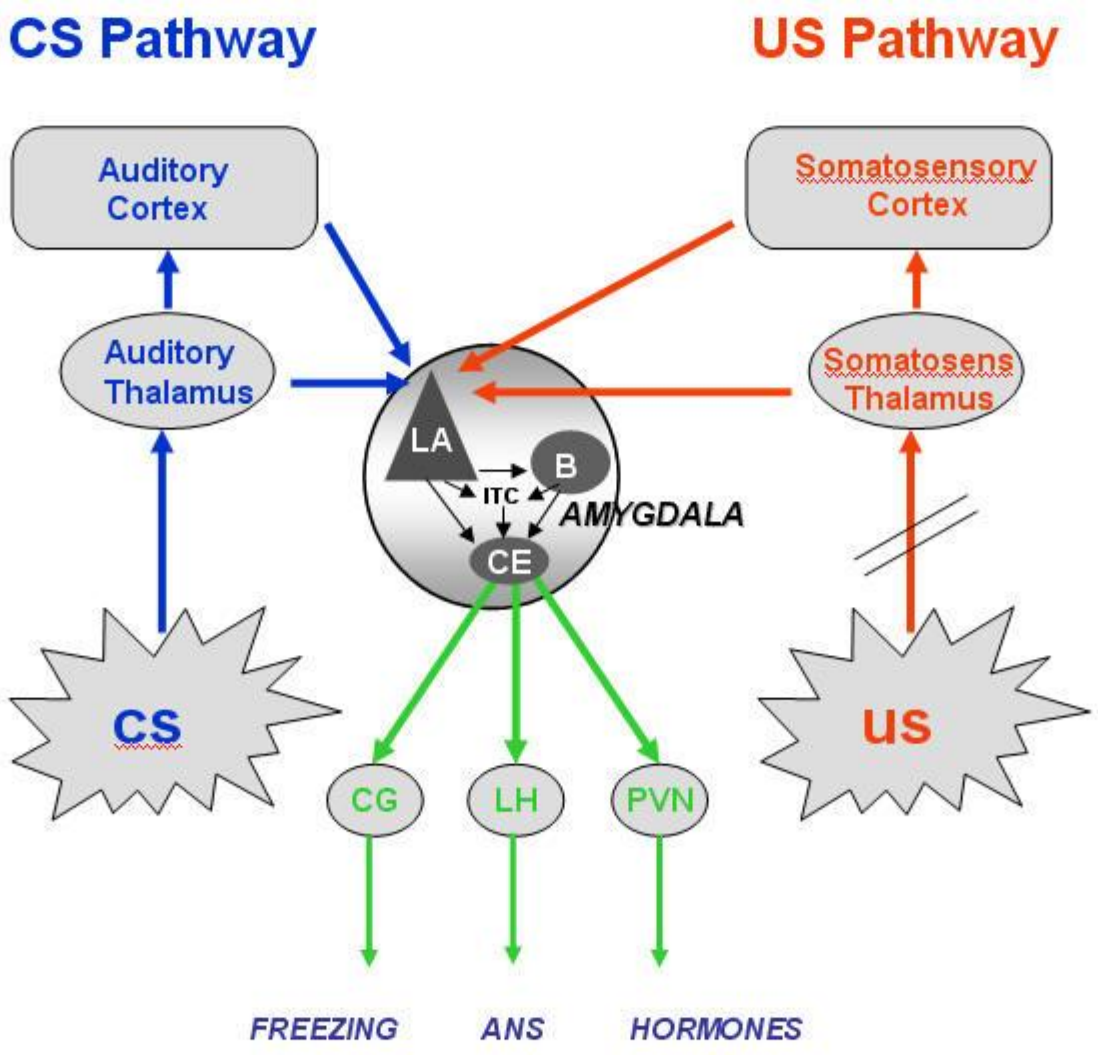
Classical fear conditioning





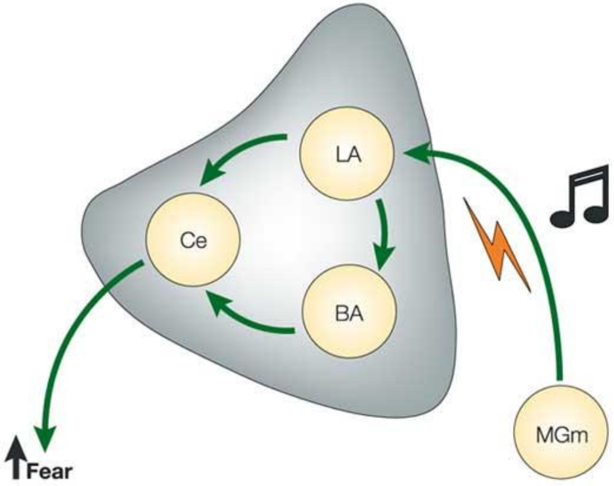
CS-US pairing
Tone = conditioned stimulus (CS)
Foot-shock = unconditioned stimulus (US)
Freezing = conditioned response (CR-UR)

Fear circuit

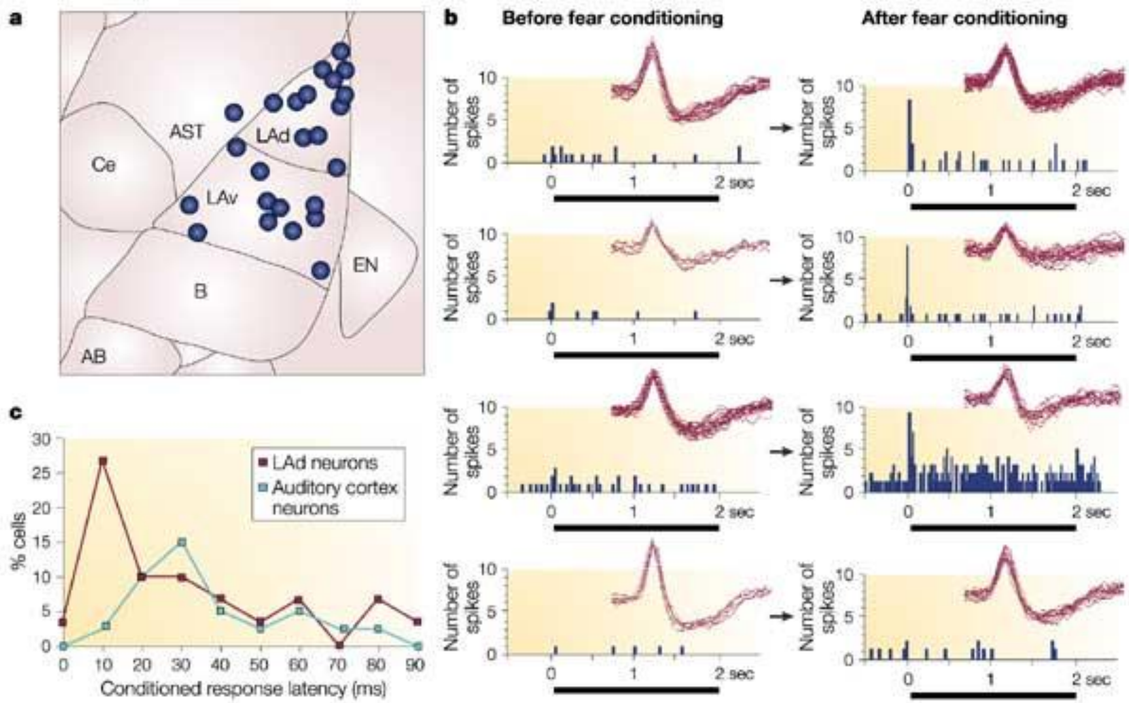


Ledoux JE

Neurons acquire tone responses after conditioning



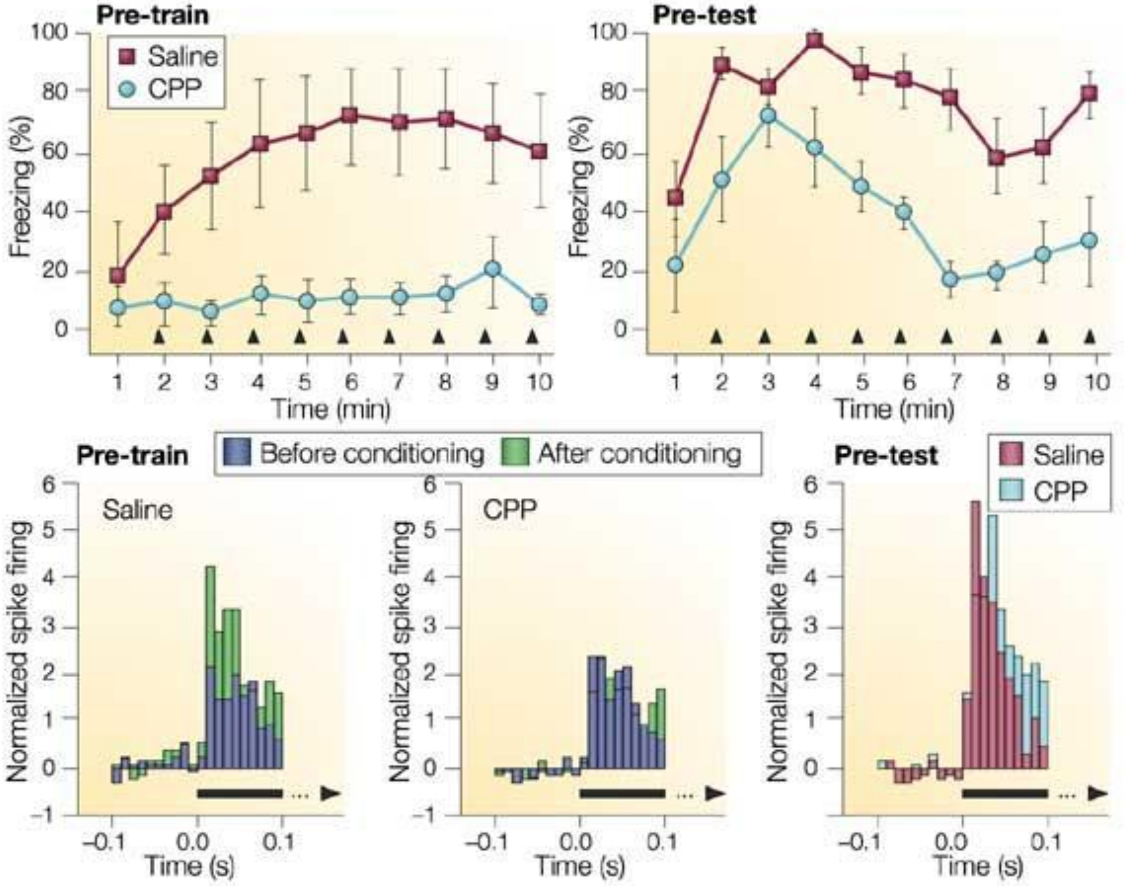
Nature Reviews | Neuroscience



Nature Reviews | Neuroscience

LTP is required

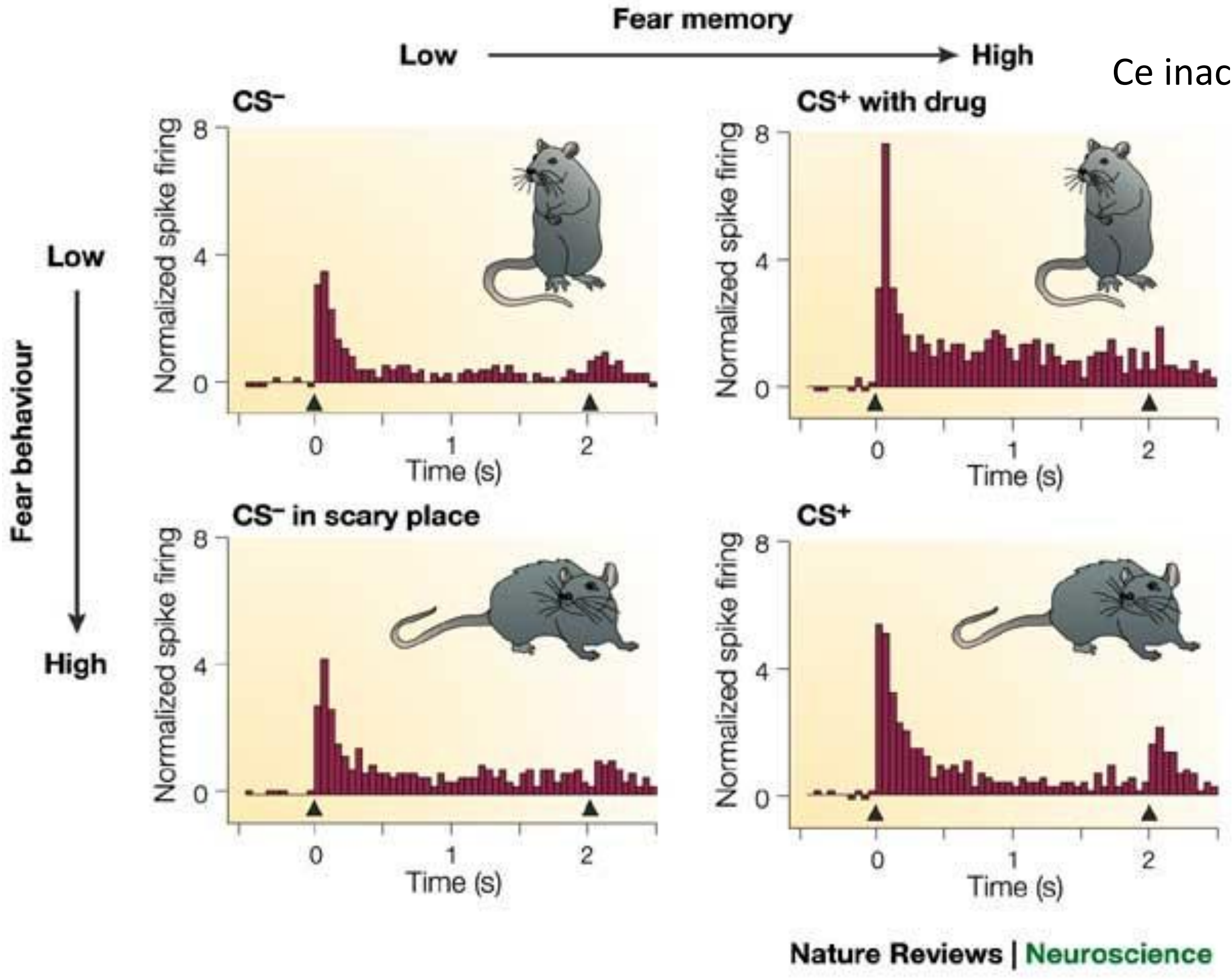
NMDA (***N***-methyl-**D**-aspartate, glutamate receptor) is involved in both the acquisition of fear memory and the induction of long-term potentiation (LTP) in the amygdala.



CPP (3-(2-carboxypiperazin-4-yl) propyl-1-phosphonic acid), a competitive NMDA-receptor antagonist

Nature Reviews | Neuroscience

LA encodes memory independent of fear behavior

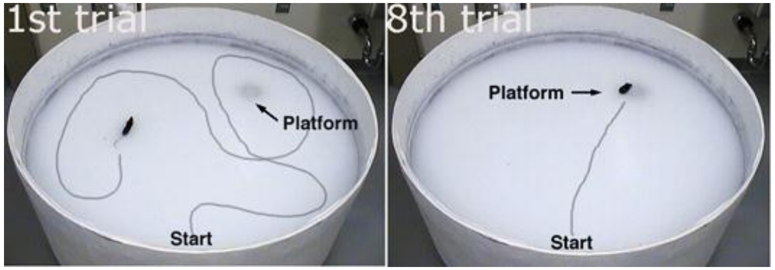


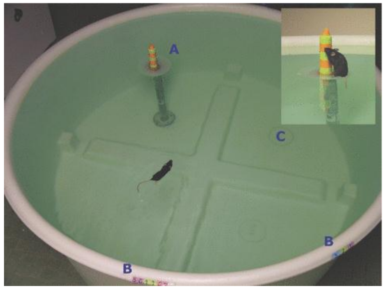
Ce inactivation

Amygdala: modulation of emotional memory

- Hippocampal dependent learning: spatial
- Striatum dependent-learning: cue-related

Morris water maze





Neurobiology: Packard *et al.*

Proc. Natl. Acad. Sci. USA 91 (1994) 8479

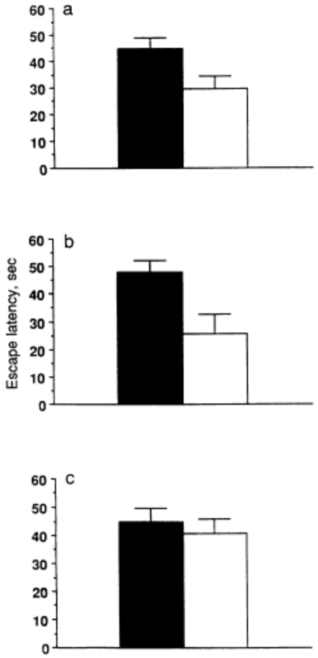


FIG. 1. Mean (±SE) escape latencies of *d*-amphetamine (10 μg) (□) and saline-treated (■) rats on the retention test trial in the spatial task. (*a*) Hippocampal injections. (*b*) Amygdala injections. (*c*) Caudate nucleus injections.

posttraining intracaudate and intrahippocampal injections of *d*-amphetamine on retention of cued and spatial learning in

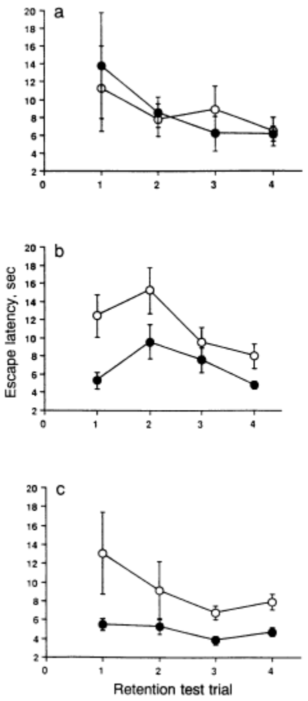
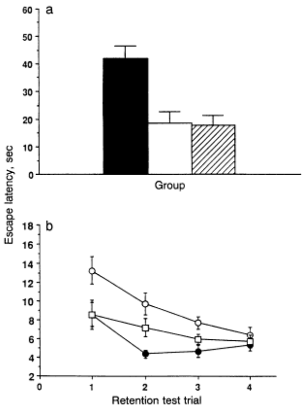


FIG. 2. Mean (±SE) escape latencies of *d*-amphetamine (10 μg) (●) and saline-treated (○) rats on the retention test trial in the cued task. (*a*) Hippocampal injections. (*b*) Amygdala injections. (*c*) Caudate nucleus injections.

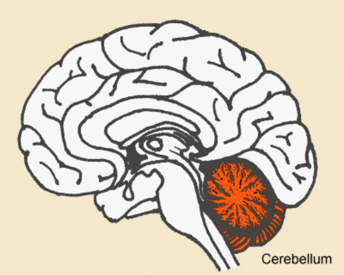


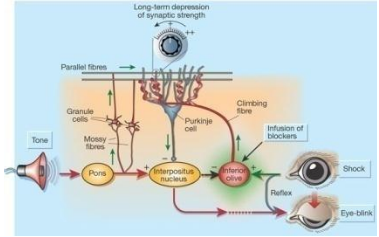
Injection of d-amphetamine into the Amygdala affects both if right after training, but not if pre-testing

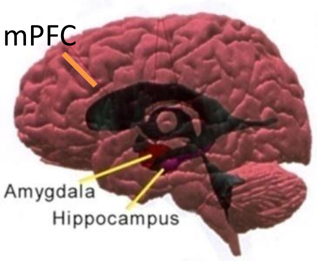
Packard, Mcgaugh

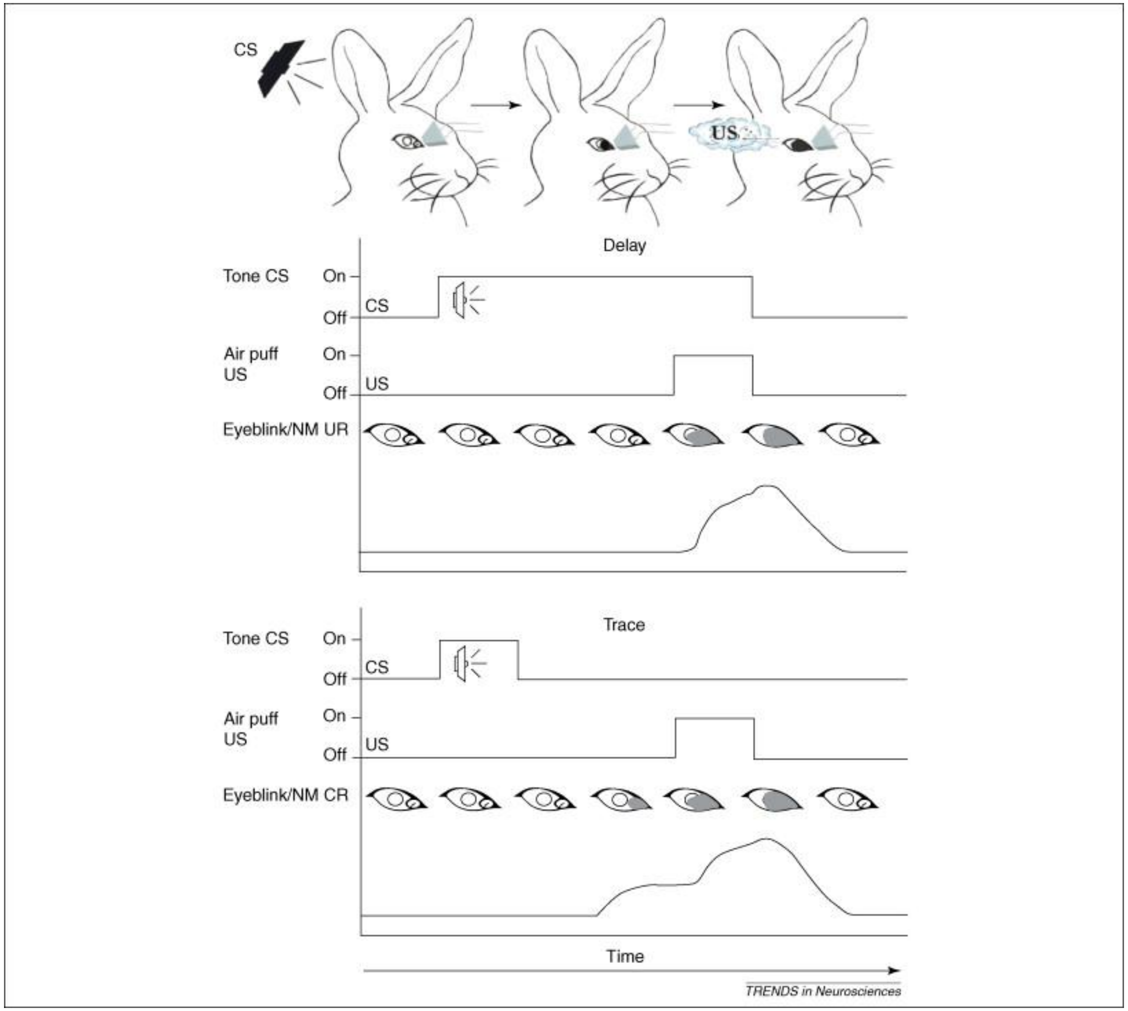
FIG. 3. Mean (±SE) escape latencies of rats receiving intraamygdala posttraining *d*-amphetamine or saline and rats receiving preretention test lidocaine or saline on the retention test trial(s) in the spatial task (*a*) and cued task (*b*). Posttraining/preretention: ■ (*a*) and ○ (*b*), saline/saline; □ (*a*) and ● (*b*), *d*-amphetamine/saline; ▨ (*a*) and □ (*b*), *d*-amphetamine/lidocaine.

Eyelid (blink) reflex conditioning



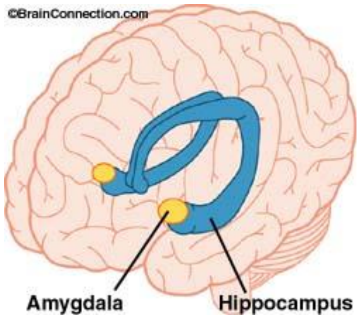


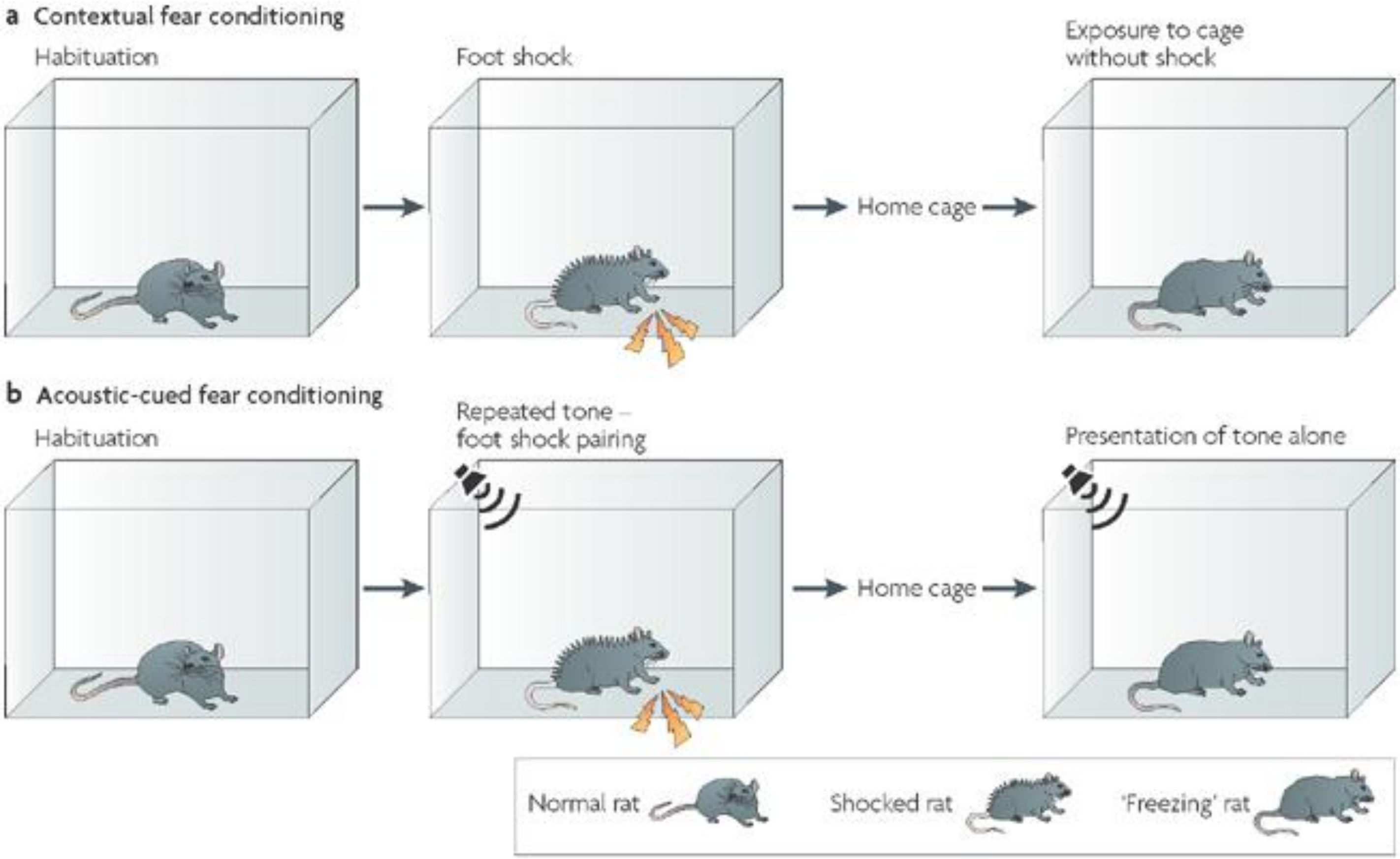




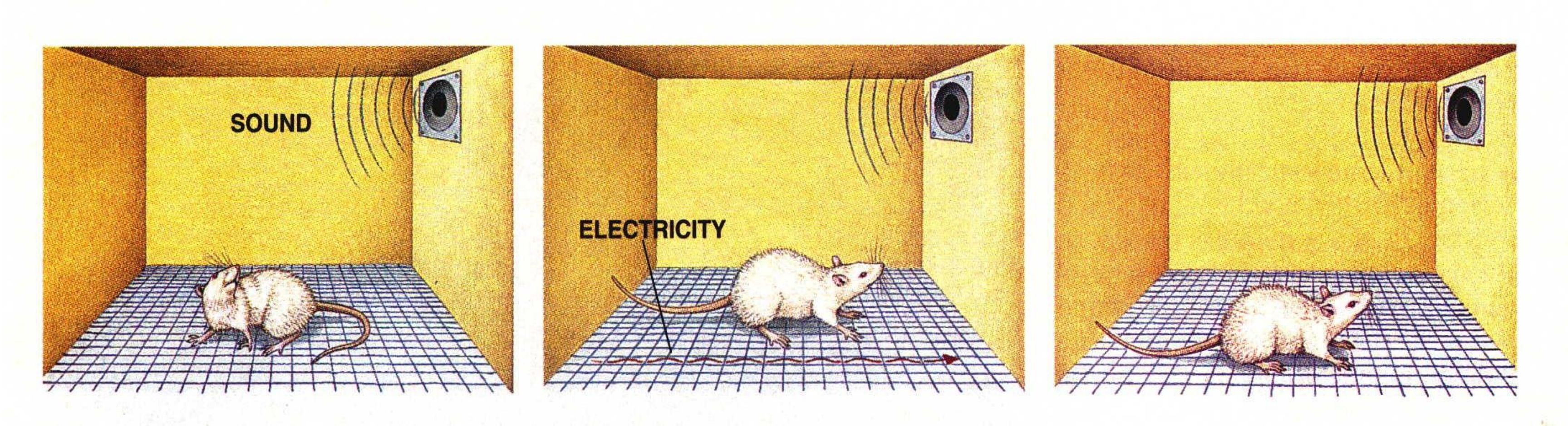
- Why is trace hippocampal-dependent?
- Maintaining the CS? Timing the trace? Harder?
- Eyelid requires ~0.3sec, and hippocampus is required when 0.5-1sec.
- In tone-shock, trace can be 3sec, and hippocampus is required for ~20sec
- This suggest context-conditioning

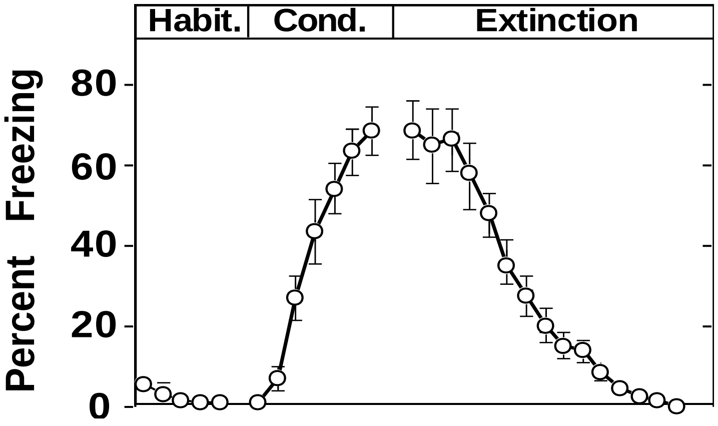
Contextual fear



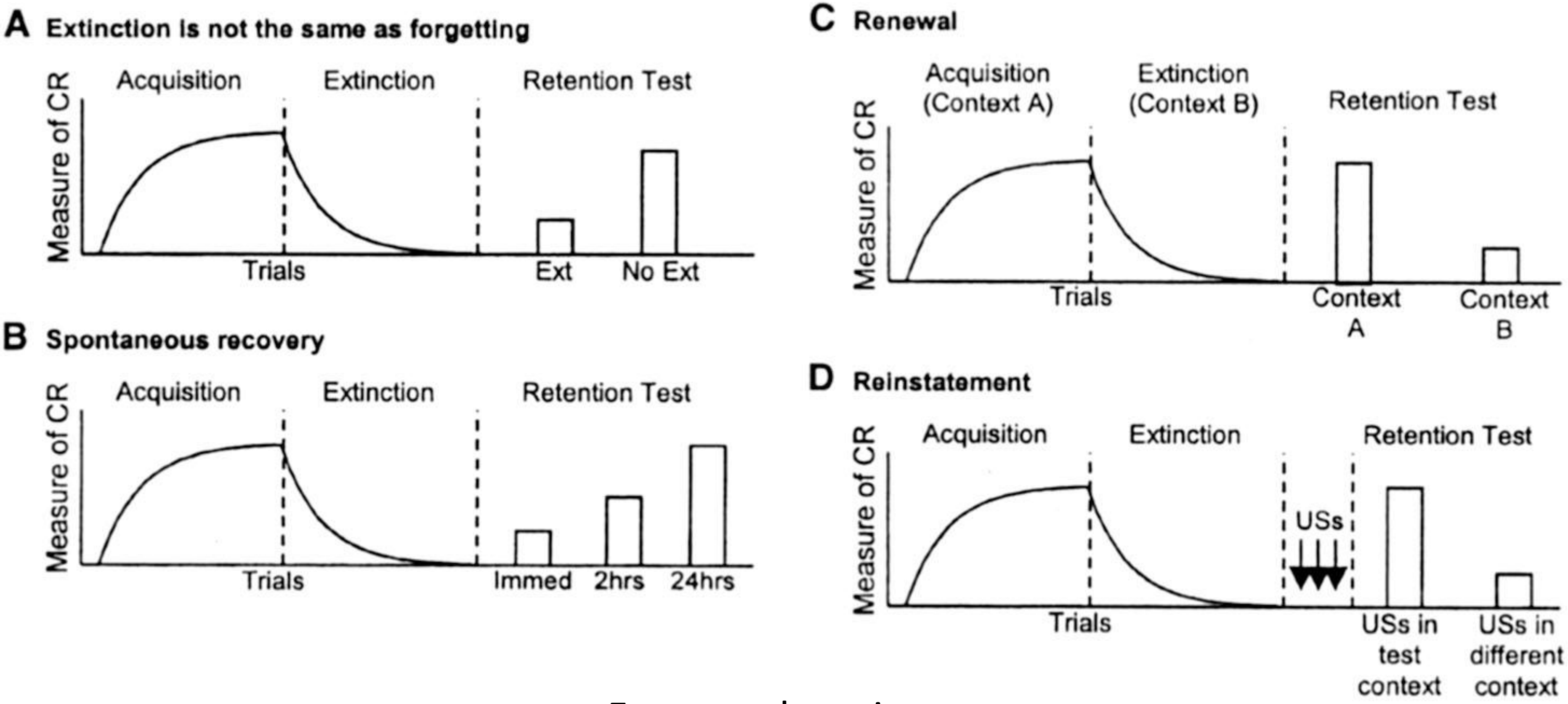

a Contextual fear conditioning
Habituation
Foot shock
Home cage
Exposure to cage without shock
b Acoustic-cued fear conditioning
Habituation
Repeated tone – foot shock pairing
Home cage
Presentation of tone alone
Normal rat
Shocked rat
'Freezing' rat

Extinction of fear-conditioning



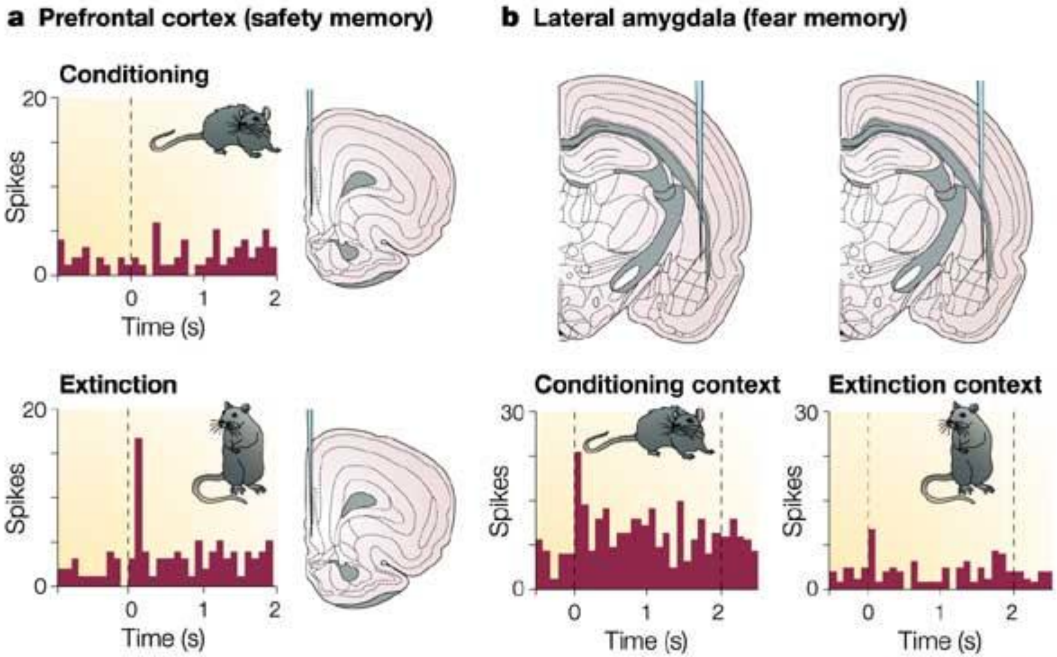


Extinction: a new learning

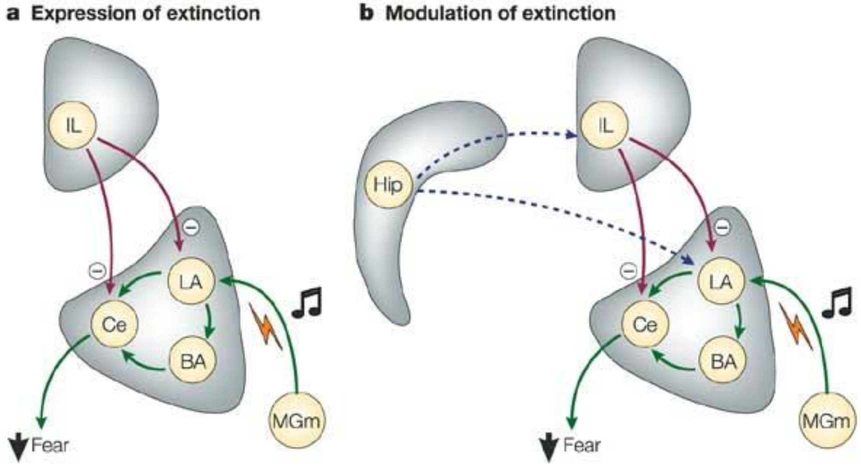


Faster re-learning

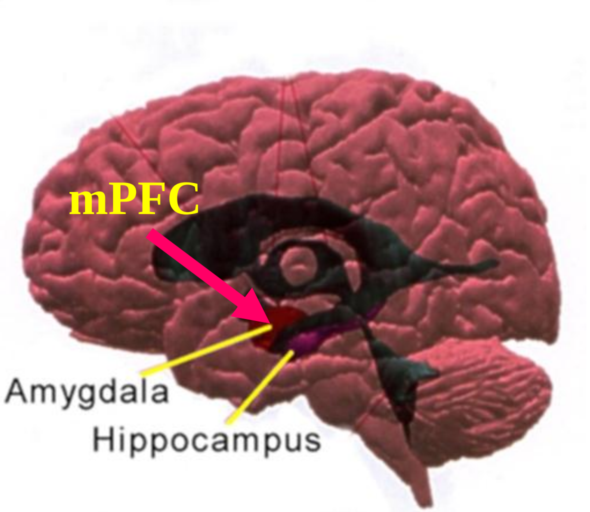
Extinction: brain mechanisms



Nature Reviews | Neuroscience



Nature Reviews | Neuroscience



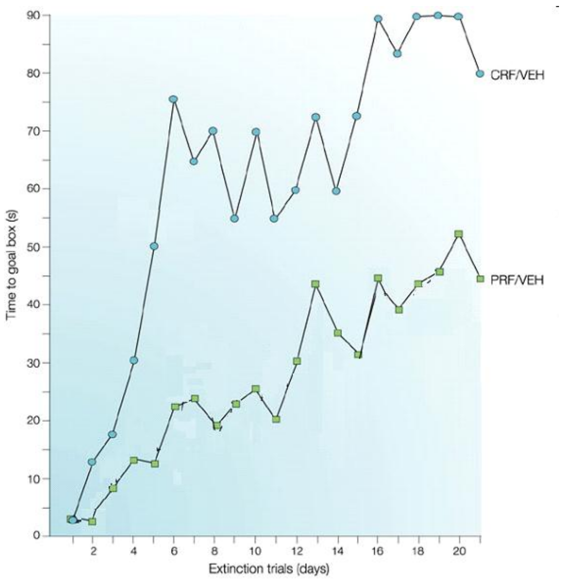
Partial reinforcement extinction effect

- Partial reinforcement
  - Fixed/variable ratio
  - Fixed/variable schedule

- Results in longer extinction learning

- Why?
  - Frustration theory (Amsel): The omission of the US induces frustration. Therefore, during extinction, the frustration predicts the US.
  - Sequential theory (Capaldi): conditioning to strings of NNNRNNNR

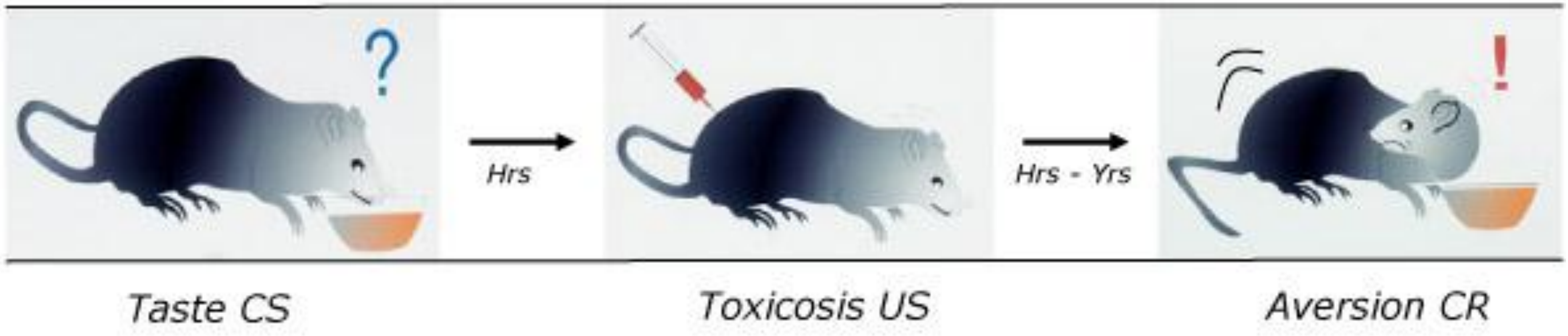
- Bad for behavior flexibility
- Good for education

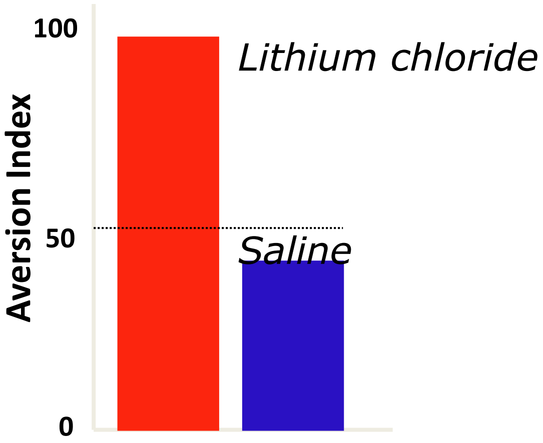


Nature Reviews | Neuroscience

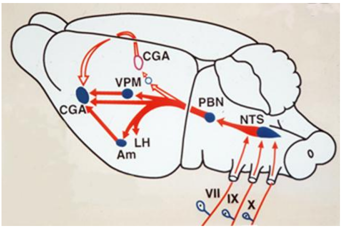
Garcia J

Conditioned Taste Aversion





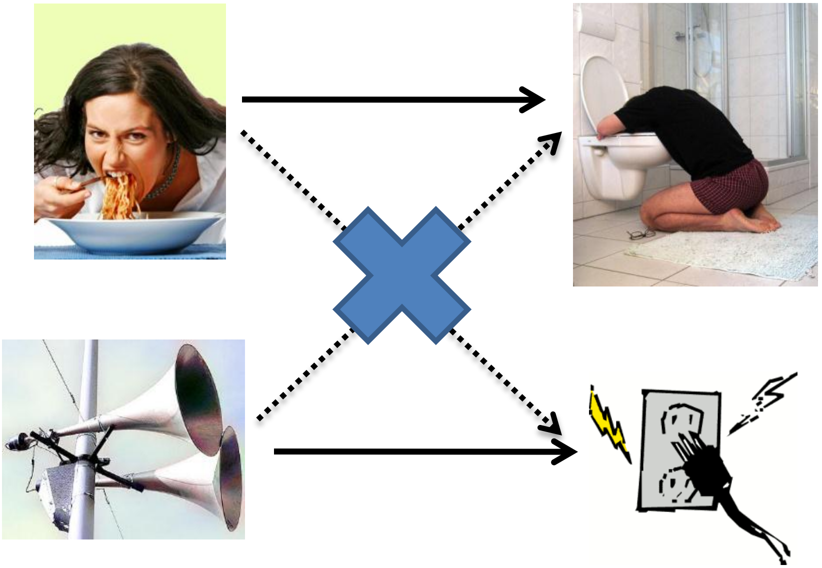
- One-trial learning
- Long-delay learning (few hours)
  - A [lack of] interference effect?
  - Still a problem for neuroscientists
- Hedonic shift: changes the CS, not its predictions



CTA

- Compound potentiation: odor + taste increase response to odor

- Preparedness:



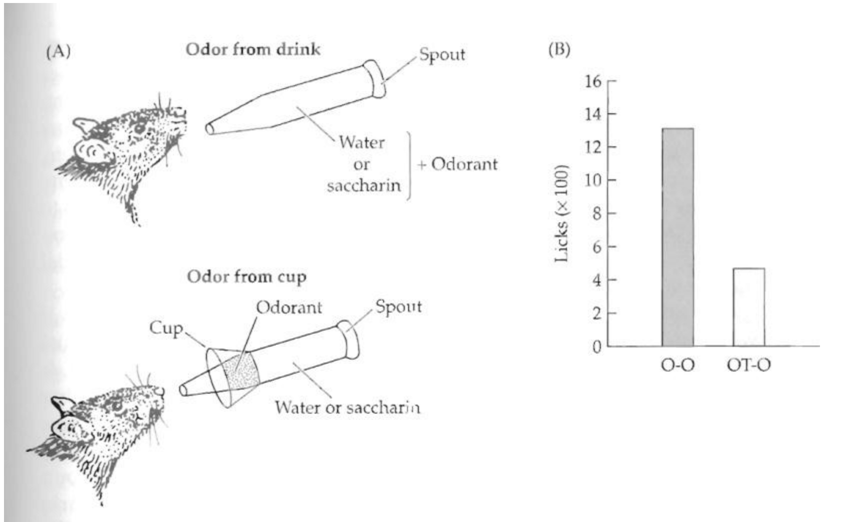
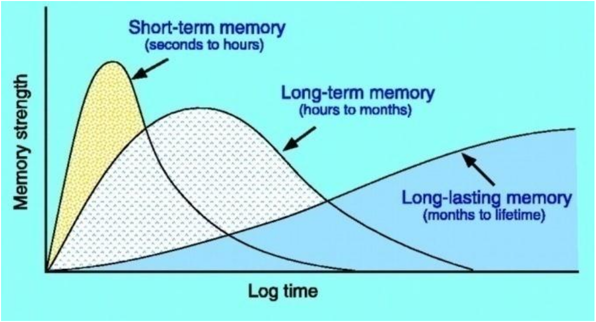
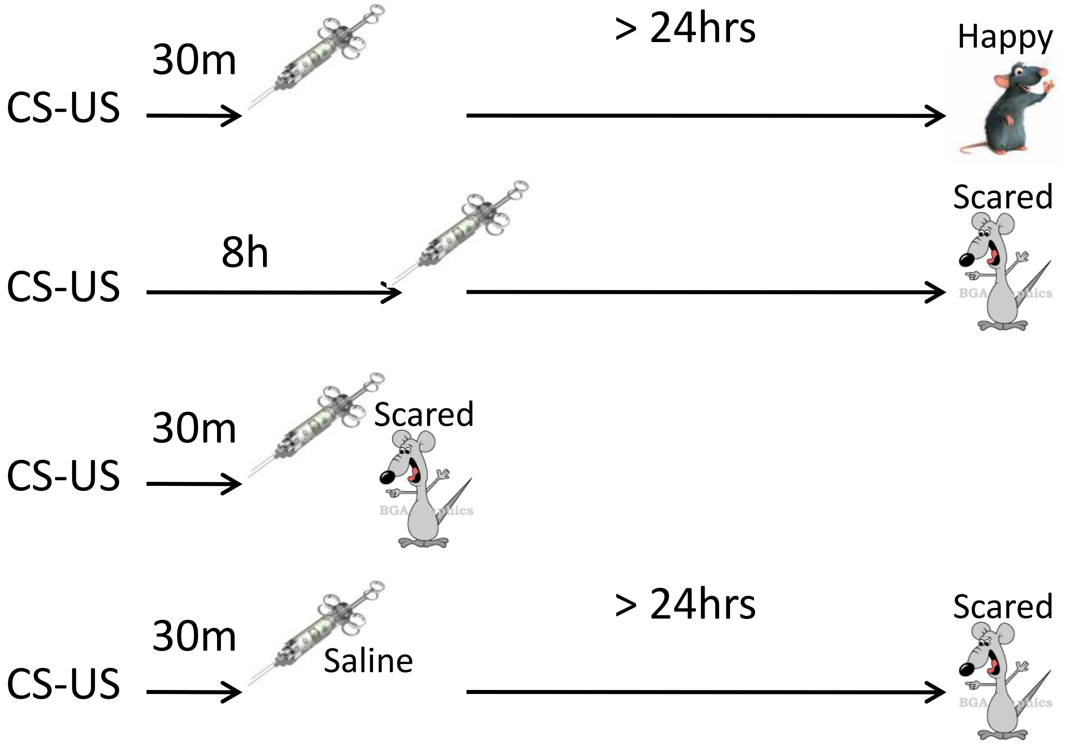


Figure 6.5 Potentiation of odor conditioning by taste. (A) A rat is given a taste in drinking water that also contains an odor. (In many experiments, the odor comes from a cup near the spout instead of being mixed in the drink.) (B) When odor is paired with illness on its own and then tested (O-O), it does not suppress consumption much. But if it has been combined with a taste on the conditioning trial (OT-O), strong odor conditioning is obtained. (A, after Inui, Shimura, & Yamamoto, 2006; B, after Palmerino et al., 1980.)

Consolidation

- Anisomycin, a protein synthesis inhibitor, into the Basolateral complex of the amygdala (BLA)
  - No effect on short-term-memory
  - No effect after XX time (rule of thumb is 6hrs)
  - But harms long-term memory below that.

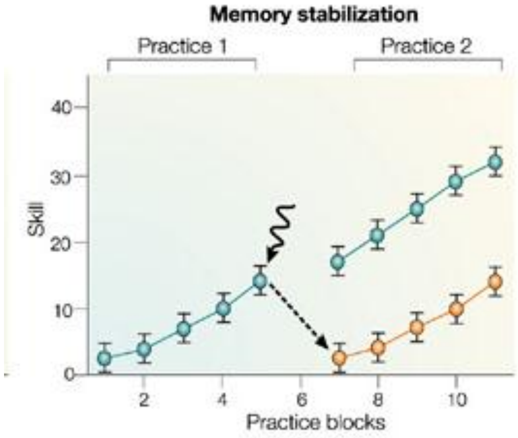



CS-US 30m > 24hrs Happy

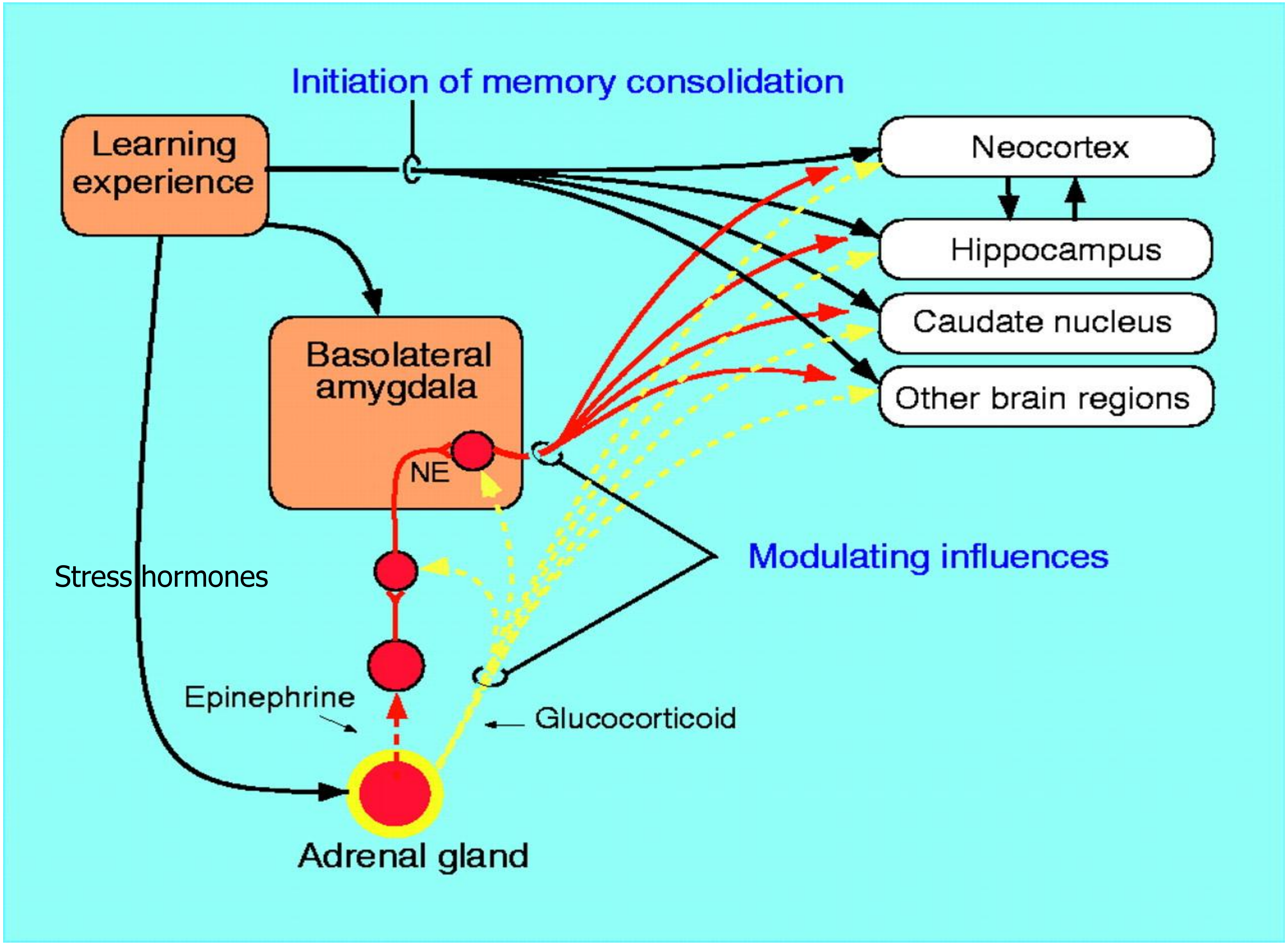
CS-US 8h Scared

CS-US 30m Scared

CS-US 30m Saline > 24hrs Scared



Nature Reviews | Neuroscience

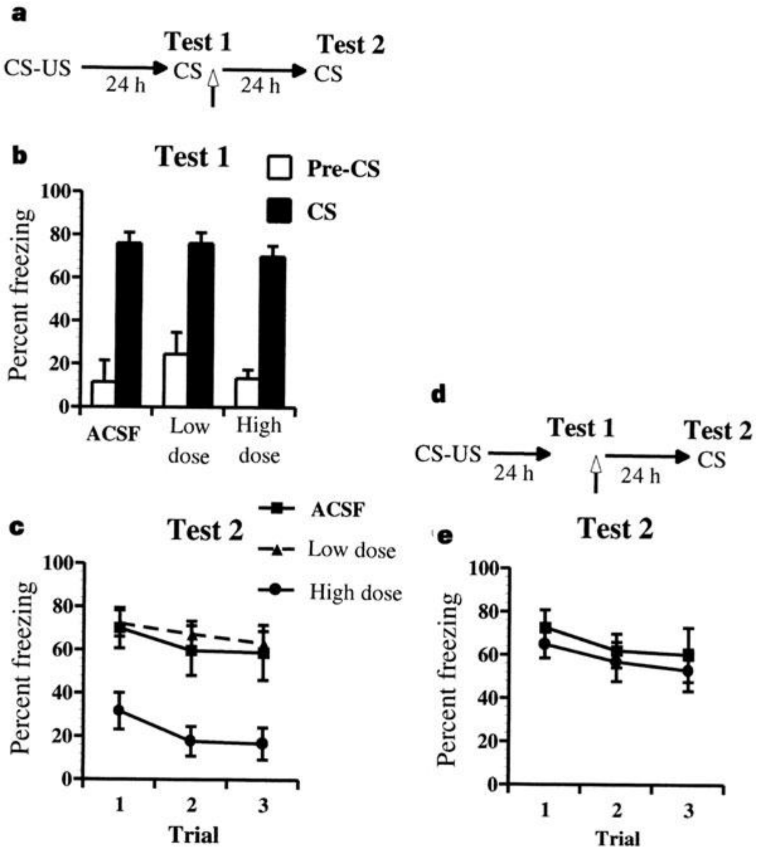


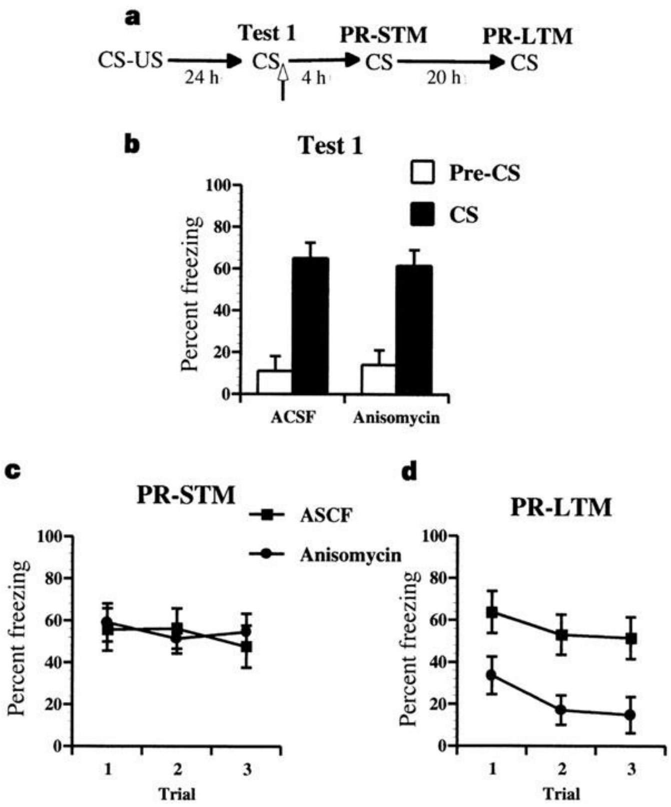
Stress hormones

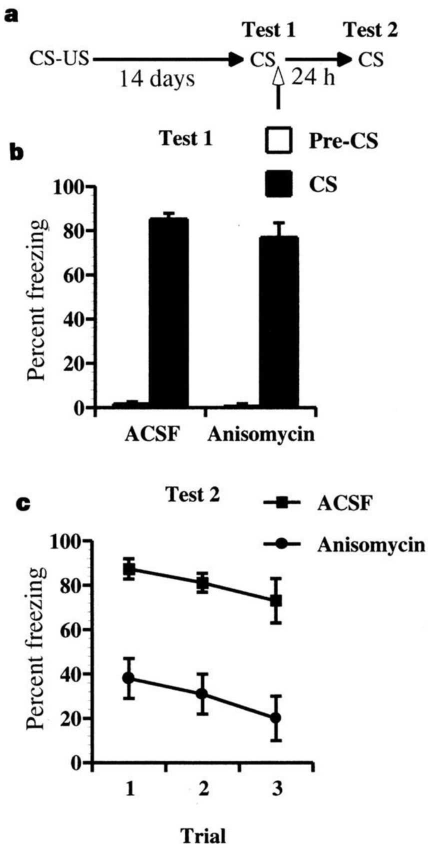
Mcgaugh JL, science, 2000

Reconsolidation

No effect on STM

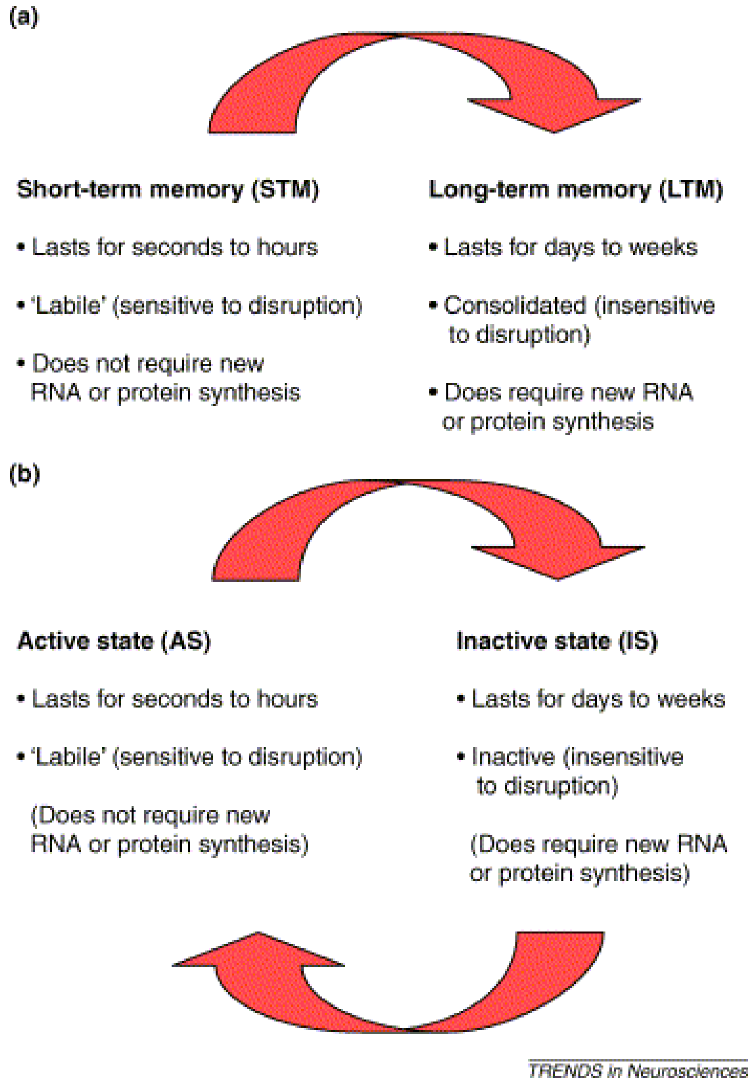






Nader, Ledoux, Nature 2000

An updated view of memories



(a)

Short-term memory (STM)

- Lasts for seconds to hours
- ‘Labile’ (sensitive to disruption)
- Does not require new RNA or protein synthesis

Long-term memory (LTM)

- Lasts for days to weeks
- Consolidated (insensitive to disruption)
- Does require new RNA or protein synthesis

(b)

Active state (AS)

- Lasts for seconds to hours
- ‘Labile’ (sensitive to disruption)

(Does not require new RNA or protein synthesis)

Inactive state (IS)

- Lasts for days to weeks
- Inactive (insensitive to disruption)

(Does require new RNA or protein synthesis)

TRENDS in Neurosciences