# Learning dynamics in error-driven learning

Dorothée Hoppe & Jacolien van Rij

October 2019, Formal Models of Cognition

# cue vs. outcome competition

# recap formula: cue competition

**weight update**

$$\Delta V_{ij}^{t} = \begin{cases} 0, & \text{cue i absent} \\ \eta(1 - a_j^t), & \text{cue i \& outcome j present} \\ \eta(0 - a_j^t), & \text{cue i present \& outcome j absent} \end{cases}$$

**activation**

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

all currently present cues

**aim:**
**every cue set fully predicts exactly one outcome**

# recap formula: cue competition

**weight update**

$$\Delta V_{ij}^{t} = \begin{cases} 0, & \text{cue i absent} \\ \eta(1 - a_j^t), & \text{cue i \& outcome j present} \\ \eta(0 - a_j^t), & \text{cue i present \& outcome j absent} \end{cases}$$

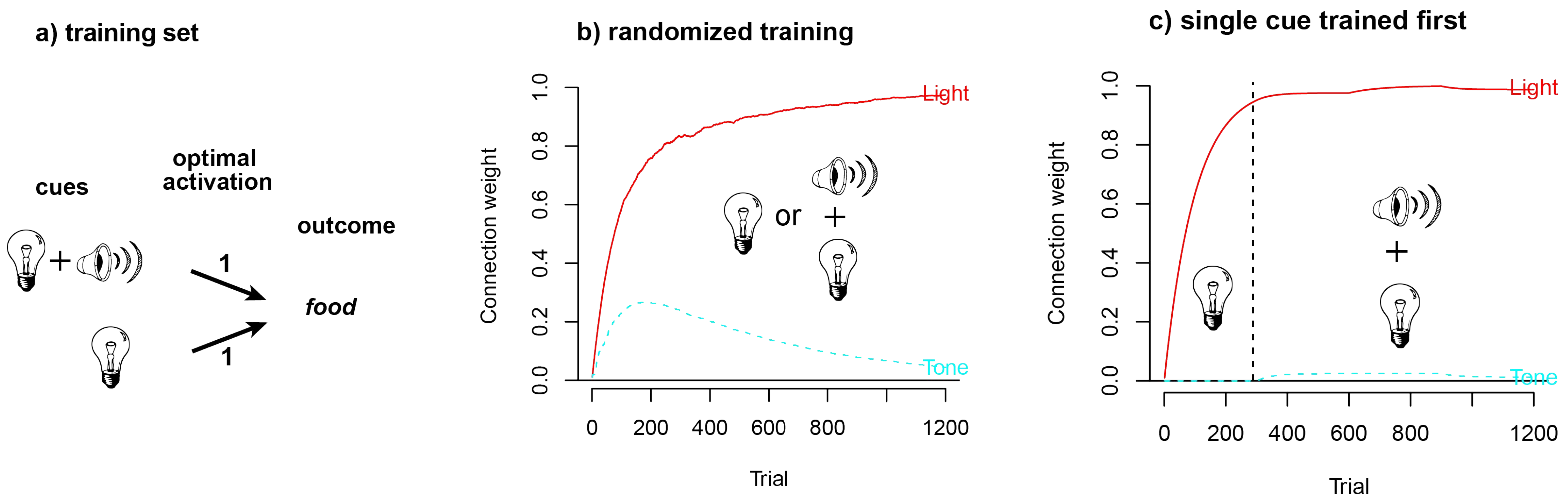
**activation**

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

all currently present cues

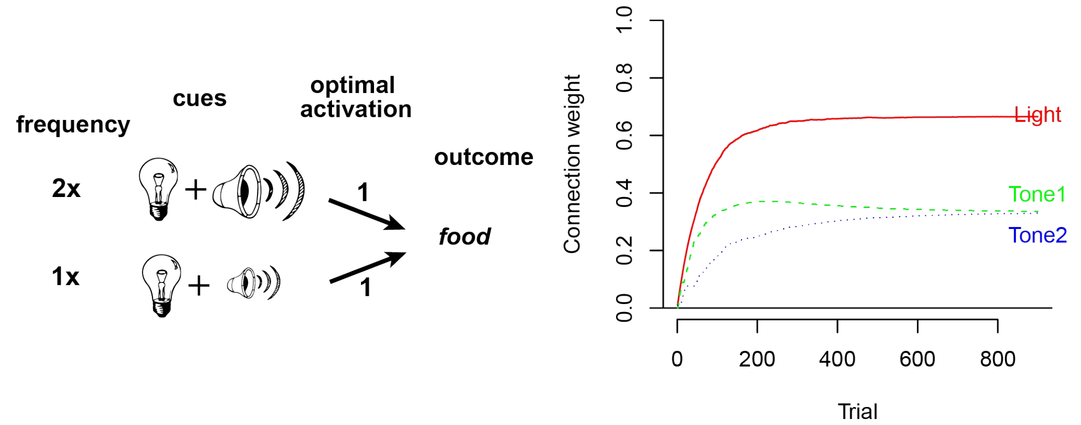
**aim:**
**maximize activation of the current outcome!**

# **cue competition: blocking**



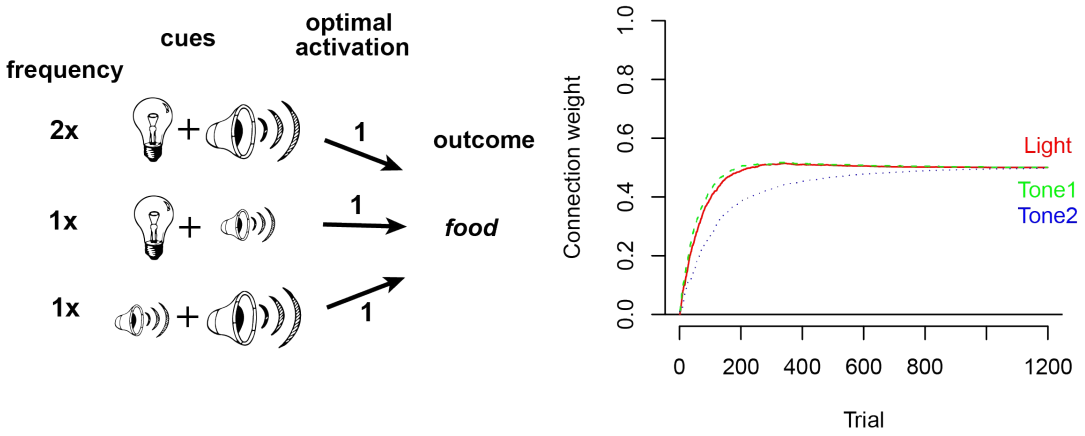

- cues that appear more frequently with an outcome are learned to be more informative (?)
- outcome activation is limited
  ⇒ **temporal order** affects cue competition

# cue competition




only cues that appear together can compete

⇒ **interaction more important than frequency**




interactions can soon get quite complex!

# recap formula: outcome competition

## weight update

$$\Delta V_{ij}^{t} = \begin{cases} 0, & \text{cue i absent} \\ \eta(1 - a_j^t), & \text{cue i \& outcome j present} \\ \eta(0 - a_j^t), & \text{cue i present \& outcome j absent} \end{cases}$$

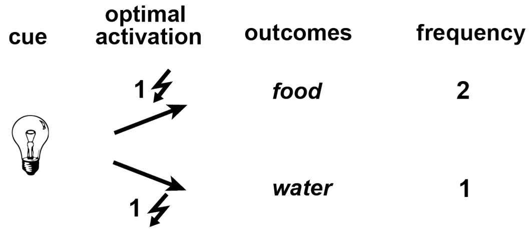
## activation

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

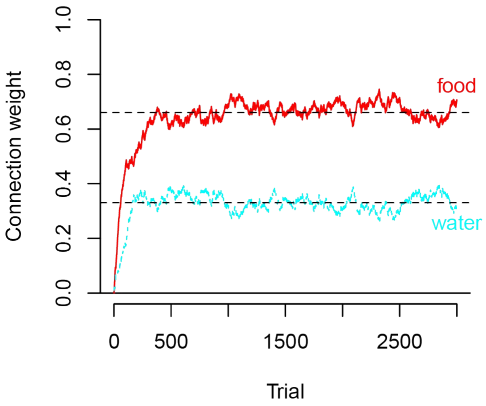
**aim:**
**every cue set fully predicts exactly one outcome**

# outcome competition

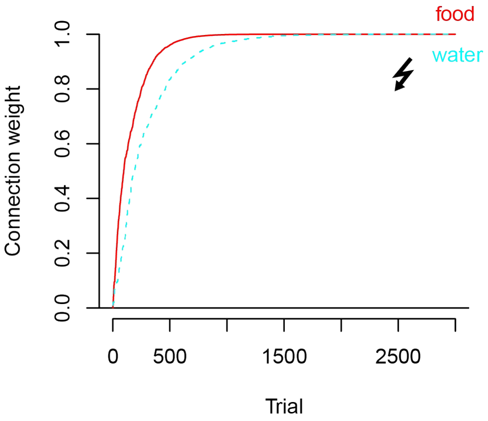
a) training set




b) with outcome competition




d) without outcome competition

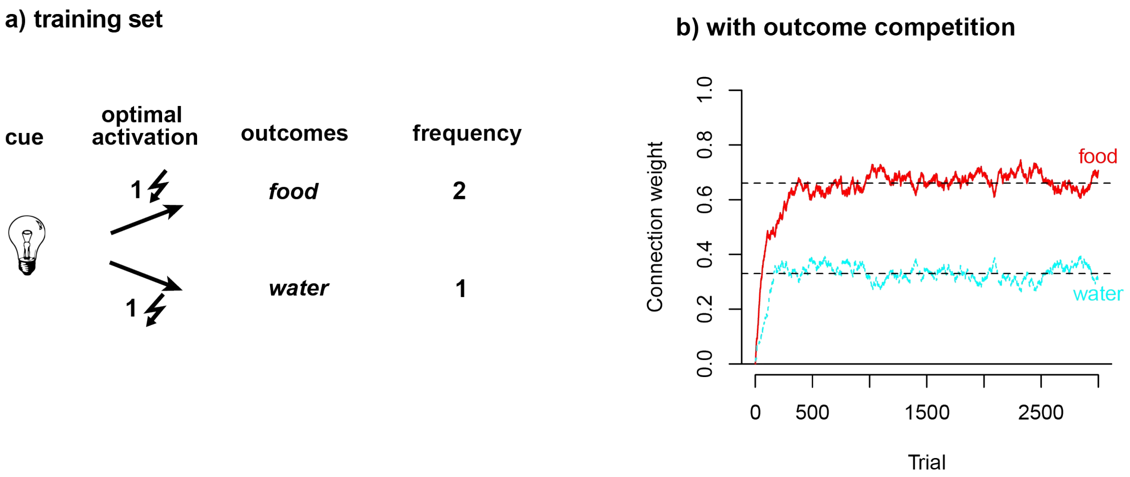



here, activations can never reach the maximum of 1!

$$\Delta V_{ij}^{t} = \begin{cases} 0, & \text{cue i absent} \\ \eta(1 - a_j^t), & \text{cue i \& outcome j present} \\ \eta(0 - a_j^t), & \text{cue i present \& outcome j absent} \end{cases}$$

$$\Delta V_{ij}^{t} = \begin{cases} 0, & \text{cue i absent} \\ \eta(1 - a_j^t), & \text{cue i \& outcome j present} \\ \cancel{\eta(0 - a_j^t),} & \cancel{\text{cue i present \& outcome j absent}} \end{cases}$$

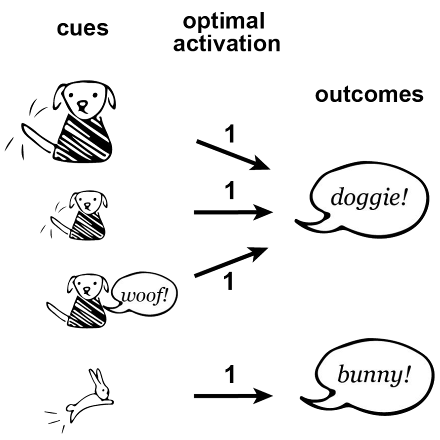
# **outcome competition**



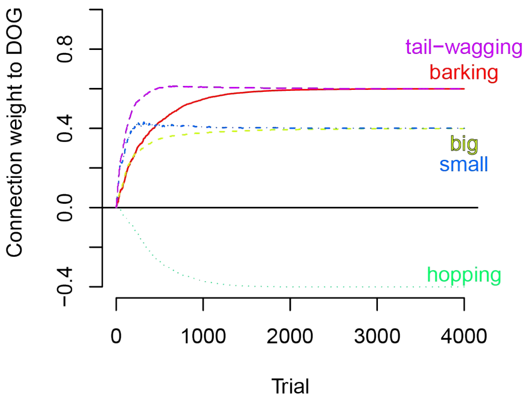

⇒ estimation of **conditional probabilities!**

# whole system

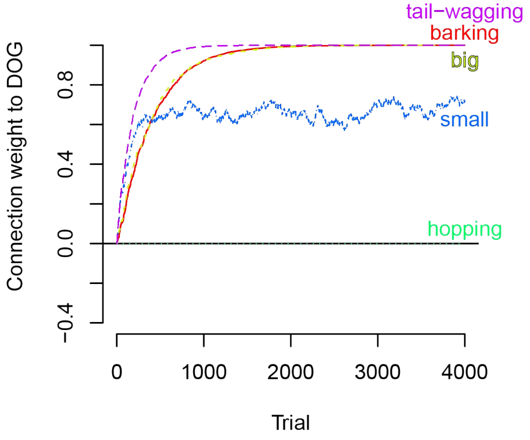
a) training set



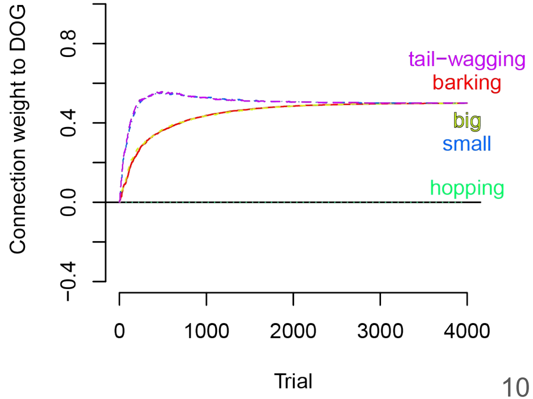

b) error-driven learning




c) without cue competition




d) without outcome competition

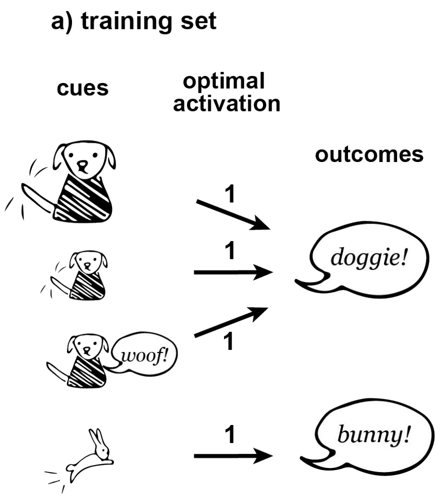


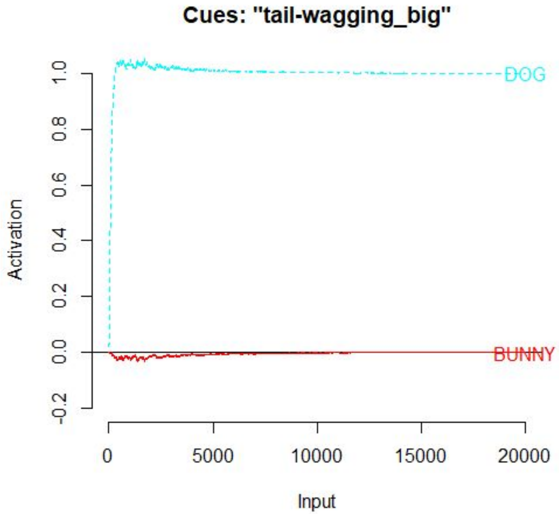

⇒ **cue** competition cancels the frequency effect (over time)

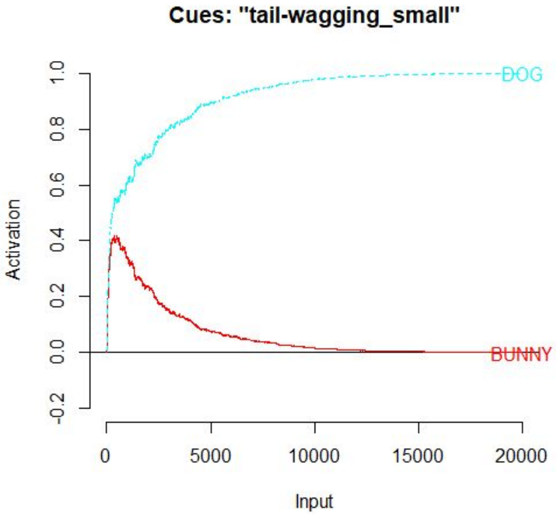
⇒ **outcome** competition finds irrelevant dimension (size)

*Beware:* c) and d) are produced with **modified** algorithms!

# whole system



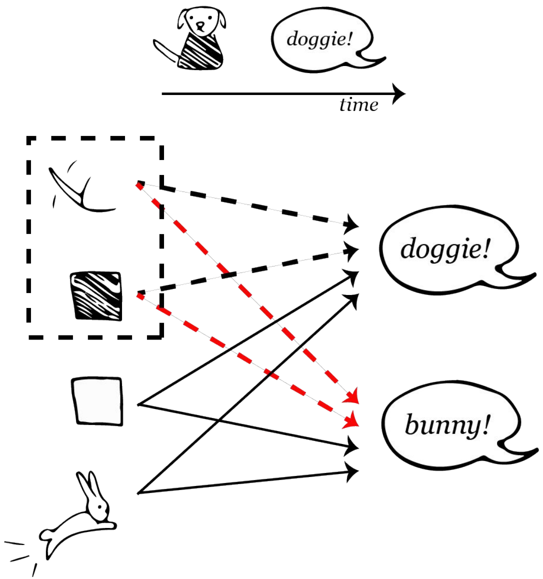




⇒ optimized outcome discrimination

# is learning asymmetric? ⇒ assignment!

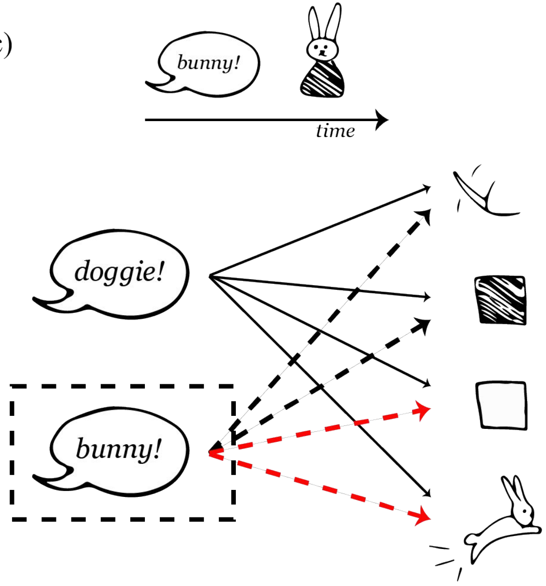
## convergent network

b)




## divergent network

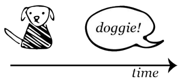
c)



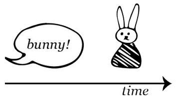
# temporal dynamics

- cue - outcome order:

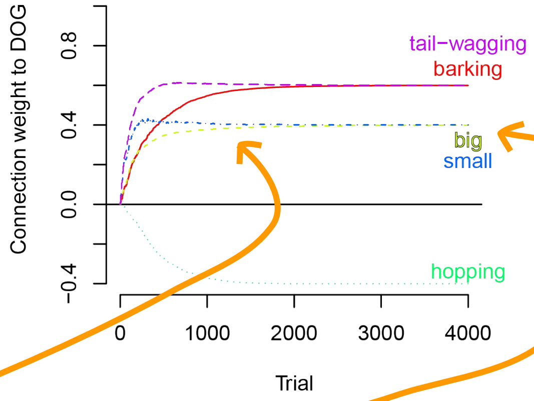
b)




c)



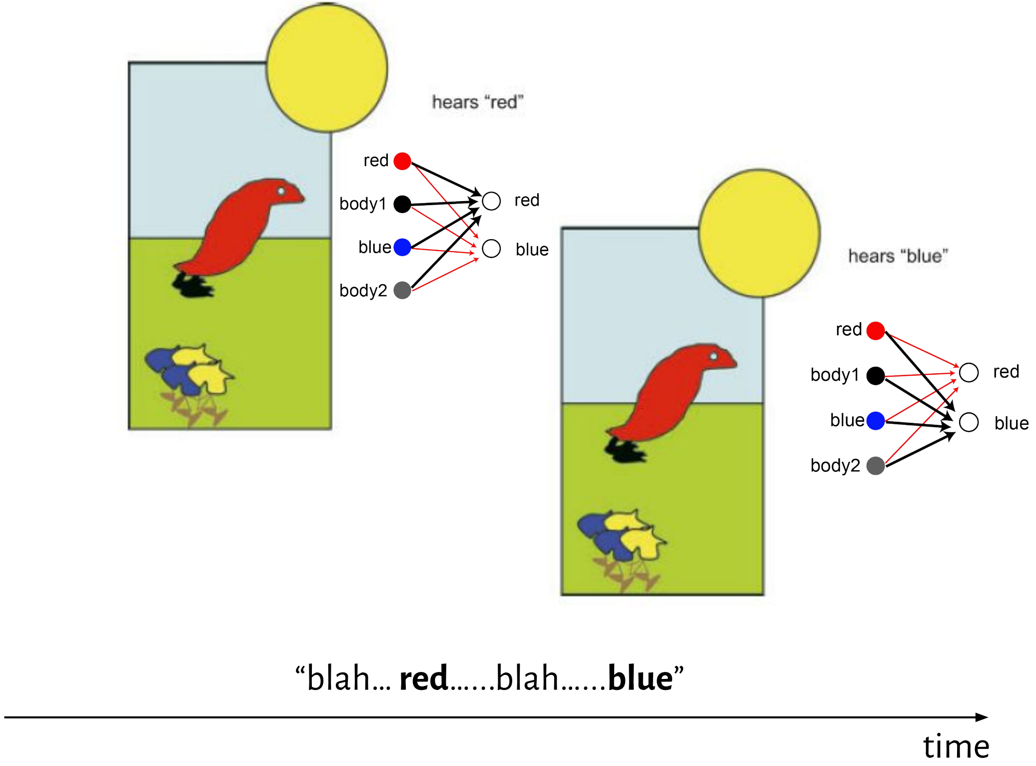

- trial order (e.g., random vs. blocked)
- training time:




- iterative vs. batch learning (Danks (2003) equations, also in package *ndl*)

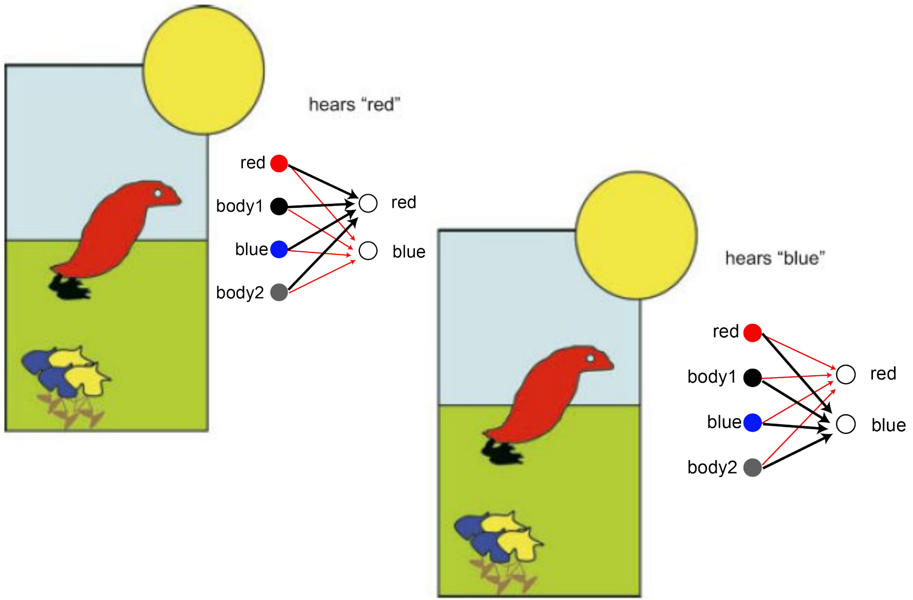
# Example: color-word learning

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010)



# **Example:** color-word learning

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010)

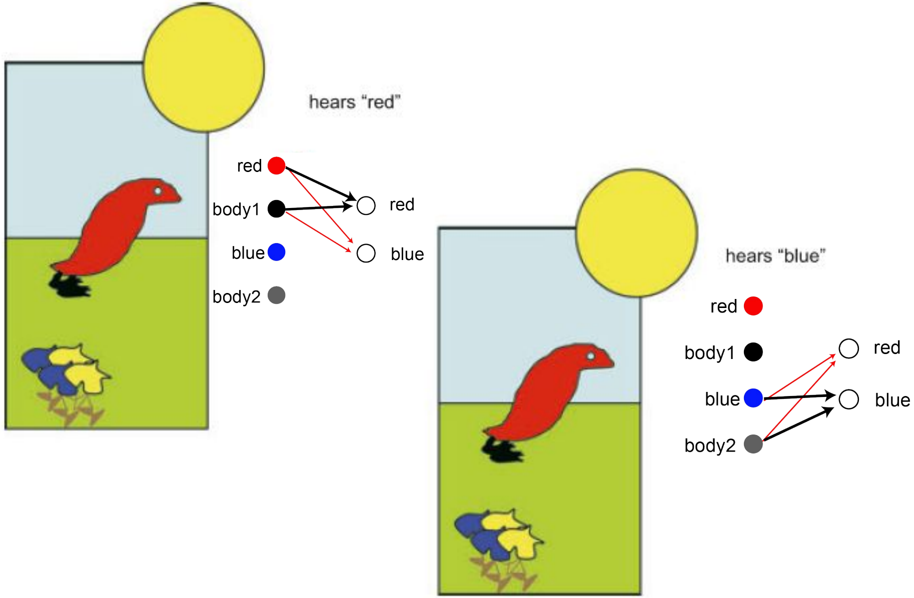



“Look, there’s a **red niz** and some **blue wugs!**”

time

# Example: color-word learning

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010)



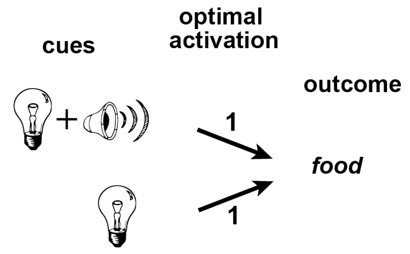

"Look, the **niz** is **red** and the **wugs** are **blue!**"

time

*Assumption*: learner already has an idea what wugs and nizzes are

# Data structure in error-driven learning models

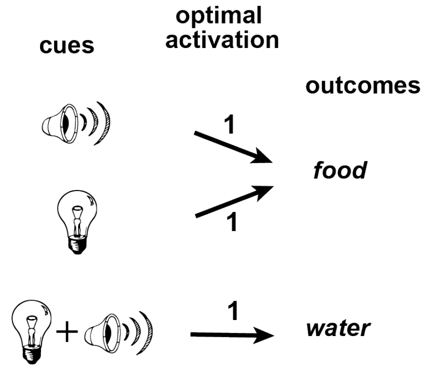
# elemental vs. configural representations




| Cues | Outcomes |
|---|---|
| light | food |
| light_tone | food |

probably too good?

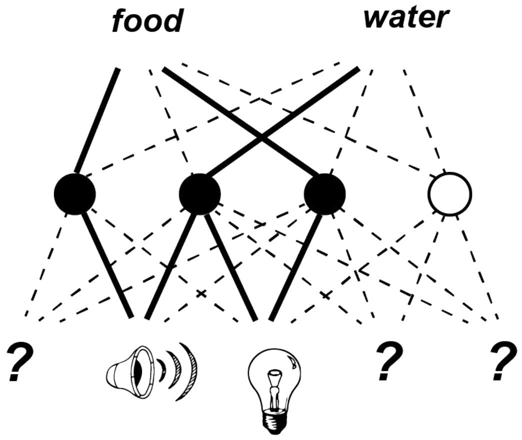
| Cues | Outcomes |
|---|---|
| light | food |
| sound | food |
| **LightTone** | water |

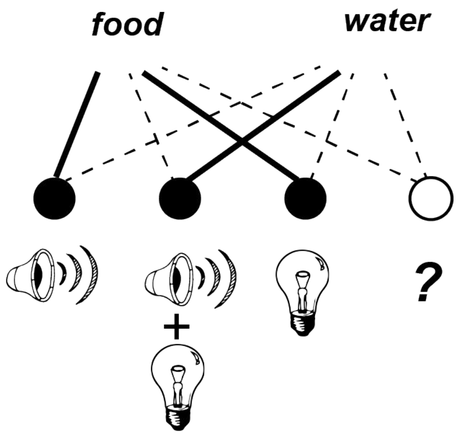



| Cues | Outcomes |
|---|---|
| light | food |
| sound | food |
| light_tone | water |

| Cues | Outcomes |
|---|---|
| light | food |
| sound | food |
| light_tone_**LightTone** | water |

# hidden layers vs. configural representations



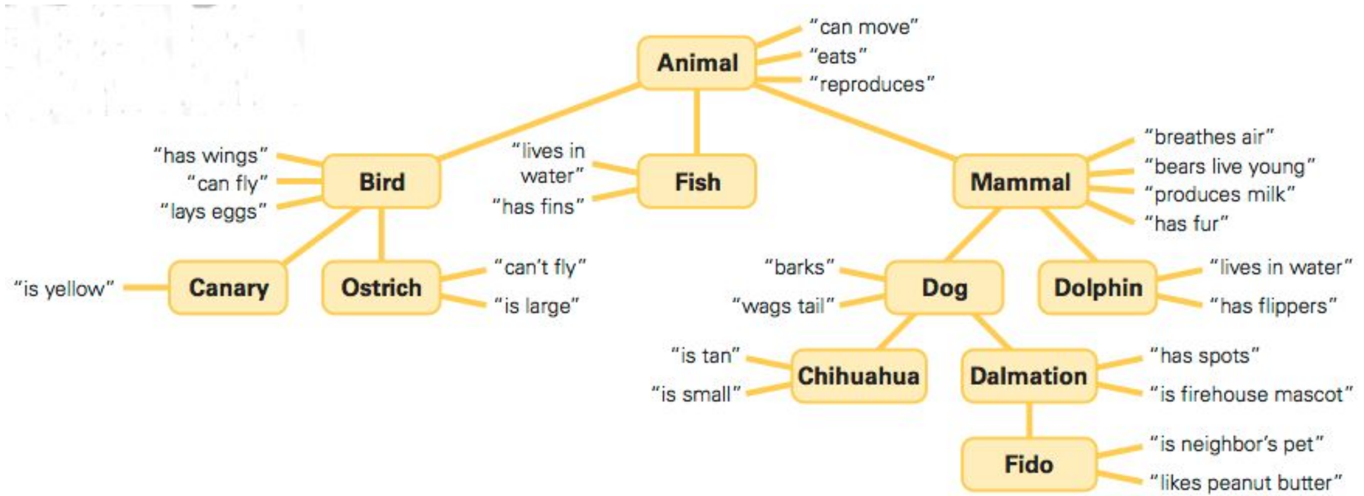




- hidden layers compute relevant configural representations!
- but hidden layers make it hard to trace what happens in the model…

# What kind of **representations**?

"All **representations** [original: *models*] are wrong but some are useful"

- adapted from George Box

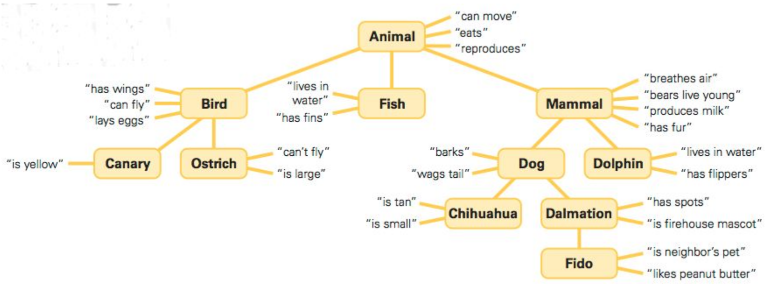



Gluck, M. A., Mercado, E., & Myers, C. E. (2009).
*Learning and memory: from brain to behavior.*

# What kind of **representations**?

"All **representations** [original: *models*] are wrong but some are useful"

- adapted from George Box




- Best case: allow the model to discover the relevant levels of abstraction for a given task!
- We can never have too many representations
- But, we cannot include all possible features because of lack of knowledge/time/computational power/etc., thus, we still have to make **informed choices**