The Rescorla-Wagner Learning Model
(and one of its descendants)

Computational Models of Neural Systems

Lecture 5.1

David S. Touretzky

Based on notes by Lisa M. Saksida

November, 2019

Outline

- Classical and instrumental conditioning
- The Rescorla-Wagner model
  - Assumptions
  - Some successes
  - Some failures
- A real-time extension of R-W: Temporal Difference Learning
  - Sutton and Barto, 1981, 1990

Classical (Pavlovian) Conditioning

- CS = initially neural stimulus (tone, light, can opener)
  - Produces no innate response, except orienting
- US = innately meaningful stimulus (food, shock)
  - Produces a hard-wired response, e.g., salivation in response to food
- CS preceding US causes an association to develop, such that the CS will produce a CR (conditioned response)
- Allows the animal to learn temporal structure of its environment:
  - CS = sound of can opener
  - US = smell of cat food
  - CR = approach and/or salivation

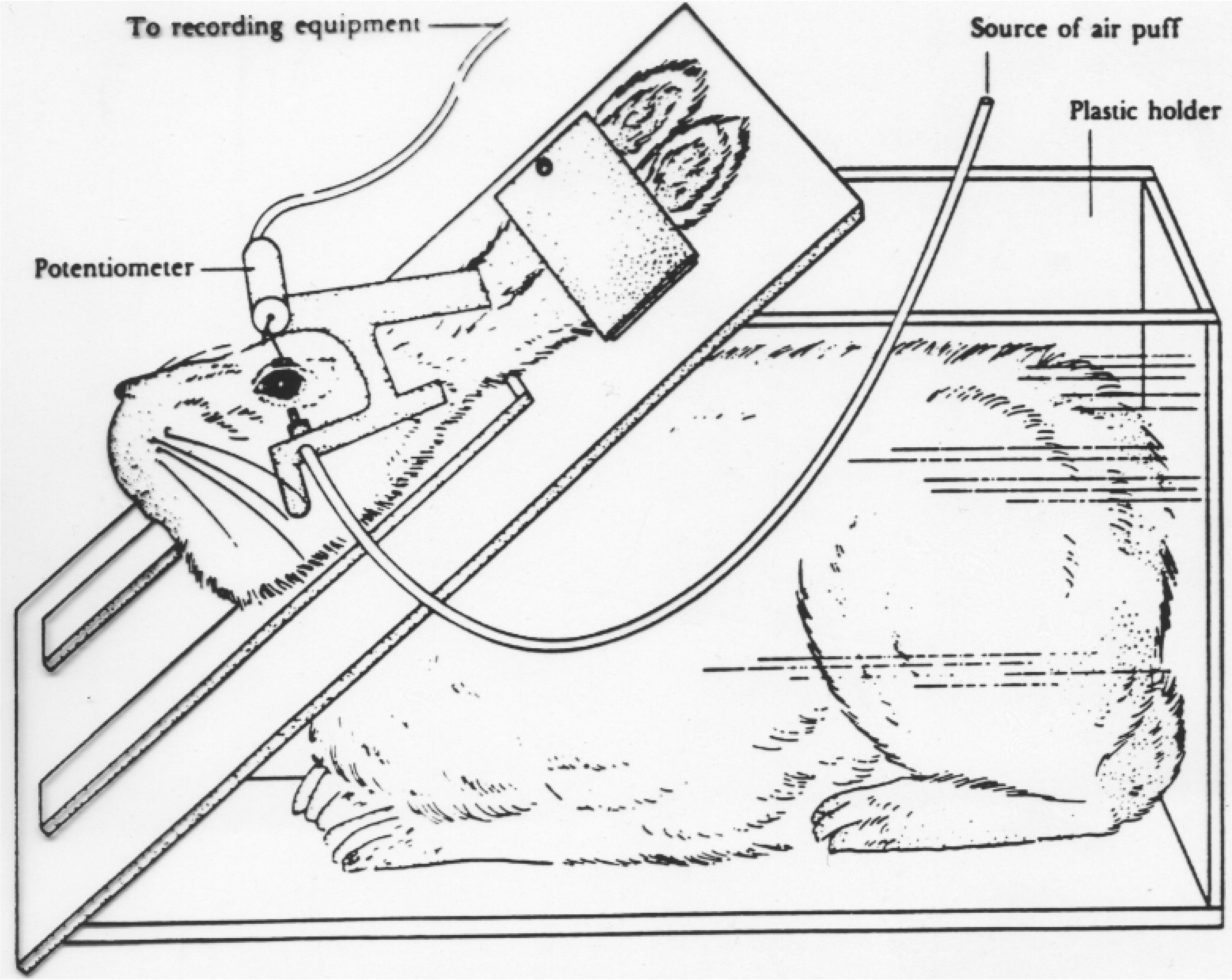
Learning in Simple Animals

- Classical conditioning has been demonstrated in invertebrates, such as the sea slug Aplysia.

Eric Kandel, 2000 Nobel Laureate

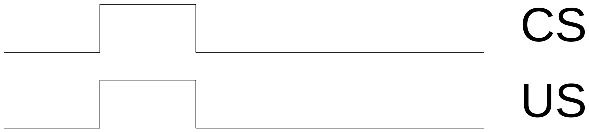
- What synaptic learning rules govern invertebrate learning?

Classical NMR (Nictitating Membrane Response) Conditioning

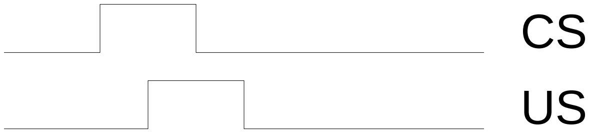


Excitatory Conditioning Processes

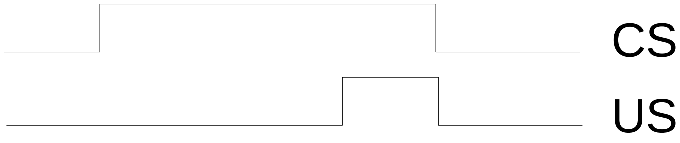
Simultaneous conditioning



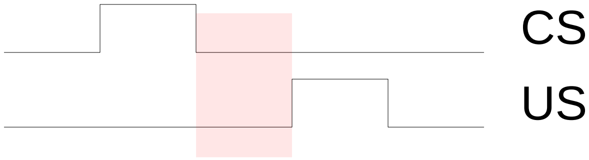
Short-delayed conditioning



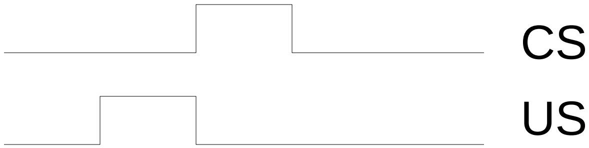
Long-delayed conditioning



Trace conditioning



Backwards conditioning



Instrumental (Operant) Conditioning

- Association between action (A) and outcome (O)
- Mediated by *discriminative stimuli* (lets the animal know when the contingency is in effect).
- Must wait for the animal to emit the action, then reinforce it.
- Unlike Pavlovian CR, the action is voluntary.
- Training a dog to sit on command:
  - Discriminative stimulus: say "sit"
  - Action = dog eventually sits down
  - Outcome = food or praise

The Rescorla-Wagner Model

- Trial-level description of changes in associative strength between CS and US, i.e., how well CS predicts US.
- Learning happens when events violate expectations, i.e., amount of reward/punishment differs from prediction.
- As the discrepancy between predicted and actual US decreases, less learning occurs.
- First model to take into account the effects of multiple CSs.

Rescorla-Wagner Learning Rule

$$\bar{V} = \sum V_i X_i \qquad \Delta V_i = \alpha_i \; \beta \; \left(\lambda - \bar{V}\right) X_i$$

$\bar{V}$ = strength of response

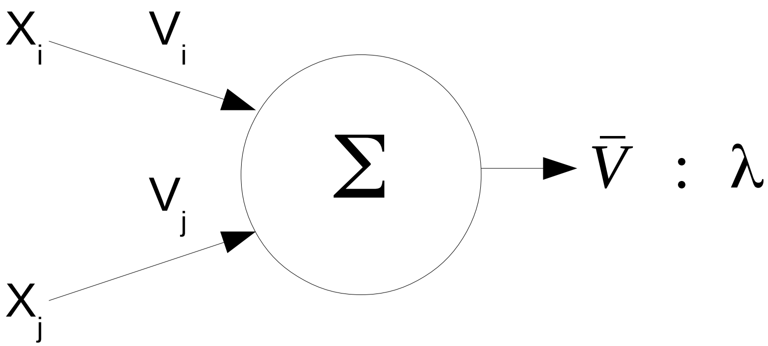
V_i = associative strength of CS *i* (predicted value of US)

X_i = presence of CS *i*

α_i = innate salience of CS *i*

β = associability of the US

λ = strength (intensity and/or duration) of the US

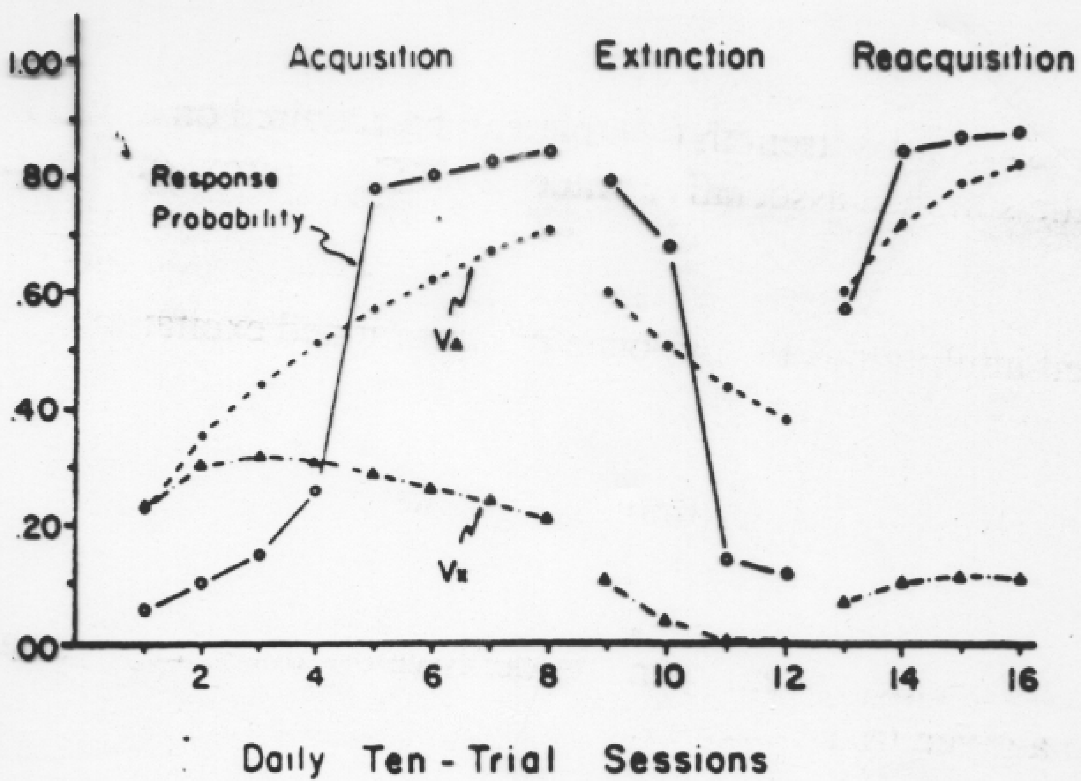


This is essentially the same as the LMS or CMAC learning rules.

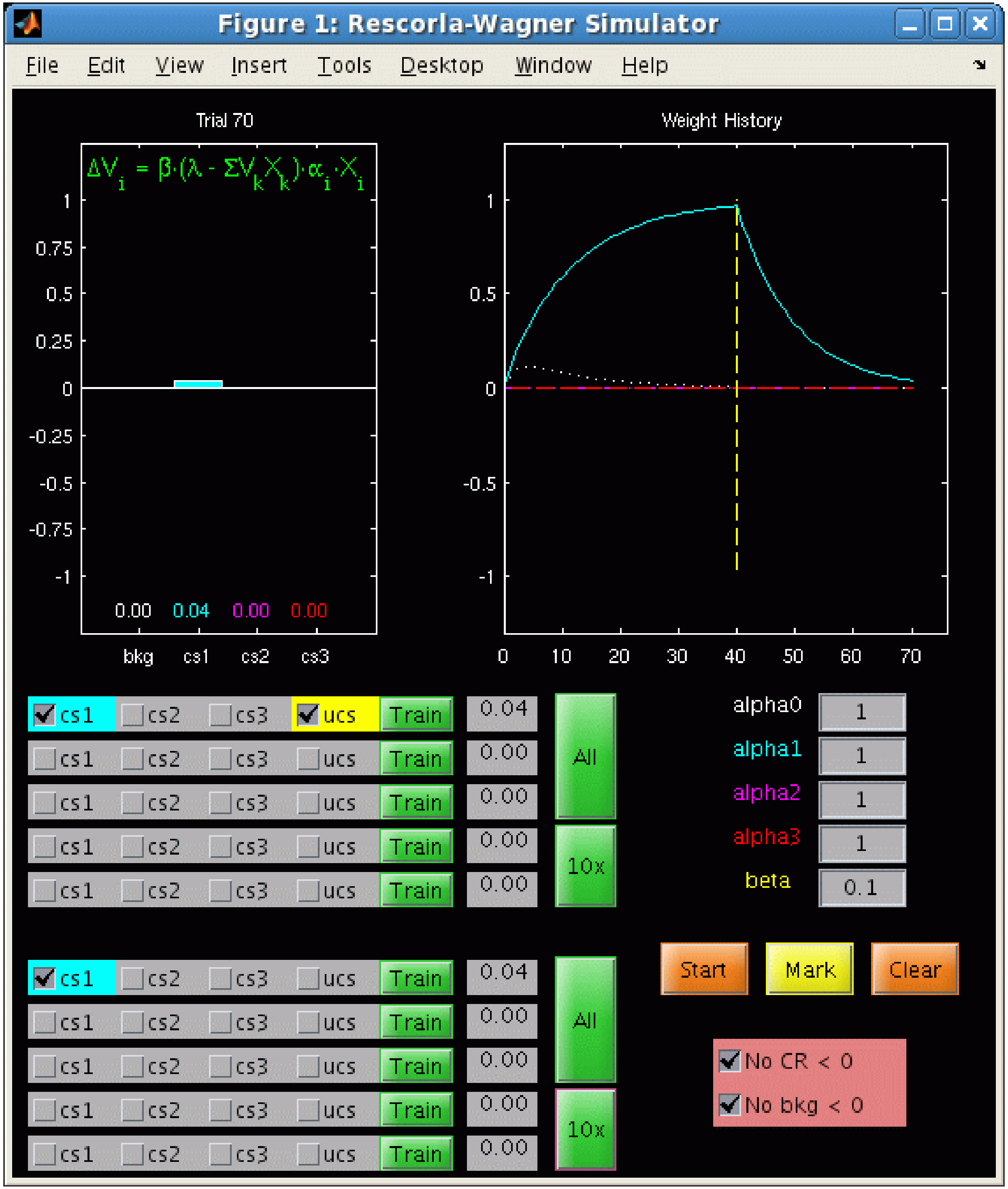
Rescorla-Wagner Assumptions

1. Amount of associative strength $\overline{V}$ that can be acquired on a trial is limited to the summed associative values of all CSs present on the trial.
2. Conditioned inhibition is the opposite of conditioned excitation.
3. Salience (α_ι) of a stimulus is constant.
4. New learning is independent of the associative history of any stimulus present on a given trial.
5. Monotonic relationship between learning and performance, i.e., associative strength ($\overline{V}$) is monotonically related to the observed CR.

Success: Acquisition/Extinction Curves



- **Acqusition:** deceleration of learning as (λ – $\overline{V}$) decreases
- **Extinction:** loss of responding to a trained CS after non-reinforced CS presentations
  - RW assumes that $\lambda = 0$ during extinction, so extinction is explained in terms of absolute loss of $\overline{V}$.
  - See later why this is not an adequate explanation.



☑ cs1 ☐ cs2 ☐ cs3 ☑ ucs Train 0.04

☐ cs1 ☐ cs2 ☐ cs3 ☐ ucs Train 0.00

☐ cs1 ☐ cs2 ☐ cs3 ☐ ucs Train 0.00

☐ cs1 ☐ cs2 ☐ cs3 ☐ ucs Train 0.00

☐ cs1 ☐ cs2 ☐ cs3 ☐ ucs Train 0.00

☑ cs1 ☐ cs2 ☐ cs3 ☐ ucs Train 0.04

☐ cs1 ☐ cs2 ☐ cs3 ☐ ucs Train 0.00

☐ cs1 ☐ cs2 ☐ cs3 ☐ ucs Train 0.00

☐ cs1 ☐ cs2 ☐ cs3 ☐ ucs Train 0.00

☐ cs1 ☐ cs2 ☐ cs3 ☐ ucs Train 0.00

alpha0 1

alpha1 1

alpha2 1

alpha3 1

beta 0.1

☑ No CR < 0

☑ No bkg < 0

Success: Stimulus Generalization/Discrimination

- **Generalization** between two stimuli increases as the number of stimulus elements common to the two increases.
- **Discrimination:**
  - Two similar CSes presented: CS+ with US, and CS- with no US
  - Subjects initially respond to both, then reduce responding to CS– and increase response to CS+
  - Model assumes some stimulus elements are unique to each CS, and some are shared.
  - Initially, all CS+ elements become excitatory, causing generalization to CS–
  - Then CS– elements become inhibitory; eventually common elements become neutral.

Success: Overshadowing and Blocking

- **Overshadowing:**
  - Novel stimulus A presented with novel stimulus B and a US.
  - Testing on A produces smaller CR than if A were trained alone.
  - Greater overshadowing by stimuli with higher salience (α_i).

- **Blocking:**
  - Train on A plus US until asymptote
  - Then present A and B together plus US
  - Test with B: find little or no CR
  - Pre-training with A causes US to “lose effectiveness”.

- **Unblocking with increased US:**
  - When intensity of US is increased, unblocking occurs.

Success: Patterning

- Positive patterning:

  A → no US

  B → no US

  AB → US

- Discrimination solved when animal responds to AB but not to A or B alone.
- Rescorla-Wagner solves this with a hack:
  - Compound stimulus consists of 3 stimuli: A, B, and X (configural cue)
  - X is true whenever A and B are both true
- After many trials, X has all the associative strength; A and B have none.

Success: Conditioned Inhibition

- “Negative summation” and “retardation” are tests for conditioned inhibitors.
- **Negative summation** test: CS passes if presenting it with a conditioned exciter reduces the level of responding.
  - R-W: this is due to the negative V of the CS summing with the positive value of the exciter.
- **Retardation** test: CS passes if it requires more pairings with the US to become a conditioned exciter than if the CS were novel.
  - R-W: inhibitor starts the training with a negative V, so it takes longer to become an exciter than if it had started from 0.

Success: Relative Validity of Cues

- AX → US and BX → no US
  - X becomes a weak elicitor of conditioned response
- AX → US on ½ of trials and BX → US on ½ of trials
  - X becomes a strong elicitor of conditioned responding
- In both cases, X has been reinforced on 50% of presentations.
  - In the first condition, A gains most of the associative strength because X loses strength on BX trials, is then reinforced again on AX trials.
  - In the second condition, A and B are also reinforced on only 50% of presentations so they don't overpower X, which is seen twice as often.
- Rescorla-Wagner model is successful if β for reinforced trials is greater than β for non-reinforced trials.

Failure 1: Recovery From Extinction

1. Spontaneous recovery (seen over long retention intervals).
2. External disinhibition: temporary recovery when a physically intense neutral CS precedes the test CS.
3. Reminder treatments: present a cue from training (either CS or US) without providing a complete trial.

- Recovery of a strong but extinguished association usually leads to a stronger response, which suggests that extinction is not due to a permanent loss of associative strength.
- Failure is due to the assumption of “path independence”: that subjects know only the current associative strengths and retain no knowledge of past associative history.

Failure 2: Facilitated and Retarded Reacquisition After Extinction

- Reacquisition is usually much faster than initial learning.
- Could be due to residual CS-US association: R-W can handle this if we add a threshold for behavioral response.
- Retarded acquisition has been seen – due to massive overextinction (continued trials after responding has stopped.)
- Retarded reaqcuisition is inconsistent with the R-W prediction that an extinguished association should be reacquired at the same rate as a novel one.
- Another example of the (incorrect) path independence assumption.

Failure 3: Failure to Extinguish A Conditioned Inhibitor

- R-W predicts that V for both conditioned exciters and inhibitors moves toward 0 on non-reinforced presentations of the CS.
- However, presentations of a conditioned inhibitor alone either have no effect, or increase its inhibitory potential.
- Failure of the theory is due to the assumption that extinction and inhibition are symmetrical opposites.
- Later we will see a simple solution to this problem.

Failure 4: CS-Preexposure (Latent Inhibition)

- Learning about a CS occurs more slowly when the animal has had non-reinforced pre-exposure to it.
- Seen in both excitatory and inhibitory conditioning, so it is not due to acquisition of inhibition.
- R-W predicts that, since no US is present during pre-exposure, no learning should occur.
- Usual explanation: slower learning is due to a decrease in α_i (salience of CS), but R-W says this value is constant.
- Failure due to the assumption of fixed associability (α_ι and β are constants).

Failure 5: Second-Order Conditioning

1. Train A → US until asymptote
2. Train B → A
3. Test B → CR ?

- R-W: because B → A trials do not involve US presentation, B should become a conditioned inhibitor.
- Subjects expect US because of the presentation of A, so V is positive and λ is 0 so V decreases.
- Not due to any one assumption of the model – it would have to undergo major revisions to account for this.

Extensions to the Rescorla-Wagner Model

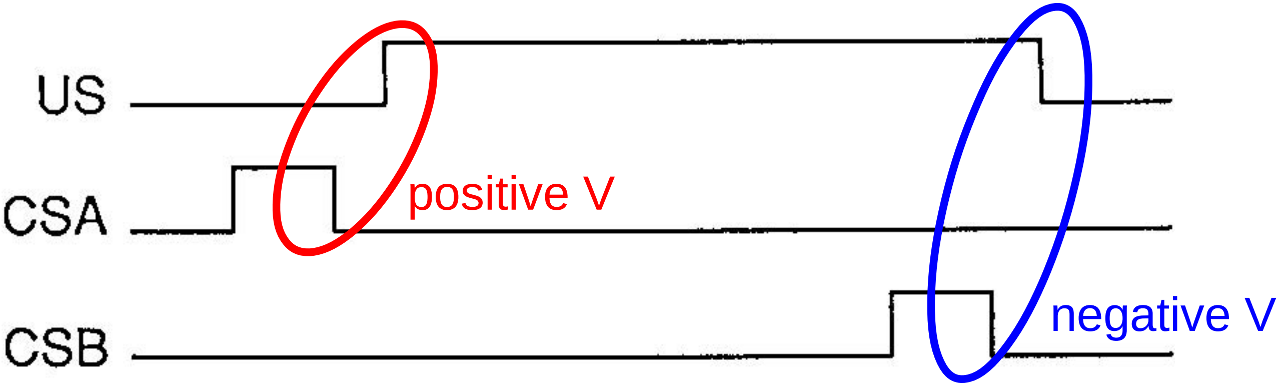
1. Changing attention to simulus (α) can model latent inhibition
   - Pearce and Hall (1980), Mackintosh (1975)

2. Learning-performance distinction (nonlinear mapping from associative strength to CR)
   - Bouton (1993), Miller and Mazel (1988)

3. Within-trial processing can model ISI and 2nd order effects
   - Wagner (1981), Sutton and Barto (1981, 1990)

Real-Time Models

- Updated at every time step.
- Stimulus trace models originated by Hull (1939) – internal representation of CS persists for several seconds.
- Can look at within-trial effects (e.g., λ varies within trials – US produces opposite signed reinforcement at onset and offset.)|



- Key idea: changes in US level determine reinforcement.

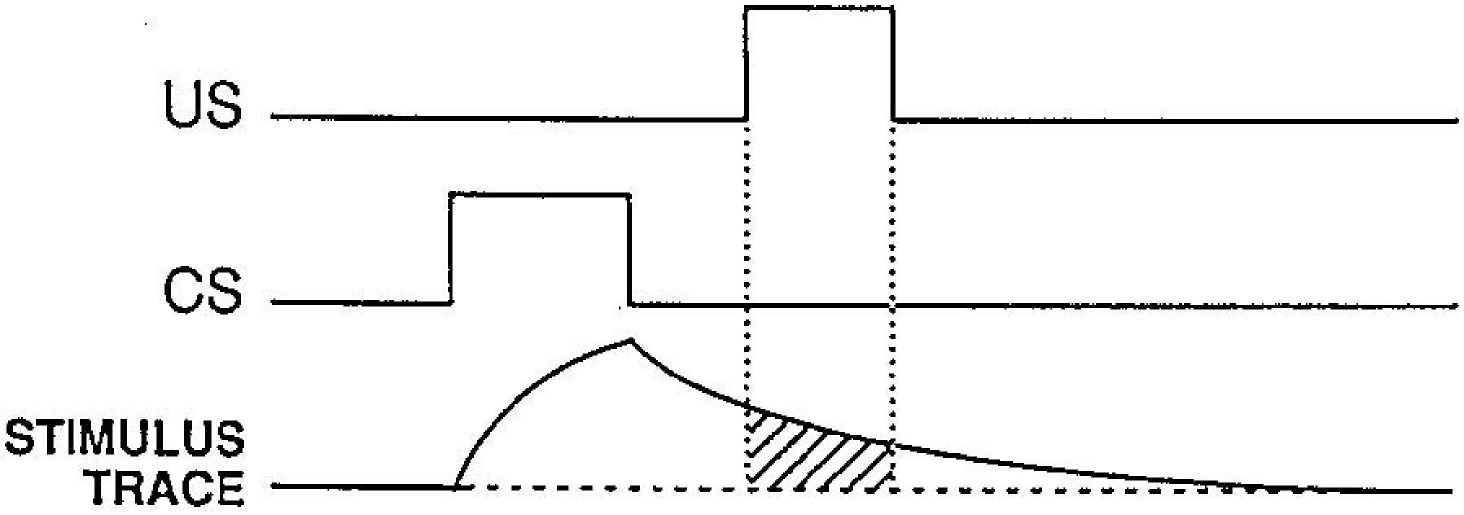
Real-Time Theory of Reinforcement $(\dot{Y})$

- Assume all stimuli generate reinforcement at onset (+) and at offset (–).
- Y(t) = sum of all V's – this changes across the trial as stimuli are added and removed.
- $\dot{Y}$ is the change in Y over time: $\dot{Y}(t) = Y(t) - Y(t-\Delta t)$
- If all CSs have simultaneous onsets and offsets, we have R-W.
- CS onset yields no learning because reinforcement precedes the CS.
- CS offset coincides with US onset so $V = V_{US} - \bar{V}$
- Negative reinforcement from US offset is not a problem as long as US is long and has poor temporal correlation with the CS.

Real-Time Theory of Eligibility

- Trace interval = interval between CS and US when no stimuli are present.
  - Conditioning takes longer as this interval increases.

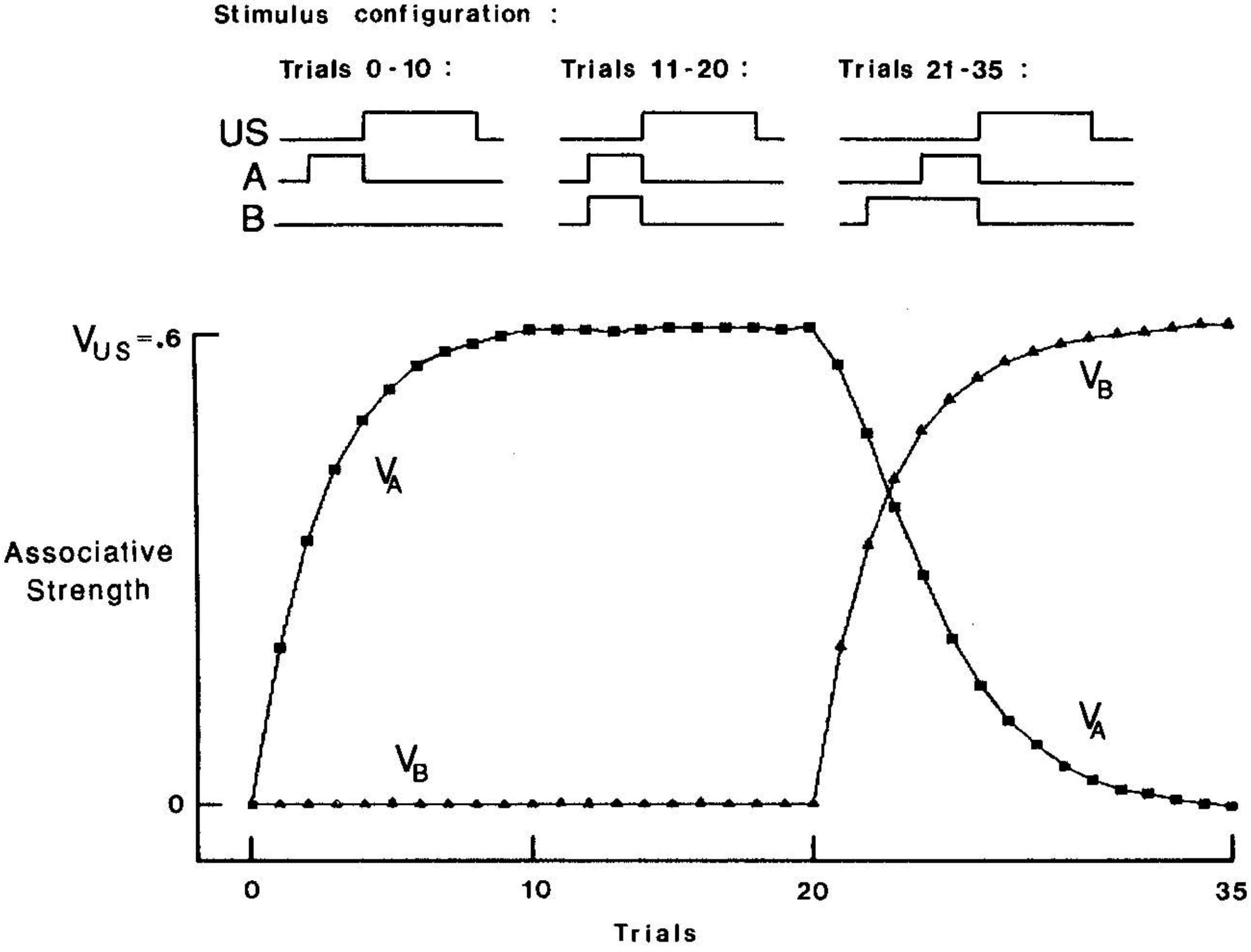
- Sutton and Barto use an *eligibility trace* as the CS representation:



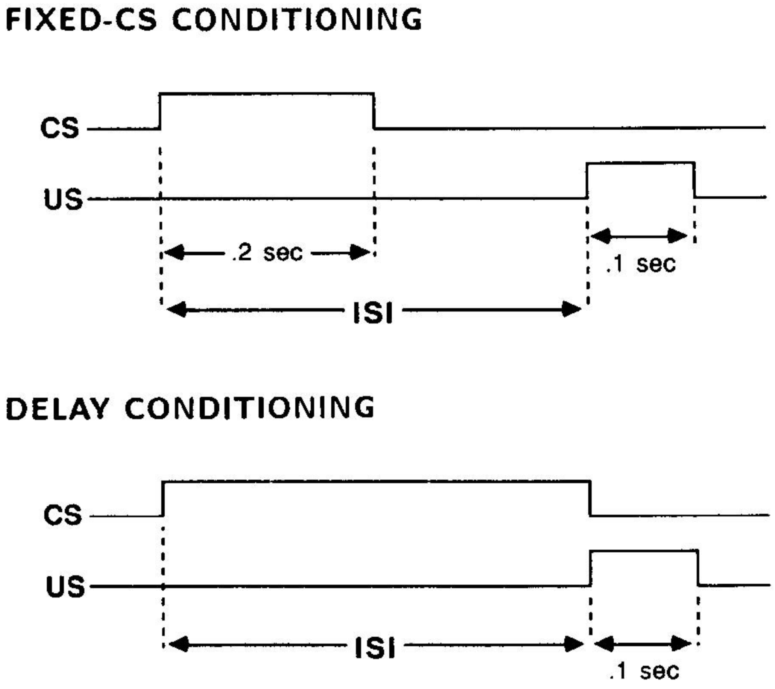
Sutton and Barto (1981)

- This model can produce effects that the Rescora-Wagner model cannot capture.

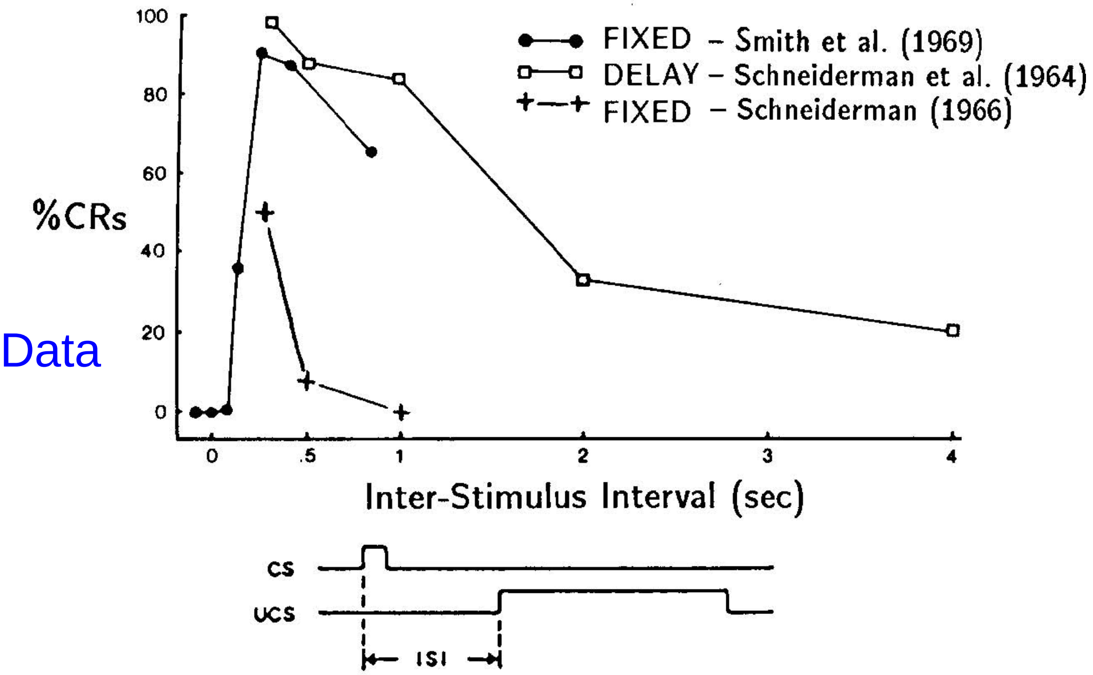
$$\Delta V_i = \beta \dot{Y} \cdot \alpha_i \bar{x}_i$$



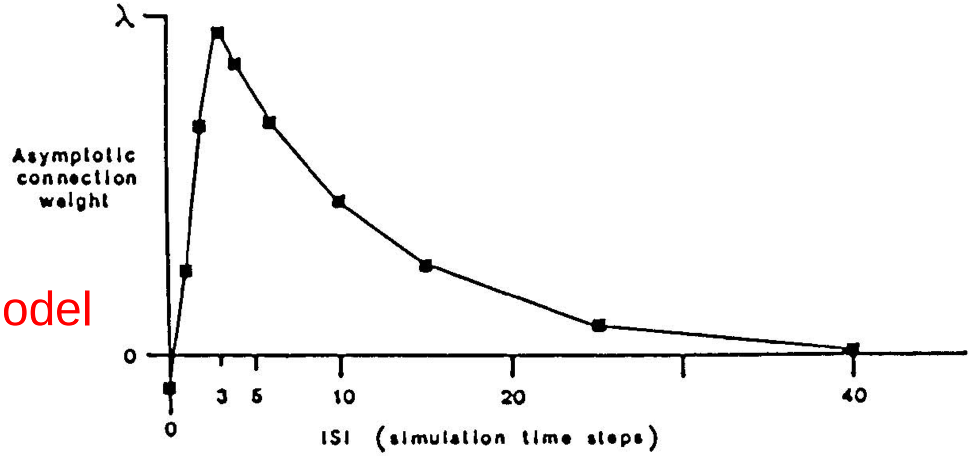
Reproducing ISI Effects: Rabbit NMR



Data

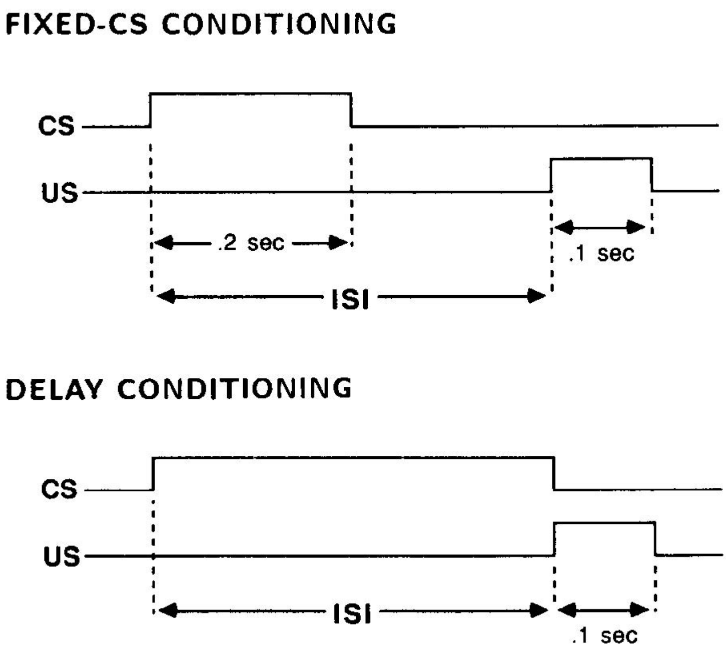


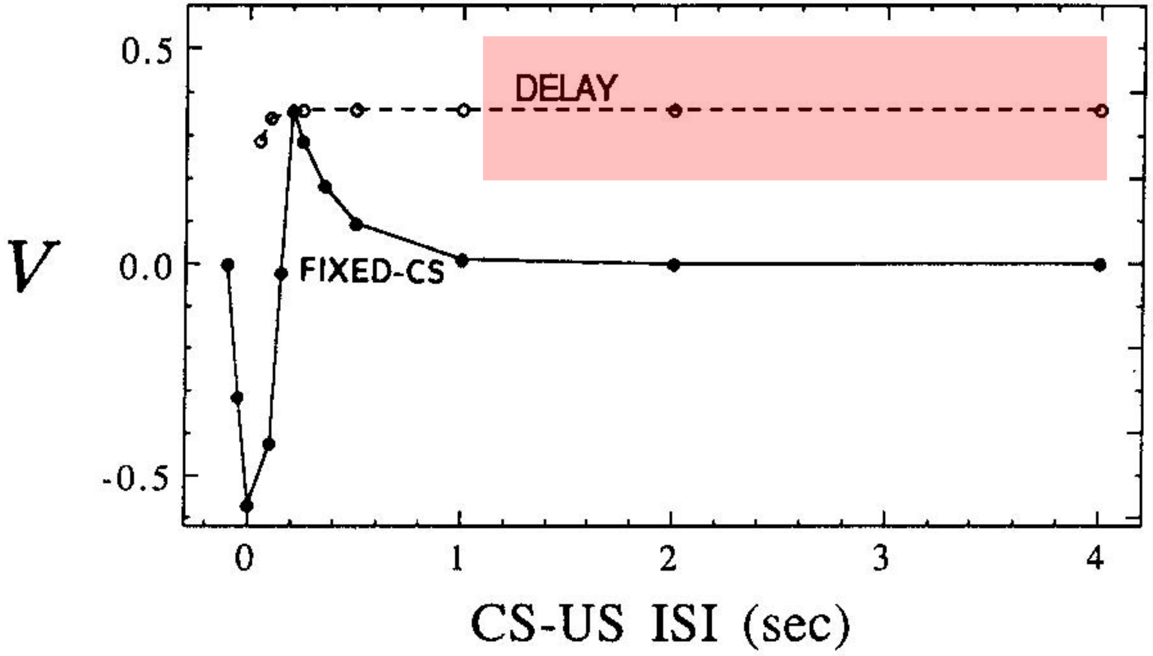
Model



Using a Realistic US Causes Problems

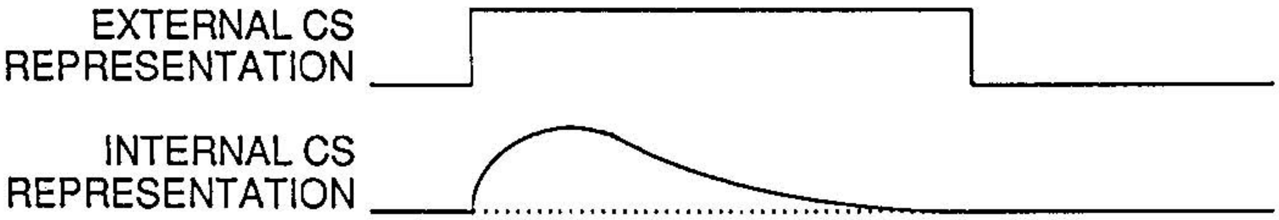
- Sutton and Barto originally used a 1500 msec US, but a typical US in real experiments lasts 100 msec.
- But using a more realistic US results in delay conditioning problems.

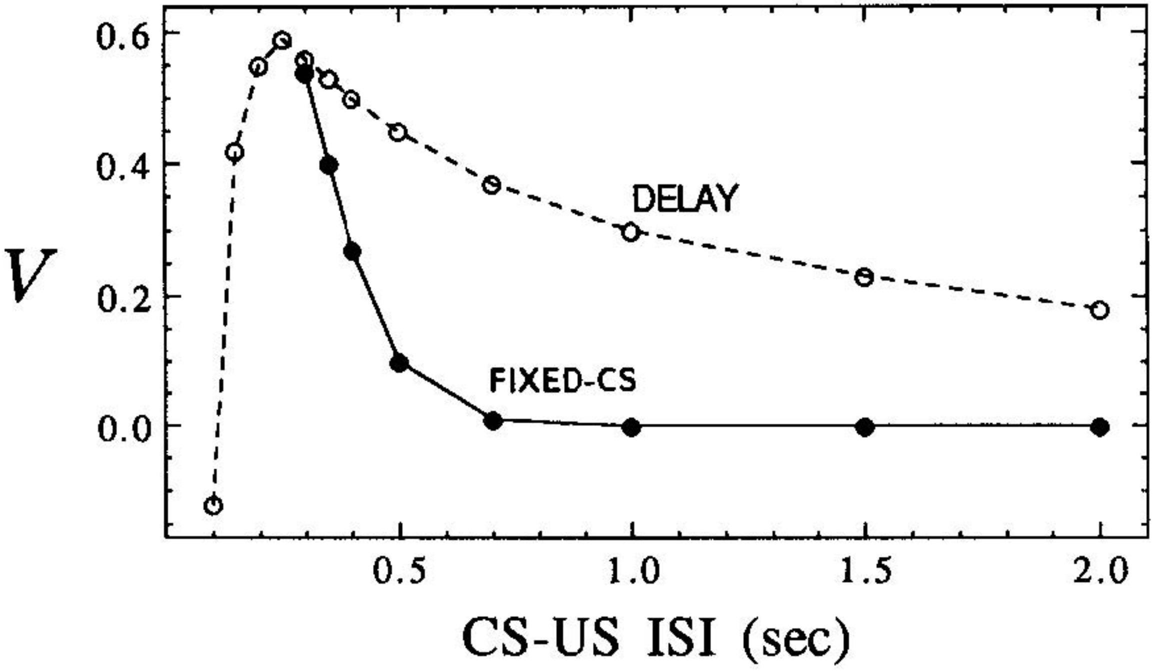




Fixing the Delay Problem

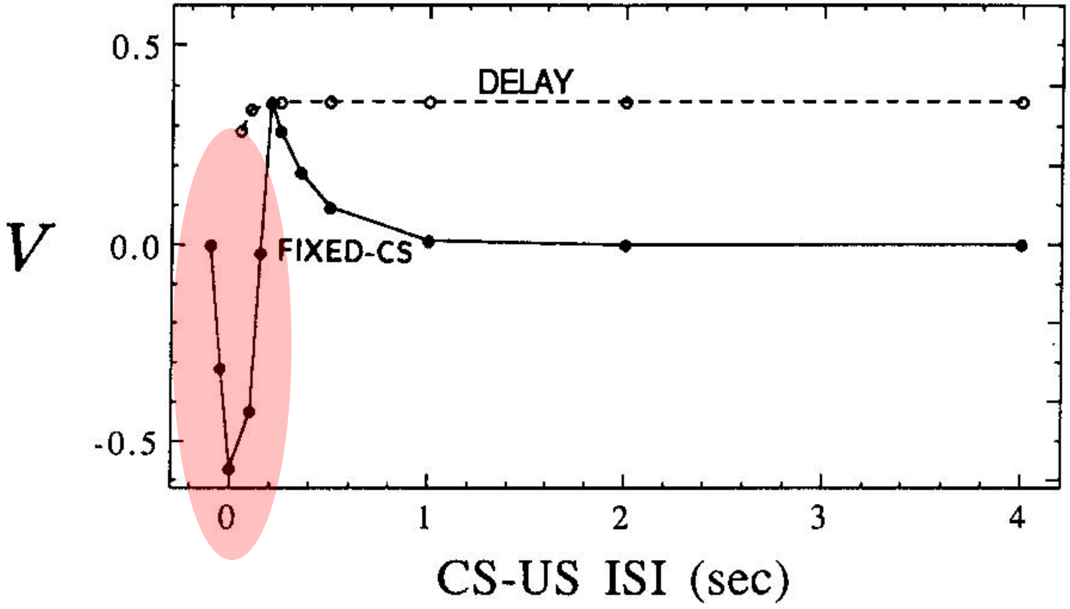
- Assuming an internal CS that decreases with time fixes the delay problem.





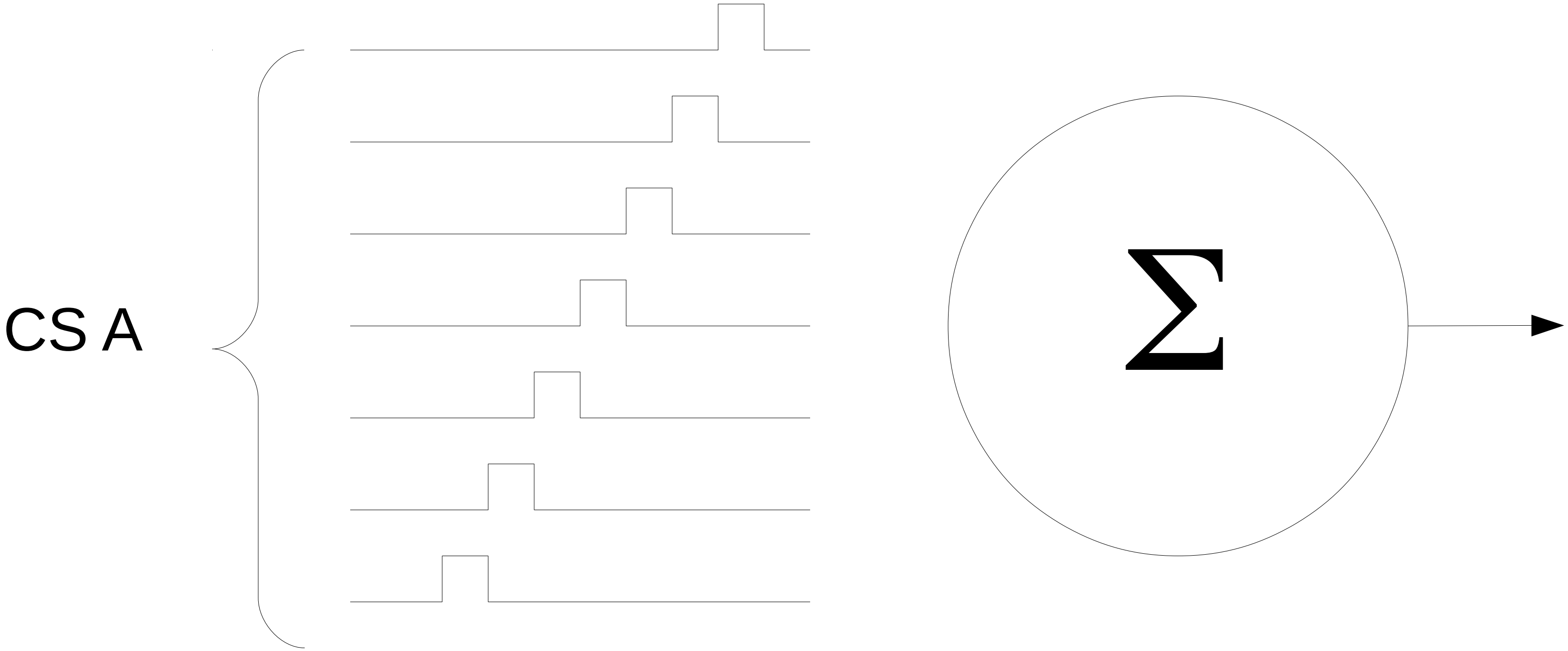
Fixed CS Also Has A Problem

- According to the SB model, inhibition is predicted whenever the CS and the US overlap because of the good temporal relationship between the CS and the US offset.
- This causes inhibitory conditioning for small ISIs and for backward conditioning, but the animal data show mild excitatory conditioning in these situations.



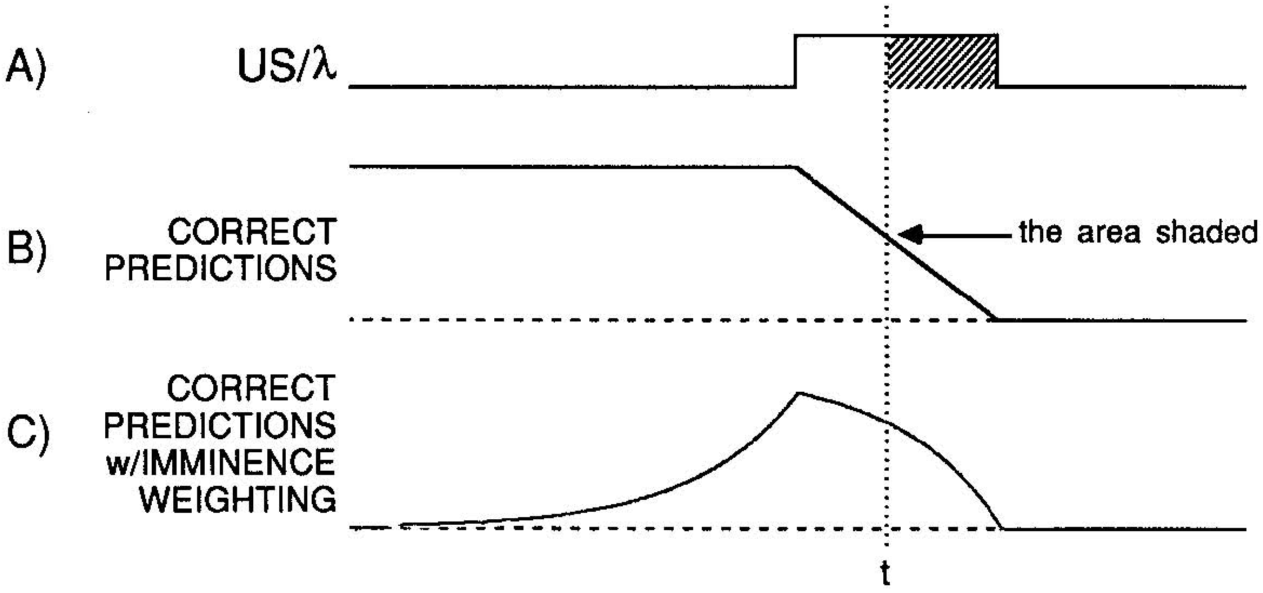
Complete Serial Compound (CSC) Stimuli

- Compound stimulus covers the entire intra-trial interval.
- Used in an alternative to the Y-dot model.



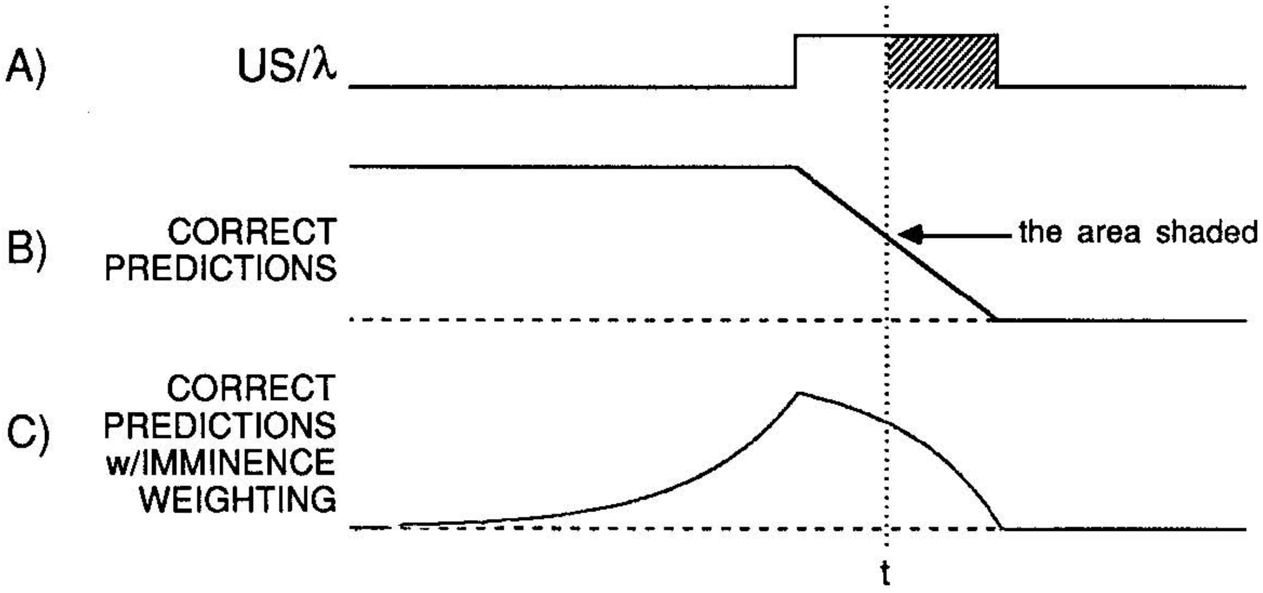
Temporal Difference Model

- Stimuli are assumed to be CSCs
- λ changes over the trial, and the area under the curve reflects the total primary reinforcement for the trial (A).
- $\bar{V}$ is the animal's prediction of the area under the λ curve, at each time step predicting only future λ's area (B).

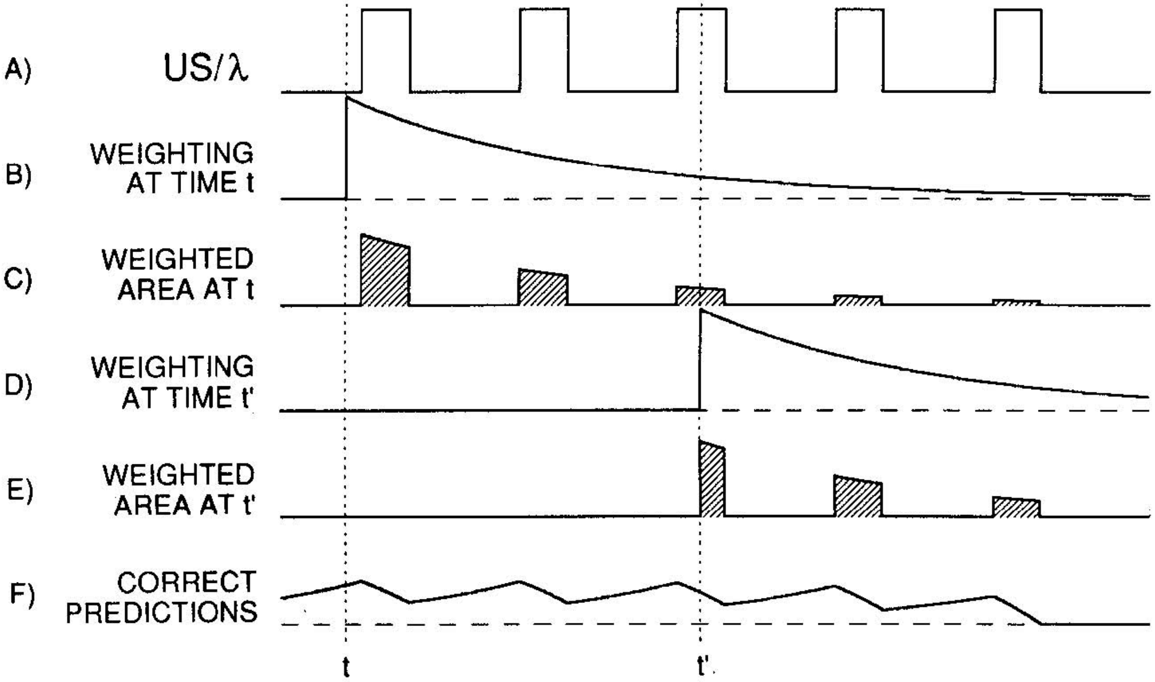


Imminence Weighting

- The prediction is equally high for all times prior to the US, but animals learn weaker associations for stimuli presented far in advance. Also, temporally remote USs should be discounted. So upcoming reinforcement should be weighted according to its **imminence** (C).



Effect of Imminence Weighting



Undiscounted: $\bar{V}_t = \lambda_{t+1} + \lambda_{t+2} + \lambda_{t+3} + \ldots$

Discounted: $\bar{V}_t = \lambda_{t+1} + \gamma\lambda_{t+2} + \gamma^2\lambda_{t+3} + \ldots$

Derivation of Reinforcement Term

- US prediction at time *t* is the sum of discounted future reinforcement:

$$\bar{V}_t = \lambda_{t+1} + \gamma\left(\lambda_{t+2} + \gamma\lambda_{t+3} + \gamma^2\lambda_{t+4} + \ldots\right)$$

- US prediction at time *t*+1 is the second term of the above:

$$\bar{V}_{t+1} = \lambda_{t+2} + \gamma\lambda_{t+3} + \gamma^2\lambda_{t+4} + \ldots$$

- The desired prediction for time *t* is in terms of reinforcement received and prediction made at *t*+1:

$$\bar{V}_t = \lambda_{t+1} + \gamma\ \bar{V}_{t+1}$$

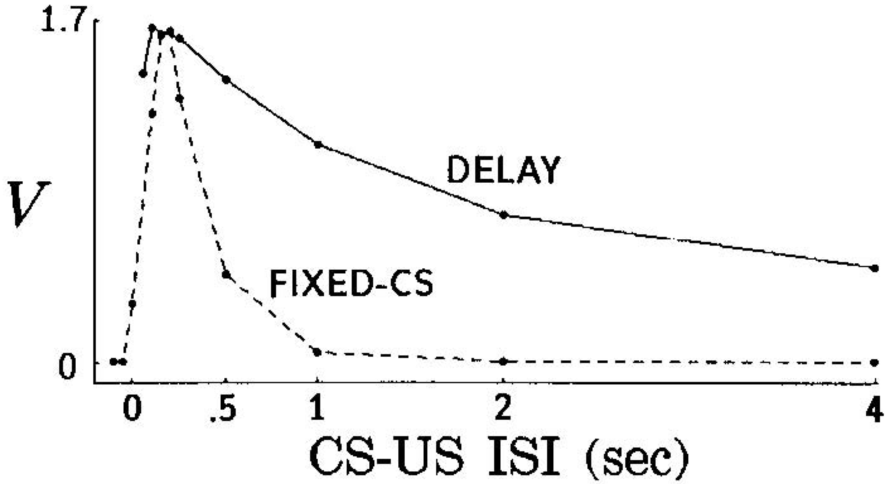
Temporal Difference Learning

- Discrepancy between the two terms is the prediction error δ:

$$\delta = (\lambda_{t+1} + \gamma \bar{V}_{t+1}) - \bar{V}_t$$

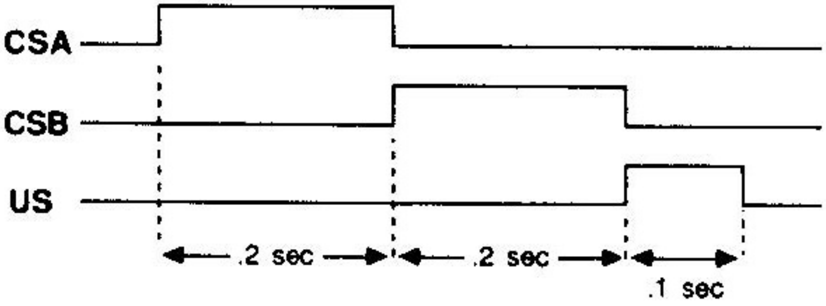
TD Learning Model

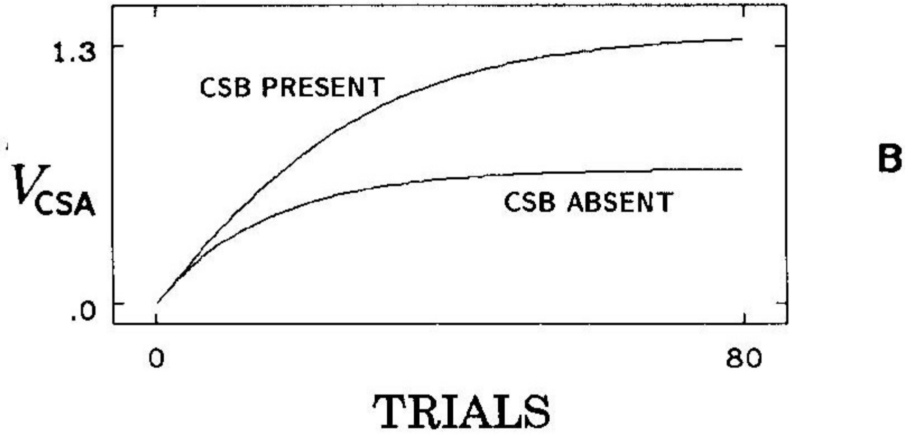
$$\Delta V_i = \beta(\lambda_{t+1} + \gamma \bar{V}_{t+1} - \bar{V}_t) \cdot \alpha_i \bar{x}_i$$



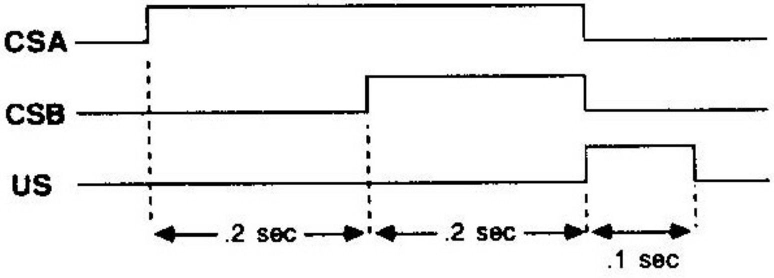
Modeling Within-Trial Effects

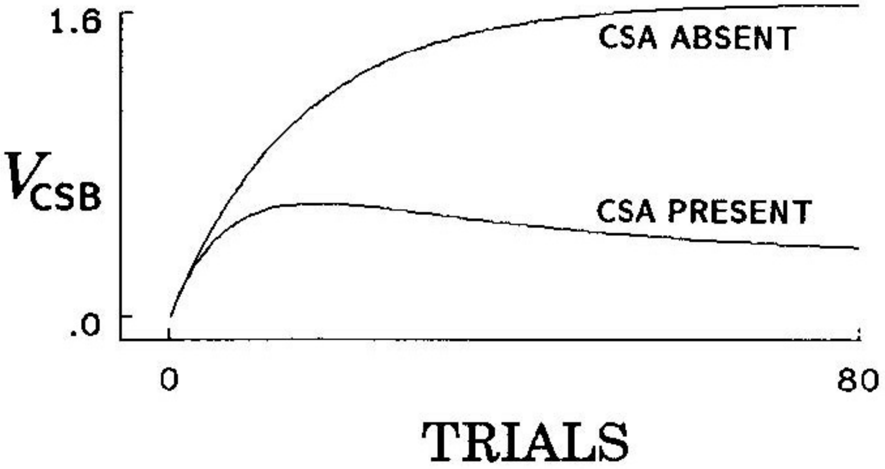
- If the ISI is filled with a second stimulus, learning is facilitated:





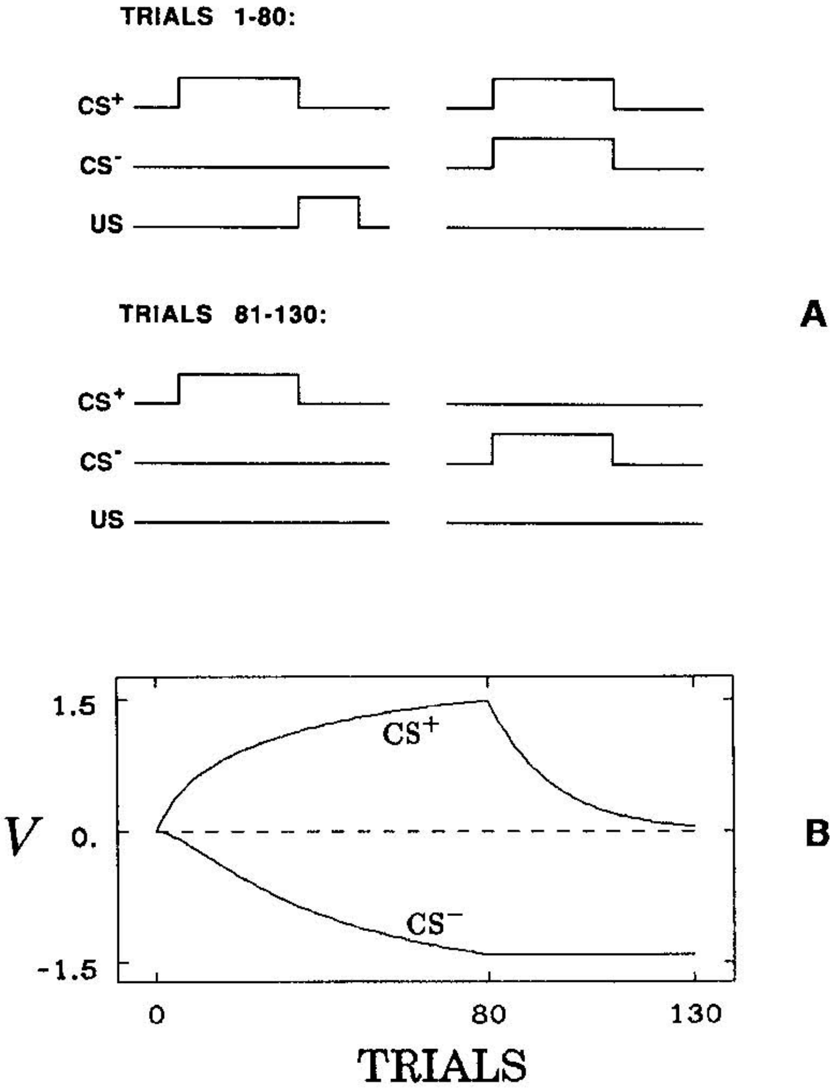
- Primacy effect: B is closer to the US, but presence of A reduces conditioning to B:





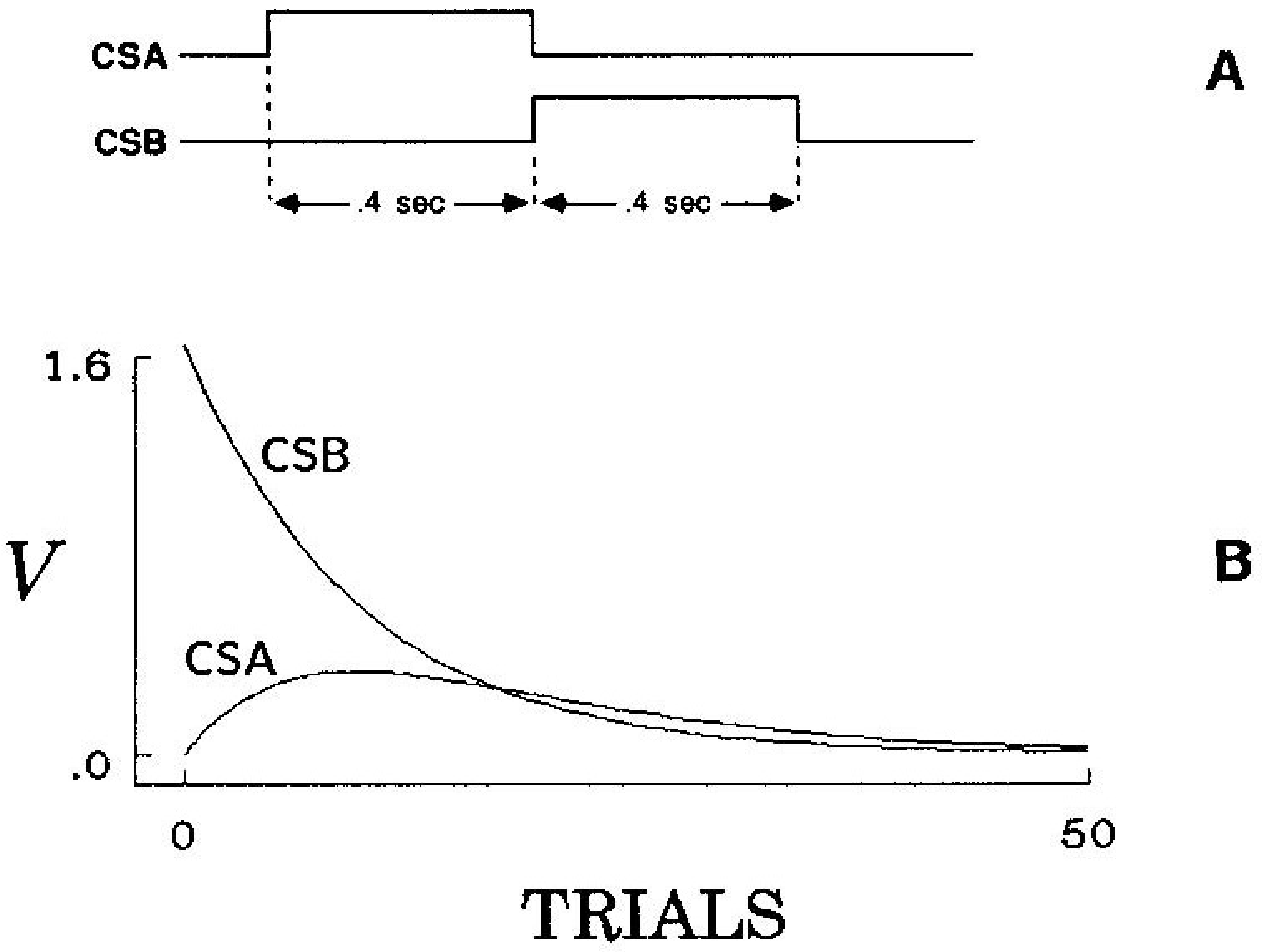
Failure to Extinguish a Conditioned Inhibitor

Assume the sum of Vs is constrained to be non-negative:



Second-Order Conditioning

- Train B → US, then train A → B



Summary

- The Rescorla-Wagner model is the prototypical example of a computational model in psychology.
  - Not intended as a neural-level model, although it uses one "neuron".
- It is very abstract, but that is a strength.
  - Neatly captures a variety of important conditioning phenomena in just two equations.
  - Makes testable predictions (some of which falsify the model).
- Sutton & Barto's TD learning extends R-W to the temporal domain, but is in other respects still very abstract.
- Is there TD learning in the brain?
  - Possibly in basal ganglia.