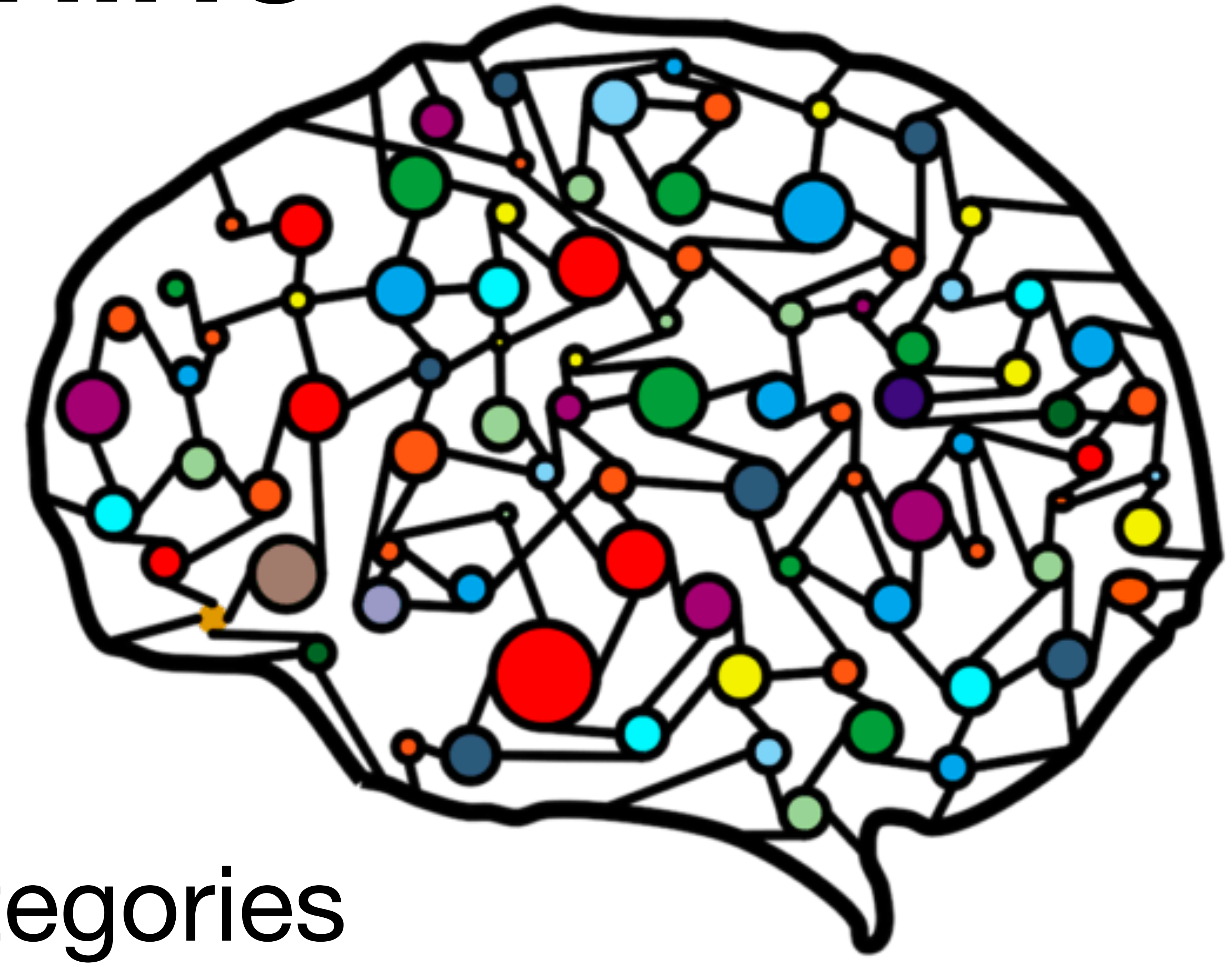
General Principles of Human and Machine Learning



Lecture 8: Concepts and Categories

Dr. Charley Wu

https://hmc-lab.com/GPHML.html

Quiz clarification

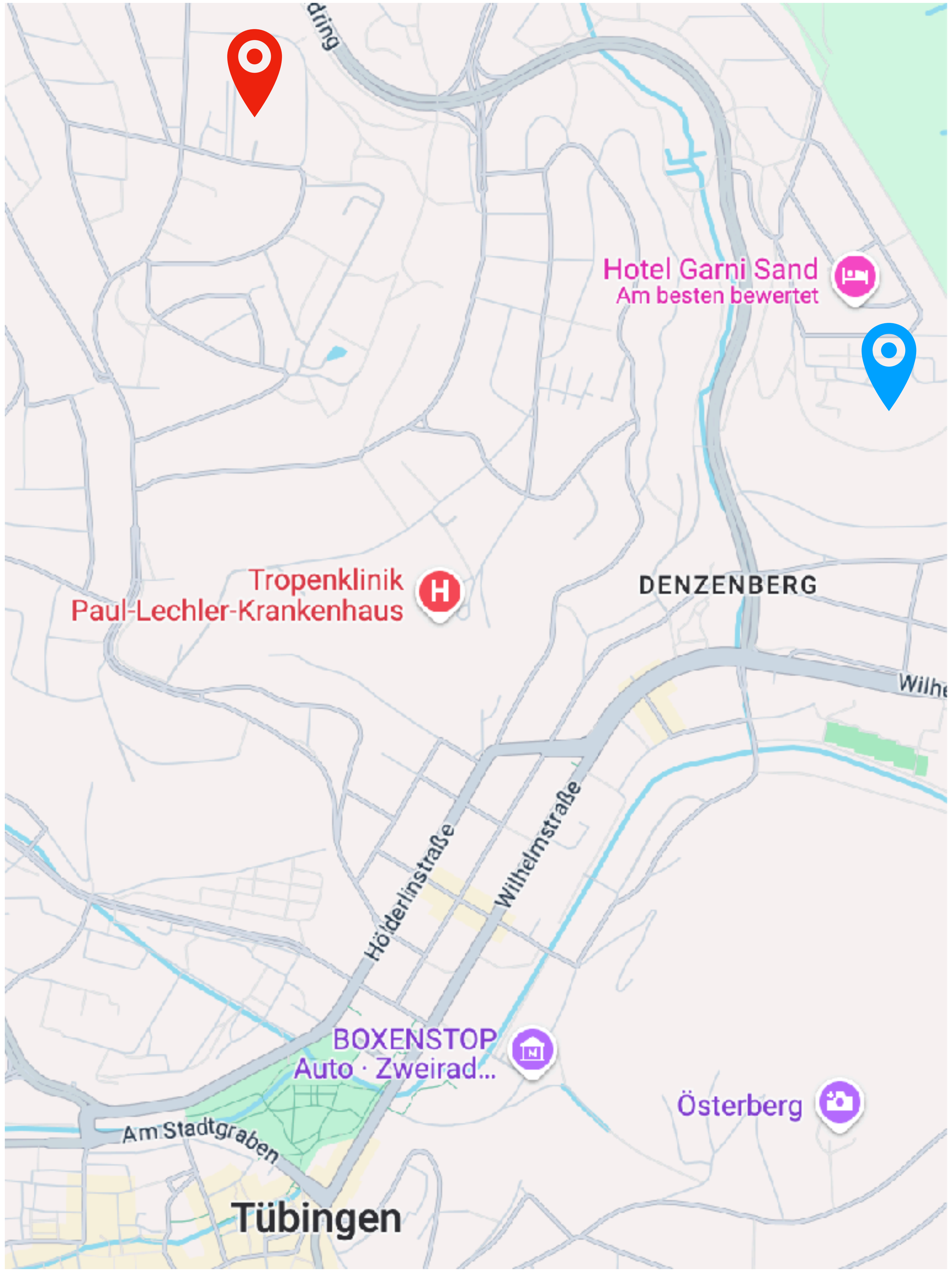
- Quiz #2 further reduced to be out of 16
  - Avg score = 83%
- Quizzes
  - try not to have overlapping content
  - minimal content from the same week's lecture

Date	Remarks	Lecture	Tutorial	TA	Readings
Week 1:		Oct 15: Introduction (slides)	Oct 16 (slides)	Alex	Spicer & Sanborn (2019). What does the mind learn?
Week 2:		Oct 22: Origins of biological and artificial learning (slides)	Oct 23 (slides)	Turan	[1] Behaviorism [2] What is a perceptron? (Blog post)
Week 3:		Oct 29: Symbolic AI and Cognitive maps (slides)	Oct 30 (Quiz #1)	Alex	[1] Garnelo & Shanahan (2019) [2] Boorman et al., 2021
Week 4:		Nov 5: Introduction to RL (slides)	Nov 6 (slides)	Turan	Sutton & Barton (Ch. 1 & 2)
Week 5:		Nov 12: Advances in RL (slides)	Nov 13 (Quiz #2)	Turan	Neftci & Averbeck (2019)
Week 6:	Guest lecturer: Alexandra Witt	Nov 19: Social learning (slides)	Nov 20 (slides)	Alex	Witt et al., (2024)
Week 7:	Guest lecturer: Dr. David Nagy	Nov 26: Compression and resource constraints (slides)	Nov 27	David	Nagy et al., (2020)
Week 8:		Dec 3: Concepts and Categories	Dec 4	Hanqi	Murphy (2023)
Week 9:		Dec 10: Supervised and Unsupervised learning	Dec 11	Hanqi	Bishop (Ch. 4)
	Holiday break				
Week 10:		Jan 14: Function learning	Jan 15	Alex	Wu, Meder, & Schulz (2024)
Week 11:		Jan 21: No Lecture	Jan 22: No Tutorial		
Week 12:		Jan 28: Language and semantics	Jan 29	TBD	Kamath et al., (2024)
Week 13:		Feb 4: General Principles	Feb 5	Charley	Gershman (2023)

Quiz #3?

Exam times now confirmed

Exam 1	13:00-15:00 21.02.2025 Hörsaal 1 F119 (SAND)
Exam 2	12:00-14:00 11.04.2025 Ground floor lecture room, AI building, Maria-von-Linden-Str. 6, D-72076 Tübingen



Course feedback

- Randomness of when quizzes occur
  - My logic for the current format is that it should be less stressful than mid-term tests with fixed dates
  - The content is designed to be much easier since it is random, whereas fixed date tests would have to be much harder
  - Also means I don't need to take attendance and can get a regular pulse about what people are struggling with and what is relatively easy
- Take-home tests are not ideal in a post-GPT world and would not resemble exam
- Ultimately, the pop quizzes are a small part of your grade (20% with a best 3 out of 4 evaluation scheme) and designed to help you get the best possible grade on the final exam

Course feedback

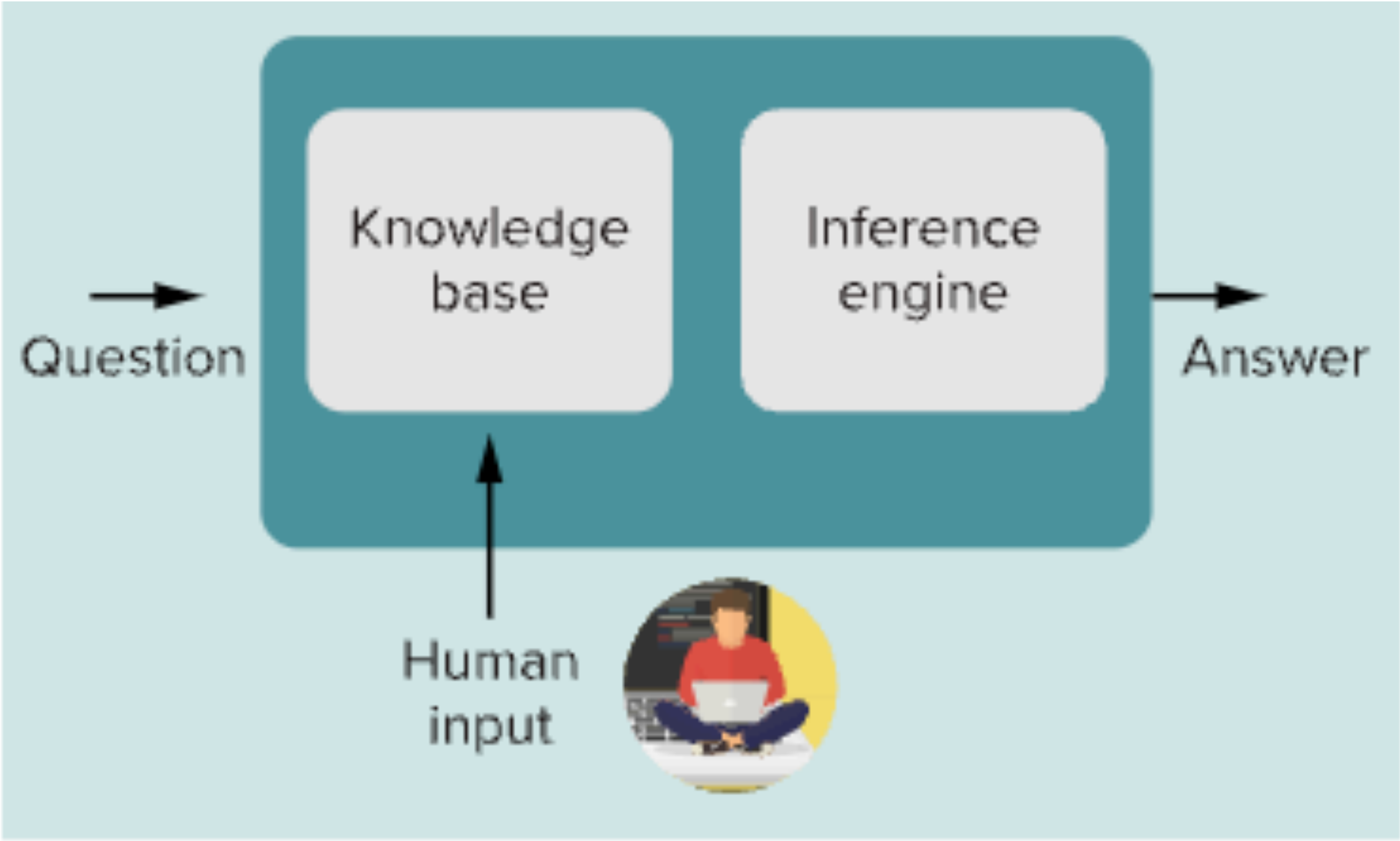
- Slides with or without animations?
  - Current I export with animations since I sometimes rely on animations to tell some part of the story
  - But I guess the pdfs are bigger
- Inconsistency of grading between quiz #1 and #2?
  - This was mentioned in the feedback
  - Let’s chat if you had this issue

Recap of the story so far…

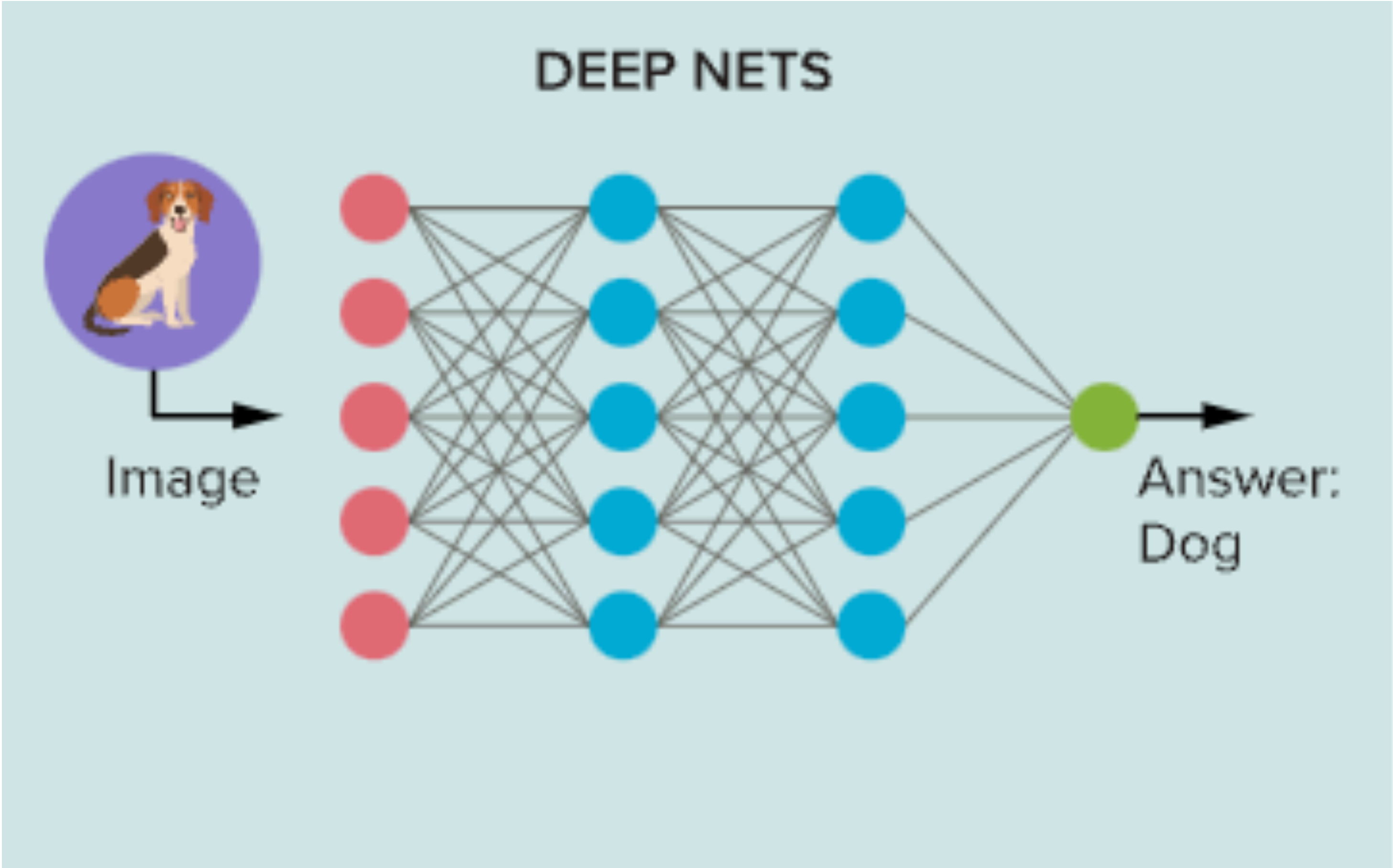
Symbolic vs. Subsymbolic AI

Symbolic vs. Subsymbolic AI

Symbolic AI



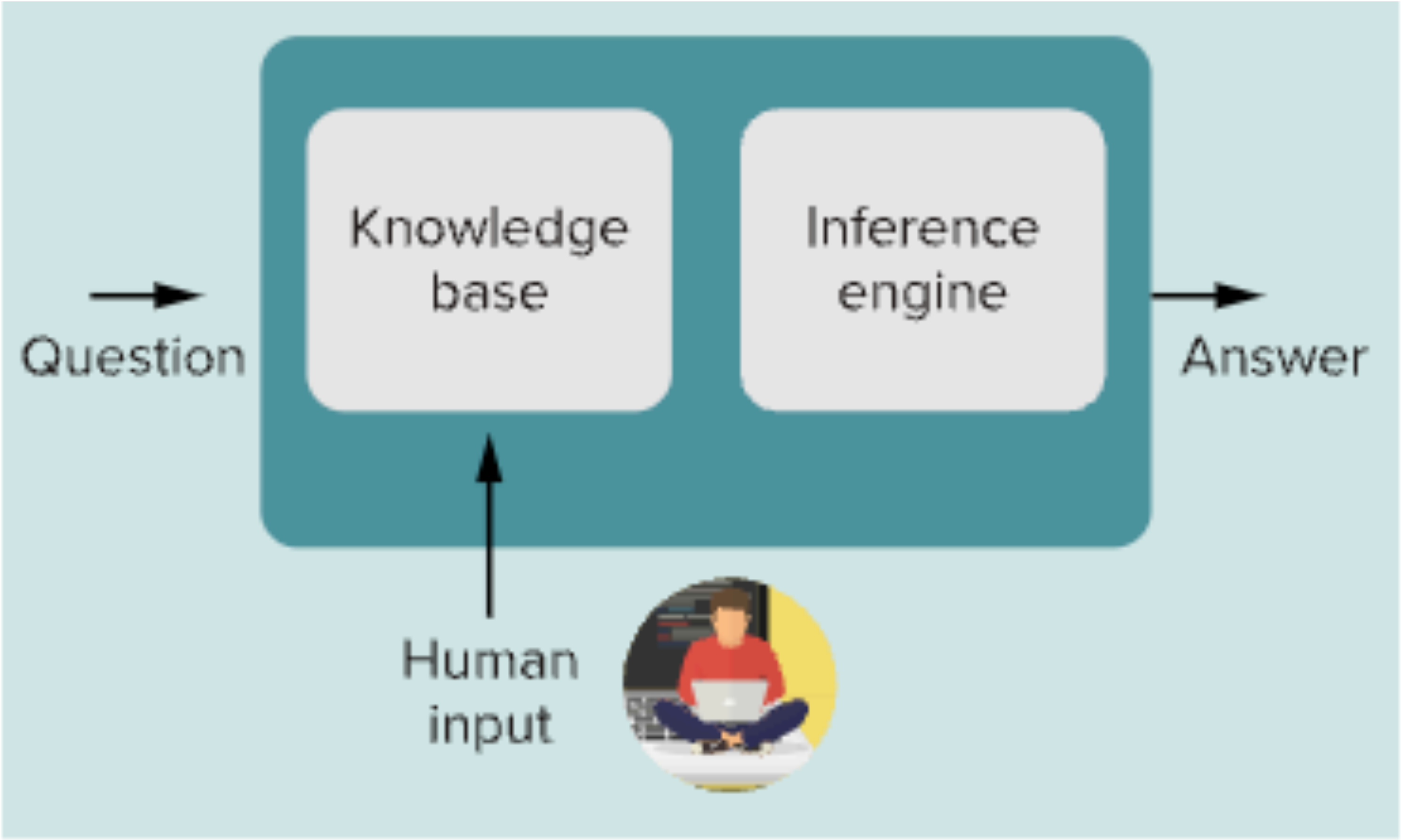
Subsymbolic AI



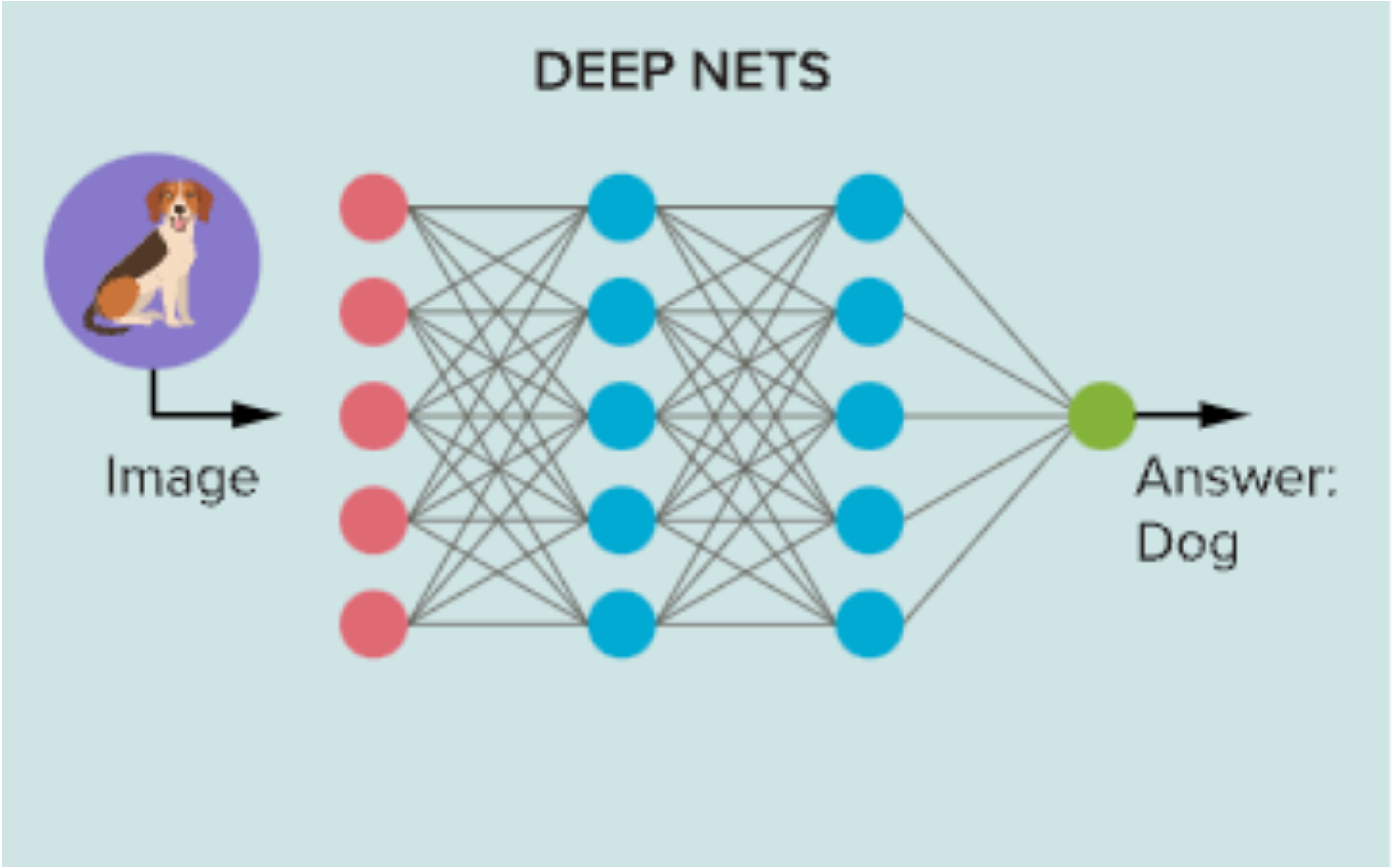
Symbolic vs. Subsymbolic AI

Physical symbol system hypothesis: manipulating **symbols** and **relations**

Symbolic AI



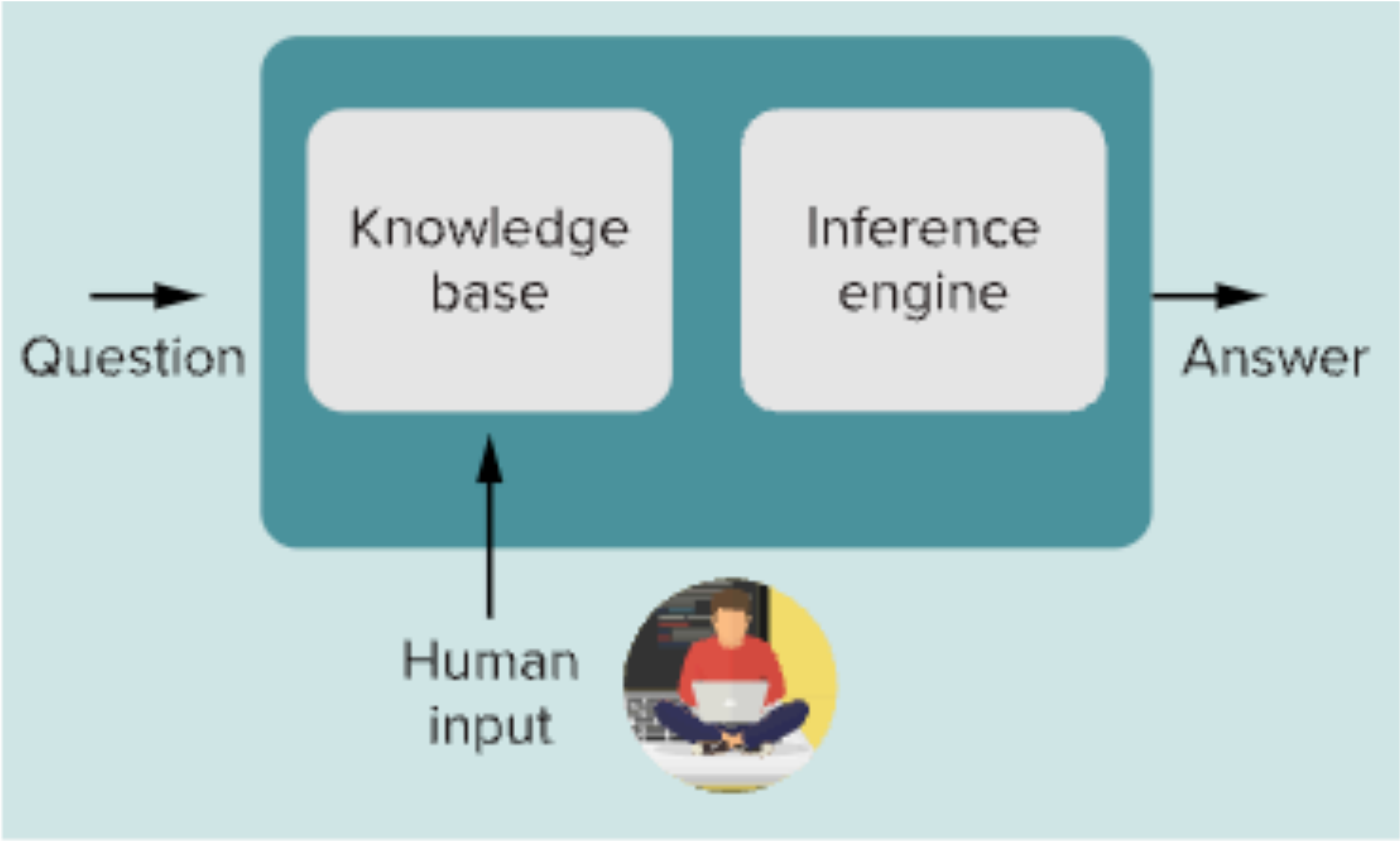
Subsymbolic AI



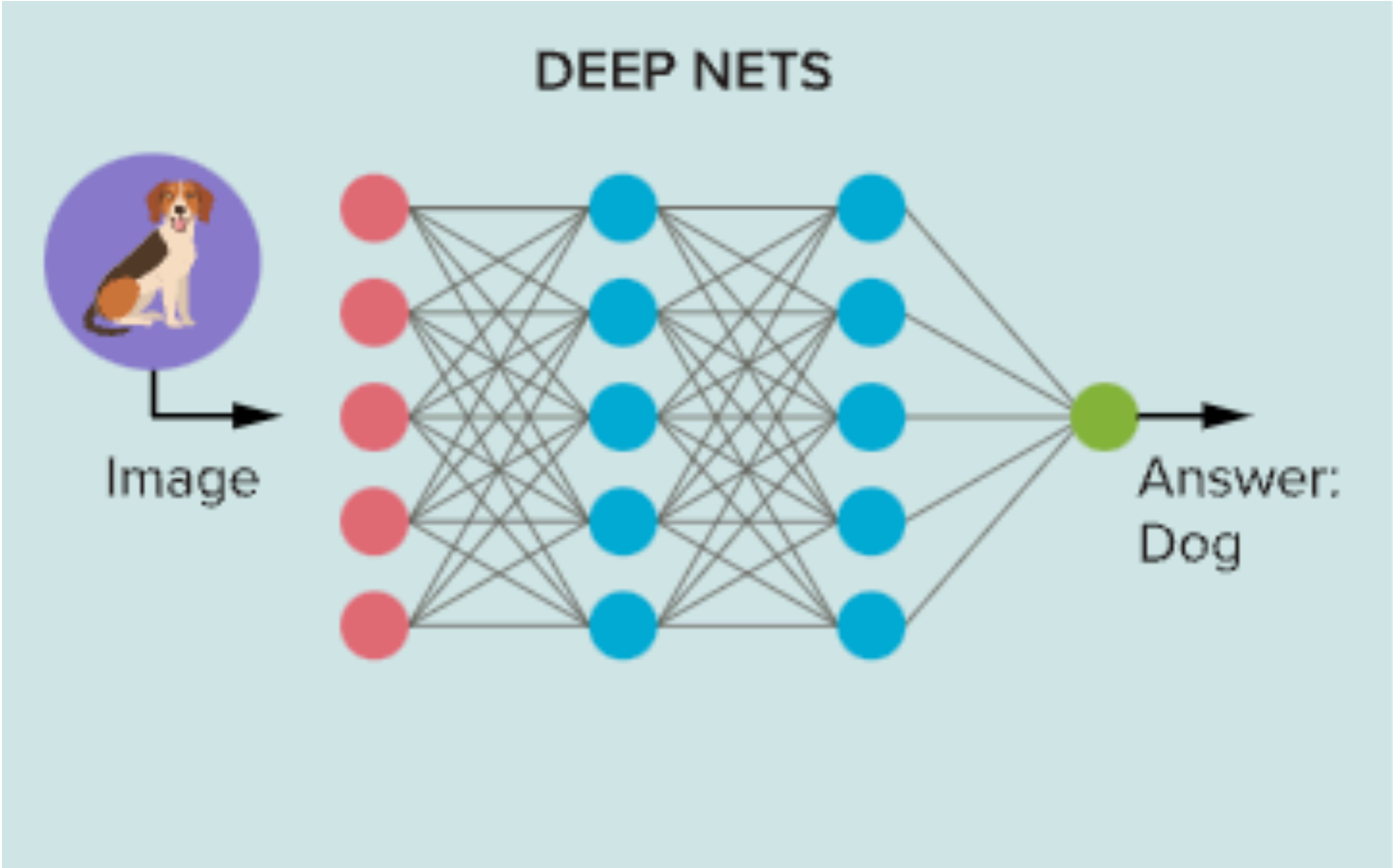
Symbolic vs. Subsymbolic AI

Physical symbol system hypothesis: manipulating **symbols** and **relations**

Symbolic AI



Subsymbolic AI

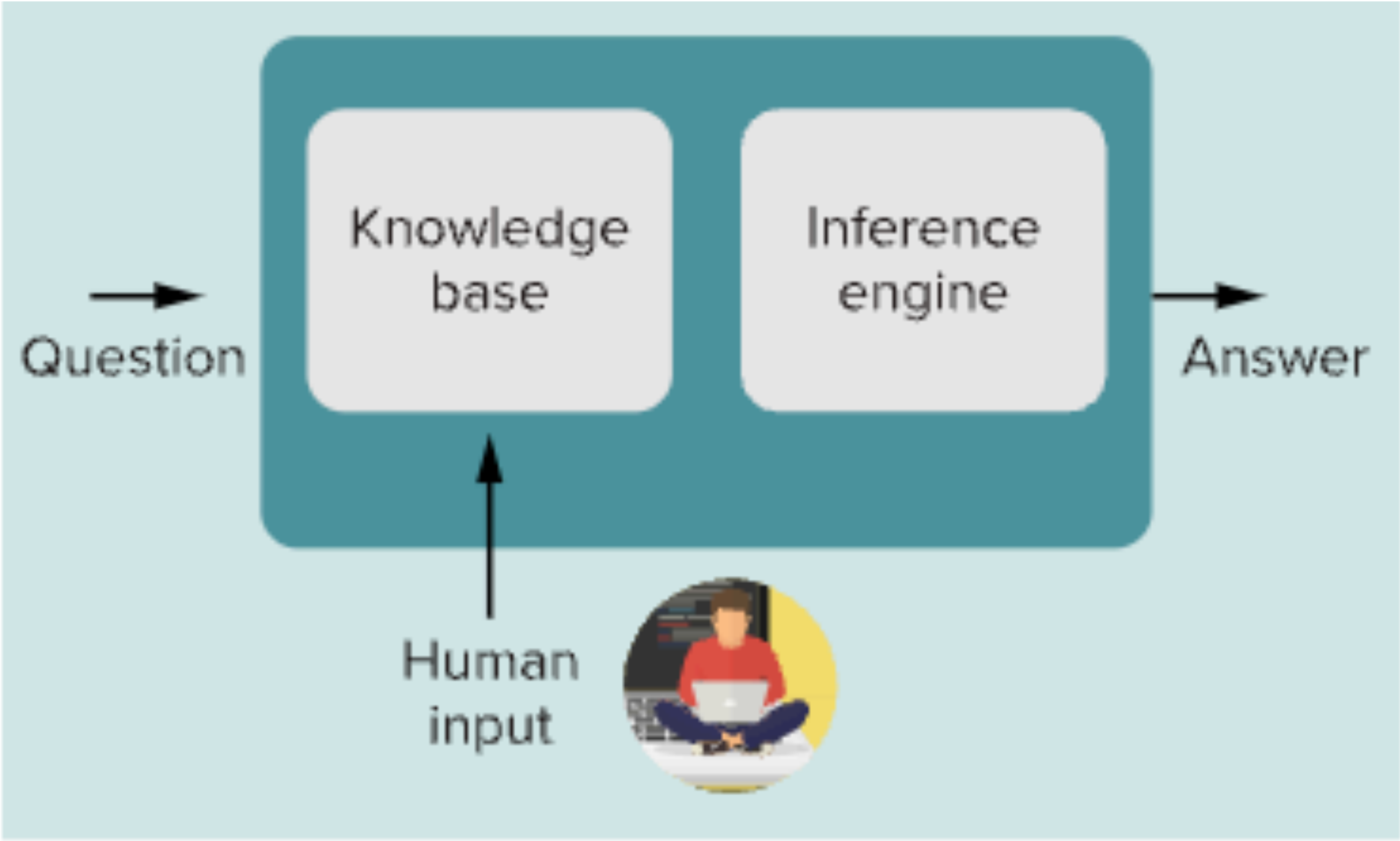


Representations are **distributed** across the weights of the neural network

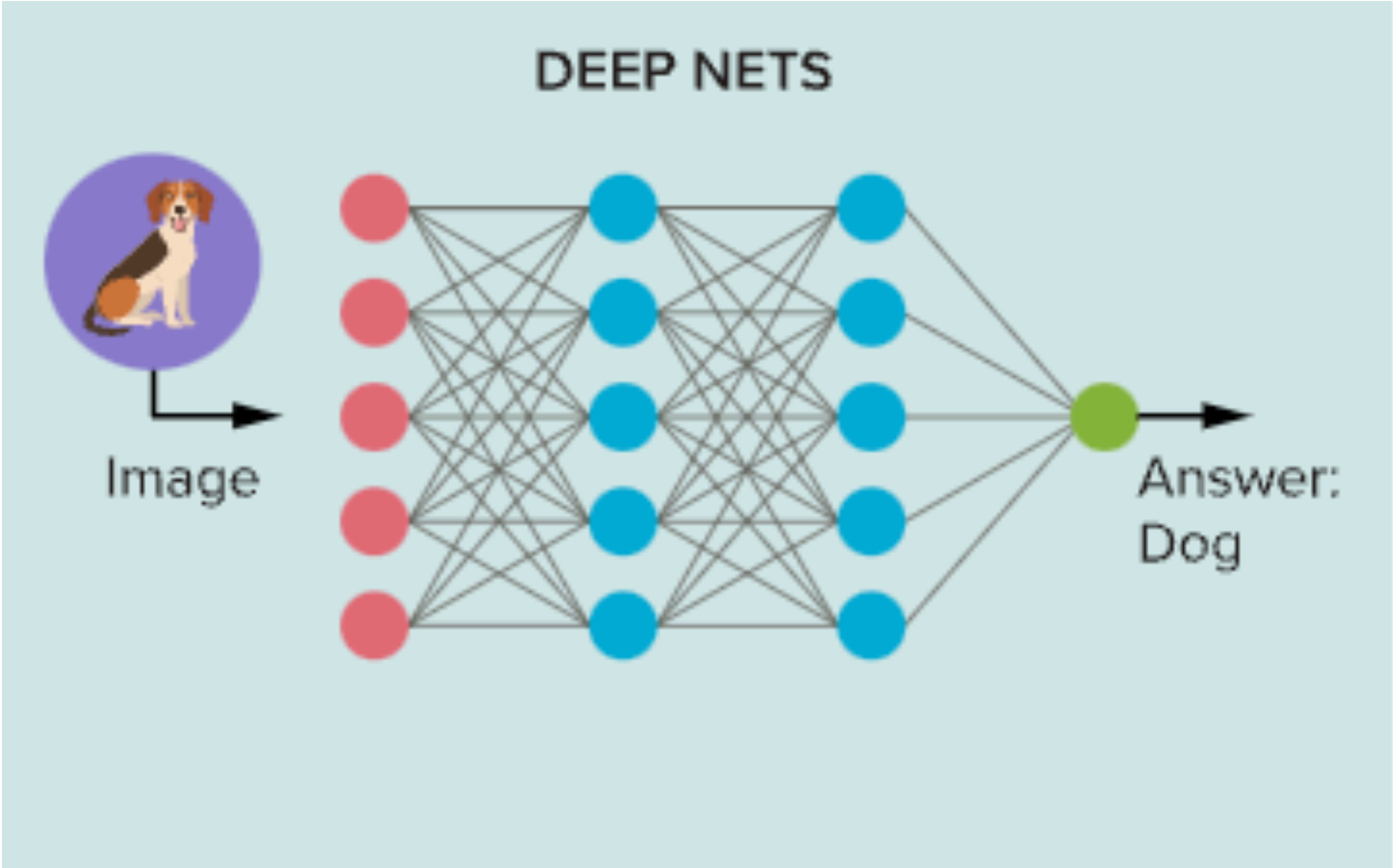
Symbolic vs. Subsymbolic AI

Physical symbol system hypothesis: manipulating **symbols** and **relations**

Symbolic AI



Subsymbolic AI



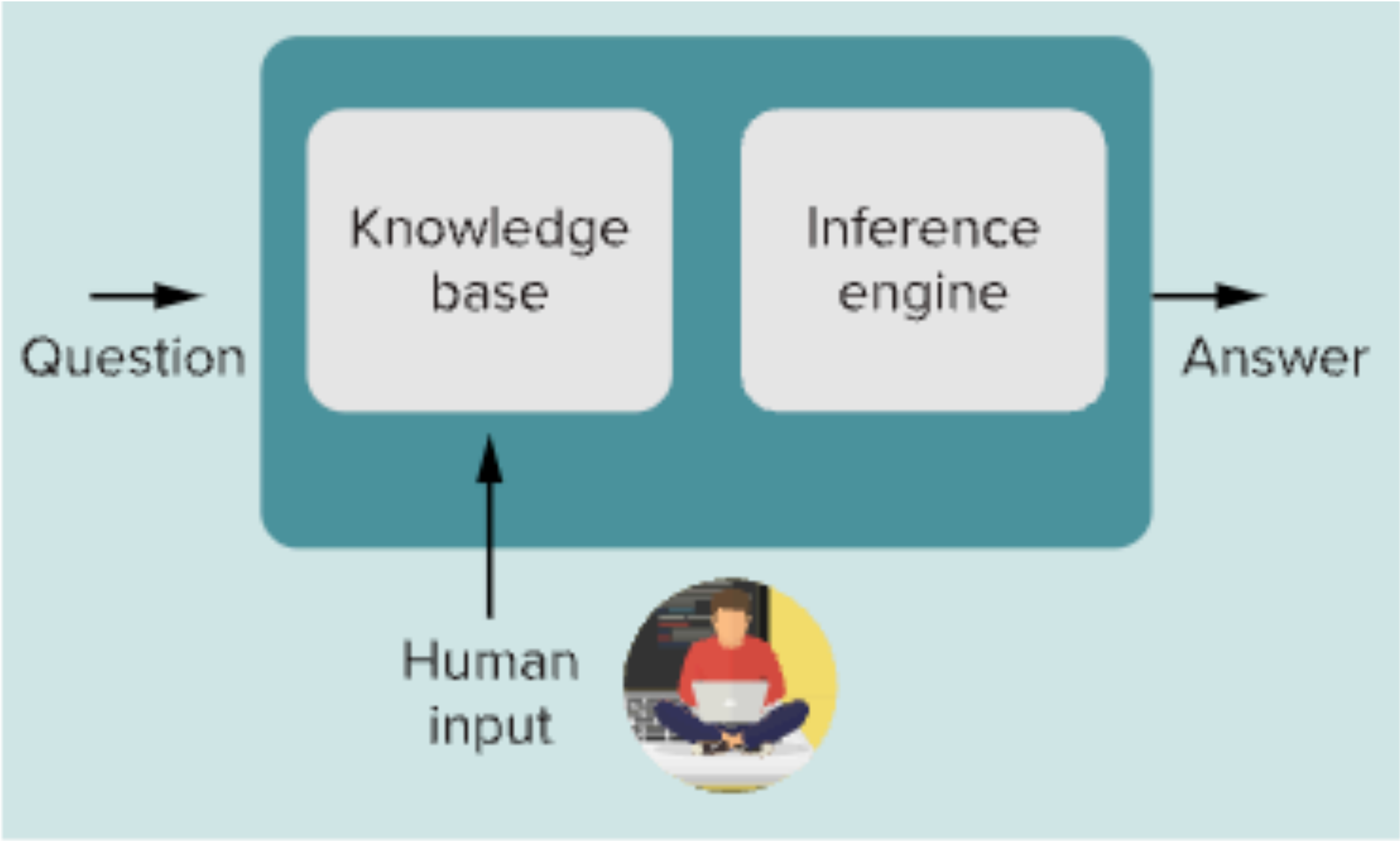
Representations are **distributed** across the weights of the neural network

Subsymbolic: the units of representation (i.e., weights) don't represent anything themselves

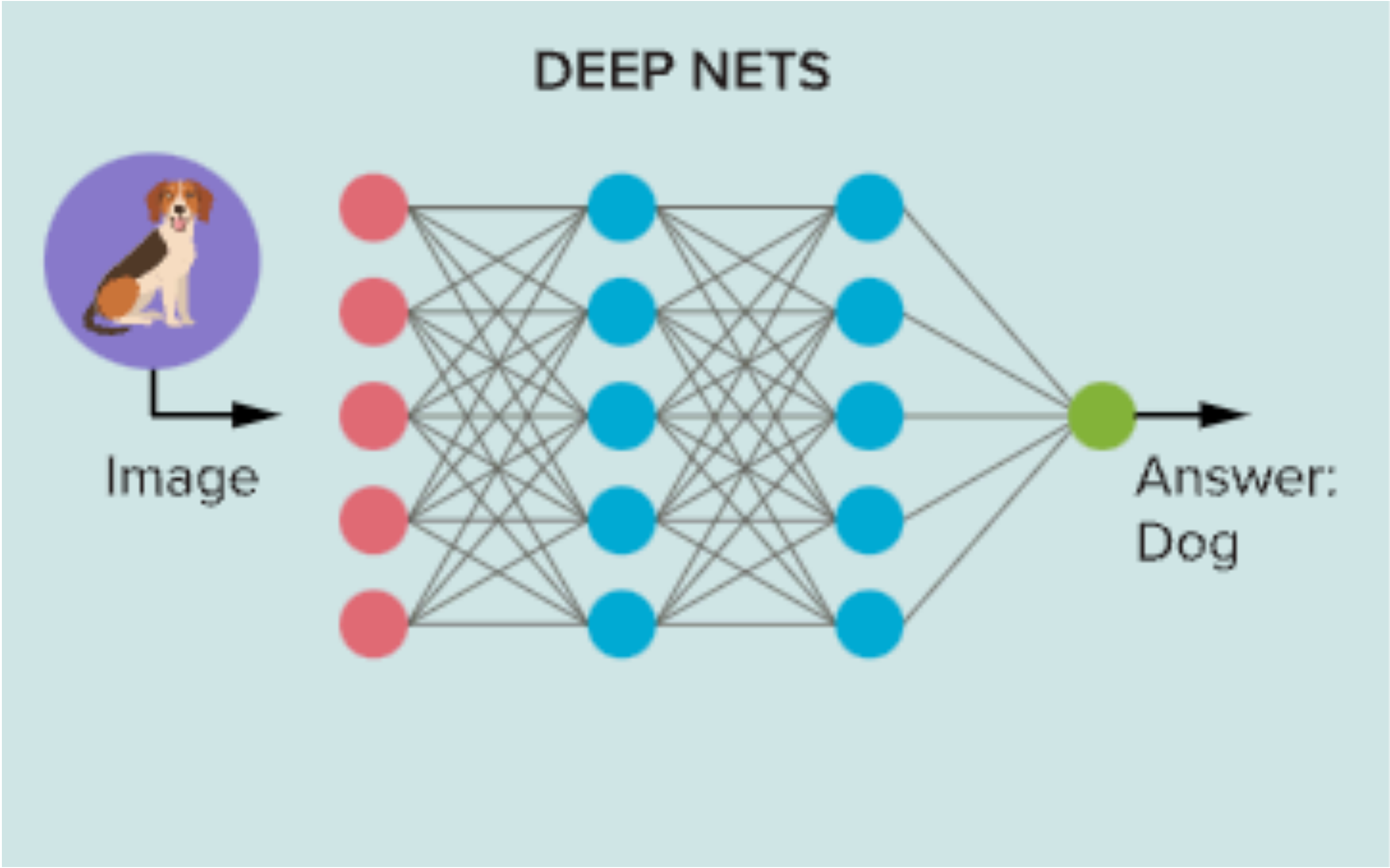
Symbolic vs. Subsymbolic AI

Physical symbol system hypothesis: manipulating **symbols** and **relations**

Symbolic AI



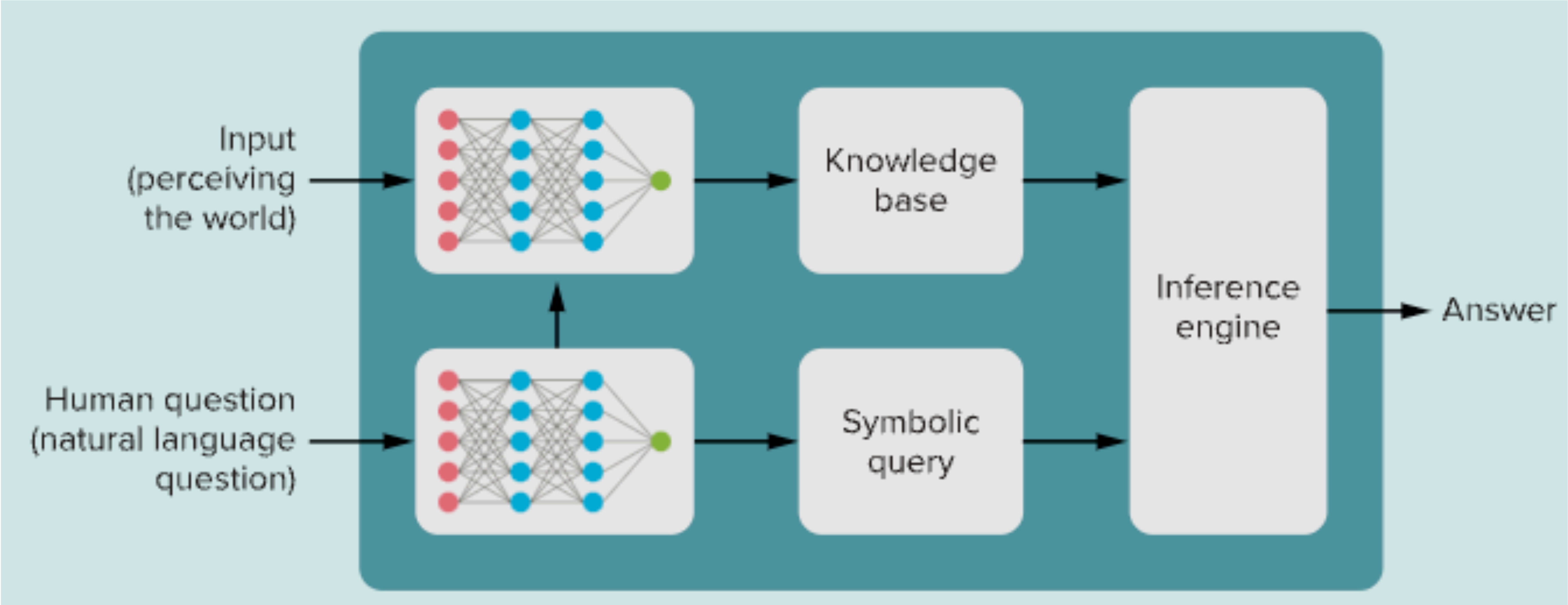
Subsymbolic AI



Representations are **distributed** across the weights of the neural network

Subsymbolic: the units of representation (i.e., weights) don't represent anything themselves

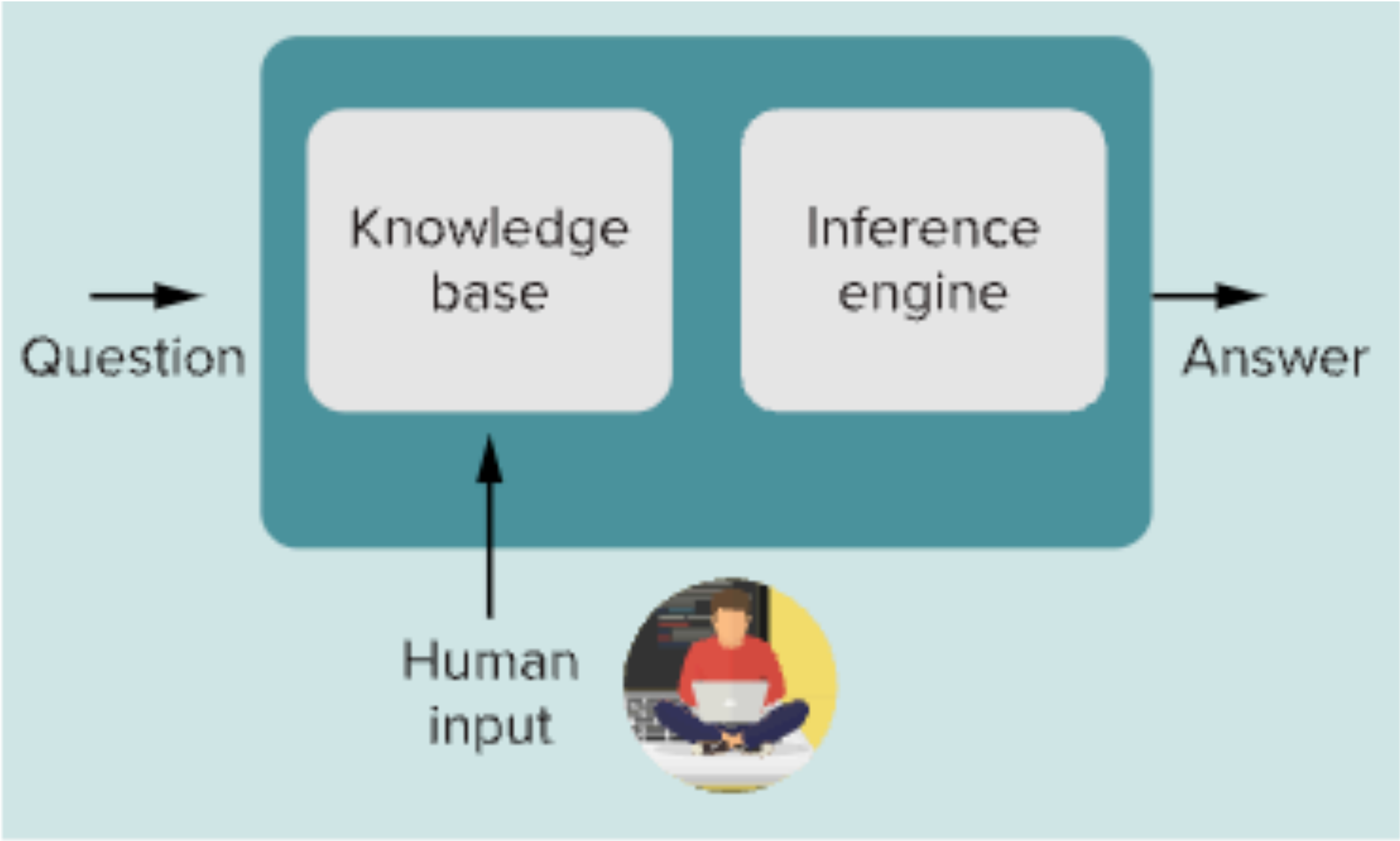
Hybrid systems



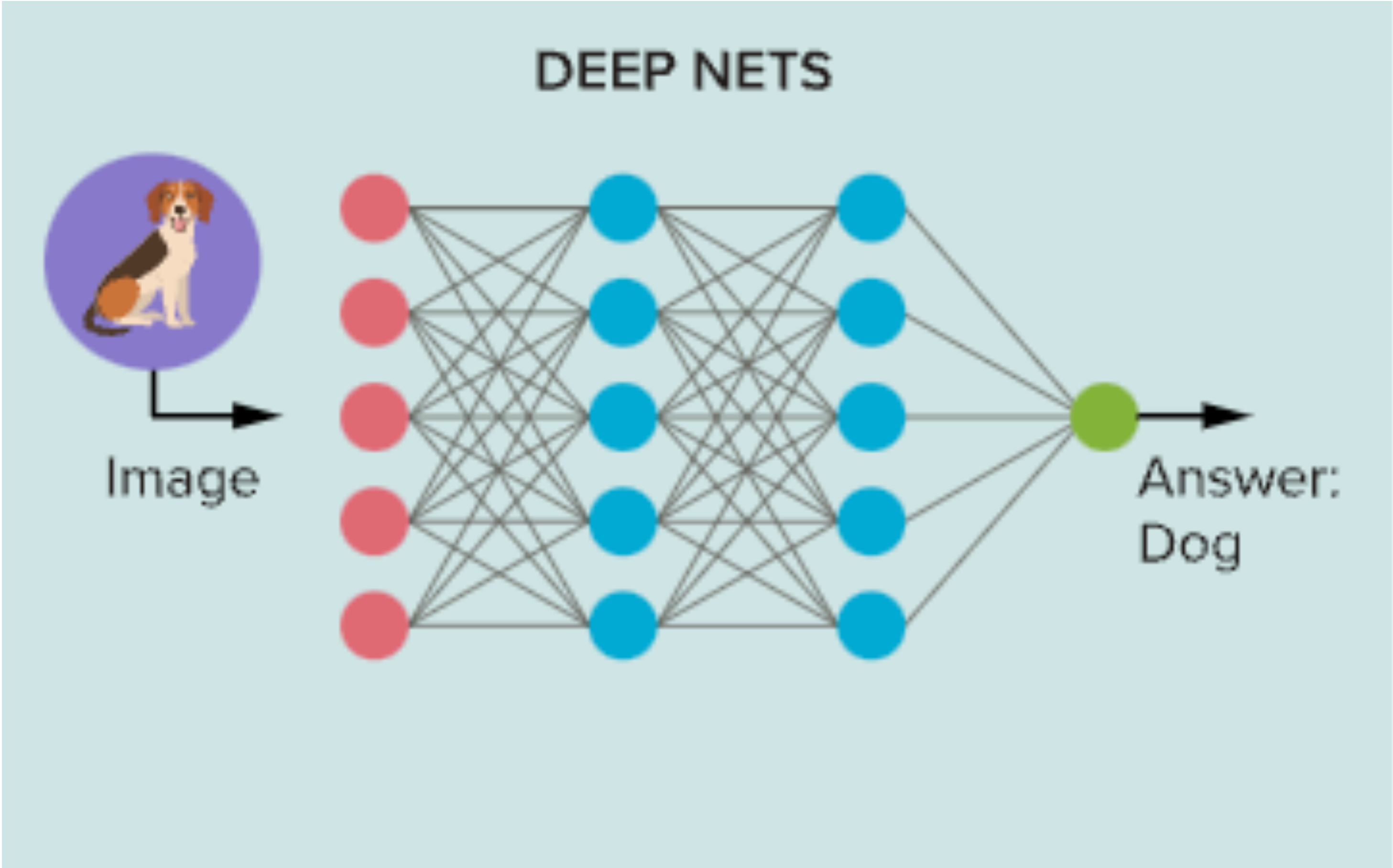
Symbolic vs. Subsymbolic AI

Physical symbol system hypothesis: manipulating **symbols** and **relations**

Symbolic AI



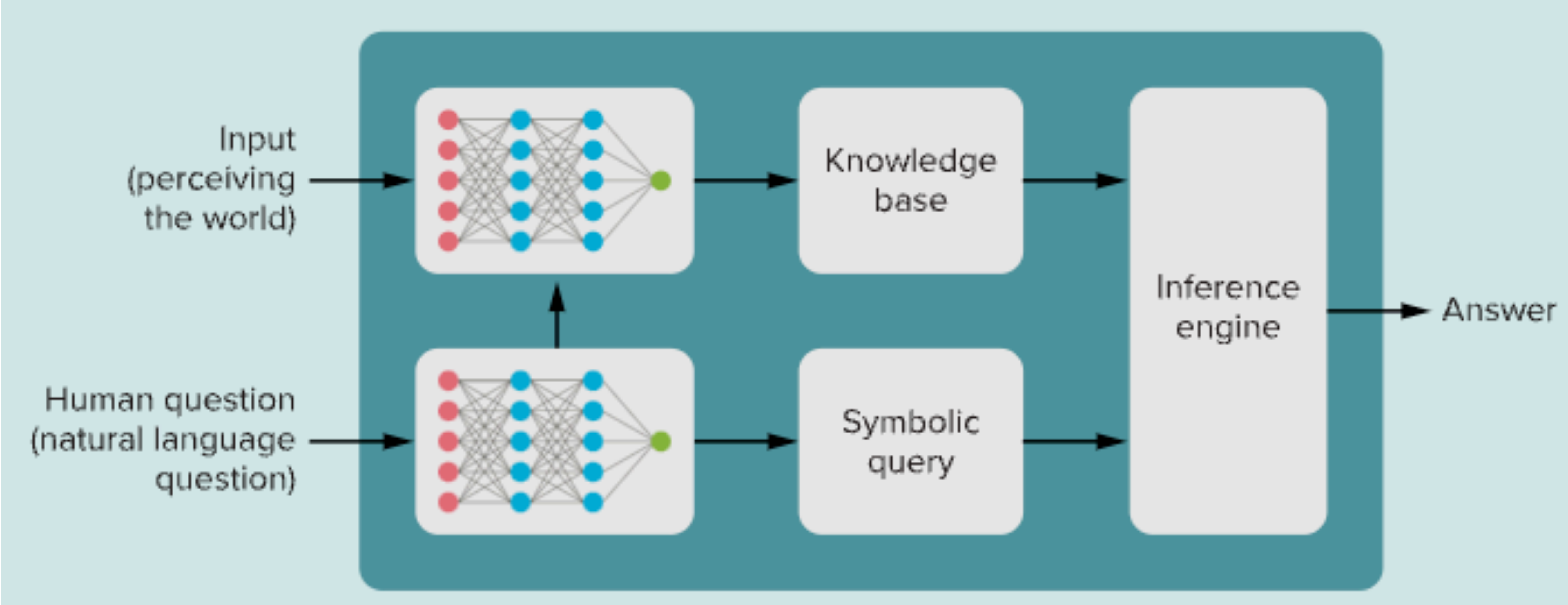
Subsymbolic AI



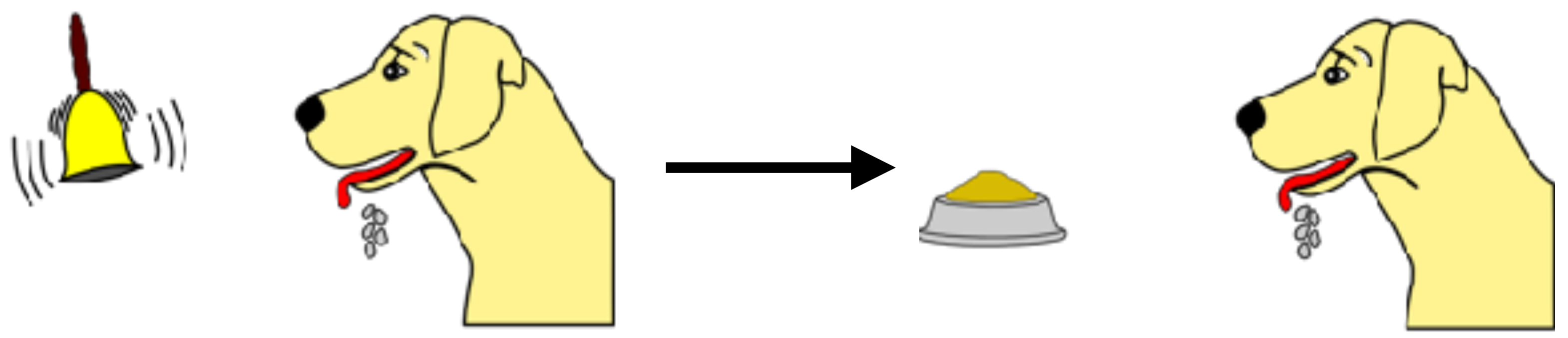
Representations are **distributed** across the weights of the neural network

Subsymbolic: the units of representation (i.e., weights) don't represent anything themselves

Hybrid systems

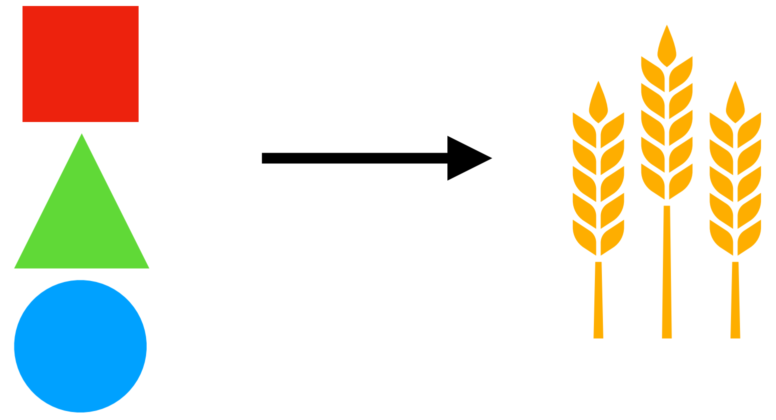


Pavlovian (classical) conditioning



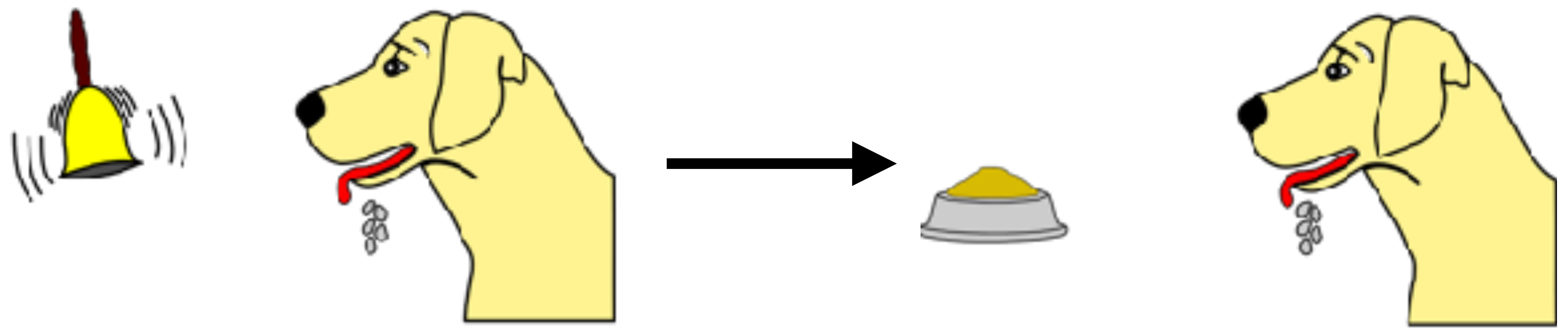
Learn which environmental cues *predict* reward

Operant (instrumental) conditioning

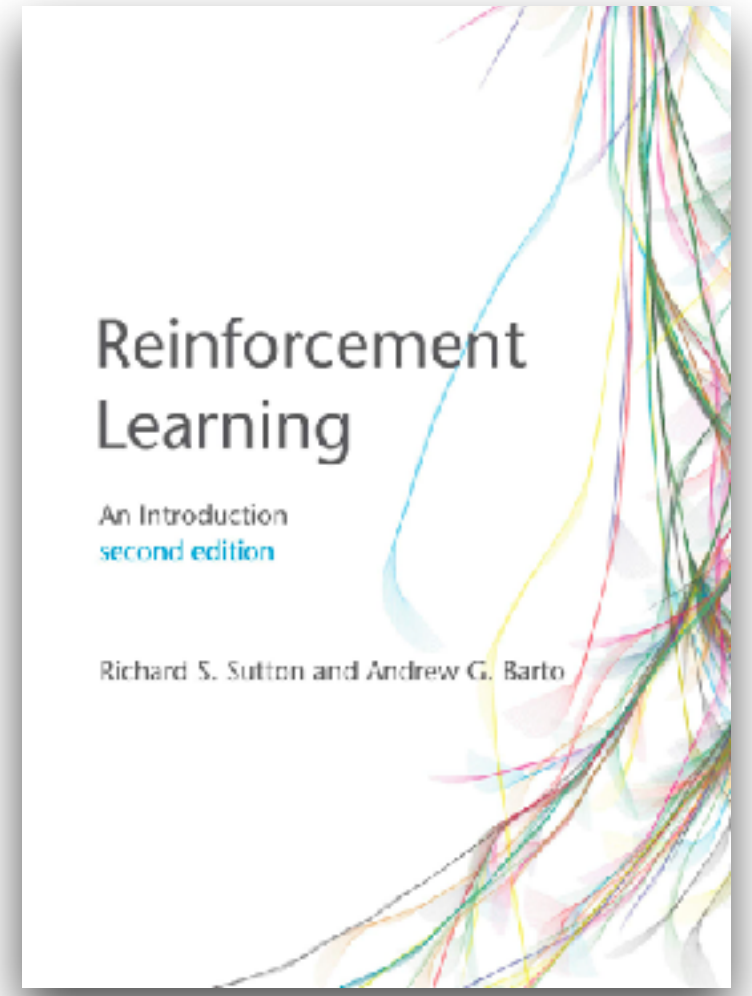


Learn which actions *predict* reward

Pavlovian (classical) conditioning

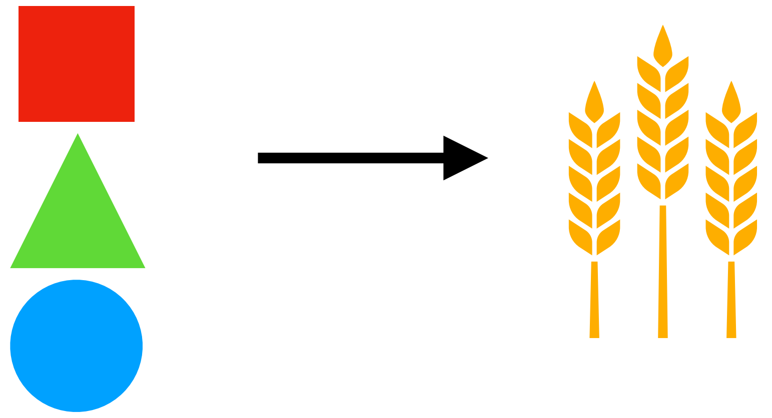


Learn which environmental cues *predict* reward



Reinforcement Learning

Operant (instrumental) conditioning

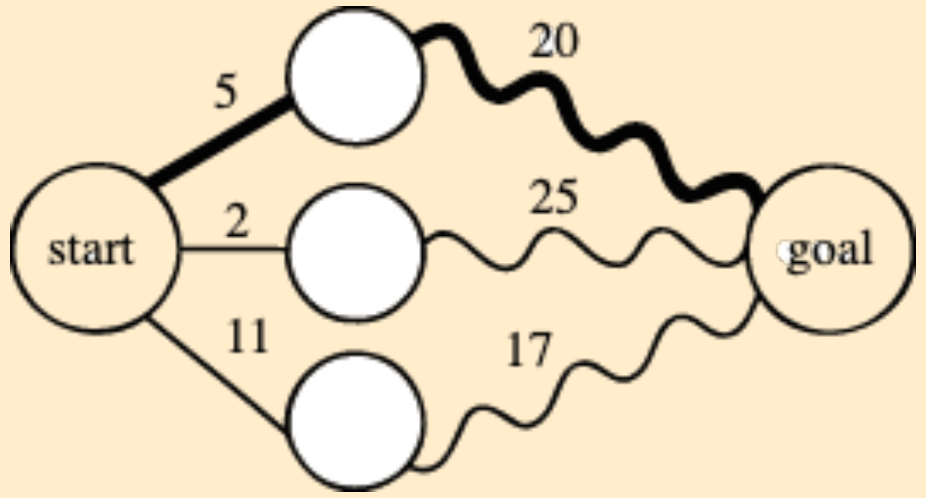


Learn which actions *predict* reward

Neuro-dynamic programing

Bertsekas & Tsitsiklis (1996)

Stochastic approximations to dynamic programing problems



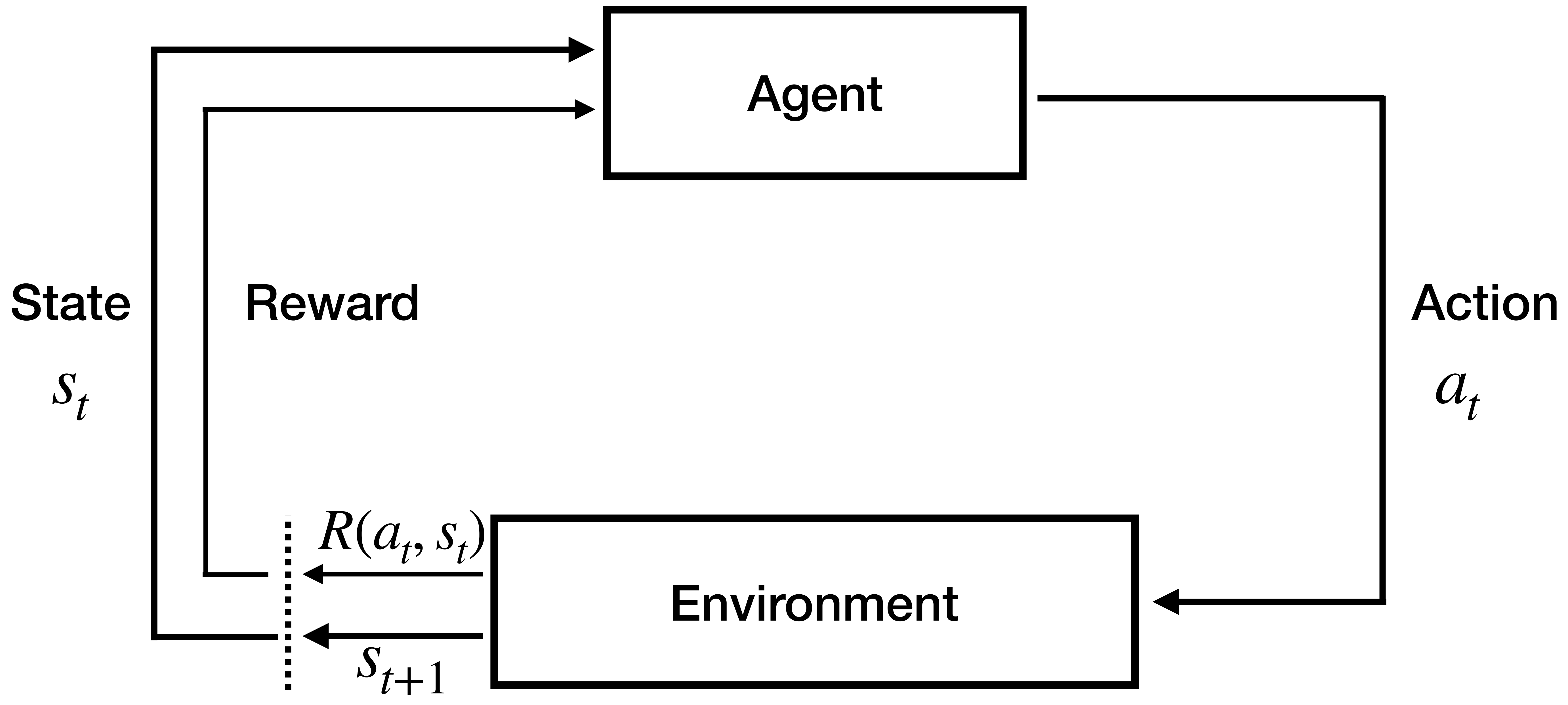
Reinforcement Learning

The Agent:

- Iteratively selects actions a_t based on a policy π
- Receives feedback from the environment in terms of new states s_{t+1} and rewards $R(a_t, s_t)$
- Updates internal representations
  - value $Q(s, a)$ or $V(s)$
  - model of the environment
    - reward function R
    - transitions $T(s' \mid s)$

The Environment:

- governs the transition between states $s_t \rightarrow s_{t+1}$
- provides rewards $R(a_t, s_t)$



Sutton and Barto (2018 [1998])

Reinforcement Learning

The Agent:

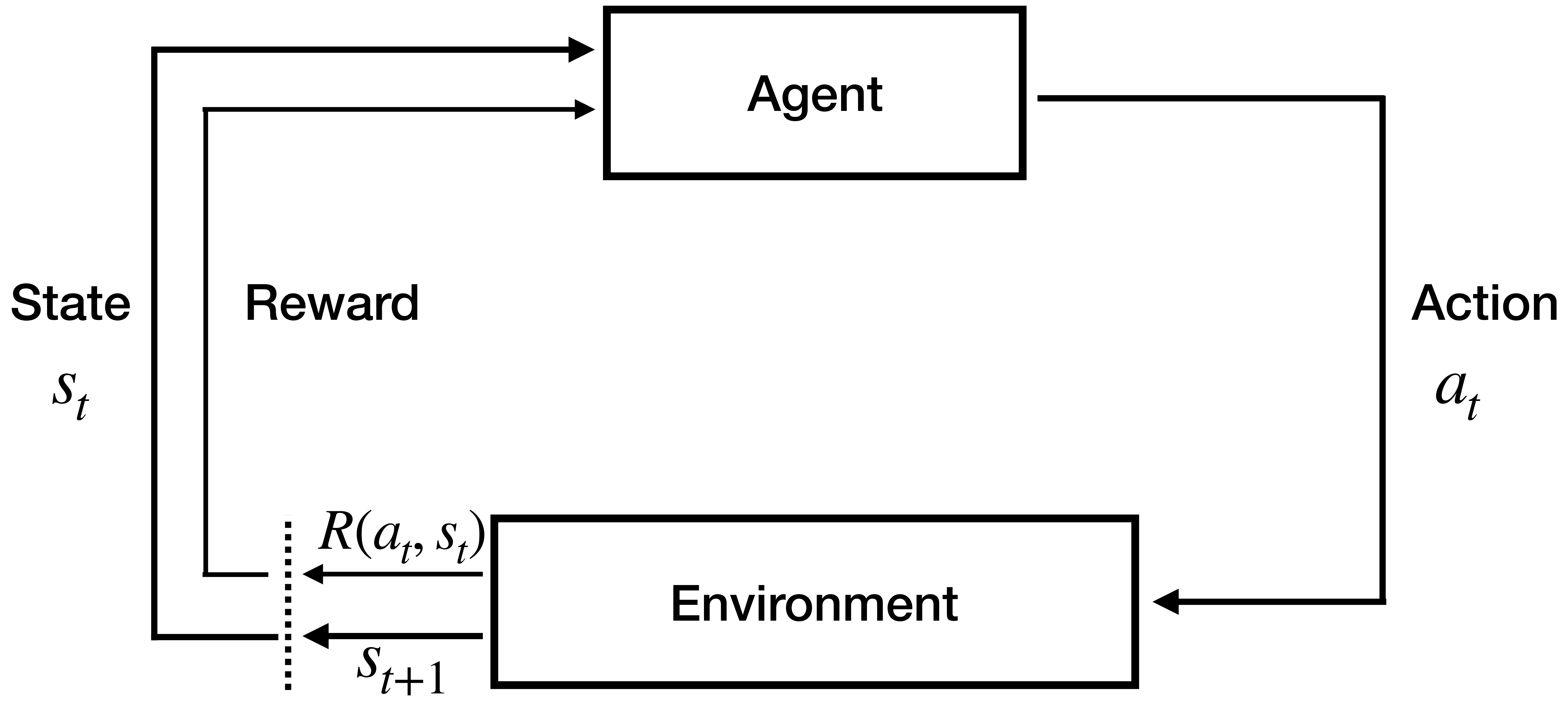
- Iteratively selects actions a_t based on a policy π
- Receives feedback from the environment in terms of new states s_{t+1} and rewards $R(a_t, s_t)$
- Updates internal representations
    - value $Q(s, a)$ or $V(s)$
    - model of the environment
        - reward function R
        - transitions $T(s' | s)$

The Environment:

- governs the transition between states $s_t \rightarrow s_{t+1}$
- provides rewards $R(a_t, s_t)$

Delta-rule of learning

Belief-updates are proportional to the magnitude of the reward predition error (RPE)

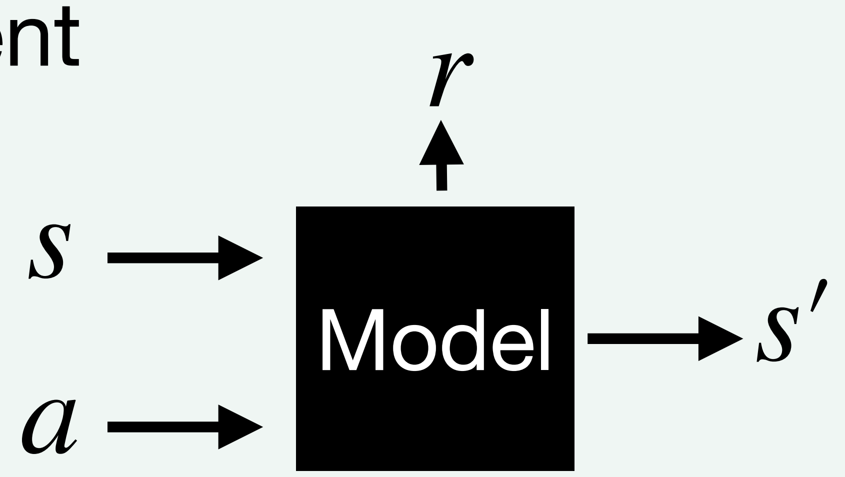


Sutton and Barto (2018 [1998])

Reinforcement Learning

The Agent:

- Iteratively selects actions a_t based on a policy π
- Receives feedback from the environment in terms of new states s_{t+1} and rewards $R(a_t, s_t)$
- Updates internal representations
    - value $Q(s, a)$ or $V(s)$
    - model of the environment
        - reward function R
        - transitions $T(s' \mid s)$

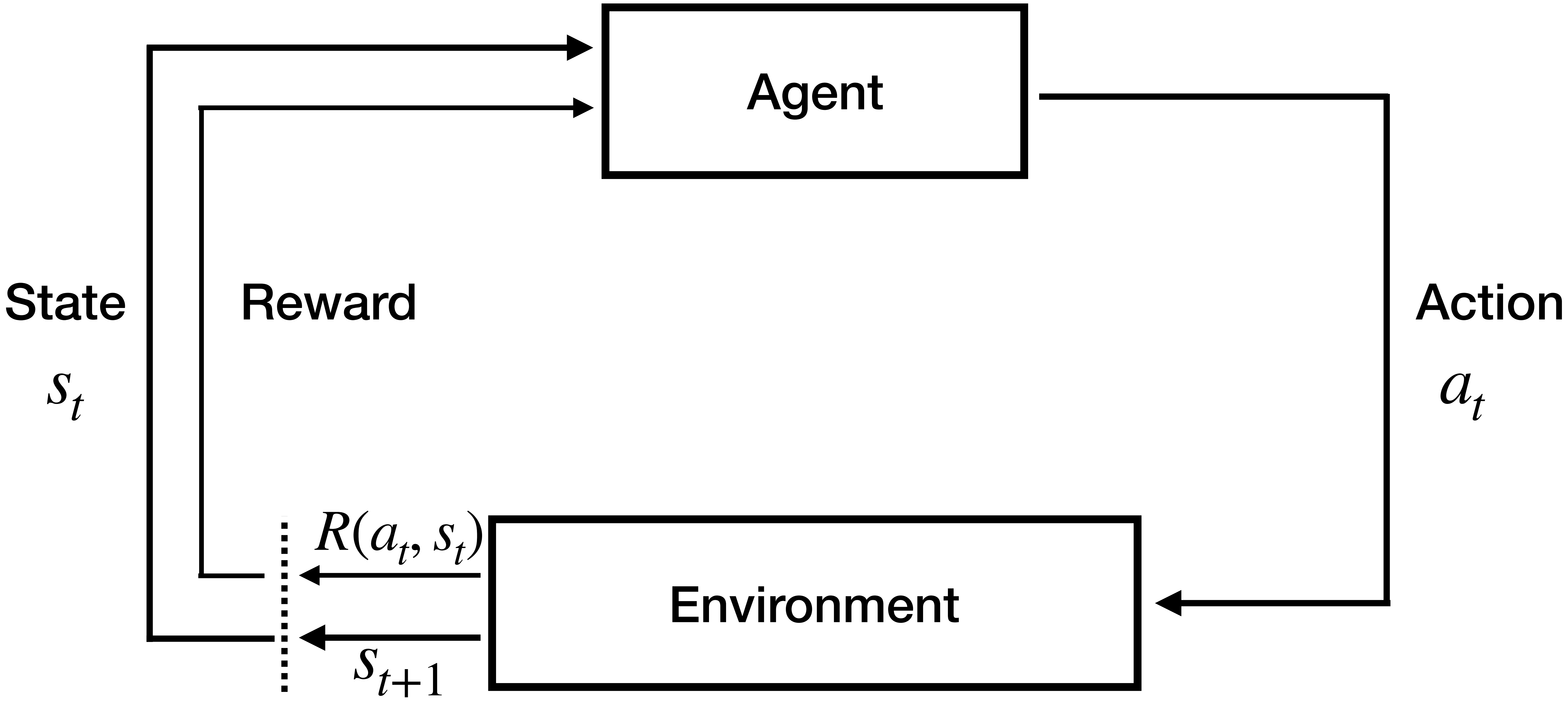


The Environment:

- governs the transition between states $s_t \rightarrow s_{t+1}$
- provides rewards $R(a_t, s_t)$

Delta-rule of learning

Belief-updates are proportional to the magnitude of the reward predition error (RPE)



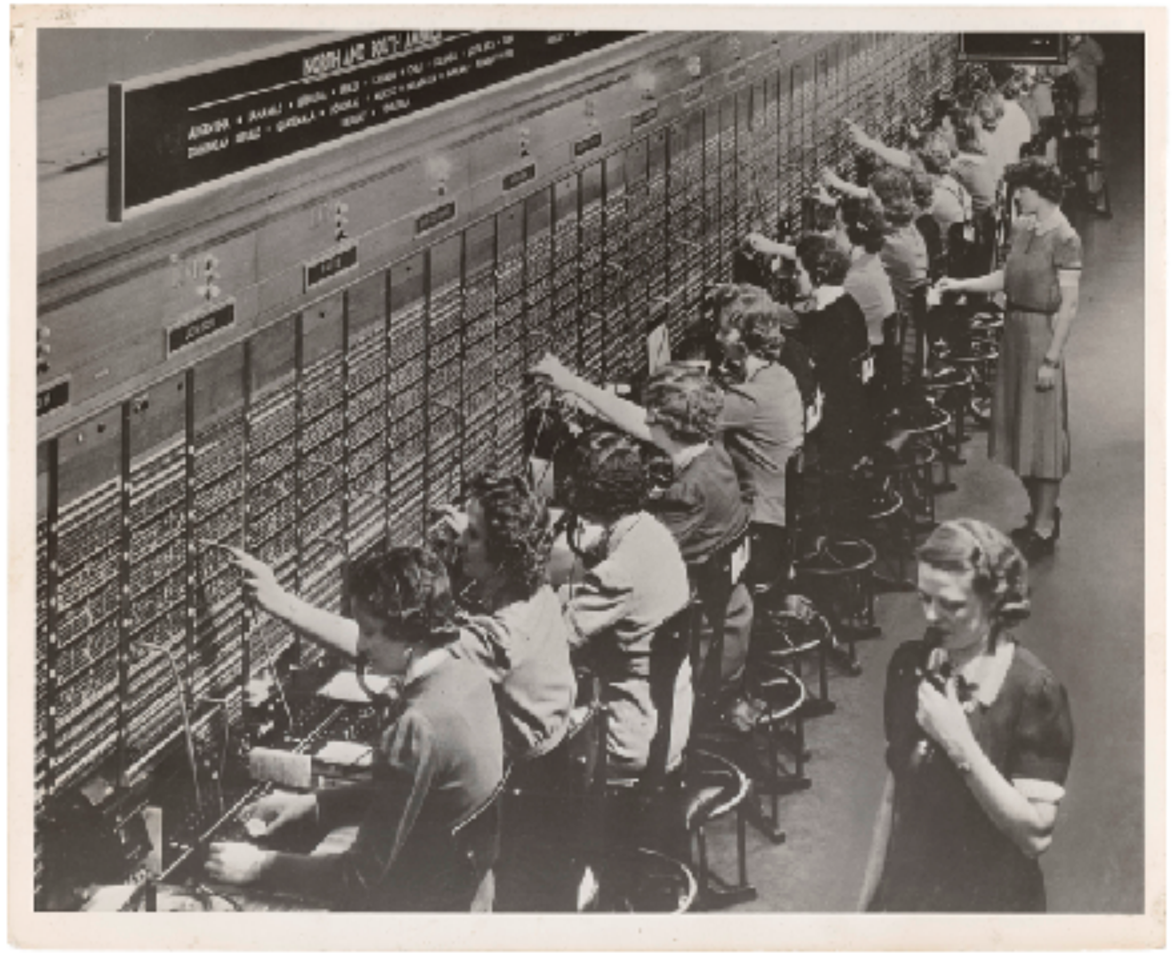
Sutton and Barto (2018 [1998])

Model-free Model-based

Model-free

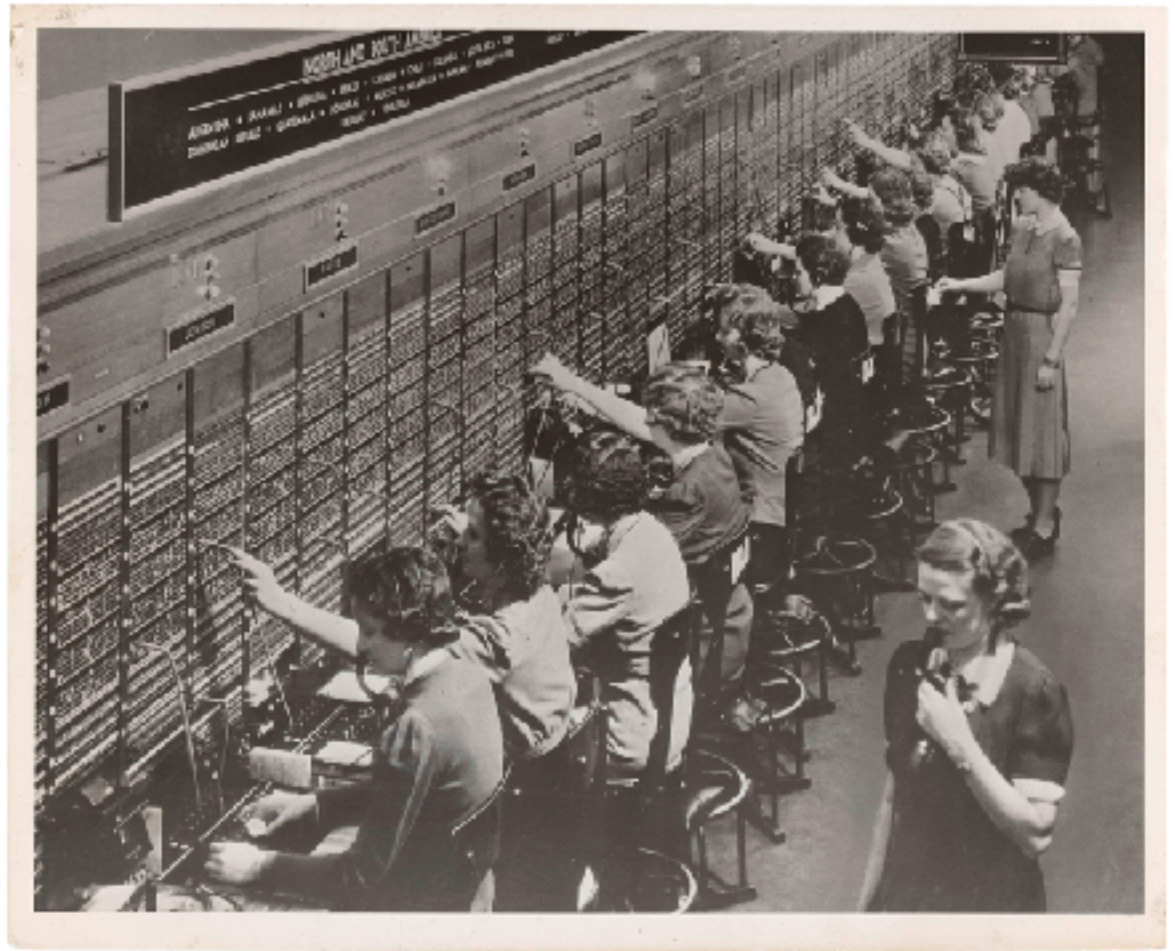
Model-based

S-R learning



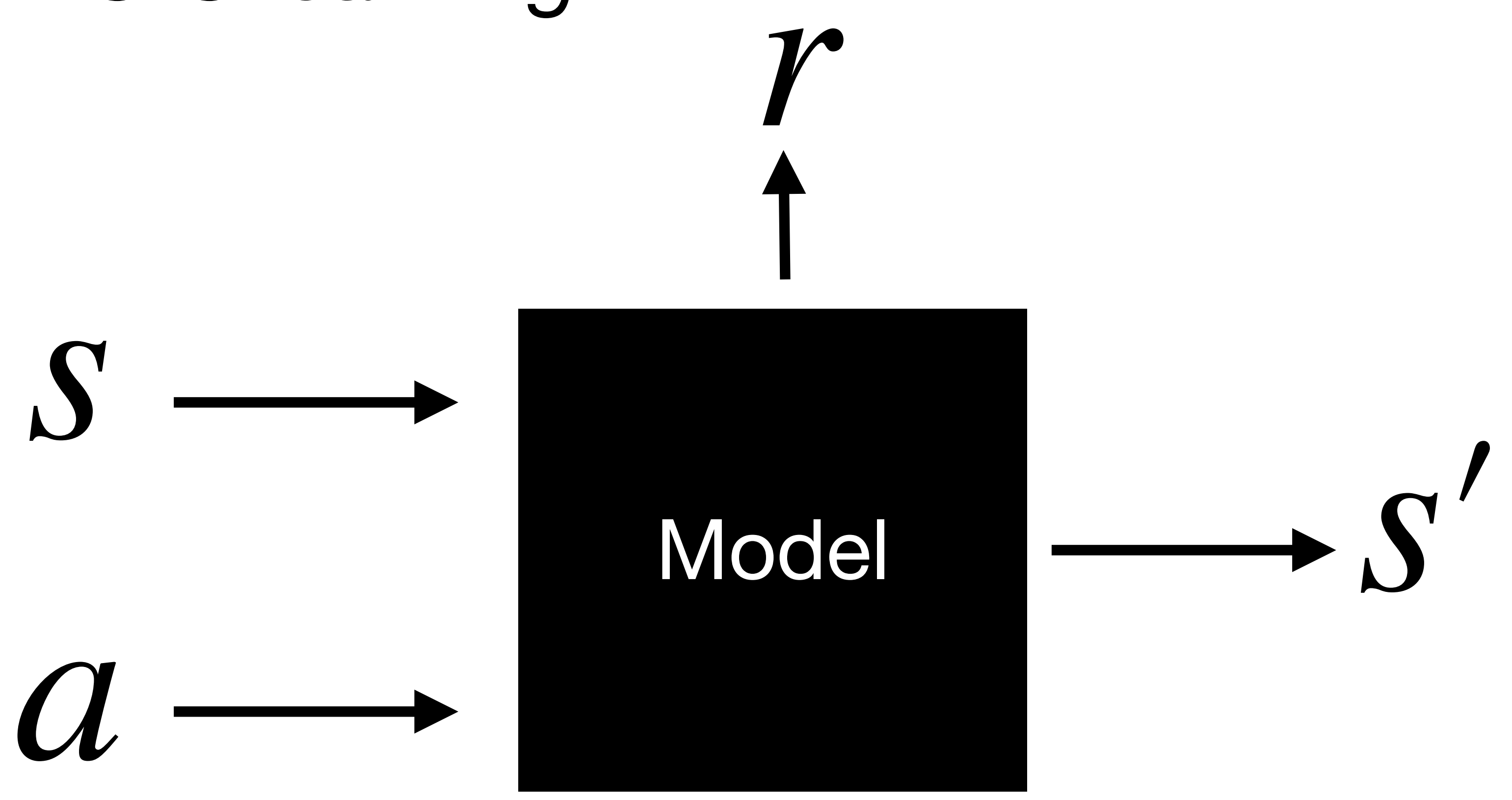
Model-free

S-R learning



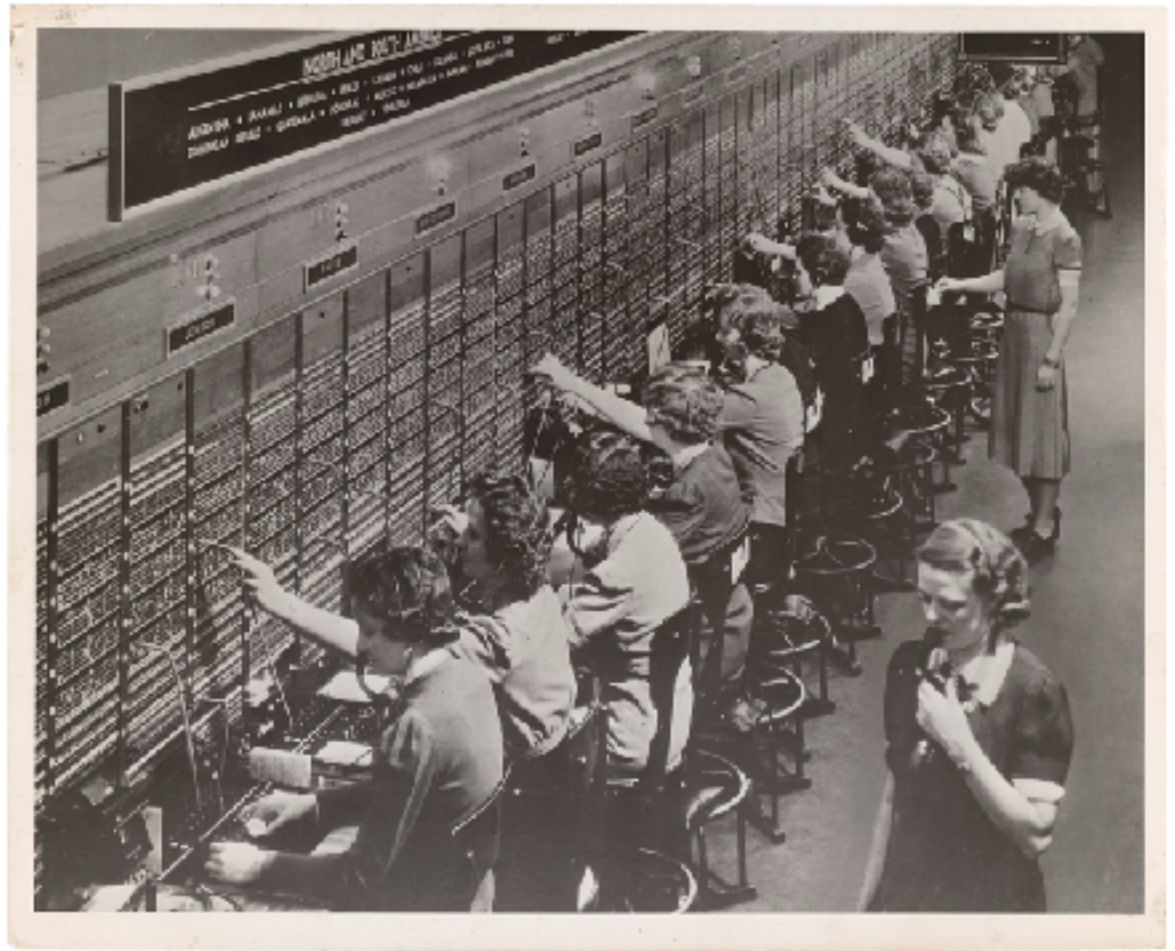
Model-based

S-S learning



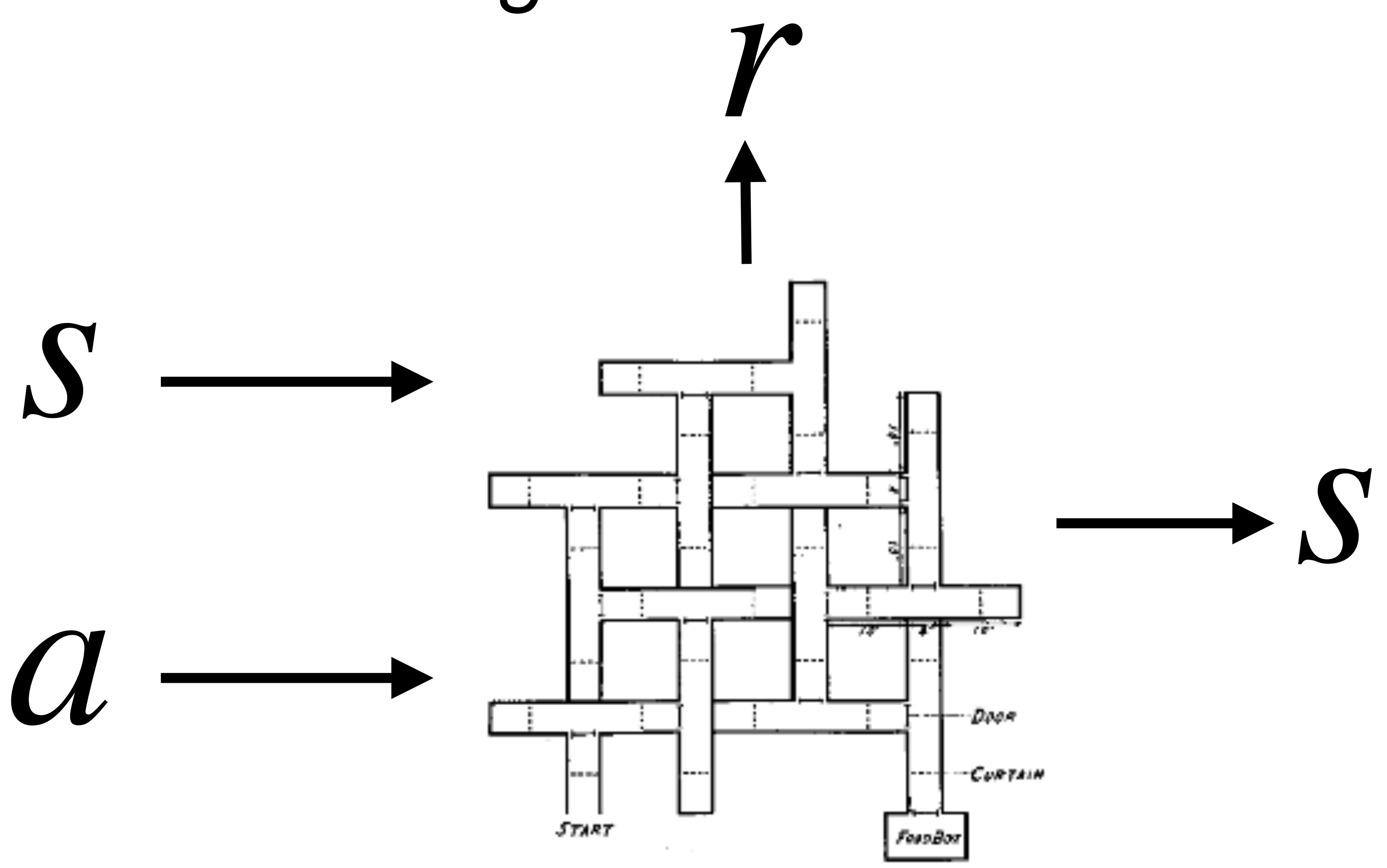
Model-free

S-R learning



Model-based

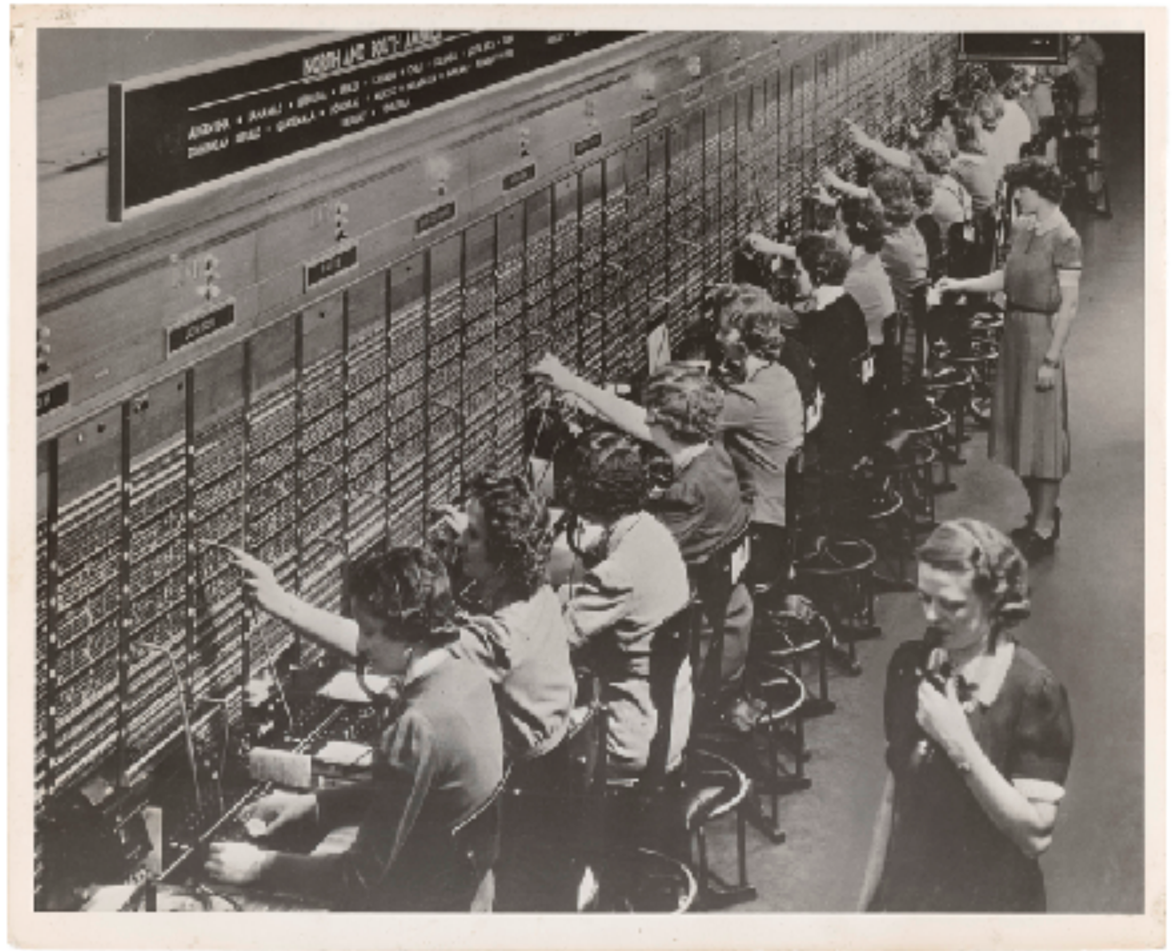
S-S learning



Tolman (1948)

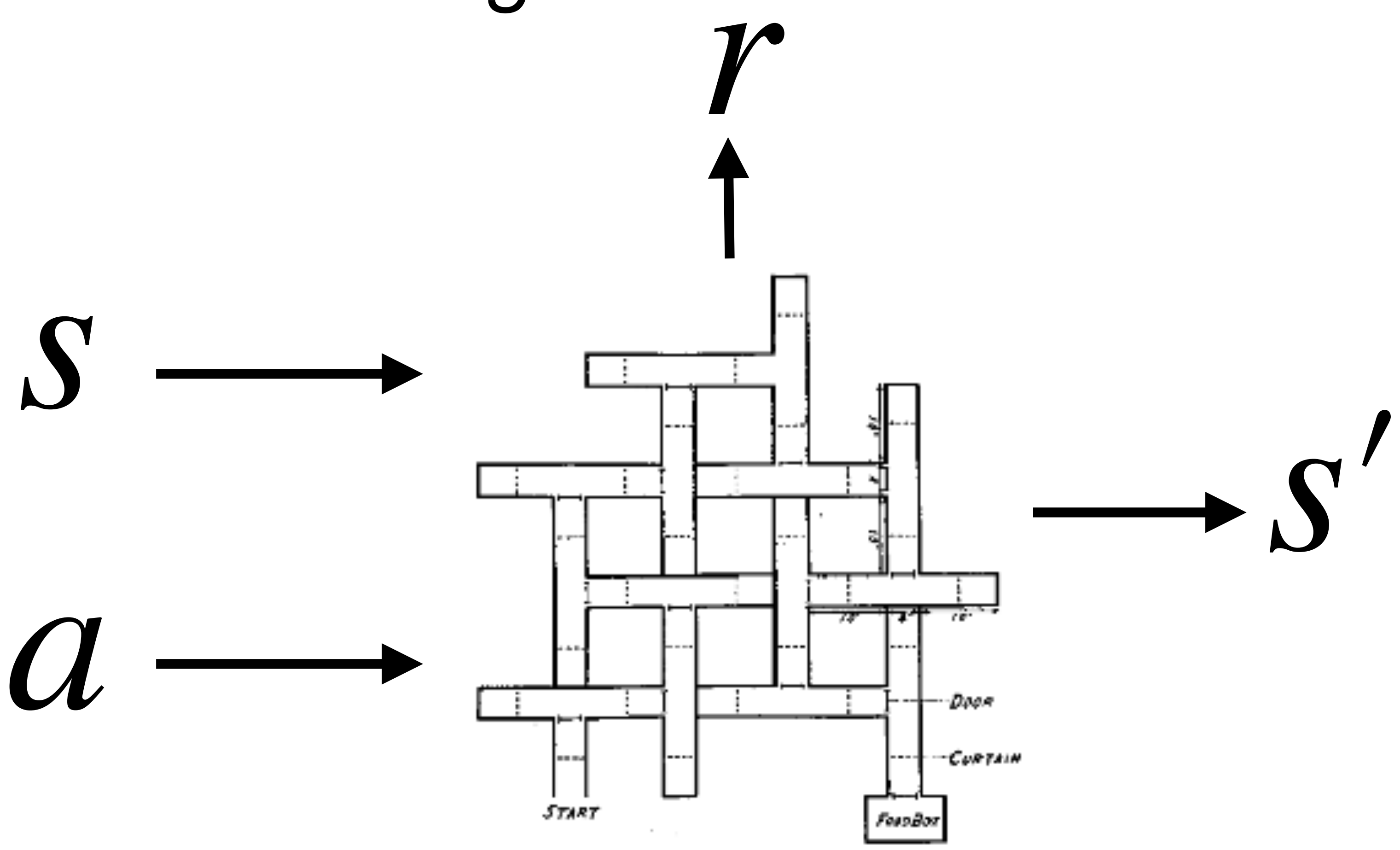
Model-free

S-R learning



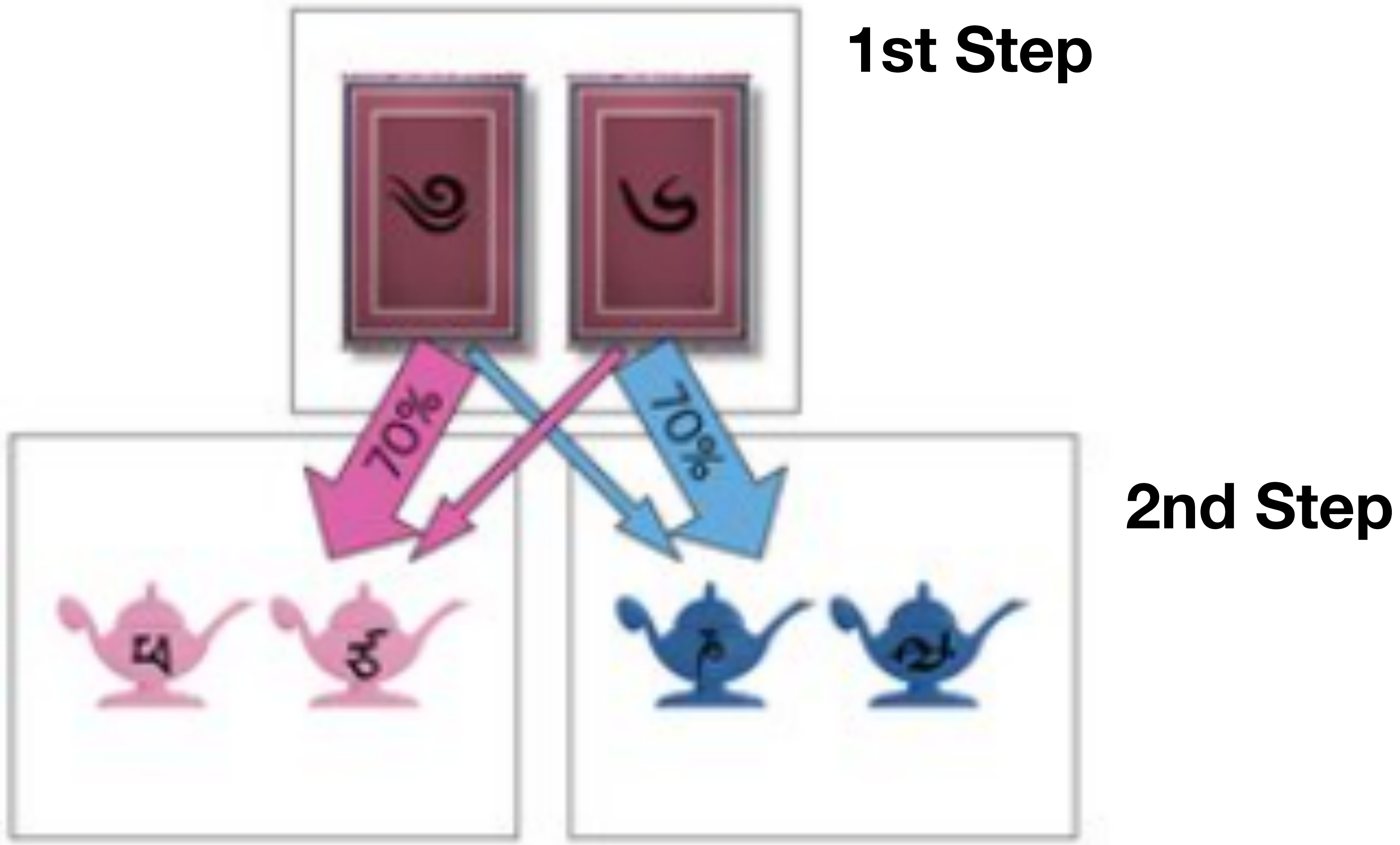
Model-based

S-S learning



Tolman (1948)

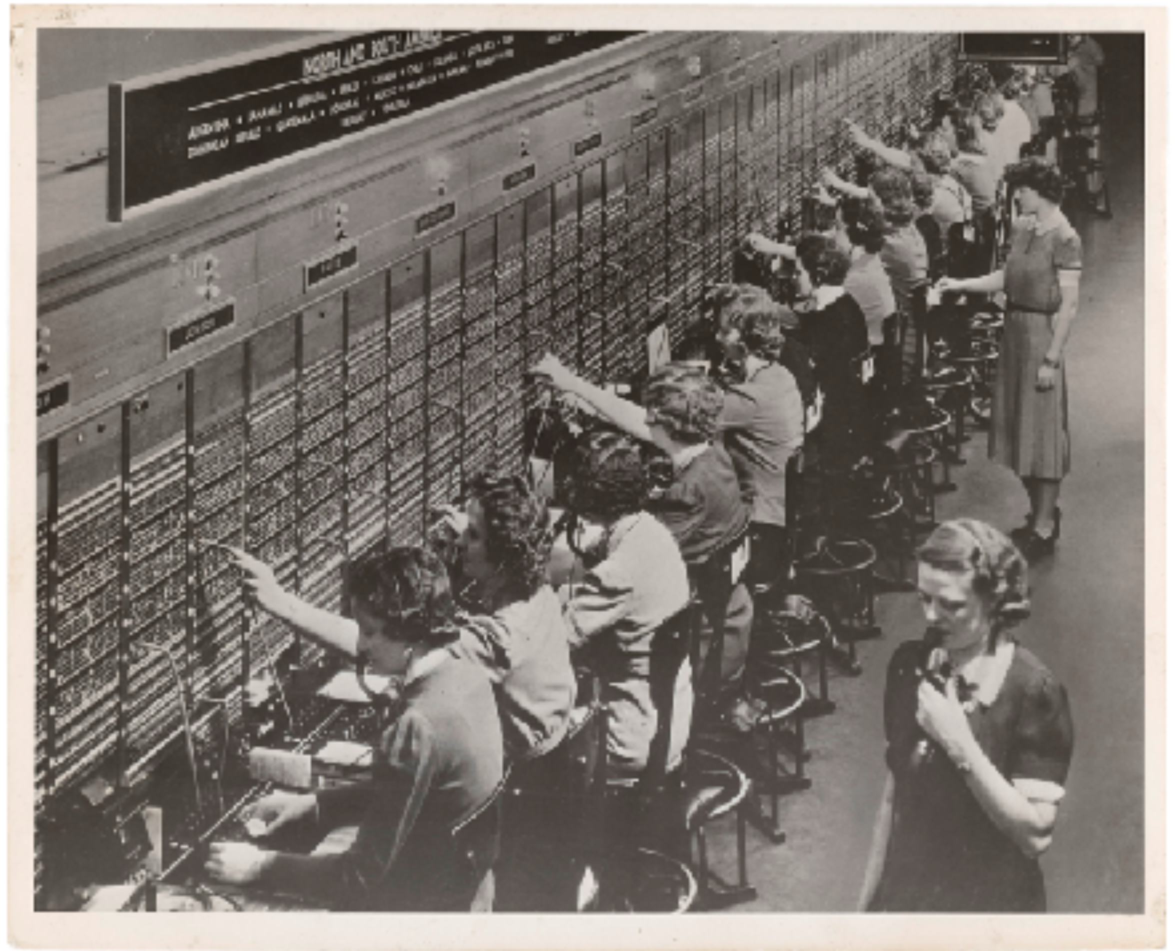
2-step task

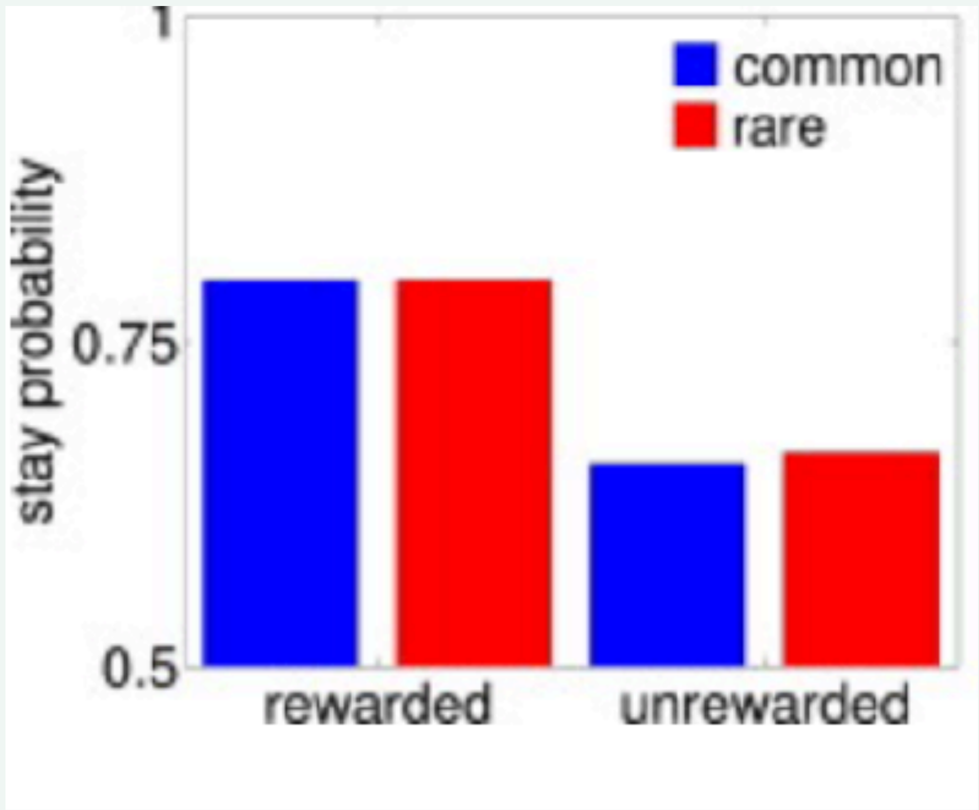


Feher da Silva et al., (2023); Daw et al., (2011)

Model-free

S-R learning

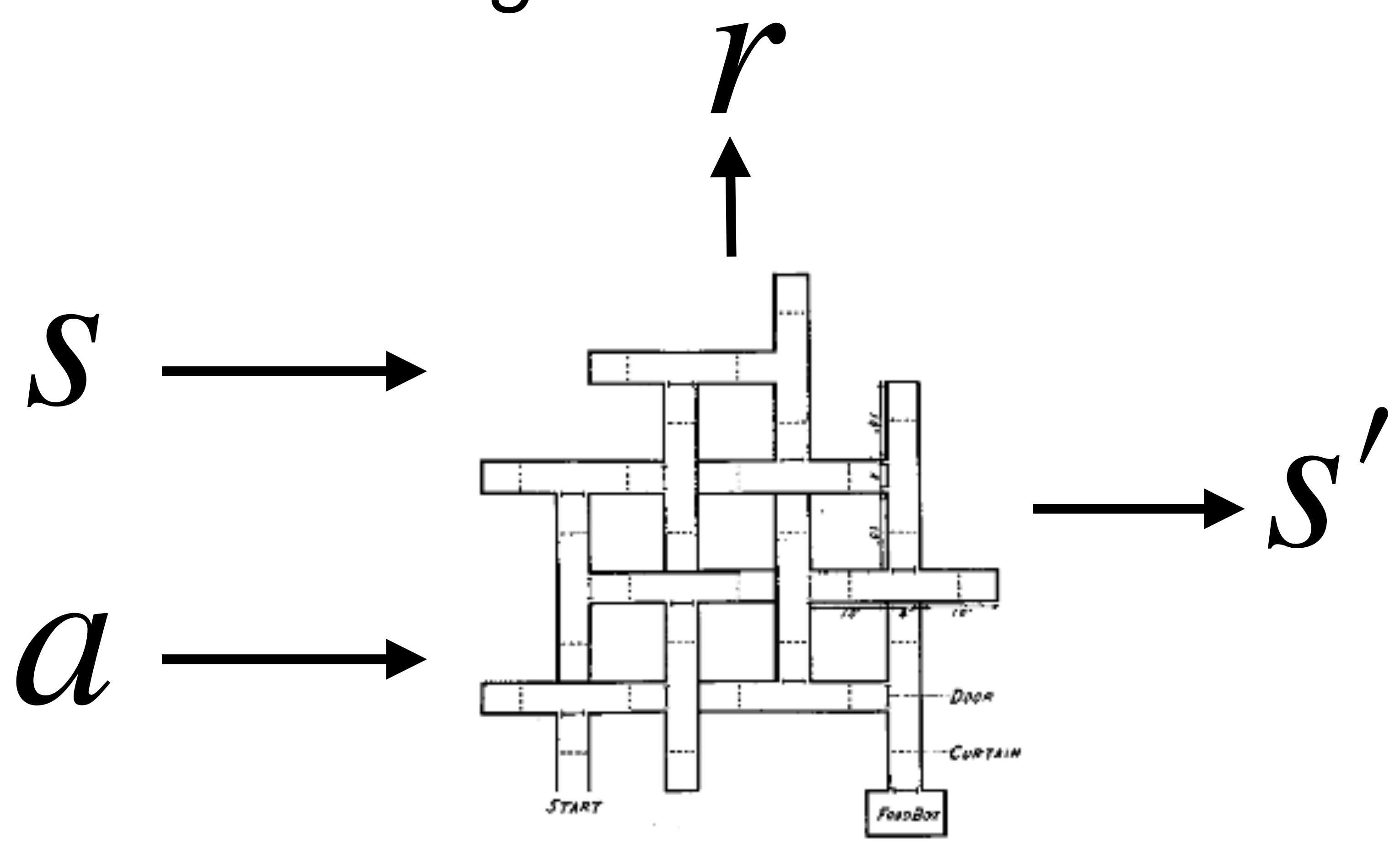




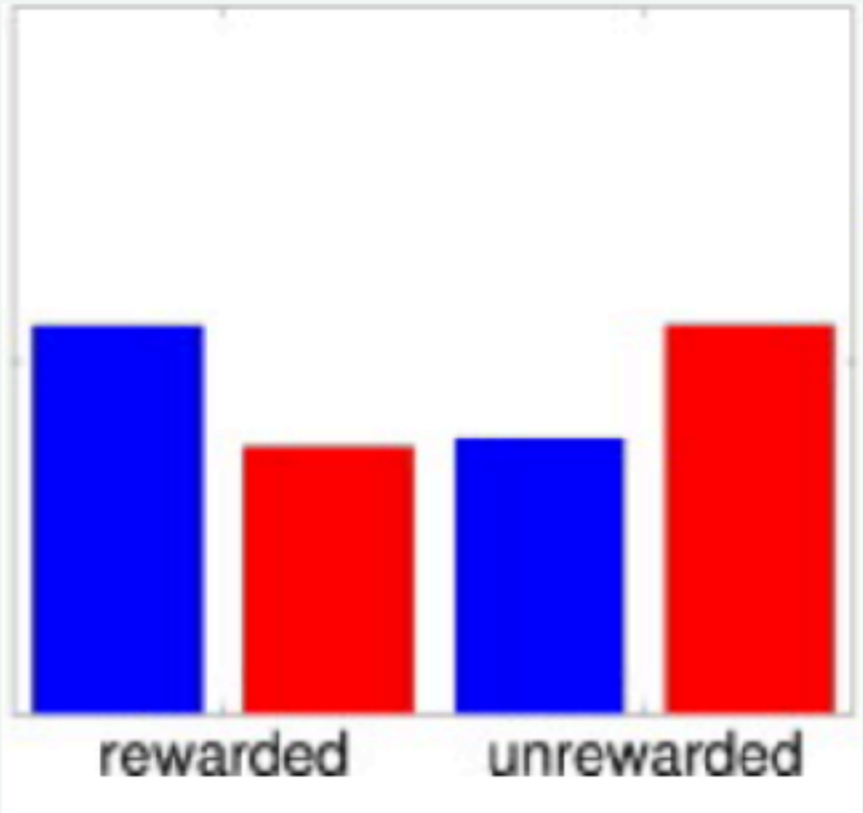
Only cares whether actions were rewarded

Model-based

S-S learning

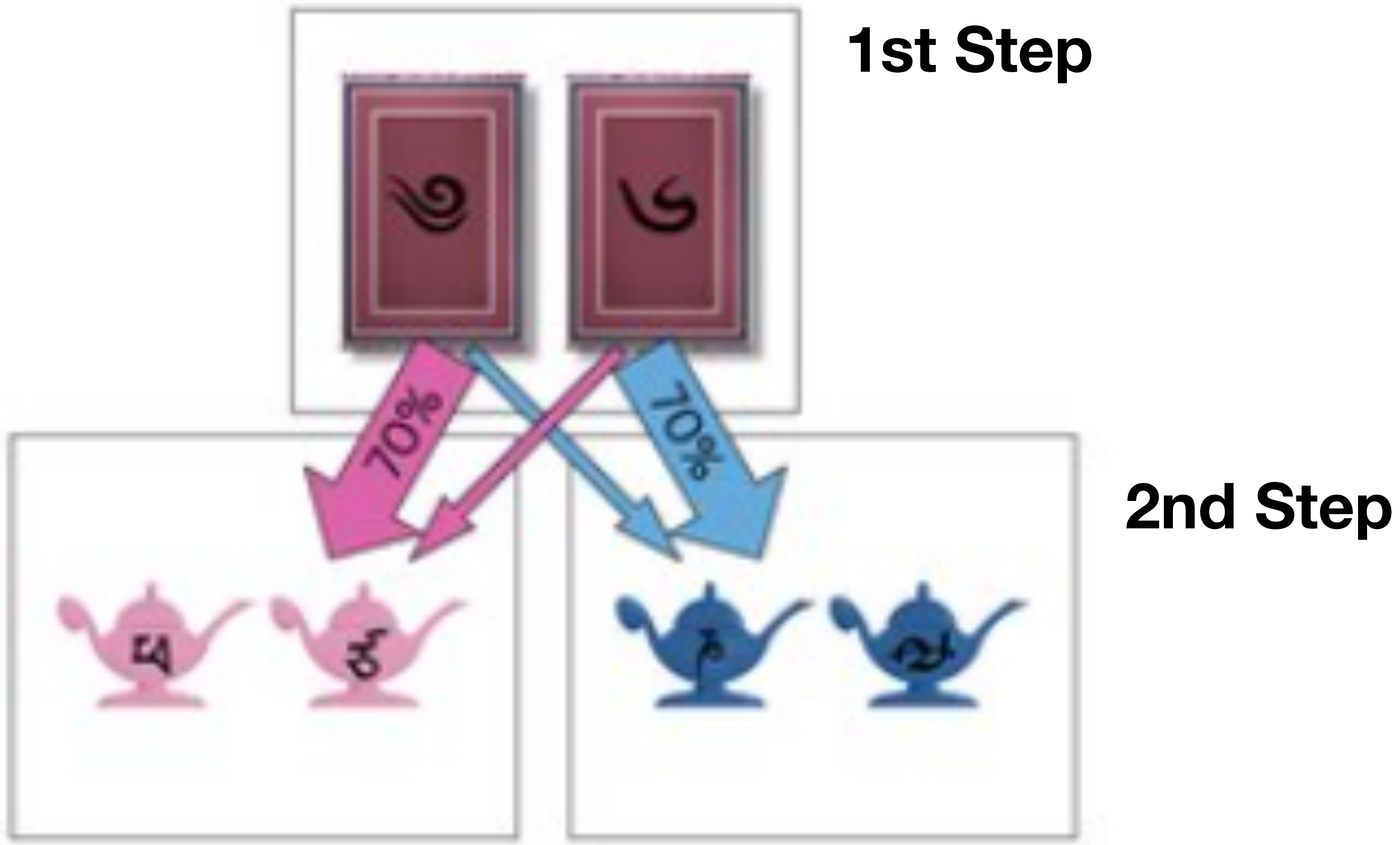


Tolman (1948)



Sensitive to *structure*, whether reward followed common vs. rare transition

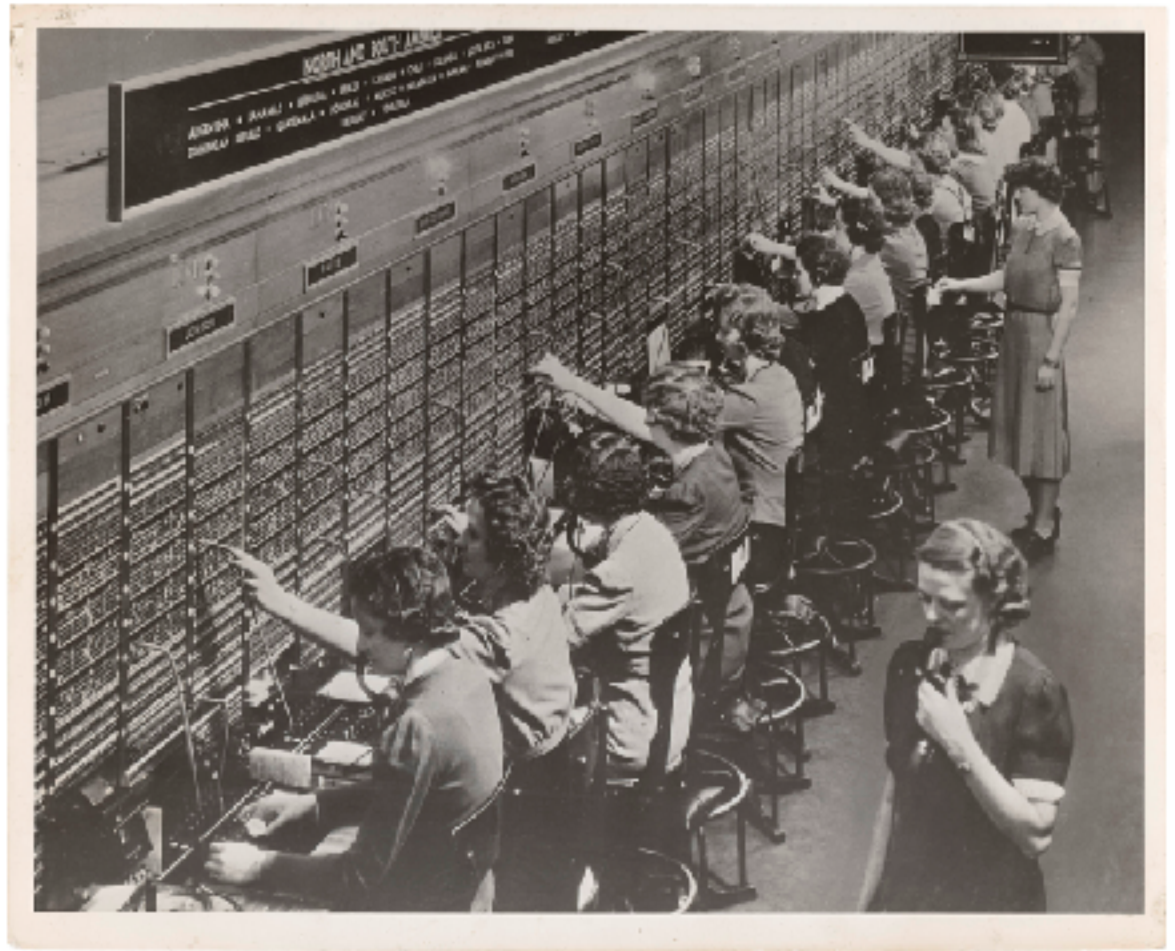
2-step task



Feher da Silva et al., (2023); Daw et al., (2011)

Model-free

S-R learning



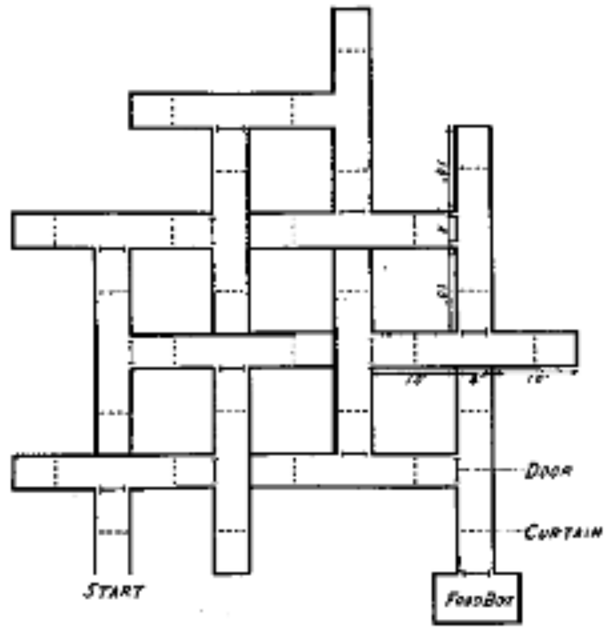
Model-based

S-S learning

r

$s \rightarrow$

$a \rightarrow$

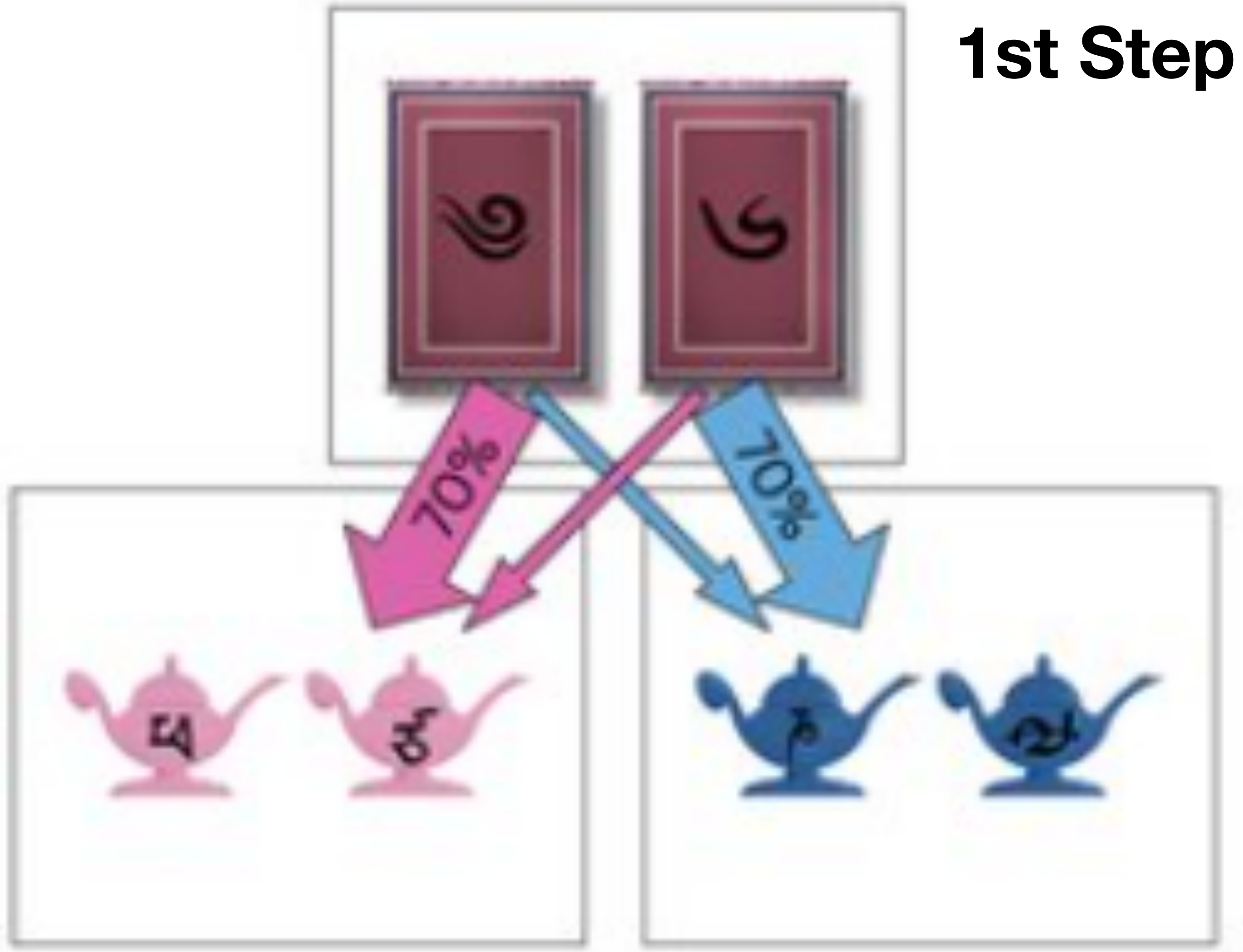


$\rightarrow s'$

Tolman (1948)

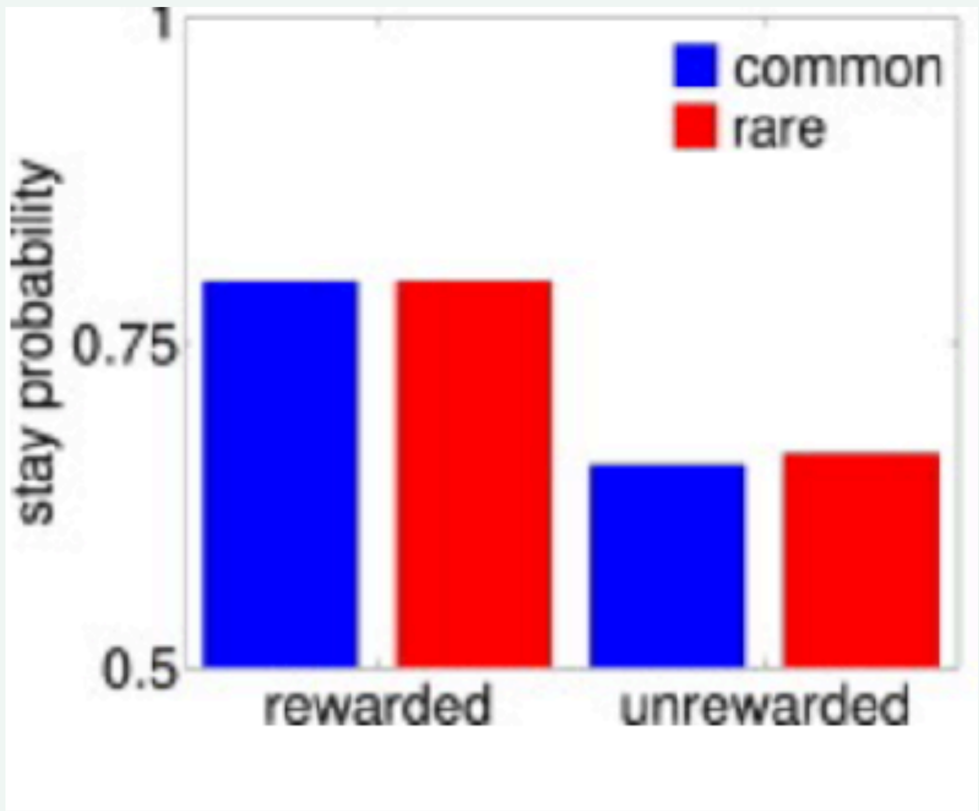
2-step task

1st Step

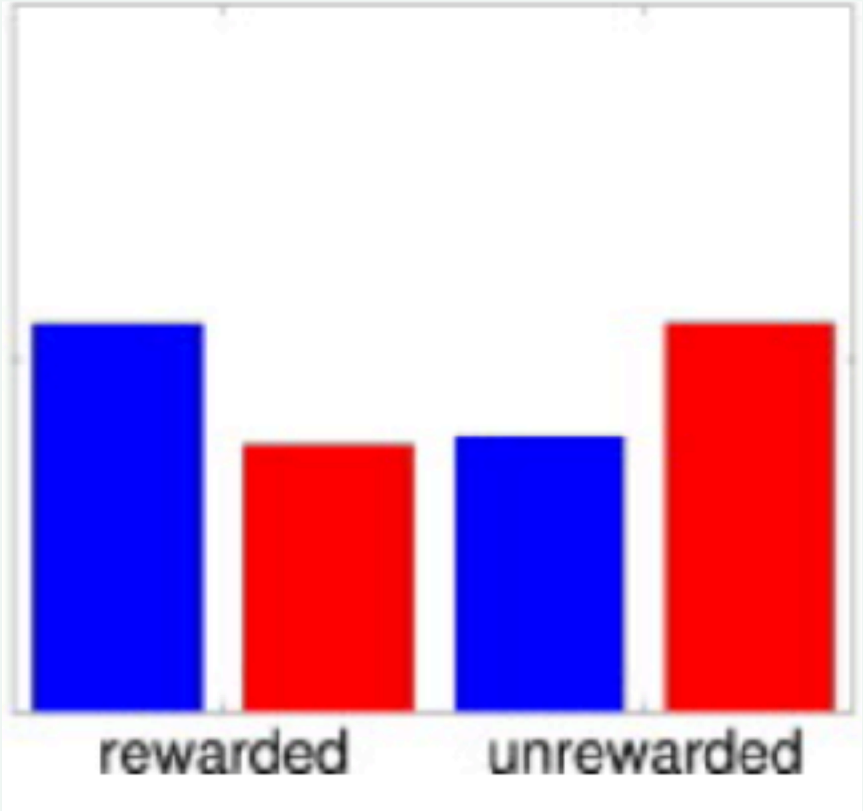


2nd Step

Feher da Silva et al., (2023); Daw et al., (2011)



Only cares whether actions were rewarded



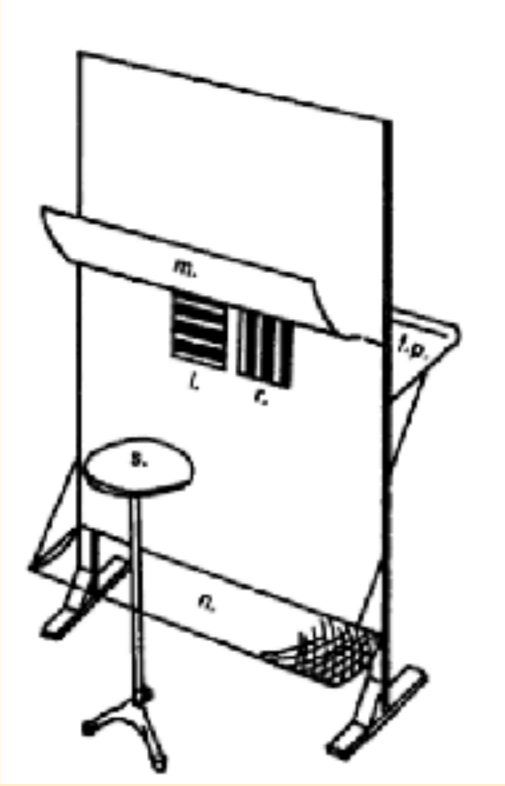
Sensitive to *structure*, whether reward followed common vs. rare transition

Vicarious trial and error (VTE): hesitating, looking-back-and-forth behavior observed in rats when confronted with a choice

VTE as active hypothesis testing

↑VTE = ↑Learning

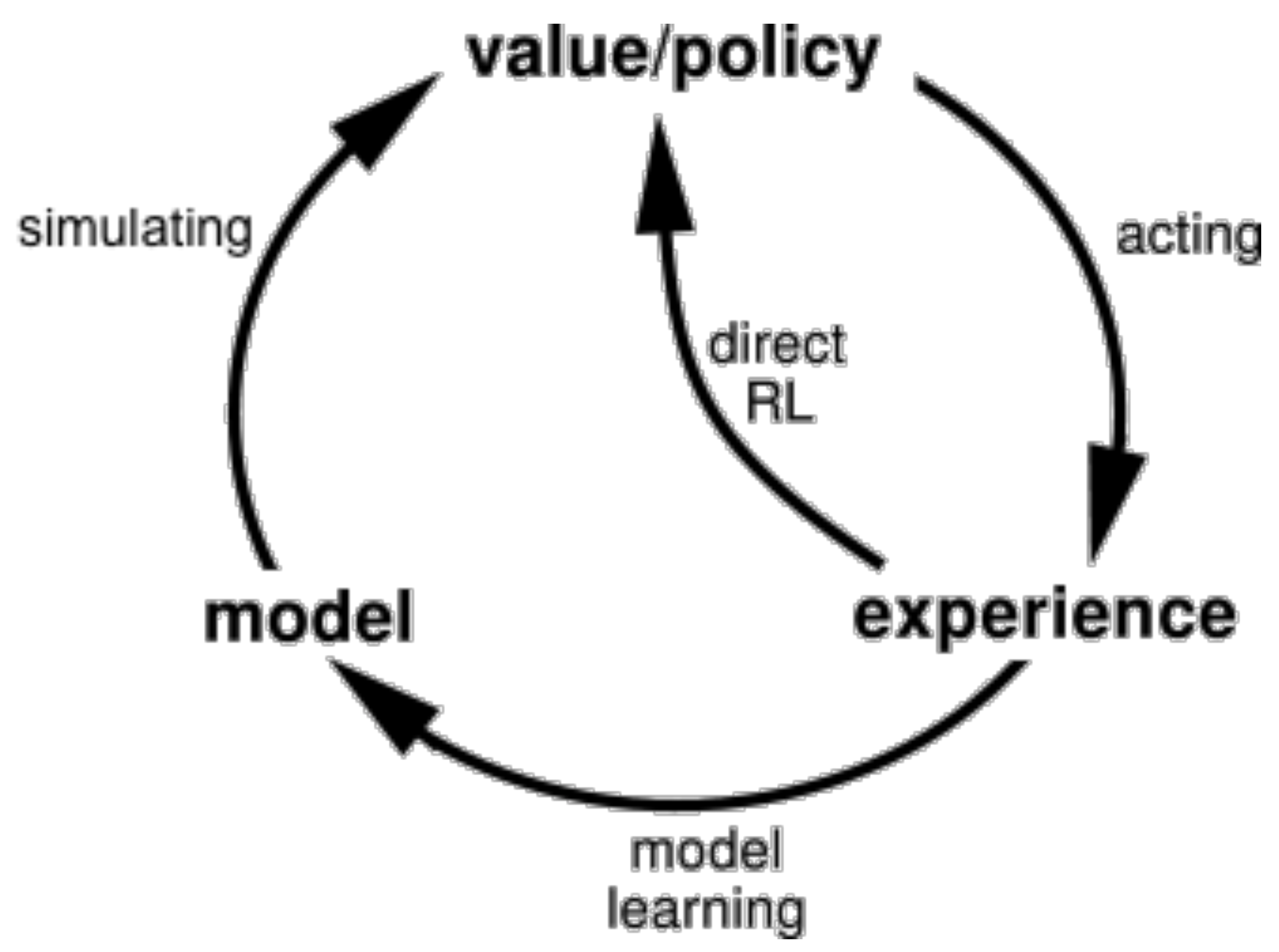
Won't come back for the exam



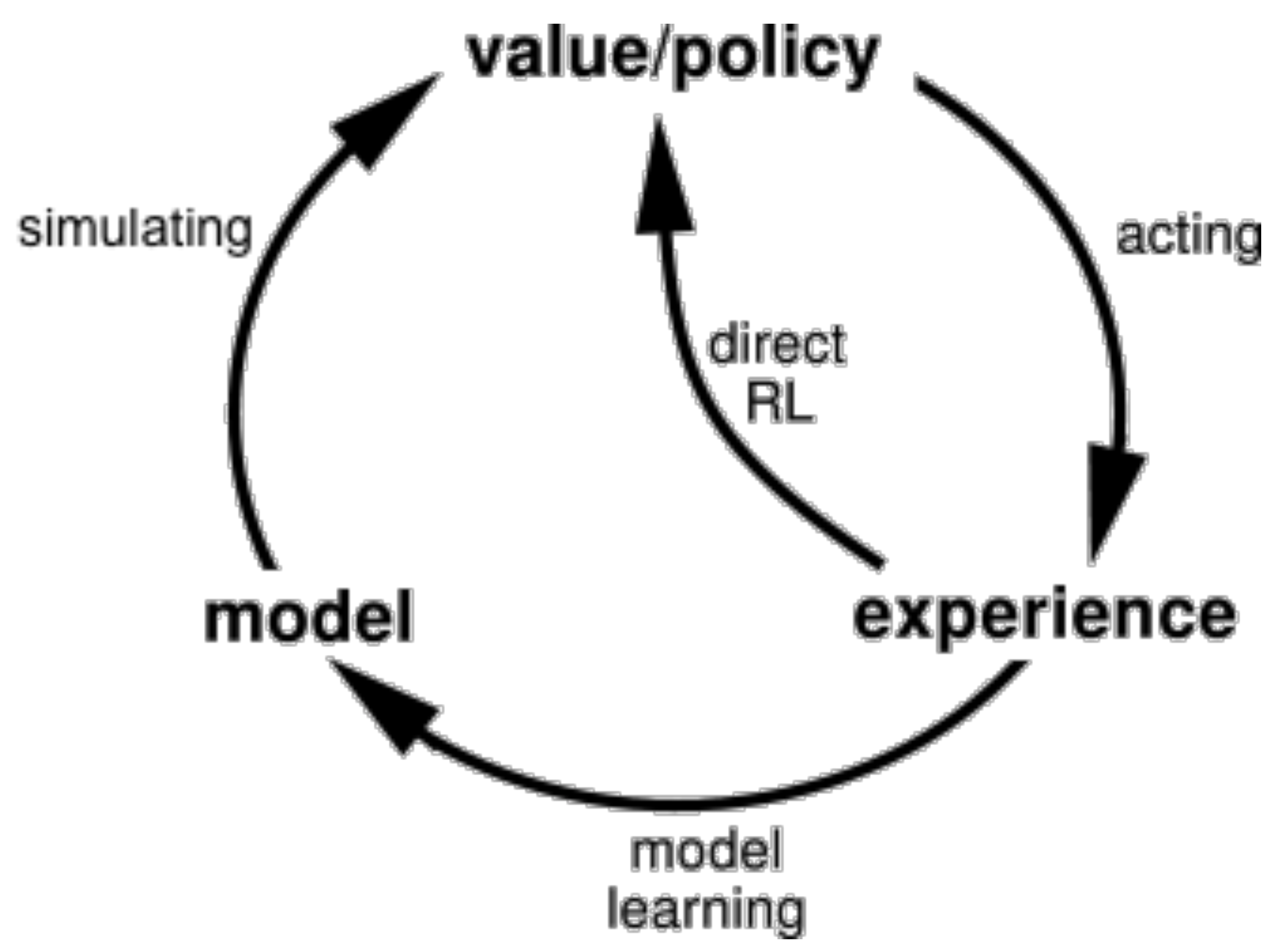
Model-based planning

The model can be used to **simulate experiences** for updating the value/policy

These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**



Model-based planning



The model can be used to **simulate experiences** for updating the value/policy

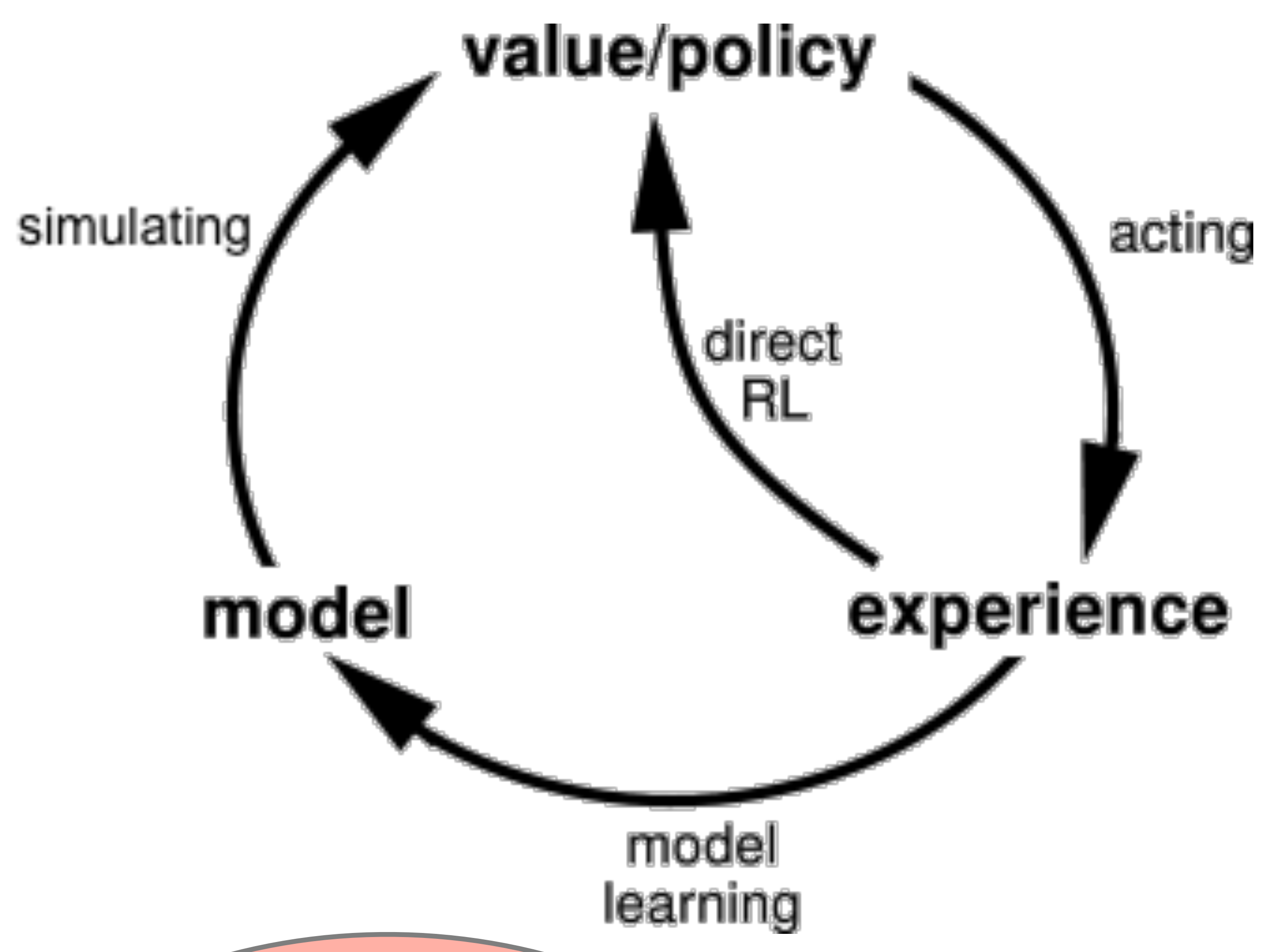
These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**

- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor-Critic**

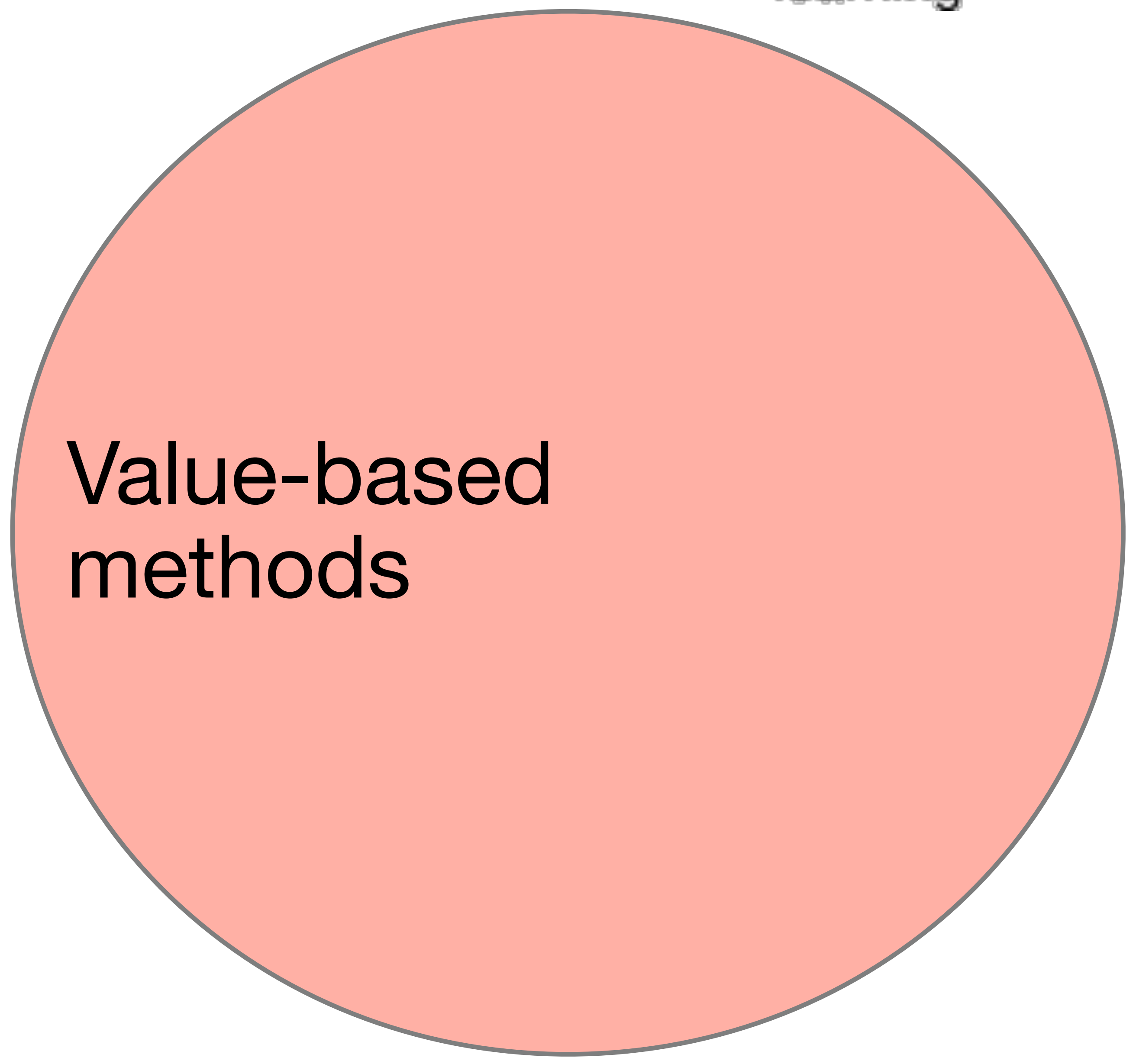
Model-based planning

The model can be used to **simulate experiences** for updating the value/policy

These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**



- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor**-**Critic**

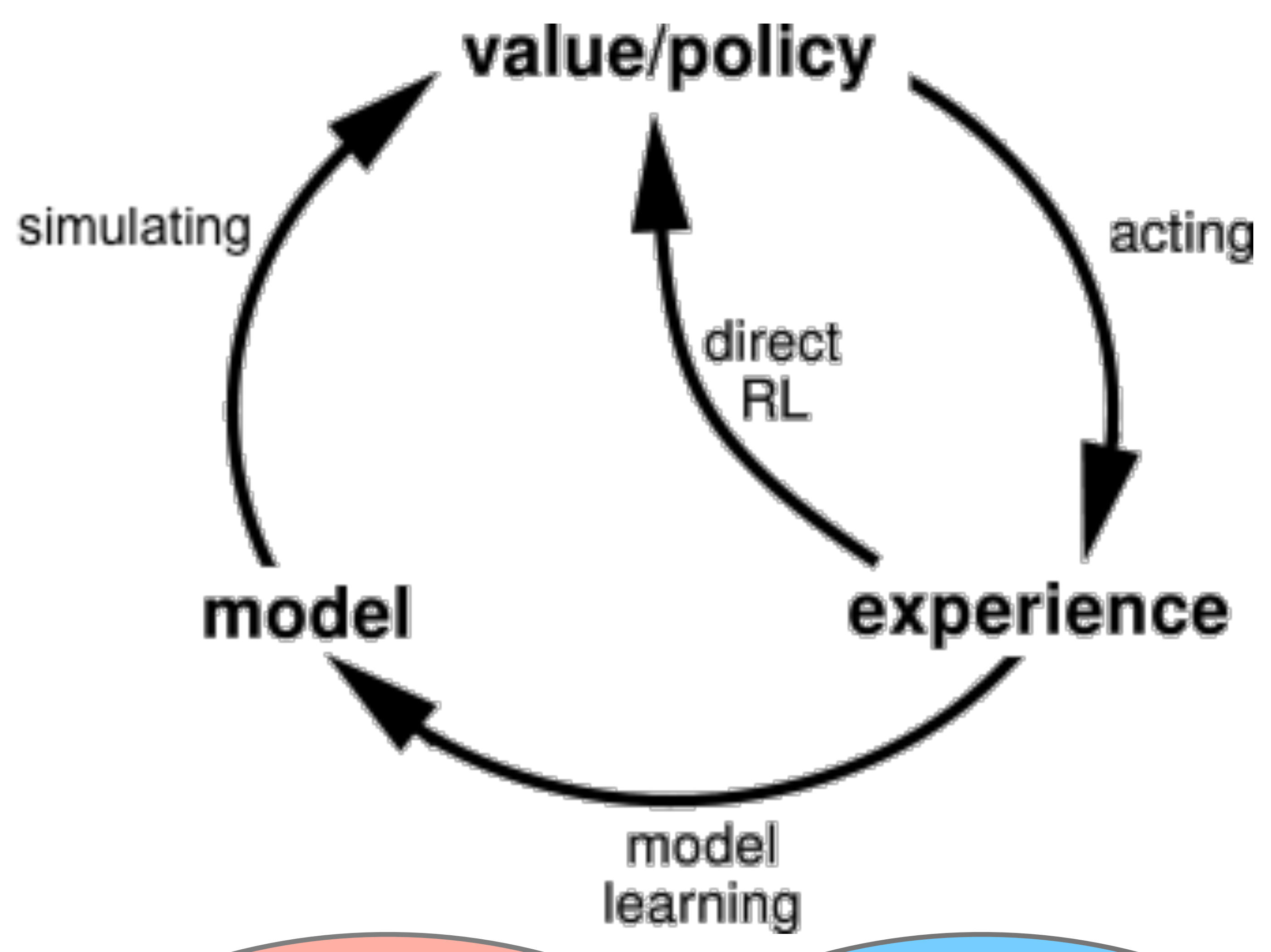


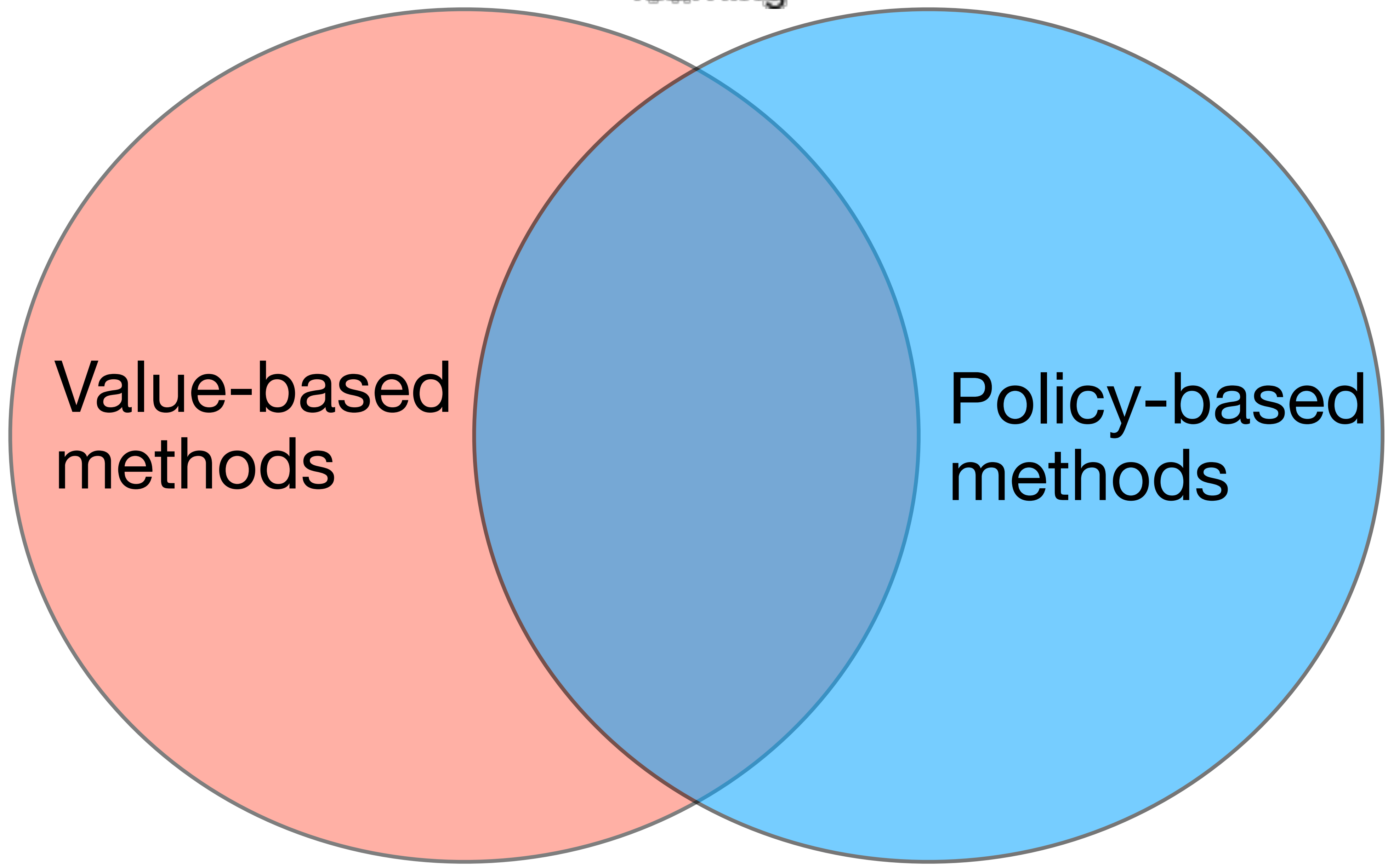
Model-based planning

The model can be used to **simulate experiences** for updating the value/policy

These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**

- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor-Critic**

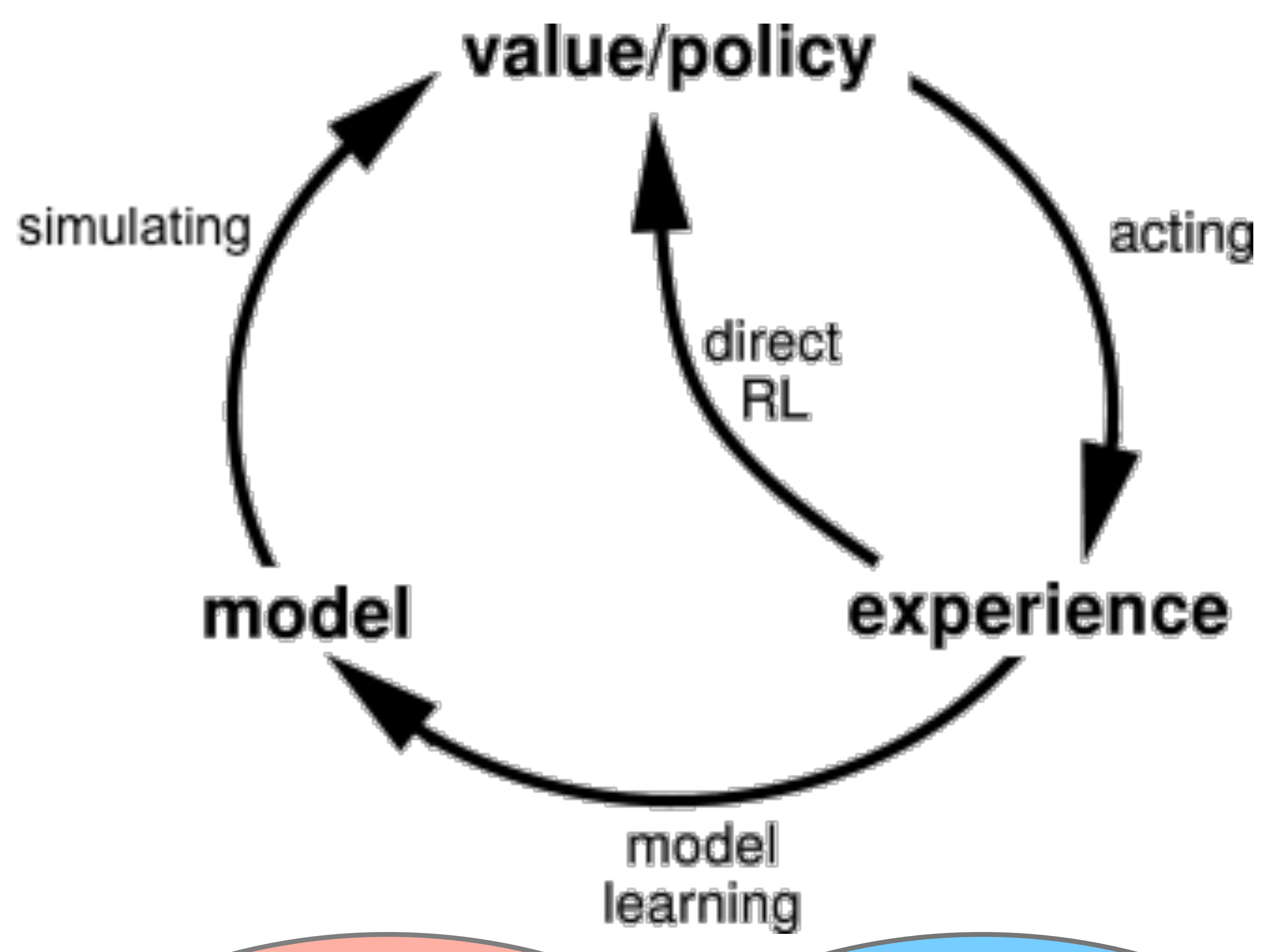




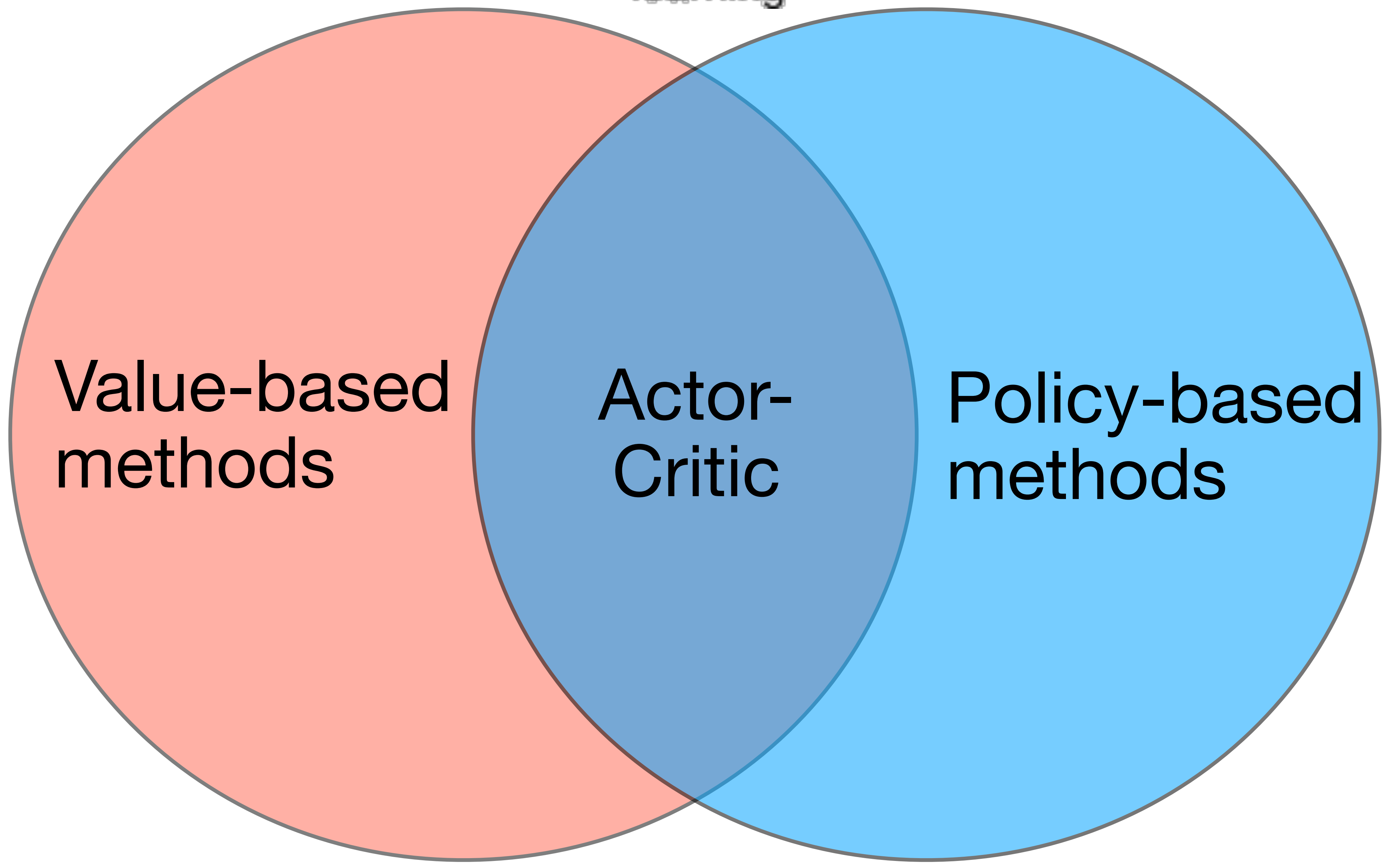
Model-based planning

The model can be used to **simulate experiences** for updating the value/ policy

These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**



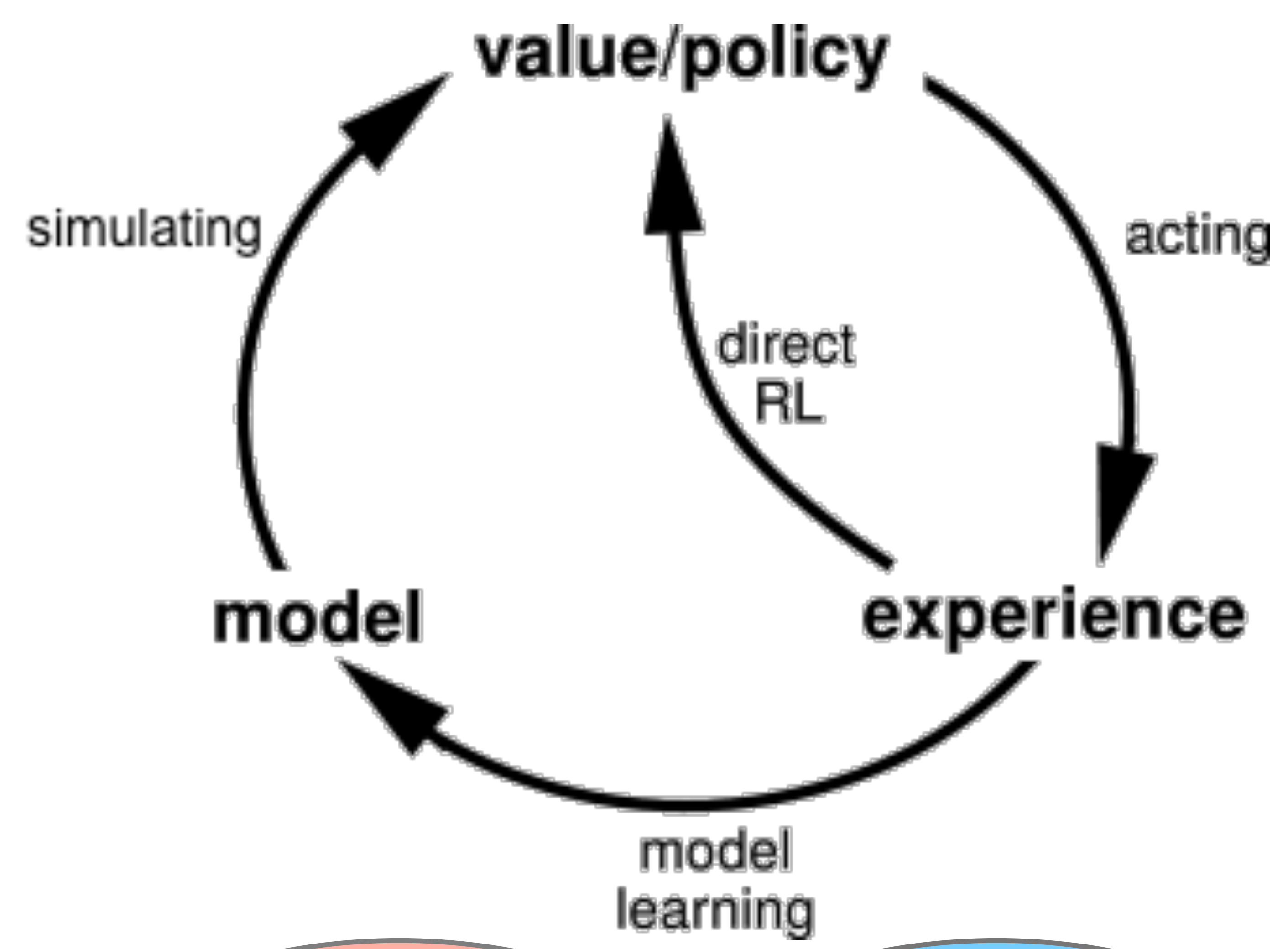
- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor-Critic**



Model-based planning

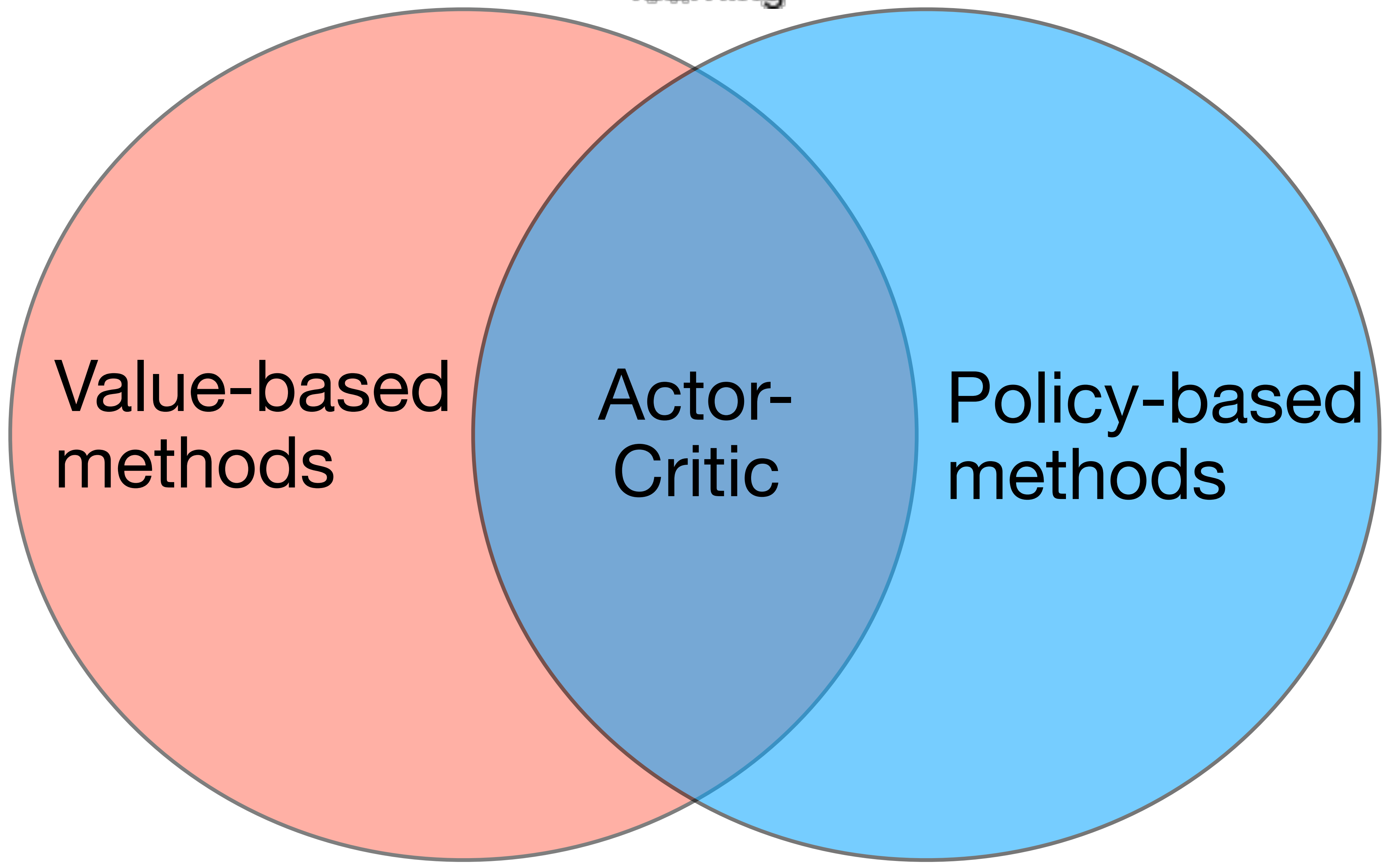
The model can be used to **simulate experiences** for updating the value/ policy

These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**



- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor-Critic**

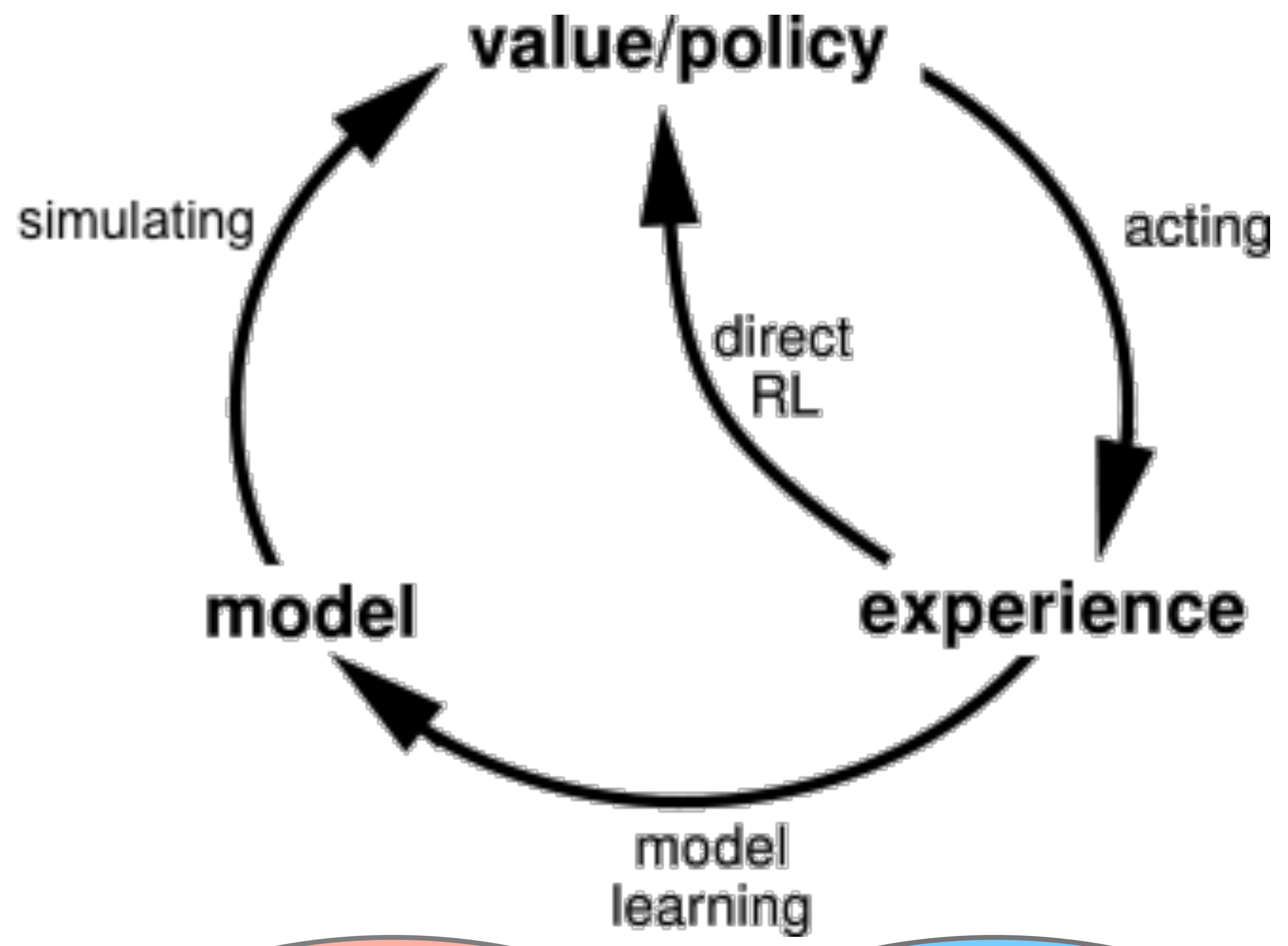
    Deep Q-learning



Model-based planning

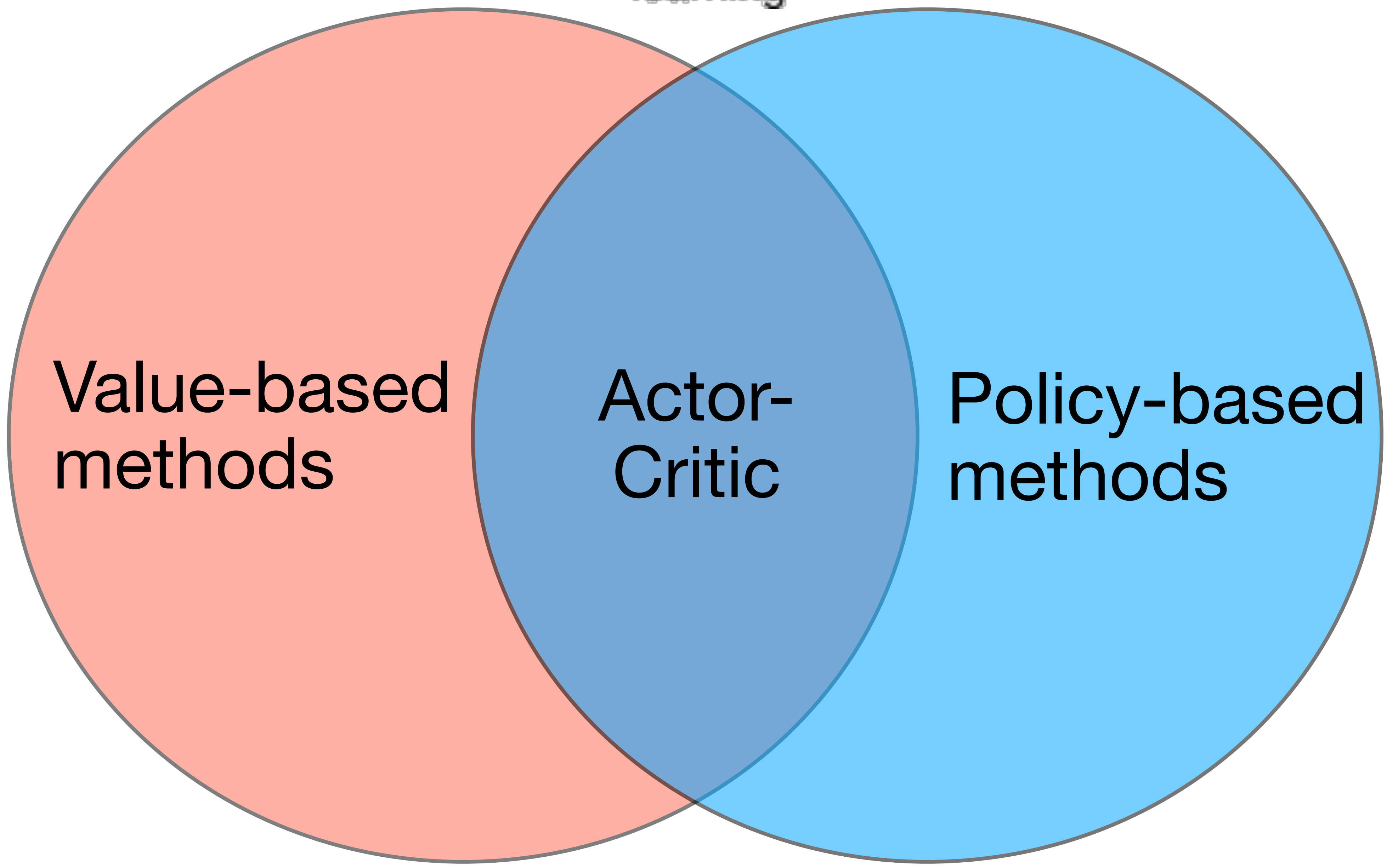
The model can be used to **simulate experiences** for updating the value/policy

These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**



- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor-Critic**

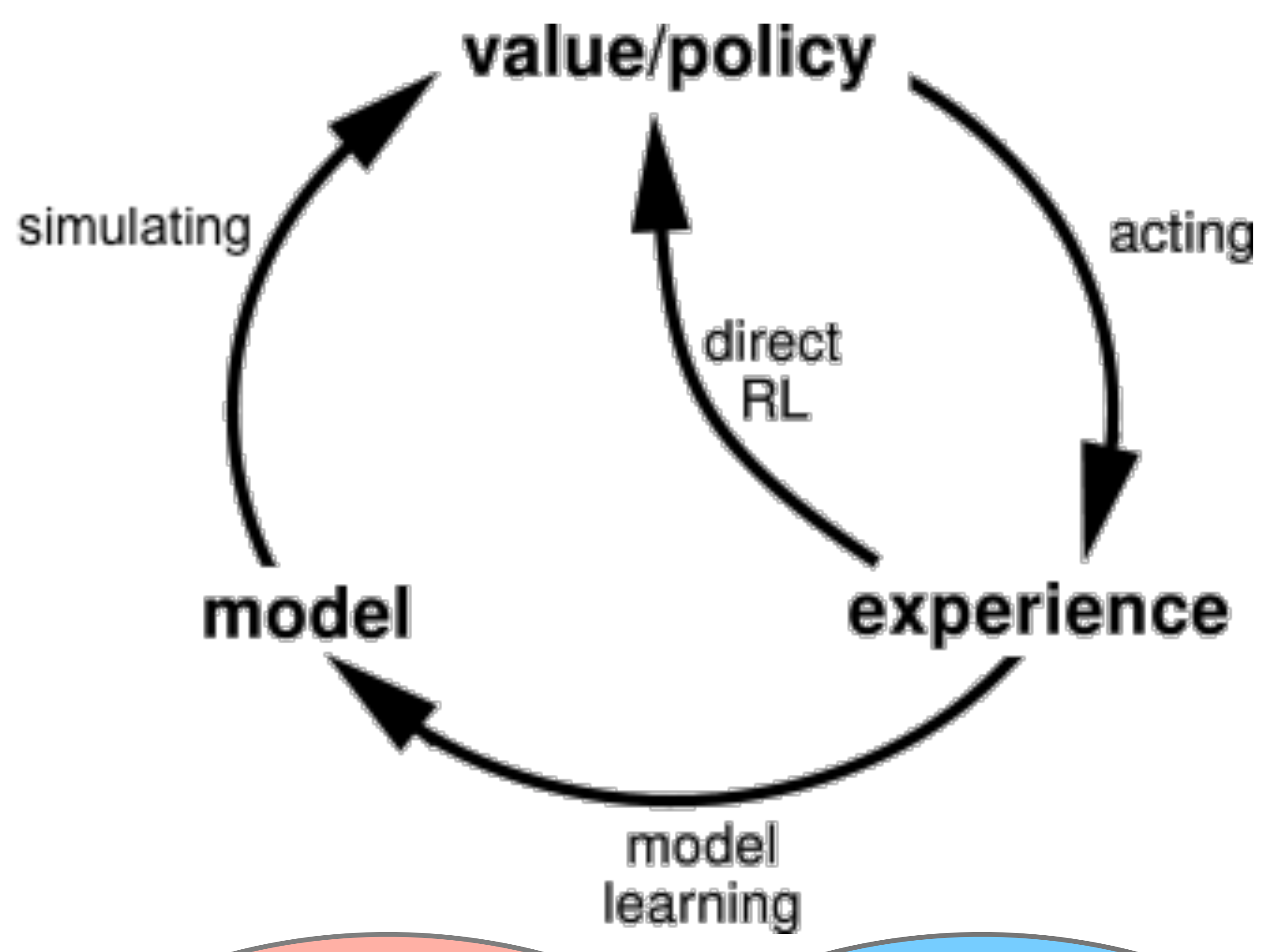
  Deep Q-learning Policy gradient



Model-based planning

The model can be used to **simulate experiences** for updating the value/ policy

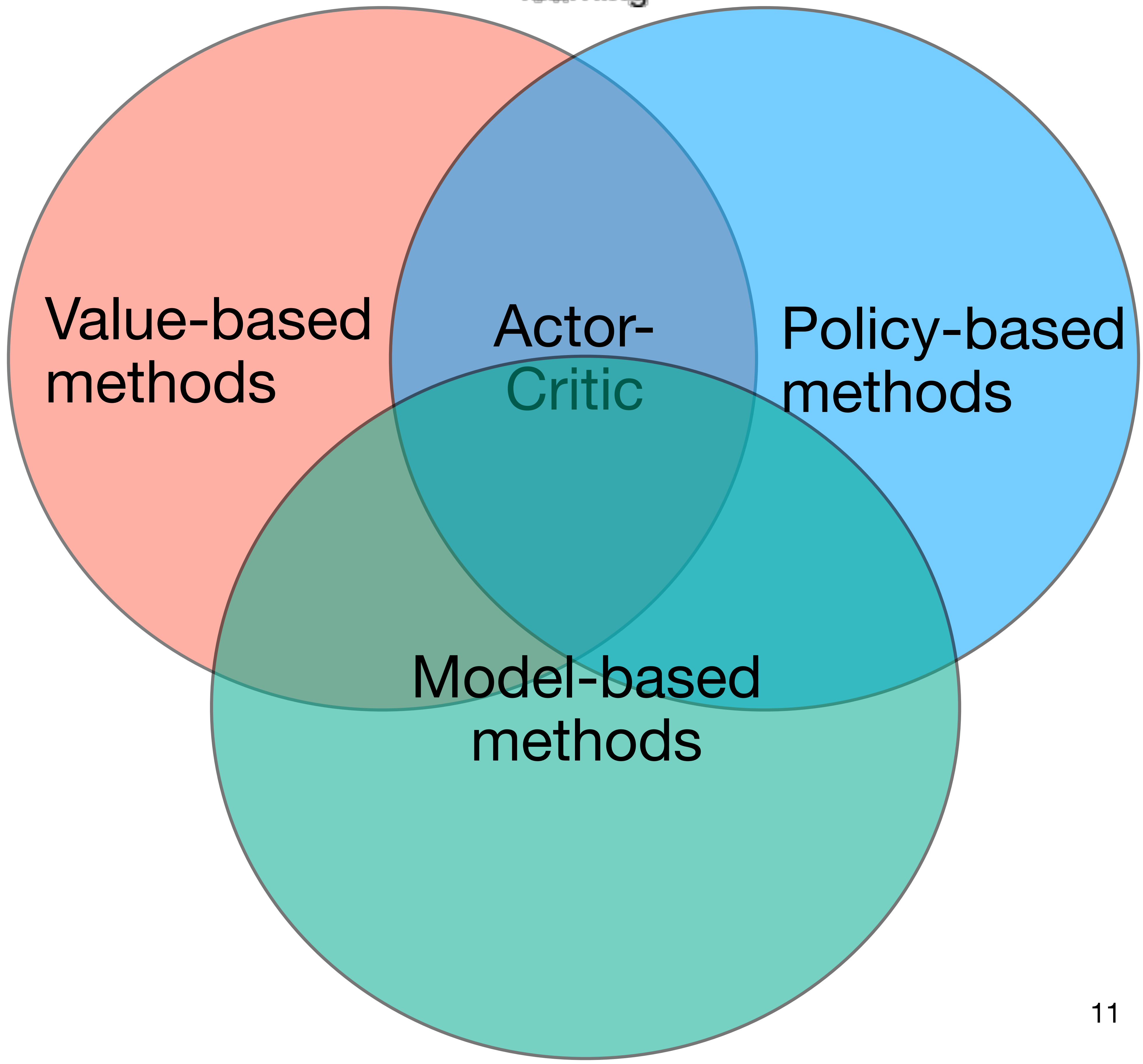
These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**



- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor-Critic**

  Deep Q-learning Policy gradient

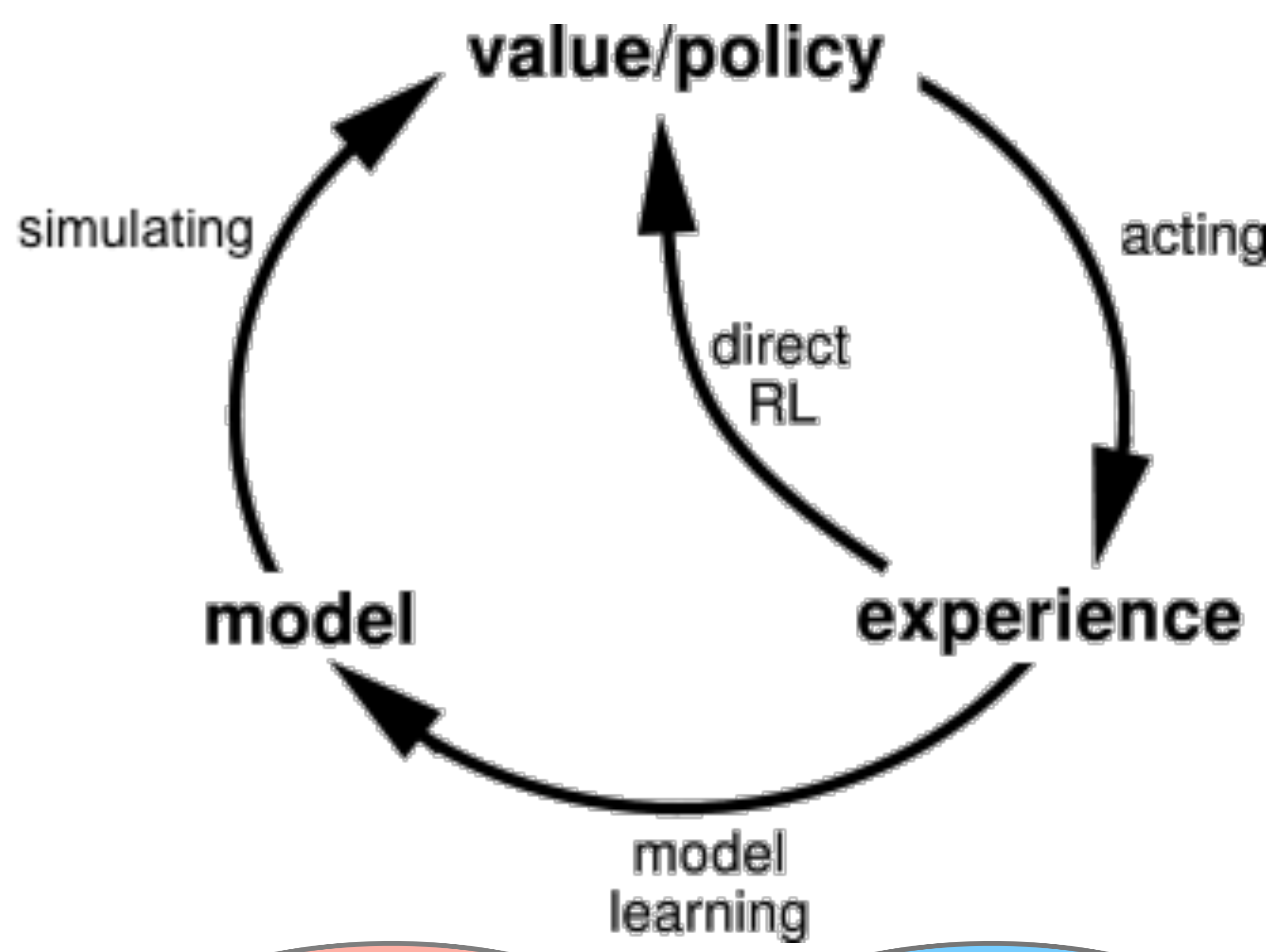
- Model-based methods can as well…



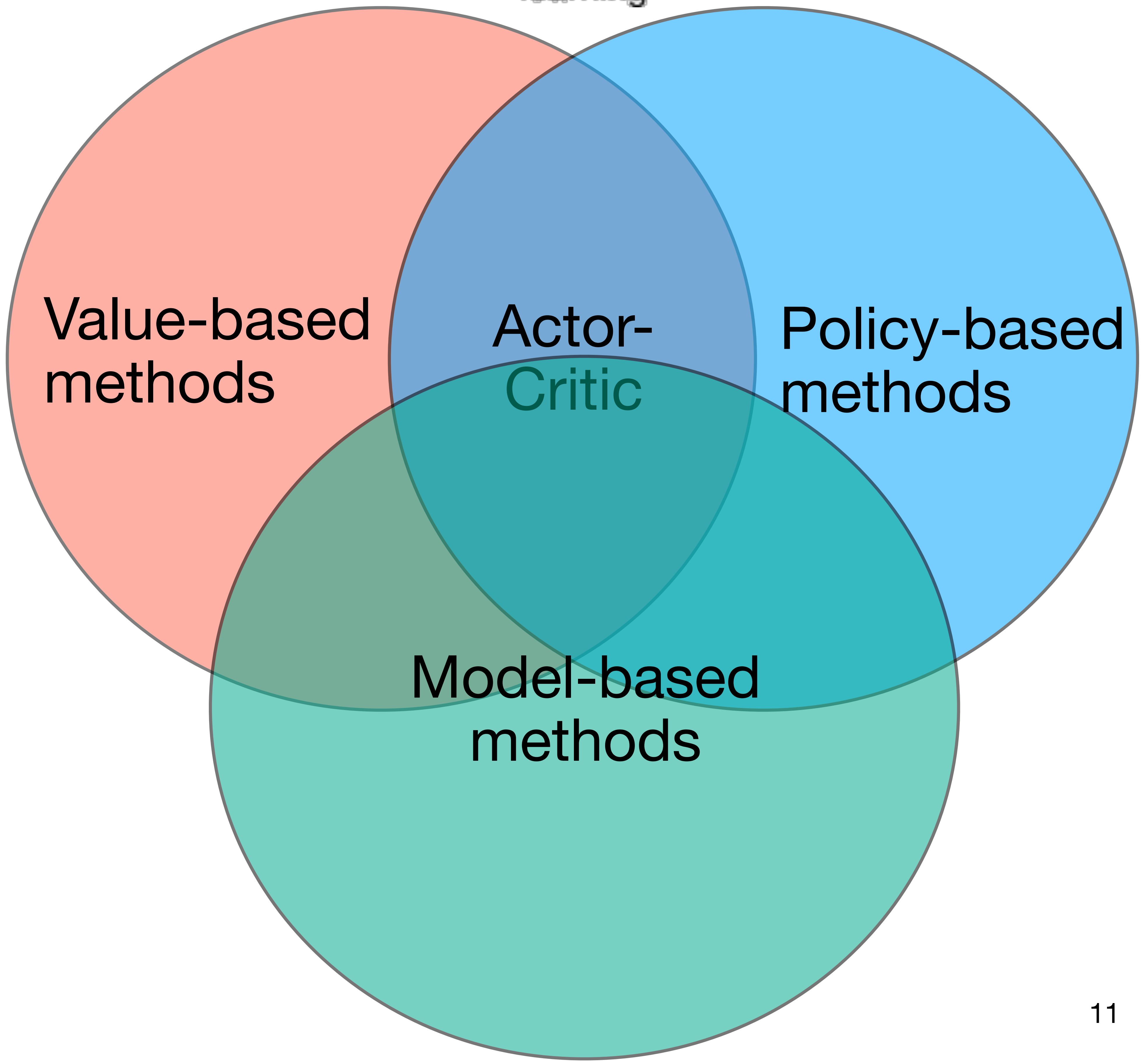
Model-based planning

The model can be used to **simulate experiences** for updating the value/policy

These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**



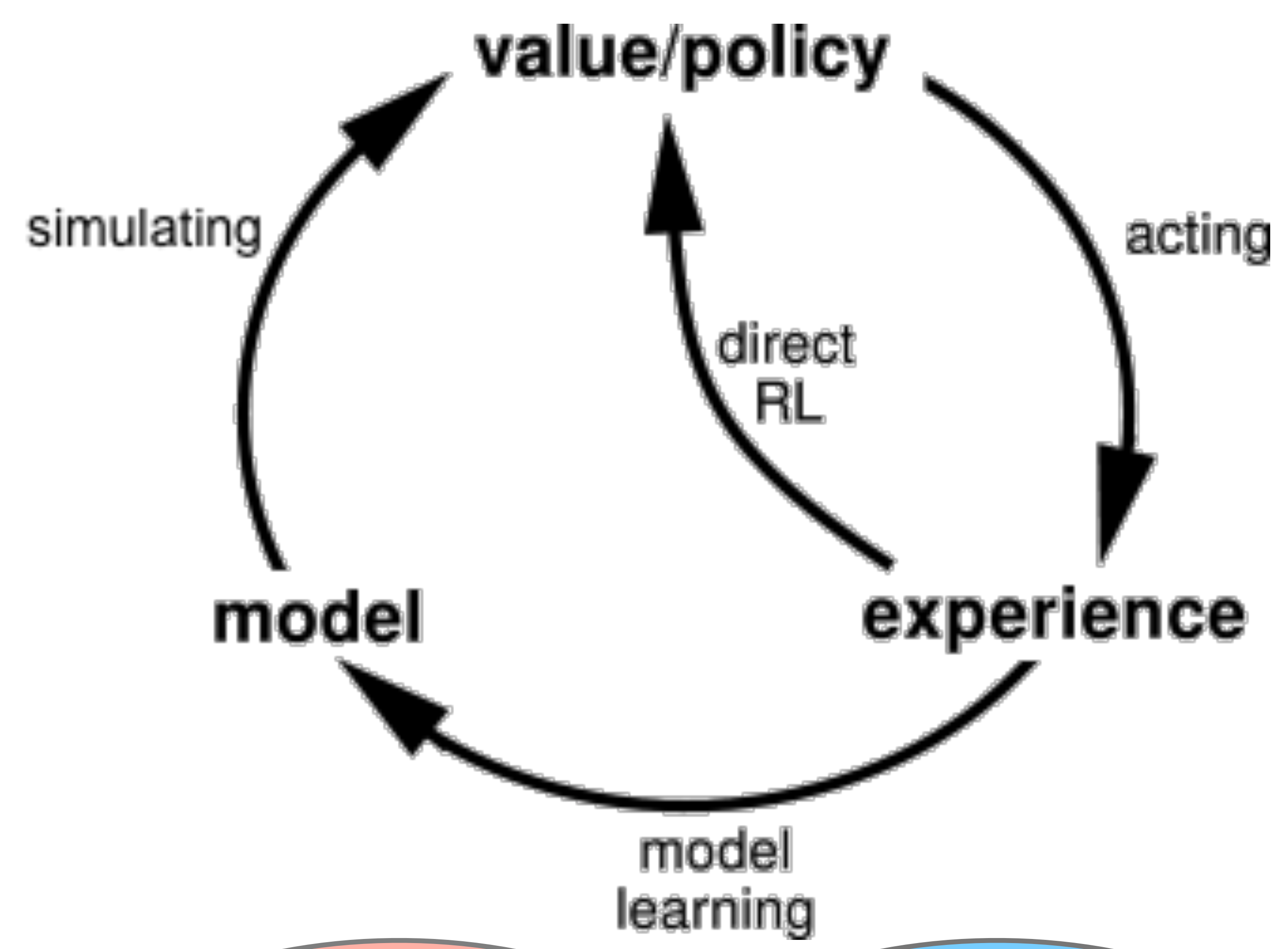
- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor-Critic**
  Deep Q-learning   Policy gradient
- Model-based methods can as well…
  - DYNA (**Model** & **Value**)



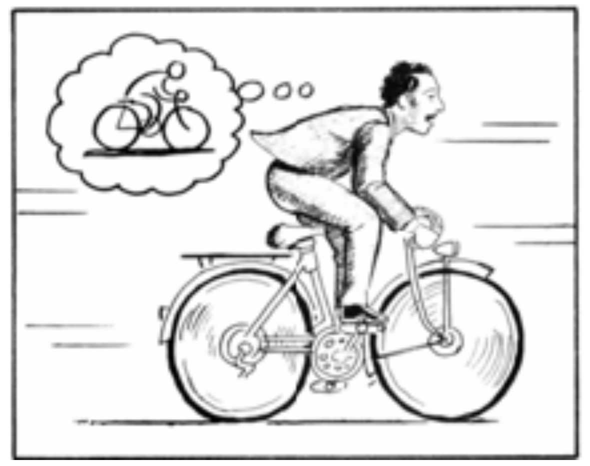
Model-based planning

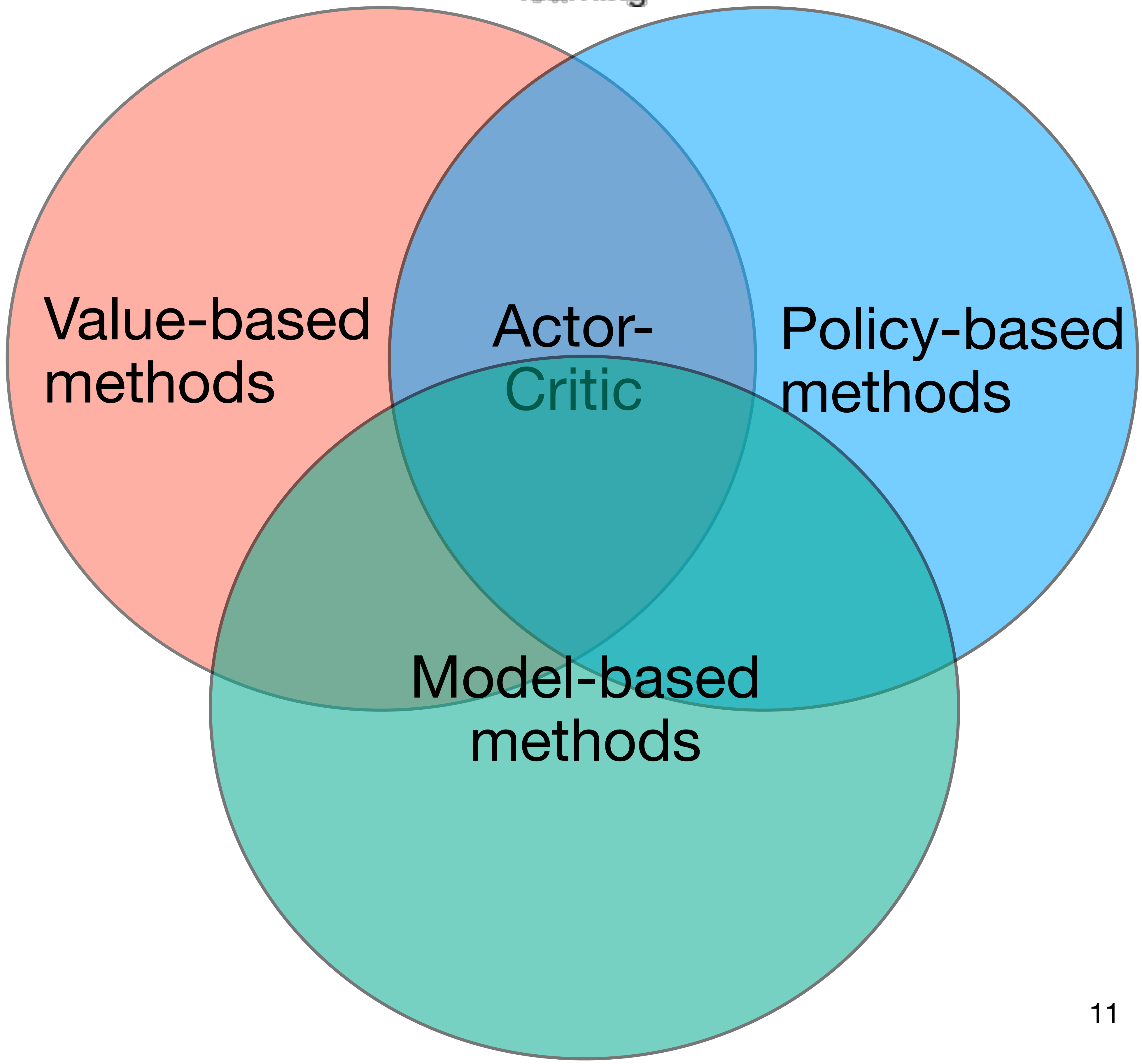
The model can be used to **simulate experiences** for updating the value/ policy

These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**



- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor-Critic**
  Deep Q-learning Policy gradient
- Model-based methods can as well…
  - DYNA (**Model** & **Value**)
  - World Models (**Model** & **Policy**)

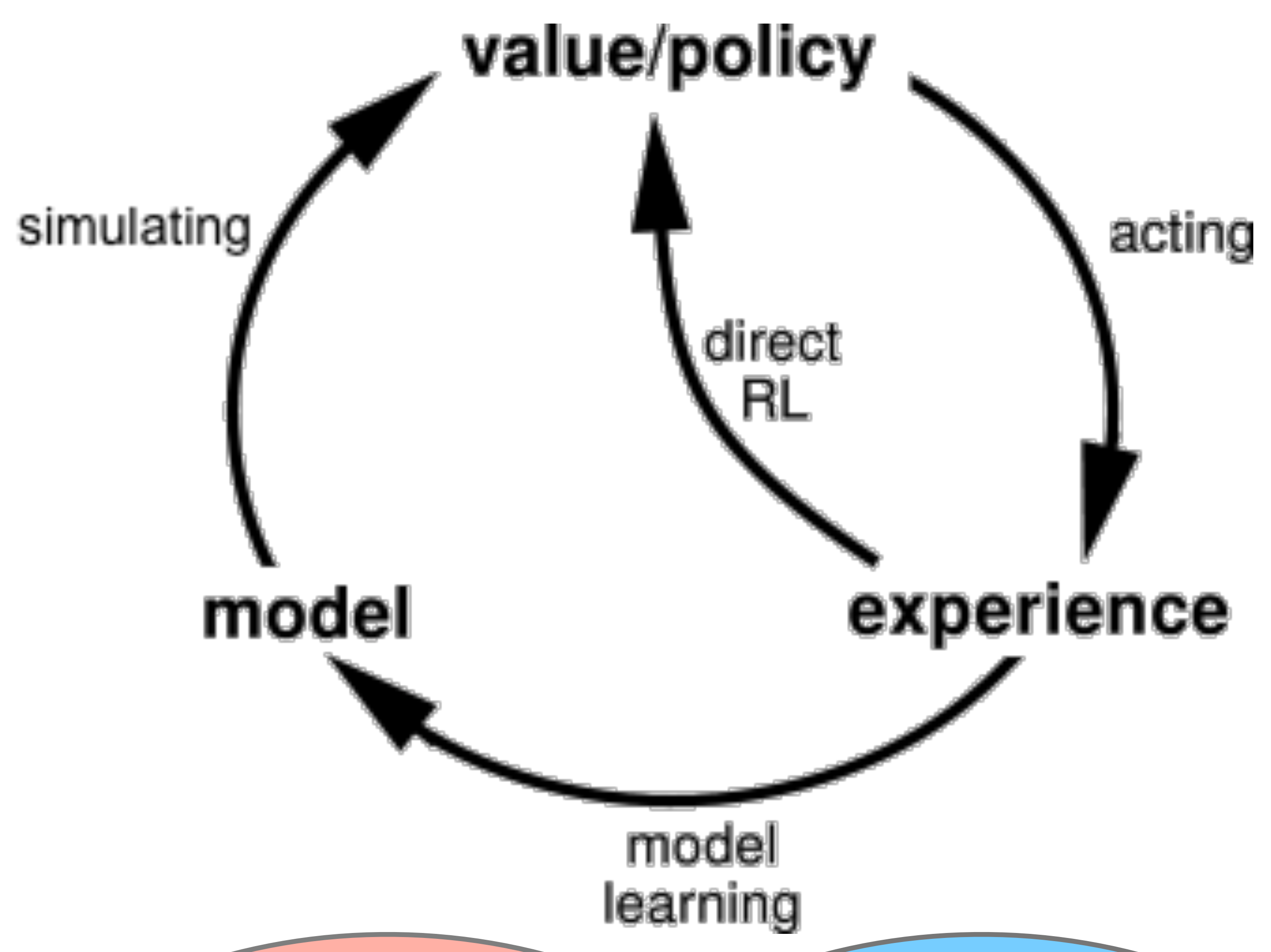




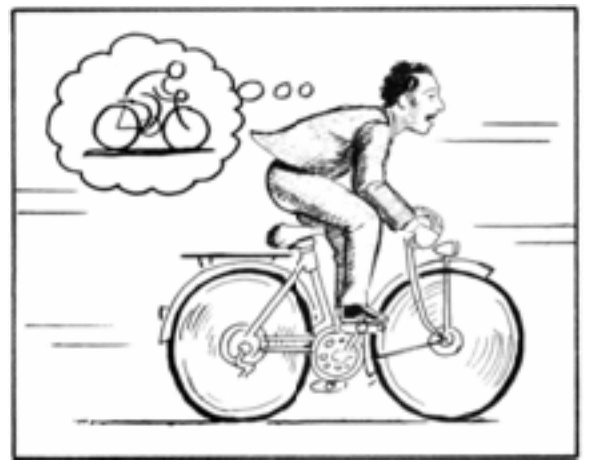
Model-based planning

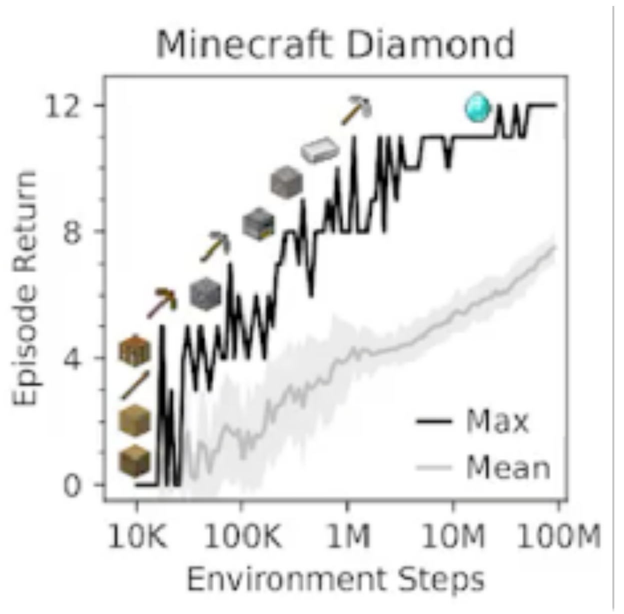
The model can be used to **simulate experiences** for updating the value/ policy

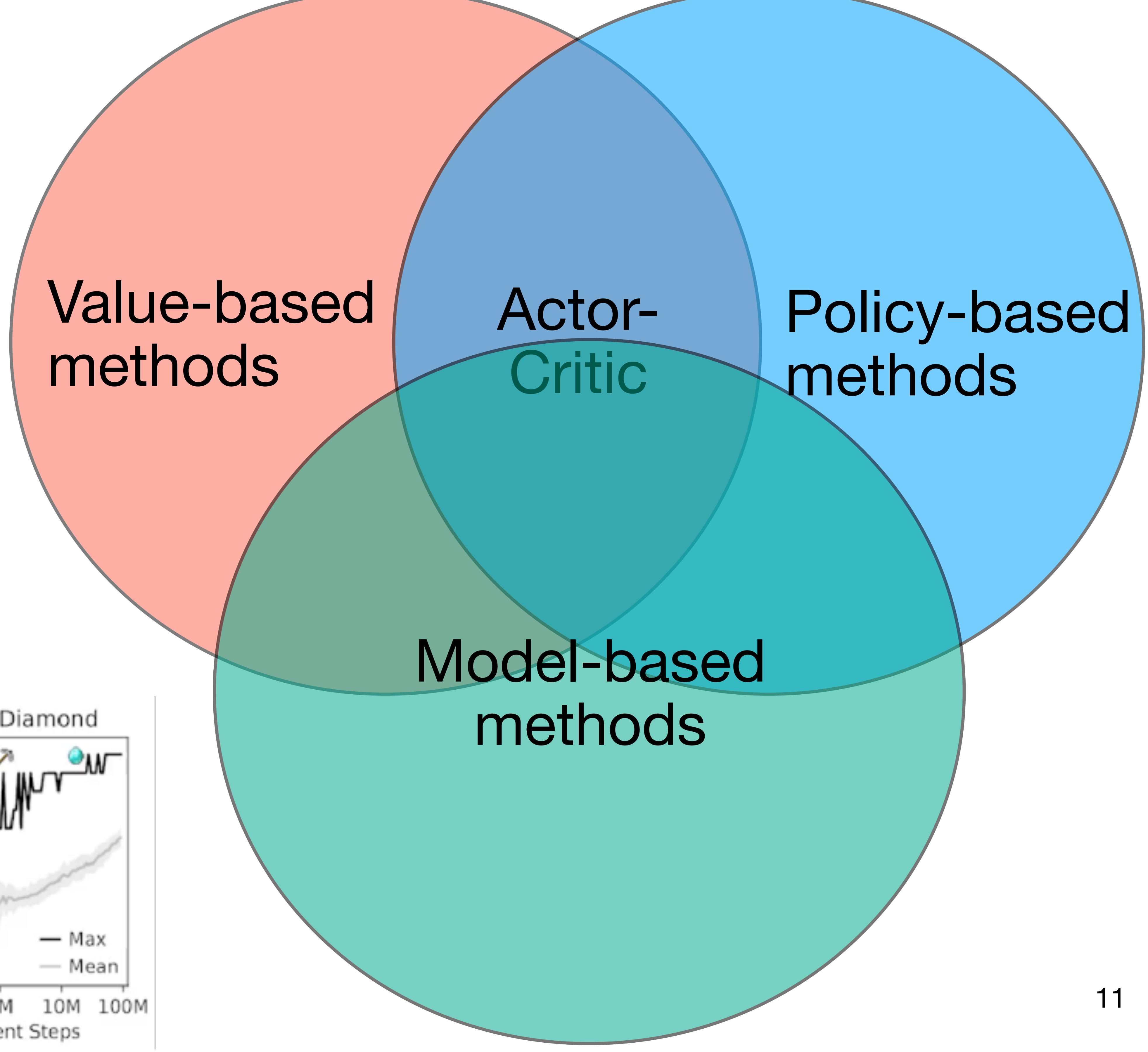
These simulations are **computationally costly**, but supplement direct RL, leading to **faster learning** and **greater flexibility**



- Recall that model-free methods can be categorized as **Value-based**, **Policy-based**, or **Actor-Critic**
  Deep Q-learning Policy gradient
- Model-based methods can as well…
  - DYNA (**Model** & **Value**)
  - World Models (**Model** & **Policy**)
  - Dreamer (**Model** & **Actor-Critic**)

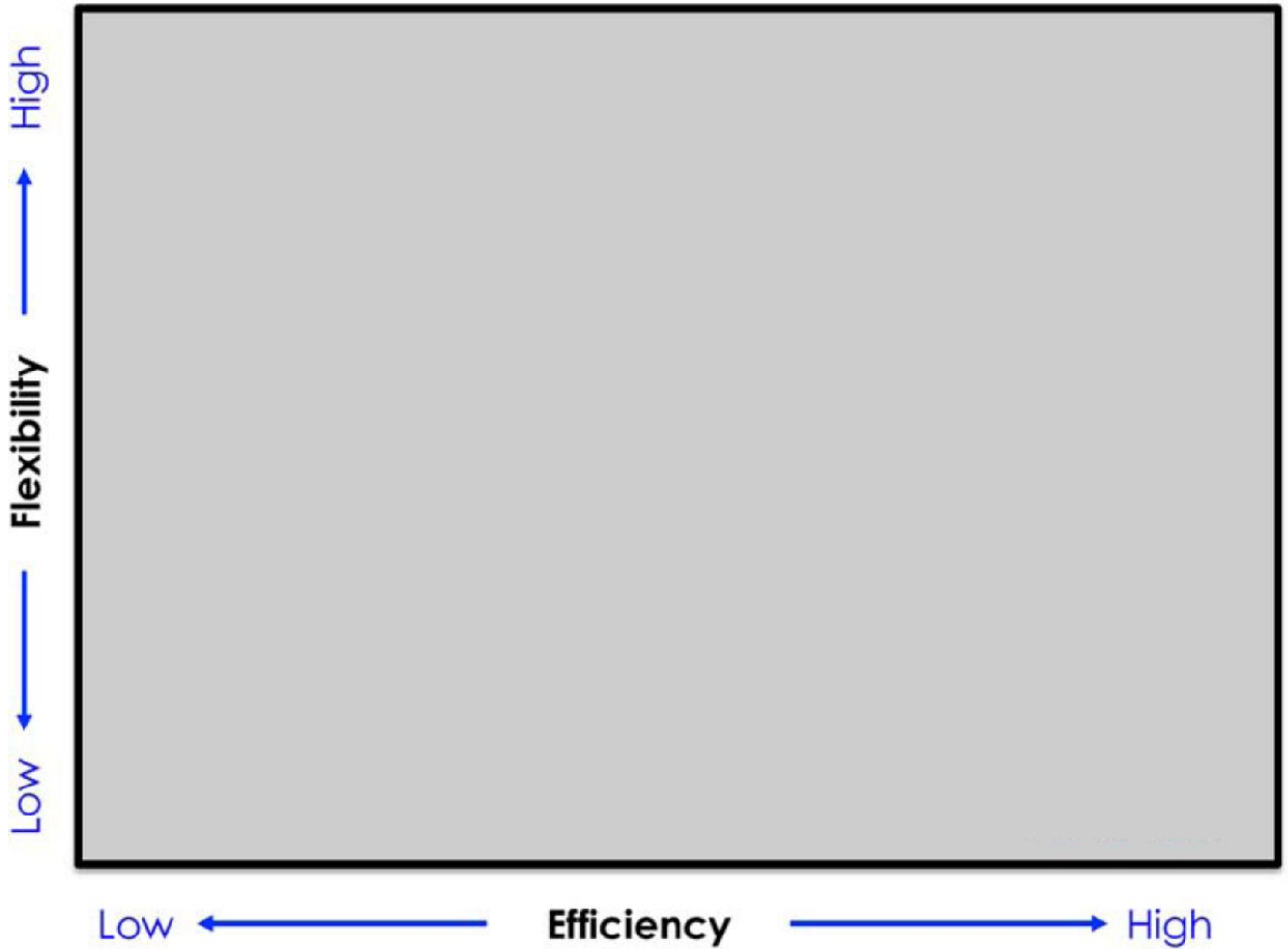






Balancing flexibility and efficiency

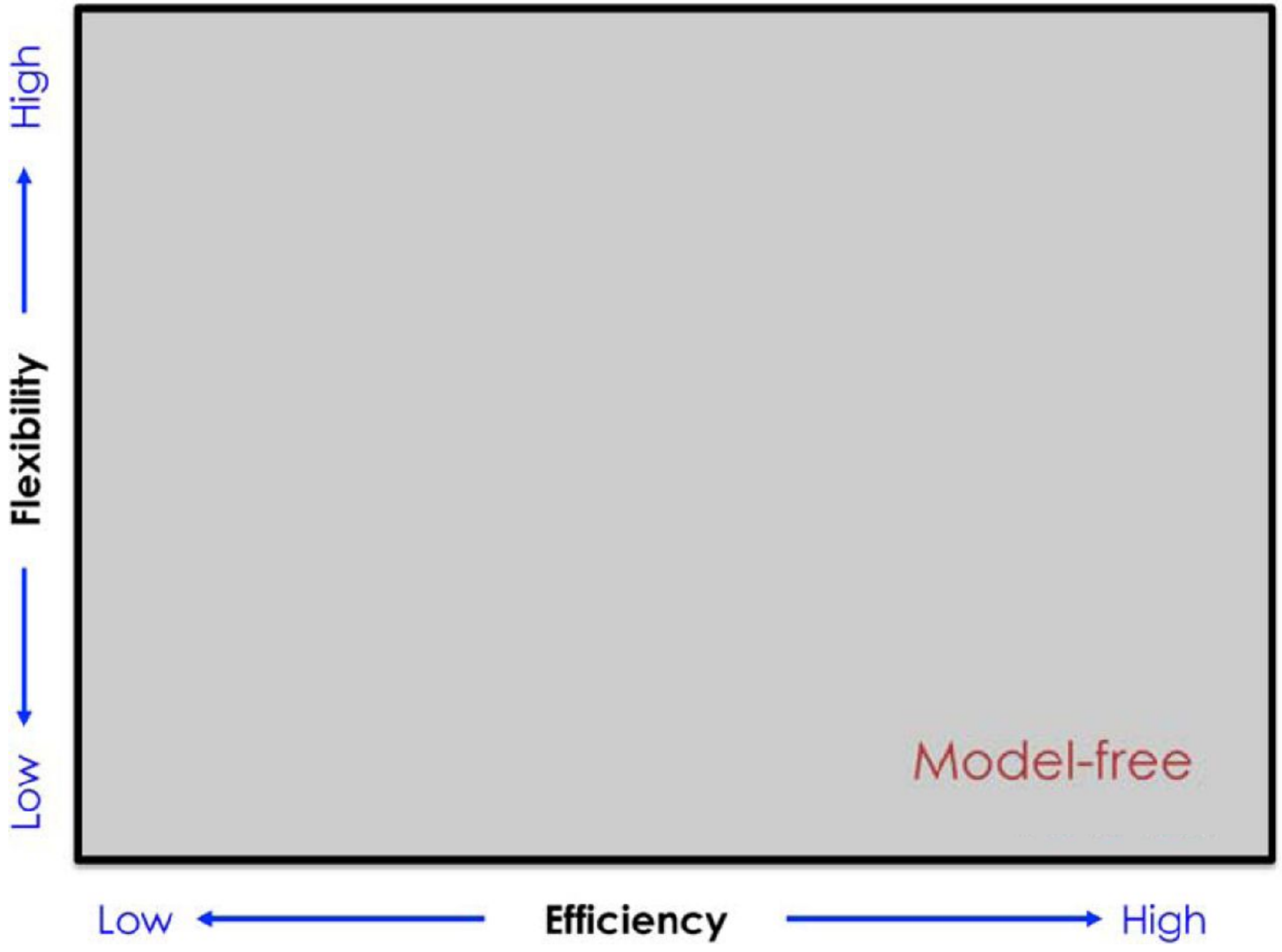
- Model-free methods
  - **Computationally efficient**
  - But **lack flexibility** to changes in the environment: *value is coupled to policy*
- Model-based methods
  - **Highly flexible:** *value and policy can be quickly recomputed* via simulations
  - But performing simulations are **computationally costly**
- SR falls in between



Gershman (2018)

Balancing flexibility and efficiency

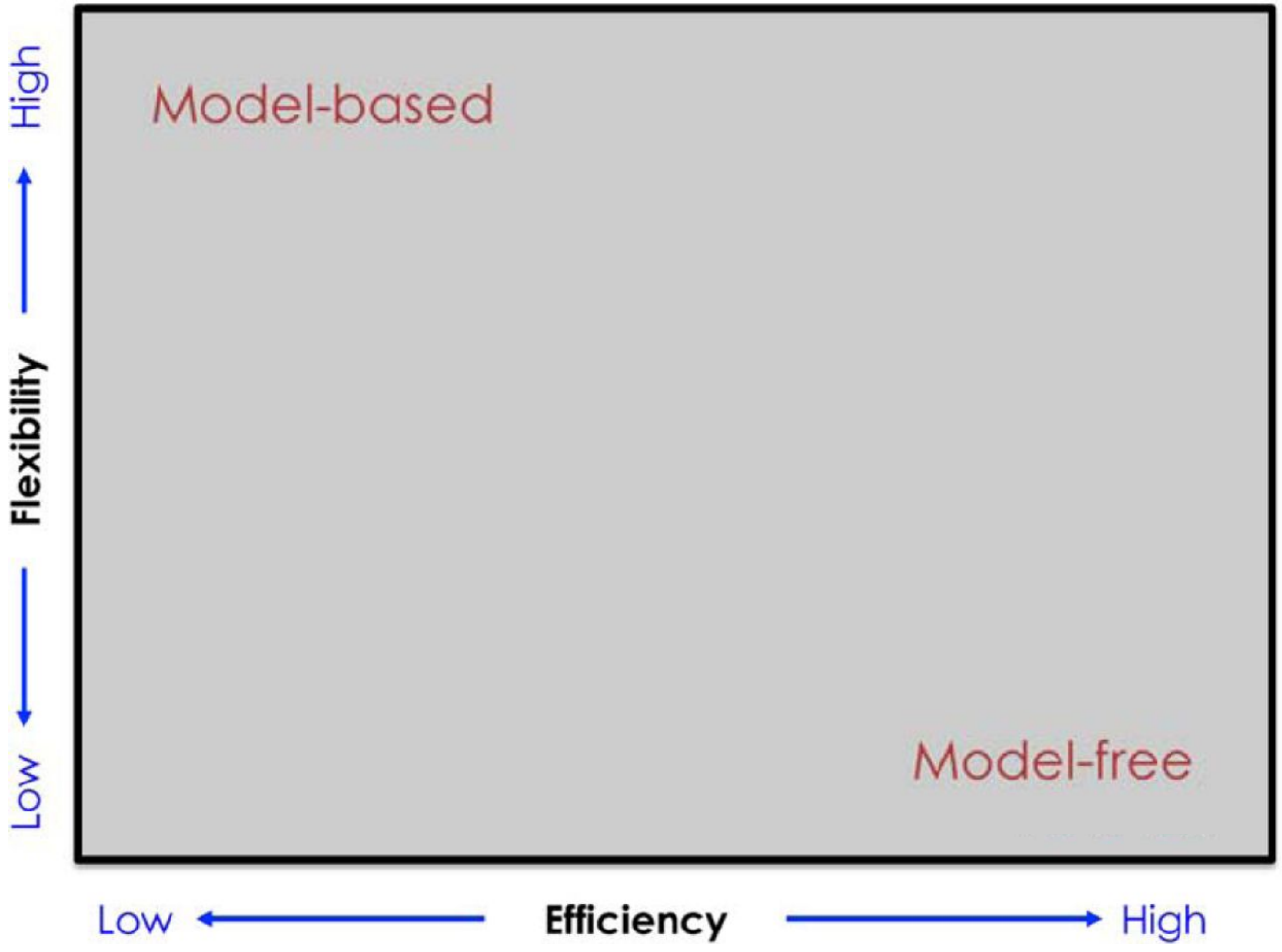
- Model-free methods
  - **Computationally efficient**
  - But **lack flexibility** to changes in the environment: *value is coupled to policy*
- Model-based methods
  - **Highly flexible:** *value and policy can be quickly recomputed* via simulations
  - But performing simulations are **computationally costly**
- SR falls in between



Gershman (2018)

Balancing flexibility and efficiency

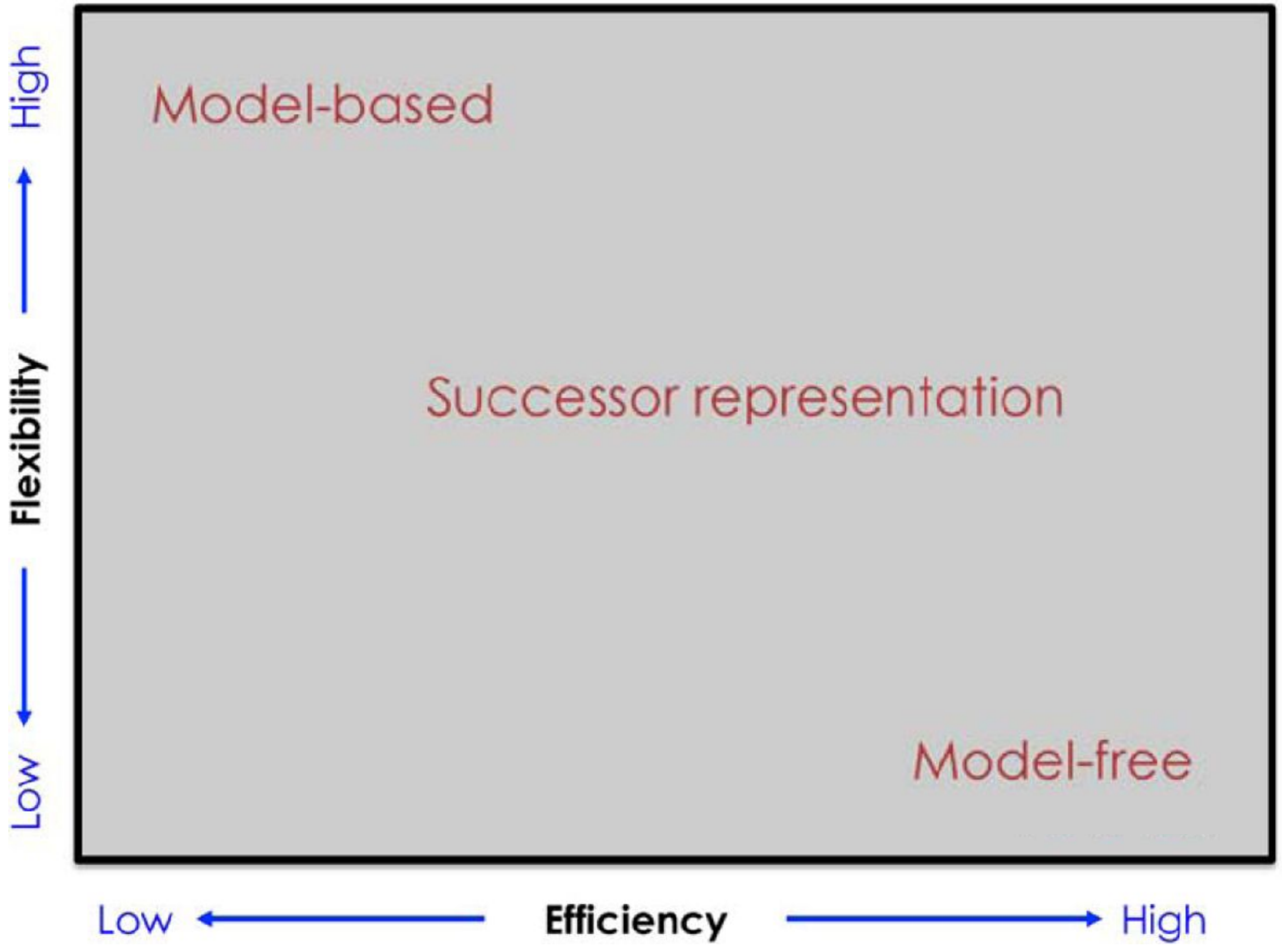
- Model-free methods
  - **Computationally efficient**
  - But **lack flexibility** to changes in the environment: *value is coupled to policy*
- Model-based methods
  - **Highly flexible:** *value and policy can be quickly recomputed* via simulations
  - But performing simulations are **computationally costly**
- SR falls in between



Gershman (2018)

Balancing flexibility and efficiency

- Model-free methods
  - **Computationally efficient**
  - But **lack flexibility** to changes in the environment: *value is coupled to policy*
- Model-based methods
  - **Highly flexible:** *value and policy can be quickly recomputed* via simulations
  - But performing simulations are **computationally costly**
- SR falls in between



Gershman (2018)

Successor Representation

$$V^{\pi}(s) = \sum_{s'} M(s, s')r(s')$$

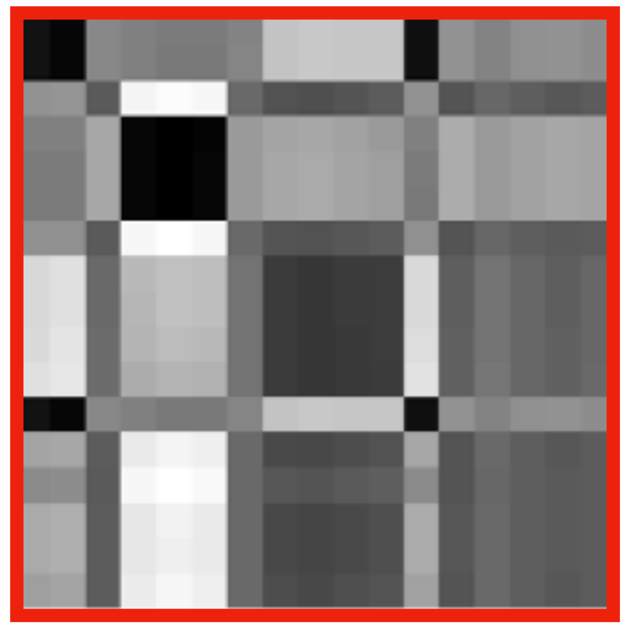
"Successor" as in succession: which states are likely to follow the current state, under a given policy

Decomposition of the TD value function into **SR** and **reward** components, allowing for faster generalization to changes in reward

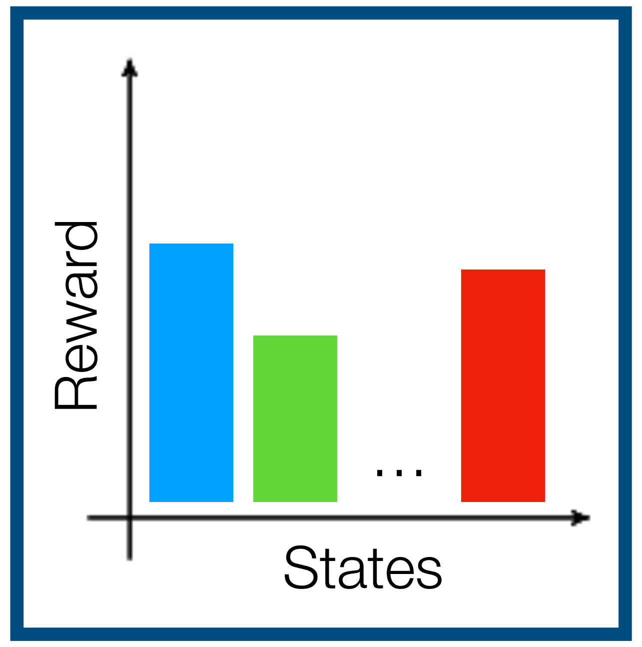
The SR is sensitive to policy, with representations skewed towards goals

In practice, computed using either ***off-policy*** (assuming a random policy) or ***on-policy*** methods (using delta-rule)

The SR naturally identifies ***subgoals*** (via Eigenvectors)



Successor Representation



Reward Values

Social learning

Learning is not only from environmental feedback, but also from social sources

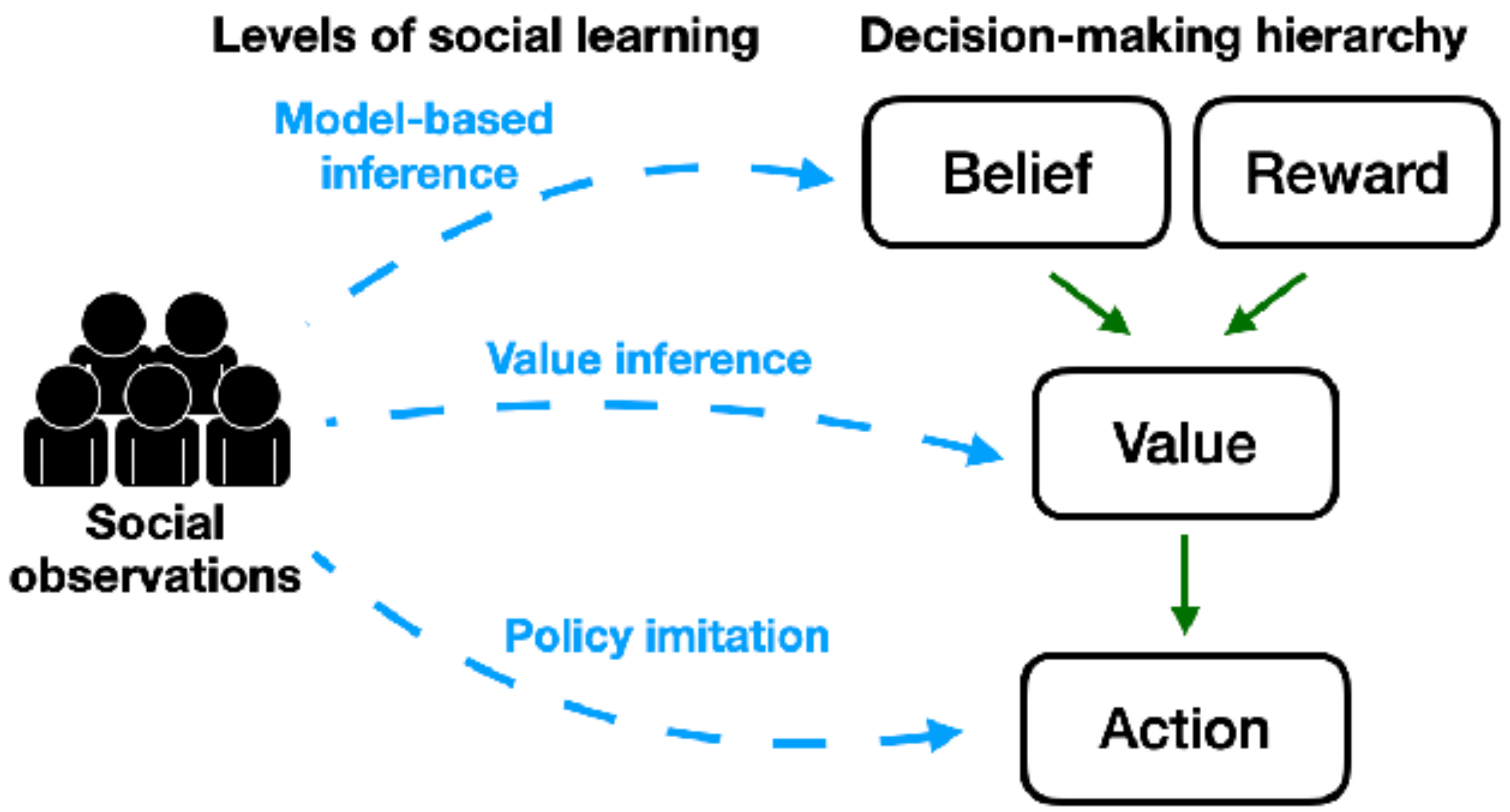
Imitation via observational learning, where **social learning strategies (SLS)** define various who, what, when

Theory of mind (ToM) involves inferring hidden mental states from observable behavior

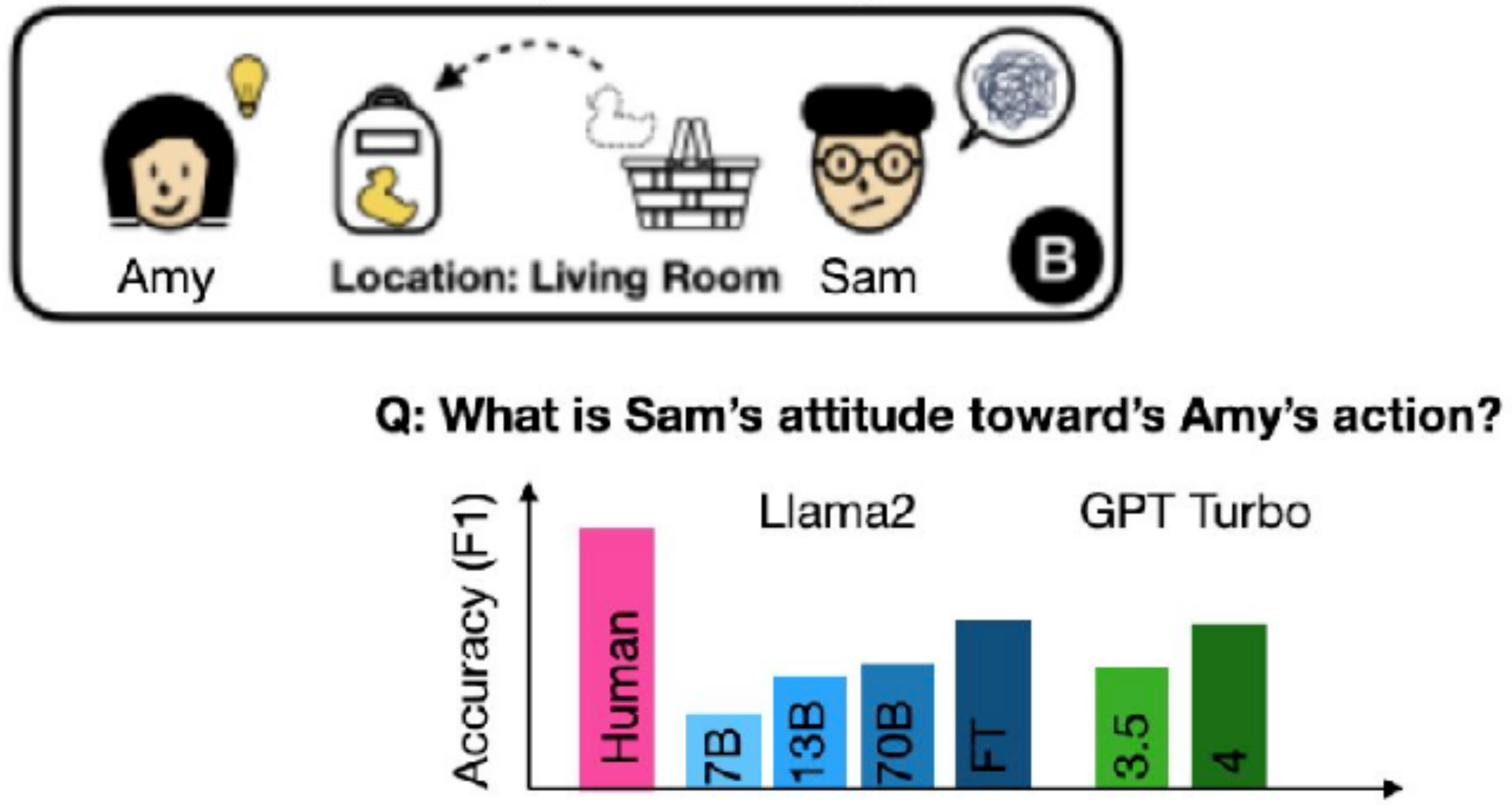
Various Bayesian formalisms of ToM, but typically intractable and a key limitation of current AI

Bandura (1961)

Wu, Vélez, & Cushman (2022



OpenToM Benchmark (Xu et al., 2024)



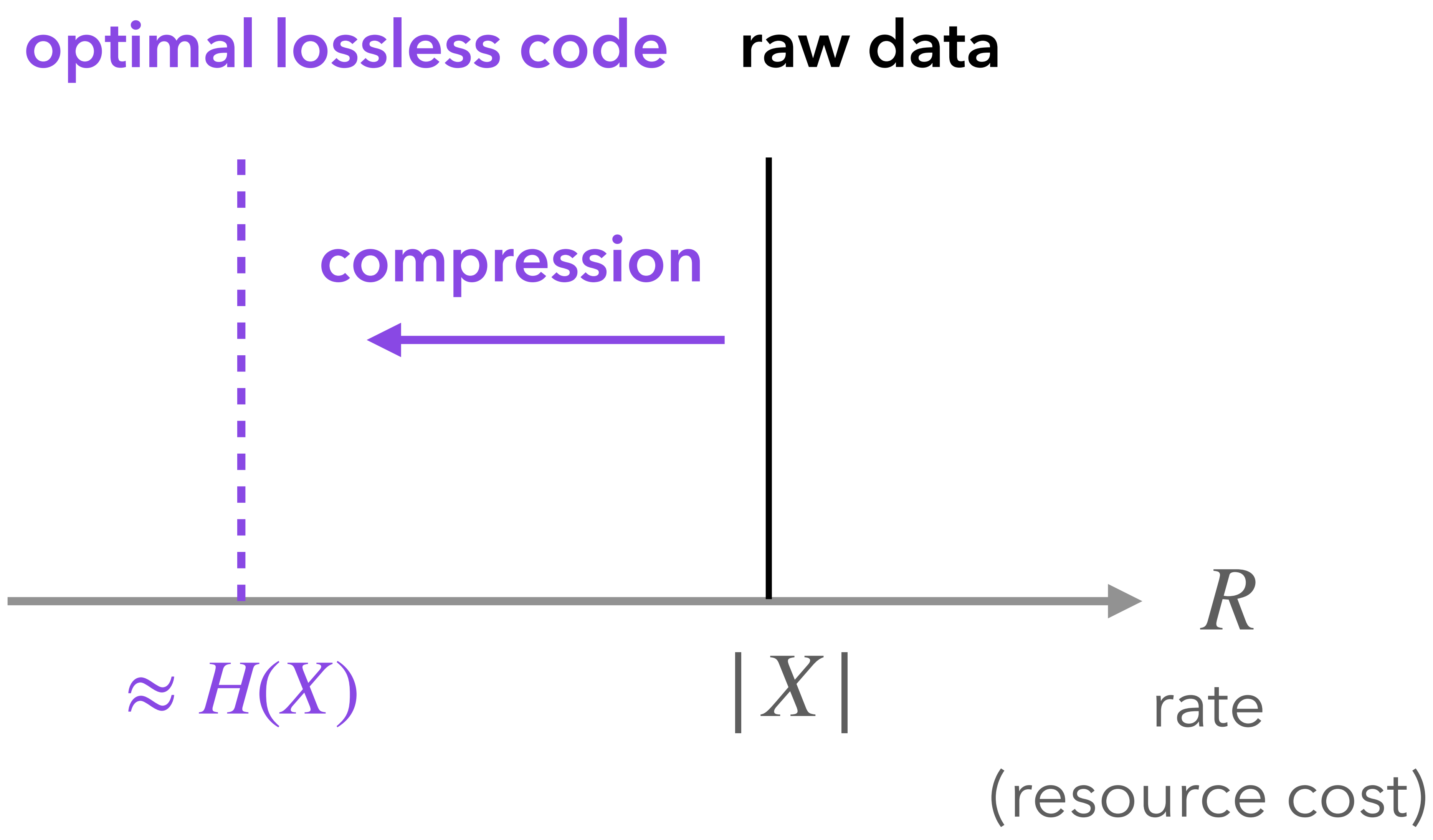
Compression

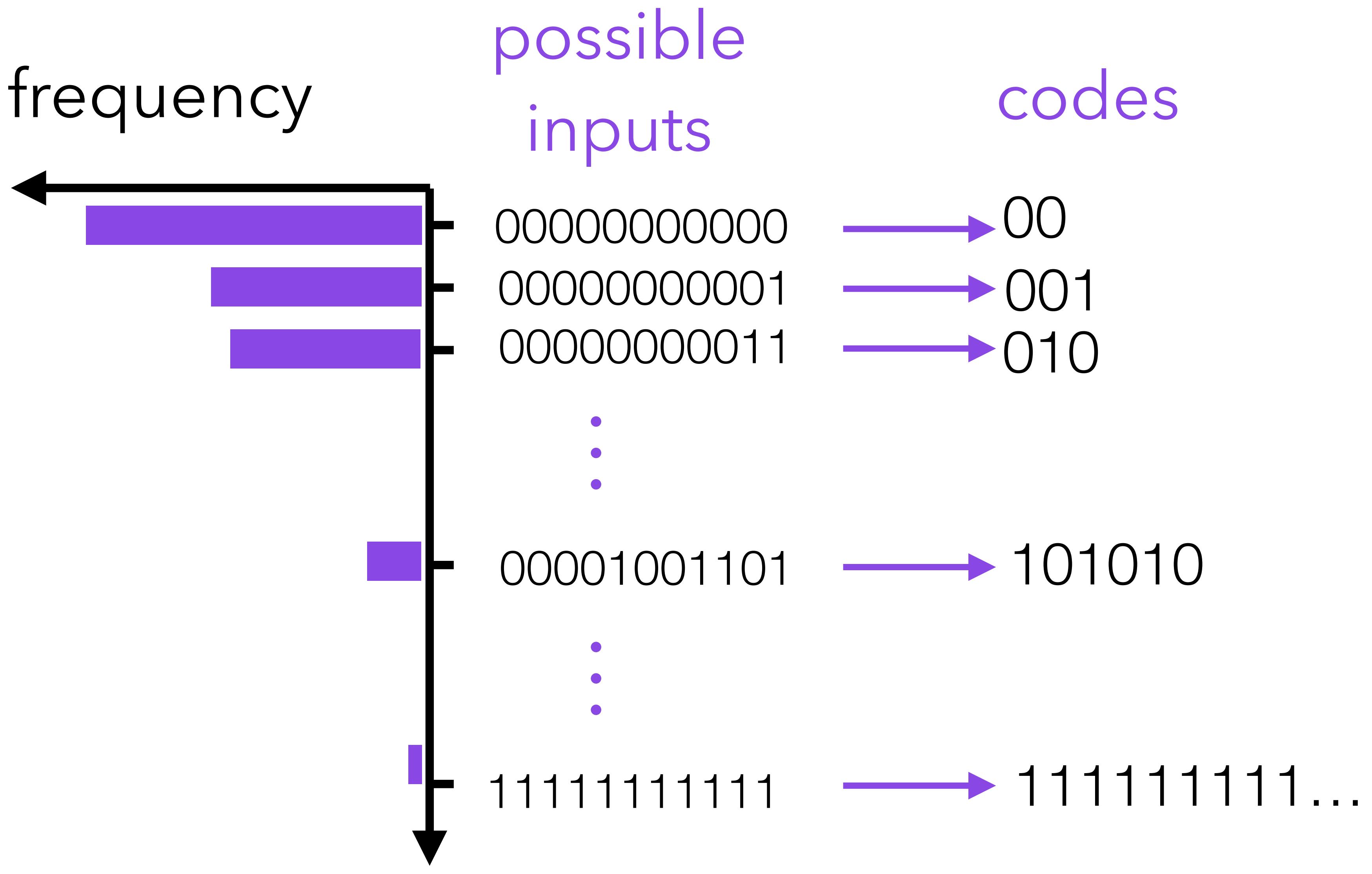
Compression decreases the resources R required to store data

Lossless compression is without loss of information

The **optimal lossless code** is based on assigning the shortest codes to the most frequent inputs: *source coding theorem*

Even greater compression is possible by allowing for distortions: **lossy compression**





Agenda for today

1. What is a concept?
2. Rule-based theories
3. Similarity-based theories
4. Hybrid approaches

What is a concept?

What is a concept?

Conceptual art

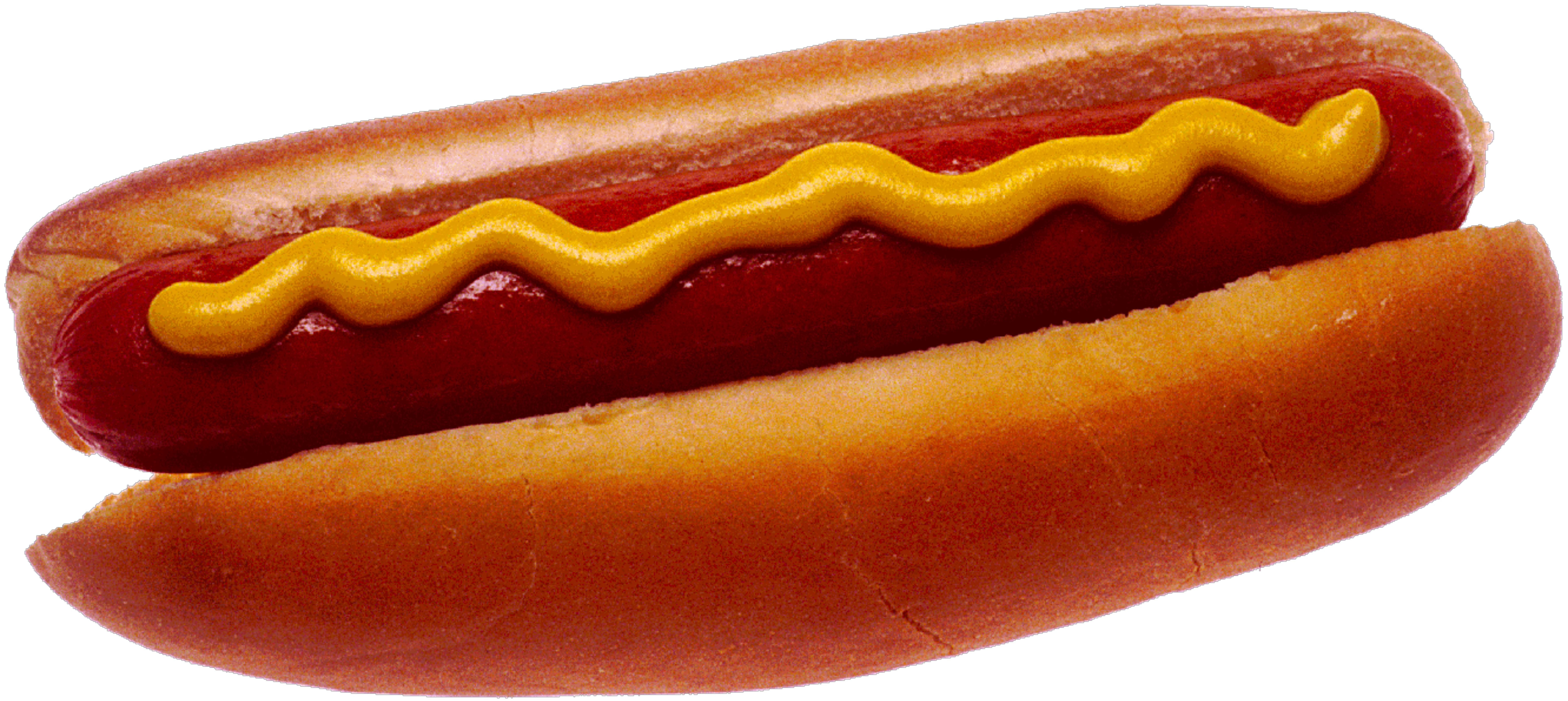

Maurizio Cattelan

Marcel Duchamp


Shutterstock

What is a “Sandwich?”

What is a “Sandwich?”



Is a hotdog a sandwich?

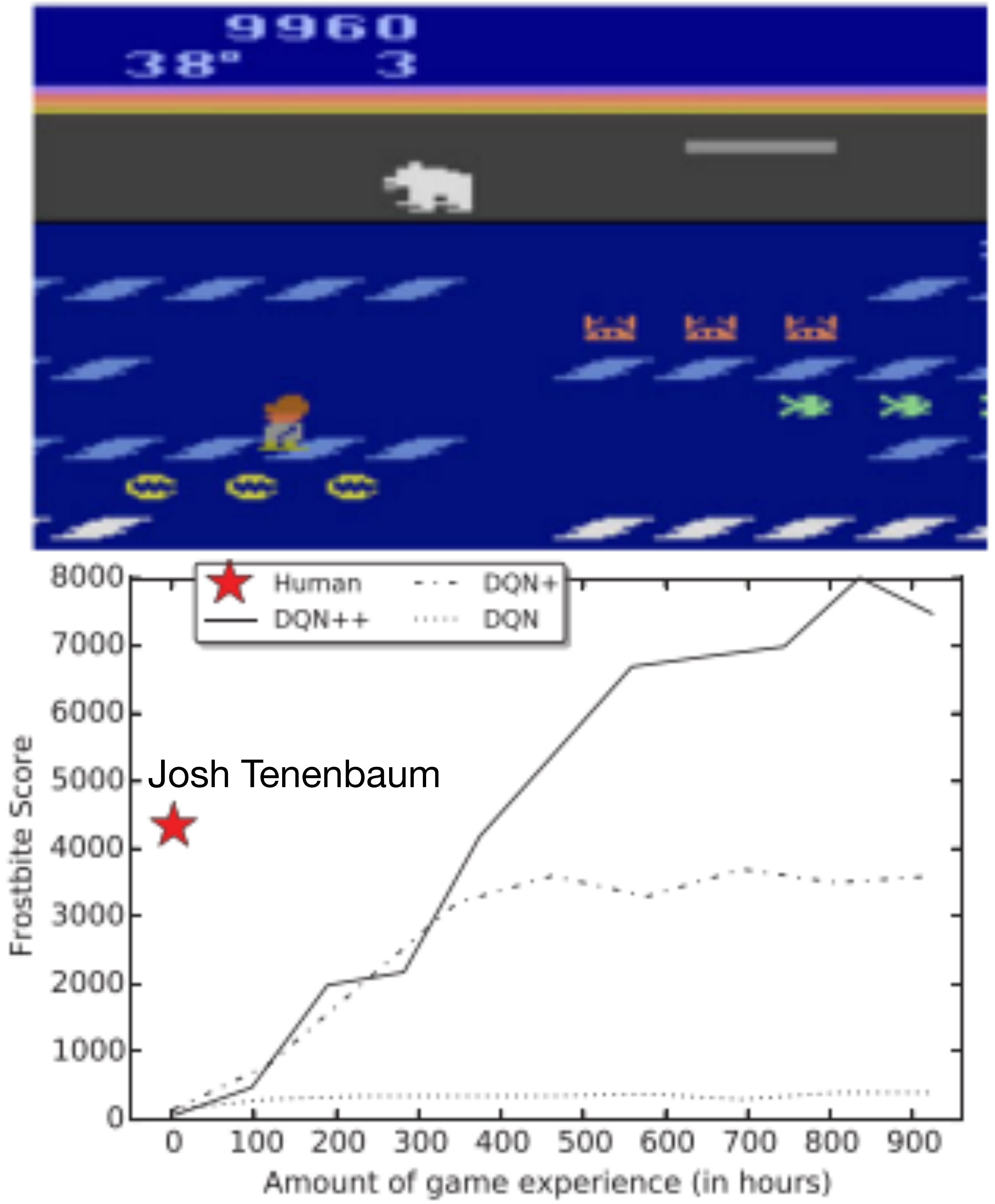
Concept learning is at the heart of many key aspects of intelligence

One-shot generalization

Lake et al., (2015); Lake et al., (2017)

Creative composition

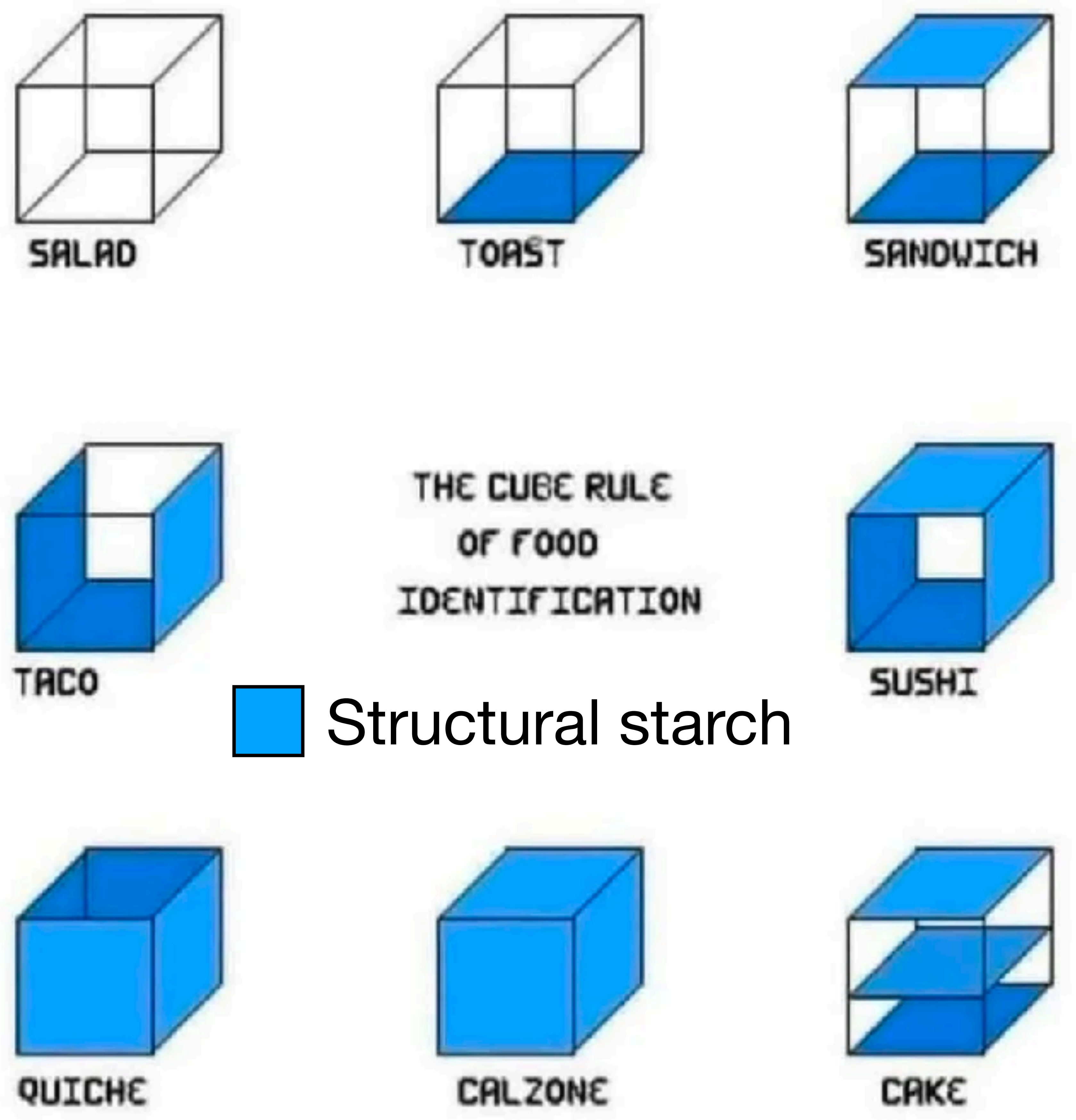
Rapid transfer



The study of categories and concepts

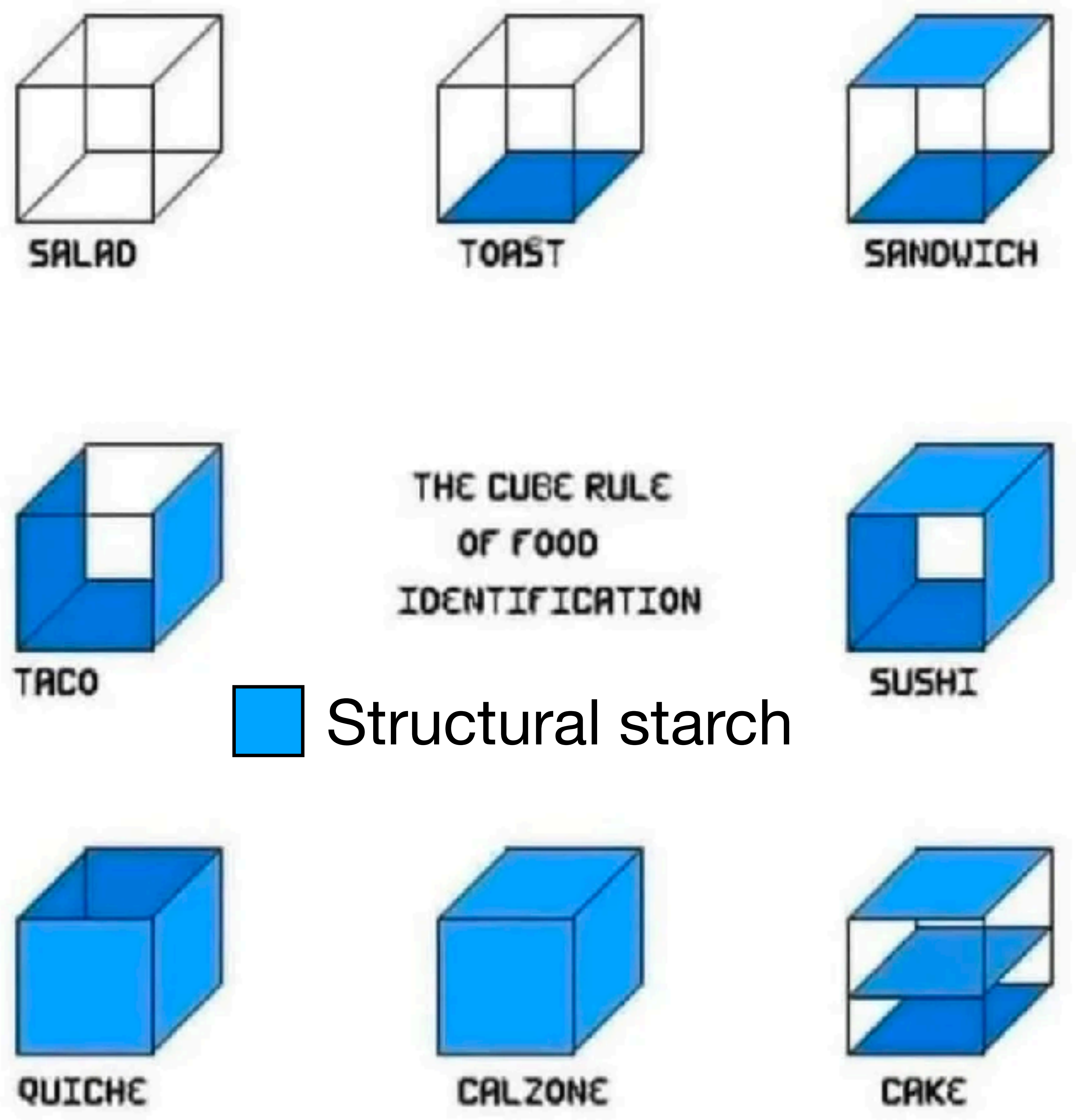
- A *category* is a set of objects in the world and a *concept* is a mental representation of a category
  - We will use the two interchangeably
- **Classical View** (Bruner et al., 1967):
  1. Concepts defined based on *necessary* and *sufficient* conditions for category membership
  2. Membership is all-or-nothing. All members are equally good
- This perspective dates to Aristotelian “forms” and Logical positivist philosophy (e.g., Quine, Popper, etc…)
- What are the necessary and sufficient conditions for something to be a sandwich?

Different approaches to defining concepts



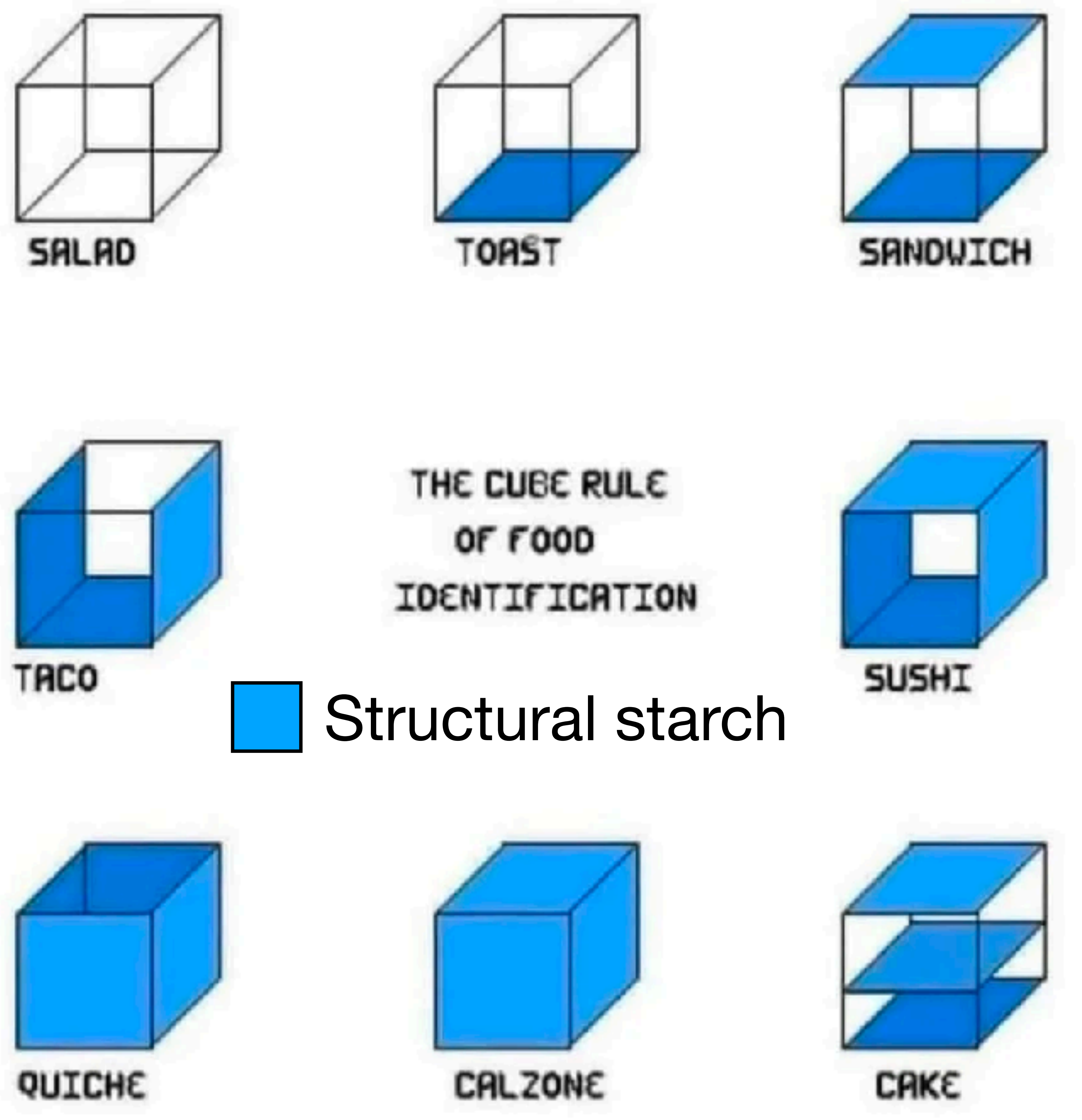
Different approaches to defining concepts

Rule-based approaches



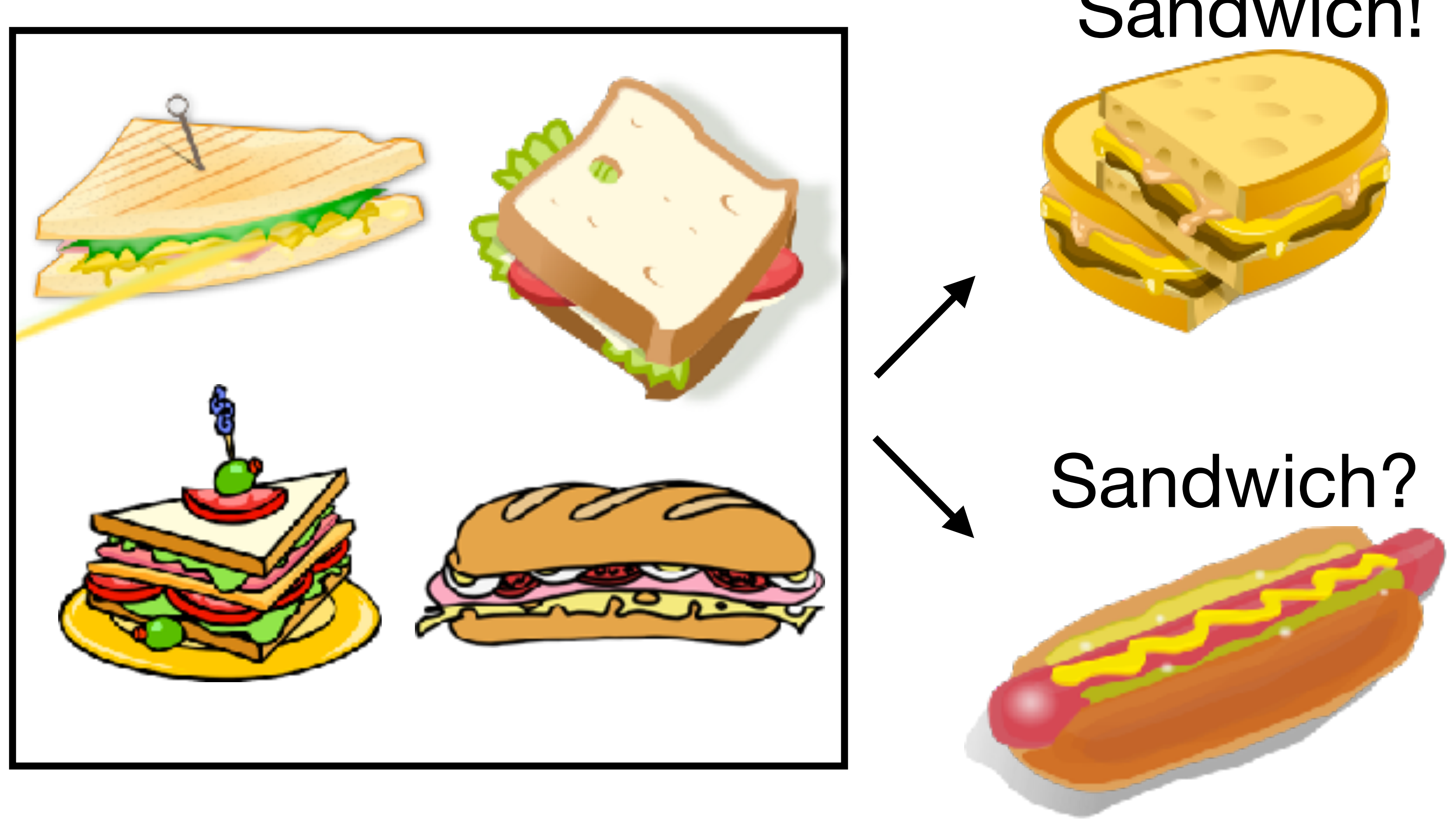
Different approaches to defining concepts

Rule-based approaches



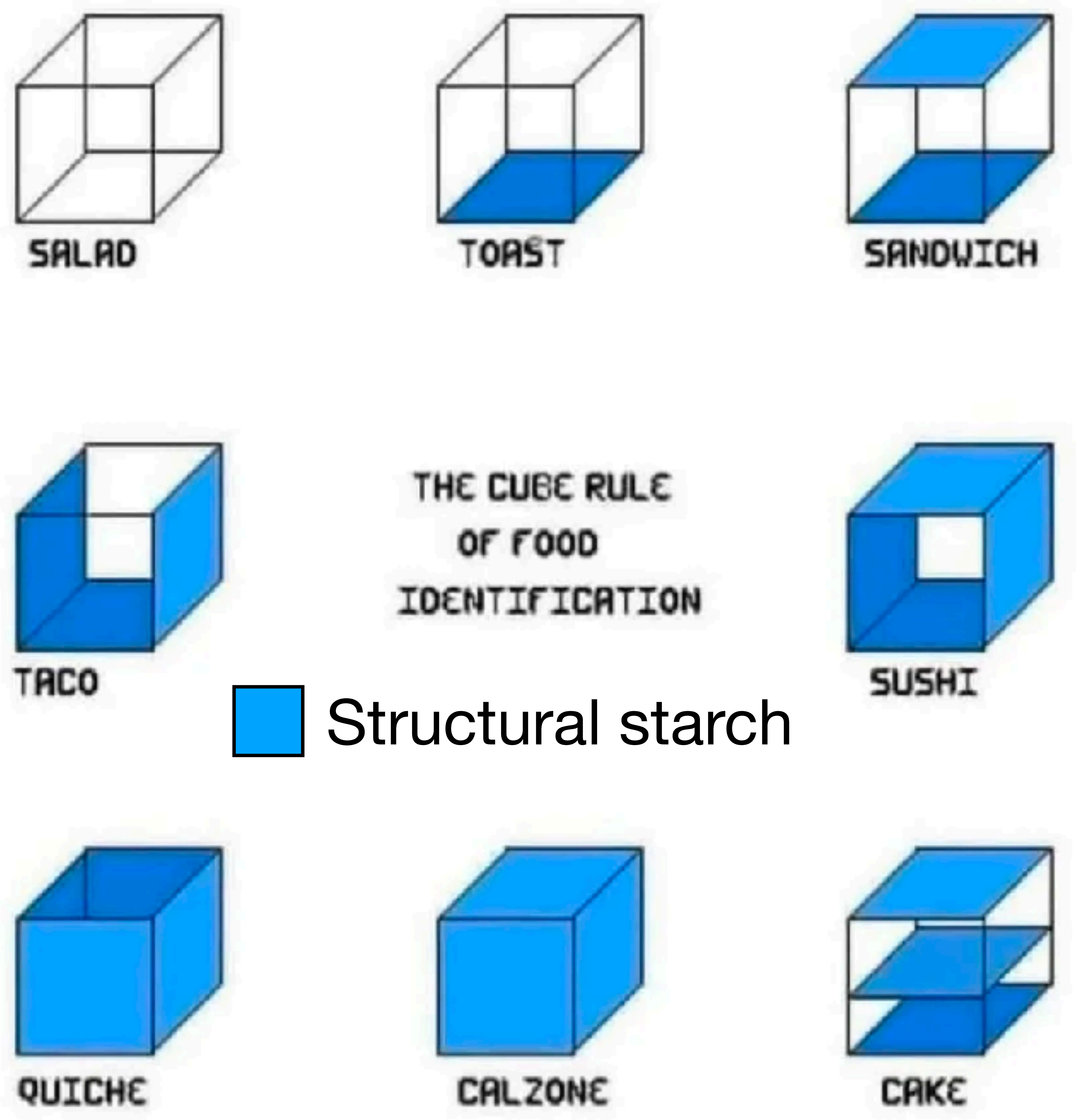
Similarity-bases approaches

Previous Experiences



Different approaches to defining concepts

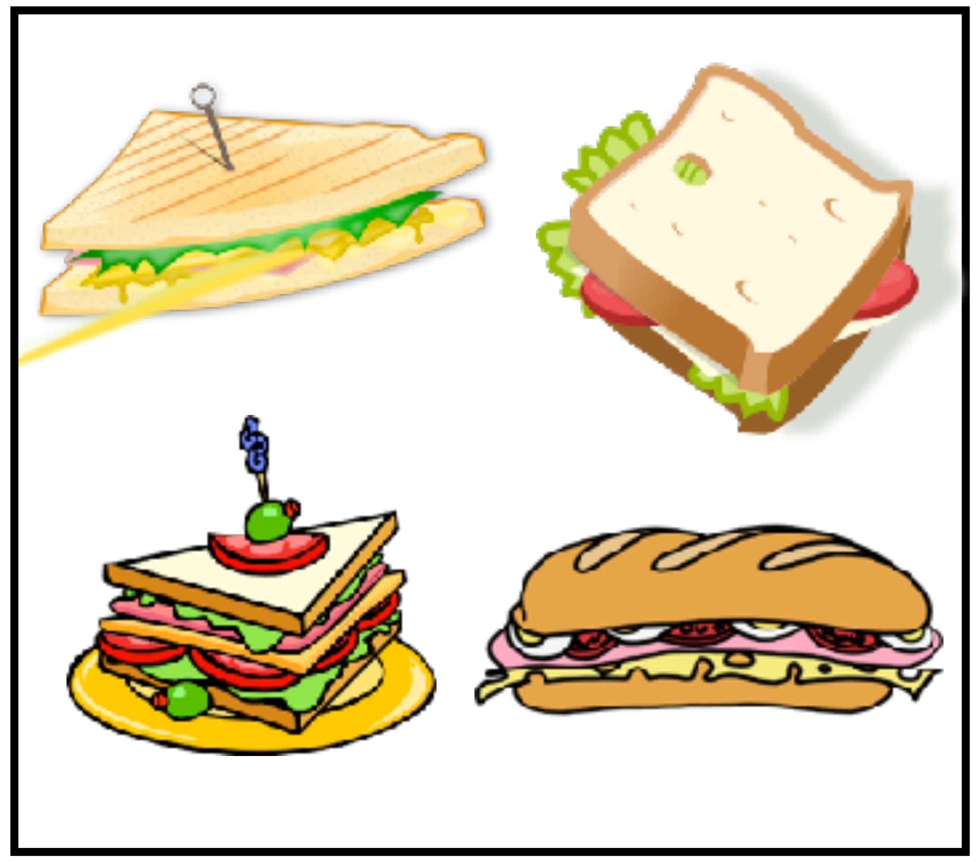
Rule-based approaches



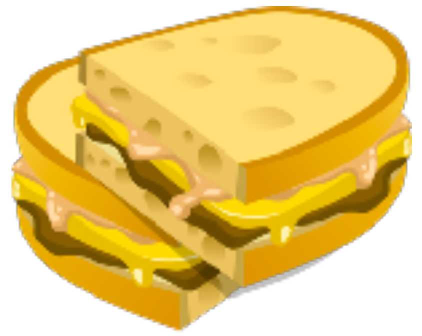
Structural starch

Similarity-bases approaches

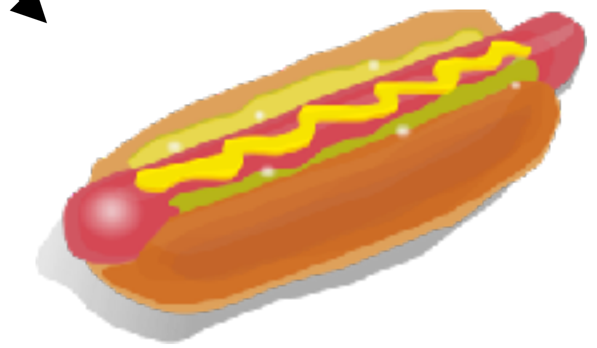
Previous Experiences

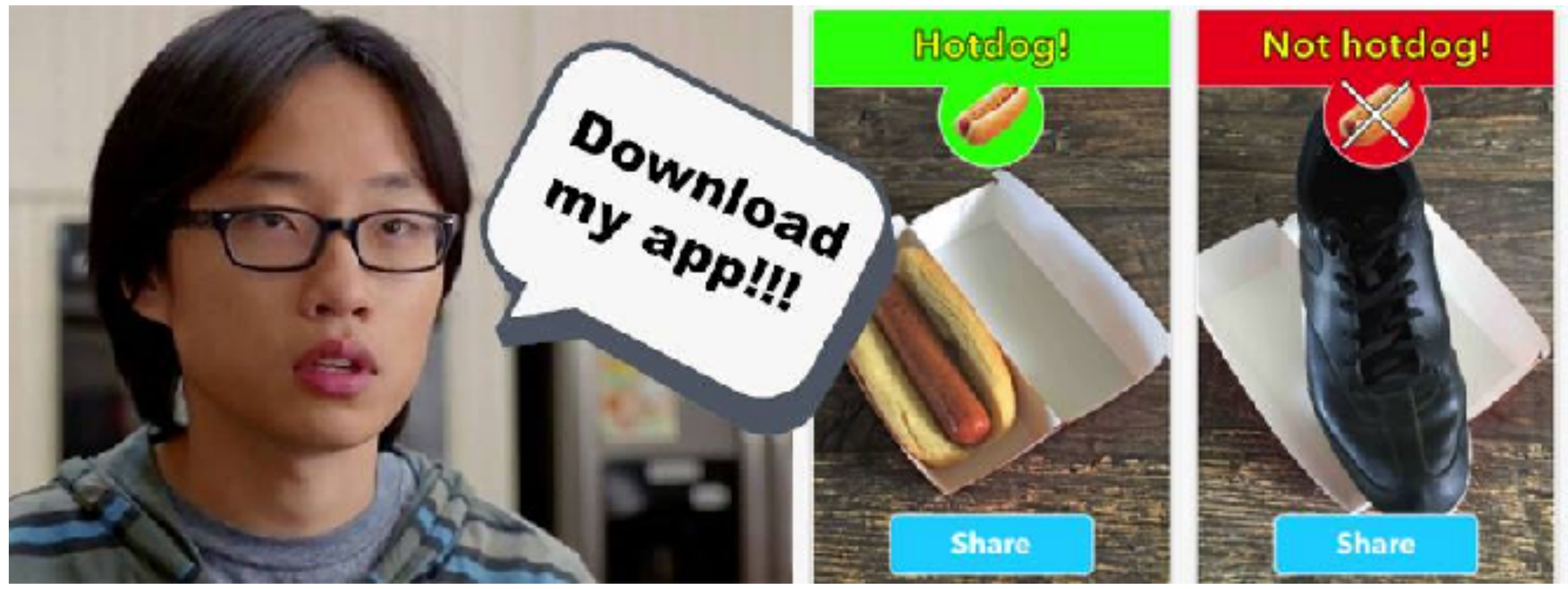


Sandwich!



Sandwich?





Rule-based theories

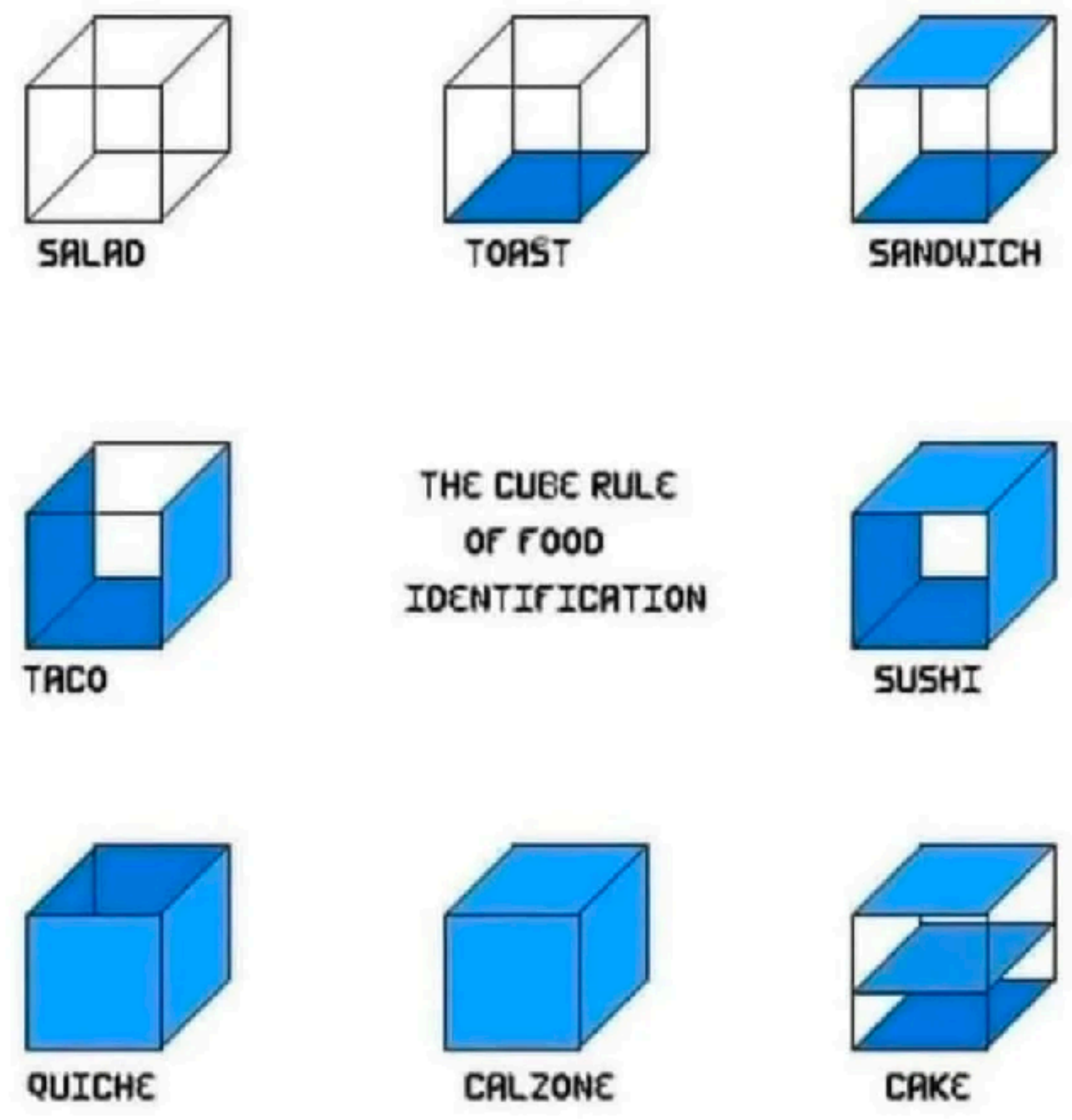
- Category membership defined by explicit rule-based boundaries (Ashby & Gott, 1988)
- The *specificity* of rules facilitates rapid generalization
- Rules can be *combined compositionally,* making them infinitely productive (Goodman et al., 2008)
- Yet *rigidity* makes them inflexible
  - What about root beer? Or open-faced sandwiches?
- Even when accounting for exceptions to rules (Nosofsky et al,. 1994), rule-based methods can only really explain human behavior when paired with other learning mechanisms (Erickson & Krushke, 1998; Ashby et al,. 1998; Love et al., 2004)

Furthermore, we wish to emphasize that in future in all cities, market-towns and in the country, ***the only ingredients used for the brewing of beer must be Barley, Hops and Water. -*** Reinheitsgebot (1516)

Rule-based theories

- Category membership defined by explicit rule-based boundaries (Ashby & Gott, 1988)
- The *specificity* of rules facilitates rapid generalization
- Rules can be *combined compositionally,* making them infinitely productive (Goodman et al., 2008)
- Yet *rigidity* makes them inflexible
    - What about root beer? Or open-faced sandwiches?
- Even when accounting for exceptions to rules (Nosofsky et al,. 1994), rule-based methods can only really explain human behavior when paired with other learning mechanisms (Erickson & Krushke, 1998; Ashby et al,. 1998; Love et al., 2004)

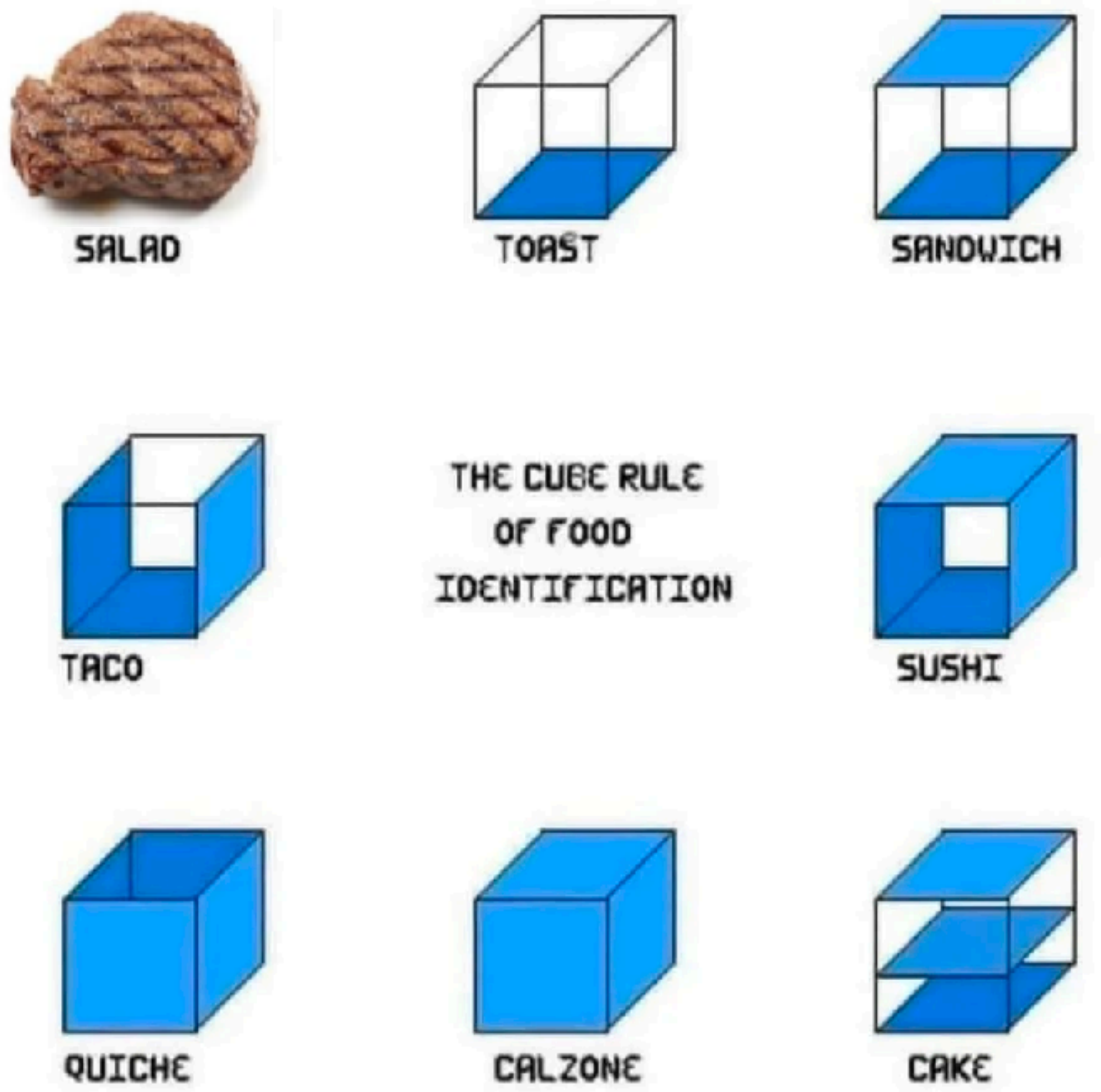
Furthermore, we wish to emphasize that in future in all cities, market-towns and in the country, ***the only ingredients used for the brewing of beer must be Barley, Hops and Water. -*** Reinheitsgebot (1516)



Rule-based theories

- Category membership defined by explicit rule-based boundaries (Ashby & Gott, 1988)
- The *specificity* of rules facilitates rapid generalization
- Rules can be *combined compositionally,* making them infinitely productive (Goodman et al., 2008)
- Yet *rigidity* makes them inflexible
  - What about root beer? Or open-faced sandwiches?
- Even when accounting for exceptions to rules (Nosofsky et al,. 1994), rule-based methods can only really explain human behavior when paired with other learning mechanisms (Erickson & Krushke, 1998; Ashby et al,. 1998; Love et al., 2004)

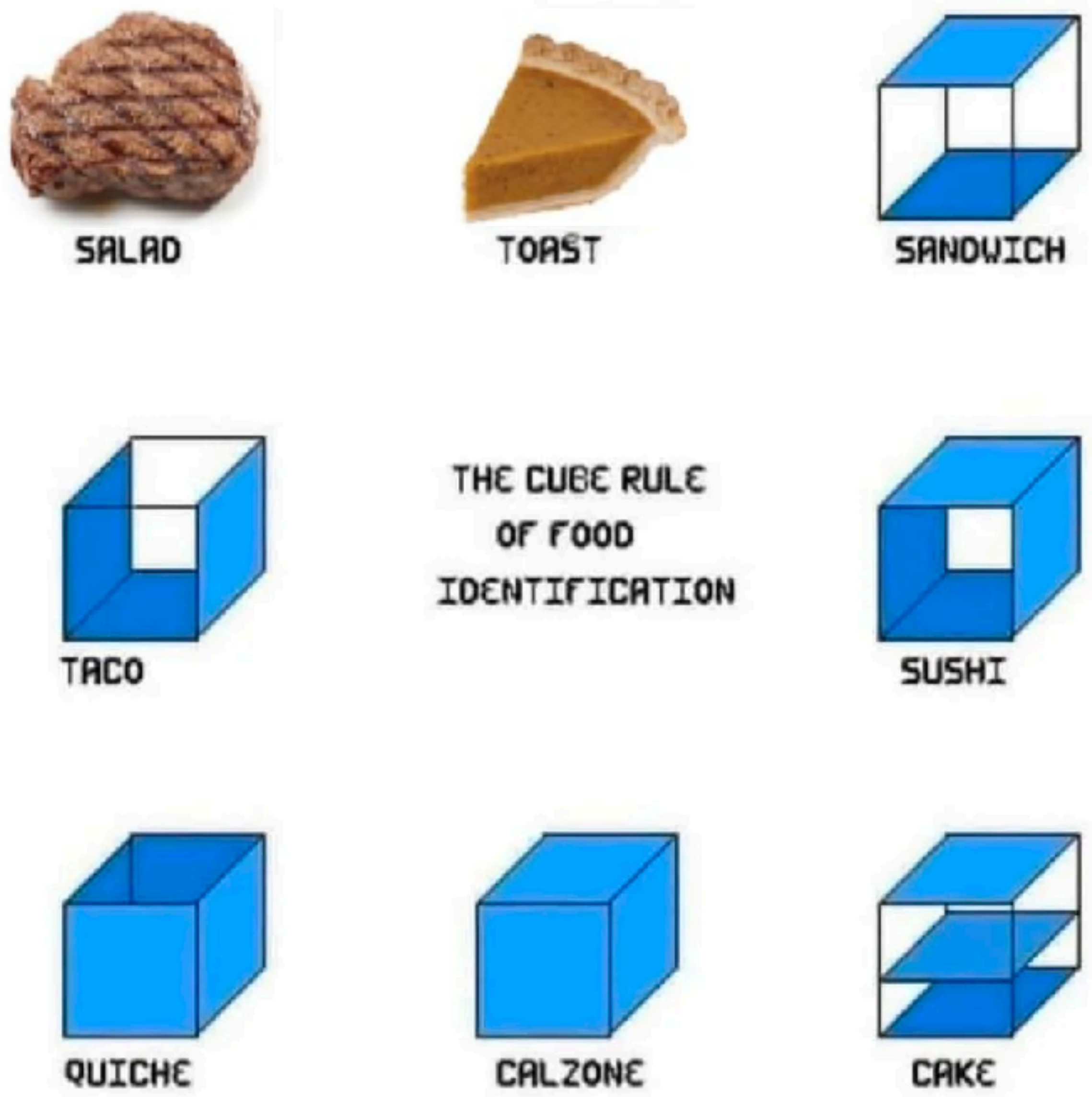
Furthermore, we wish to emphasize that in future in all cities, market-towns and in the country, ***the only ingredients used for the brewing of beer must be Barley, Hops and Water. -*** Reinheitsgebot (1516)



Rule-based theories

- Category membership defined by explicit rule-based boundaries (Ashby & Gott, 1988)
- The *specificity* of rules facilitates rapid generalization
- Rules can be *combined compositionally,* making them infinitely productive (Goodman et al., 2008)
- Yet *rigidity* makes them inflexible
  - What about root beer? Or open-faced sandwiches?
- Even when accounting for exceptions to rules (Nosofsky et al,. 1994), rule-based methods can only really explain human behavior when paired with other learning mechanisms (Erickson & Krushke, 1998; Ashby et al,. 1998; Love et al., 2004)

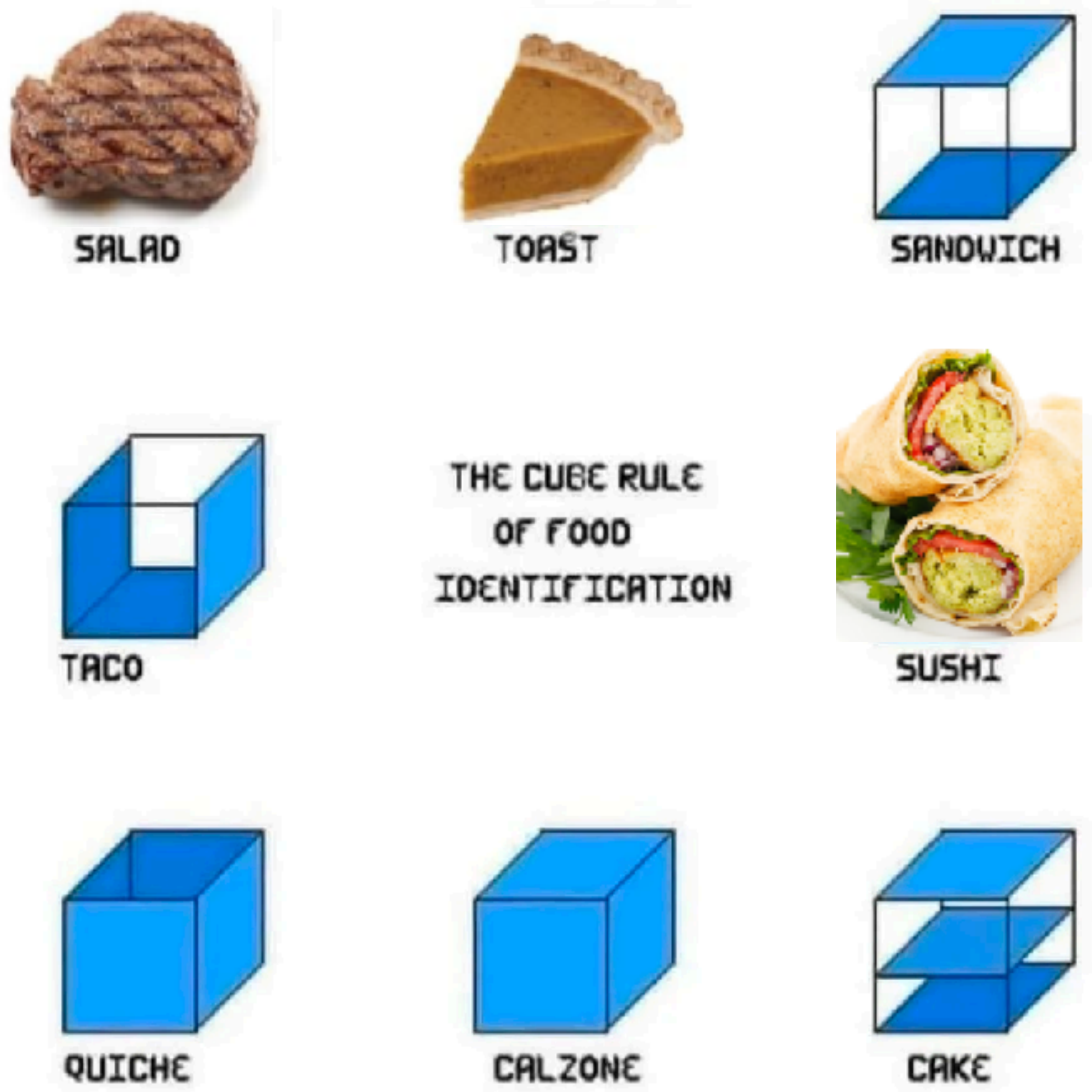
*Furthermore, we wish to emphasize that in future in all cities, market-towns and in the country, **the only ingredients used for the brewing of beer must be Barley, Hops and Water.* -** Reinheitsgebot (1516)



Rule-based theories

- Category membership defined by explicit rule-based boundaries (Ashby & Gott, 1988)
- The *specificity* of rules facilitates rapid generalization
- Rules can be *combined compositionally,* making them infinitely productive (Goodman et al., 2008)
- Yet *rigidity* makes them inflexible
    - What about root beer? Or open-faced sandwiches?
- Even when accounting for exceptions to rules (Nosofsky et al,. 1994), rule-based methods can only really explain human behavior when paired with other learning mechanisms (Erickson & Krushke, 1998; Ashby et al,. 1998; Love et al., 2004)

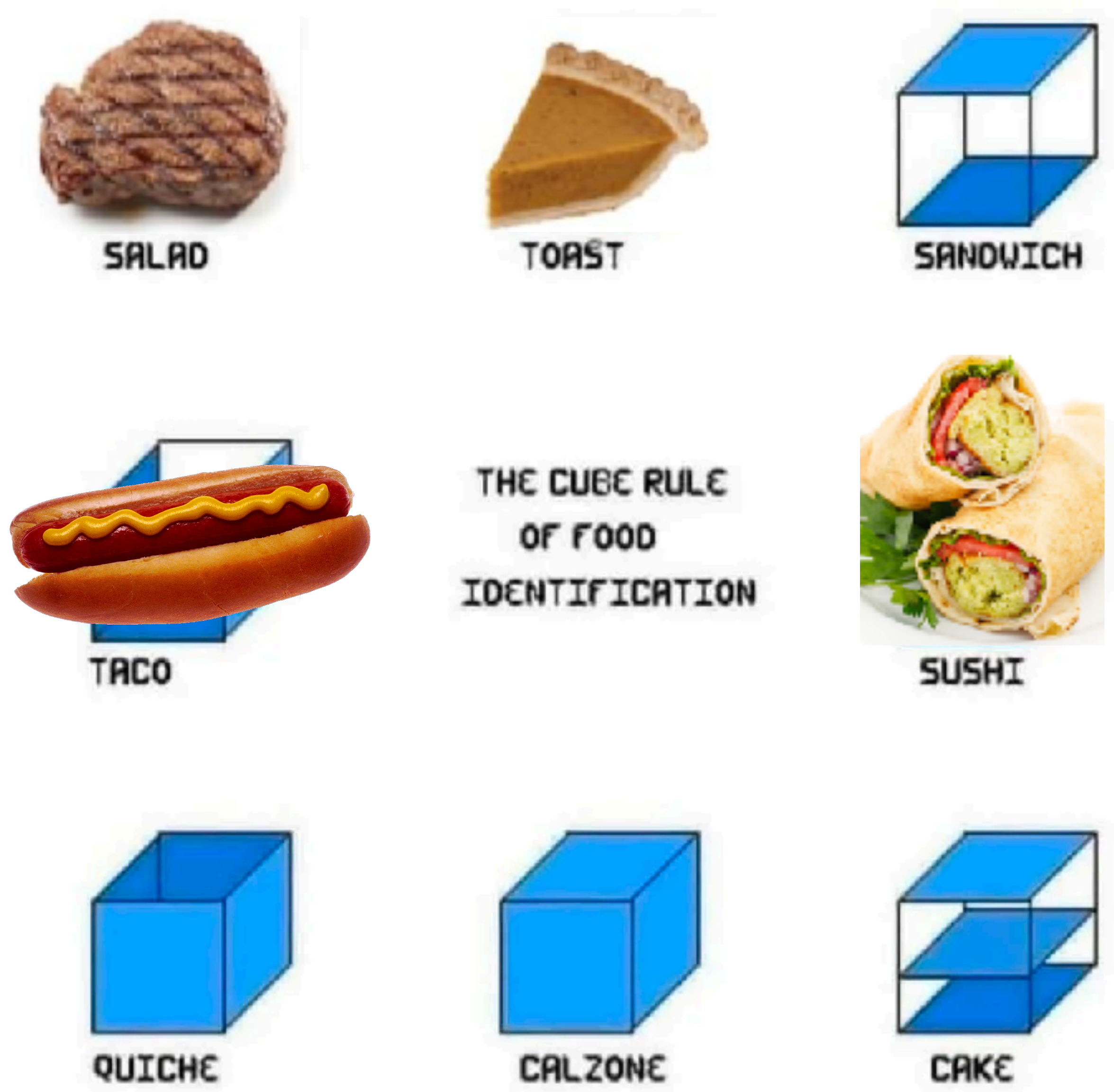
Furthermore, we wish to emphasize that in future in all cities, market-towns and in the country, ***the only ingredients used for the brewing of beer must be Barley, Hops and Water. -*** Reinheitsgebot (1516)



Rule-based theories

- Category membership defined by explicit rule-based boundaries (Ashby & Gott, 1988)
- The *specificity* of rules facilitates rapid generalization
- Rules can be *combined compositionally,* making them infinitely productive (Goodman et al., 2008)
- Yet *rigidity* makes them inflexible
    - What about root beer? Or open-faced sandwiches?
- Even when accounting for exceptions to rules (Nosofsky et al,. 1994), rule-based methods can only really explain human behavior when paired with other learning mechanisms (Erickson & Krushke, 1998; Ashby et al,. 1998; Love et al., 2004)

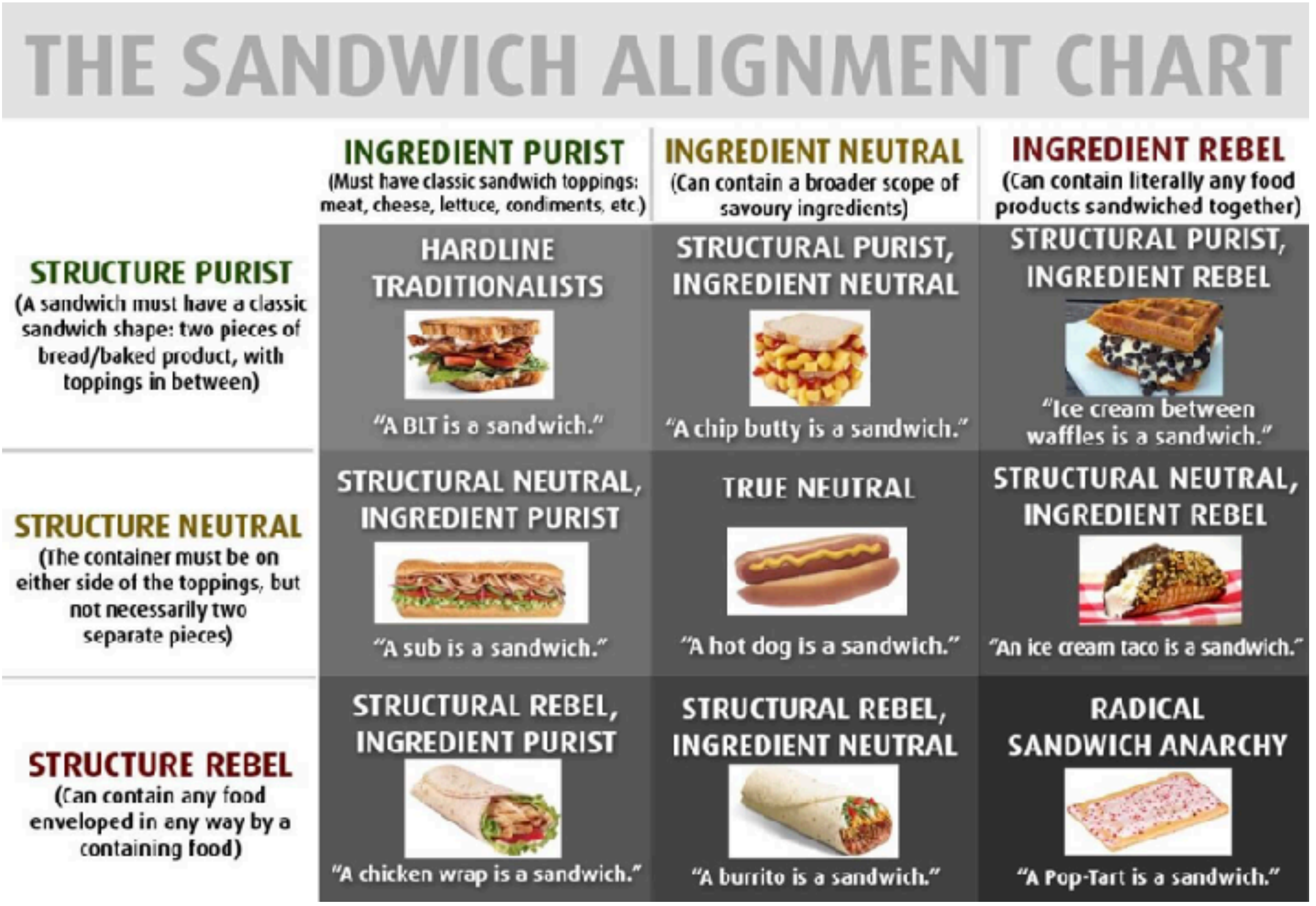
Furthermore, we wish to emphasize that in future in all cities, market-towns and in the country, ***the only ingredients used for the brewing of beer must be Barley, Hops and Water.*** **-** Reinheitsgebot (1516)



Rule-based theories

- Category membership defined by explicit rule-based boundaries (Ashby & Gott, 1988)
- The *specificity* of rules facilitates rapid generalization
- Rules can be *combined compositionally,* making them infinitely productive (Goodman et al., 2008)
- Yet *rigidity* makes them inflexible
  - What about root beer? Or open-faced sandwiches?
- Even when accounting for exceptions to rules (Nosofsky et al,. 1994), rule-based methods can only really explain human behavior when paired with other learning mechanisms (Erickson & Krushke, 1998; Ashby et al,. 1998; Love et al., 2004)

*Furthermore, we wish to emphasize that in future in all cities, market-towns and in the country, **the only ingredients used for the brewing of beer must be Barley, Hops and Water. -*** Reinheitsgebot (1516)



Borderline items

- Early psychological experiments showed that people didn't have well-defined categories (Hampton, 1979; Rosch & Mervis, 1975) and were even inconsistent when labeling the same object twice (McCloskey & Glucksberg, 1978)
- People's intuitive category boundaries seem to be fuzzy, can shift over time, and are sensitive to context

Borderline items

- Early psychological experiments showed that people didn't have well-defined categories (Hampton, 1979; Rosch & Mervis, 1975) and were even inconsistent when labeling the same object twice (McCloskey & Glucksberg, 1978)
- People's intuitive category boundaries seem to be fuzzy, can shift over time, and are sensitive to context

Is an olive a fruit?

Borderline items

- Early psychological experiments showed that people didn't have well-defined categories (Hampton, 1979; Rosch & Mervis, 1975) and were even inconsistent when labeling the same object twice (McCloskey & Glucksberg, 1978)
- People's intuitive category boundaries seem to be fuzzy, can shift over time, and are sensitive to context

Is an olive a fruit?

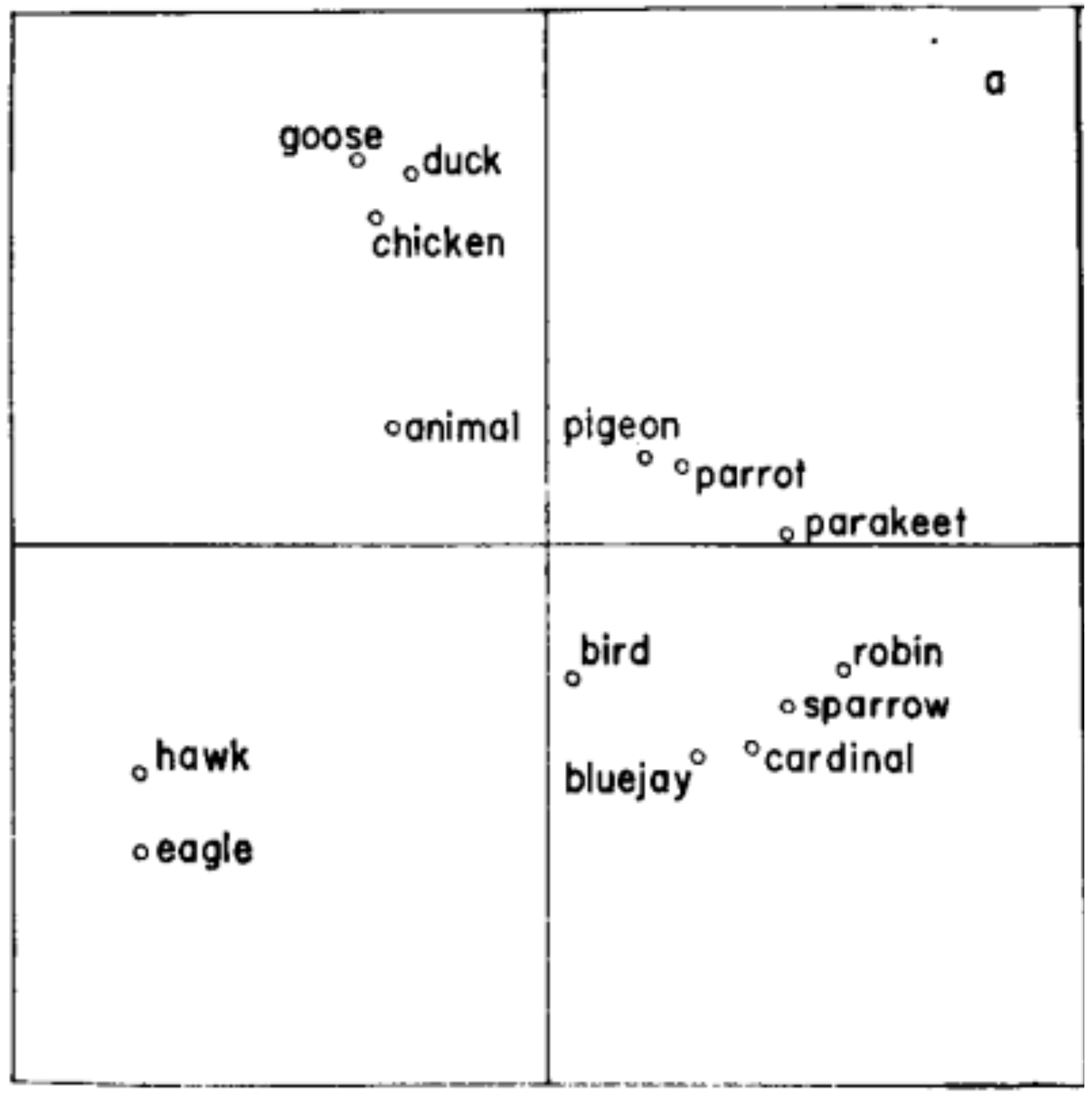
Borderline items

- Early psychological experiments showed that people didn't have well-defined categories (Hampton, 1979; Rosch & Mervis, 1975) and were even inconsistent when labeling the same object twice (McCloskey & Glucksberg, 1978)
- People's intuitive category boundaries seem to be fuzzy, can shift over time, and are sensitive to context
    - ~~Concepts can be defined based on necessary and sufficient conditions for category membership~~

Is an olive a fruit?


Typicality

- Some objects seem to fit better into categories than others
  - Some are more typical than others
- *Family resemblance theory* (Rosch & Mervis, 1975):
  - Items are typical if they
    a) have features frequent in the category
    b) don't have features frequent in other categories
- Thus, rather than hard & fast rules, similarity to typical items seems to matters

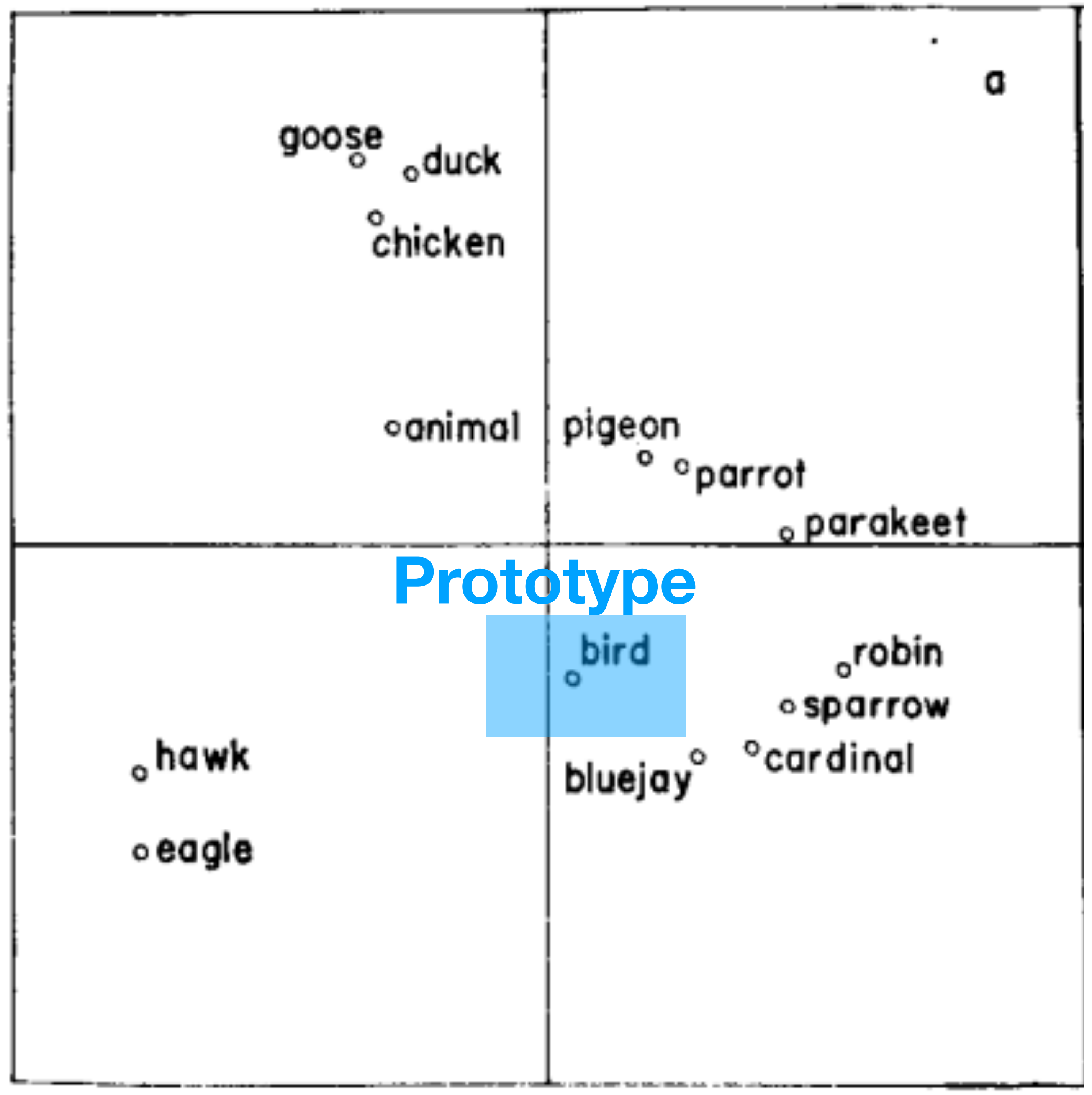


Multi-dimensional scaling of similarity ratings from Rips, Shoben, & Smith (1973)

Armstrong, Gleitman, & Gleitman (1983)

Typicality

- Some objects seem to fit better into categories than others
    - Some are more typical than others
- *Family resemblance theory* (Rosch & Mervis, 1975):
    - Items are typical if they
      a) have features frequent in the category
      b) don't have features frequent in other categories
- Thus, rather than hard & fast rules, similarity to typical items seems to matters

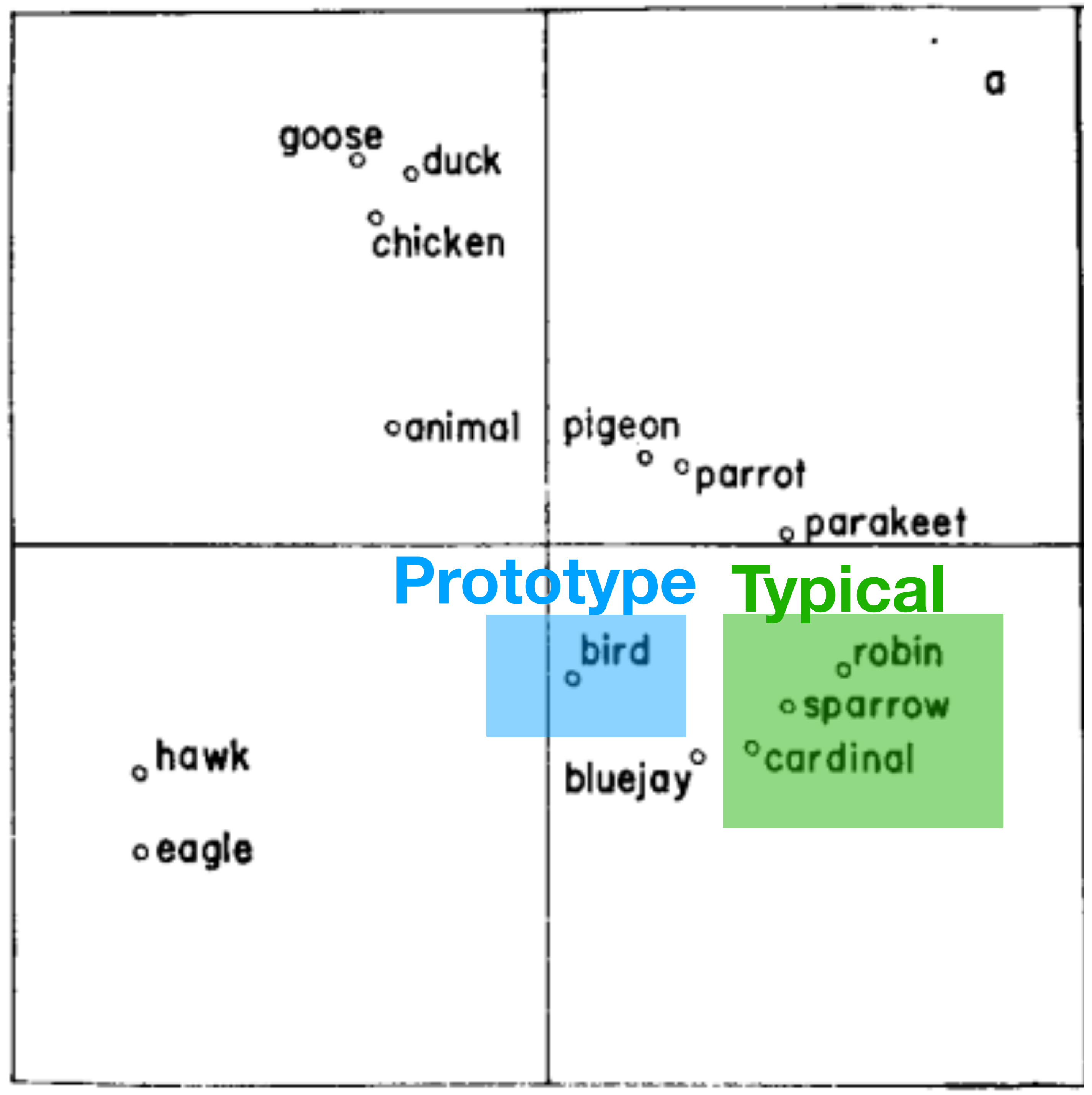


Multi-dimensional scaling of similarity ratings from Rips, Shoben, & Smith (1973)

Armstrong, Gleitman, & Gleitman (1983)

Typicality

- Some objects seem to fit better into categories than others
  - Some are more typical than others
- *Family resemblance theory* (Rosch & Mervis, 1975):
  - Items are typical if they
    a) have features frequent in the category
    b) don't have features frequent in other categories
- Thus, rather than hard & fast rules, similarity to typical items seems to matters

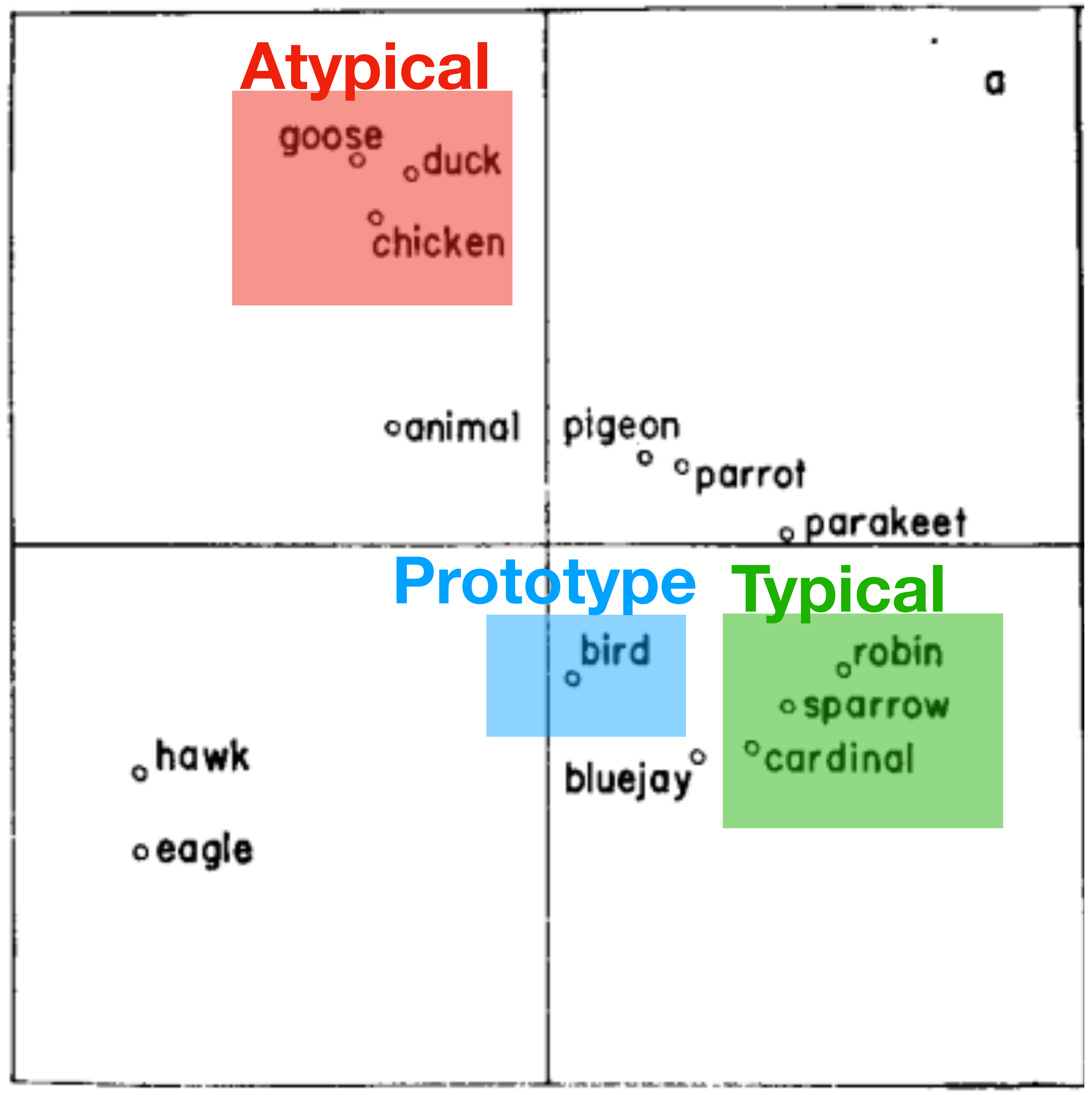


Multi-dimensional scaling of similarity ratings from Rips, Shoben, & Smith (1973)

Armstrong, Gleitman, & Gleitman (1983)

Typicality

- Some objects seem to fit better into categories than others
  - Some are more typical than others
- *Family resemblance theory* (Rosch & Mervis, 1975):
  - Items are typical if they
    a) have features frequent in the category
    b) don't have features frequent in other categories
- Thus, rather than hard & fast rules, similarity to typical items seems to matters

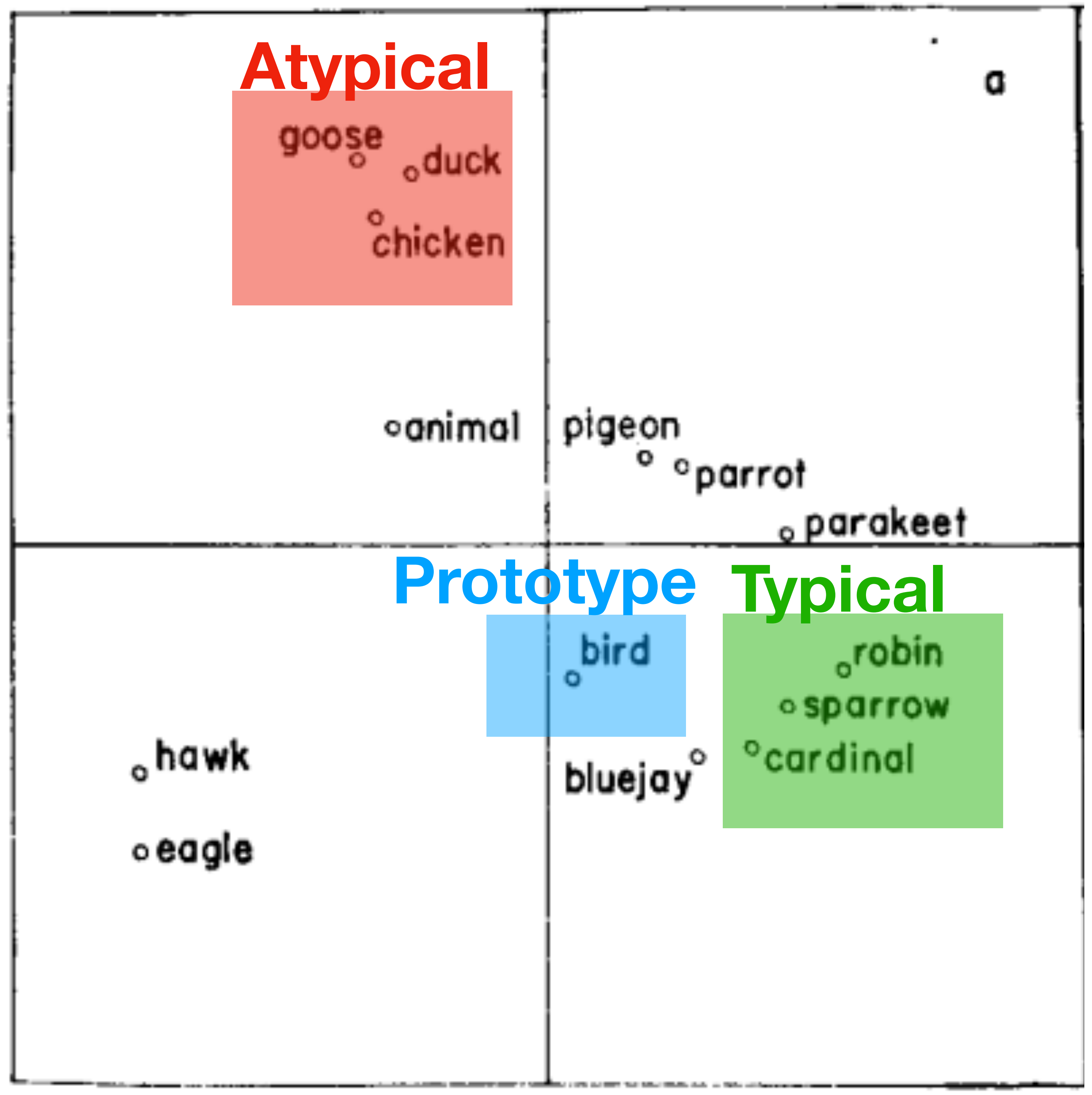


Multi-dimensional scaling of similarity ratings from Rips, Shoben, & Smith (1973)


Armstrong, Gleitman, & Gleitman (1983)

Typicality

- Some objects seem to fit better into categories than others
  - Some are more typical than others
- *Family resemblance theory* (Rosch & Mervis, 1975):
  - Items are typical if they
    a) have features frequent in the category
    b) don't have features frequent in other categories
- Thus, rather than hard & fast rules, similarity to typical items seems to matters

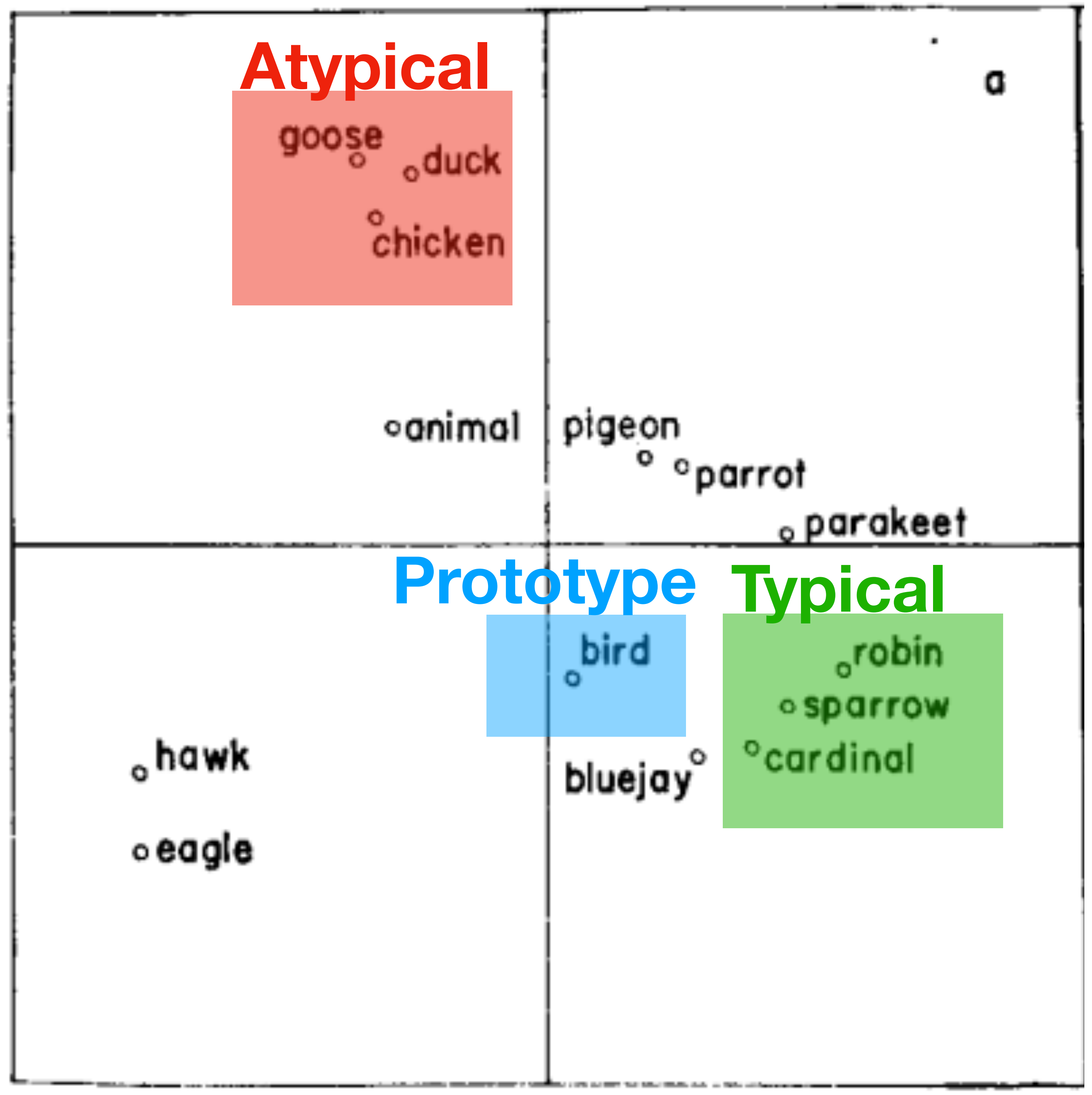


Multi-dimensional scaling of similarity ratings from Rips, Shoben, & Smith (1973)

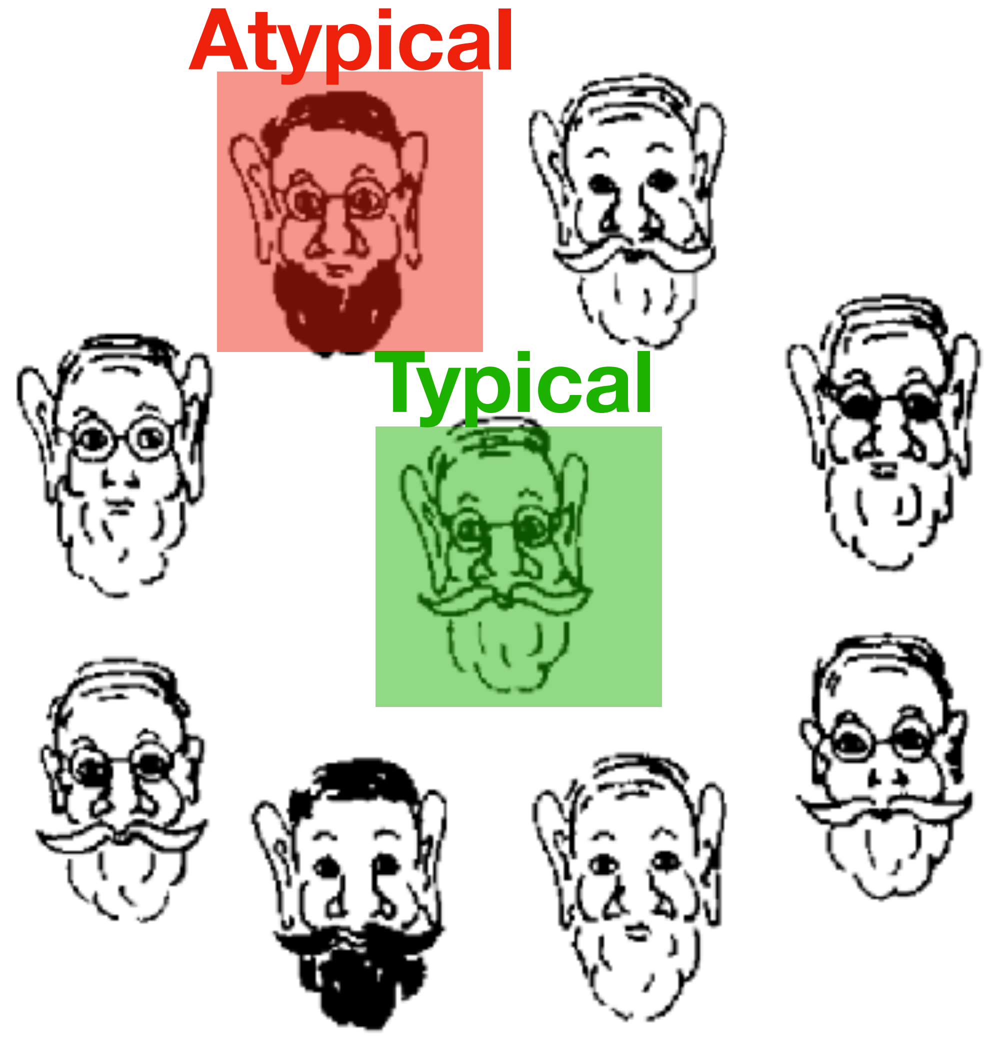
Armstrong, Gleitman, & Gleitman (1983)

Typicality

- Some objects seem to fit better into categories than others
  - Some are more typical than others
- *Family resemblance theory* (Rosch & Mervis, 1975):
  - Items are typical if they
    a) have features frequent in the category
    b) don't have features frequent in other categories
- Thus, rather than hard & fast rules, similarity to typical items seems to matters



Multi-dimensional scaling of similarity ratings from Rips, Shoben, & Smith (1973)

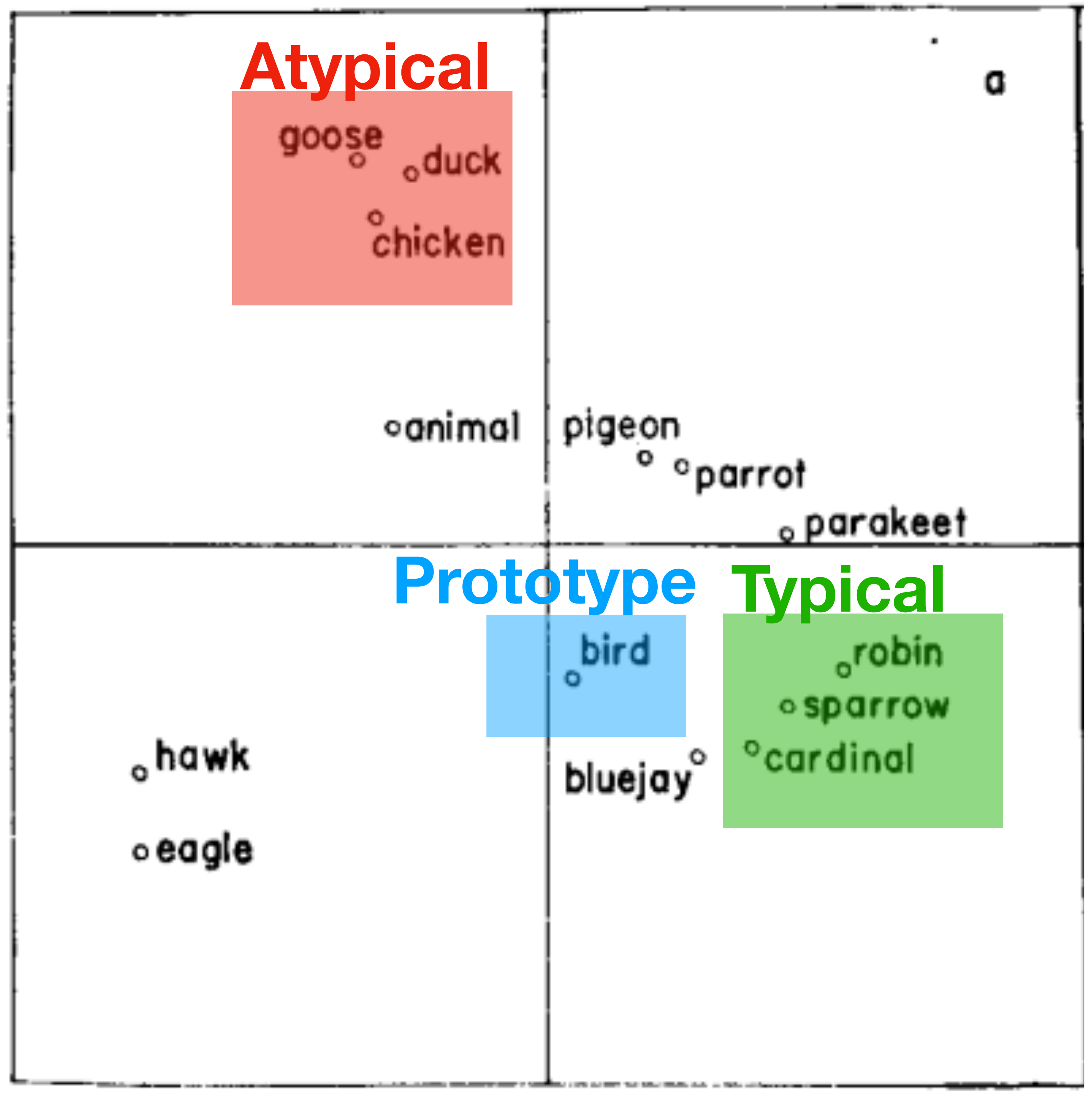


Armstrong, Gleitman, & Gleitman (1983)

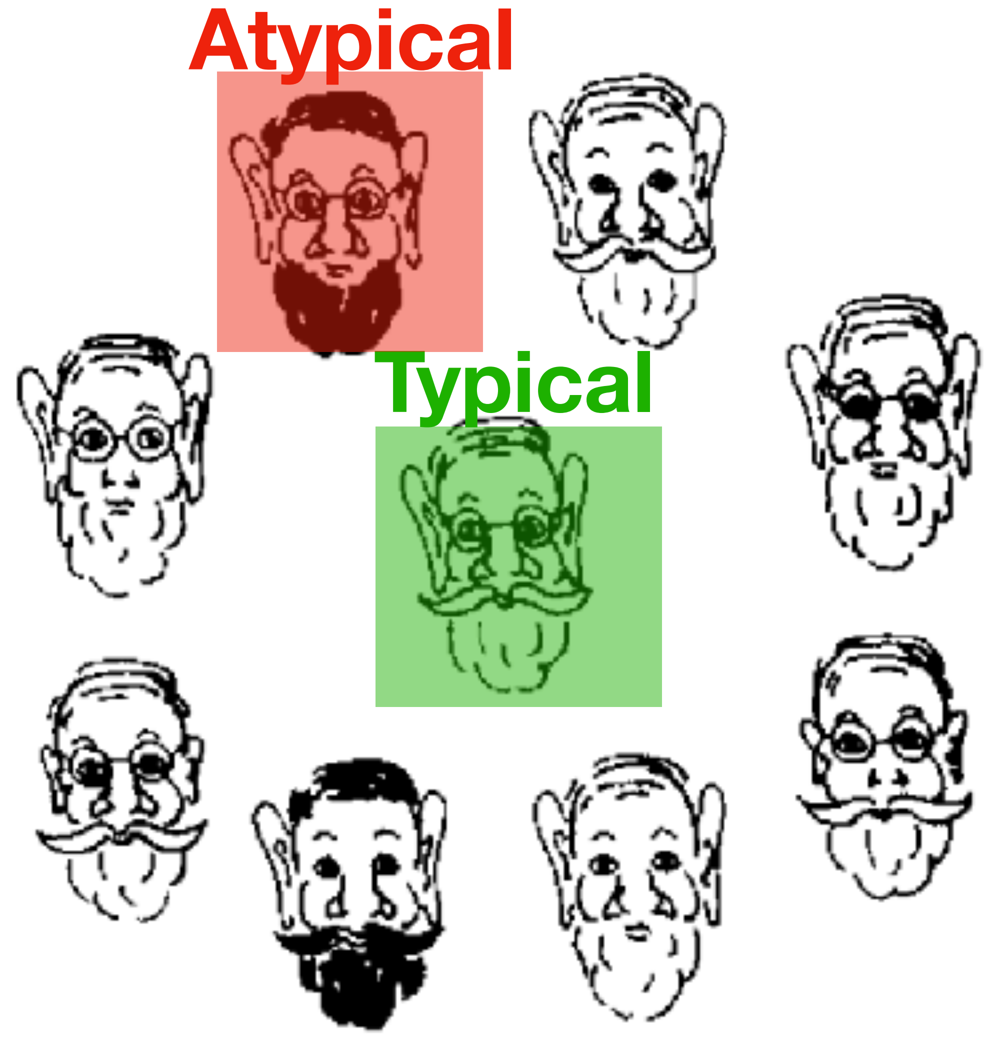
Typicality

- Some objects seem to fit better into categories than others
  - Some are more typical than others
- *Family resemblance theory* (Rosch & Mervis, 1975):
  - Items are typical if they
    a) have features frequent in the category
    b) don't have features frequent in other categories
- Thus, rather than hard & fast rules, similarity to typical items seems to matters

- ~~Membership is all-or-nothing. All members are equally good~~



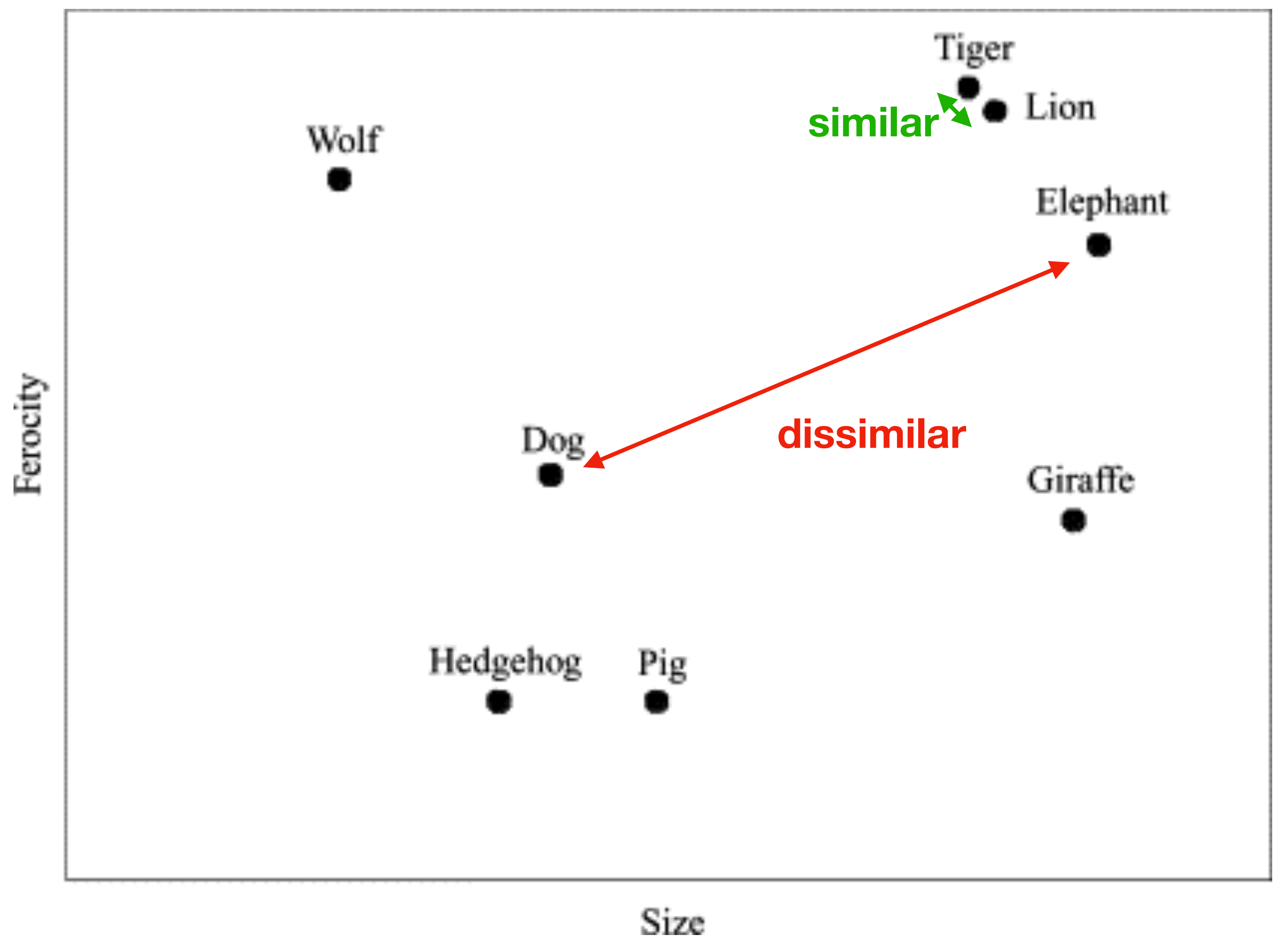
Multi-dimensional scaling of similarity ratings from Rips, Shoben, & Smith (1973)

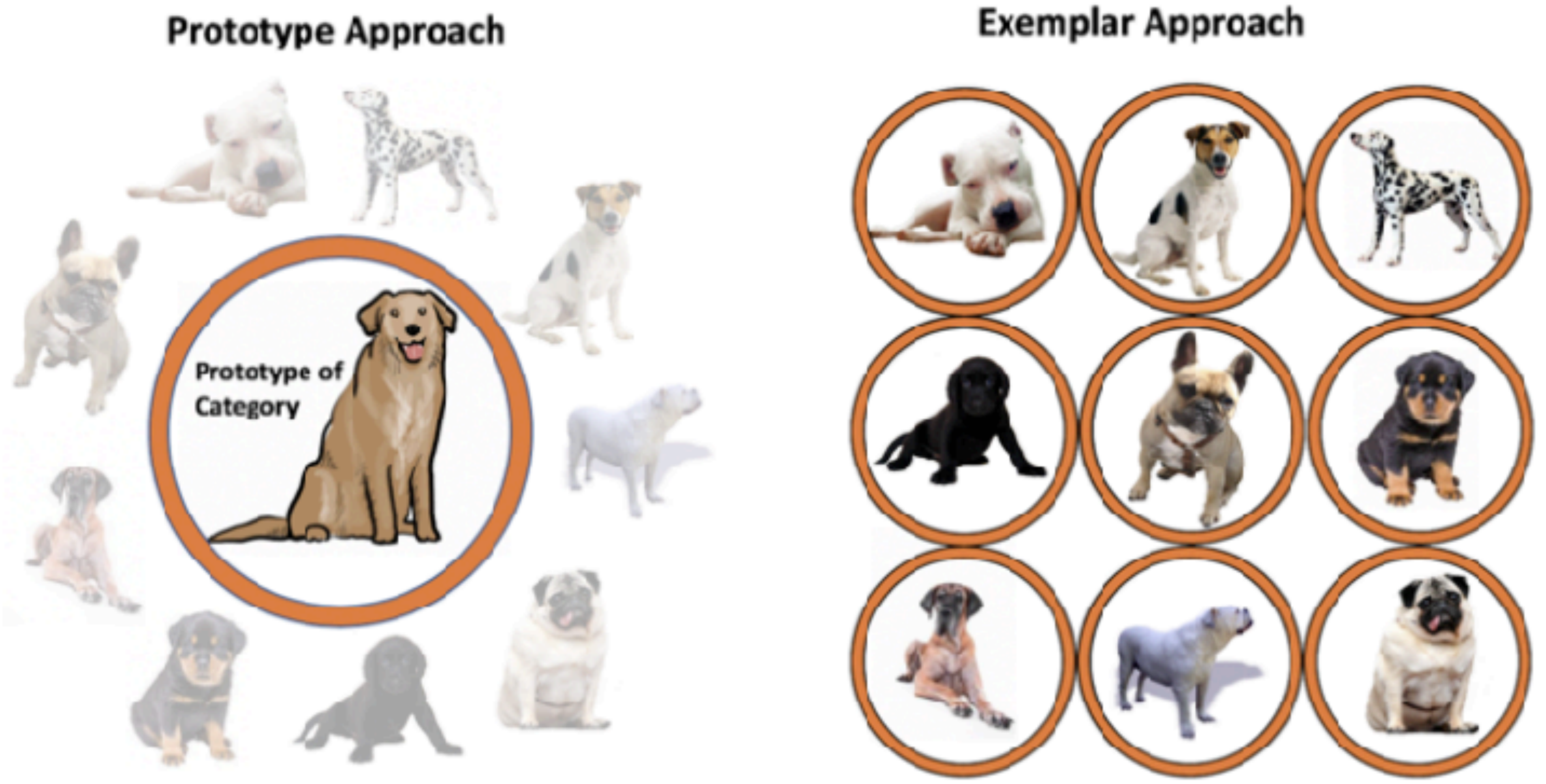


Armstrong, Gleitman, & Gleitman (1983)

Similarity-based theories

- Rather than hard and fast rules, perhaps we use **similarity comparisons** to make on the fly generalizations about new objects
- **Similarity theories**: Stimuli with similar features are more likely to belong to the same category
    - Distance in feature space provides a simple quantification of similarity
- Two main camps:
    - Category membership based on comparison to previously learned *prototypes* or *exemplars*





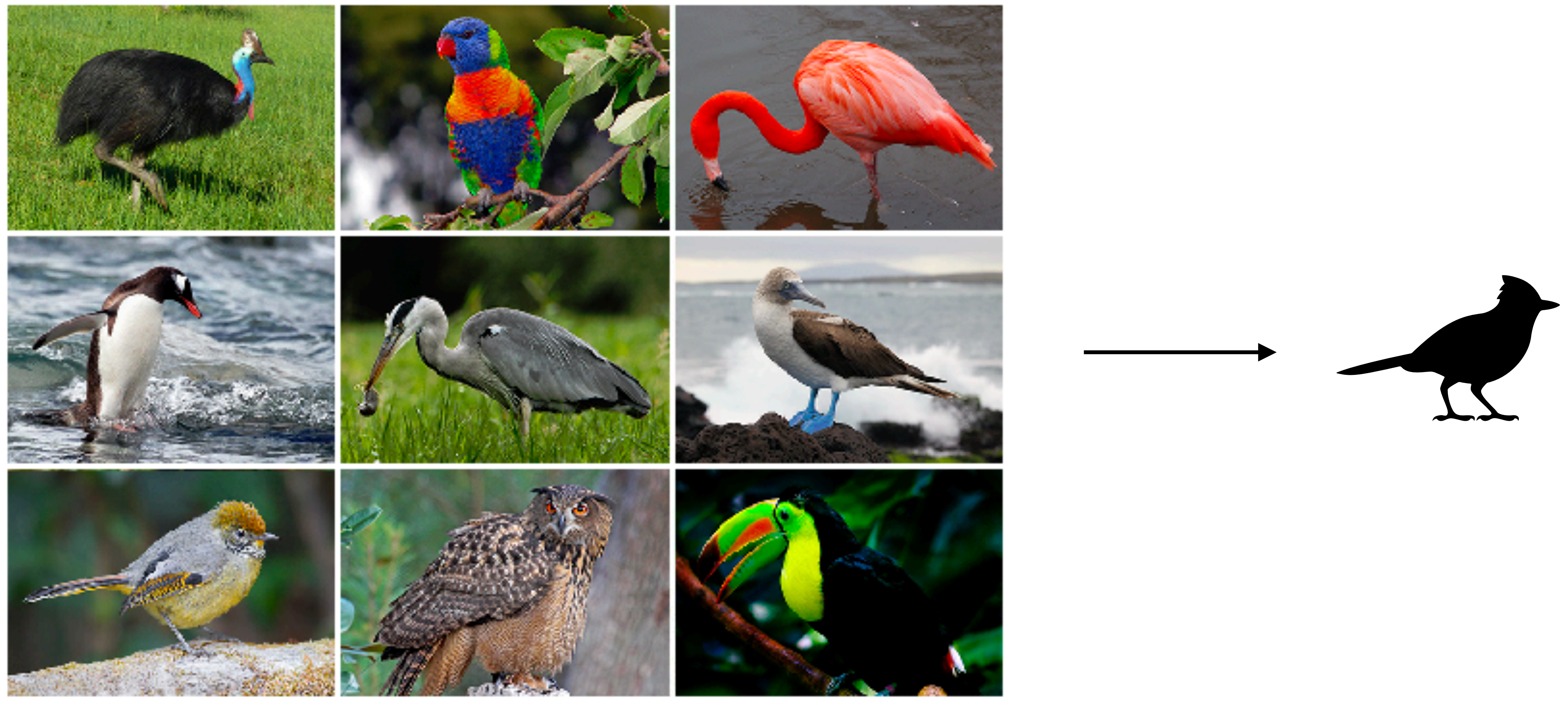
Levering & Kurtz (2019)

Prototype theory

- Prototypes are *summary representations* of a category (Rosch, 1973)
  - Typicality can be explained by items being closer to our learned prototype
- Prototypes can be constructed based weighted features (Smith & Medin, 1981)
  - Some features are more important: Birds have wings (1.0), usually fly (0.8), some sing songs (0.3), and a few eat worms (0.1)
- Categories are thus defined by similarity to the prototype

Which is the most prototypical chair?

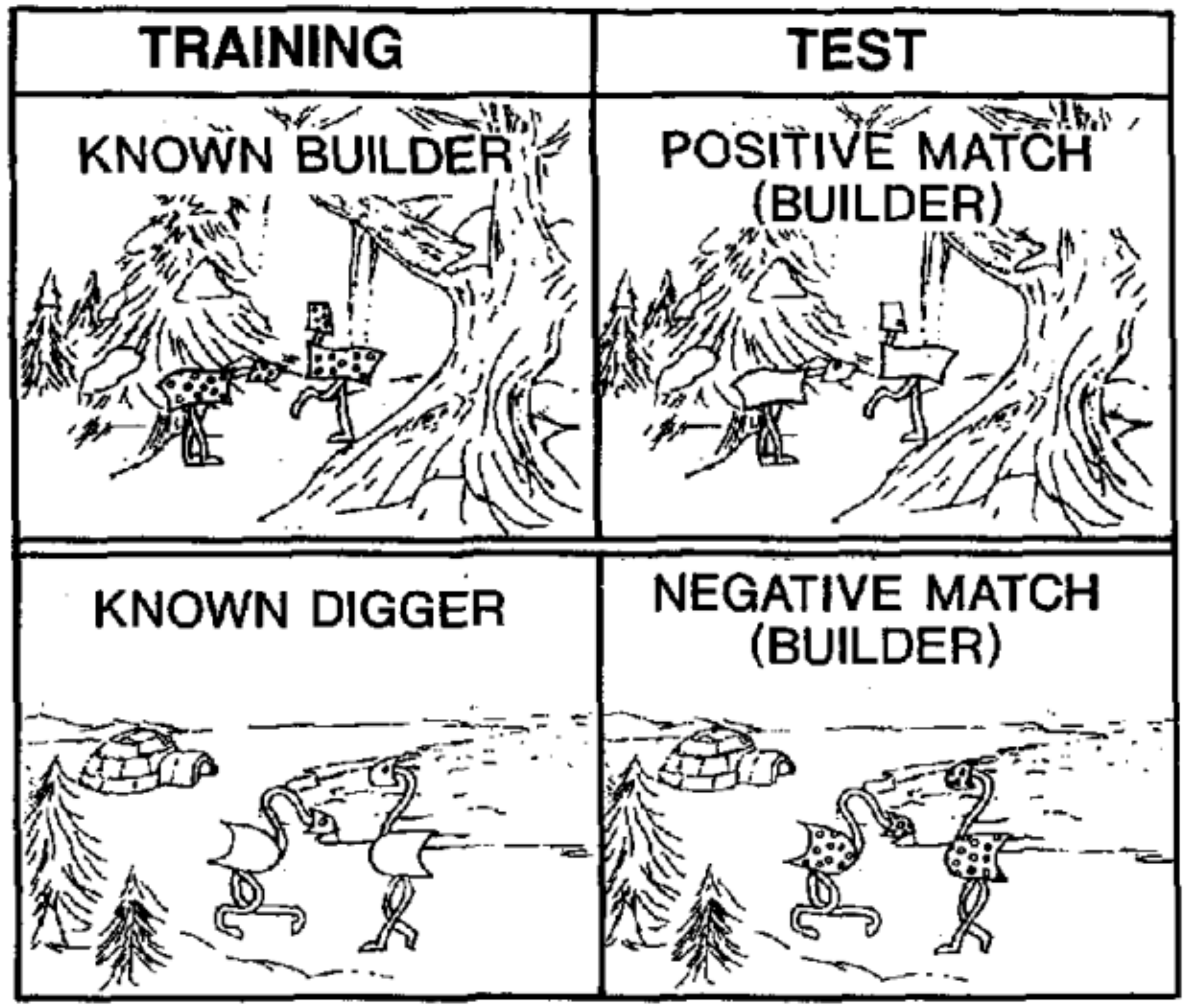
Constructing a prototype by weighing important features

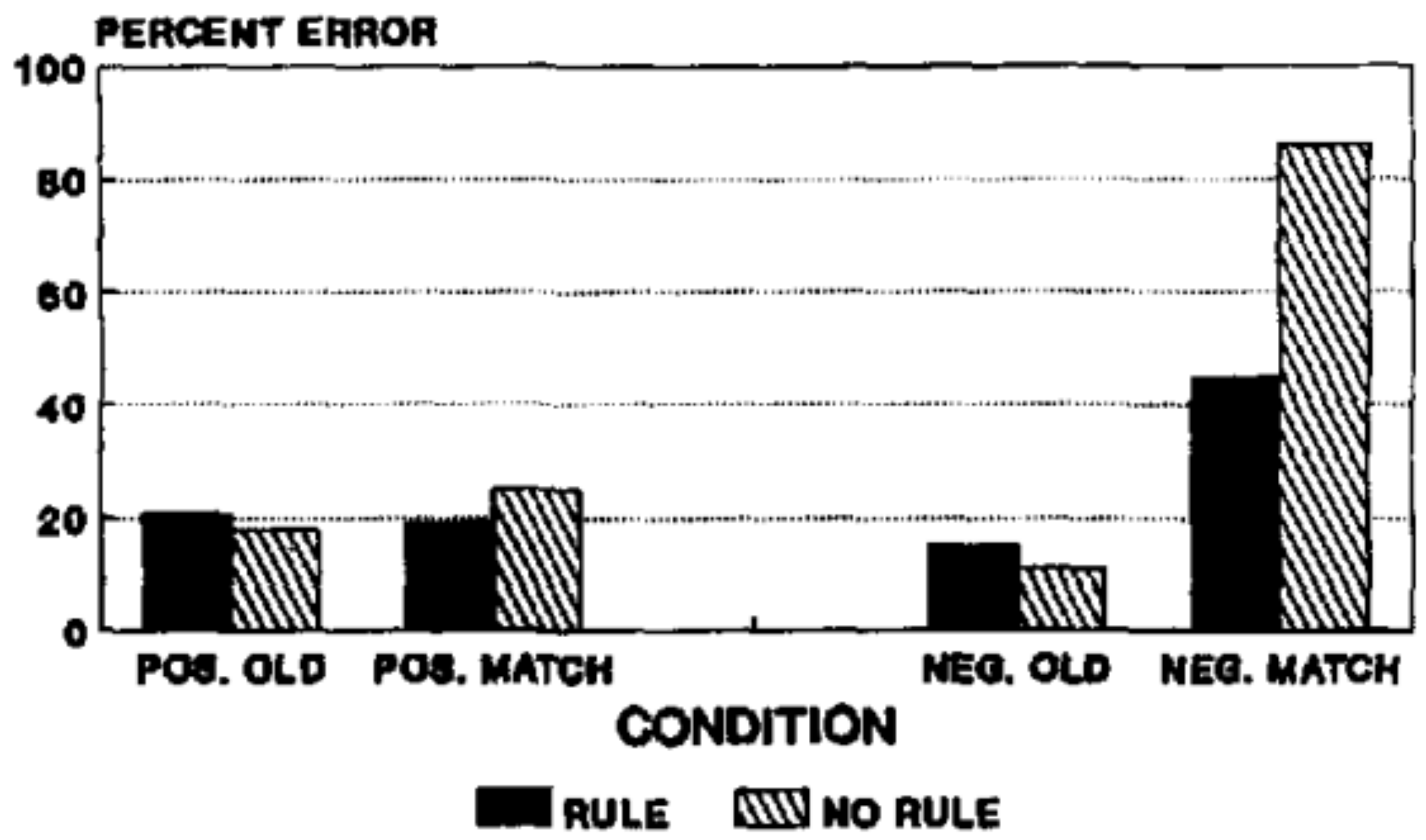


Exemplar theory

- No summary representation
  - We remember each exemplar (i.e., each instance) of a concept, and we compare new instances to these past memories (Medin & Schaffer, 1978)
- Close similarity to well-remembered stimuli has a strong effect on classification:
  - Participants were often fooled by the negative match (with spots), even when body and legs didn't match
  - Interpreted as evidence the dots from training exemplars had a large influence, even when the rule was explicitly told to participants
- Categories are thus defined by similarity to past exemplars

RULE: AT LEAST TWO OF (LONG LEGS, ANGULAR BODY, SPOTS) ⟶ BUILDER

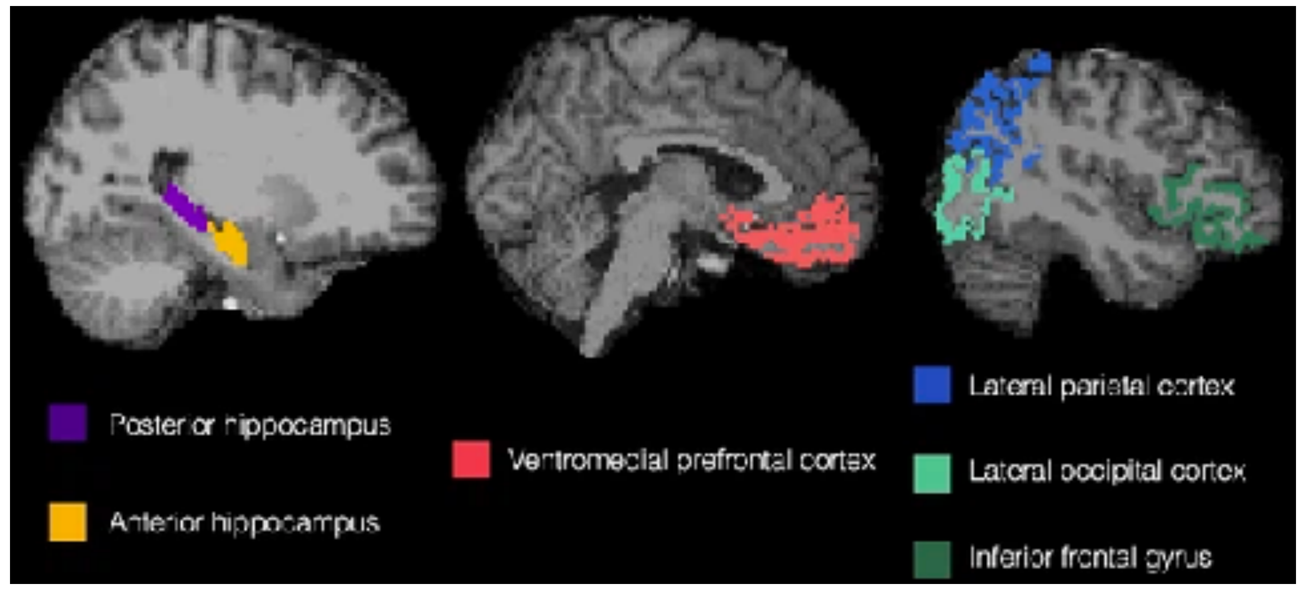


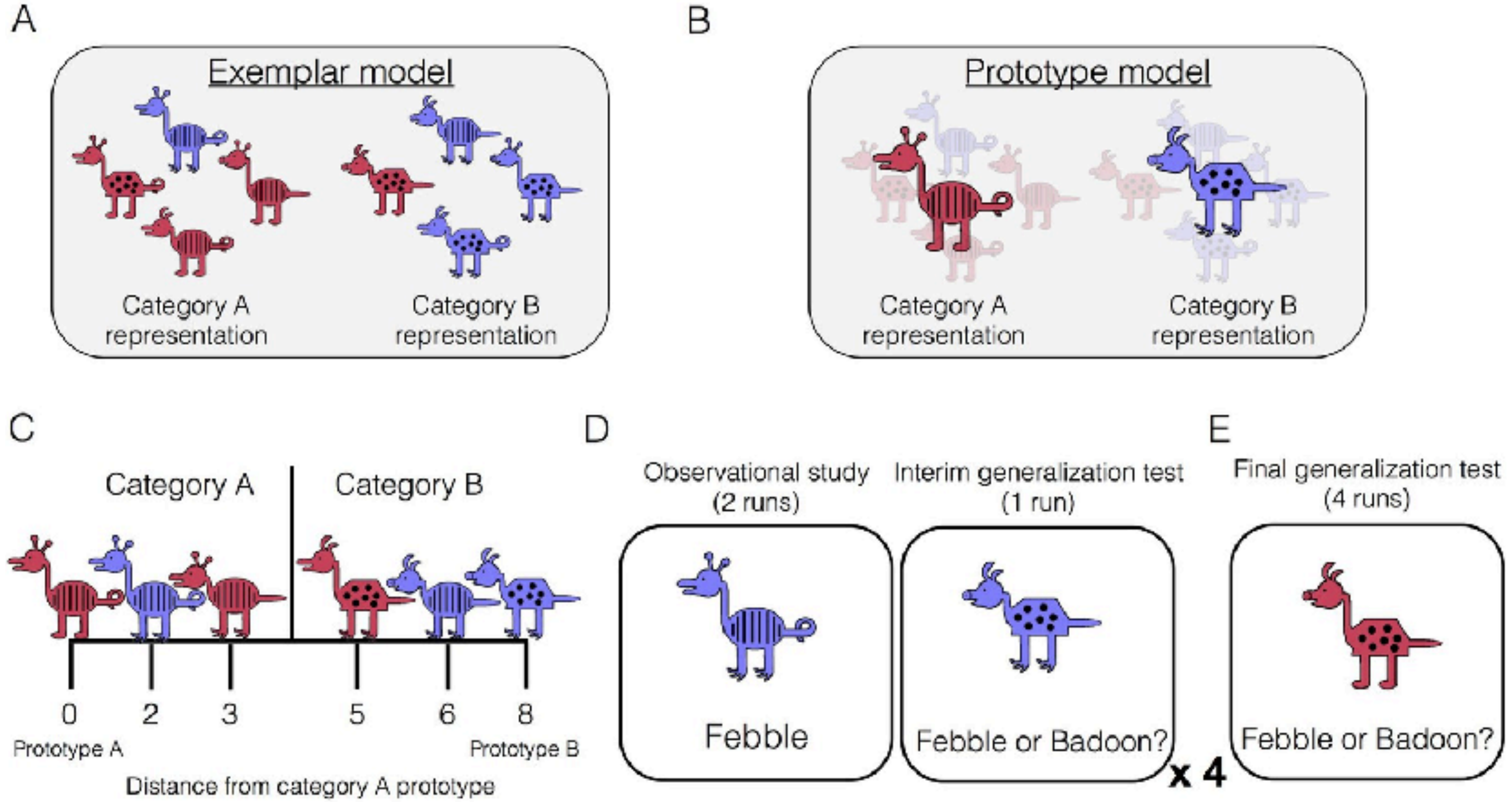


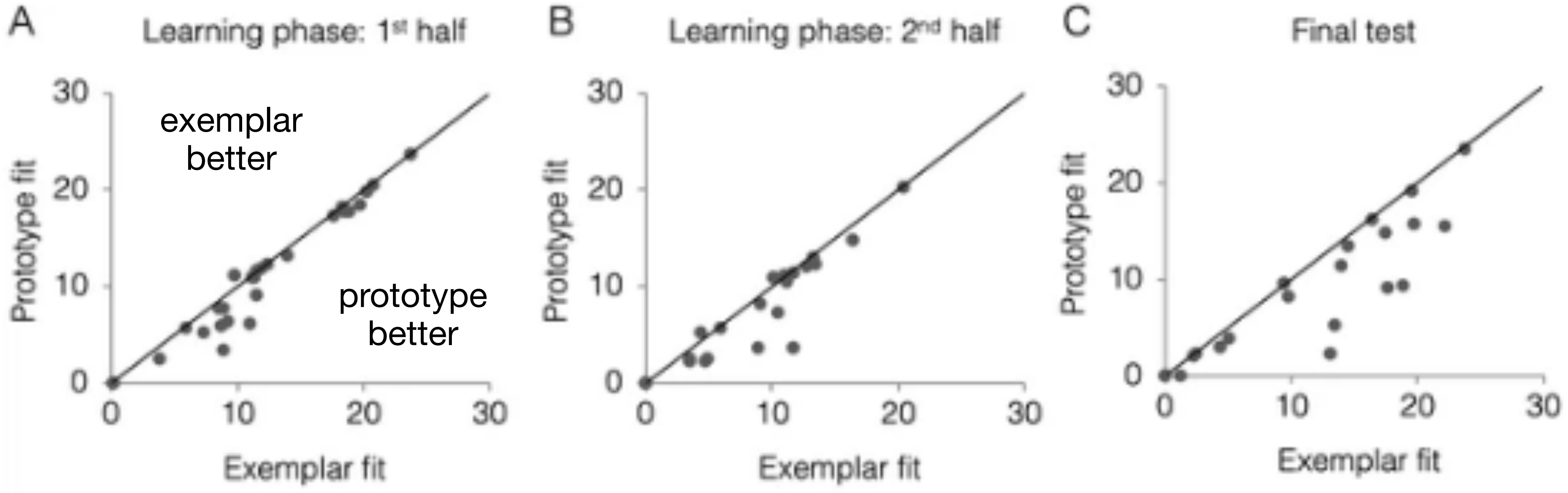
Allen & Brooks (1991)

Prototype or exemplar?

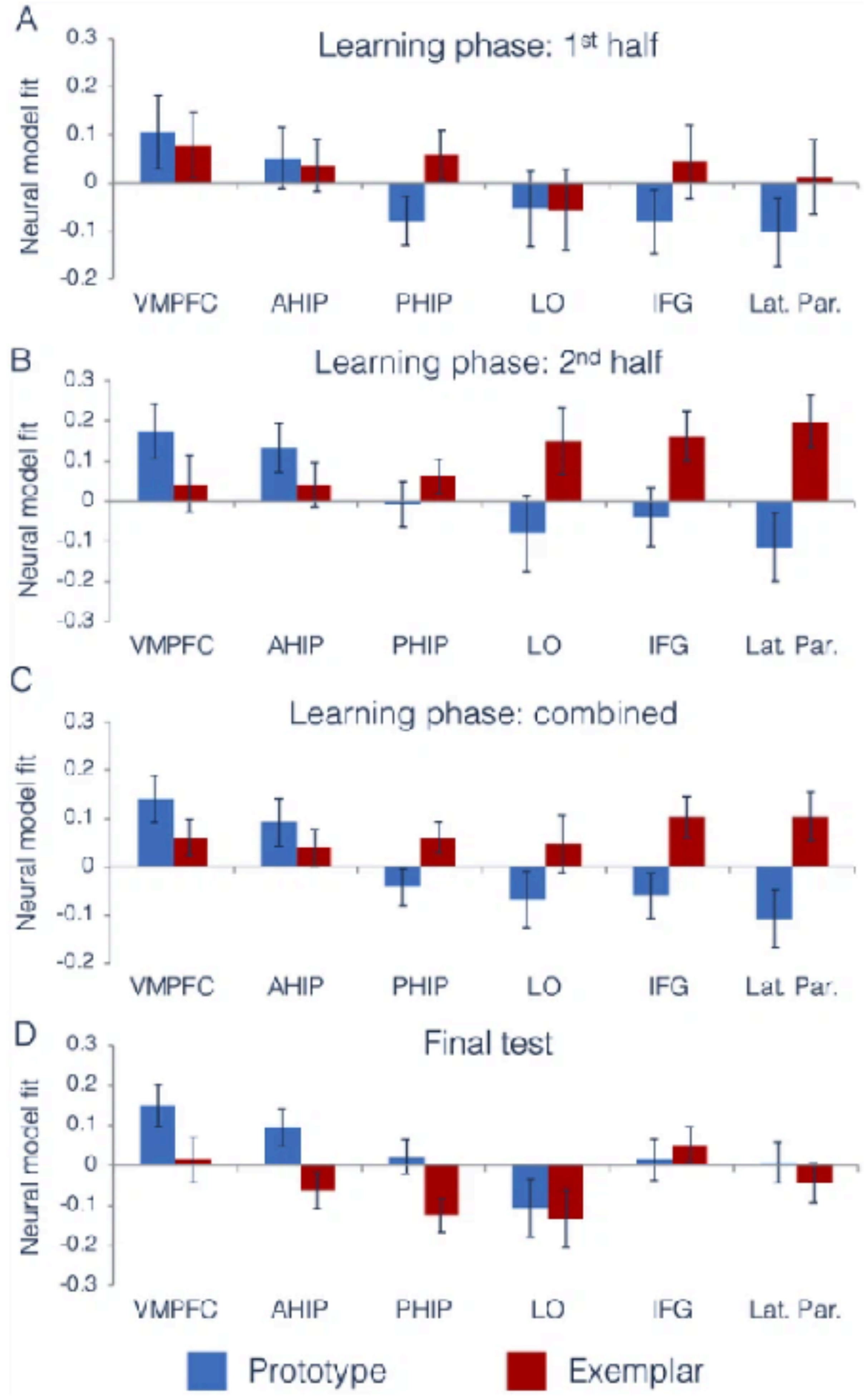
- Still an open debate
- Prototype was dominant during the final test
- But neural signatures of both throughout







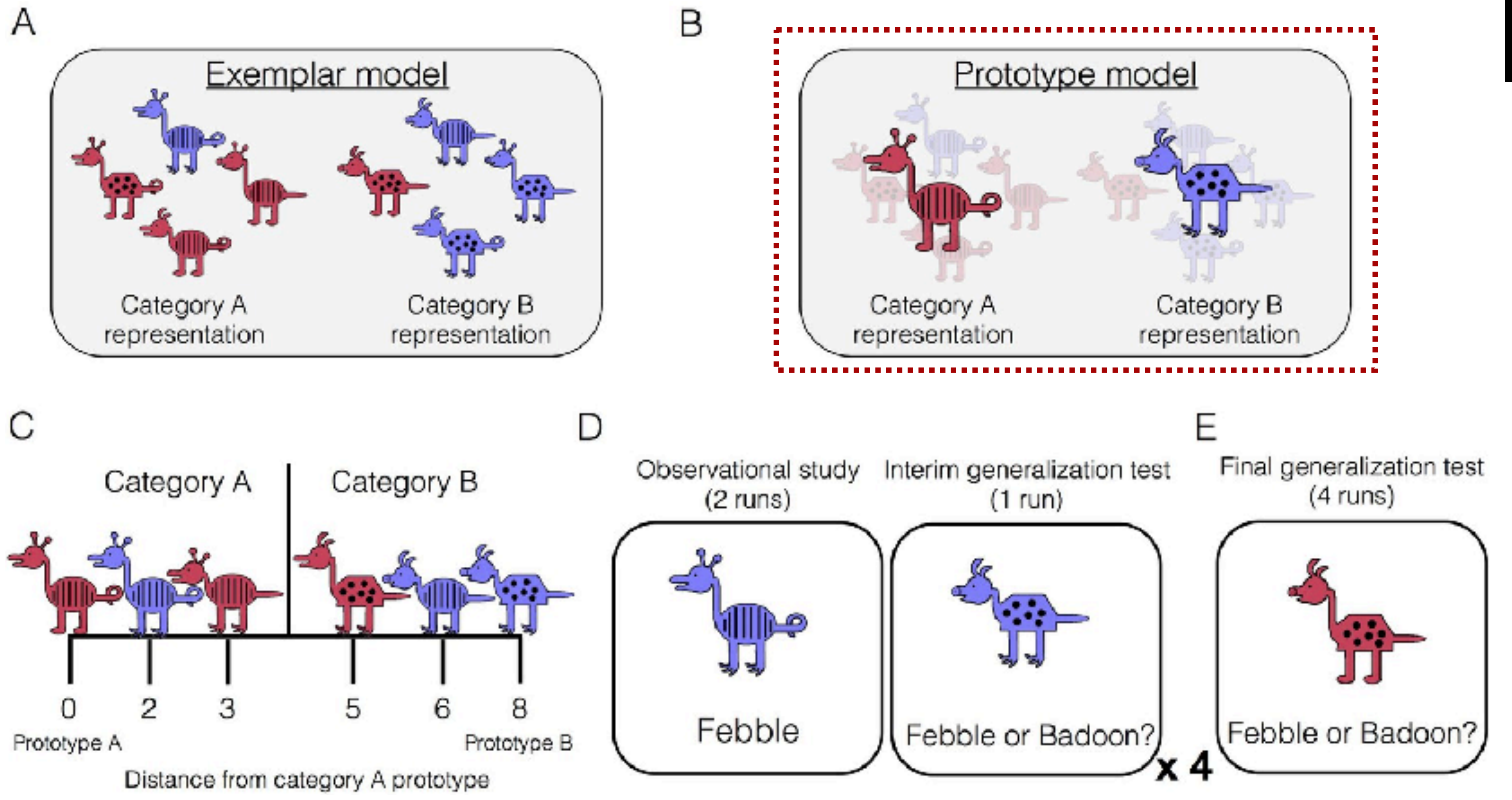
*lower is better

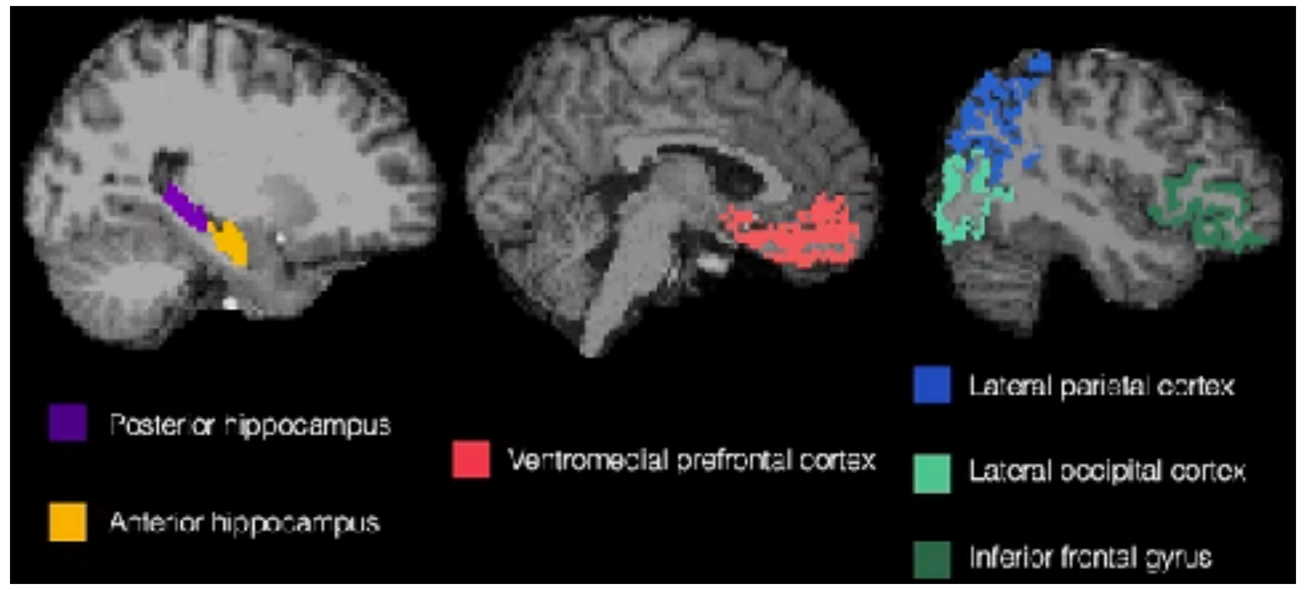


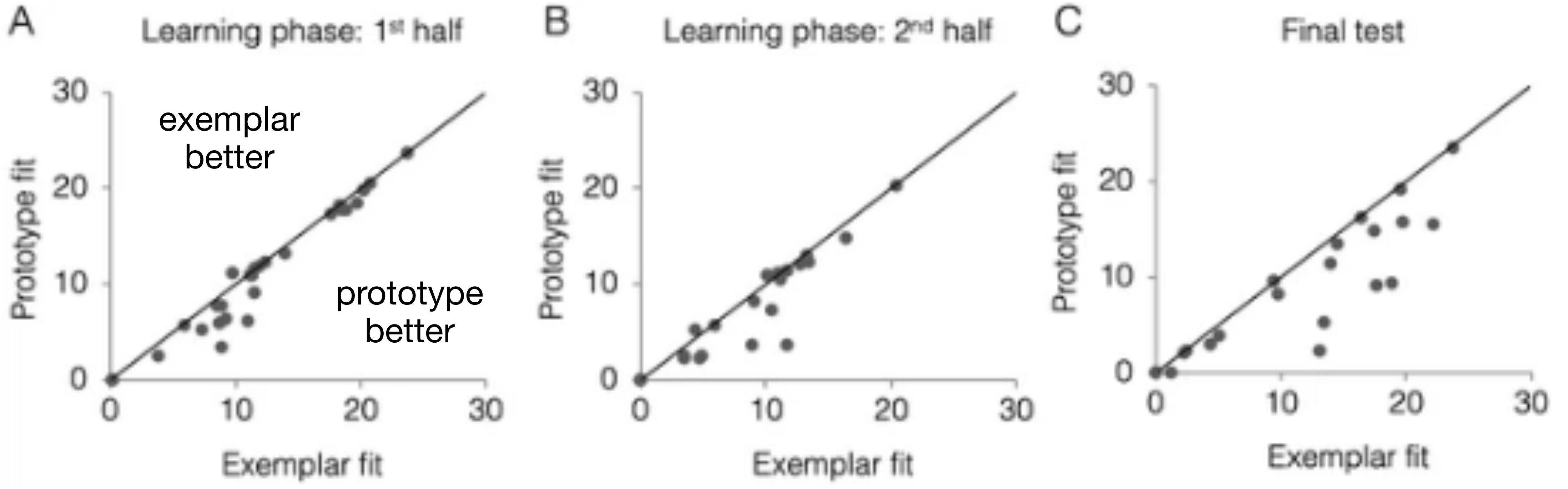
Bowman, Iwashita, & Zeithamova (2020)

Prototype or exemplar?

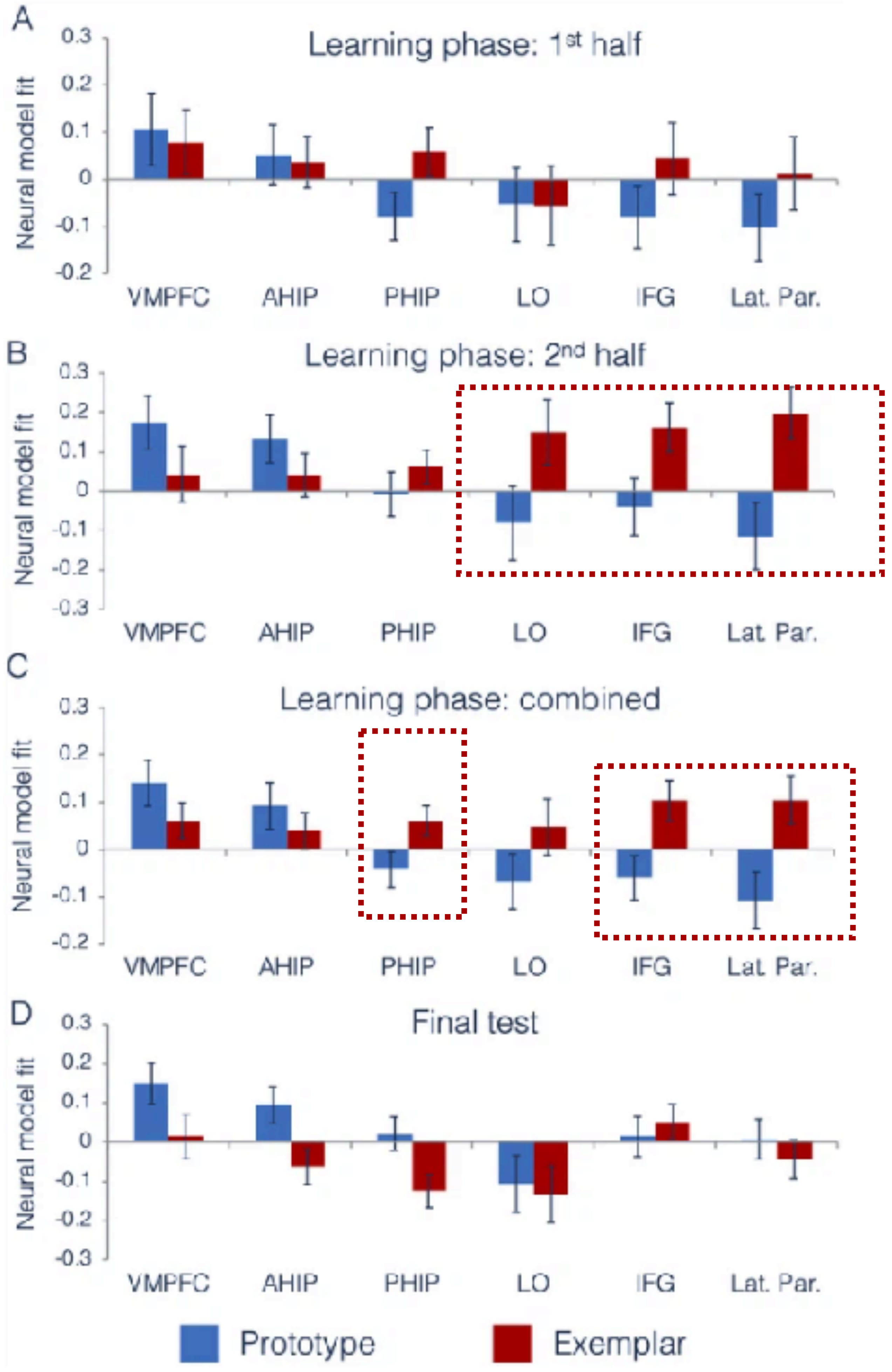
- Still an open debate
- Prototype was dominant during the final test
- But neural signatures of both throughout







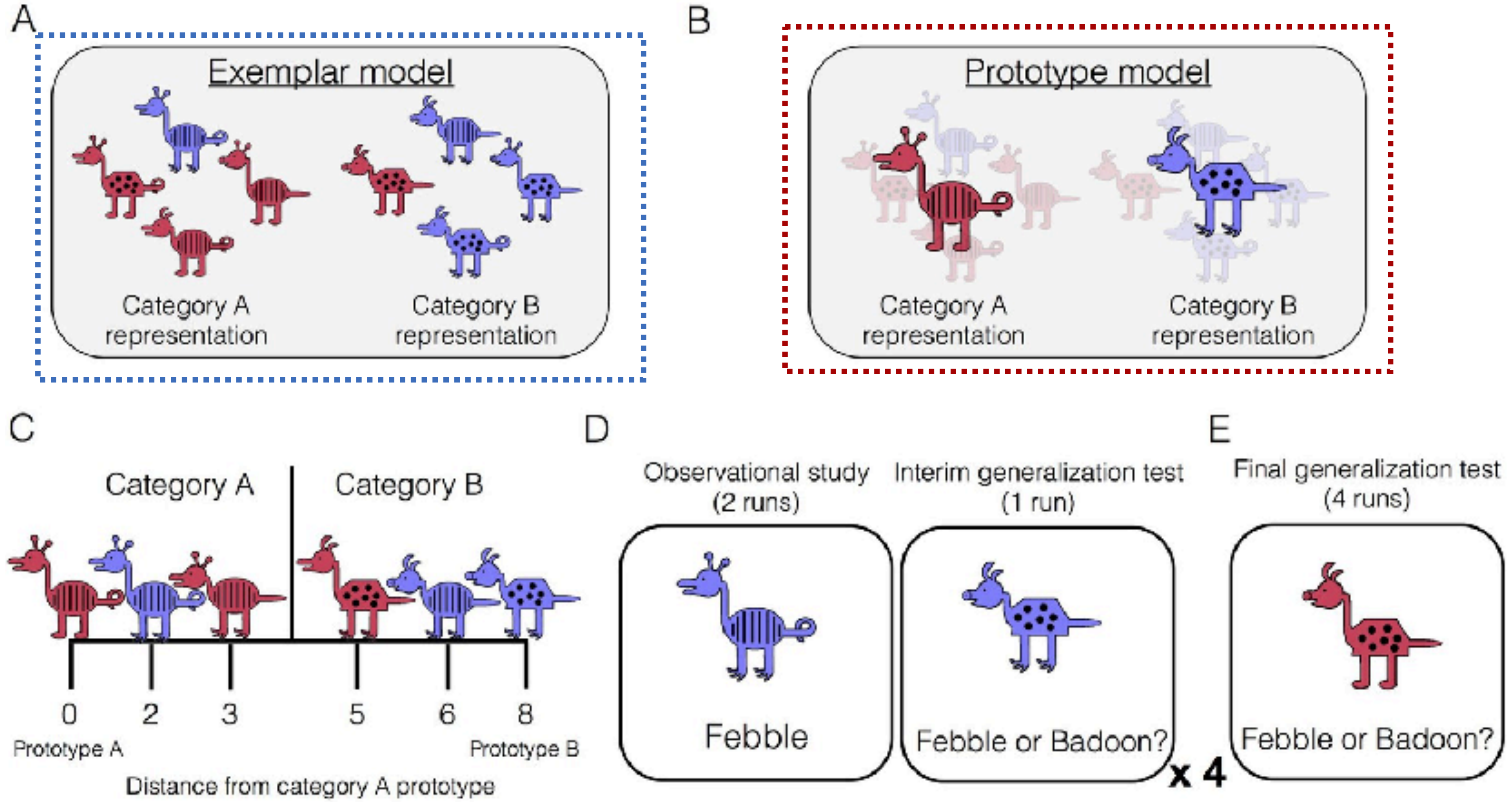
*lower is better

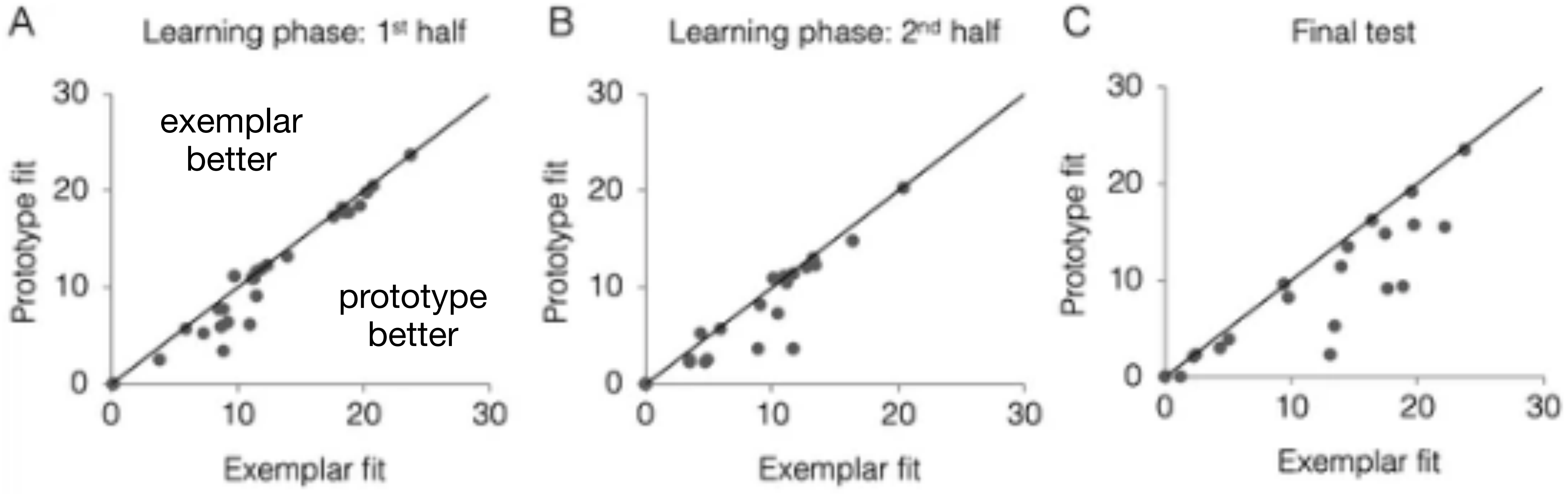


Bowman, Iwashita, & Zeithamova (2020)

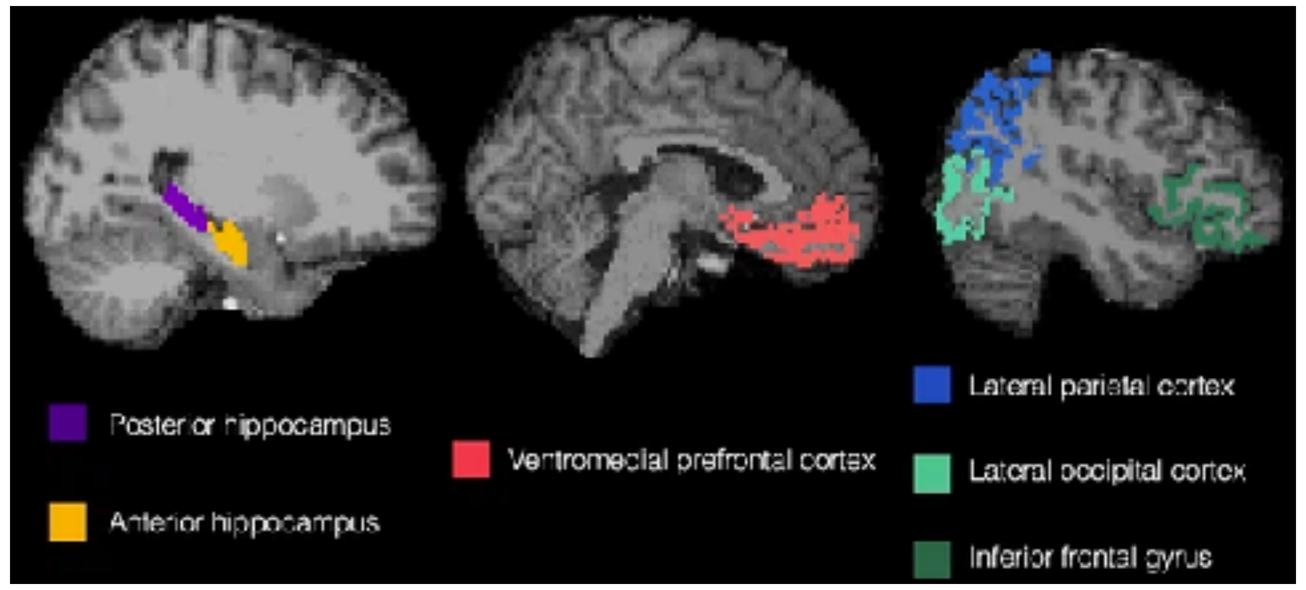
Prototype or exemplar?

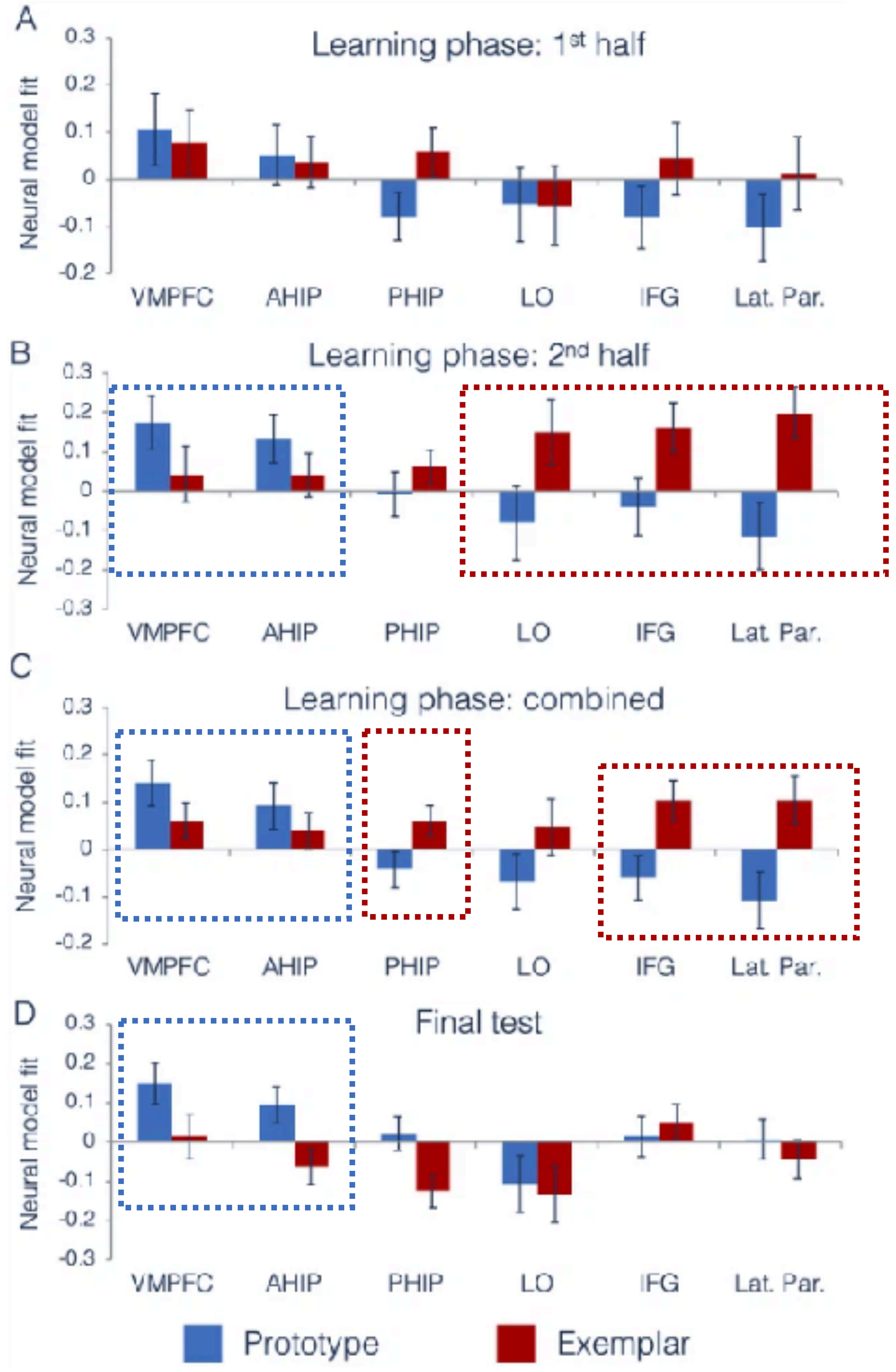
- Still an open debate
- Prototype was dominant during the final test
- But neural signatures of both throughout





*lower is better

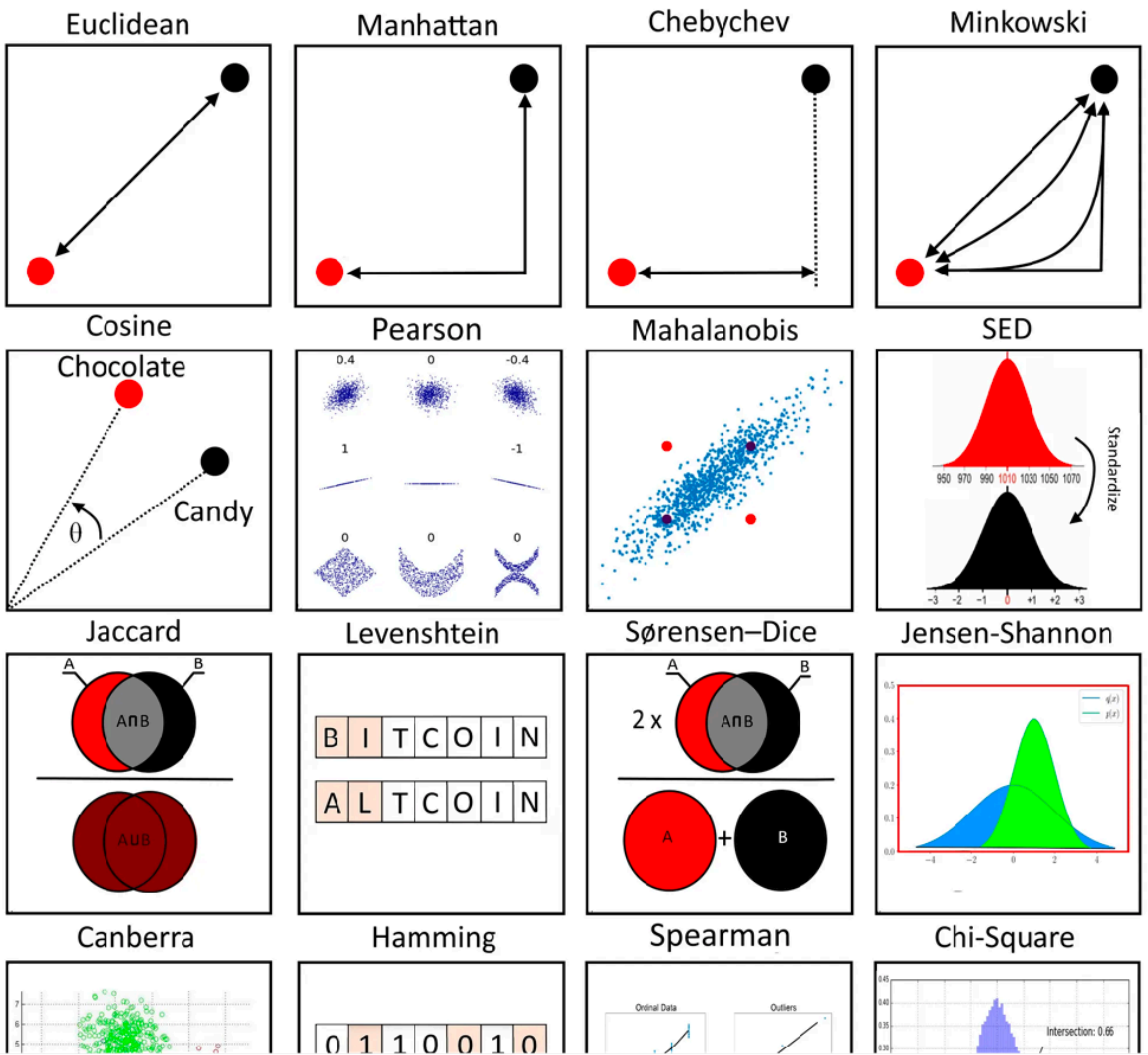




Bowman, Iwashita, & Zeithamova (2020)

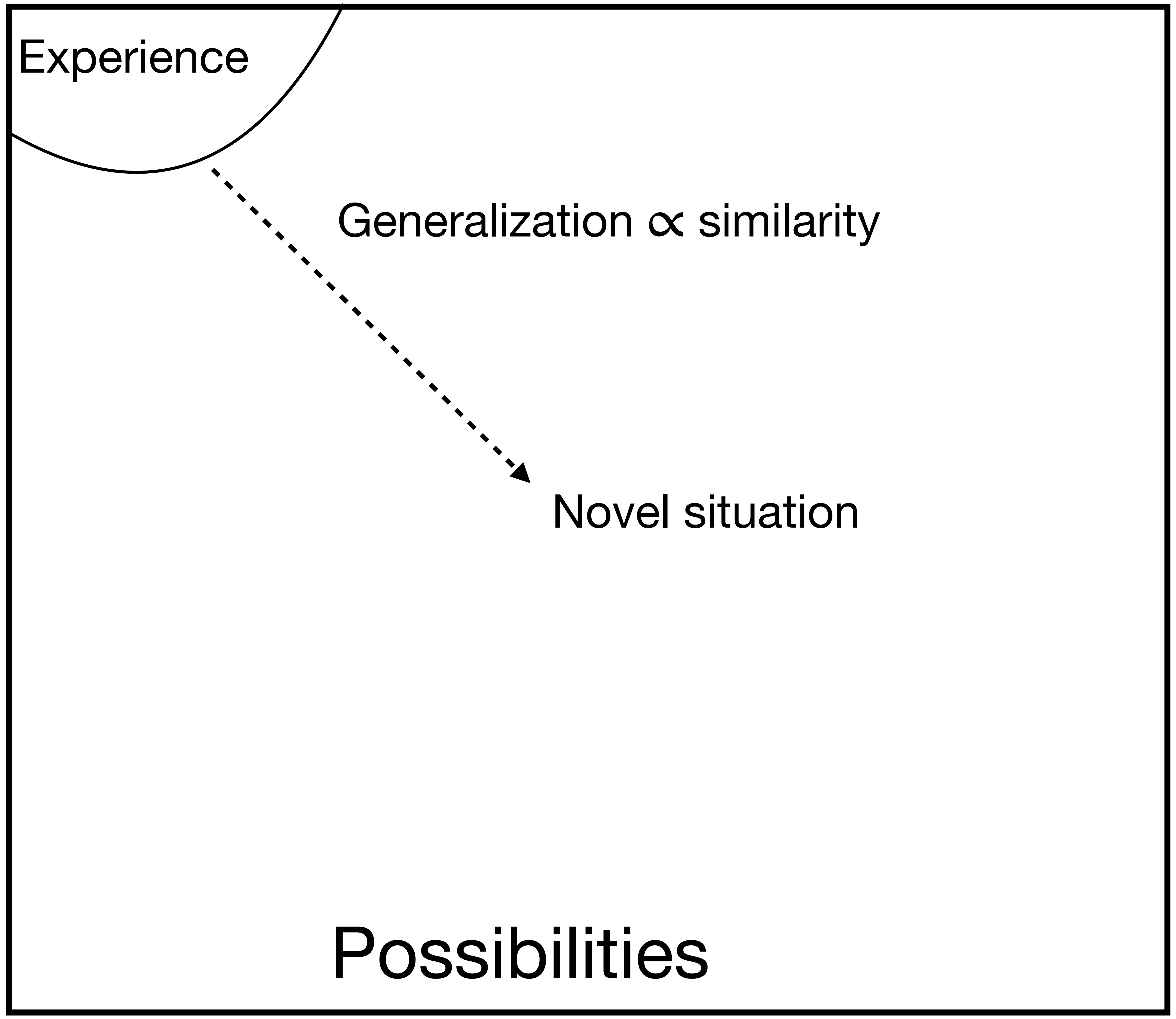
5 min break

How do we define similarity?



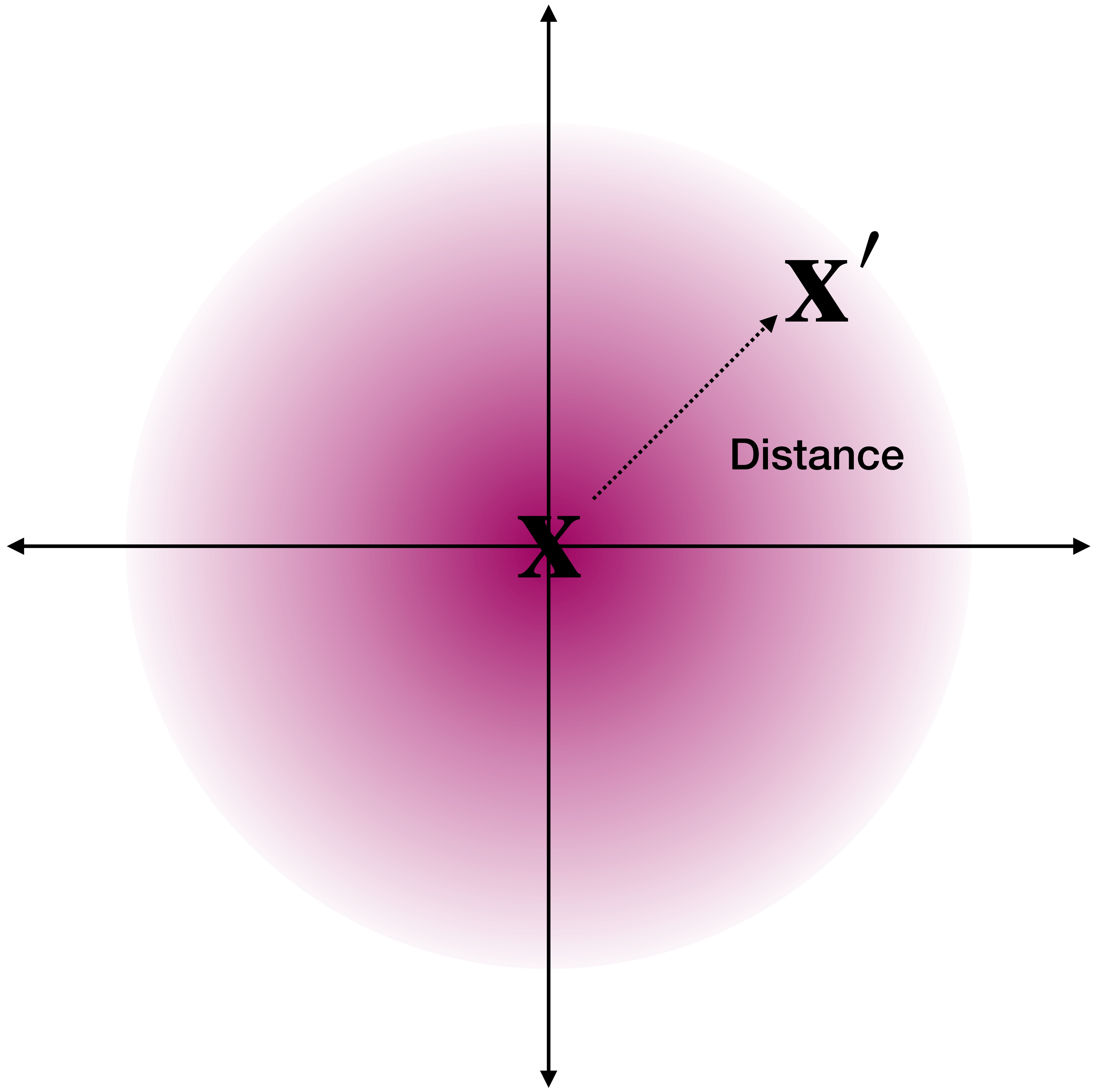
Generalization as a method to test different forms of similarity

- How do we generalize limited experience to novel situations?
  - The degree of generalization should be a function of our latent similarity computations
- The best similarity metric for predicting generalization should also reveal something about how we represent concepts



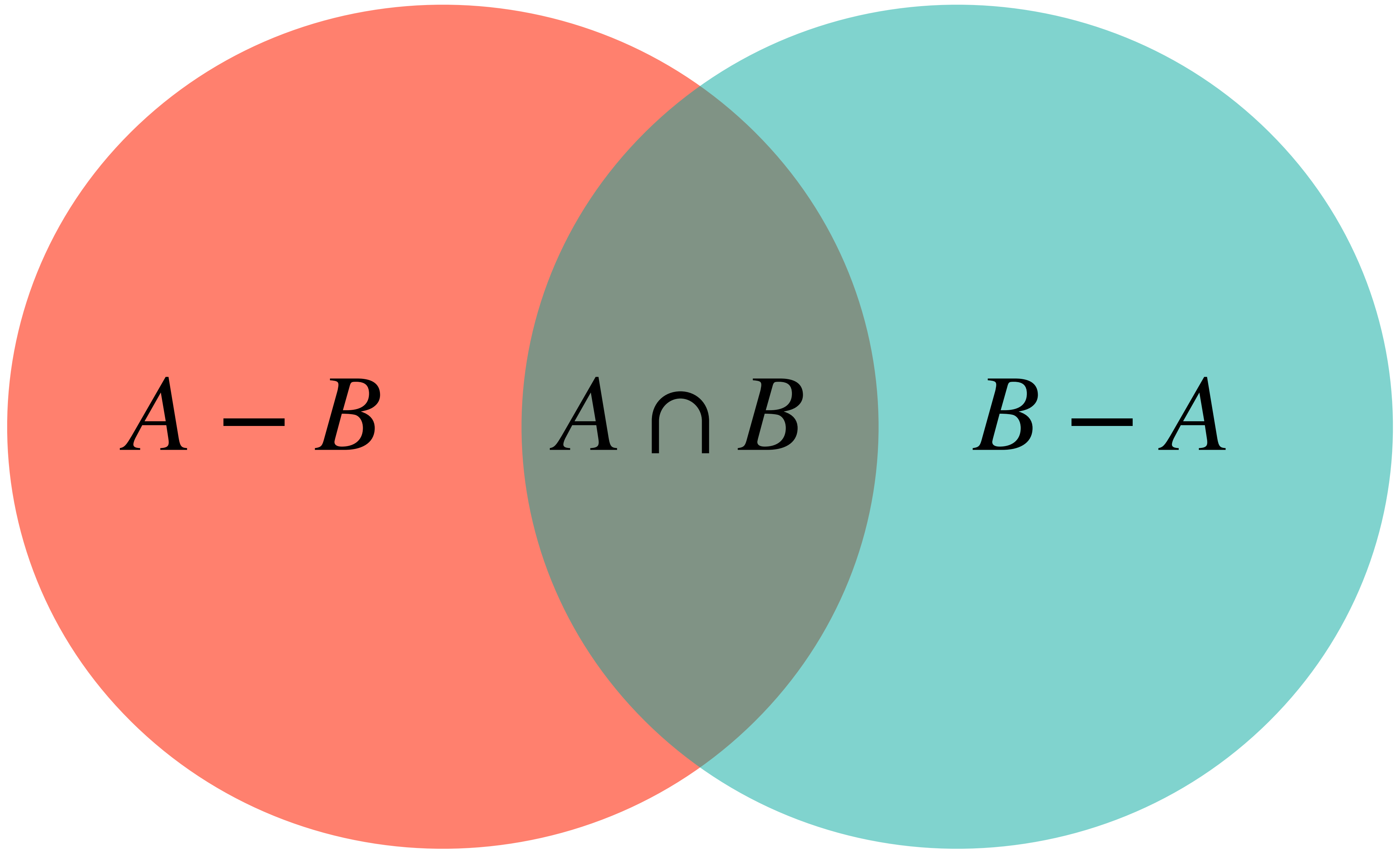
Two main approaches: Metric vs. Set

Metric



Embed data in some vector space and compute similarity as the inverse of distance

Set



Compare which features are jointly shared vs. unique (i.e., disjoint)

Shepard's (1987) Law of Generalization

STIMULUS + RESPONSE = **LEARNING**

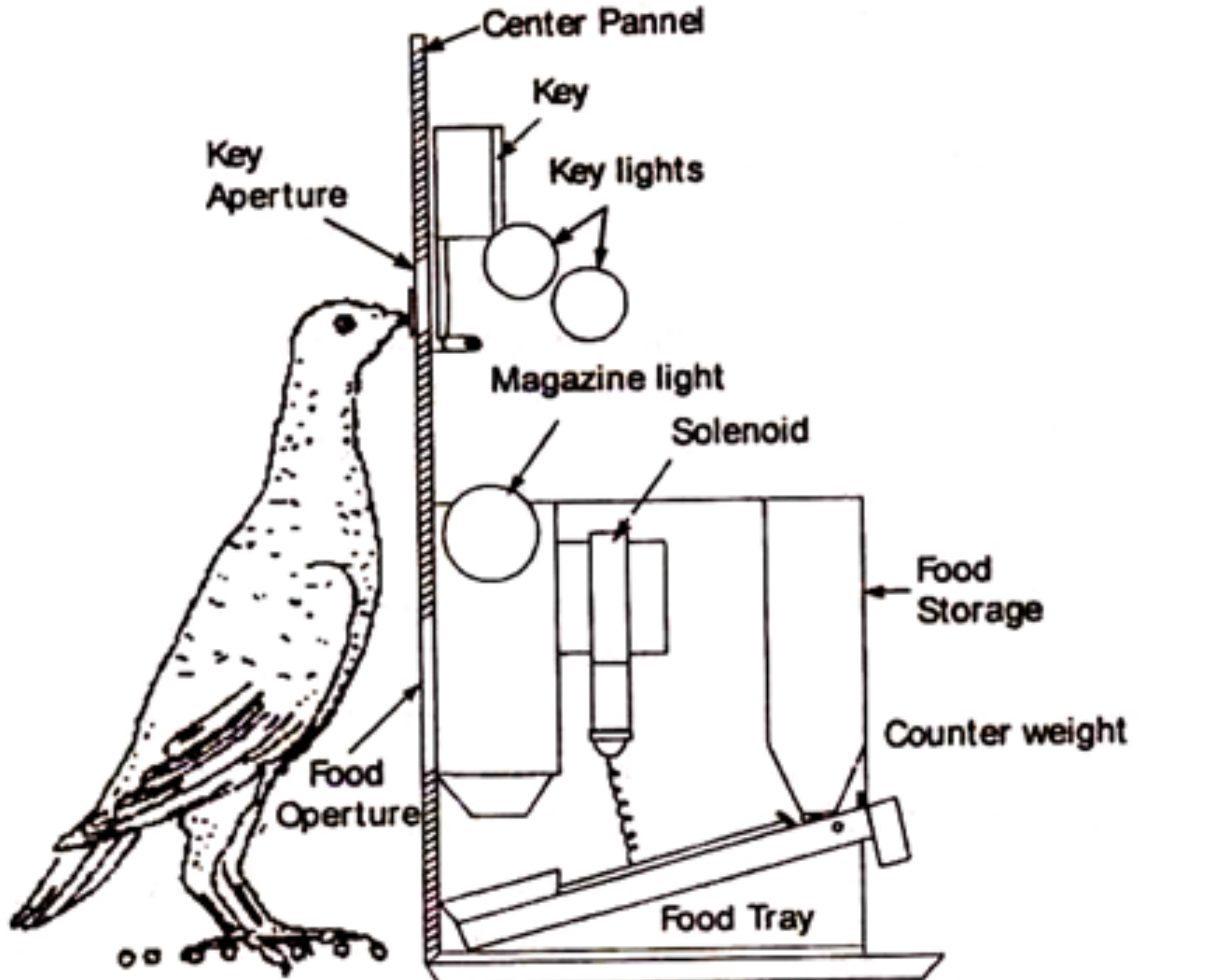
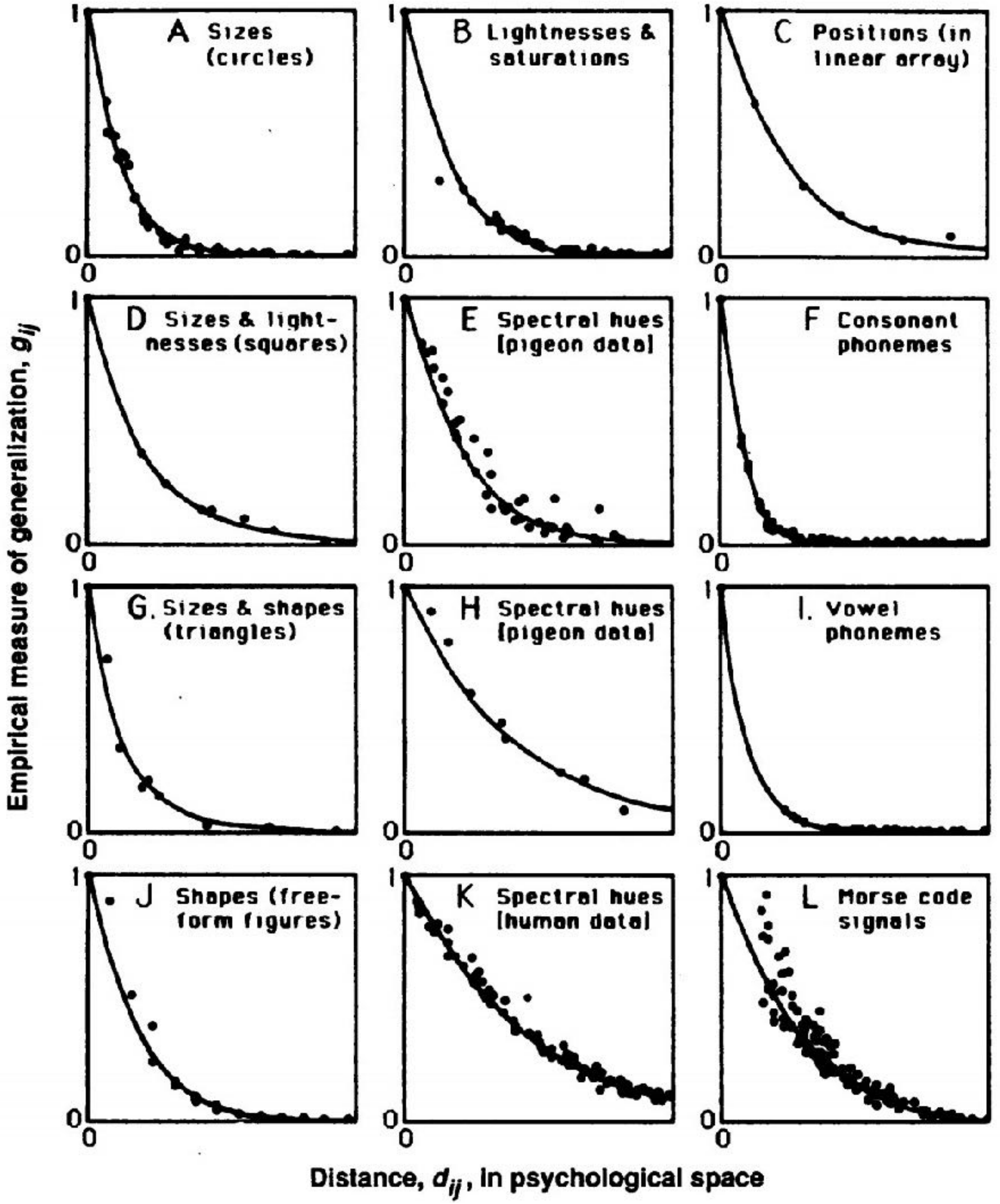


Illustration. Skinner box as adapted for the pigeon.



Shepard's (1987) Law of Generalization

STIMULUS + RESPONSE = **LEARNING**

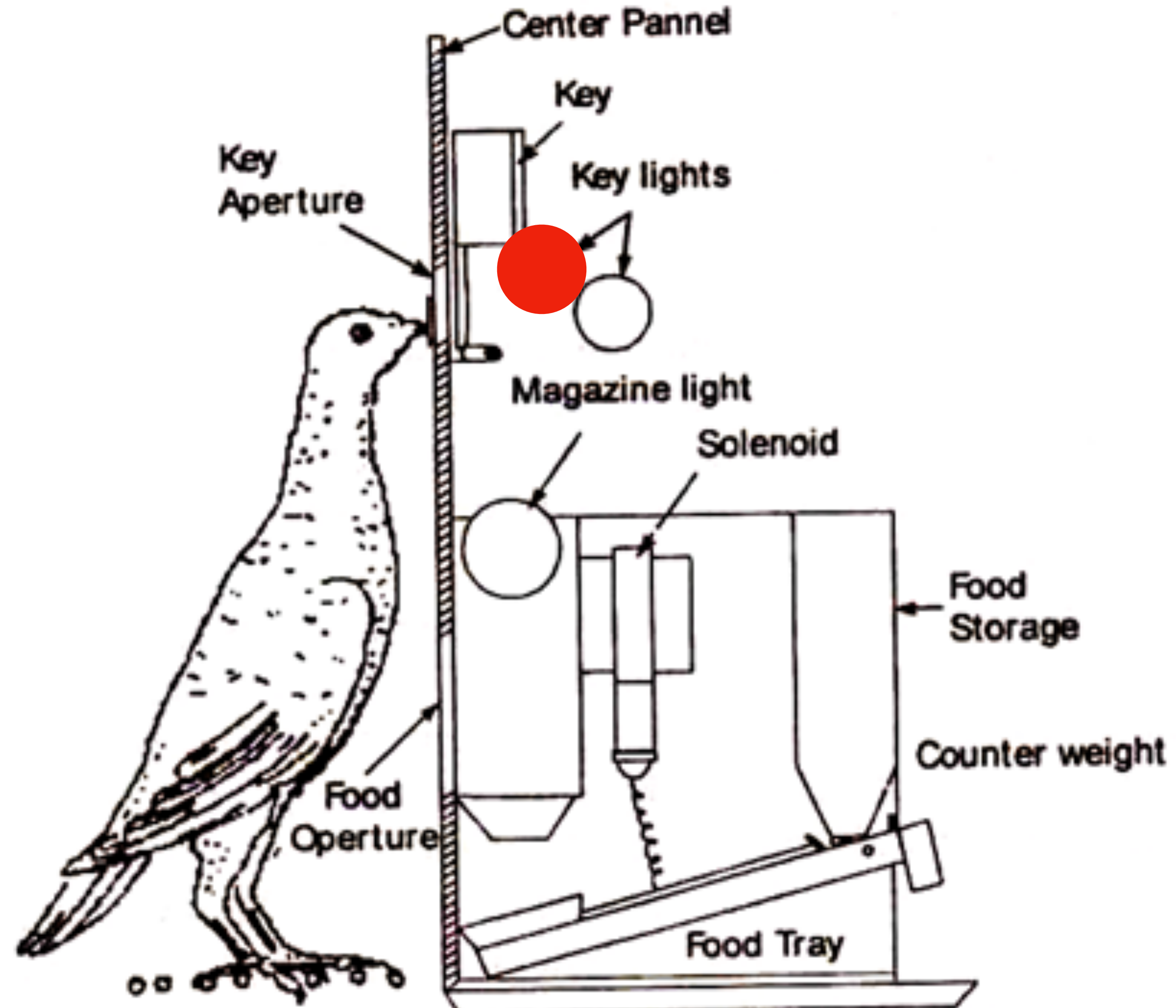
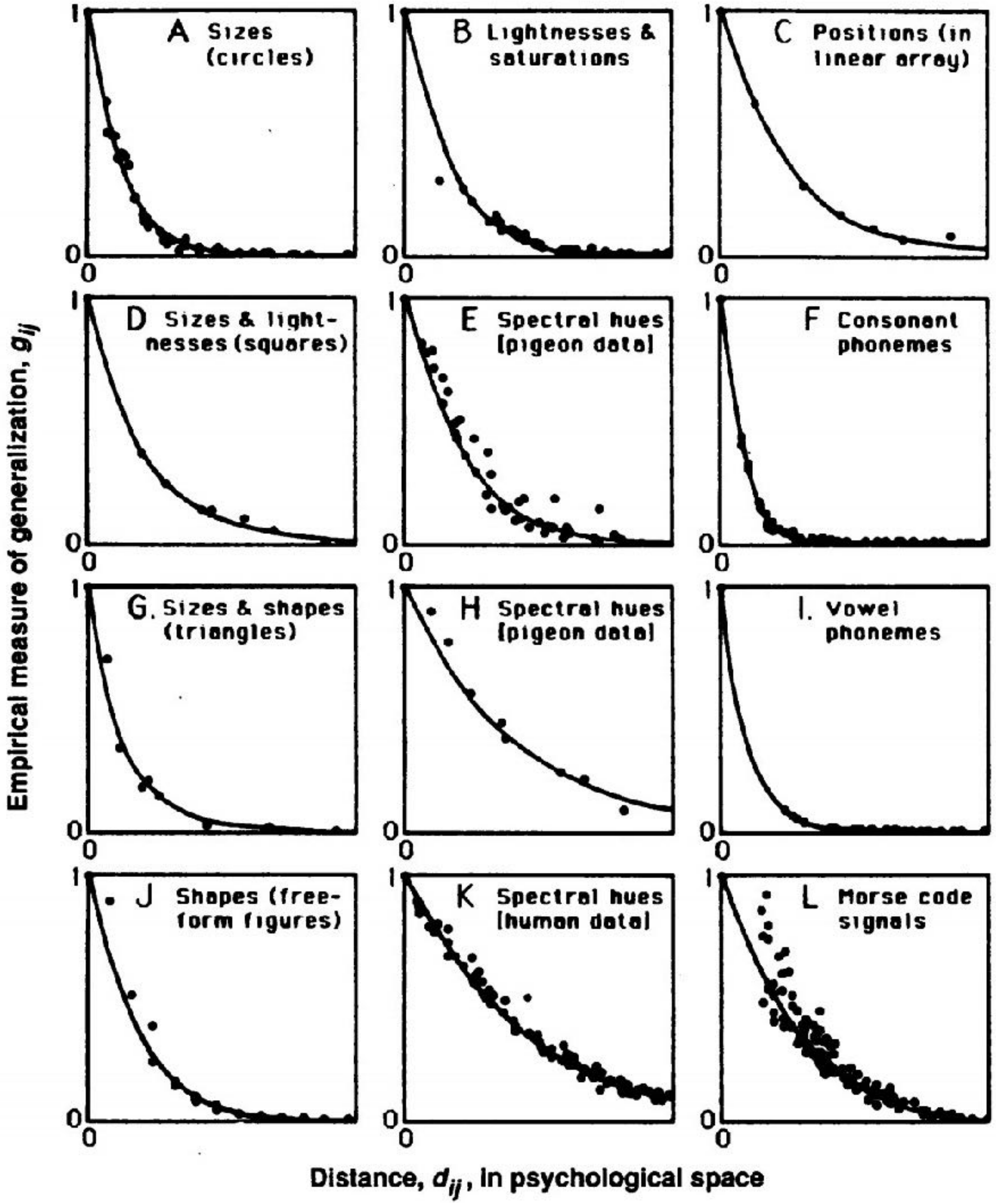
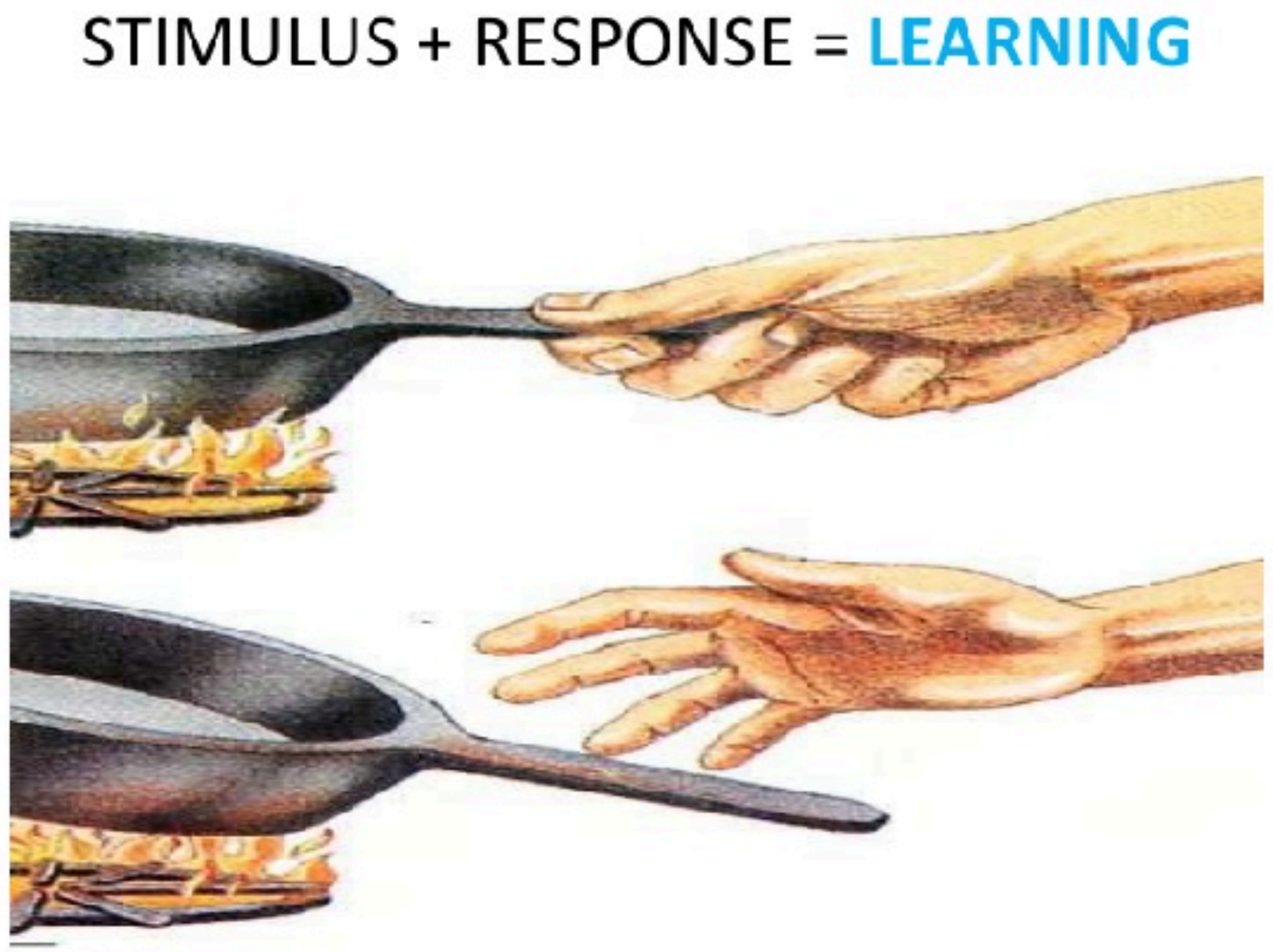


Illustration. Skinner box as adapted for the pigeon.



Shepard's (1987) Law of Generalization



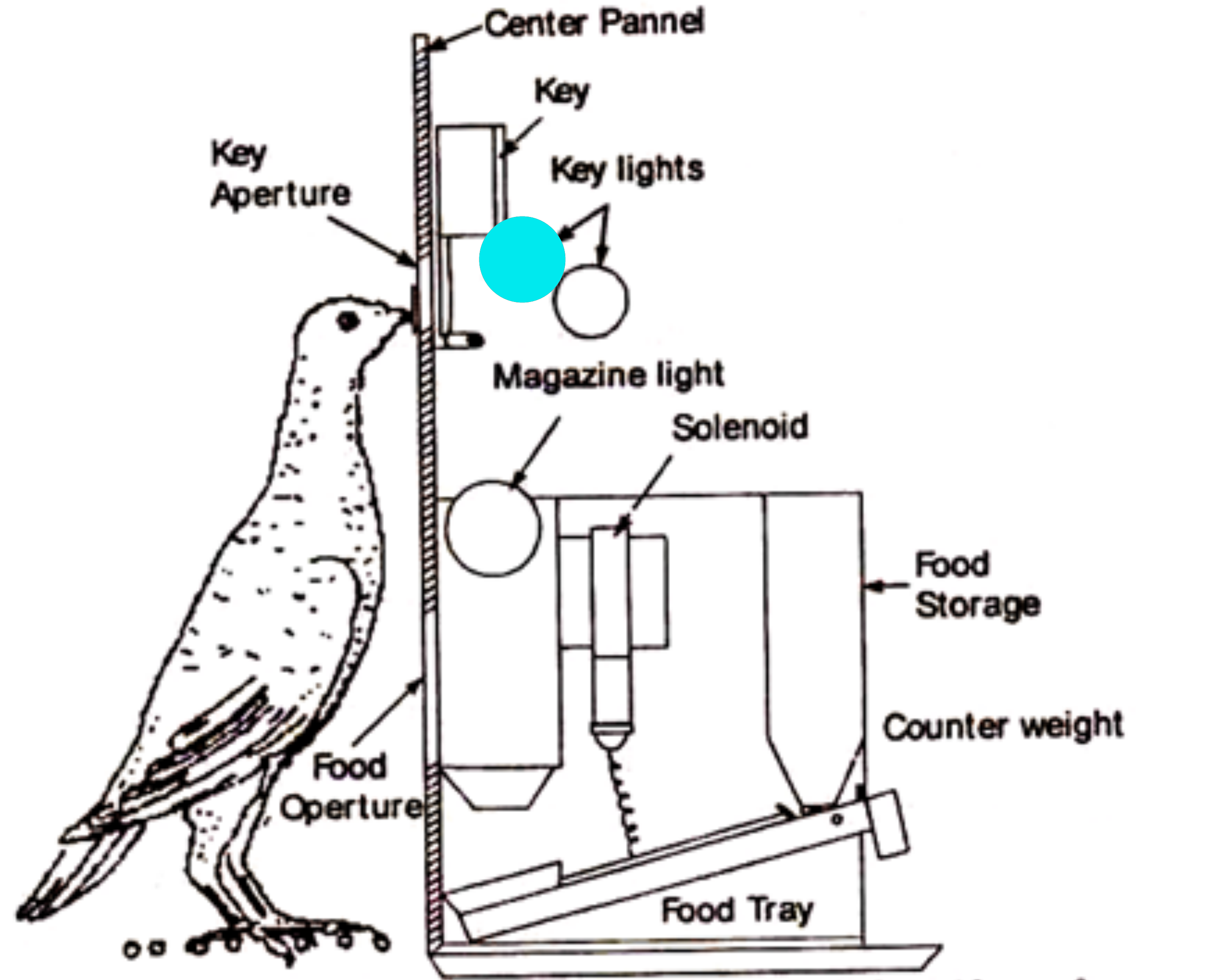
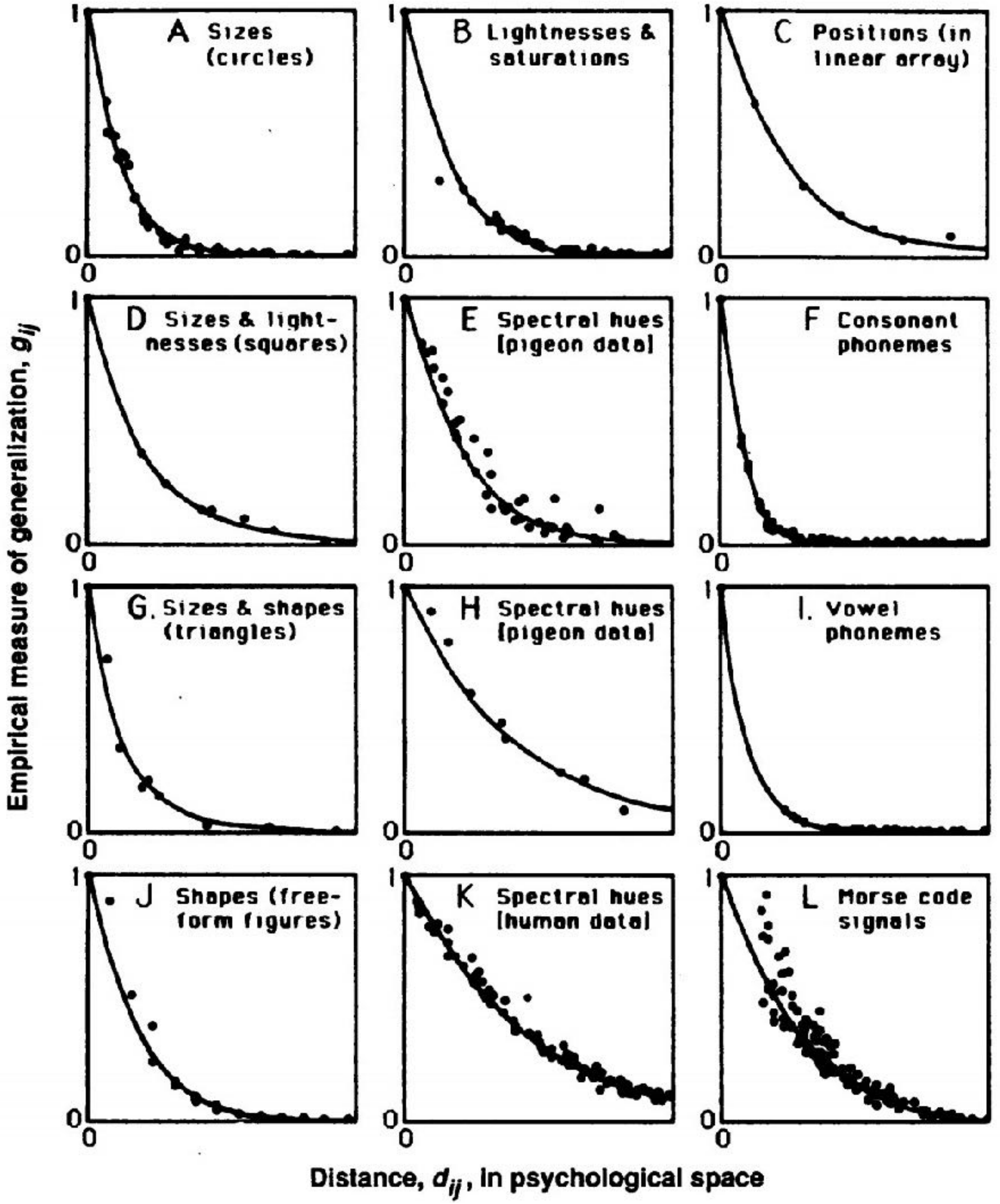


Illustration. Skinner box as adapted for the pigeon.



Shepard's (1987) Law of Generalization

STIMULUS + RESPONSE = **LEARNING**

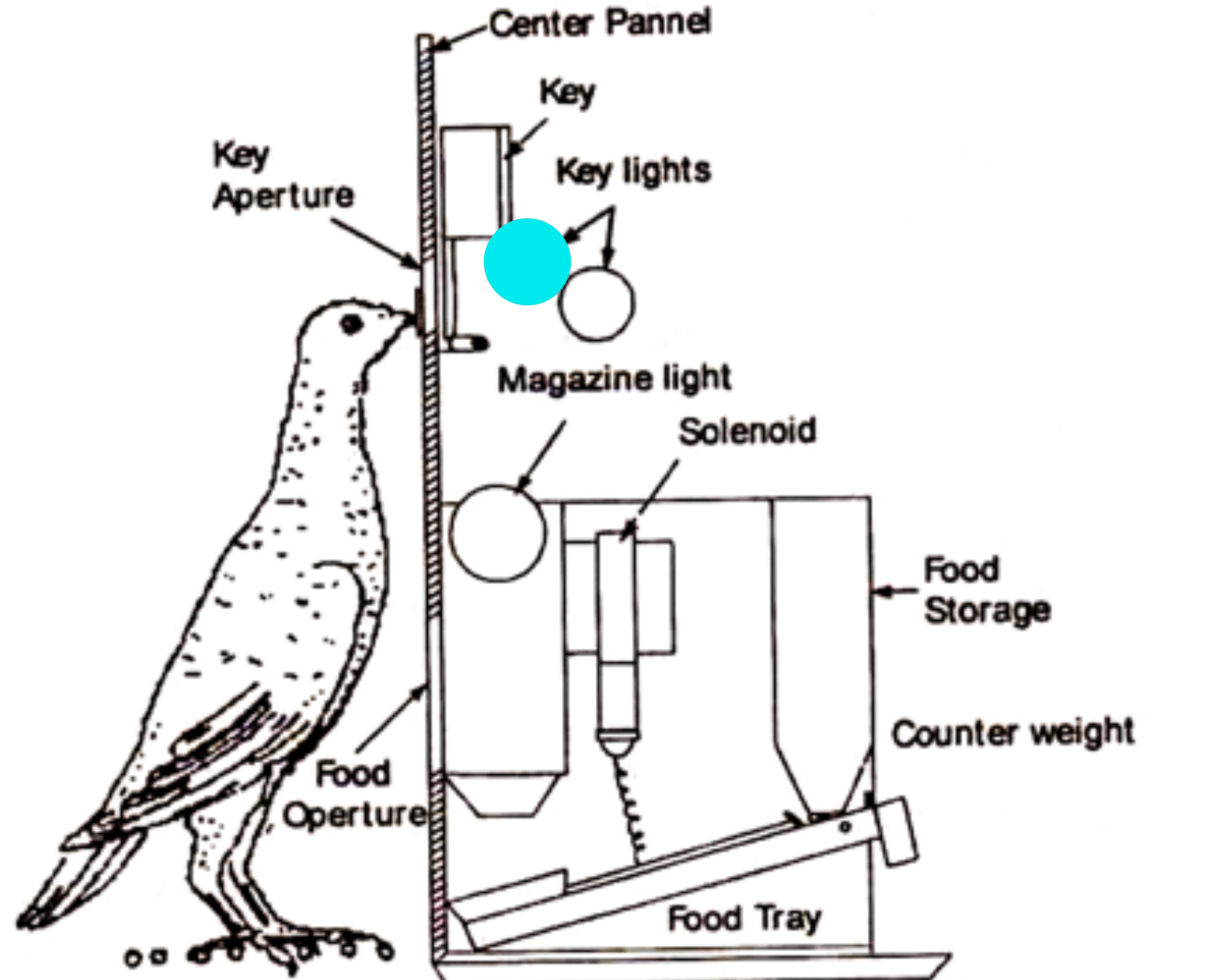
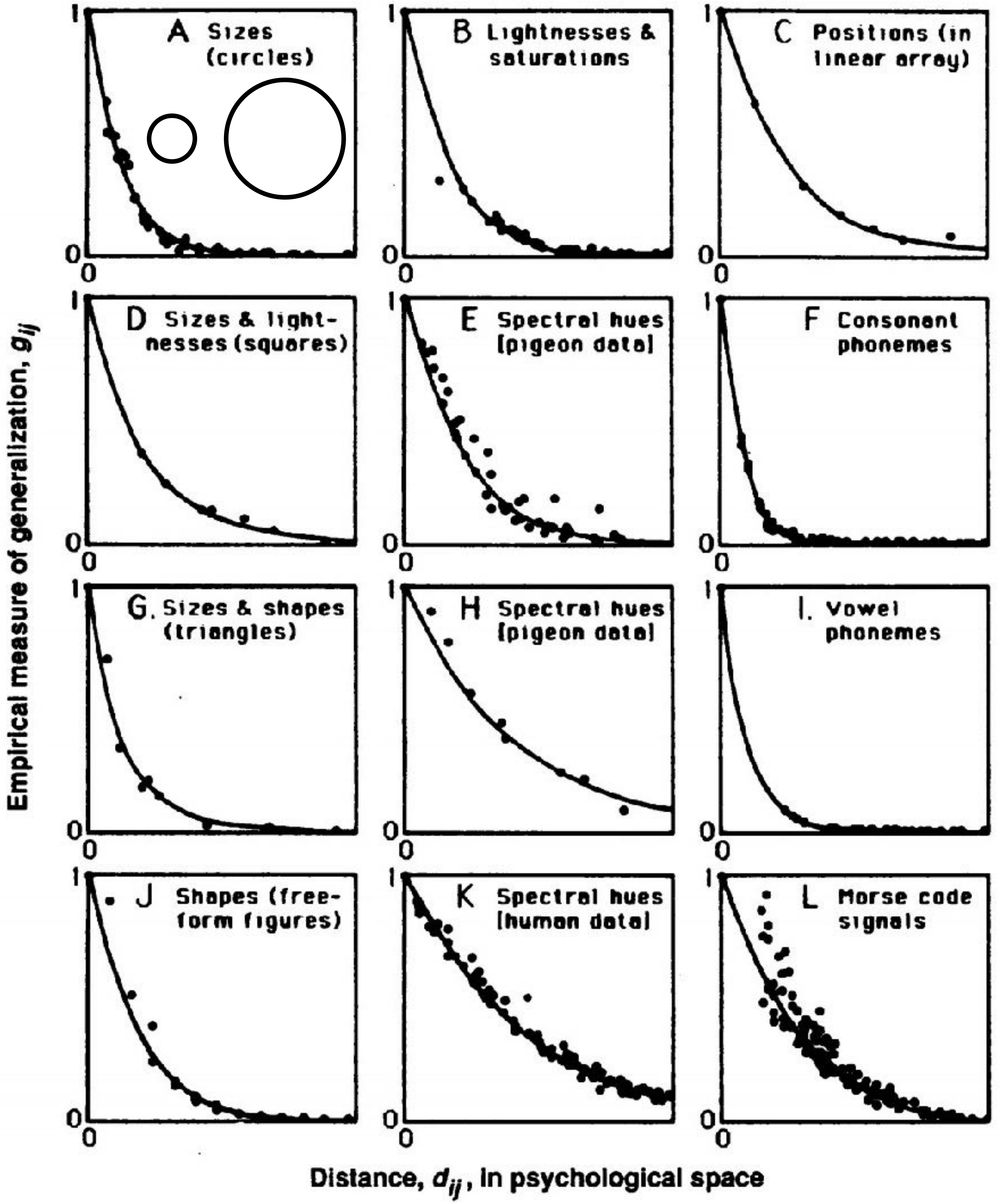


Illustration. Skinner box as adapted for the pigeon.



Shepard's (1987) Law of Generalization

STIMULUS + RESPONSE = **LEARNING**

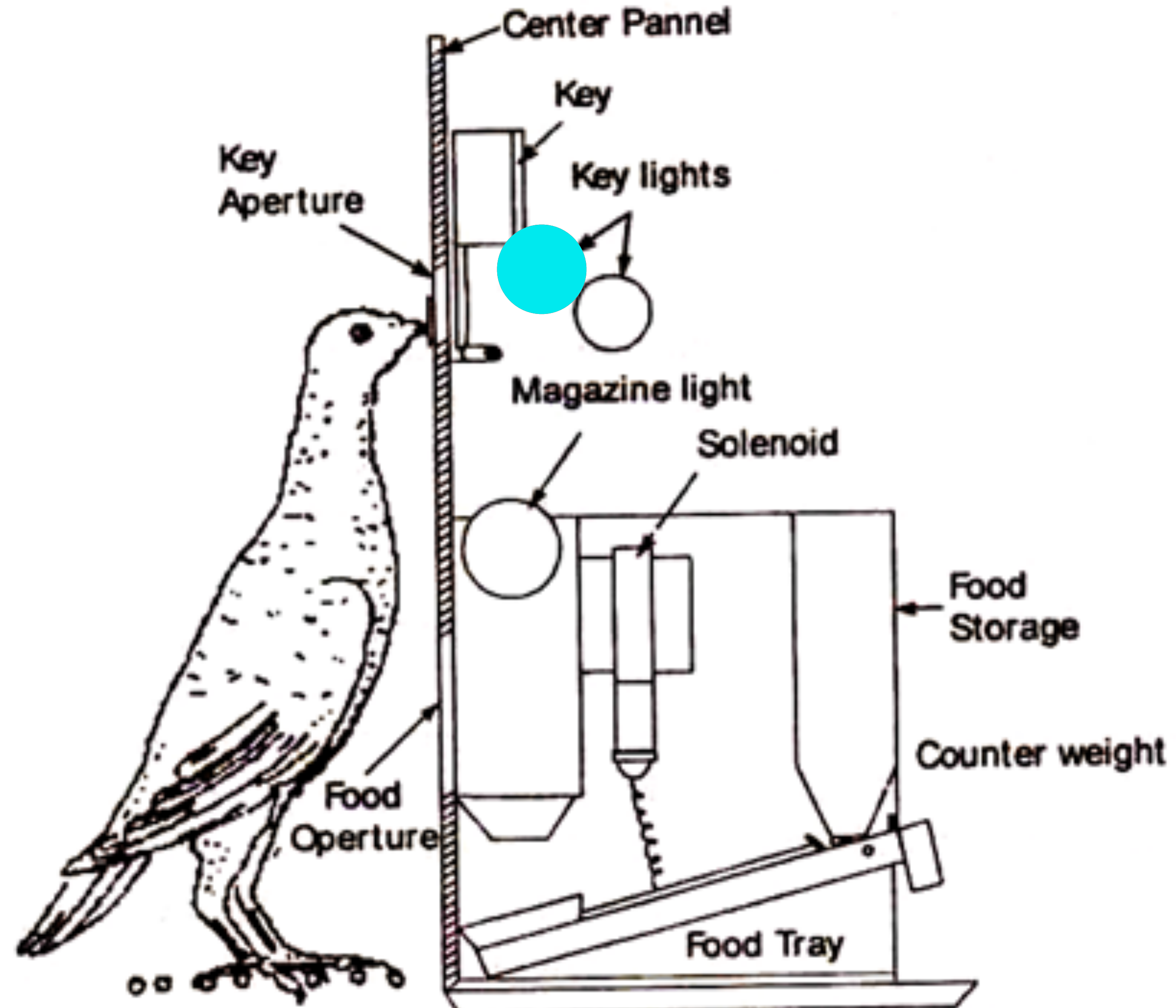
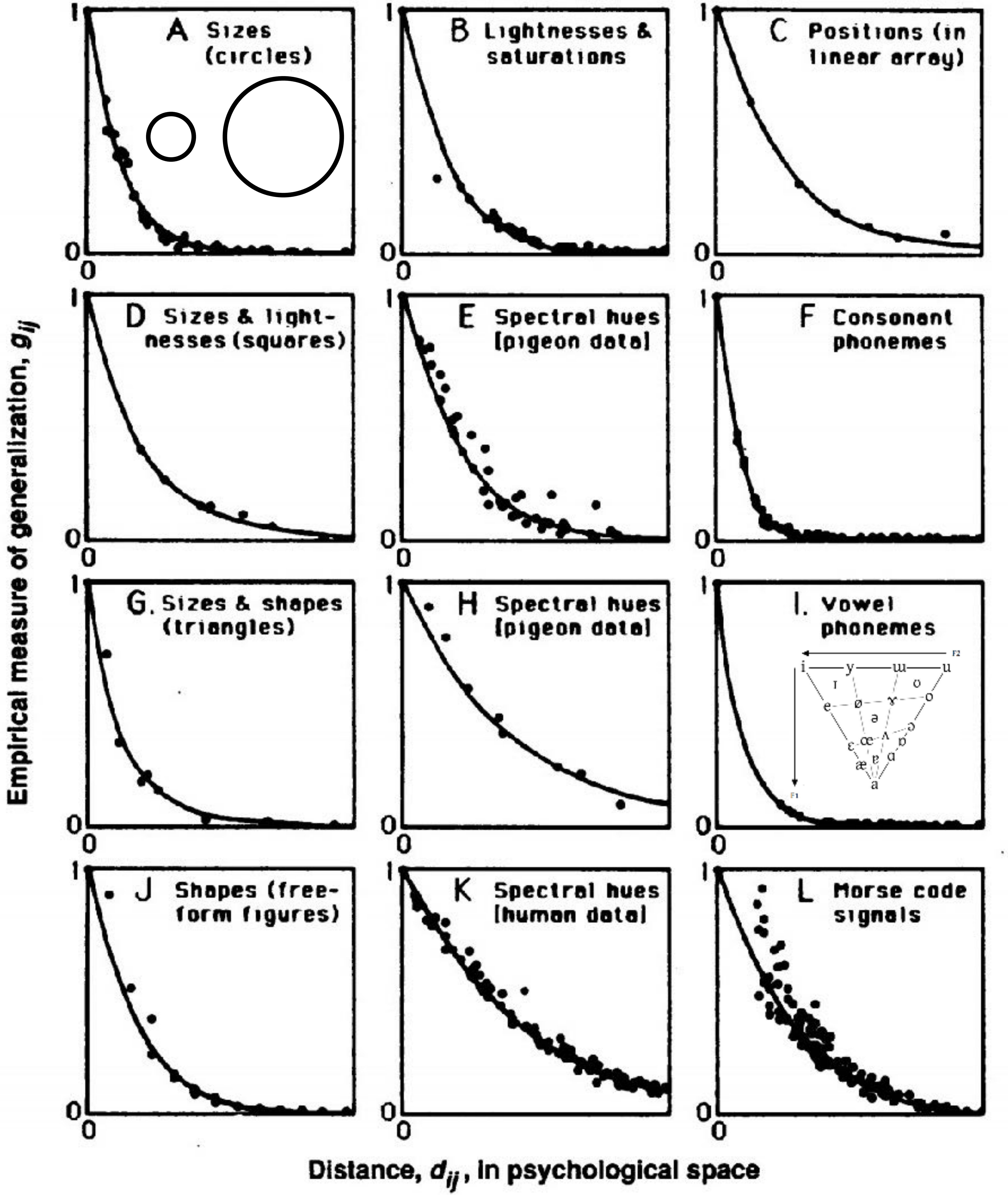
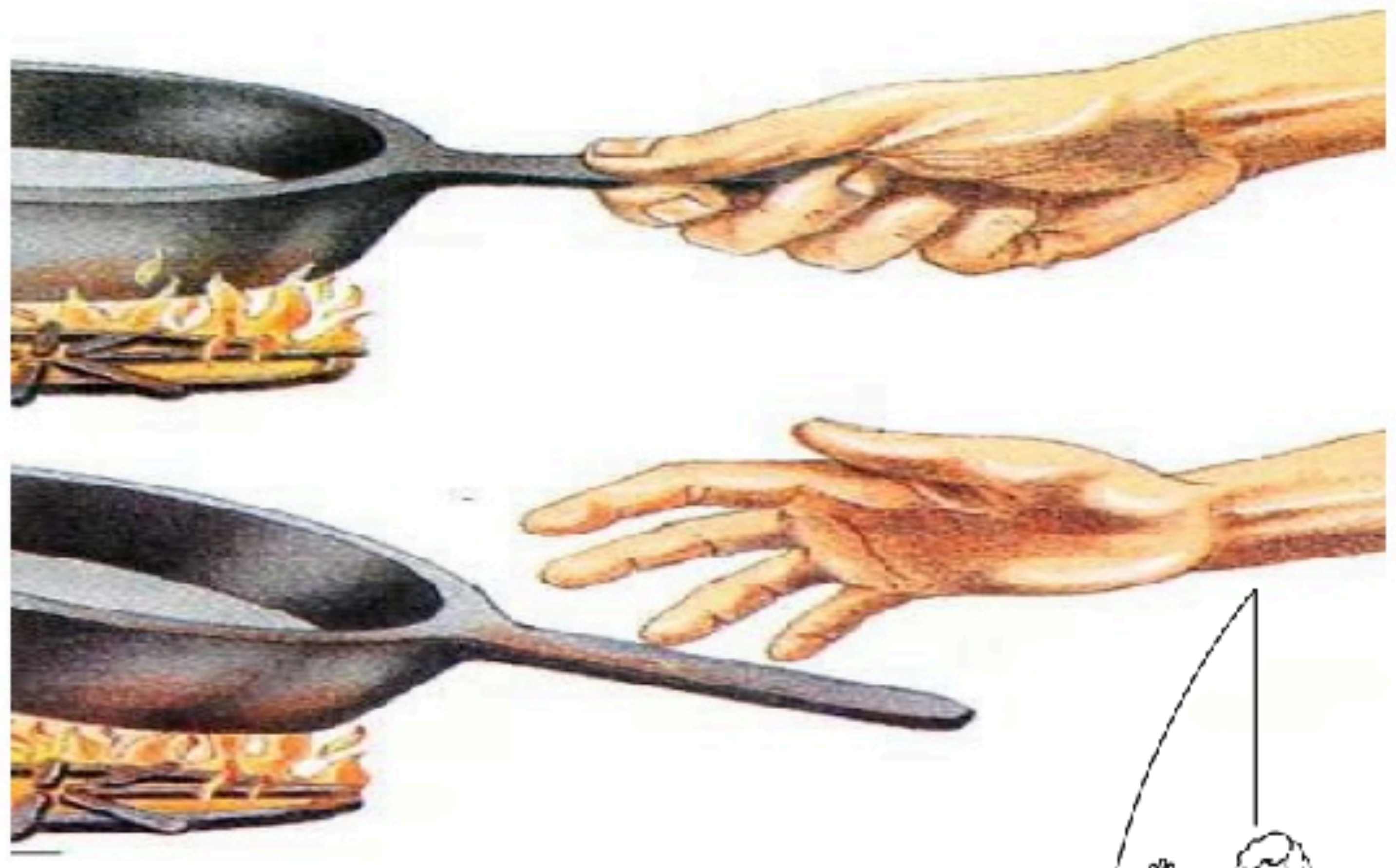


Illustration. Skinner box as adapted for the pigeon.



Shepard’s (1987) Law of Generalization

STIMULUS + RESPONSE = **LEARNING**



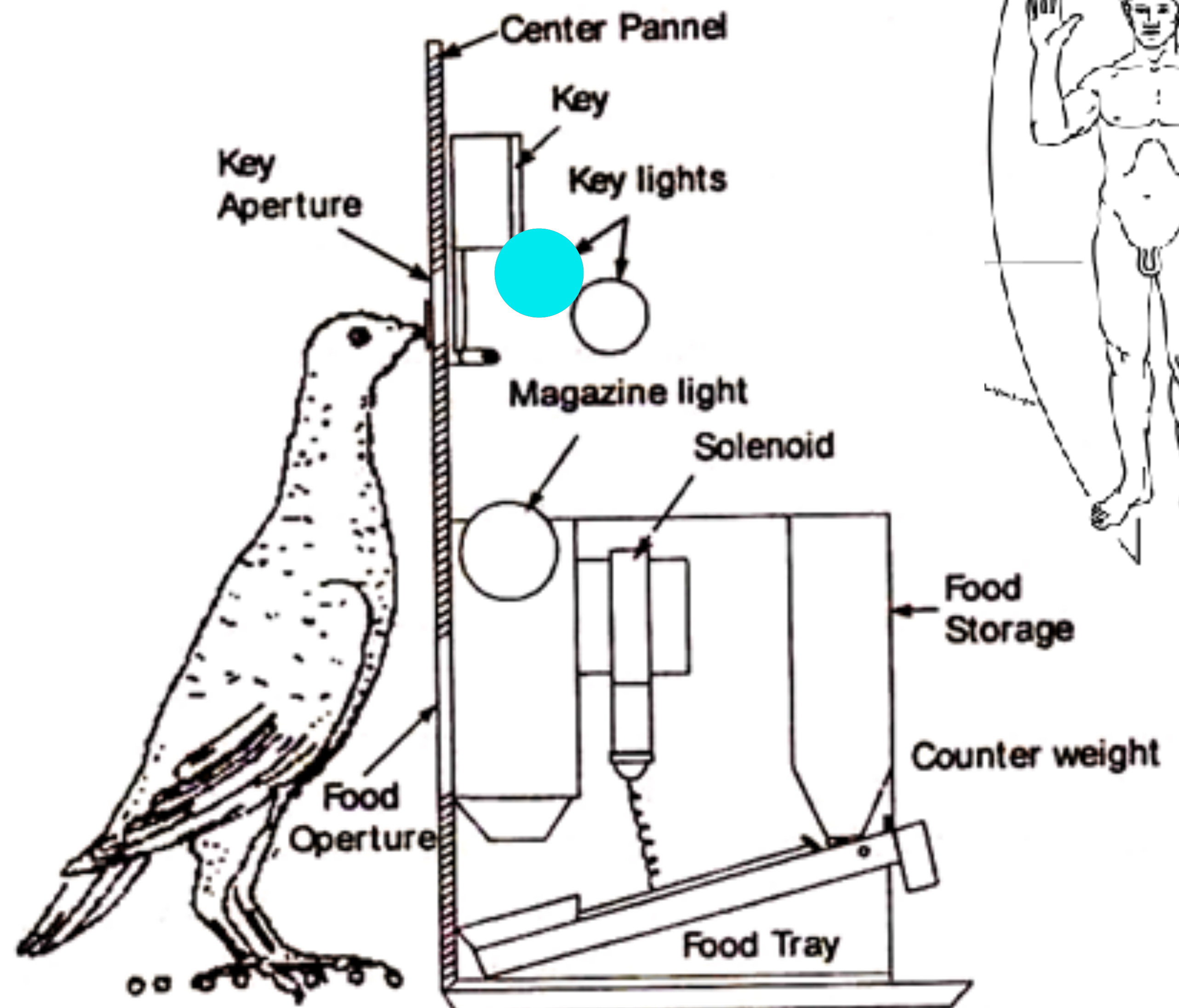
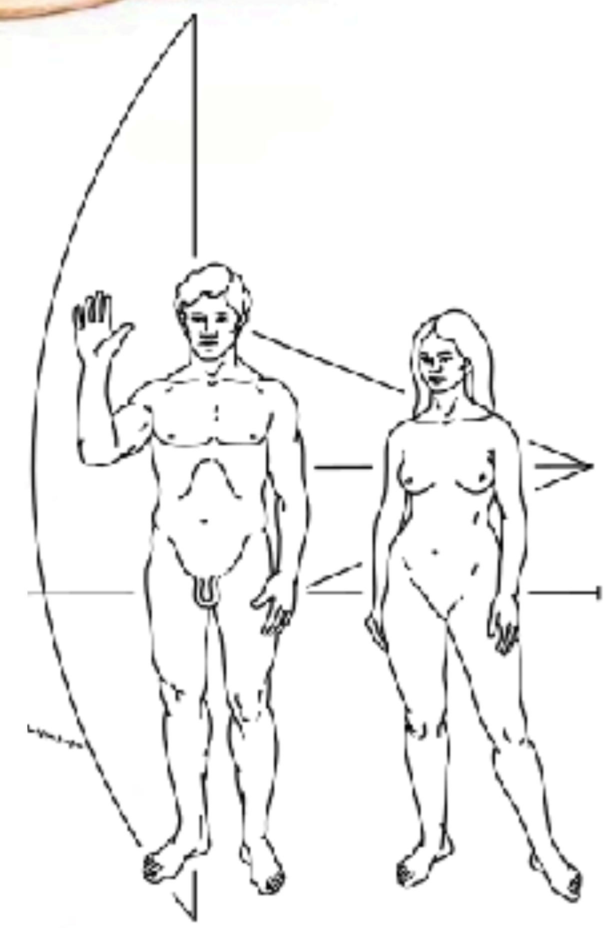
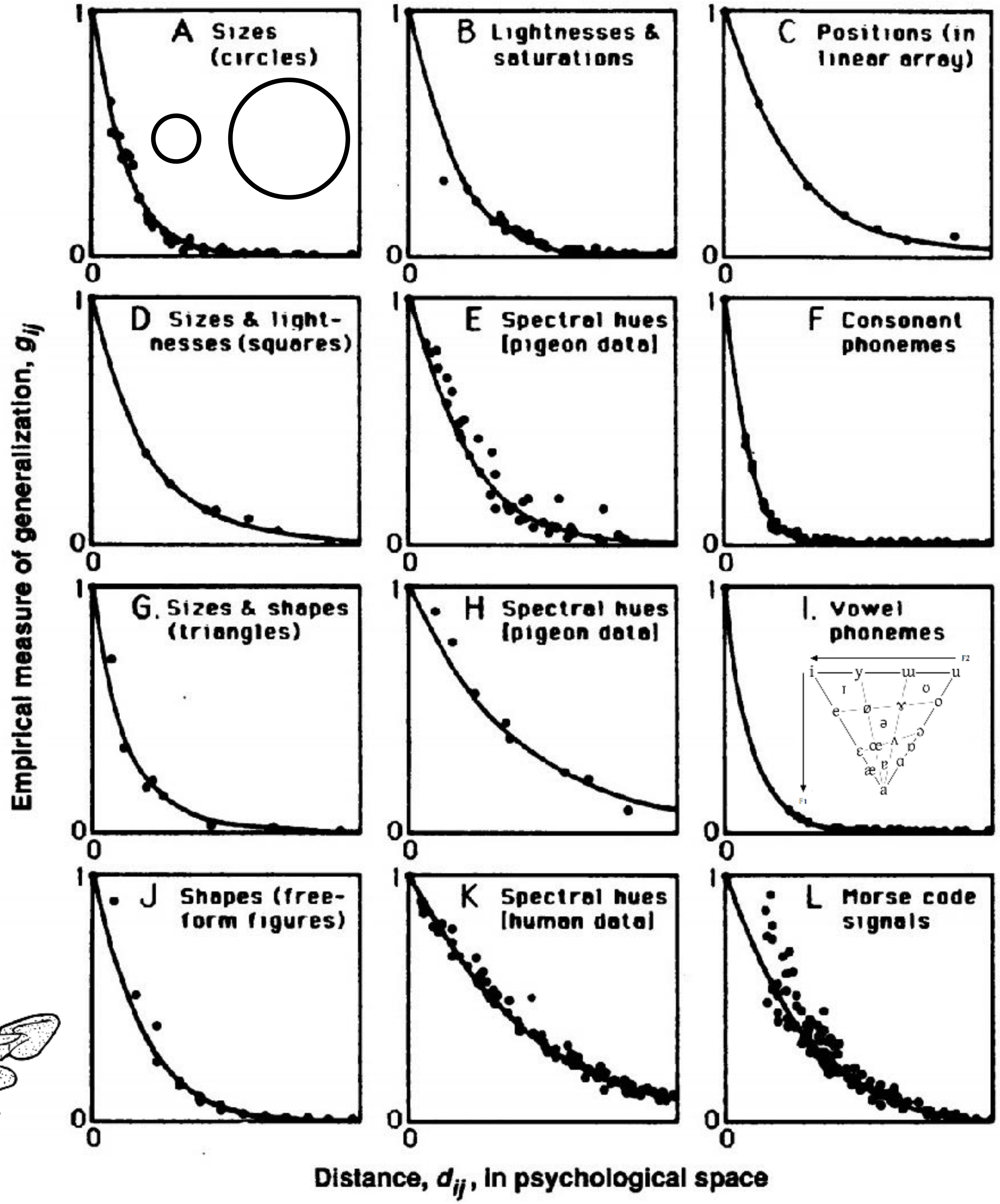


Illustration. Skinner box as adapted for the pigeon.





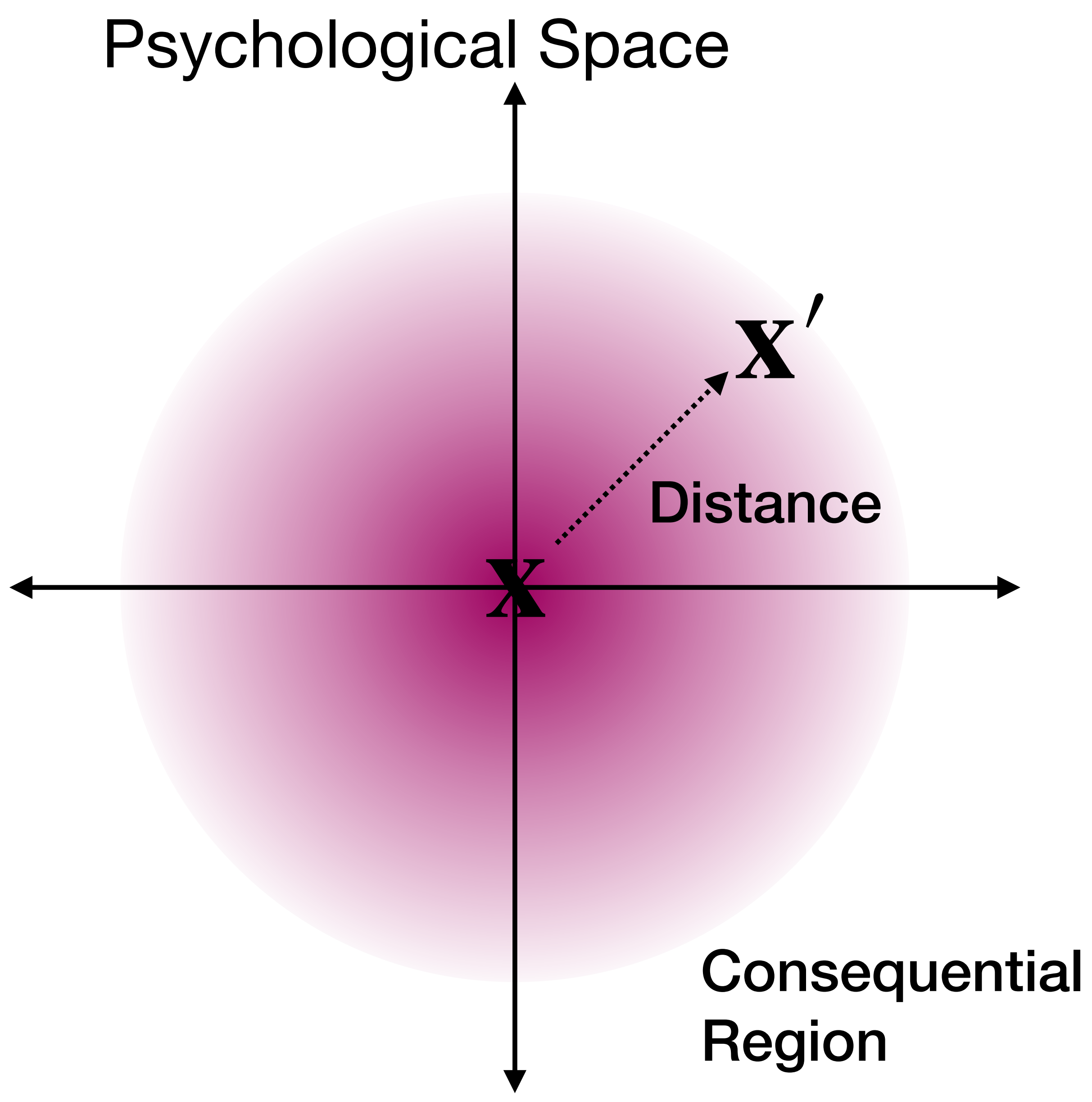
Generalization in Psychological Space

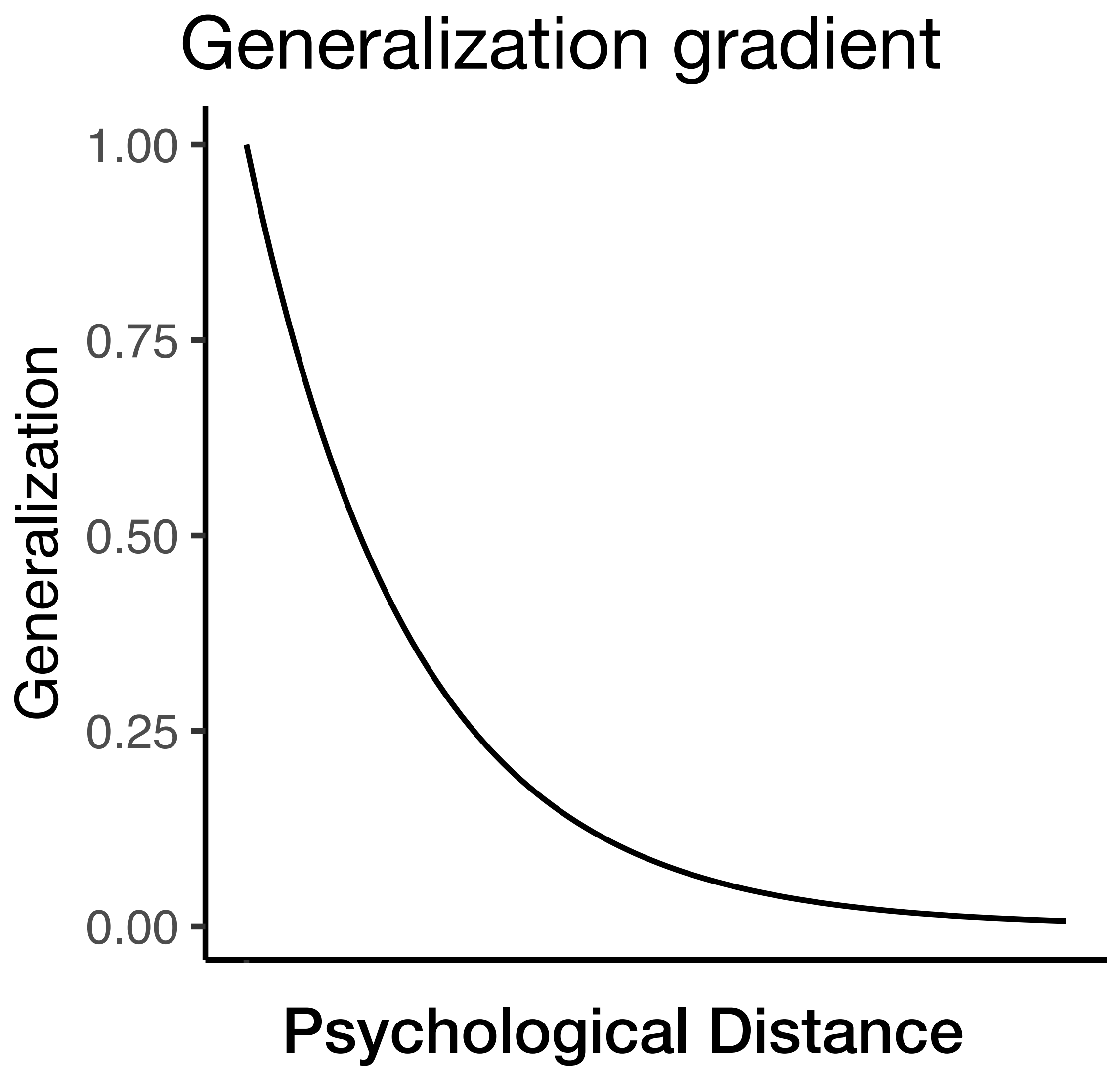
- Shepard (1987) believed that representations about categories or natural kinds correspond to a *consequential region* in psychological space

- Generalization arises from uncertainty about the extent of these regions

- As representational distance between stimuli **x** and **x**’ increases (i.e., become less **similar**), they are less likely to belong to the same region, and thus produce less similar outcomes

- This produces the smooth gradient of generalization





We generalize from one situation to another not because we cannot tell the difference between the two situations but because we judge that they are likely to belong to a set of situations having the same consequence. Generalization, which stems from uncertainty about the distribution of consequential stimuli in psychological space, is thus to be distinguished from failure of discrimination, which stems from uncertainty about the relative locations of individual stimuli in that space.

Shepard (*Science*, 1987)

Limitations of "metric" similarity

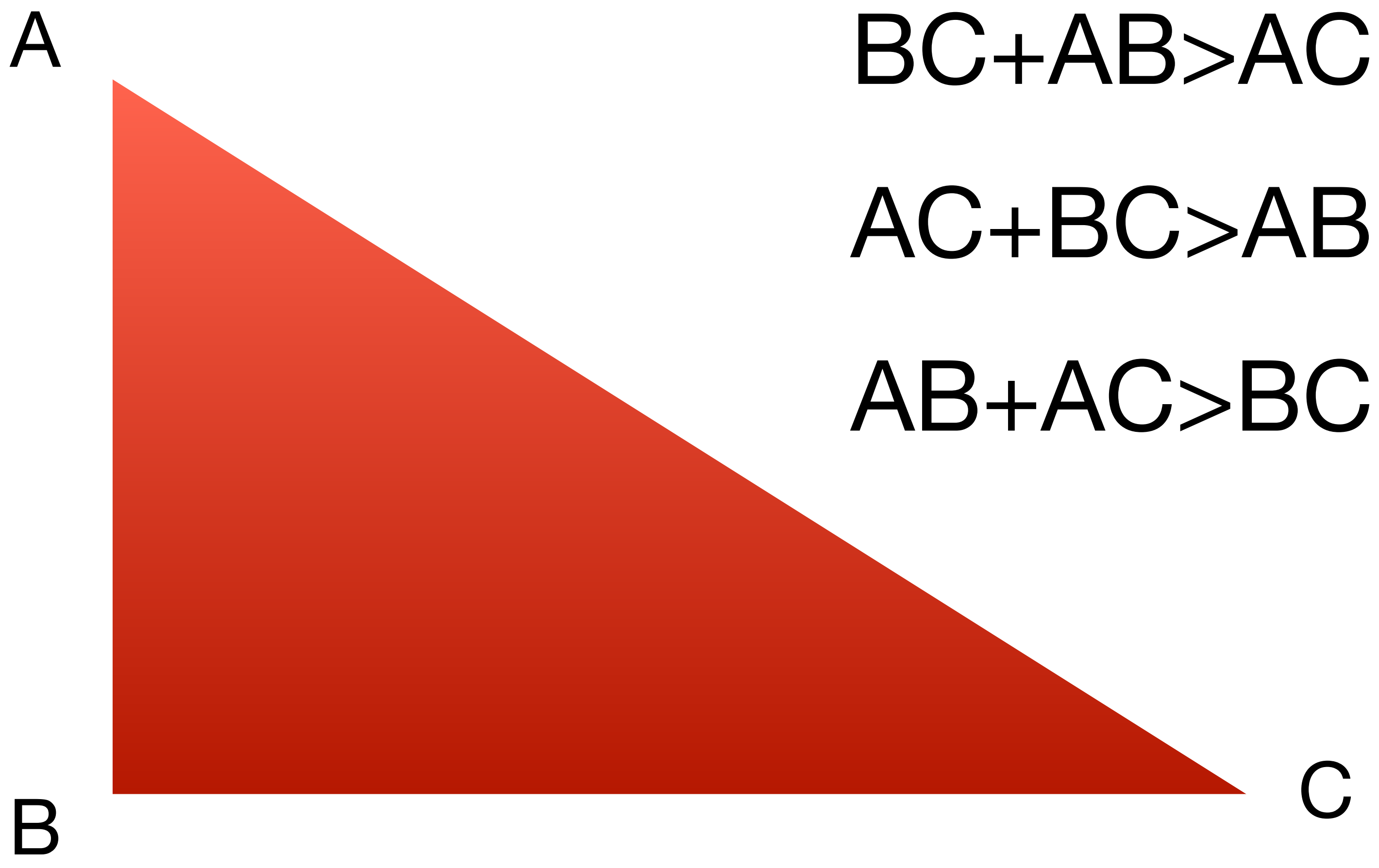
Amos Tversky
(1937-1996)

- Two definitive properties are *symmetry* and *triangle inequality*
- But they are often violated in human judgments of similarity (Tversky, 1977)

Symmetry

$$d(\mathbf{x}, \mathbf{x}') = d(\mathbf{x}', \mathbf{x})$$

Triangle Inequality



BC+AB>AC

AC+BC>AB

AB+AC>BC

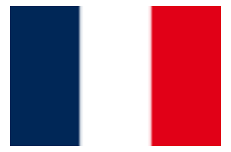
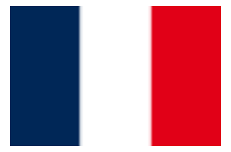
Limitations of “metric” similarity

Amos Tversky
(1937-1996)

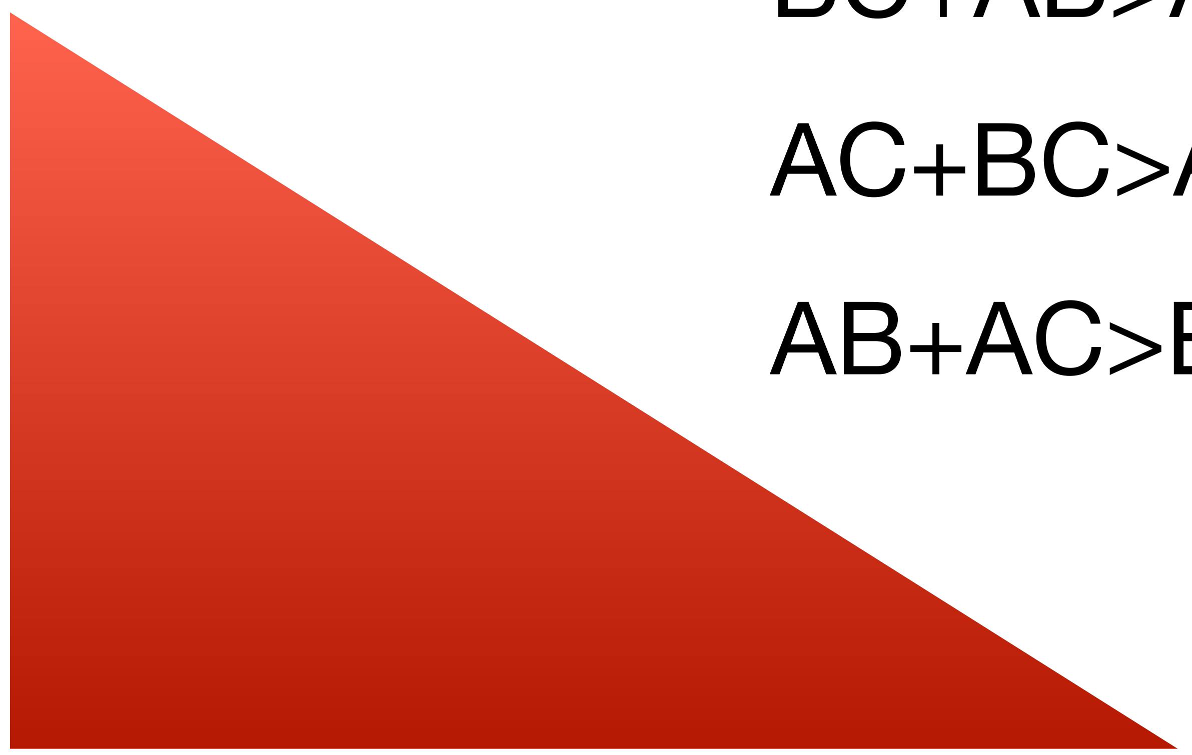
- Two definitive properties are *symmetry* and *triangle inequality*
- But they are often violated in human judgments of similarity (Tversky, 1977)

Symmetry

$$d(\mathbf{x}, \mathbf{x}') = d(\mathbf{x}', \mathbf{x})$$

$d($, $) \neq d($, $)$

Triangle Inequality



BC+AB>AC

AC+BC>AB

AB+AC>BC

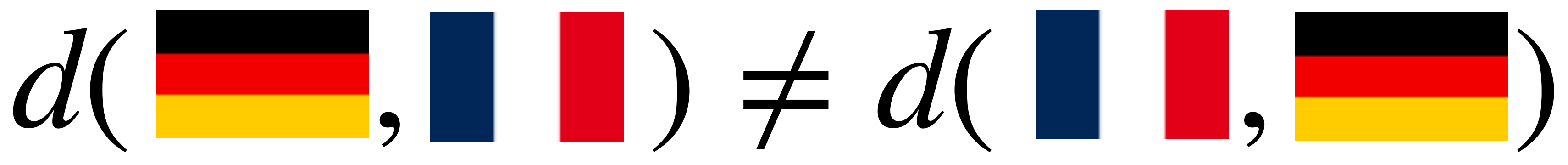
Limitations of “metric” similarity

Amos Tversky (1937-1996)

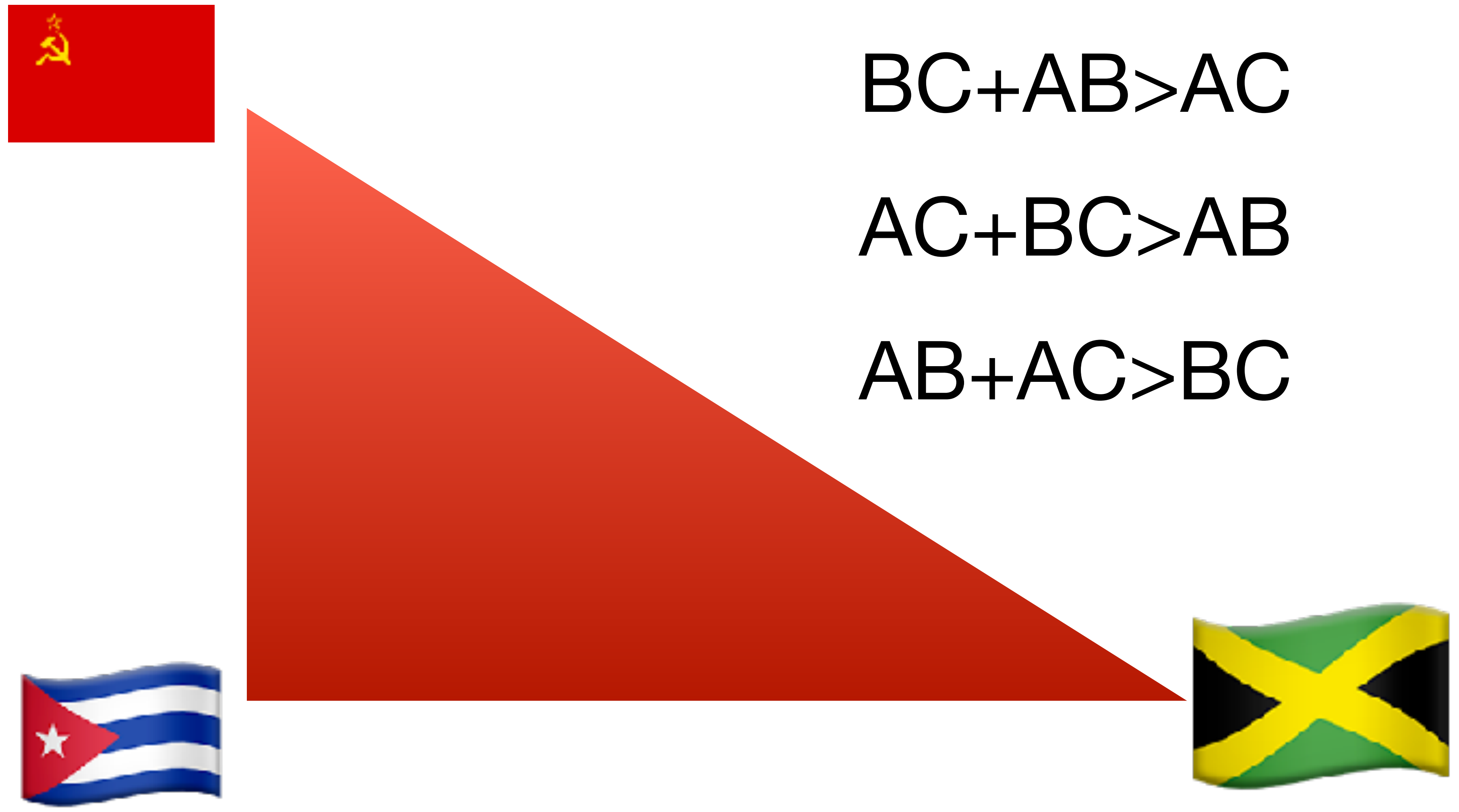
- Two definitive properties are *symmetry* and *triangle inequality*
- But they are often violated in human judgments of similarity (Tversky, 1977)

Symmetry

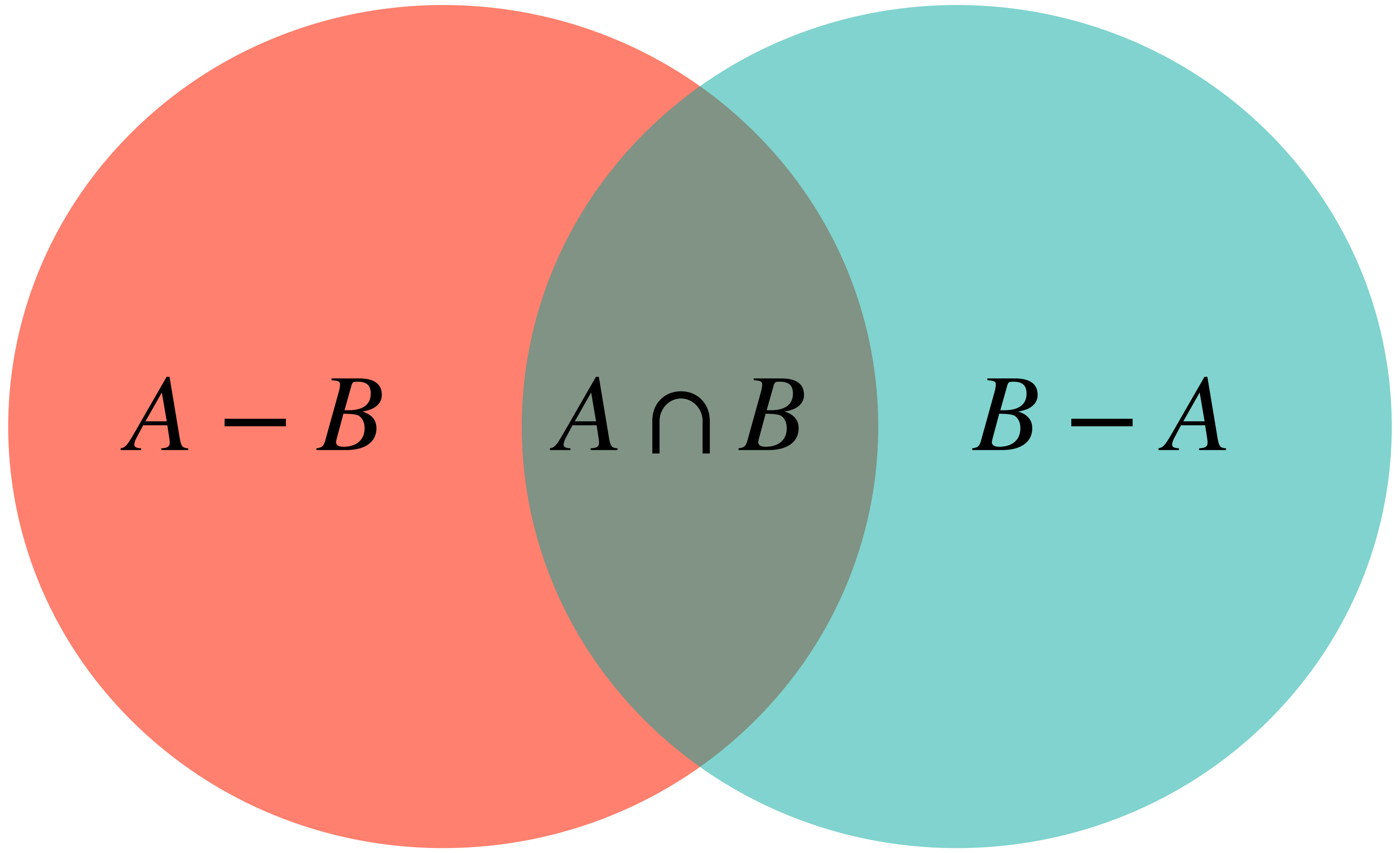
$$d(\mathbf{x}, \mathbf{x}') = d(\mathbf{x}', \mathbf{x})$$



Triangle Inequality



Contrast model

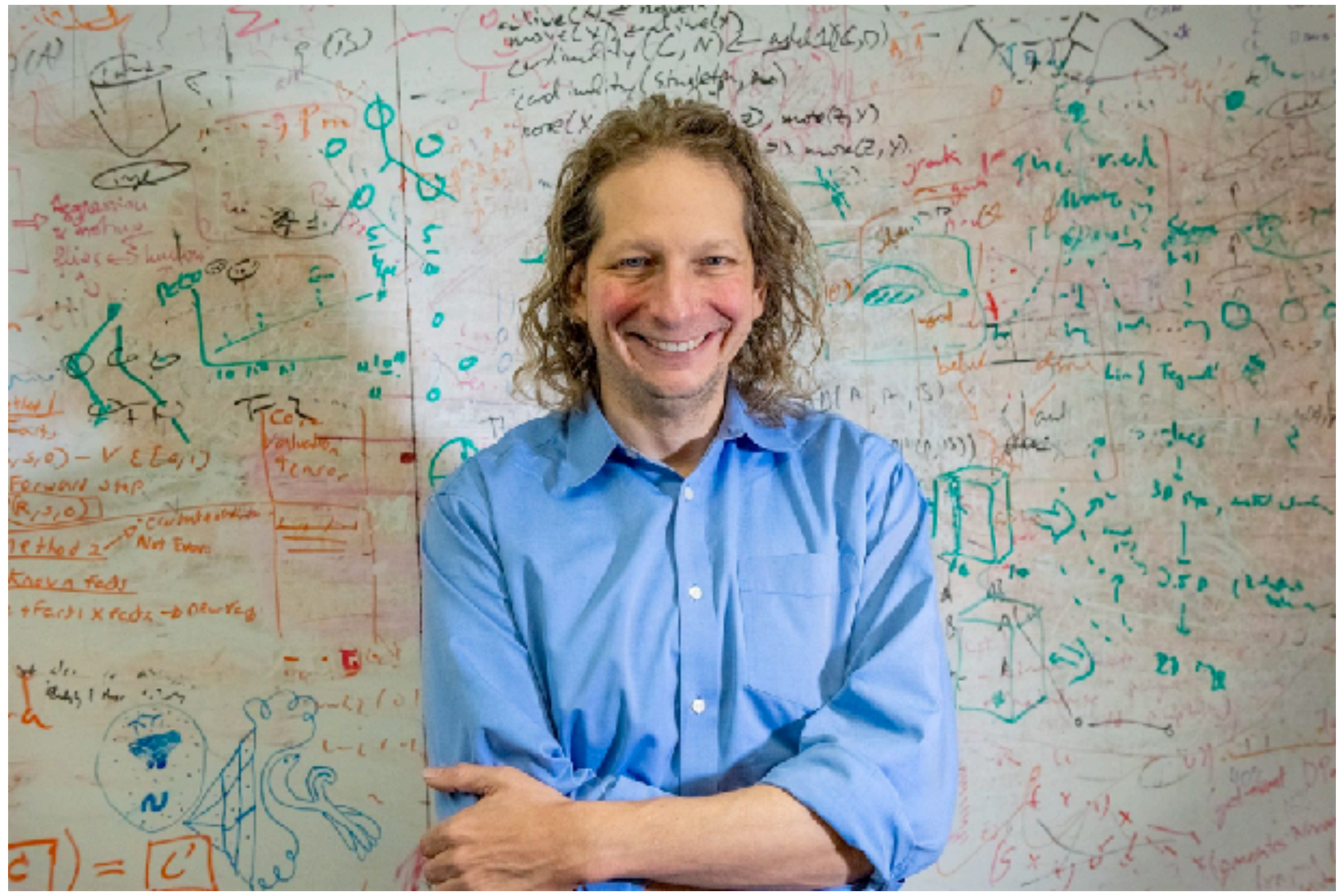


$$\text{sim}(A, B) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

- θ, α, β are free parameters
- To translate into Shepard's language, rather than consequential regions in psychological space, concepts are defined based on sets of features
  - Similar to *family resemblance theory* (Rosch & Mervis, 1975)
- Common and disjoint features may be weighted differently
- A more refined similarity theory that allows for asymmetric similarity judgments that can also violate triangle inequality

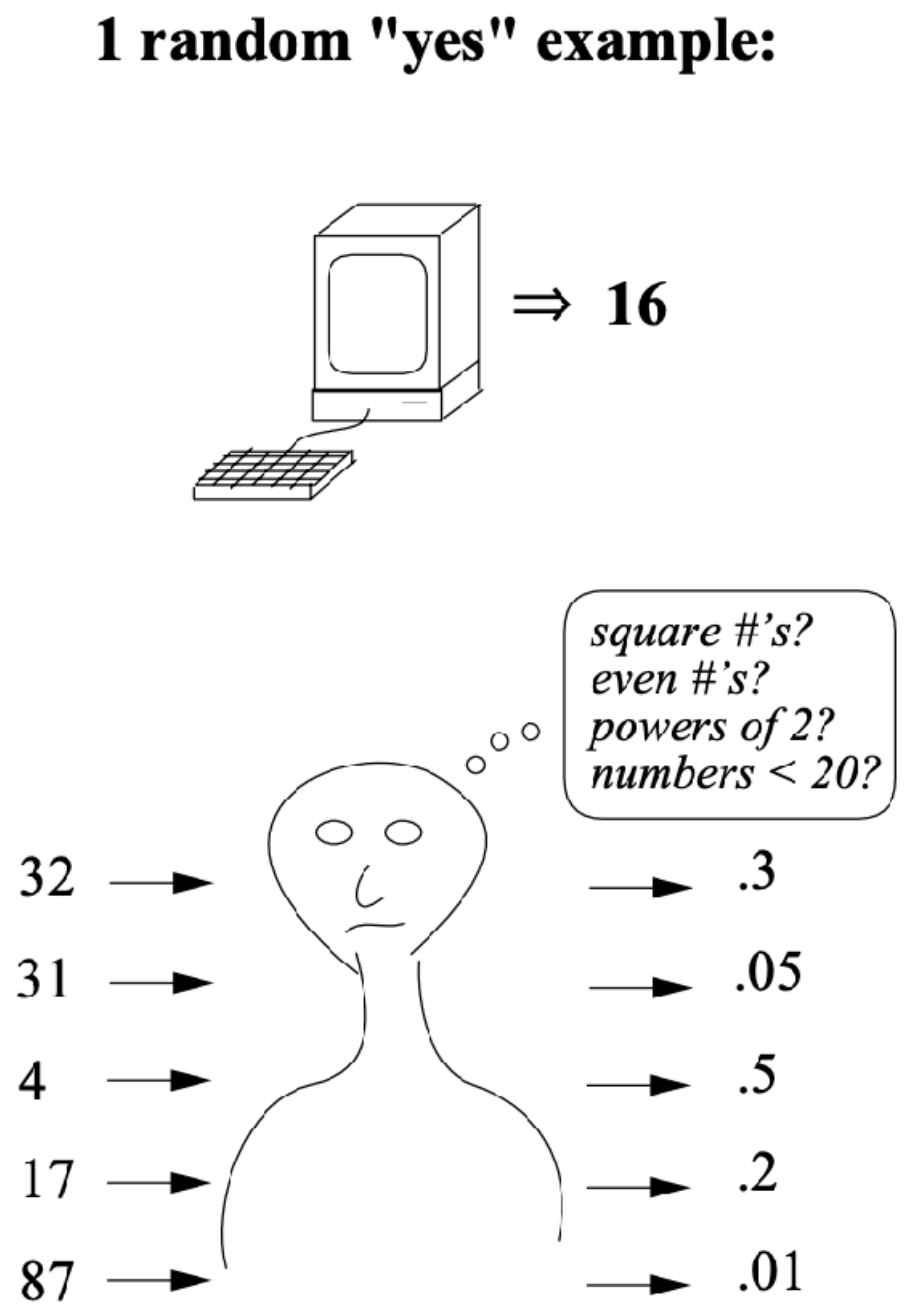
Bayesian concept learning as a hybrid approach

- Much of modern cognitive science is dominated by Bayesian inference
- Josh Tenenbaum and Tom Griffiths are two individuals who are largely responsible for it's popularity
- The same basic concept can explain a huge host of problems, from language acquisition, to structure learning, to program induction
- But it all started with a number game and a model of probablistic rule learning from Josh's PhD thesis



Number concepts

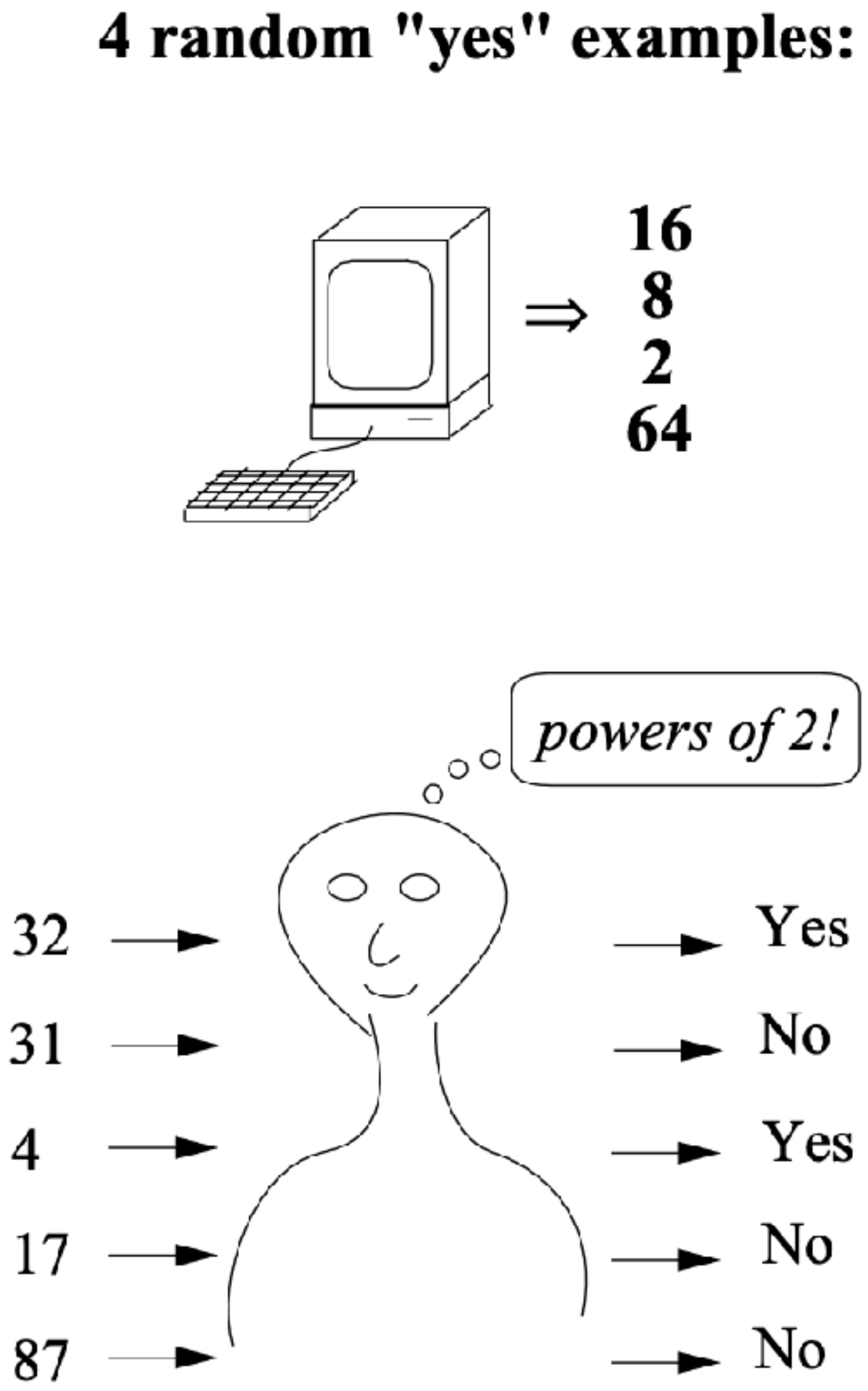
- Examples:
  - X is an even number
  - X is between 30 and 45
  - X is a prime number
- A computer generates a random number from a chosen concept, and you need to guess another number that is likely to fit



Tenenbaum (*PhD thesis* 1999)

Number concepts

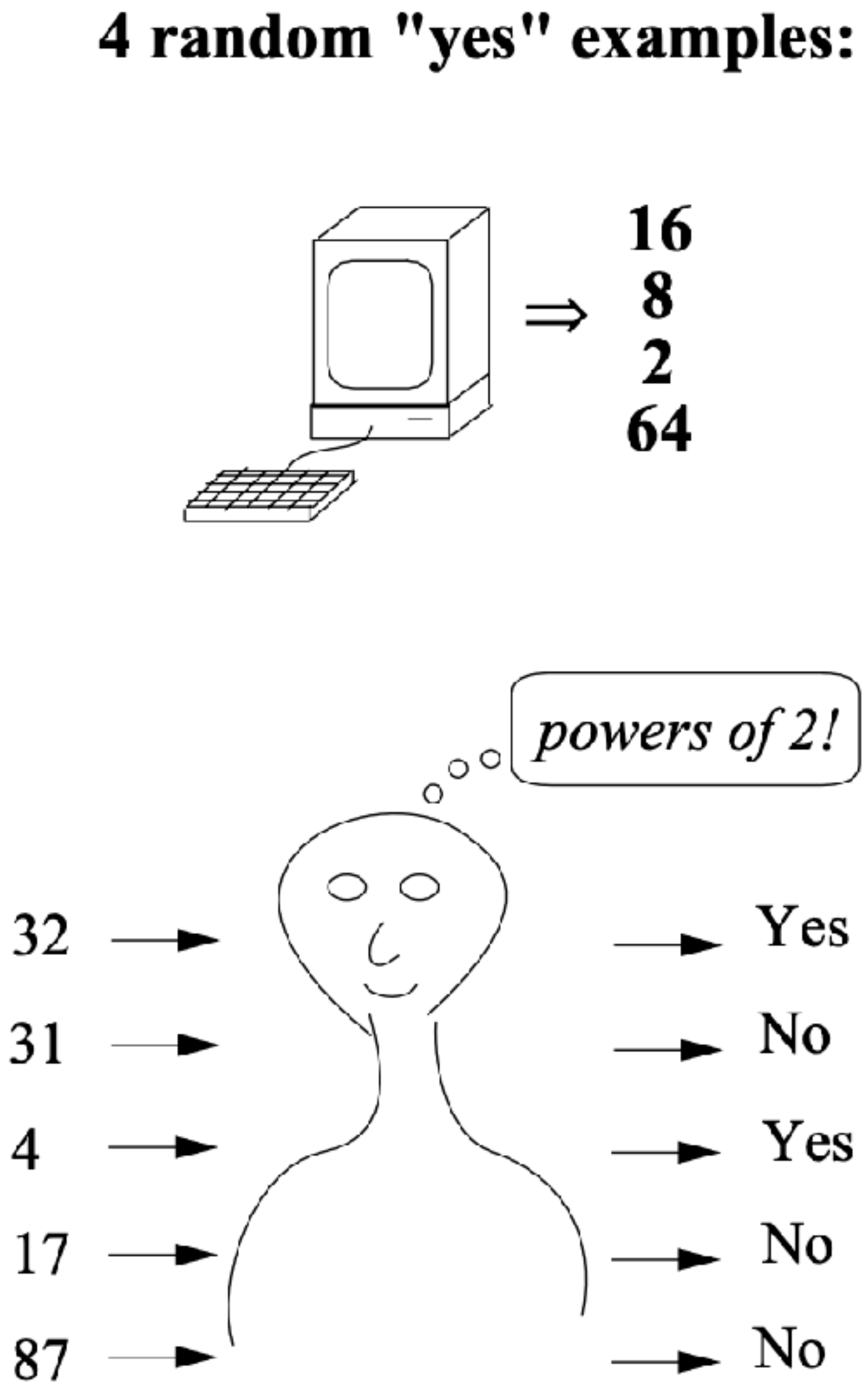
- Examples:
  - X is an even number
  - X is between 30 and 45
  - X is a prime number
- A computer generates a random number from a chosen concept, and you need to guess another number that is likely to fit



Tenenbaum (*PhD thesis* 1999)

Number concepts

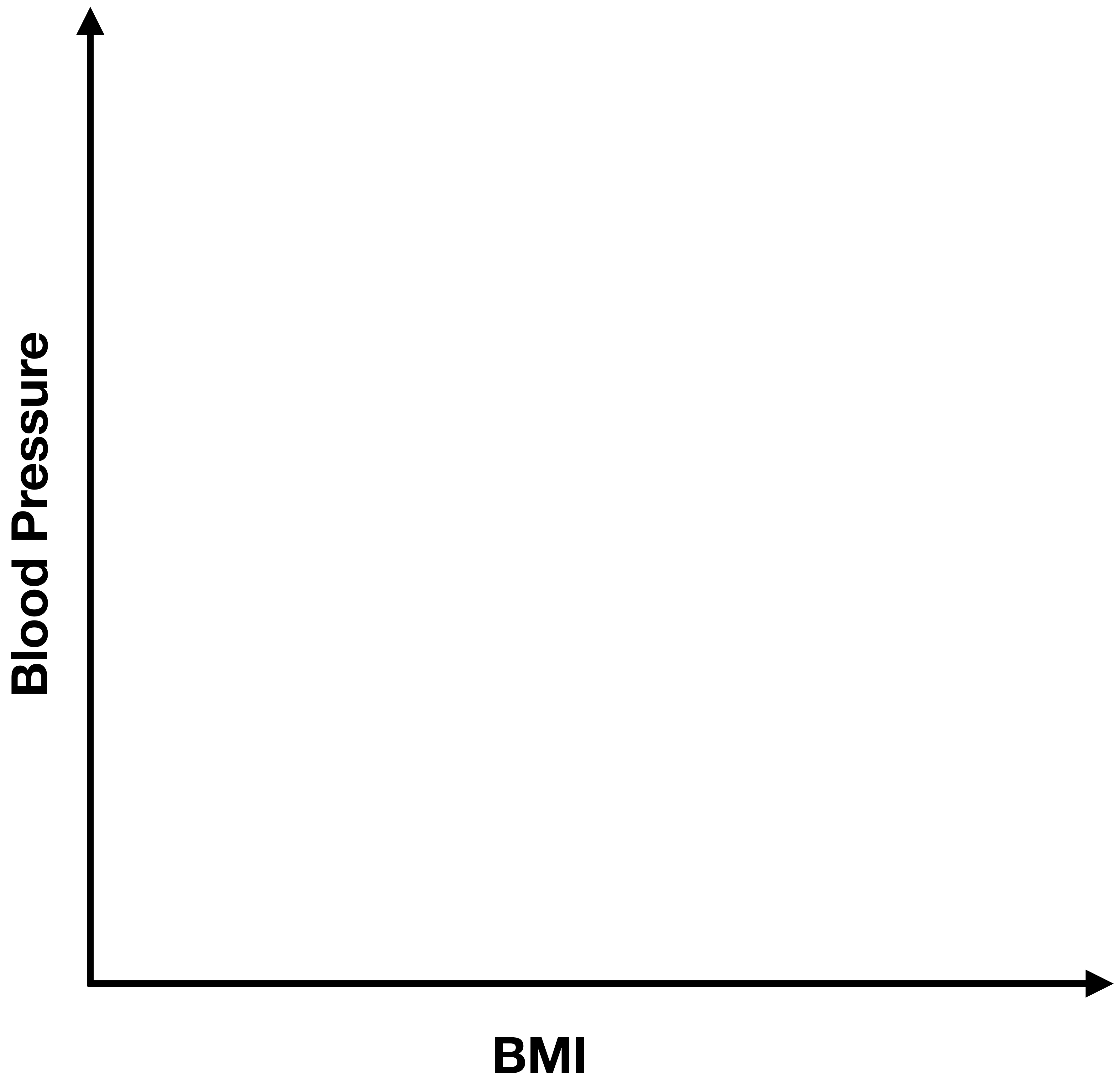
- Examples:
    - X is an even number
    - X is between 30 and 45
    - X is a prime number
- A computer generates a random number from a chosen concept, and you need to guess another number that is likely to fit
- *Even restricting the game to natural numbers between 1 and 100, there are more than a billion billion billion subsets of numbers that such a program could possibly have picked out and which are consistent with the observed “yes" examples of 16, 8, 2, and 64*



Tenenbaum (*PhD thesis* 1999)

Bayesian Concept Learning

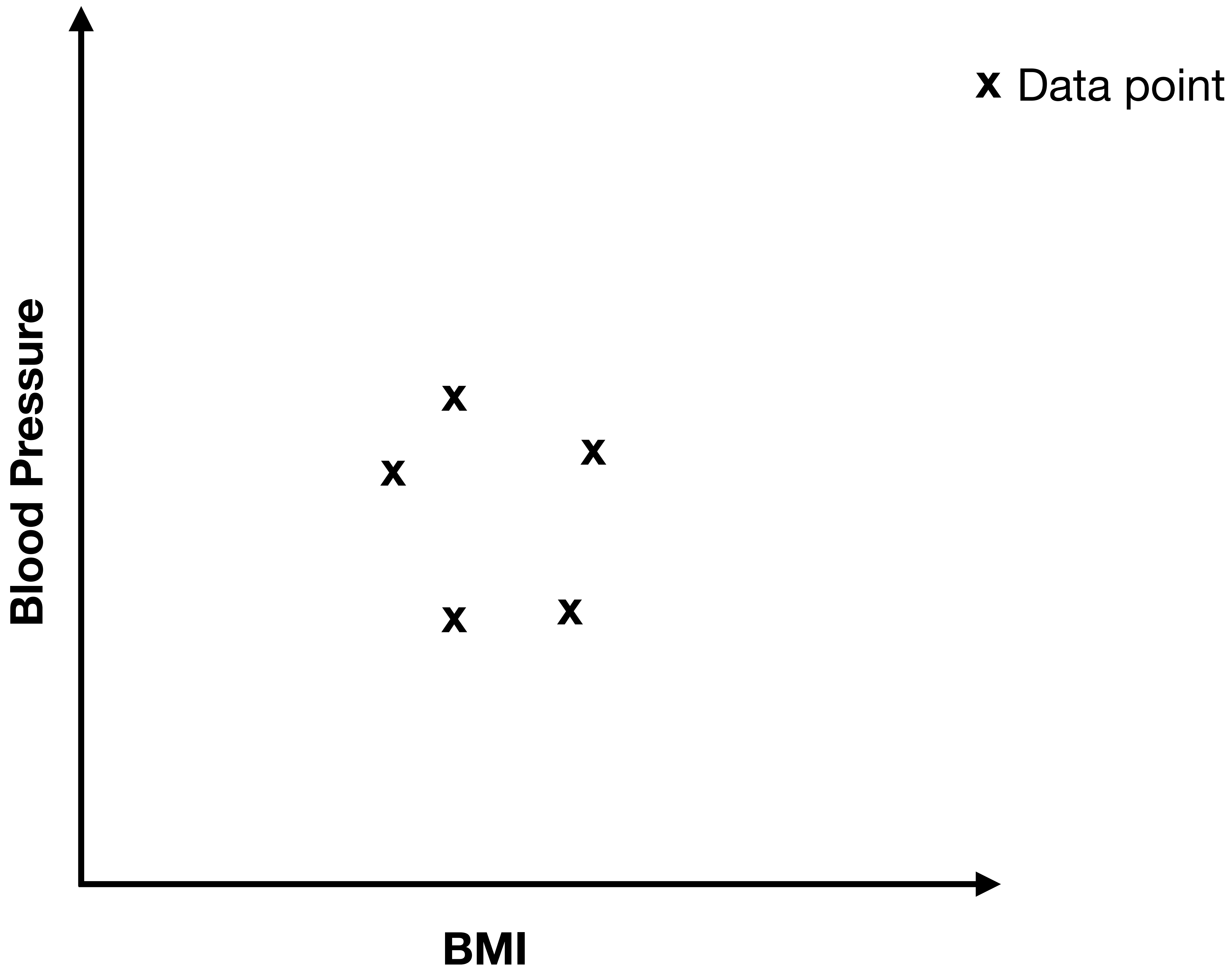
- Example: The concept of healthy person
- Problem: Given a set of examples (*x*'s in the plot), what is the probablity that some new example y will fall within consequential region *C* defining a healthy person?



Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

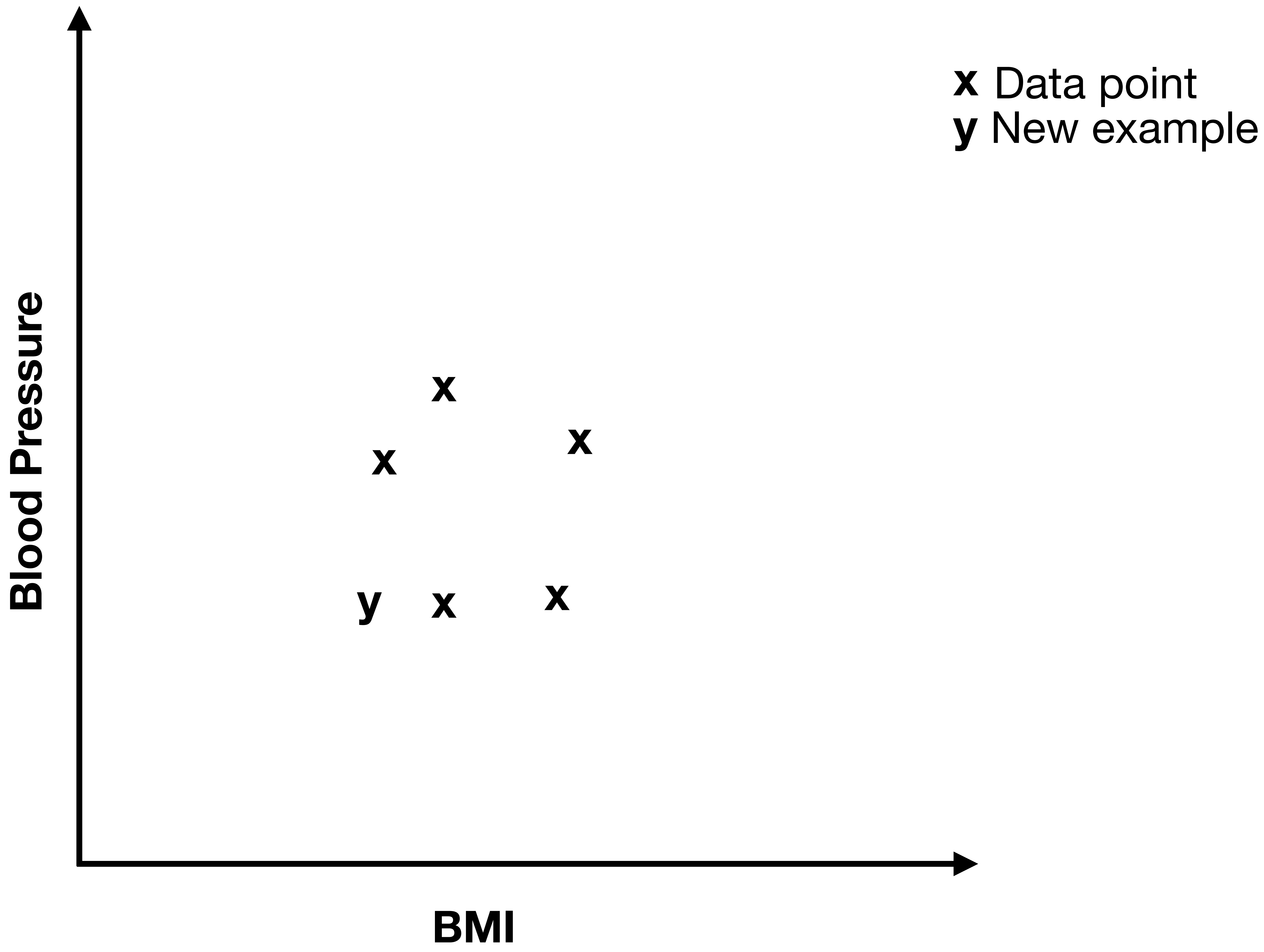
- Example: The concept of healthy person
- Problem: Given a set of examples (*x*'s in the plot), what is the probablity that some new example y will fall within consequential region *C* defining a healthy person?



Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

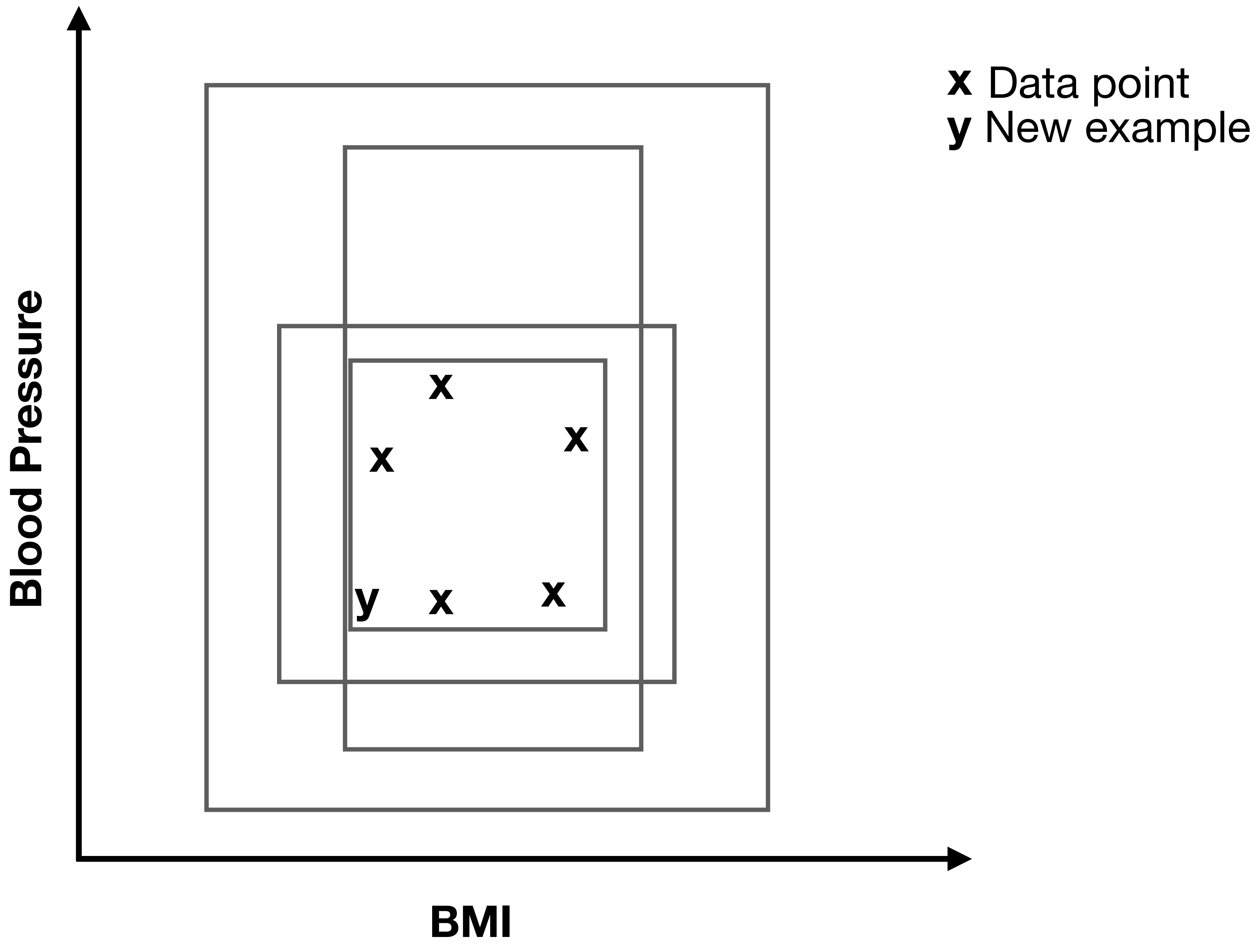
- Example: The concept of healthy person
- Problem: Given a set of examples (*x*'s in the plot), what is the probablity that some new example y will fall within consequential region *C* defining a healthy person?



Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

- Example: The concept of healthy person
- Problem: Given a set of examples (*x*'s in the plot), what is the probablity that some new example y will fall within consequential region *C* defining a healthy person?
- Solution: It depends on a distribution over hypotheses *h* (illustrated as rectangles) about the boundaries of$C$



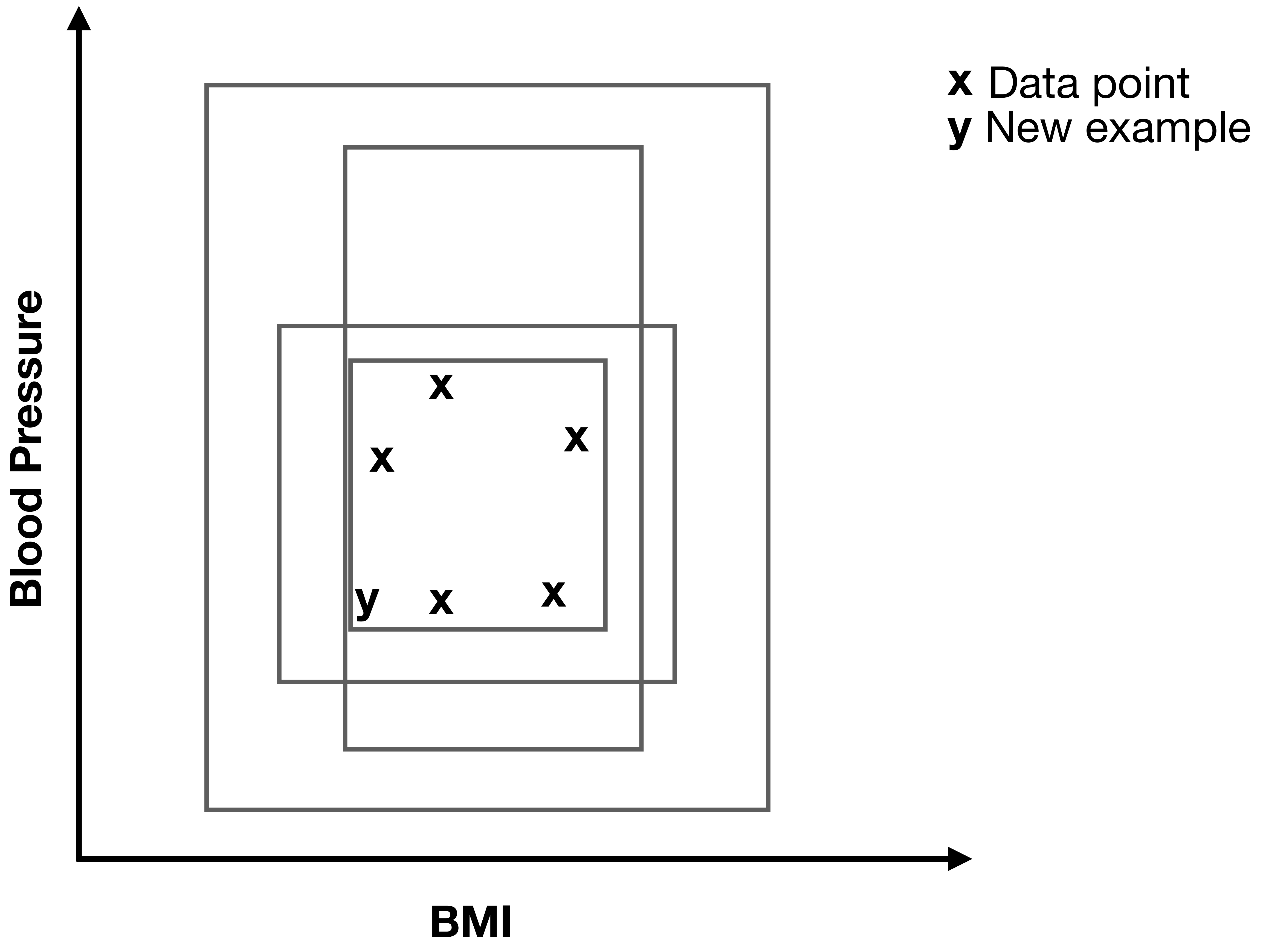
Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

- Example: The concept of healthy person
- Problem: Given a set of examples (*x*'s in the plot), what is the probablity that some new example y will fall within consequential region *C* defining a healthy person?
- Solution: It depends on a distribution over hypotheses *h* (illustrated as rectangles) about the boundaries of C

$$p(y \in C|x) = \sum_{h:y\in h} p(h|x).$$

Sum over hypotheses that include y



Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

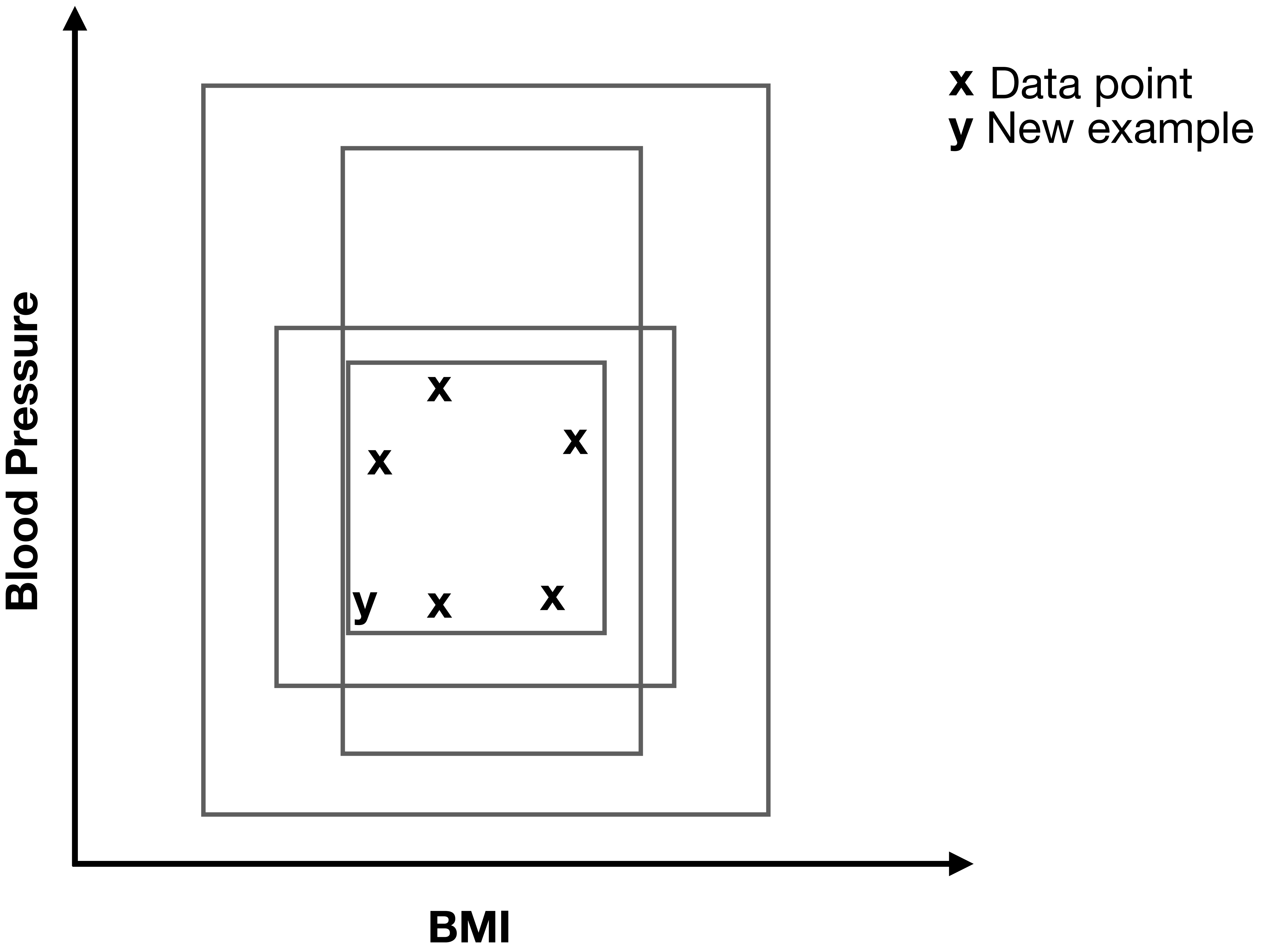
Bayesian Concept Learning

- Example: The concept of healthy person
- Problem: Given a set of examples (*x*'s in the plot), what is the probablity that some new example y will fall within consequential region *C* defining a healthy person?
- Solution: It depends on a distribution over hypotheses *h* (illustrated as rectangles) about the boundaries of$C$

$$p(y \in C|x) = \sum_{h:y \in h} p(h|x).$$

Sum over hypotheses that include y

but some hypotheses are more likely than others



Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

- Example: The concept of healthy person
- Problem: Given a set of examples (*x*'s in the plot), what is the probablity that some new example y will fall within consequential region *C* defining a healthy person?
- Solution: It depends on a distribution over hypotheses *h* (illustrated as rectangles) about the boundaries of C

$$p(y \in C|x) = \sum_{h: y \in h} p(h|x).$$

Sum over hypotheses that include y

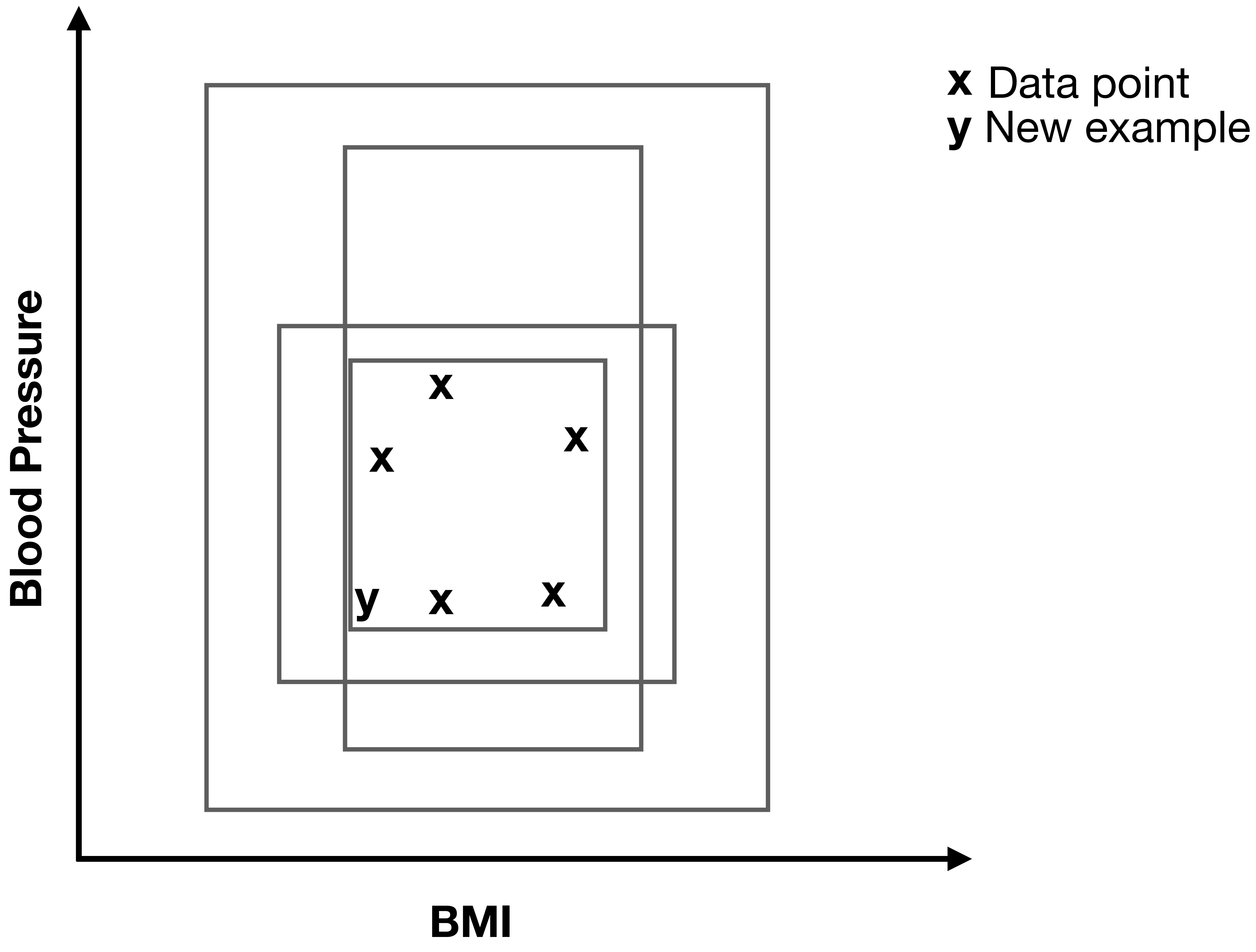
but some hypotheses are more likely than others

Bayes' rule

$$p(h|x) = \frac{p(x|h)p(h)}{p(x)}$$

likelihood * prior / evidence

$$= \frac{p(x|h)p(h)}{\sum_{h' \in \mathcal{H}} p(x|h')p(h')}.$$



Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

- Example: The concept of healthy person
- Problem: Given a set of examples (*x*'s in the plot), what is the probablity that some new example y will fall within consequential region *C* defining a healthy person?
- Solution: It depends on a distribution over hypotheses *h* (illustrated as rectangles) about the boundaries of C

$$p(y \in C|x) = \sum_{h:y\in h} p(h|x).$$

Sum over hypotheses that include y

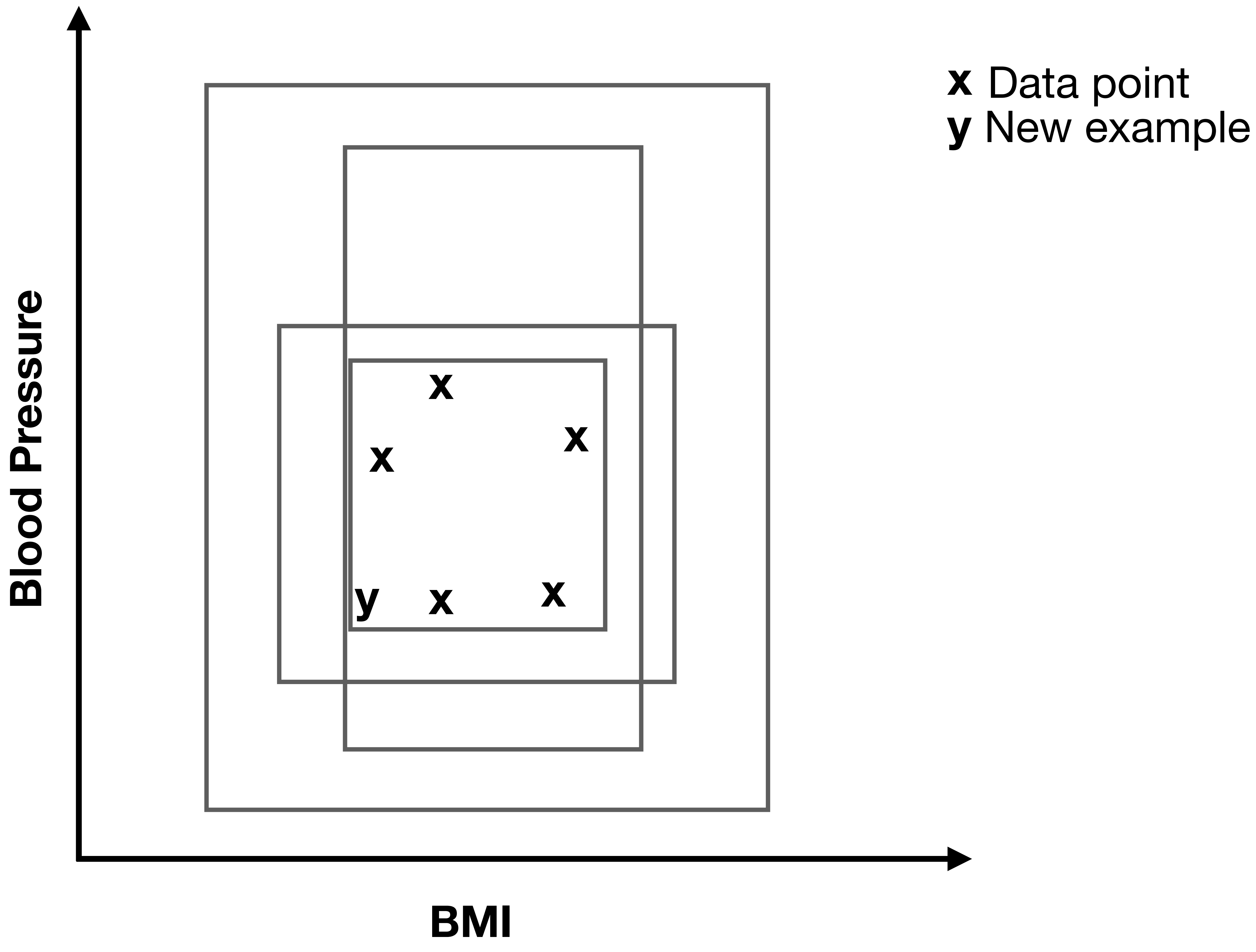
but some hypotheses are more likely than others

Bayes' rule

$$p(h|x) = \frac{p(x|h)p(h)}{p(x)}$$

likelihood * prior / evidence

$$= \frac{\boxed{p(x|h)}p(h)}{\sum_{h'\in\mathcal{H}} p(x|h')p(h')}.$$



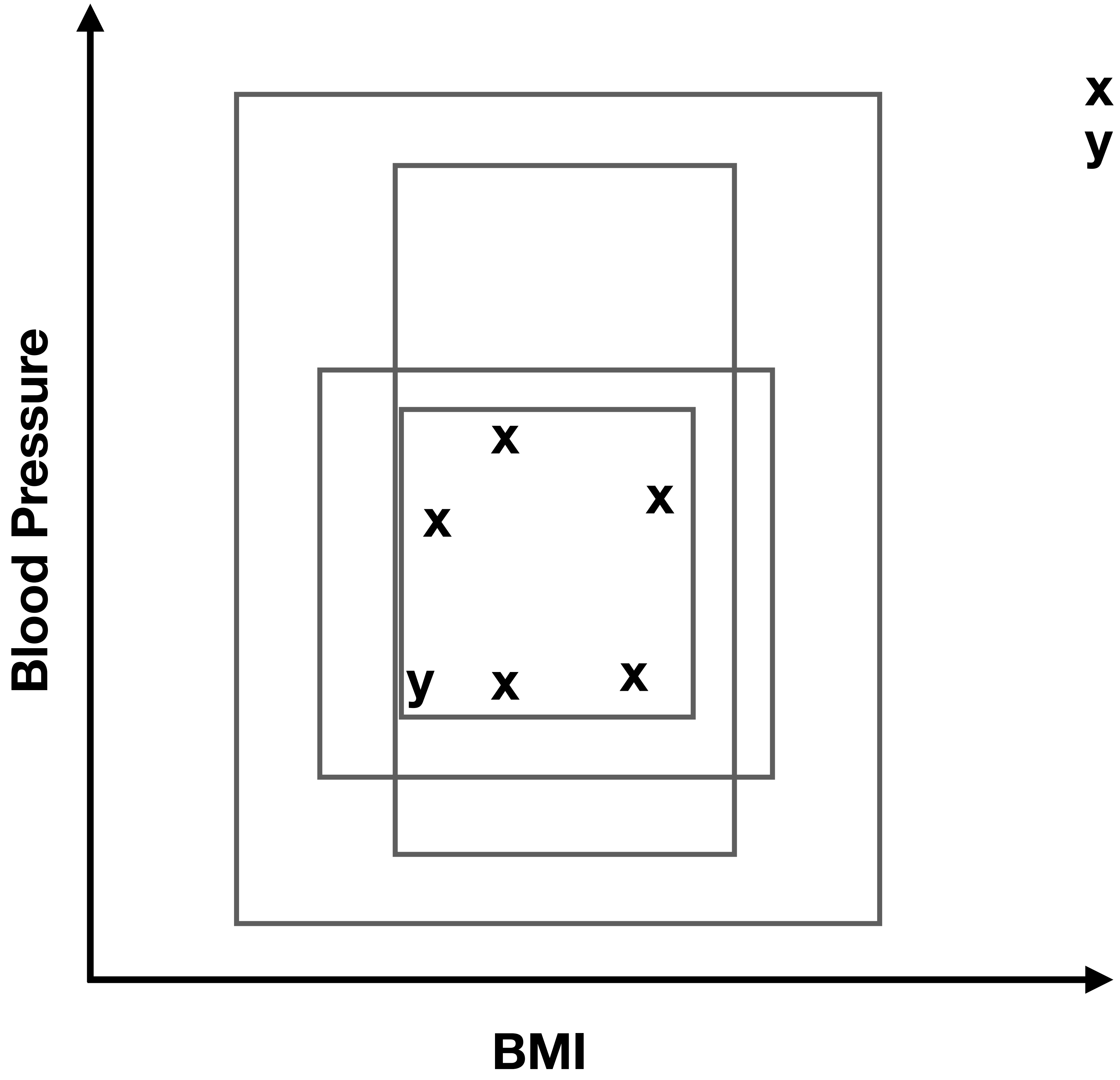
Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

Likelihood:

$$p(x|h) = \begin{cases} 1 & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases}$$

x'es generated randomly
[weak sampling].



x Data point
y new example

Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

Likelihood:

$$p(x|h) = \begin{cases} 1 & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases}$$

x'es generated randomly

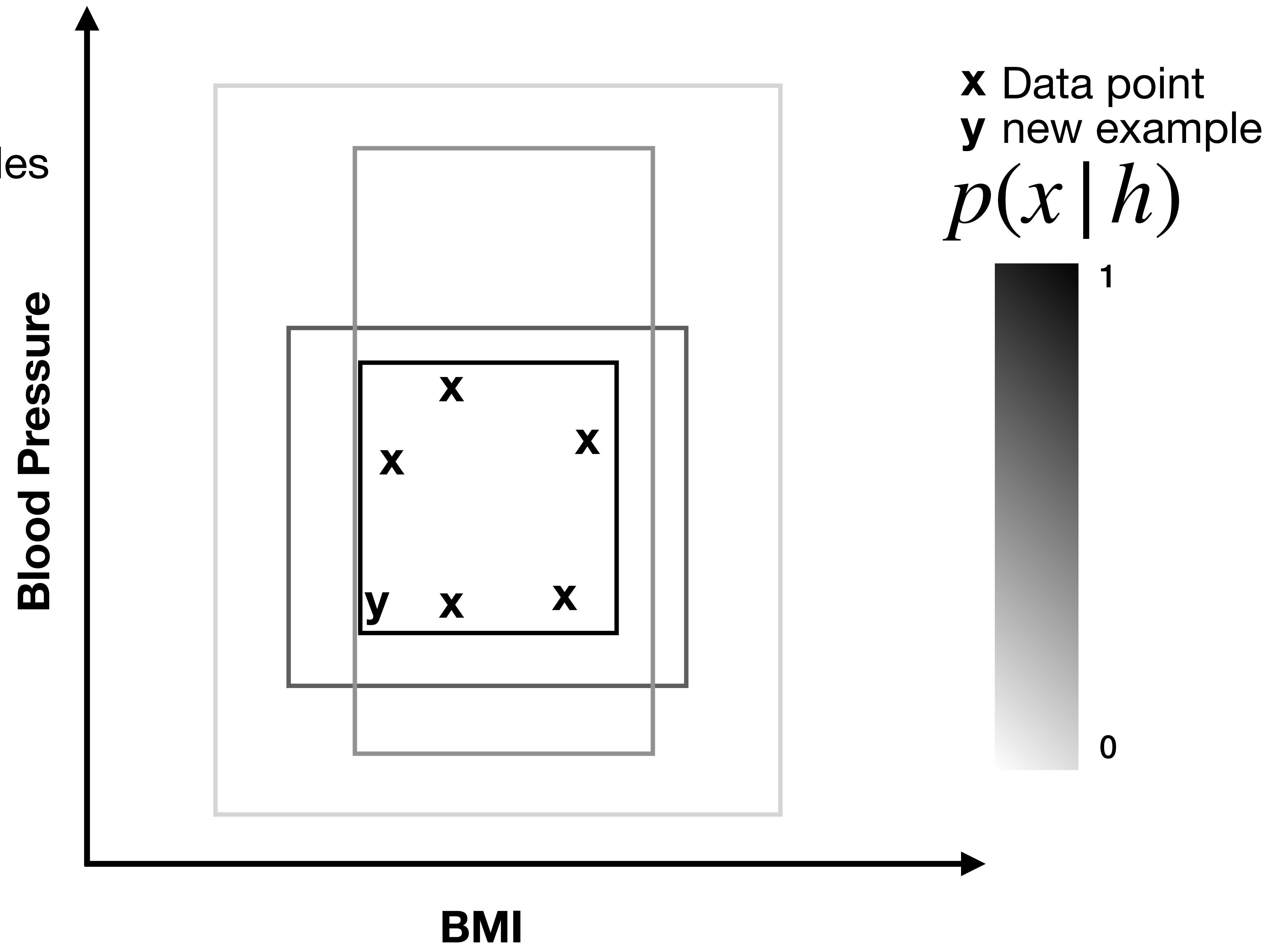
[weak sampling].

$$p(x|h) = \begin{cases} \frac{1}{|h|} & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases}$$

x'es generated to be positive examples

[strong sampling],

Bayesian size principle: under strong sampling, smaller h'es (consistent with the data) are more likely



Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

Likelihood:

$$p(x|h) = \begin{cases} 1 & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases}$$

x'es generated randomly
[weak sampling].

$$p(x|h) = \begin{cases} \frac{1}{|h|} & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases}$$

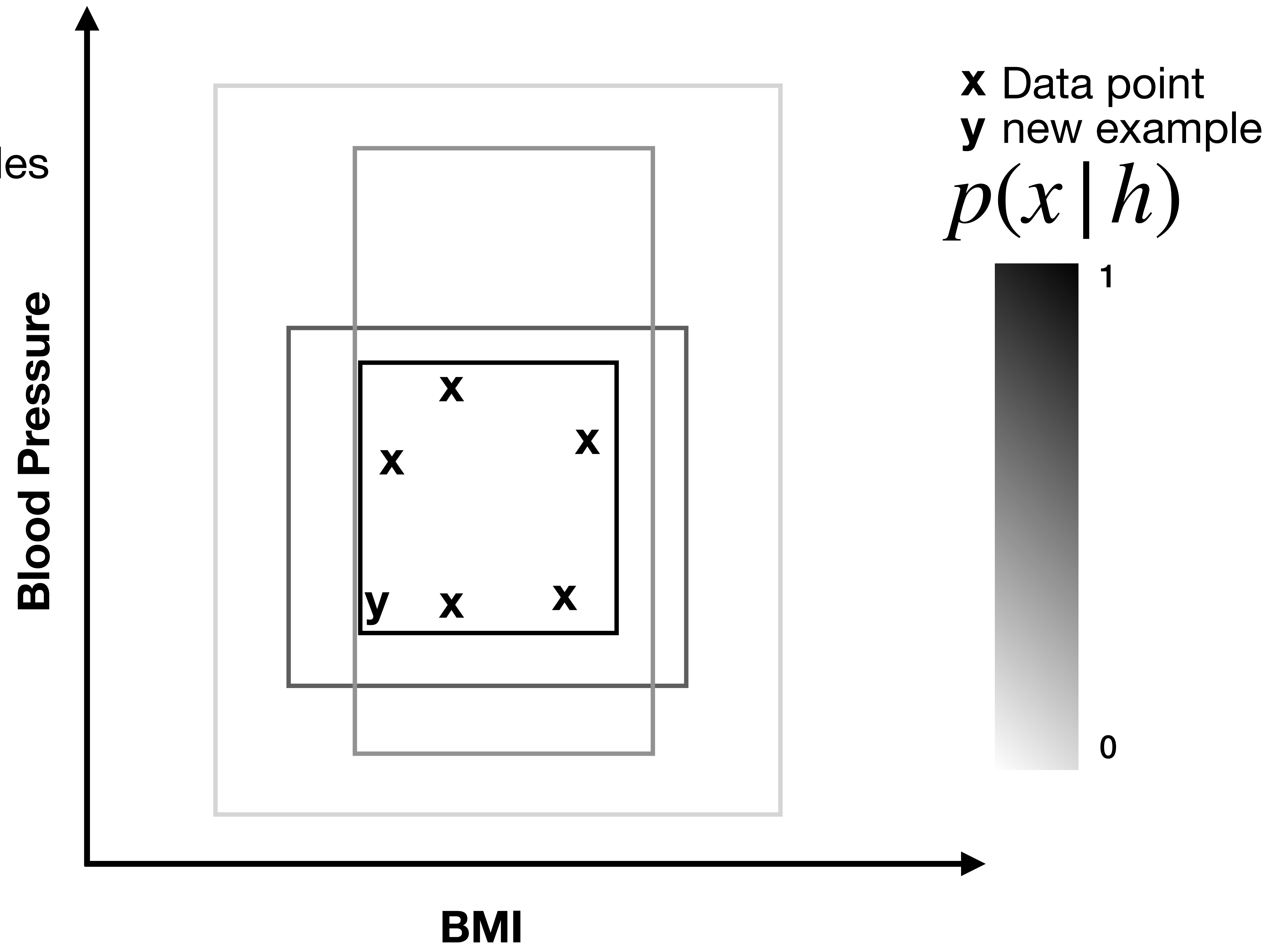
x'es generated to be positive examples
[strong sampling],

Bayesian size principle: under strong sampling, smaller h'es (consistent with the data) are more likely

Easily extended for multiple x'es with multiple features:

$$p(X|h) = \prod_i p(x_i|h)$$

$$= \begin{cases} \frac{1}{|h|^n} & \text{if } x_1, \ldots, x_n \in h \\ 0 & \text{otherwise} \end{cases}$$



Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Bayesian Concept Learning

- To summarize….

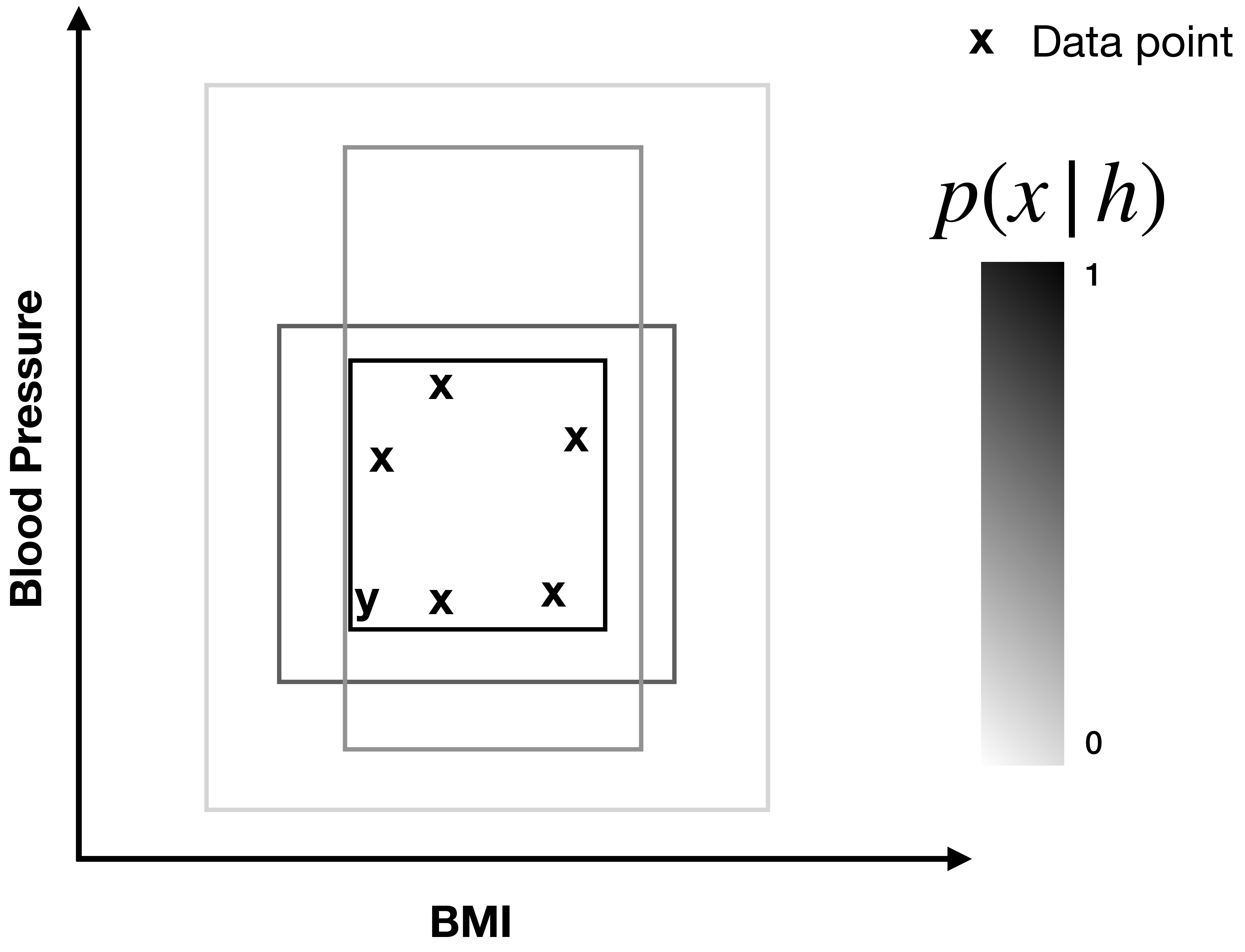
- The probability of y being in the same category of x is based on summing over all hypotheses consistent with the data

$$p(y \in C|x) = \sum_{h: y \in h} p(h|x).$$

- Where narrower hypotheses are favored under strong sampling

$$p(h|x) = \frac{p(x|h)p(h)}{p(x)}$$

$$p(x|h) = \begin{cases} \frac{1}{|h|} & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases} \qquad \text{[strong sampling]},$$



Tenenbaum (*NIPS* 1999)
Tenenbaum & Griffiths (*BBS* 2001)

Hypotheses can capture structured and arbitrary subsets of the data

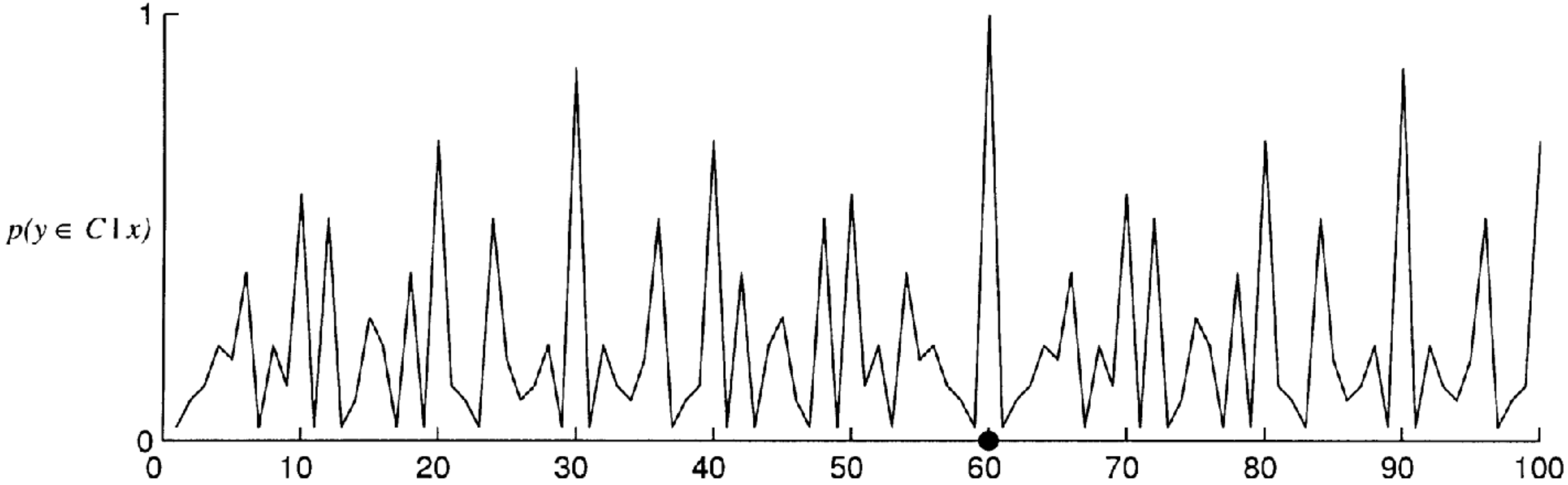
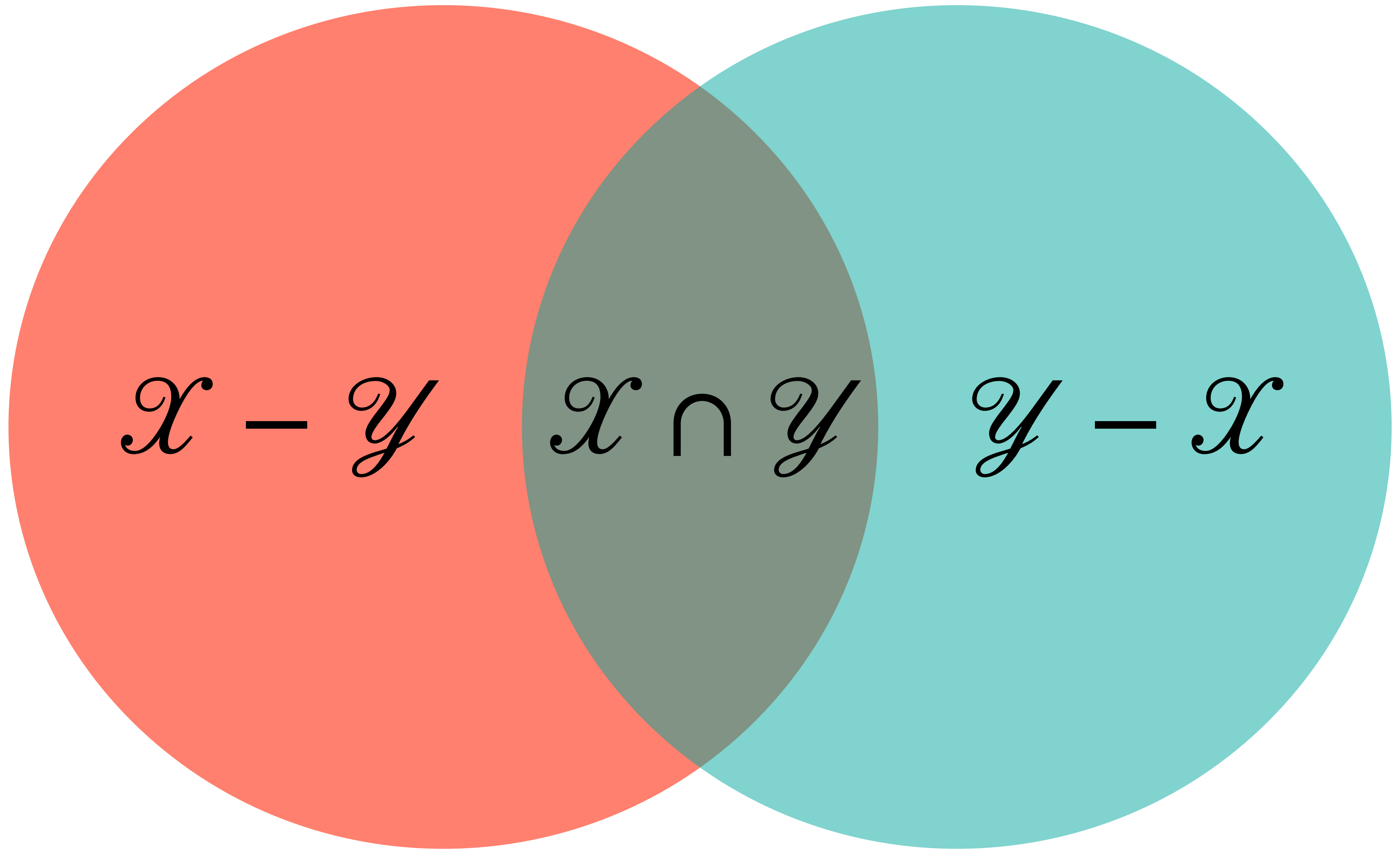
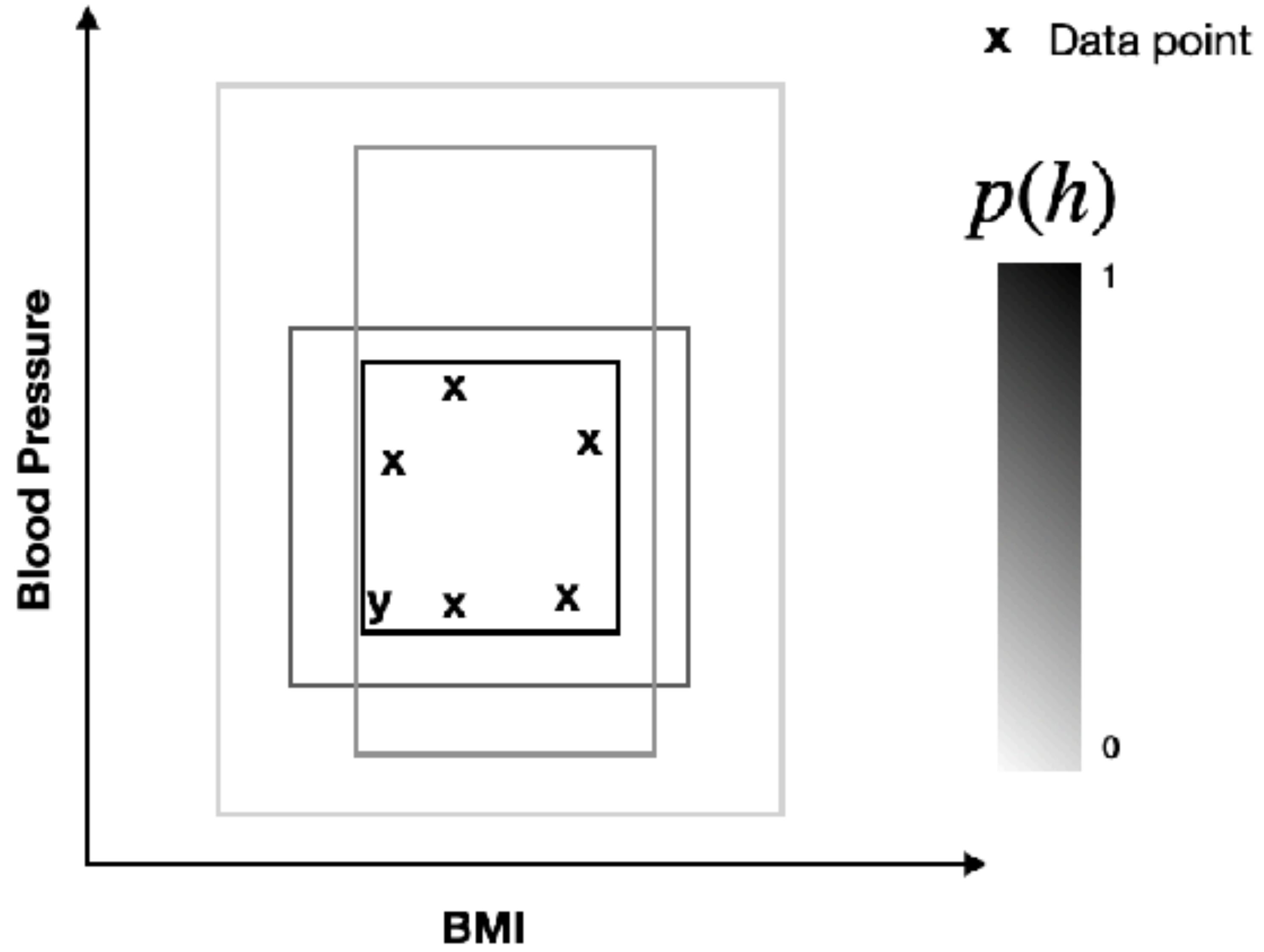


Figure 5. Bayesian generalization in the number game, given one example $x = 60$. The hypothesis space includes 33 mathematically consequential subsets (with equal prior probabilities): even numbers, odd numbers, primes, perfect squares, perfect cubes, multiples of a small number (3–10), powers of a small number (2–10), numbers ending in the same digit (1–9), numbers with both digits equal, and all numbers less than 100.

Bayesian Concept Learning Subsumes Tversky's Contrast Model





Contrast model

$$S(y,x) = \theta f(\mathcal{Y}\cap\mathcal{X}) - \alpha f(\mathcal{Y}-\mathcal{X}) - \beta f(\mathcal{X}-\mathcal{Y}),$$

Ratio model (alternative form)

$$S(y,x)=1/\left[1+\frac{\alpha f(\mathcal{Y}-\mathcal{X})+\beta f(\mathcal{X}-\mathcal{Y})}{f(\mathcal{Y}\cap\mathcal{X})}\right].$$

(equivalent when α=0 and β=1)

Bayesian concept learning

$$p(y\in C|x) = \sum_{h:y\in h} p(h|x).$$

$$=1/\left[1+\frac{\sum_{h:x\in h,y\notin h} p(h,x)}{\sum_{h:x,y\in h} p(h,x)}\right].$$

Bayesian Concept Learning Extends Shepard's Law of Generalization to Multiple Examples

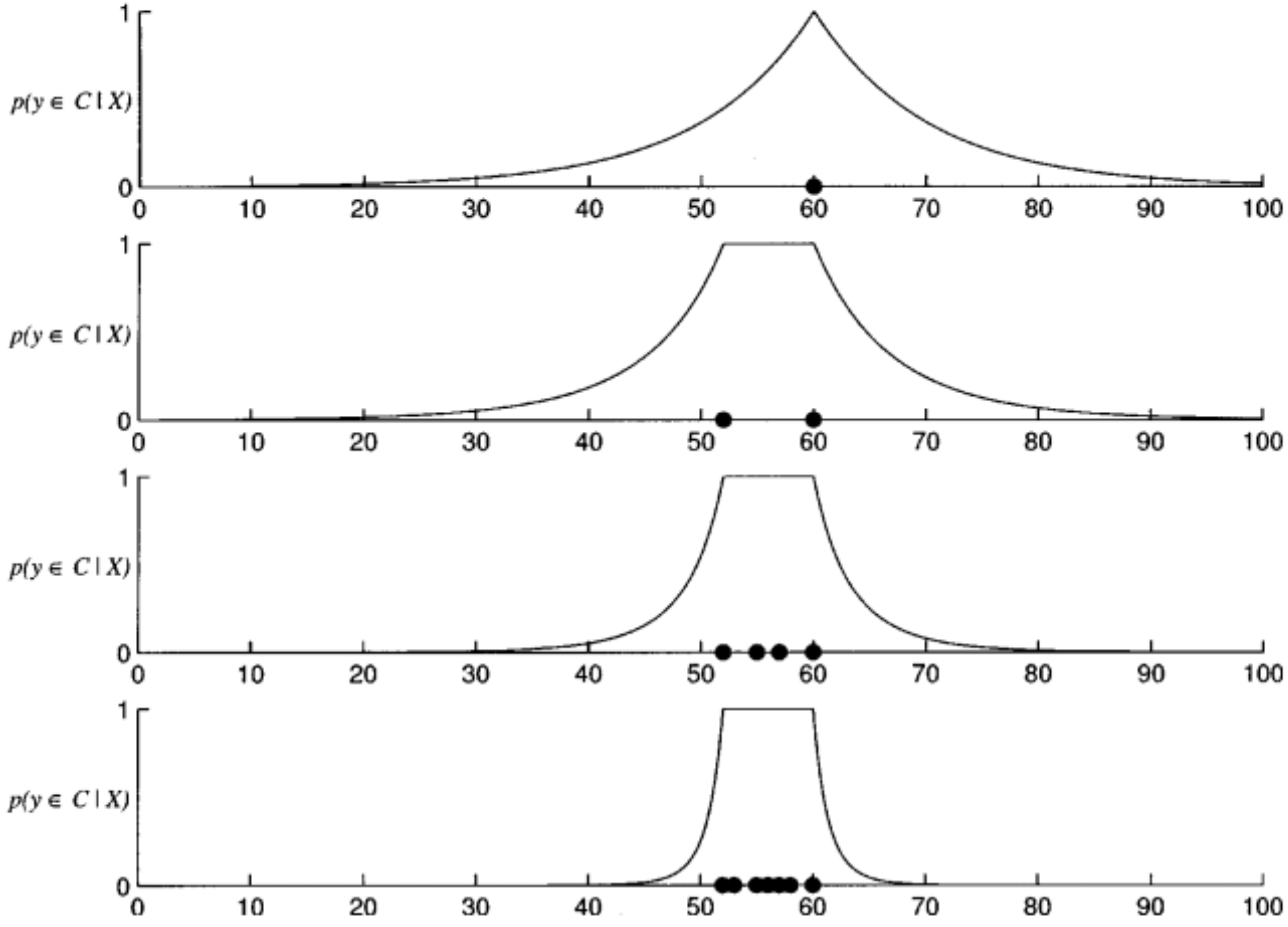
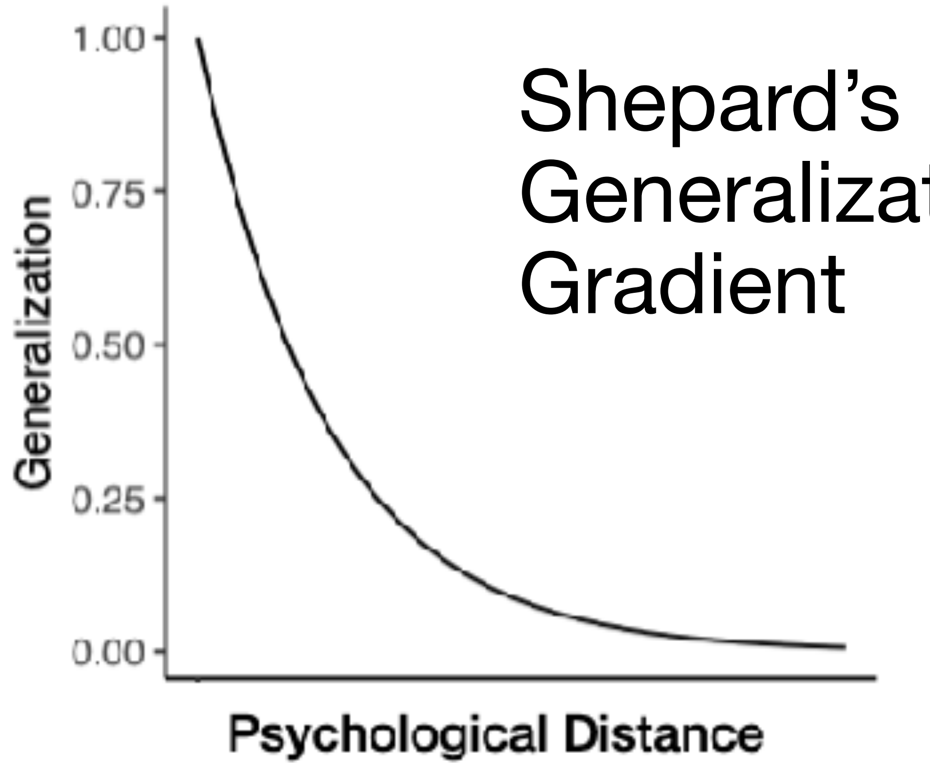
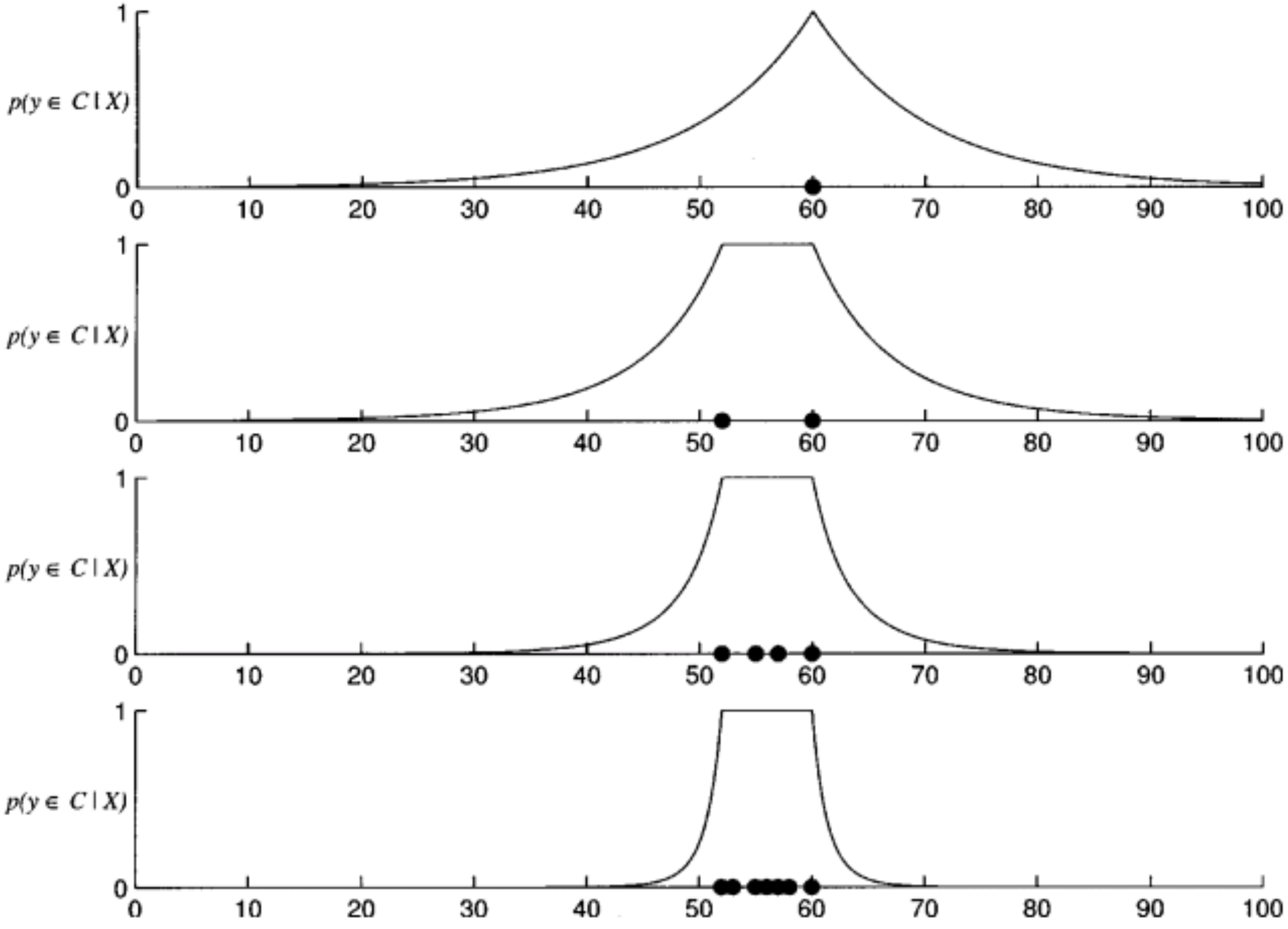


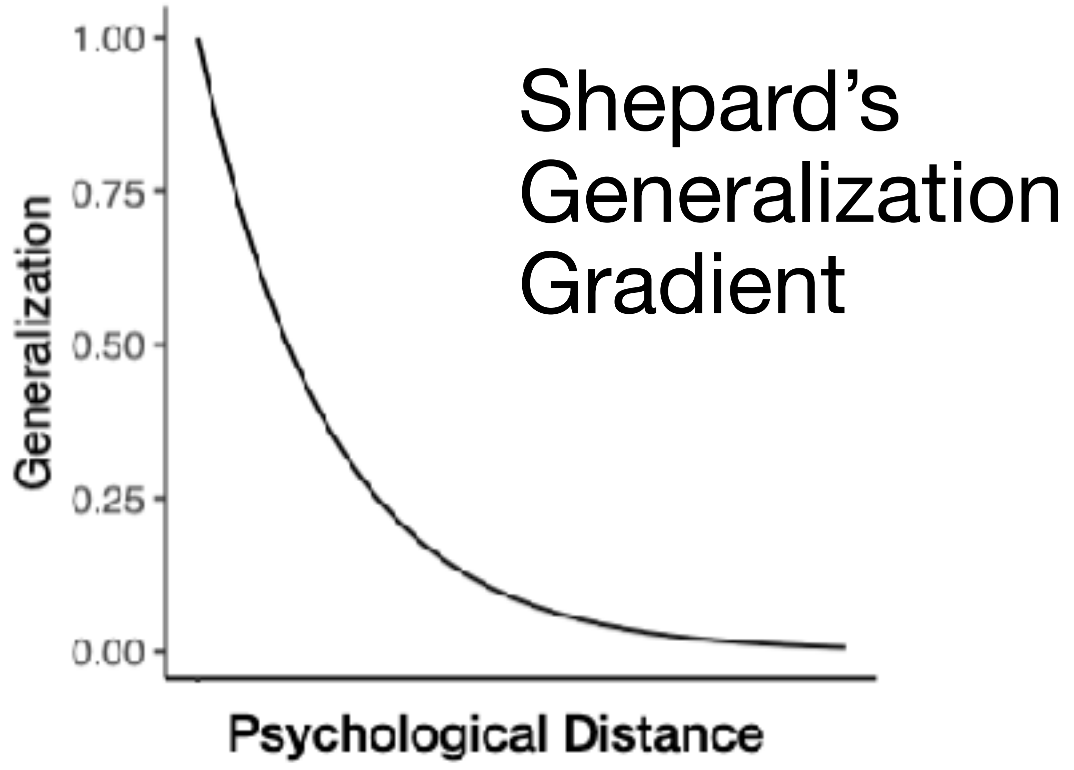
Figure 3. The effect of the number of examples on Bayesian generalization (under the assumptions of strong sampling and an Erlang prior, $\mu = 10$). Filled circles indicate examples. The first curve is the gradient of generalization with a single example, for the purpose of comparison. The remaining graphs show that the range of generalization decreases as a function of the number of examples.



Shepard's Generalization Gradient

Bayesian Concept Learning Extends Shepard's Law of Generalization to Multiple Examples



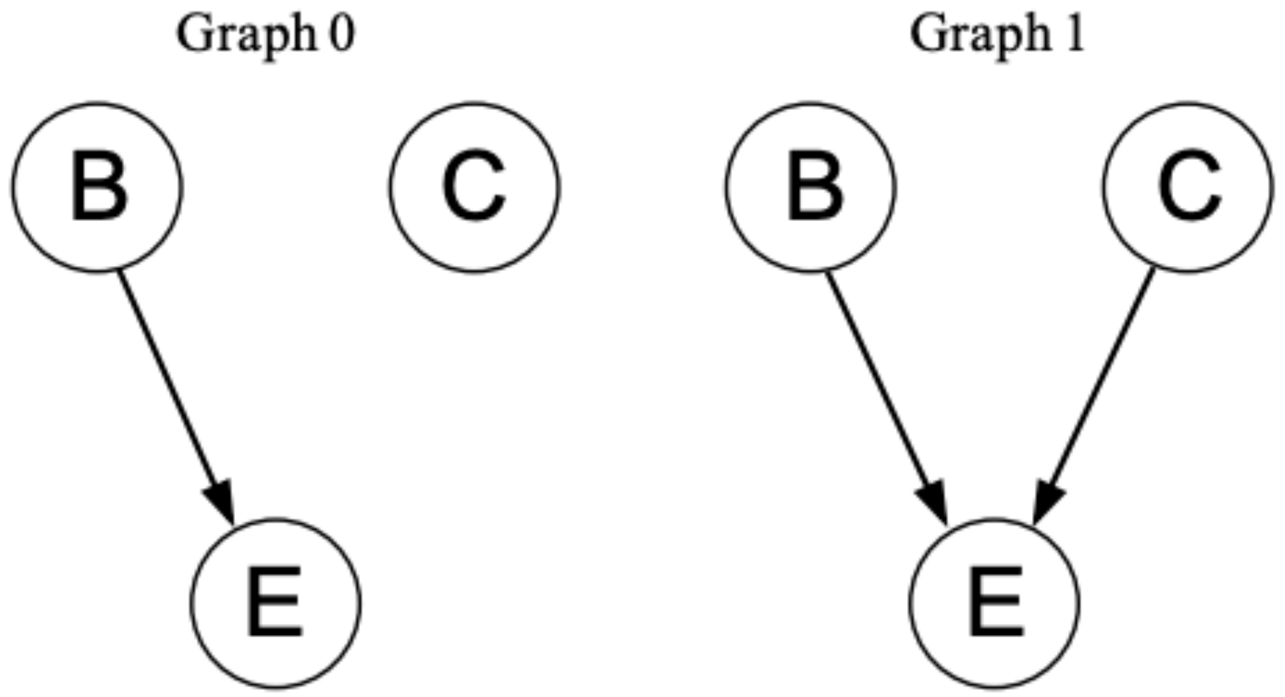


Range of generalization decreases with more examples

more examples = less uncertainty about the extent of consequential region

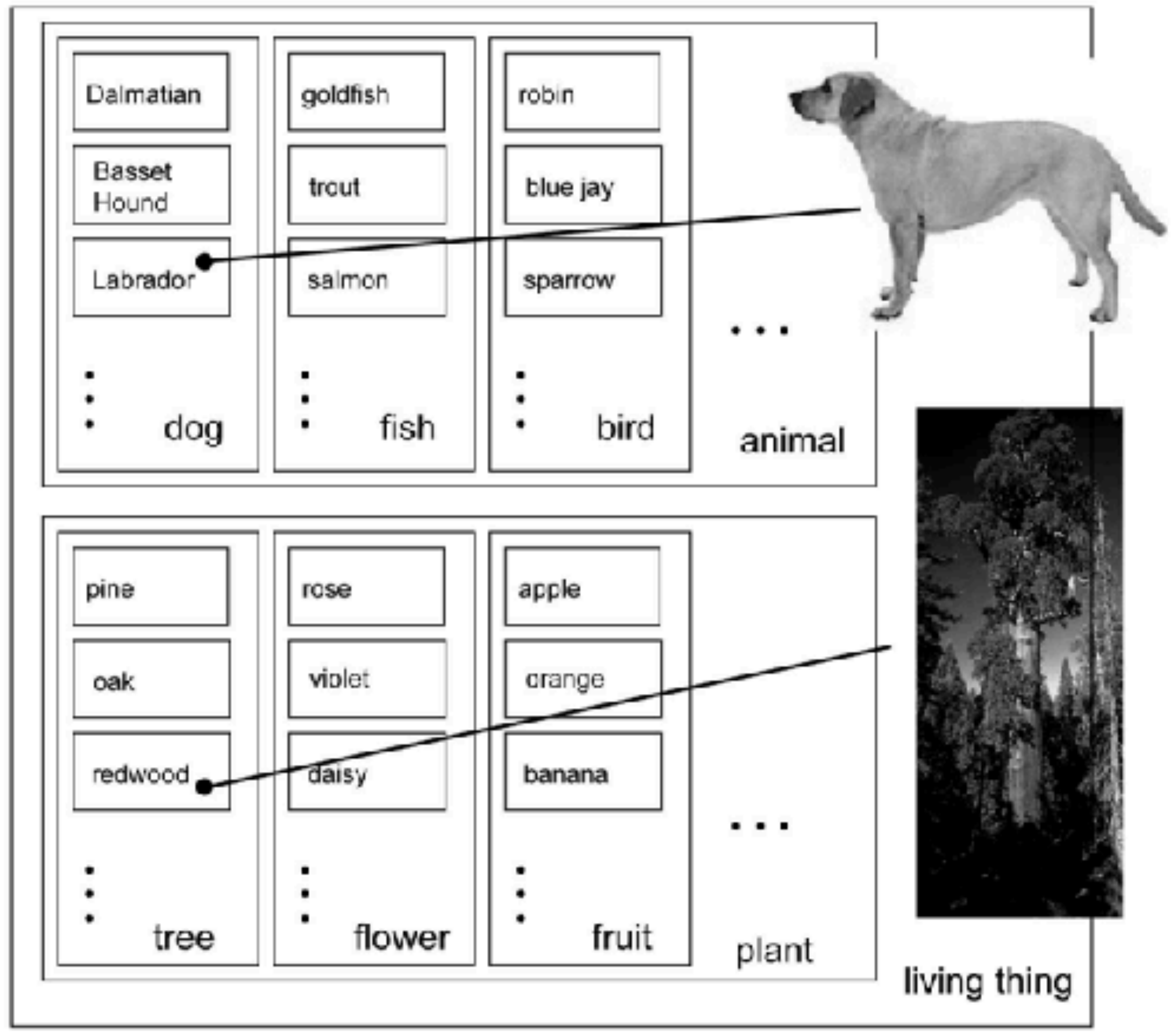
Figure 3. The effect of the number of examples on Bayesian generalization (under the assumptions of strong sampling and an Erlang prior, $\mu = 10$). Filled circles indicate examples. The first curve is the gradient of generalization with a single example, for the purpose of comparison. The remaining graphs show that the range of generalization decreases as a function of the number of examples.

Causal learning



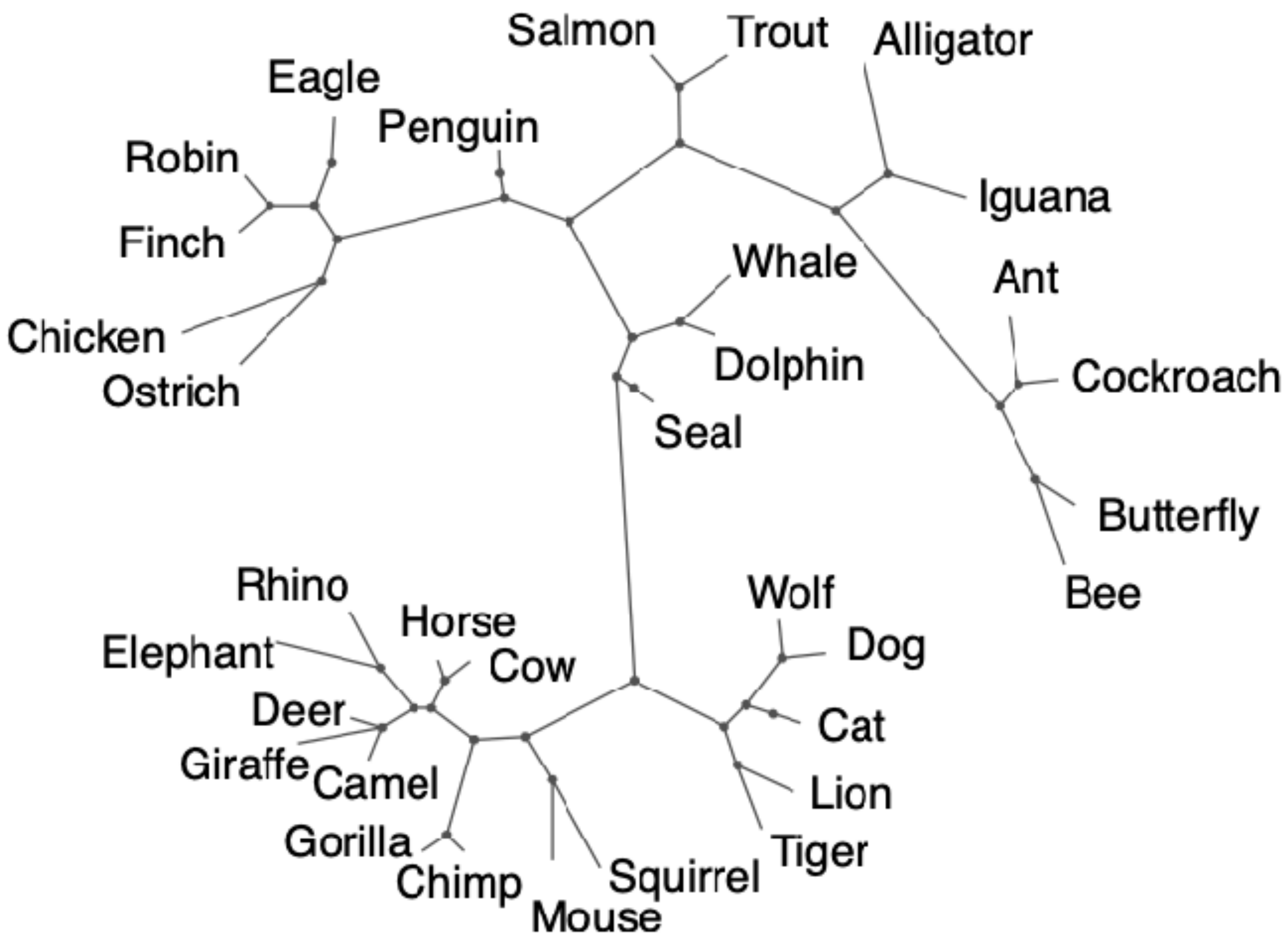
Griffiths & Tenenbaum (2005)

Word learning



Xu & Tenenbaum (2007)

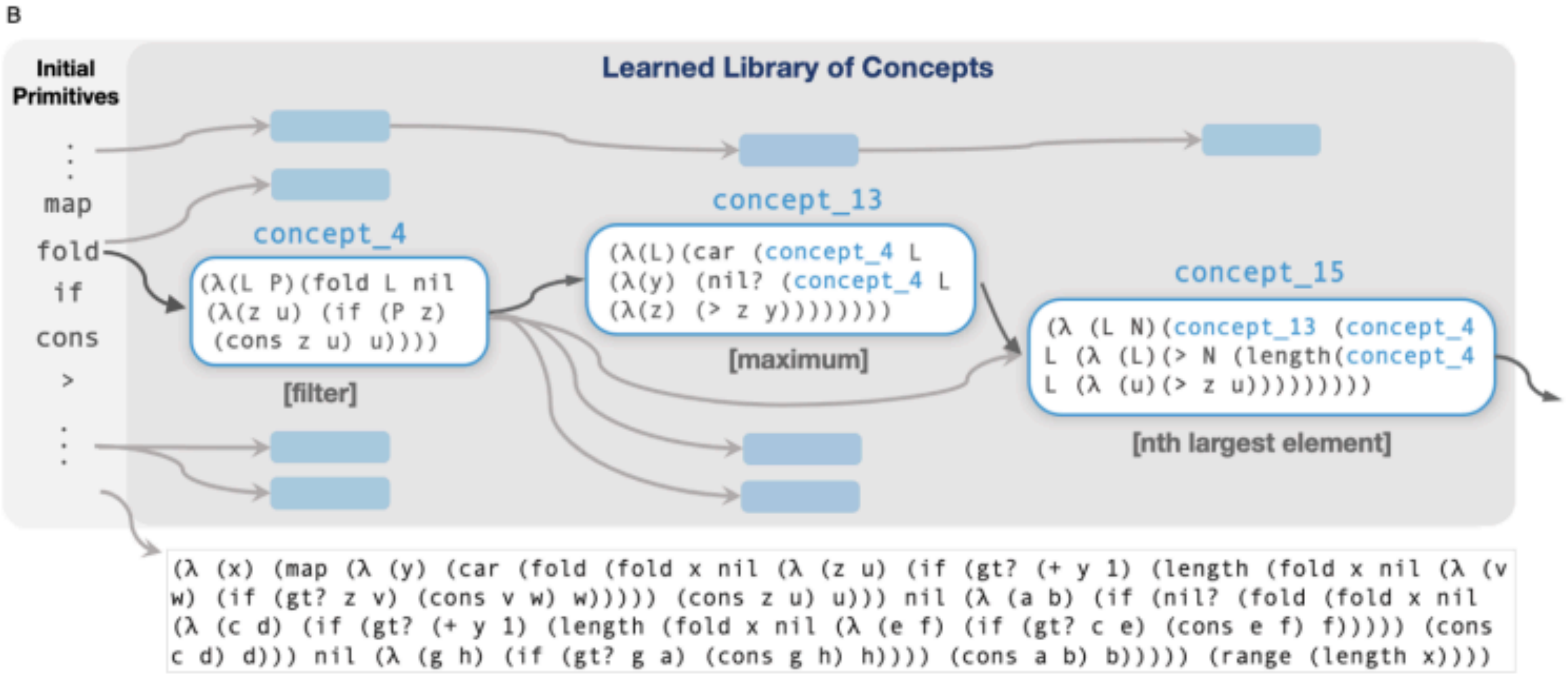
Structure learning



Kemp & Tenenbaum (2008)

Program Induction

Lake, Salakhutdinov, & Tenenbaum (2015)

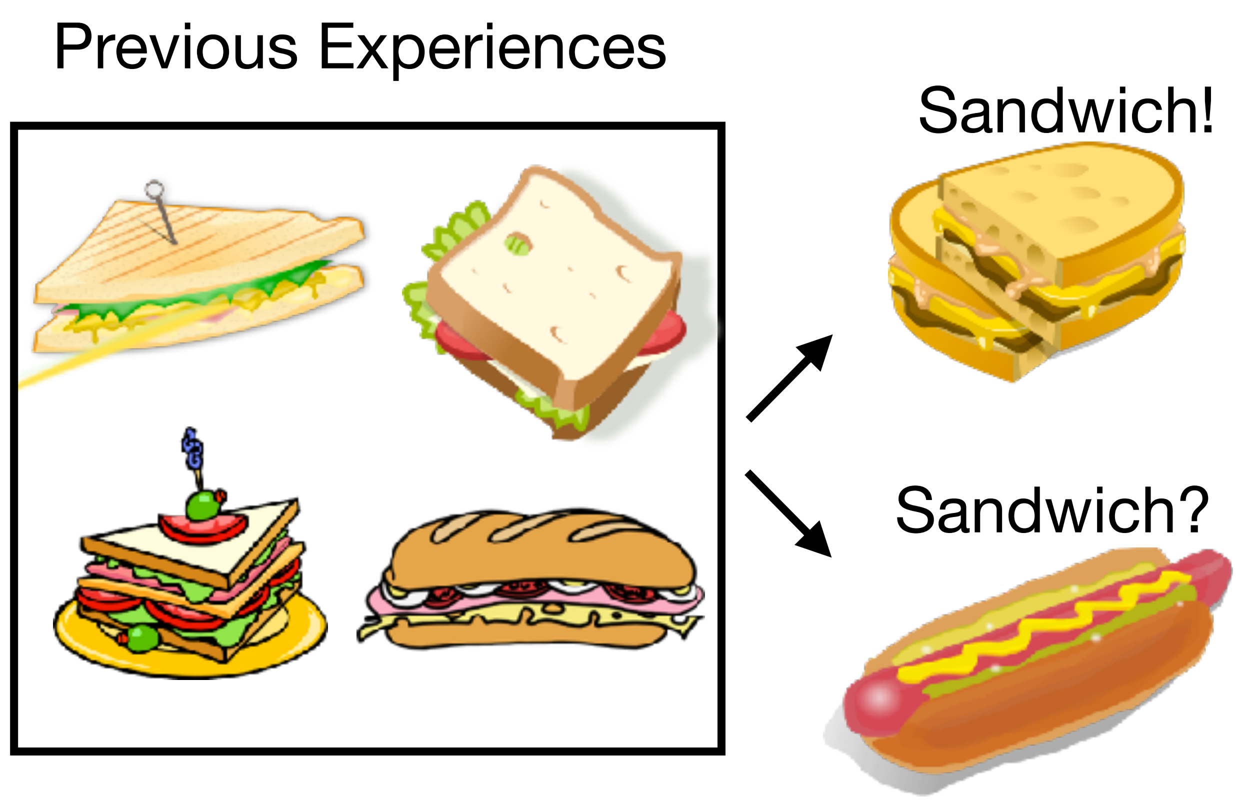


Dreamcoder: Ellis et al,. (2020)

... and many more

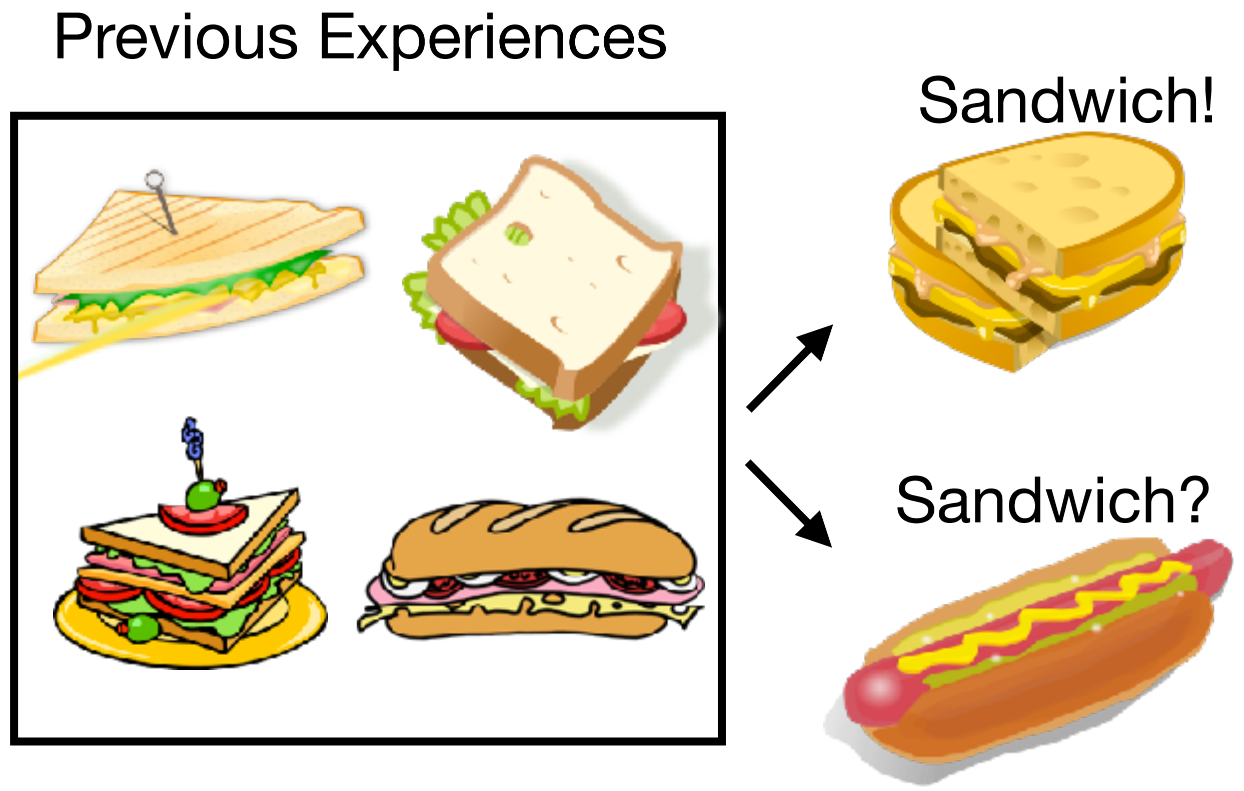
Theories of Concept Learning

Classification task

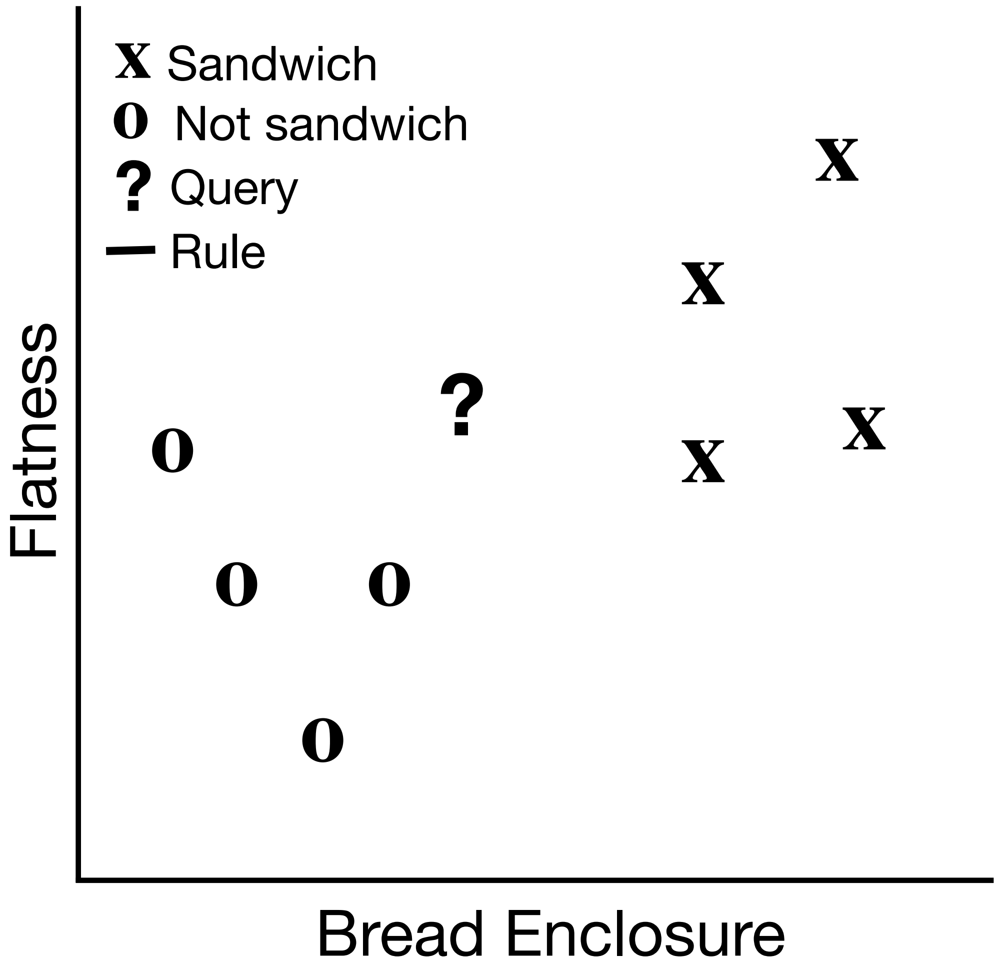


Theories of Concept Learning

Classification task

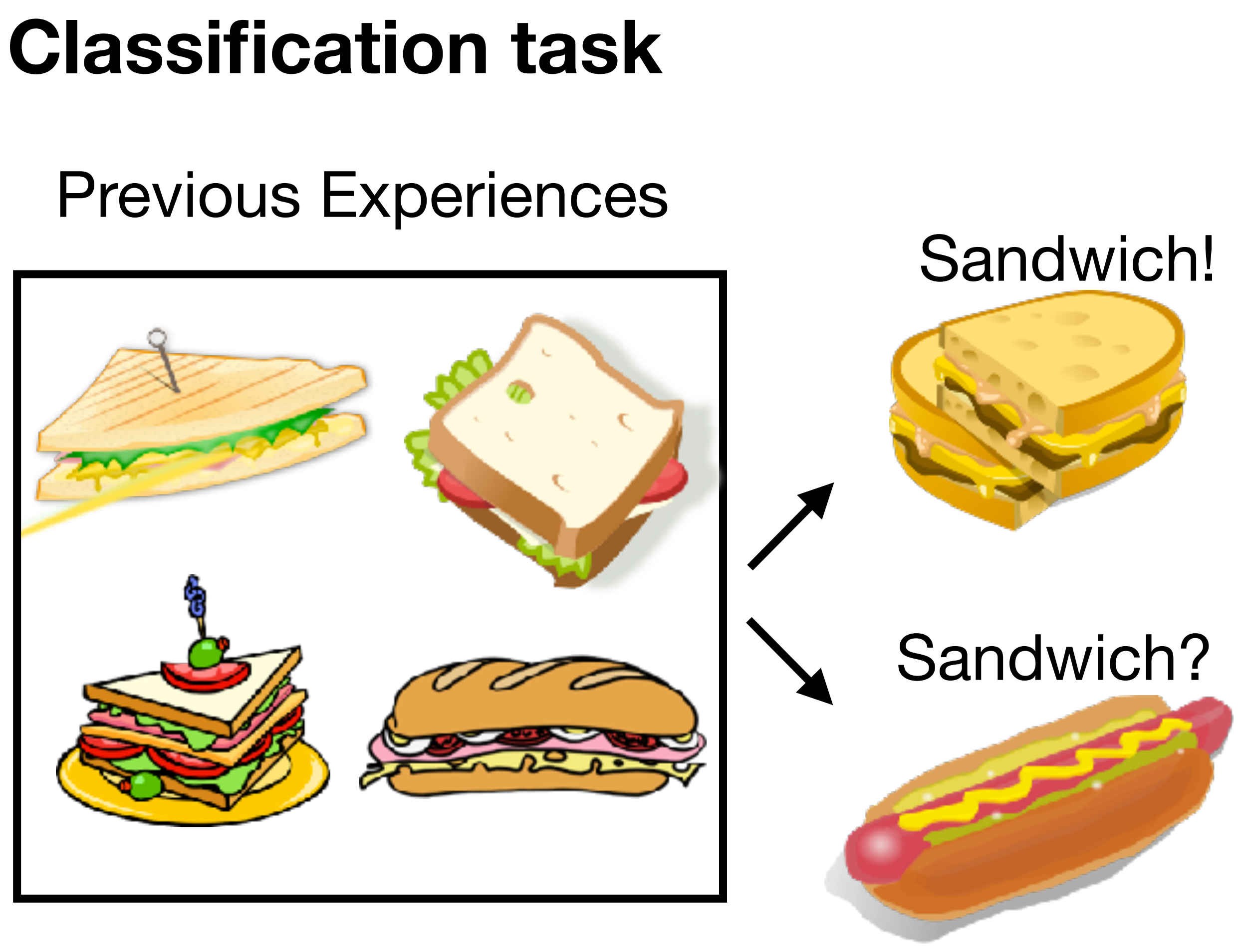


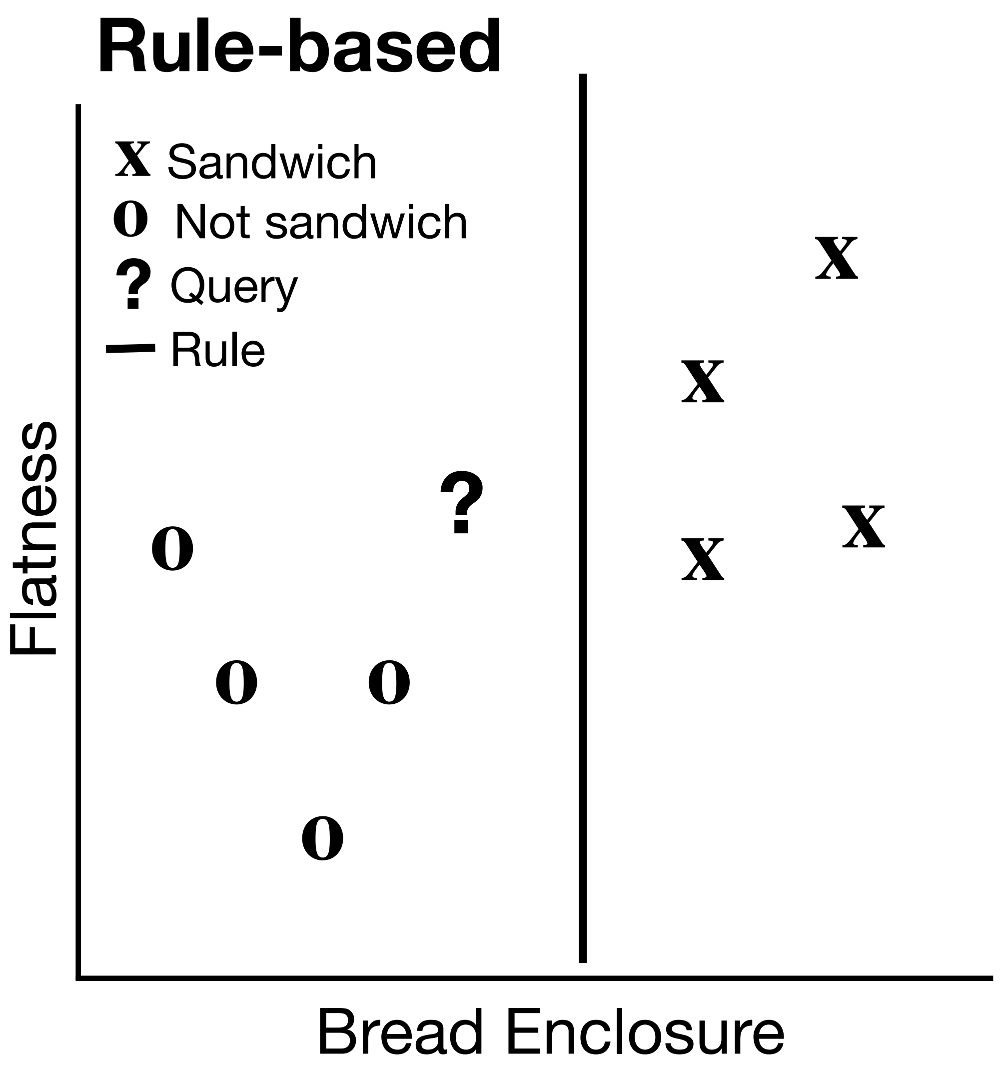
Rule-based



- *Rules* describe the explicit boundaries of category boundaries
  (Smith & Medin, 1981; Ashby & Gott, *JEP:LMC* 1988)

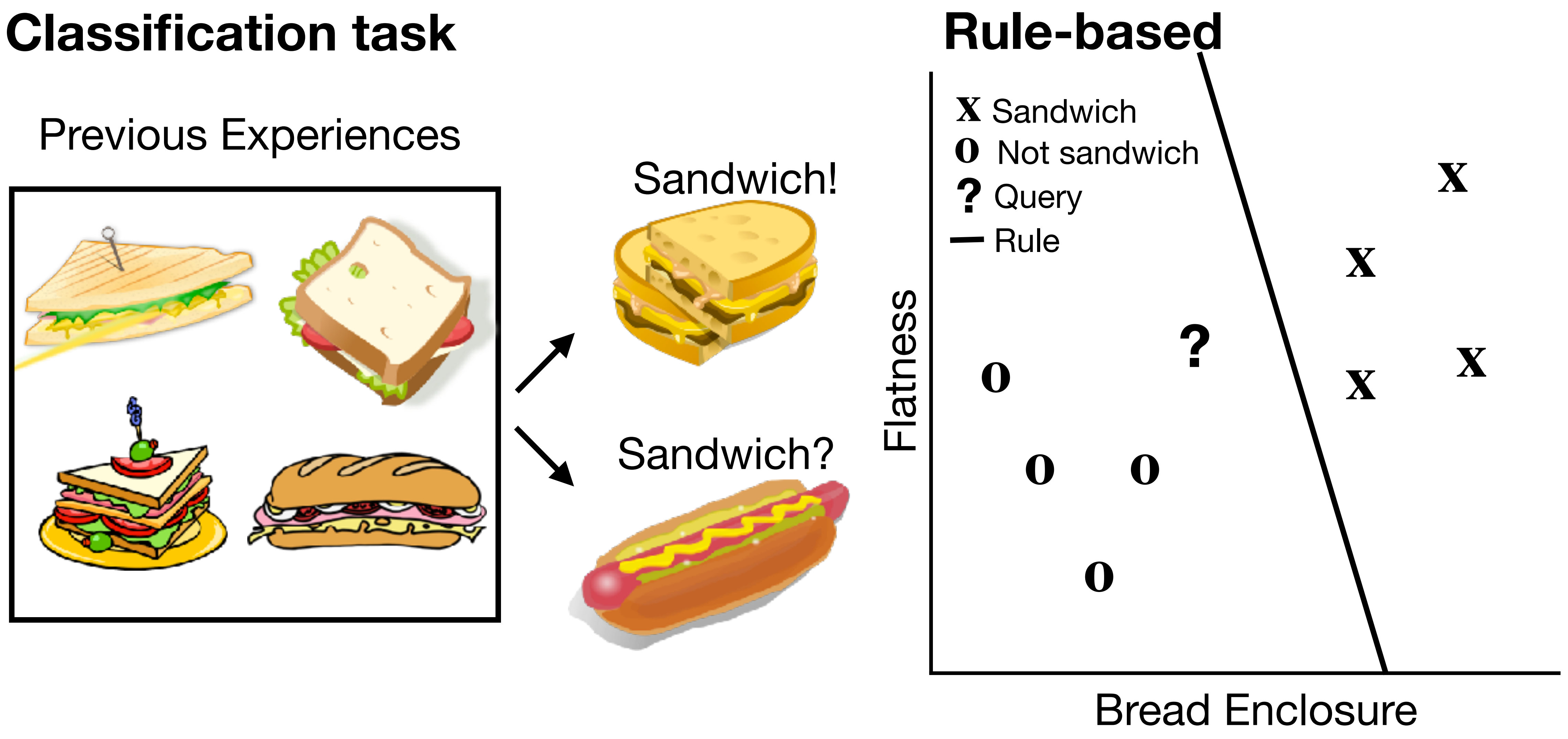
Theories of Concept Learning





- *Rules* describe the explicit boundaries of category boundaries
(Smith & Medin, 1981; Ashby & Gott, *JEP:LMC* 1988)

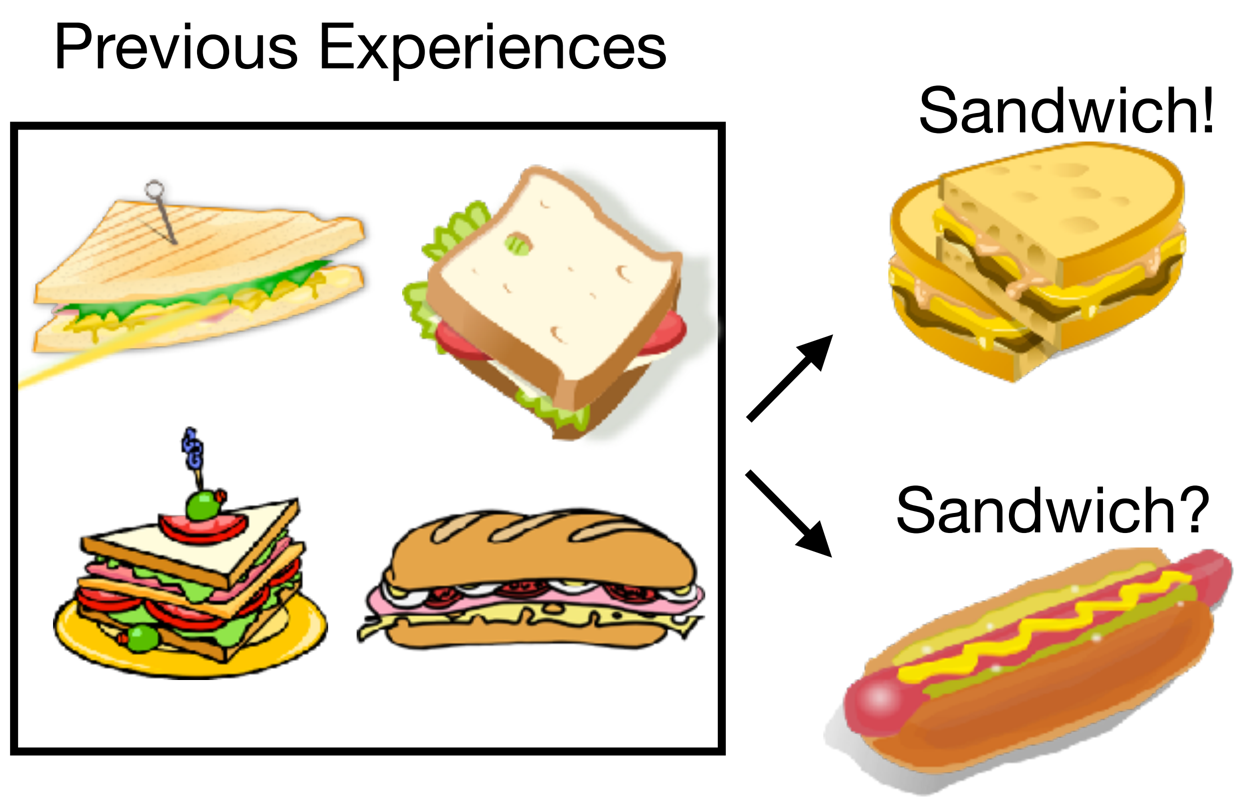
Theories of Concept Learning



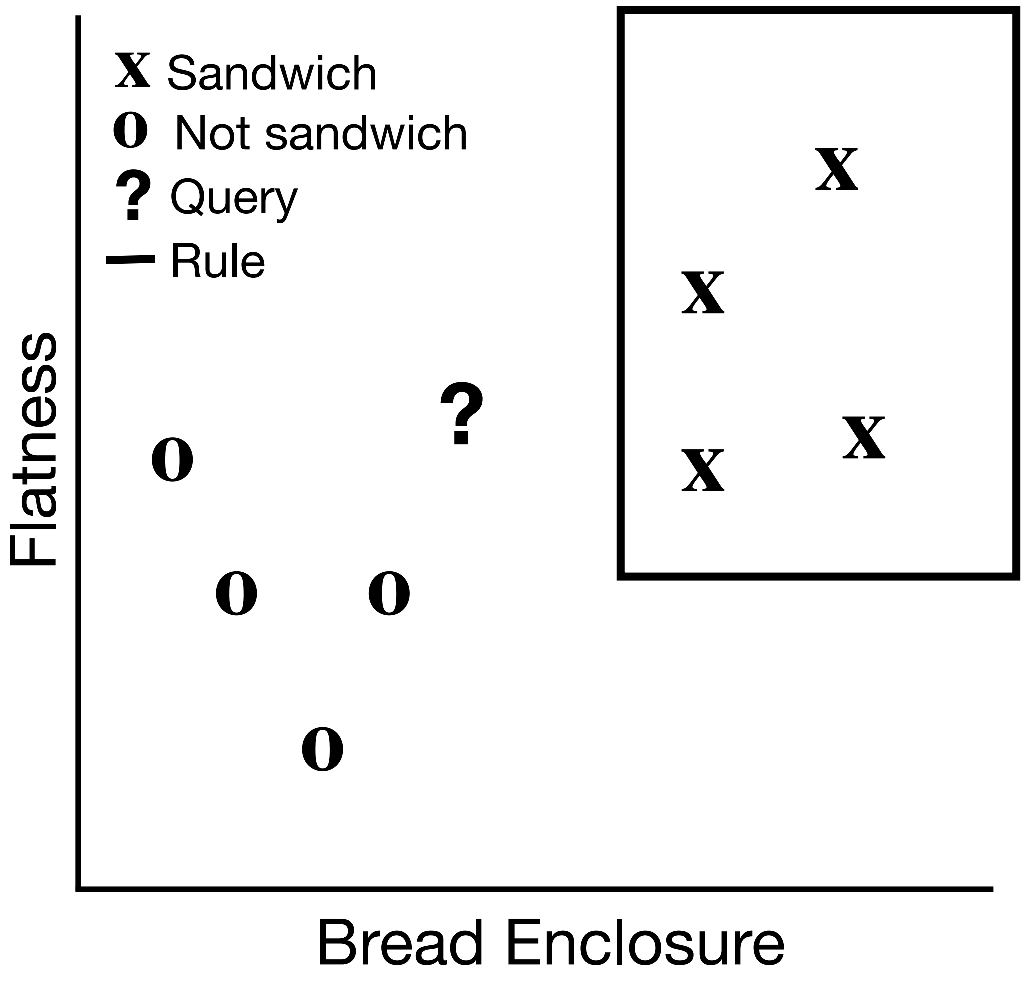
- *Rules* describe the explicit boundaries of category boundaries
(Smith & Medin, 1981; Ashby & Gott, *JEP:LMC* 1988)

Theories of Concept Learning

Classification task

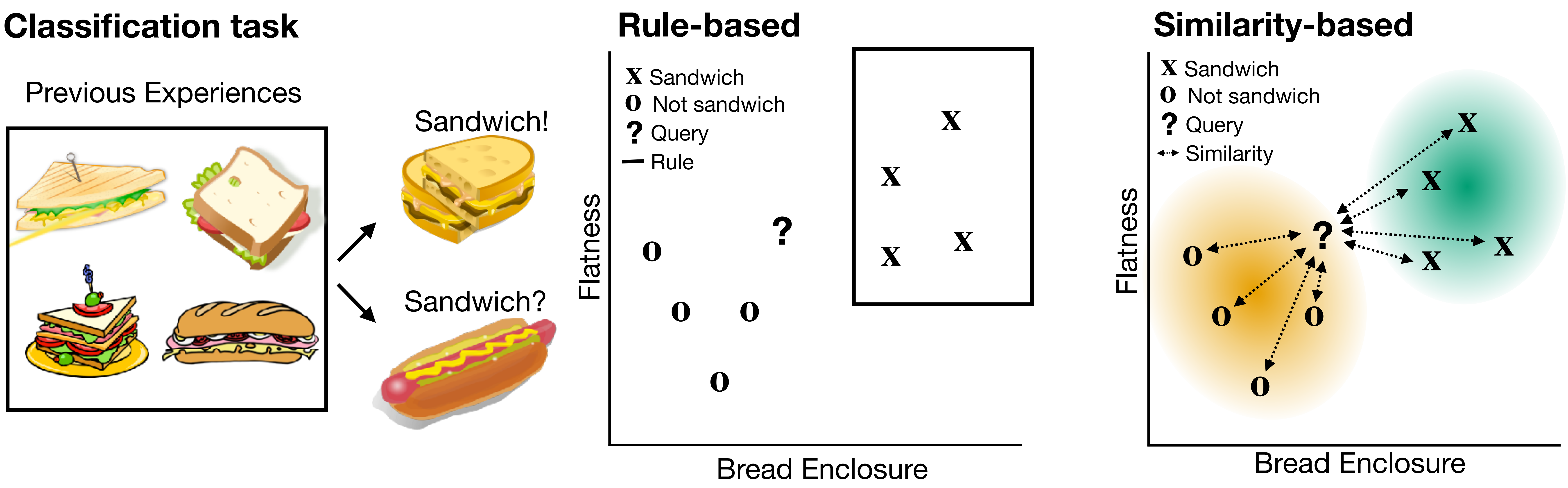


Rule-based



- *Rules* describe the explicit boundaries of category boundaries
(Smith & Medin, 1981; Ashby & Gott, *JEP:LMC* 1988)

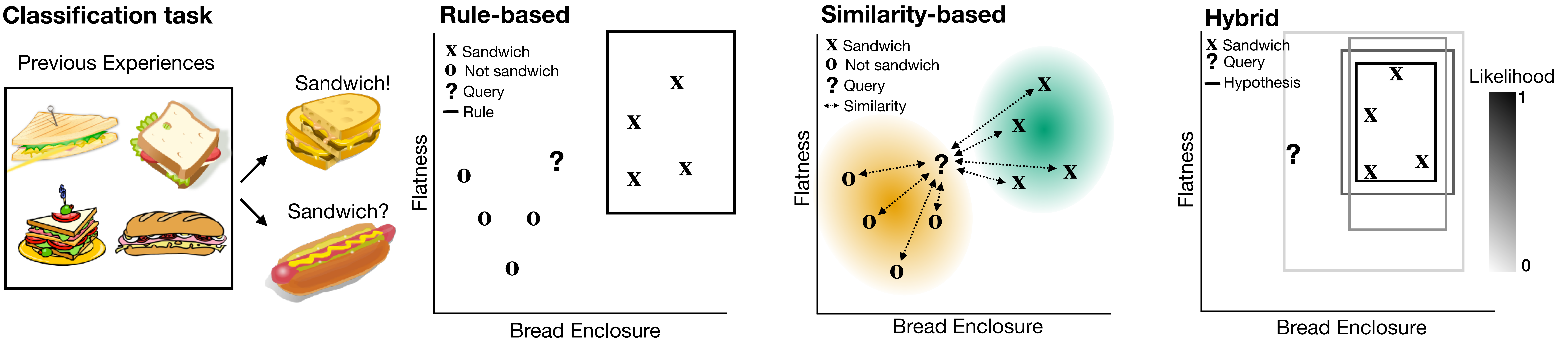
Theories of Concept Learning



- *Rules* describe the explicit boundaries of category boundaries
  (Smith & Medin, 1981; Ashby & Gott, *JEP:LMC* 1988)

- *Similarity* uses a comparison to previously encountered exemplars or a learned prototype (aggregated over multiple experiences) as the basis of generalization
  (Rosch, *CogPsy* 1973; Medin & Schaffer, *PsychRev* 1978 ; Nosofsky, *JEP:G* 1986; Smith & Minda *JEP:LMC* 1998)

Theories of Concept Learning



- *Rules* describe the explicit boundaries of category boundaries
(Smith & Medin, 1981; Ashby & Gott, *JEP:LMC* 1988)

- *Similarity* uses a comparison to previously encountered exemplars or a learned prototype (aggregated over multiple experiences) as the basis of generalization
(Rosch, *CogPsy* 1973; Medin & Schaffer, *PsychRev* 1978 ; Nosofsky, *JEP:G* 1986; Smith & Minda *JEP:LMC* 1998)

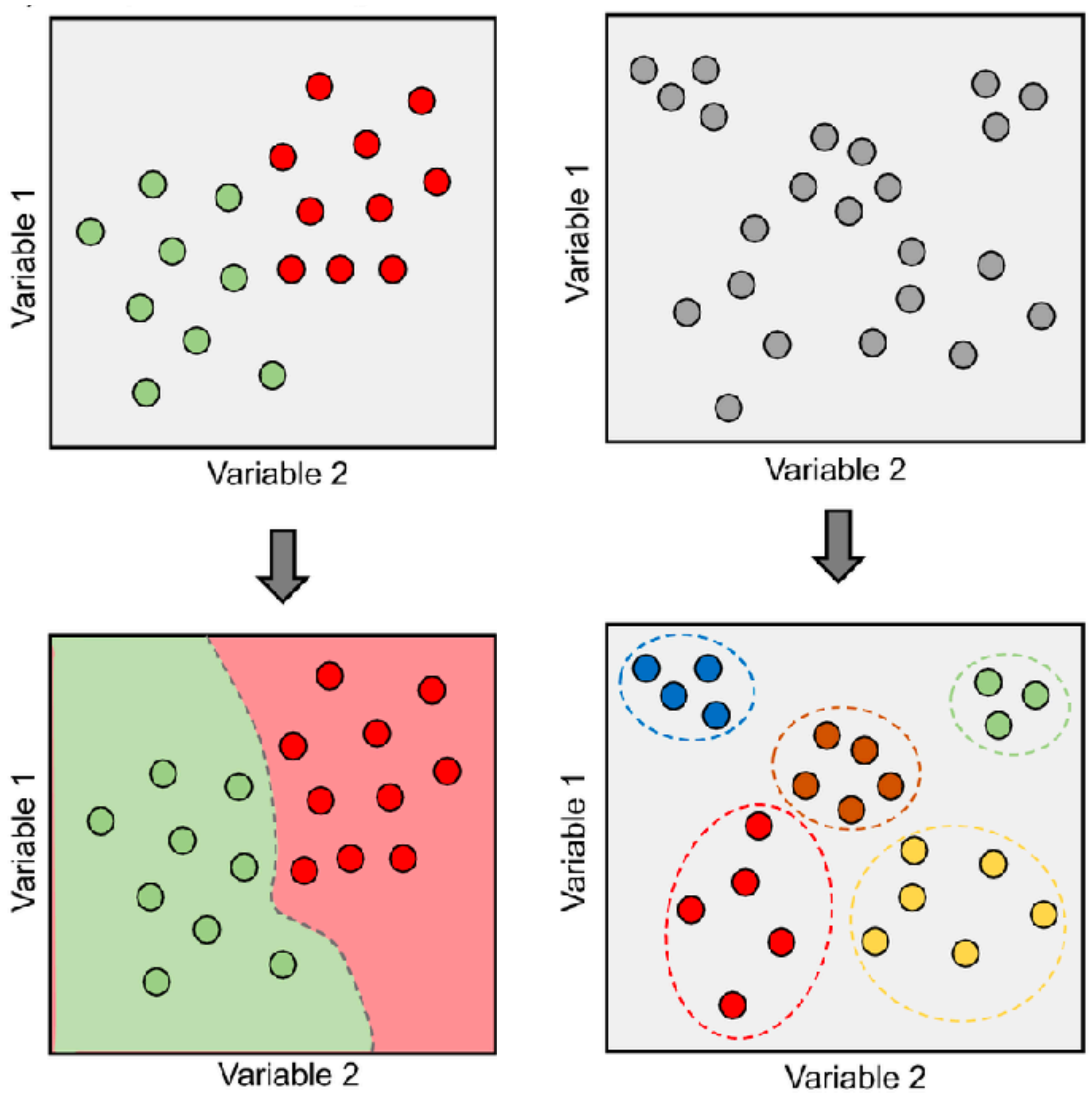
- *Hybrids* combine elements of both: Bayesian concept learning uses a distribution over rules, while reproducing predictions of two influential similarity-based approaches
(Tenenbaum & Griffiths, *BBS* 2001; Shepard, *Science* 1987; Tversky, *PsychRev* 1977)

General principles

- Again, hybrid theories combining competiting mechanisms seem to provide the best answer
    - **Rules** have a symbolic flavor, offering rapid generalization and flexible composition
    - **Similarity** has a subsymbolic flavor, where previously encountered example exert influence on generalization based on similarity-weights
    - A hybrid using Bayesian inference combines the best of both worlds
- Concepts are not just passively learned associations (model-free RL), but seem to point towards generative representations about the structure of the world (Model-based RL)

Next weeks

Supervised and unsupervised learning



Function learning