ERROR-DRIVEN LEARNING

for modeling language acquisition

Dorothée Hoppe & Jacolien van Rij
October 2019, Formal Models of Cognition

OVERVIEW

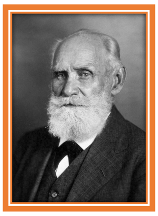
- **Today:**
  1. Introduction error-driven learning
  2. Lab session 1 / assignment

- **Friday:**
  1. Cue competition
  2. Lab session 2 / assignment

HISTORIC CONTEXT

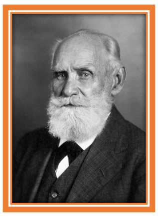
CONDITIONING

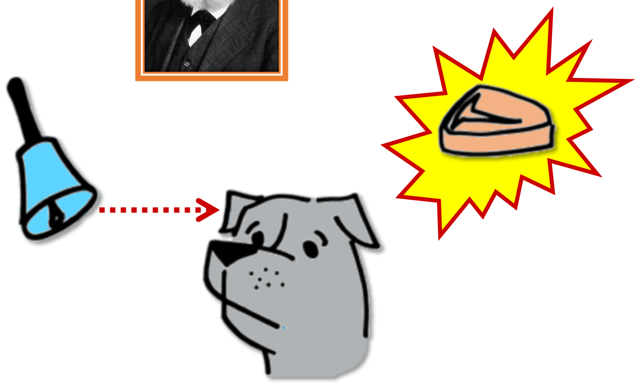
- Pavlov, I.P. (1927): *Conditional Reflexes*



CONDITIONING

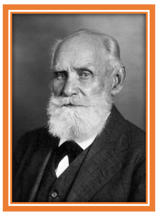
- Pavlov, I.P. (1927): *Conditional Reflexes*





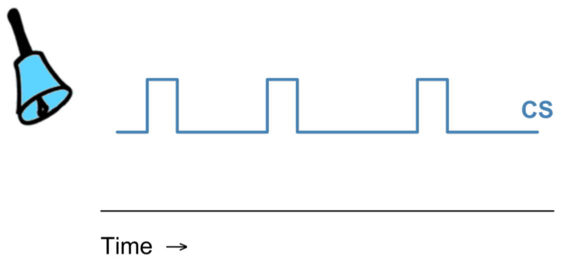
CONDITIONING

- Pavlov, I.P. (1927): *Conditional Reflexes*



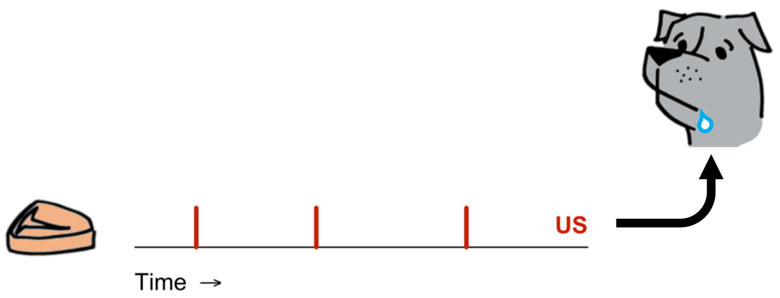
CONDITIONING

- **Conditioned stimulus (CS):** neutral stimulus, initially no response



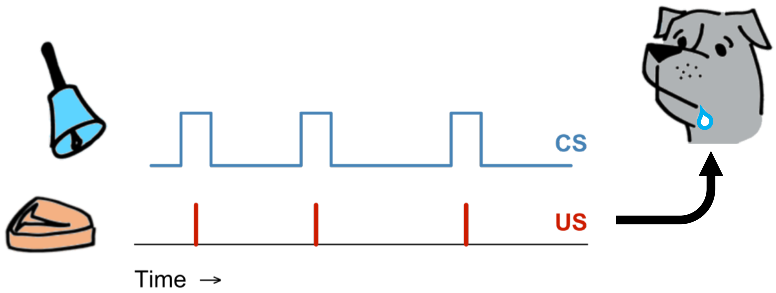
CONDITIONING

- **Conditioned stimulus (CS):** neutral stimulus, initially no response
- **Unconditioned stimulus (US):** stimulus consistently eliciting response – the unconditioned response (UR)



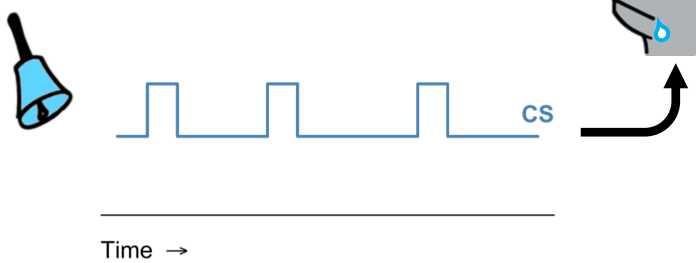
CONDITIONING

- **Conditioned stimulus (CS):** neutral stimulus, initially no response
- **Unconditioned stimulus (US):** stimulus consistently eliciting response – the unconditioned response (UR)
- Conditioned response (CR):

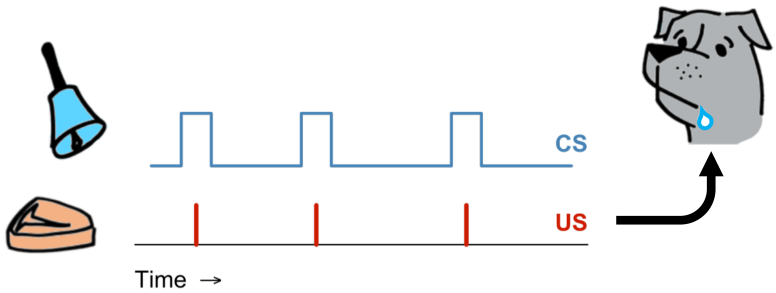


CONDITIONING

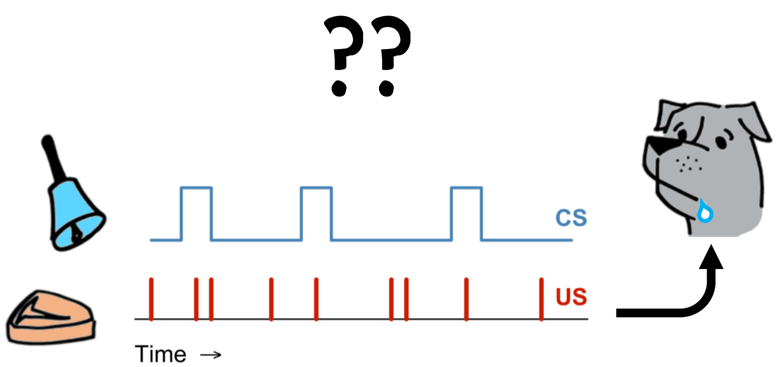
- **Conditioned stimulus (CS):** neutral stimulus, initially no response
- **Unconditioned stimulus (US):** stimulus consistently eliciting response – the unconditioned response (UR)
- Conditioned response (CR):



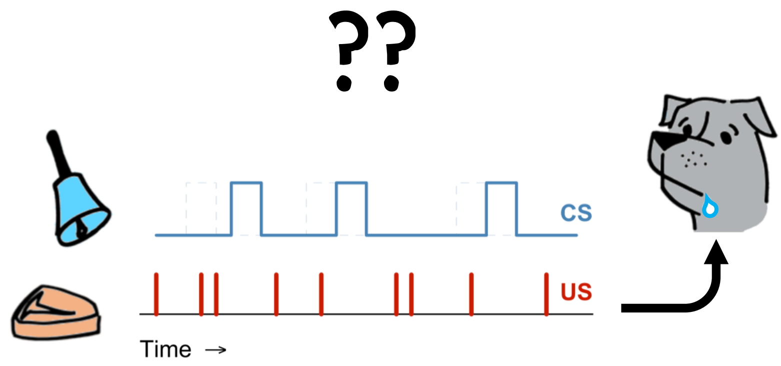
CONDITIONING



CONDITIONING

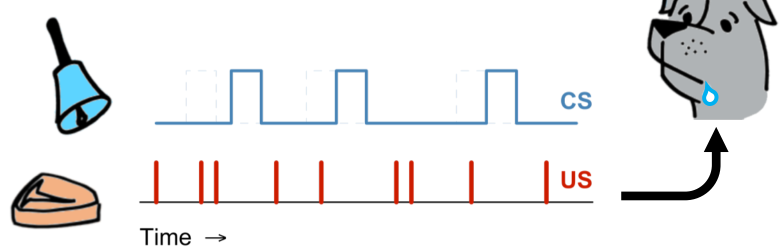


CONDITIONING

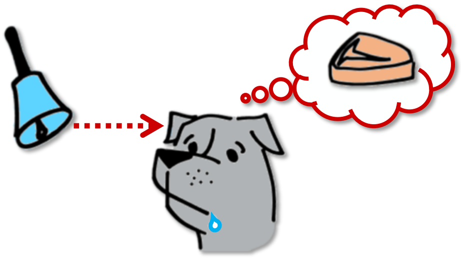


CONDITIONING

- Rescorla (1988):
  - Co-occurrence of CS and US is not sufficient
  - Co-occurrence of CS and US is not necessary



CONDITIONING REVISITED

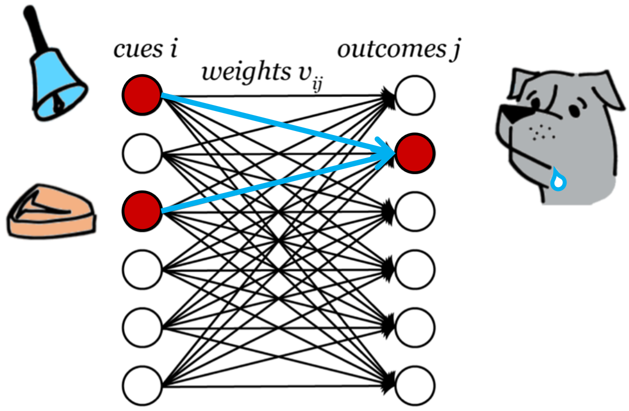


- **Pavlov (1927):** Associations increase in strength when events or objects co-occur
- **Rescorla & Wagner (1972):** Additionally, cues (events, objects that occur) influence associations with events/objects that are ***not*** present

ERROR-DRIVEN LEARNING

ERROR-DRIVEN LEARNING

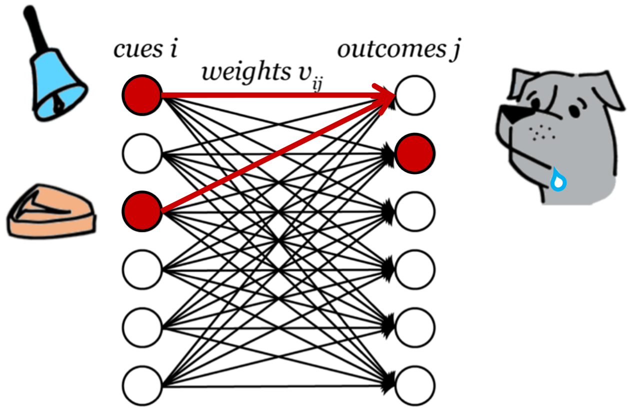
- Memory is represented as feed-forward two-layer network, fully connected



From: Hoppe et al (in preparation)

ERROR-DRIVEN LEARNING

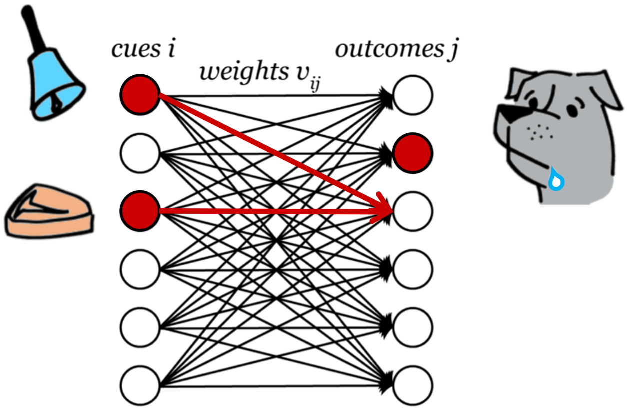
- Memory is represented as feed-forward two-layer network, fully connected



From: Hoppe et al (in preparation)

ERROR-DRIVEN LEARNING

- ❑ Memory is represented as feed-forward two-layer network, fully connected



From: Hoppe et al (in preparation)

ERROR-DRIVEN LEARNING

- ❑ Association strength V between cue i and outcome j is calculated as follows:

  $V_{ij}^{t+1} = V_{ij}^{t} + \Delta V_{ij}^{t}$

  - ΔV_{ij}^{t} depends on one of three learning situations:
    - **¬ C$_i$:** cue i is absent
    - **C$_i$ & O$_j$:** cue i is present and outcome j is present
    - **C$_i$ & ¬ O$_j$:** cue i is present and outcome j is absent

ERROR-DRIVEN LEARNING

- Association strength V between cue i and outcome j is calculated as follows:
  $V_{ij}^{t+1} = V_{ij}^{t} + \Delta V_{ij}^{t}$
  - ΔV_{ij}^{t} depends on one of three learning situations:
    - **¬ C$_i$:** 0
    - **C$_i$ & O$_i$:** $\Delta V_{ij}^{t} = \eta(1 - a_j^t)$
    - **C$_i$ & ¬ O$_i$:** $\Delta V_{ij}^{t} = \eta(0 - a_j^t)$

$\Delta V_{ij}^{t} = \eta(\lambda - a_j^t)$

ERROR-DRIVEN LEARNING

- Association strength V between cue i and outcome j is calculated as follows:
  $V_{ij}^{t+1} = V_{ij}^{t} + \Delta V_{ij}^{t}$
  - ΔV_{ij}^{t} depends on one of three learning situations:
    - **¬ C$_i$:** 0
    - **C$_i$ & O$_i$:** $\Delta V_{ij}^{t} = \eta(1 - a_j^t)$ $\eta = \alpha_i\beta_1 = 0.1\times0.1$
    - **C$_i$ & ¬ O$_i$:** $\Delta V_{ij}^{t} = \eta(0 - a_j^t)$ $\eta = \alpha_i\beta_2 = 0.1\times0.1$

ERROR-DRIVEN LEARNING

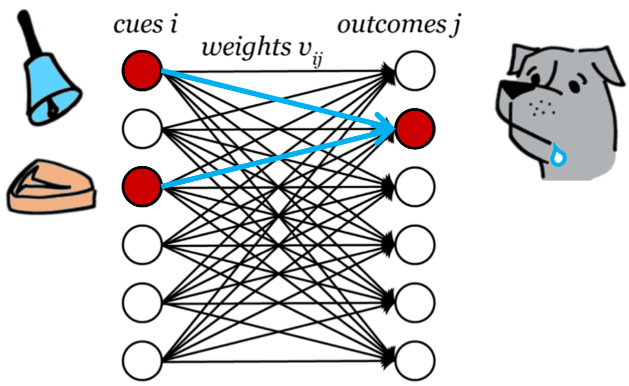
- Association strength V between cue i and outcome j is calculated as follows:

  $V_{ij}^{t+1} = V_{ij}^{t} + \Delta V_{ij}^{t}$

  - ΔV_{ij}^{t} depends on one of three learning situations:
    - **¬ C$_i$:** 0
    - **C$_i$ & O$_j$:** $\Delta V_{ij}^{t} = \eta(1 - a_j^t)$ $\quad a_j^t = \sum_{x \in cues(t)} v_{xj}^t$
    - **C$_i$ & ¬ O$_j$:** $\Delta V_{ij}^{t} = \eta(0 - a_j^t)$

ACTIVATION

- **Activation:** $a_j^t = \sum_{x \in \text{cues}(t)} v_{xj}^t$
  - Sum of weights of all cues present at current time t



From: Hoppe et al (in preparation)

ACTIVATION

- **Activation:** $a_j^t = \sum_{x \in \text{cues}(t)} v_{xj}^t$
  - Sum of weights of all cues present at current time t
  - Determined per outcome

From: Hoppe et al (in preparation)

ERROR-DRIVEN LEARNING

- Association strength V between cue i and outcome j is calculated as follows:

  $V_{ij}^{t+1} = V_{ij}^t + \Delta V_{ij}^t$

  - ΔV_{ij}^t depends on one of three learning situations:
    - **¬ C$_i$:** 0
    - **C$_i$ & O$_j$:** $\Delta V_{ij}^t = \eta(1 - a_j^t)$
    - **C$_i$ & ¬ O$_j$:** $\Delta V_{ij}^t = \eta(0 - a_j^t)$

$a_j^t = \sum_{x \in \text{cues}(t)} v_{xj}^t$

EXAMPLE

PAIRED-ASSOCIATE LEARNING

- Common experimental paradigm:
  - learning word pairings
  - many versions
    - PAL subtest of Wechsler's Memory Scale (des Rosiers & Ivison, 1986): one word acts as cue, the other as outcome
    - measure of people's ability to learn and recall information
      - performance decreases with age: slower and less accurate

Sun et al (2019)

TRAINING

- Different word pairs:

up	down
murder	crime
jury	eagle
baby	cries
…	

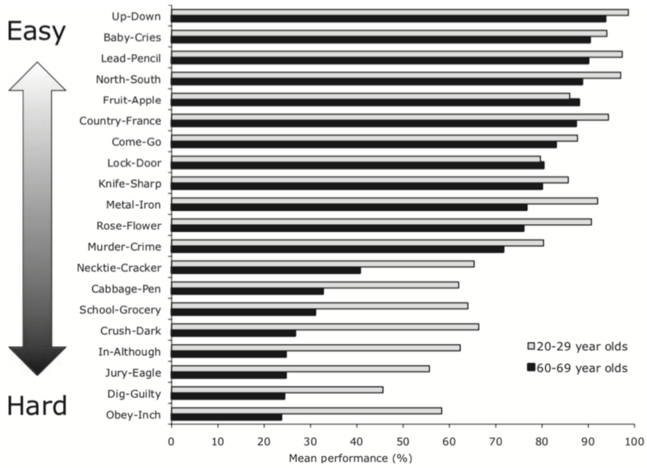
association rates / co occurrences

PAIRED-ASSOCIATE LEARNING

- Performance decreases with age
  - but only on the hard items (Ramscar et al, 2014; Sun et al, 2019)

Sun et al (2019)

PAIRED-ASSOCIATE LEARNING



Ramscar et al (2014)

PAIRED-ASSOCIATE LEARNING

- Error-driven learning model of paired-associate learning (Sun et al, 2019)
- Factors influencing learnability of the word pairs:
  1. association rates
  2. frequency of words (without other words)
  3. experience and blocking

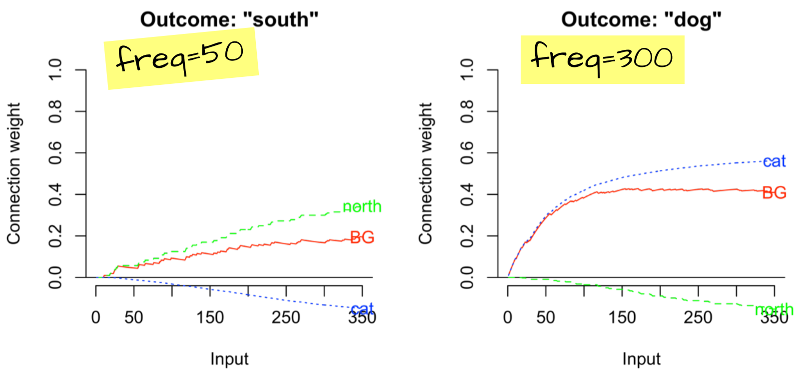
1. ASSOCIATION RATES

- Example data set (Sun et al, 2019):

Cues	Outcomes	Frequency
BG_North	South	50
BG_Cat	Dog	100

1. ASSOCIATION RATES

- Example data set (Sun et al, 2019):



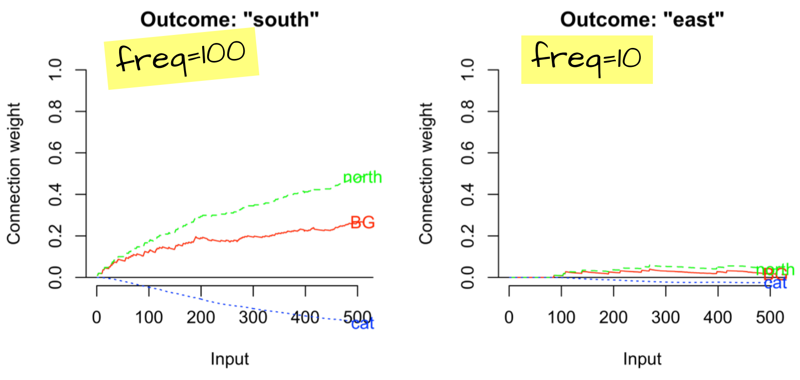
2. FREQUENCY

- Example data set (Sun et al, 2019):

Cues	Outcomes	Frequency
BG_North	South	100
BG_Cat	Dog	100
BG_North	East	10
BG_Cat	Mouse	300

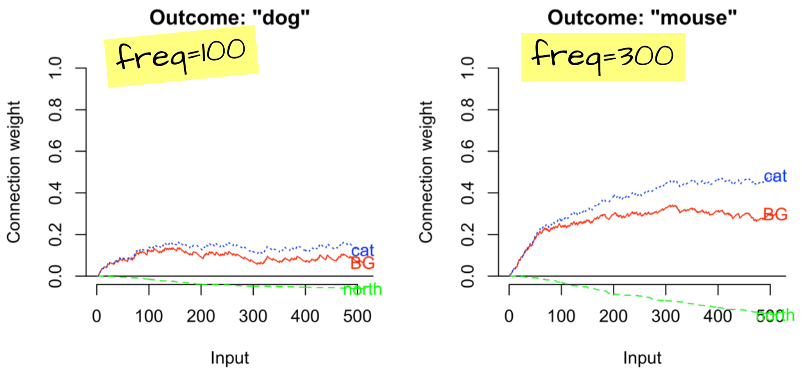
2. FREQUENCY

- Example data set (Sun et al, 2019):



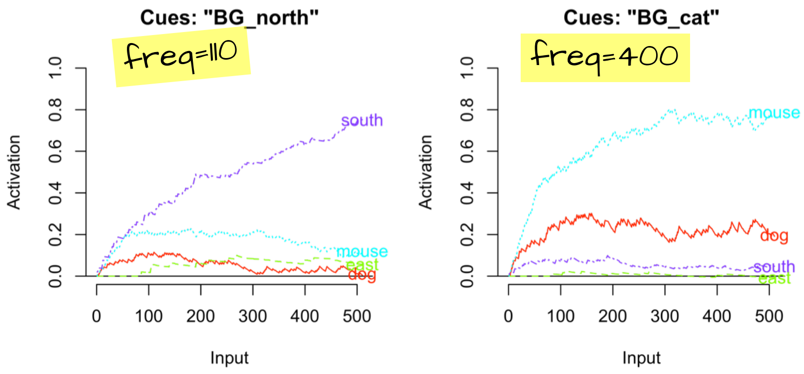
2. FREQUENCY

- Example data set (Sun et al, 2019):



2. FREQUENCY

- Activation per event (set of cues):



3. EXPERIENCE

- Example data set 1 (Sun et al, 2019):

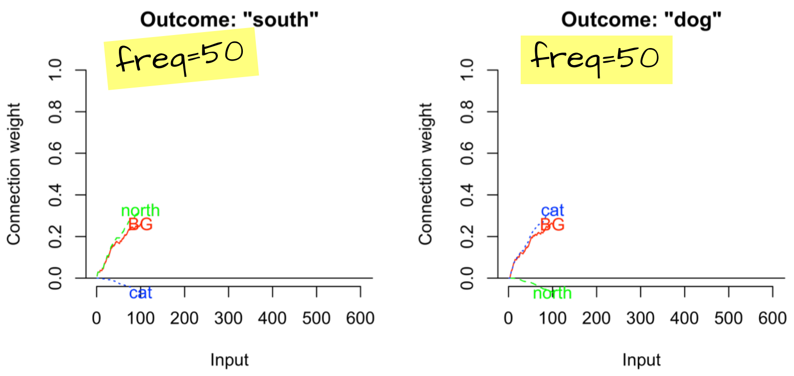
Cues	Outcomes	Frequency
BG_North	South	100 I 300
BG_Cat	Dog	100 I 300

- Example data set 2 (Sun et al, 2019):

Cues	Outcomes	Frequency
BG_North	South	1
BG_Cat	Dog	1
BG_Banana	Dog	1
BG_North	Dog	1

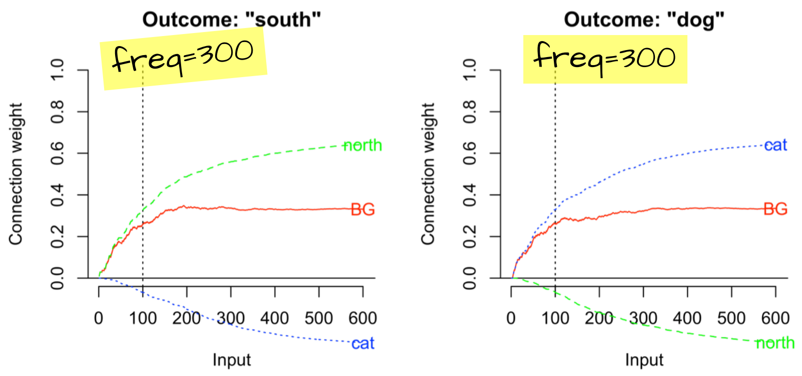
3. EXPERIENCE

- Example data set (Sun et al, 2019):



3. EXPERIENCE

- Example data set (Sun et al, 2019):



3. EXPERIENCE

- Example data set 1 (Sun et al, 2019):

Cues	Outcomes	Frequency
BG_North	South	100 | 300
BG_Cat	Dog	100 | 300

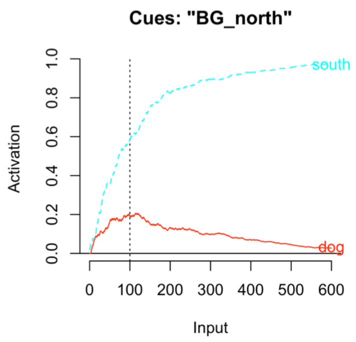
- Example data set 2 (Sun et al, 2019):

Cues	Outcomes	Frequency
BG_North	South	1
BG_Cat	Dog	1
BG_Banana	Dog	1
BG_North	Dog	1

3. EXPERIENCE

```
# Freq = 50
        south    dog
BG      0.261  0.259
north   0.328 -0.069
cat    -0.067  0.328

# Freq = 300
        south    dog
BG      0.332  0.335
north   0.642 -0.308
cat    -0.310  0.642
```



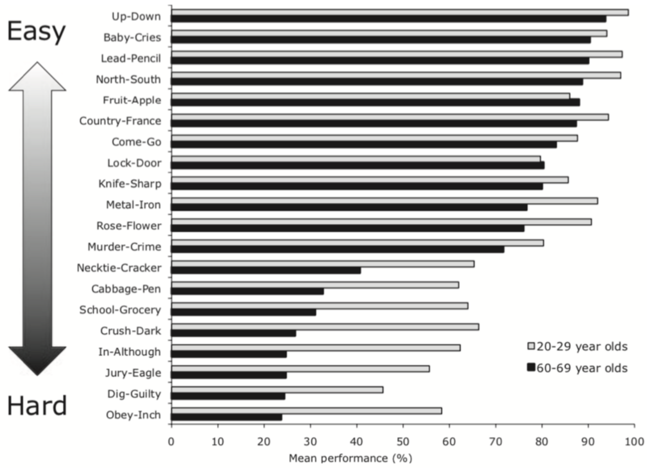
$$\Delta V_{ij}^{t+1} = \eta(\lambda - \alpha_j^t)$$

PAIRED-ASSOCIATE LEARNING

- Performance decreases with age
  - but only on the hard items (Ramscar et al, 2014; Sun et al, 2019)
    - error-driven learning simulations: more experience with language make it more difficult to learn uncommon word pairs
    - Support from PAL test in L2 (second language)

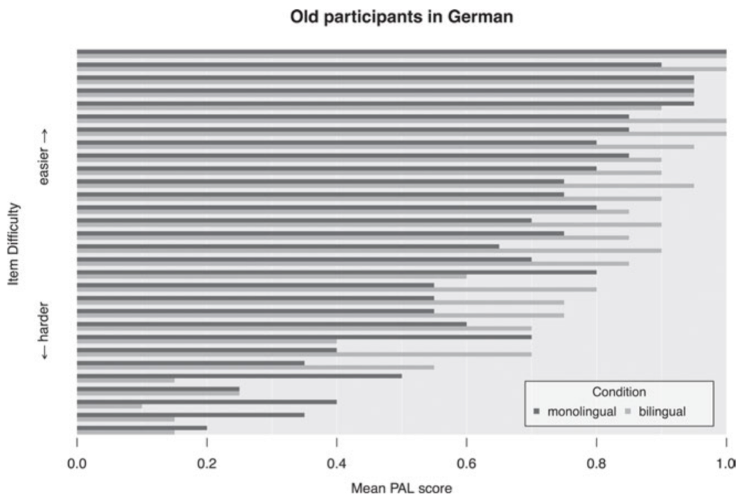
Sun et al (2019)

PAIRED-ASSOCIATE LEARNING



Ramscar et al (2014)

PAIRED-ASSOCIATE LEARNING

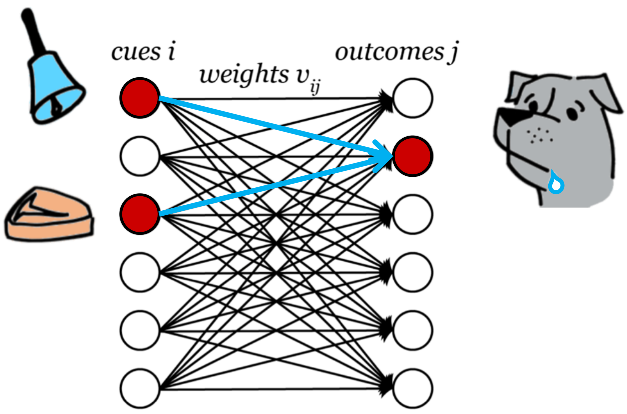


Sun et al (2019)

DISCUSSION

SUPERVISED LEARNING?

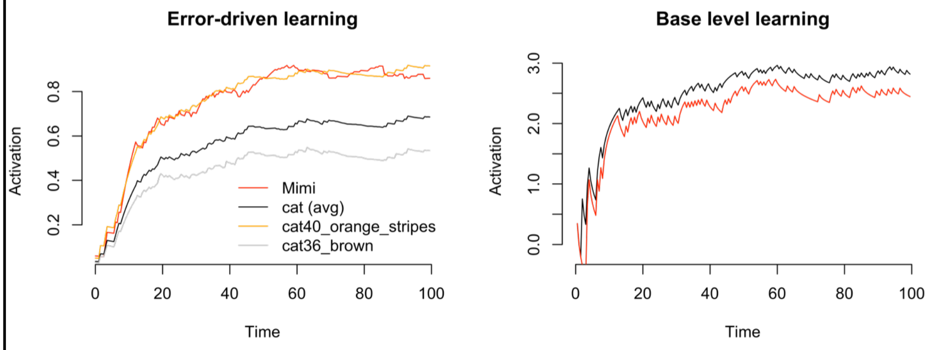
- ❑ No, adjusting predictions based on experience



From: Hoppe et al (in preparation)

ACTIVATION & TIME DECAY?

- No, activation changes irrespective of time



DISCUSSION

- Advantages:
  - Insight in how structure of input affects learning
  - No hidden layers, so learning dynamics can be studied
- Limitations:
  - Not a process model
  - Representations defined by the user. Too high/low level of abstraction may miss learning effects