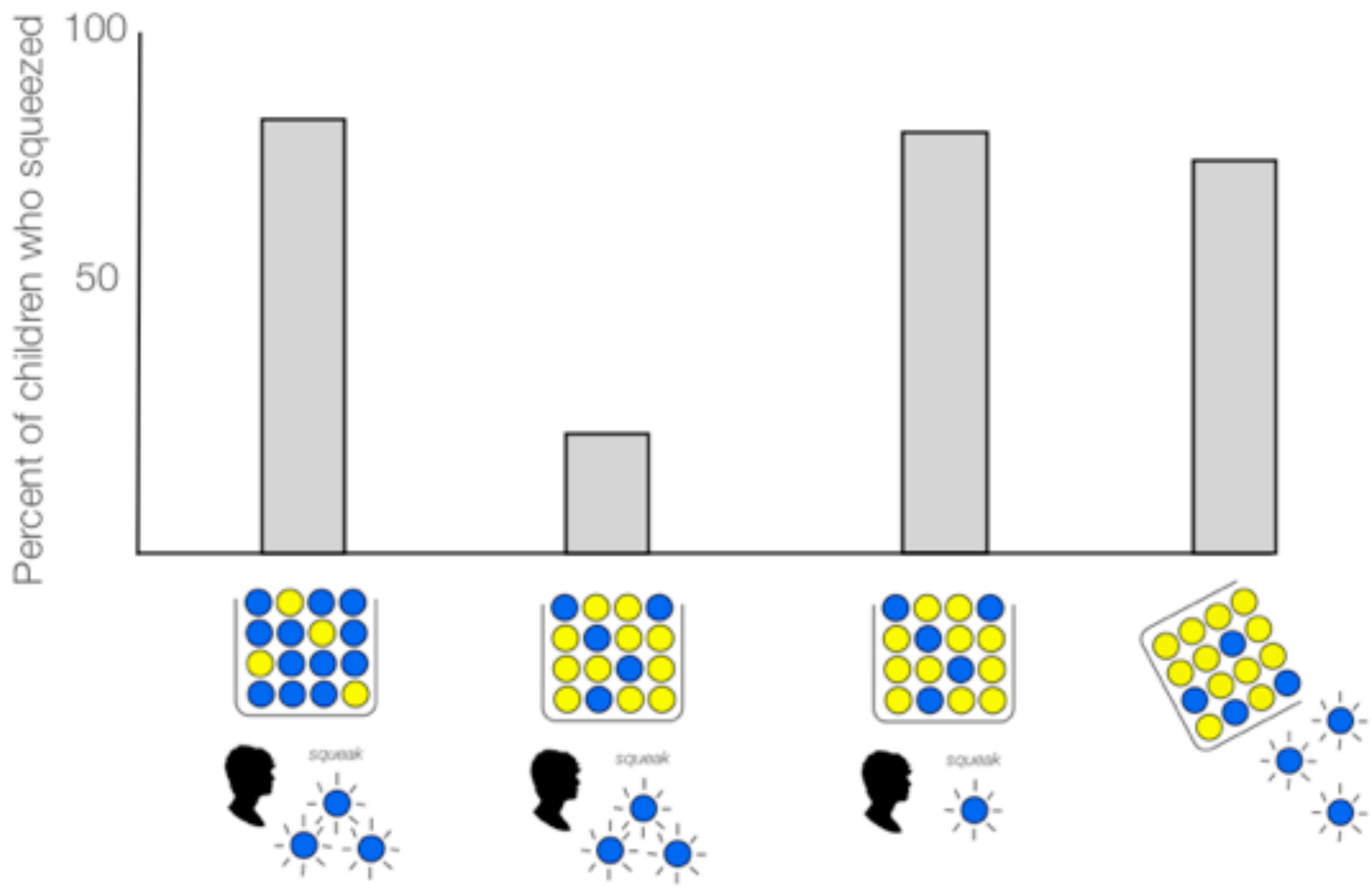
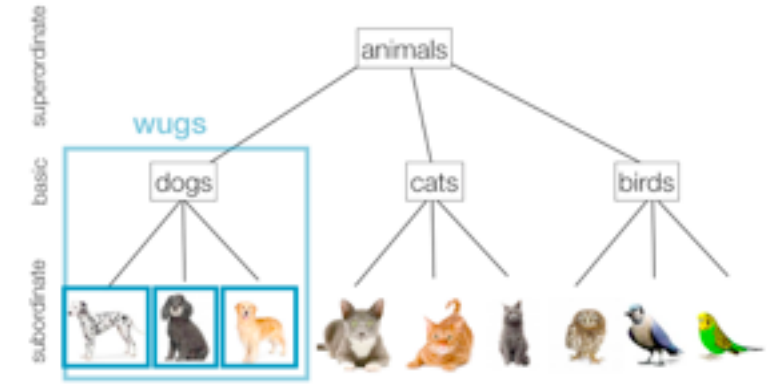
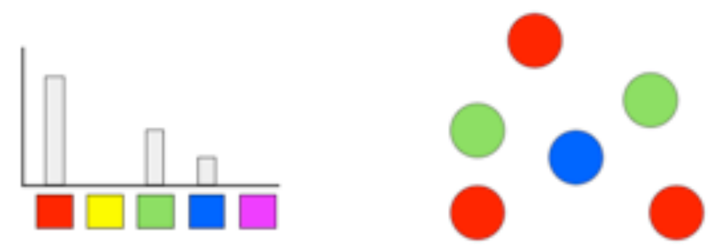
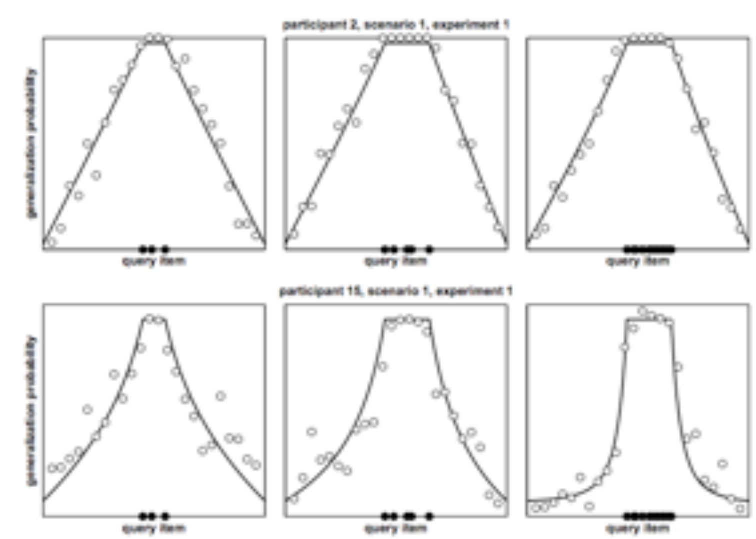
Computational Cognitive Science









Lecture 20: Strong vs weak sampling

Where we are

- So far in CCS we’ve learned about how people (and models) might learn and represent information about sequences of events or words...
- As well as how they can learn concepts that don’t change, or incorporate an element of time
- Hidden within all of this have been certain implicit assumptions about where and how this data is generated, and what kind of information people get
- These assumptions have driven the models so far but in the next few lectures we’ll revisit them

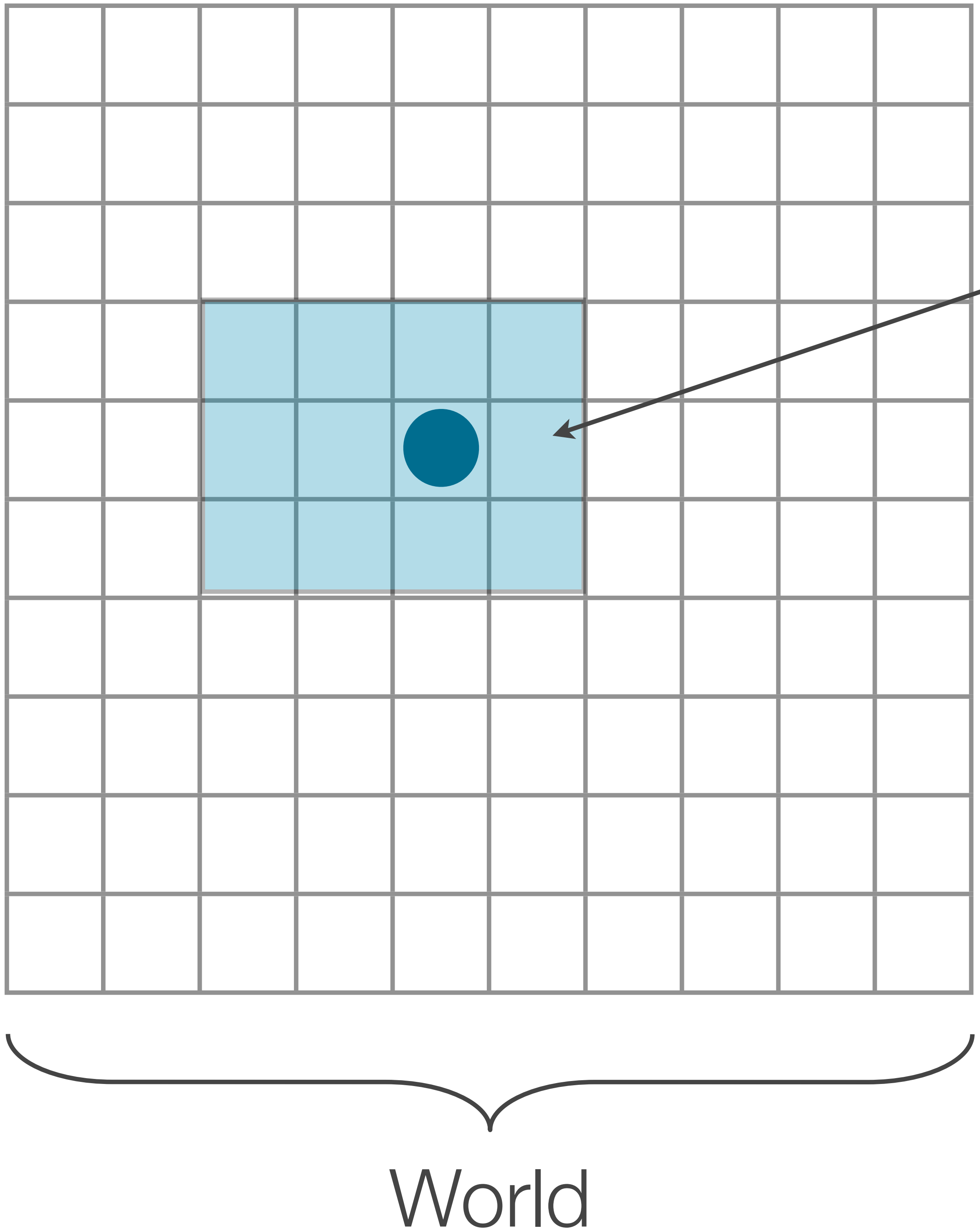
Plan

- How is the data generated?
  - Strong vs. weak sampling: the idea
  - People's sensitivity to sampling assumptions
  - Individual differences in sampling sensitivity

Plan

- ➡ How is the data generated?
  - ➡ Strong vs. weak sampling: the idea
  - People's sensitivity to sampling assumptions
  - Individual differences in sampling sensitivity

Remember Bayes' Rule...



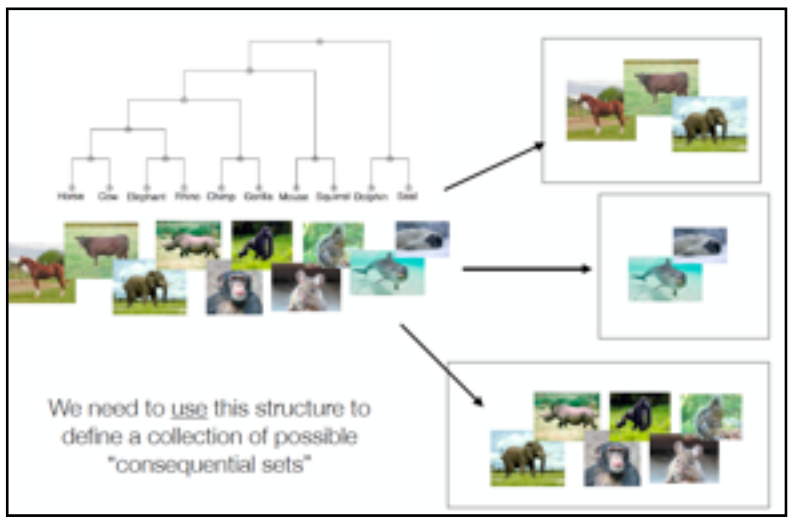
Hypothesis of size n

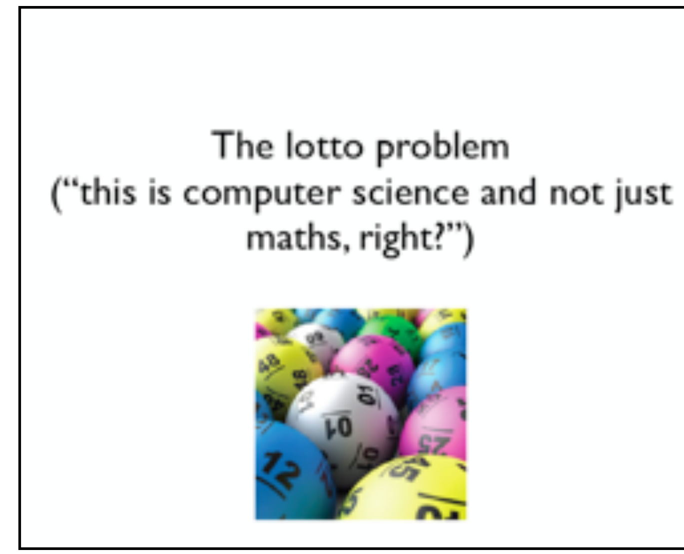
$$p(d|h) = 1/n$$
$$= 1/12$$

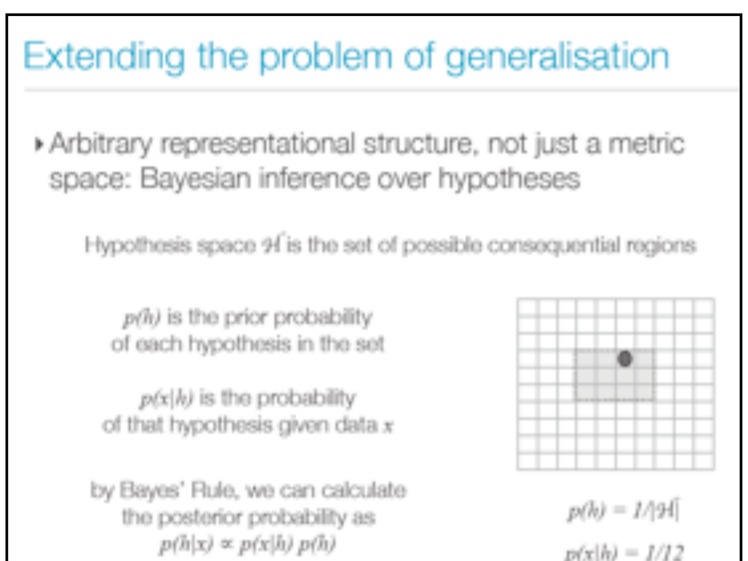
This is known as the **size principle**

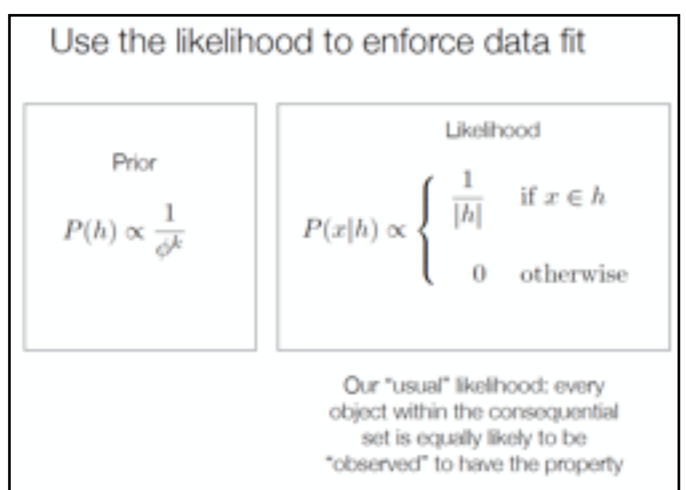
The size principle has cropped up in many places...

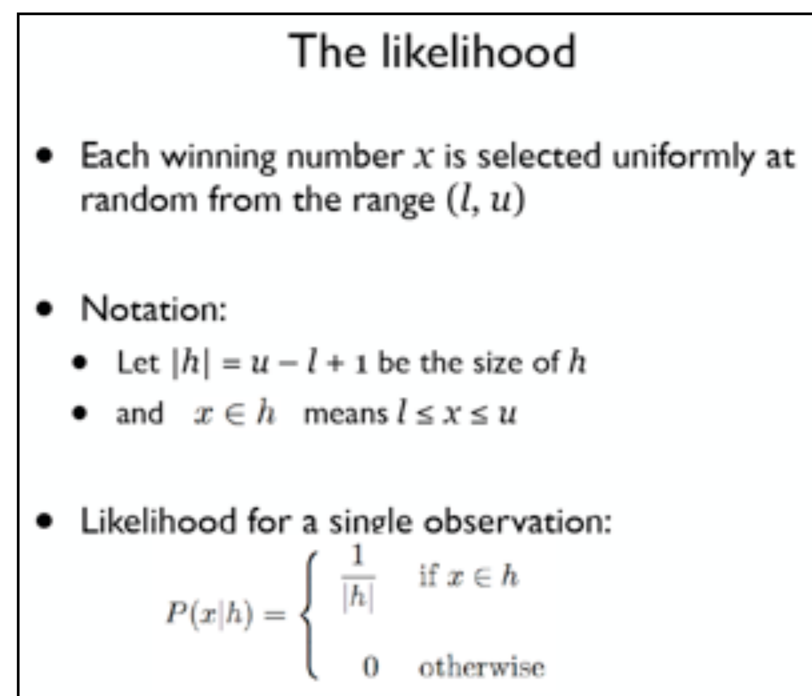
- Explicitly so when talking about the lotto problem and the problem of generalisation...





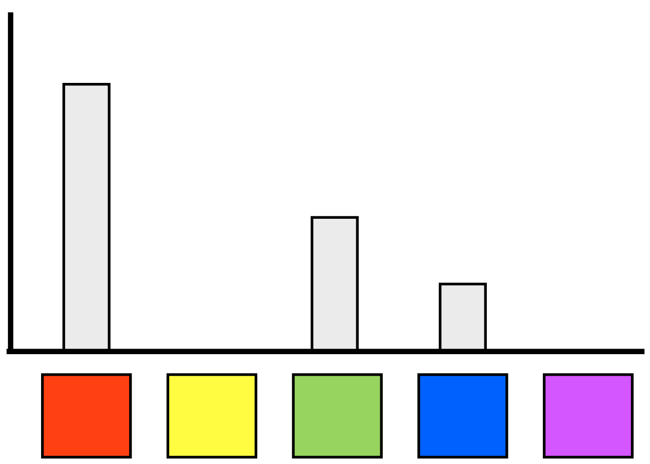


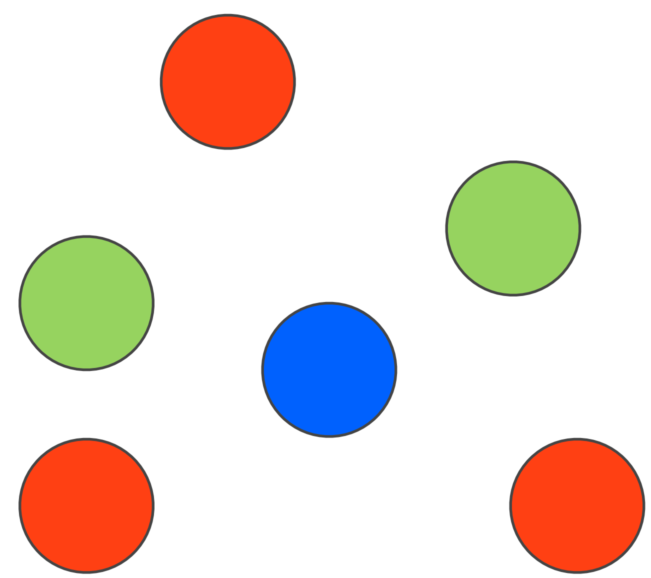




The size principle has cropped up in many places...

- But also much everywhere we have assumed that the data is drawn proportional to the probability distribution that generates it (called strong sampling)





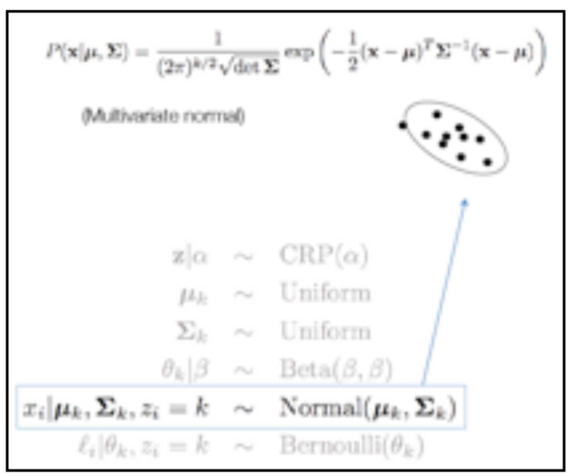
- ... and thus the probability of the data is given by the proportions under that distribution:

$$p(\bullet|h) = 50\%$$

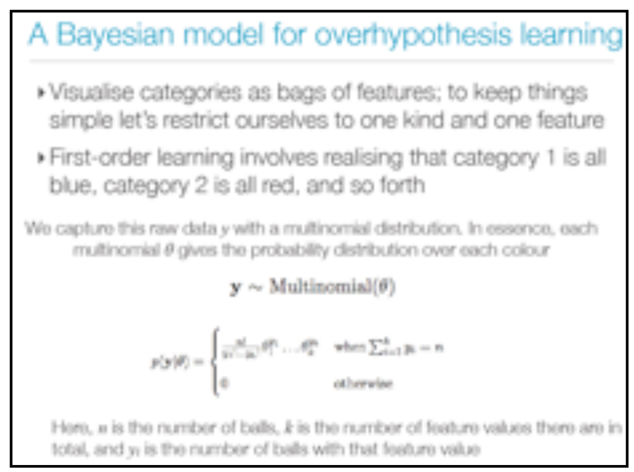
The size principle has cropped up in many places...

- But also much everywhere we have assumed that the data is drawn proportional to the probability distribution that generates it

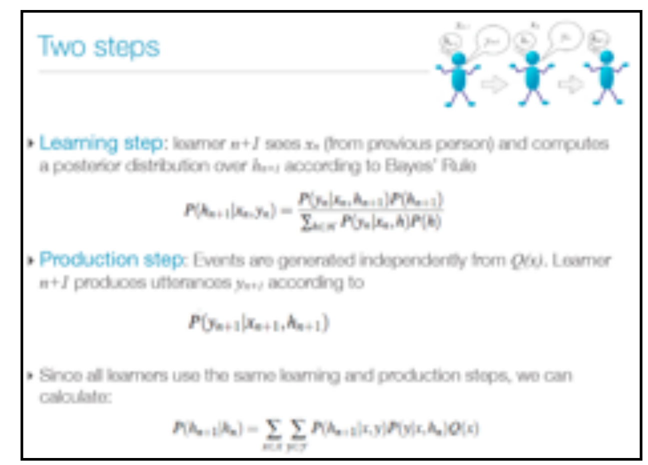
RMC



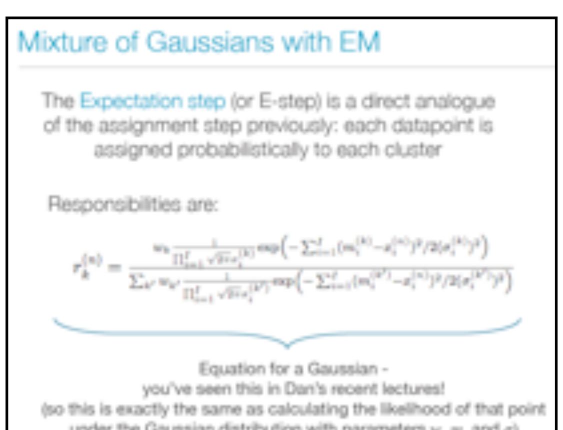
Overhypothesis models

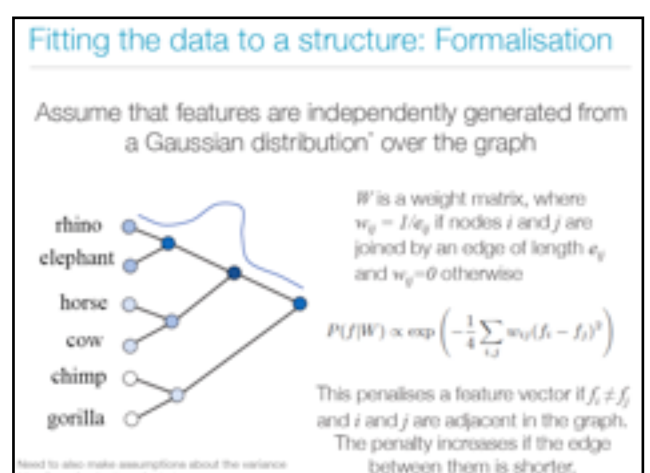


Iterated learning (most)

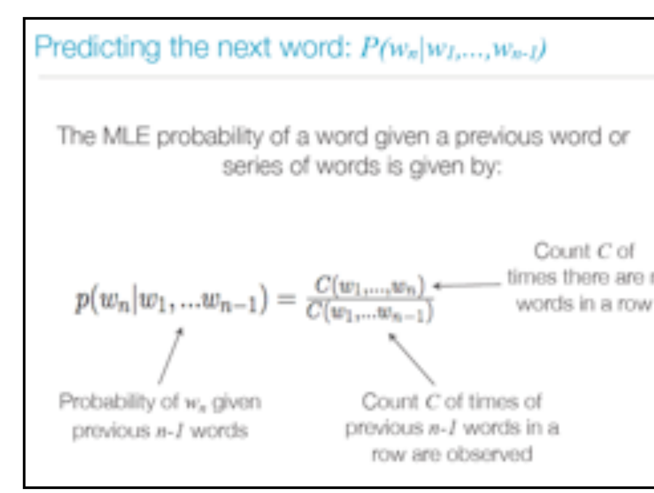


Mixture of Gaussians



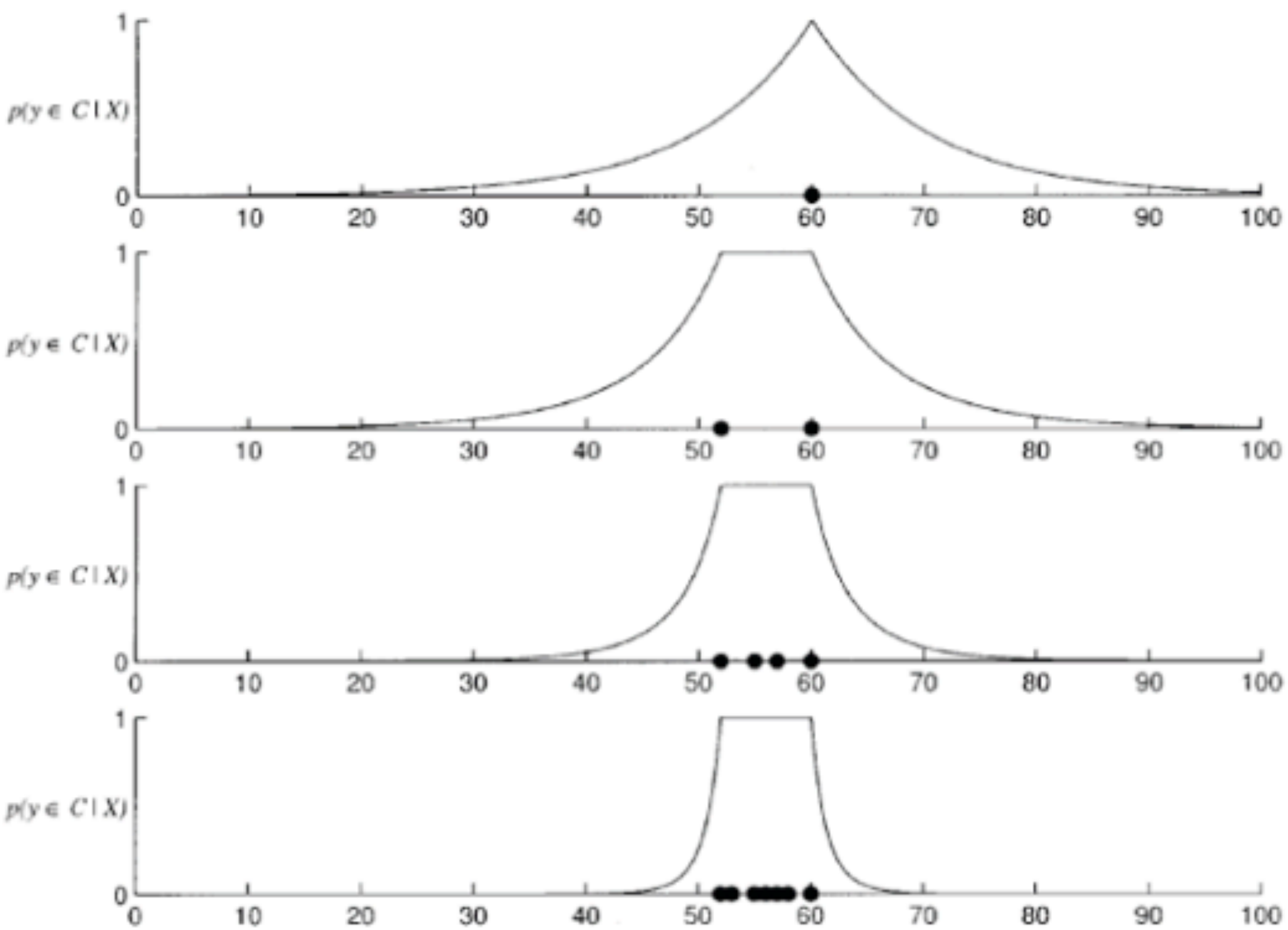


N-gram models



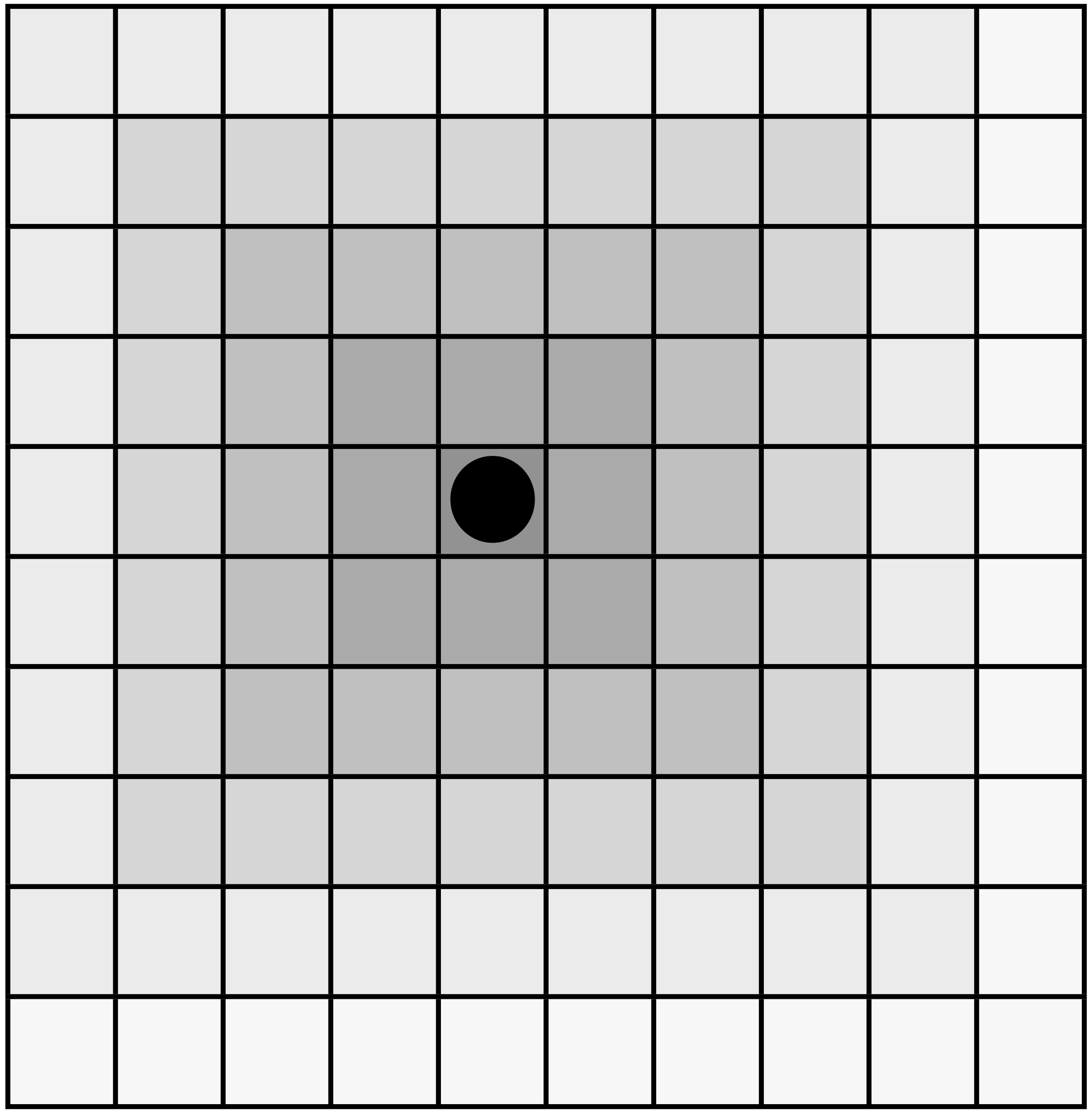
Consequence of the size principle

- It is due to the size principle that additional data points will cause generalisation curves to tighten



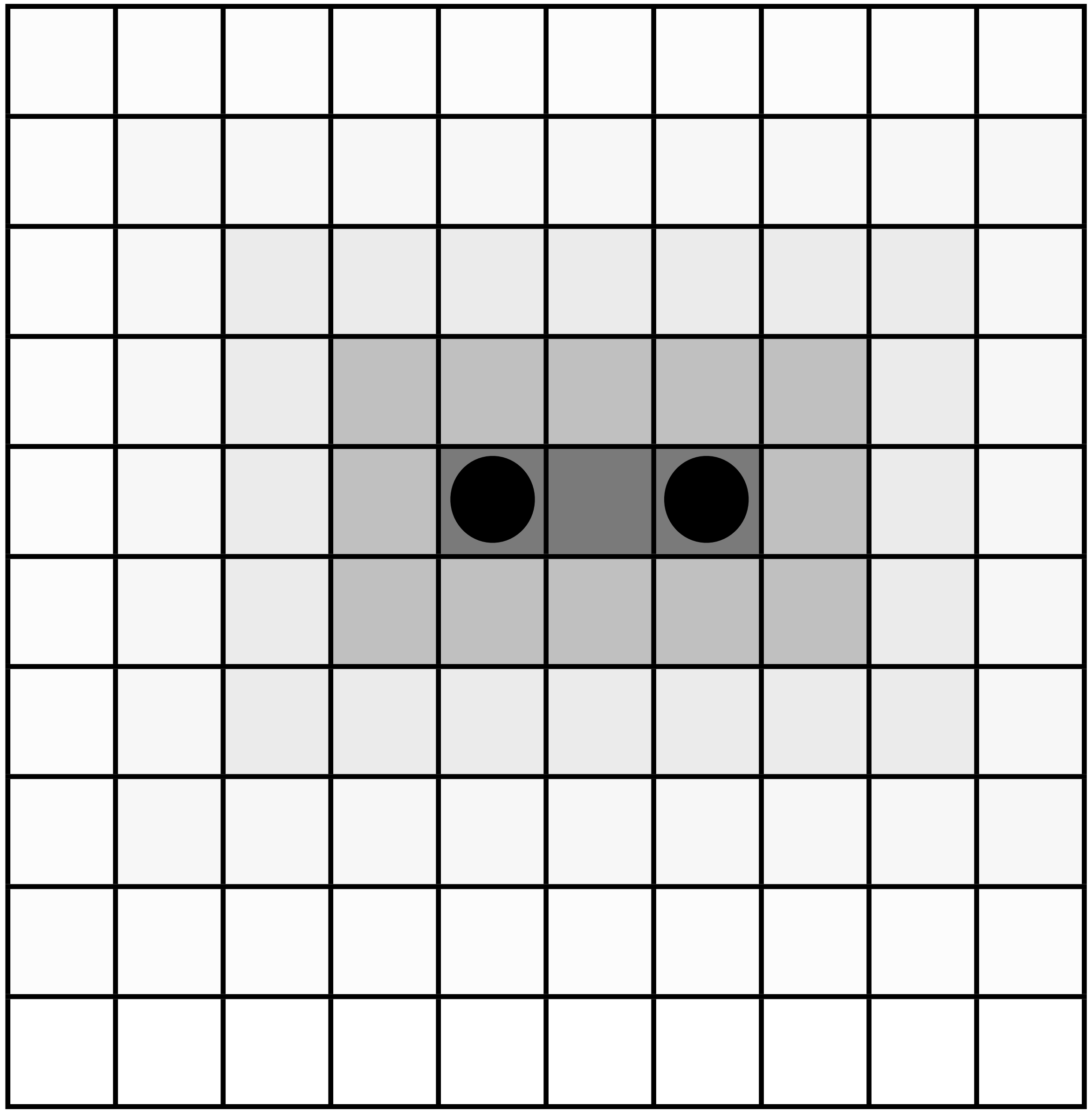
Consequence of the size principle

- It is due to the size principle that additional data points will cause generalisation curves to tighten



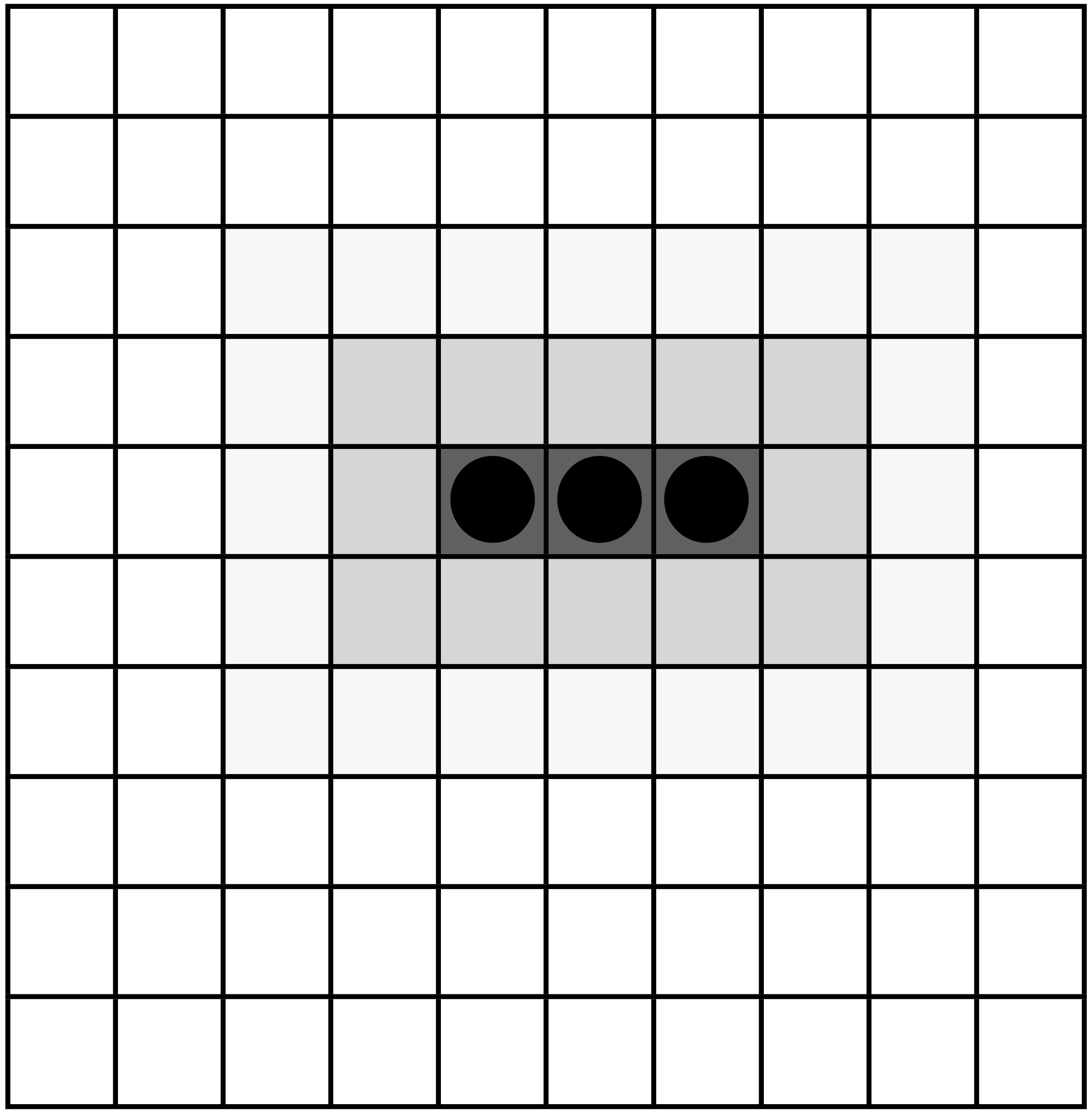
Consequence of the size principle

- It is due to the size principle that additional data points will cause generalisation curves to tighten



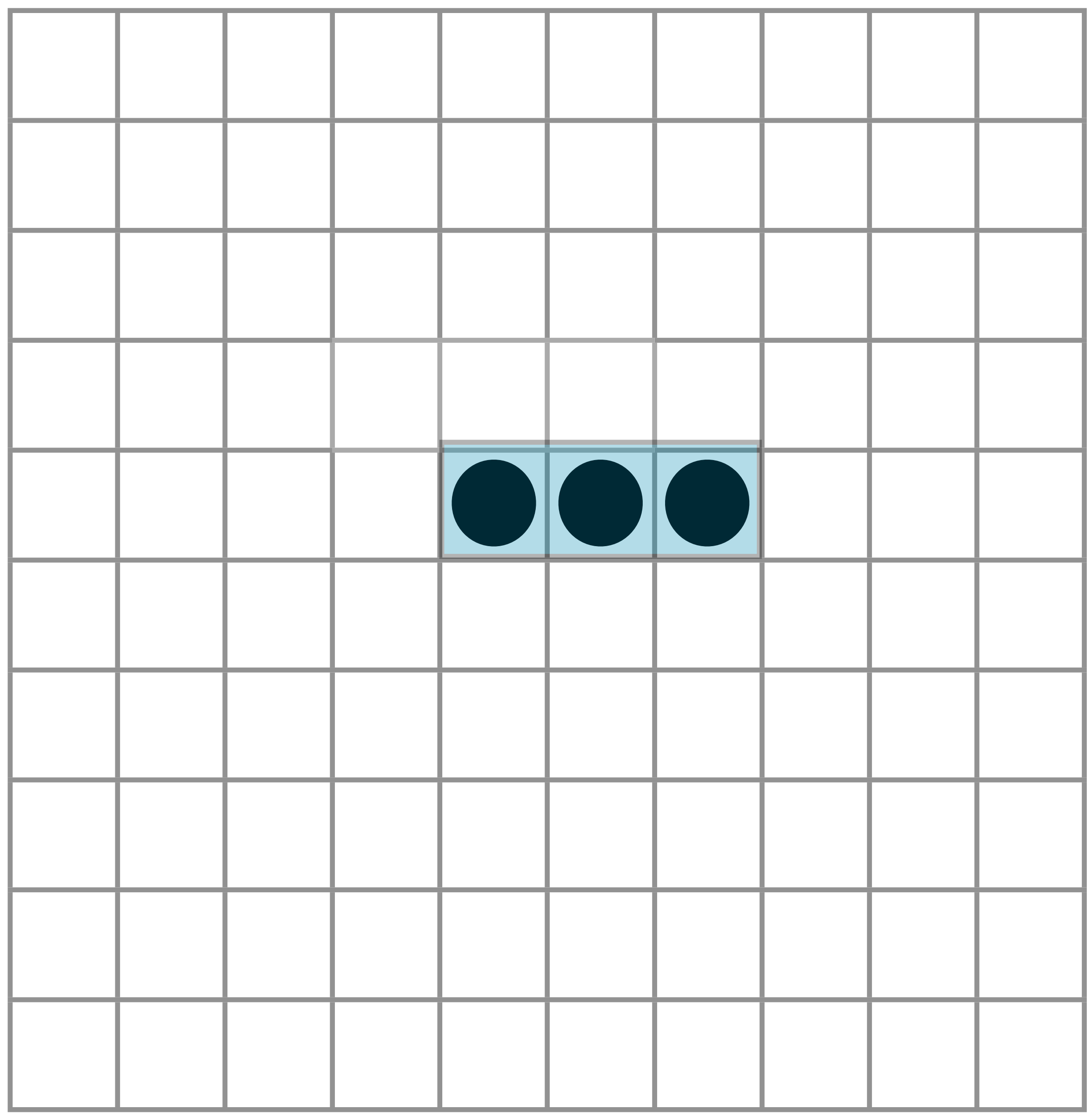
Consequence of the size principle

- It is due to the size principle that additional data points will cause generalisation curves to tighten



Consequence of the size principle

- It is due to the size principle that additional data points will cause generalisation curves to tighten

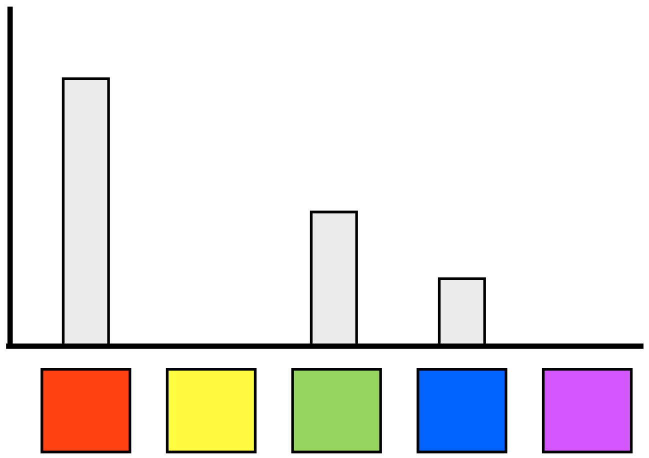


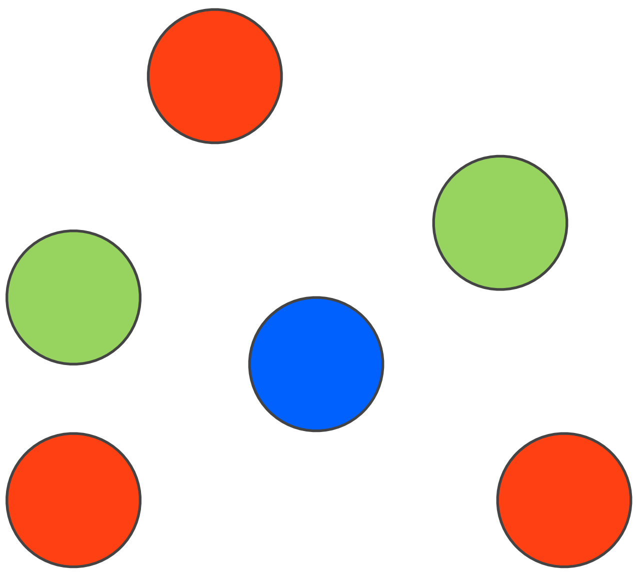
This is because it's quite a suspicious coincidence for these data points to have been generated if the true hypothesis is *not* h

The size principle is not the only way!

- It is sensible, but it follows from certain assumptions about how data were generated (or sampled)

Each point drawn independently and at random from the hypothesis



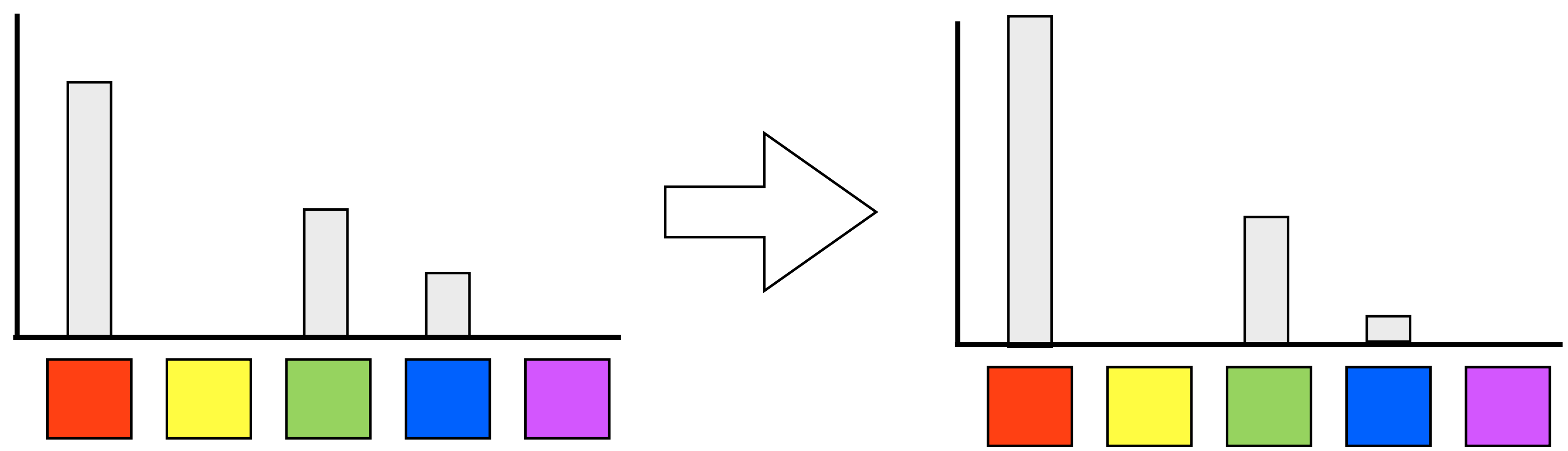


The size principle is not the only way!

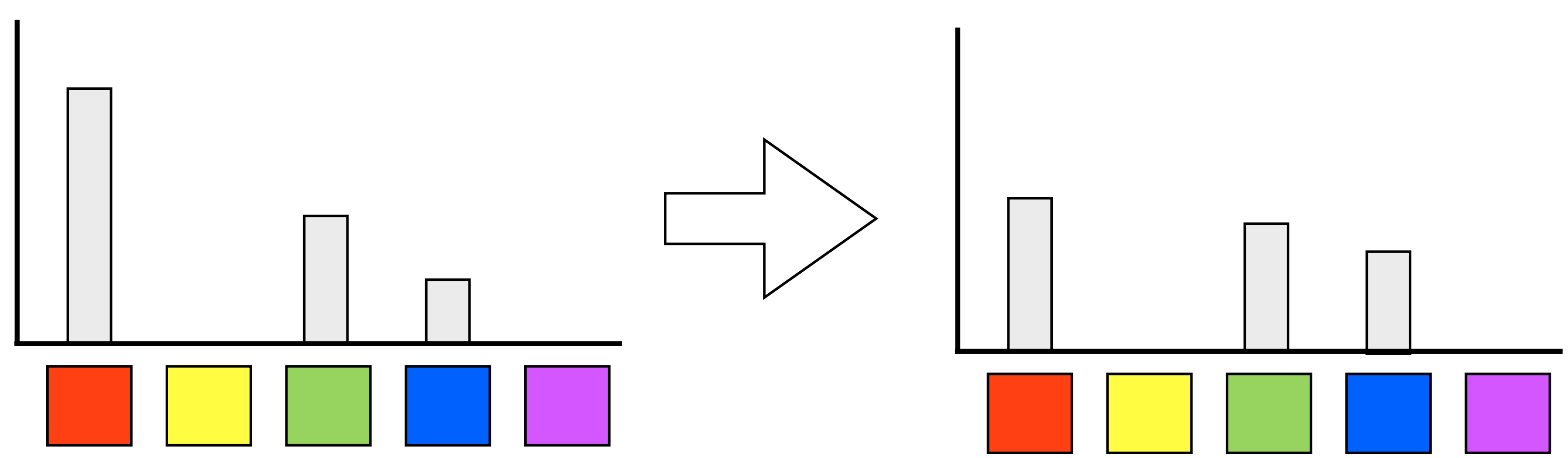
- It is sensible, but it follows from certain assumptions about how data were generated (or sampled)

Suppose data were non-independent:
$p(d_t|h)$ depended on $p(d_{t-1}|h)$.

If $p(d_t\!=\!X|h)$ is larger if $p(d_{t-1}\!=\!X|h)$, the distribution skews



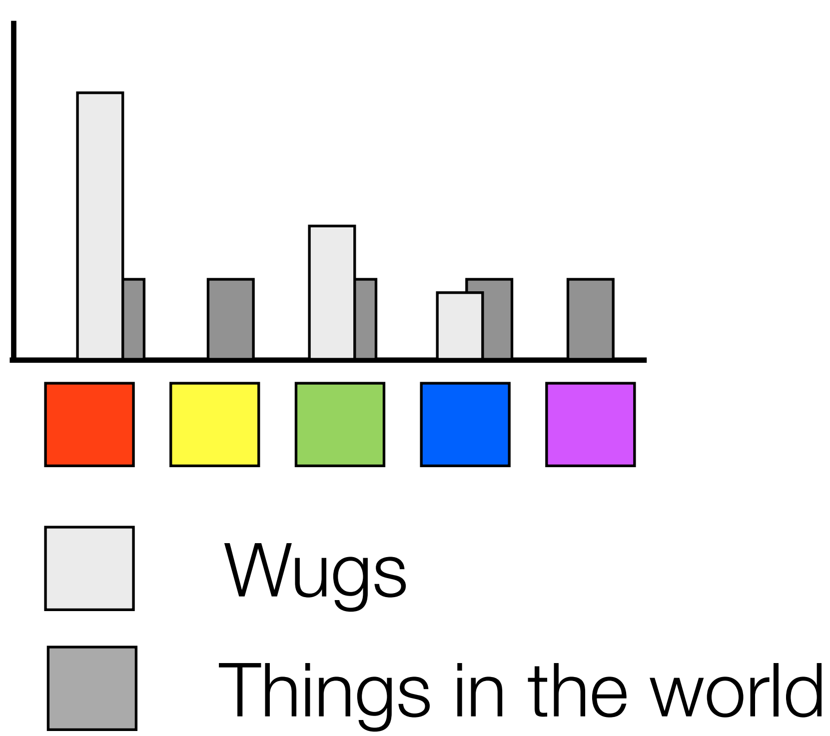
If $p(d_t\!=\!X|h)$ is smaller if $p(d_{t-1}\!=\!X|h)$, the distribution flattens

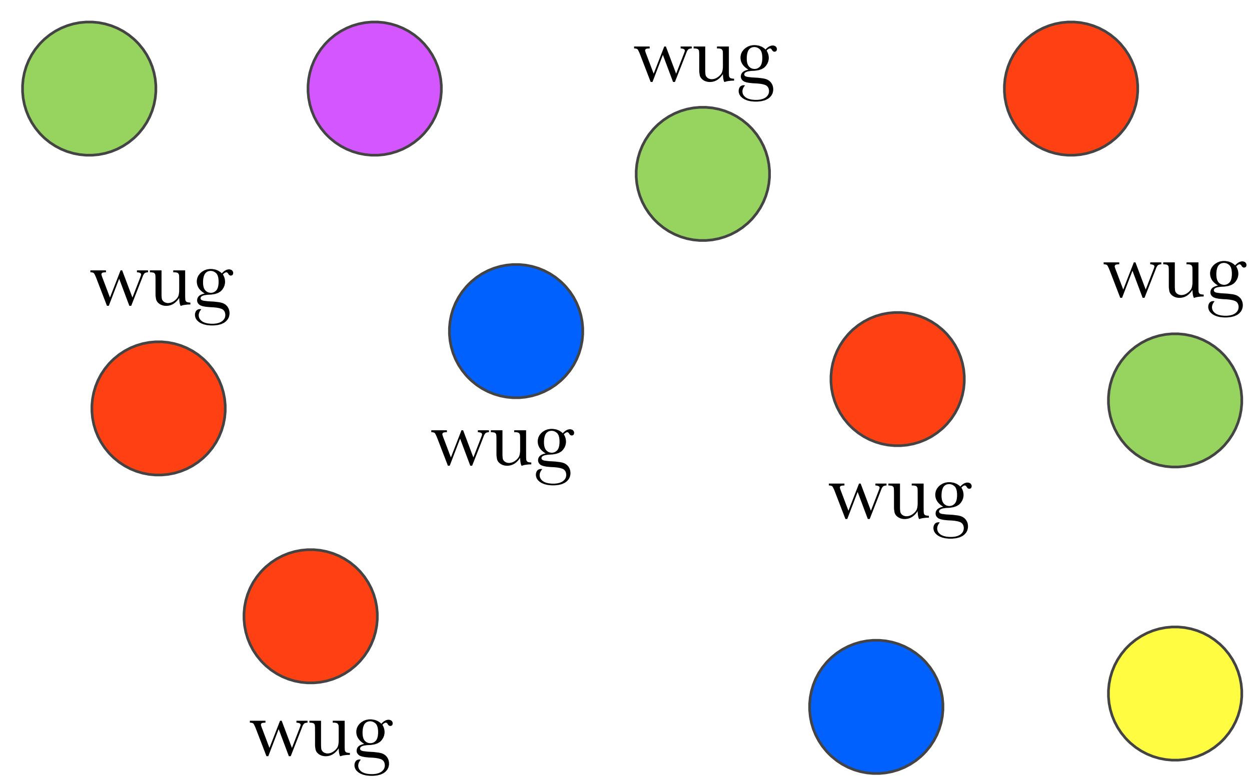


The size principle is not the only way!

- It is sensible, but it follows from certain assumptions about how data were generated (or sampled)

Suppose data could have been generated from the world in general, and only then labelled as belonging to the hypothesis (or not)

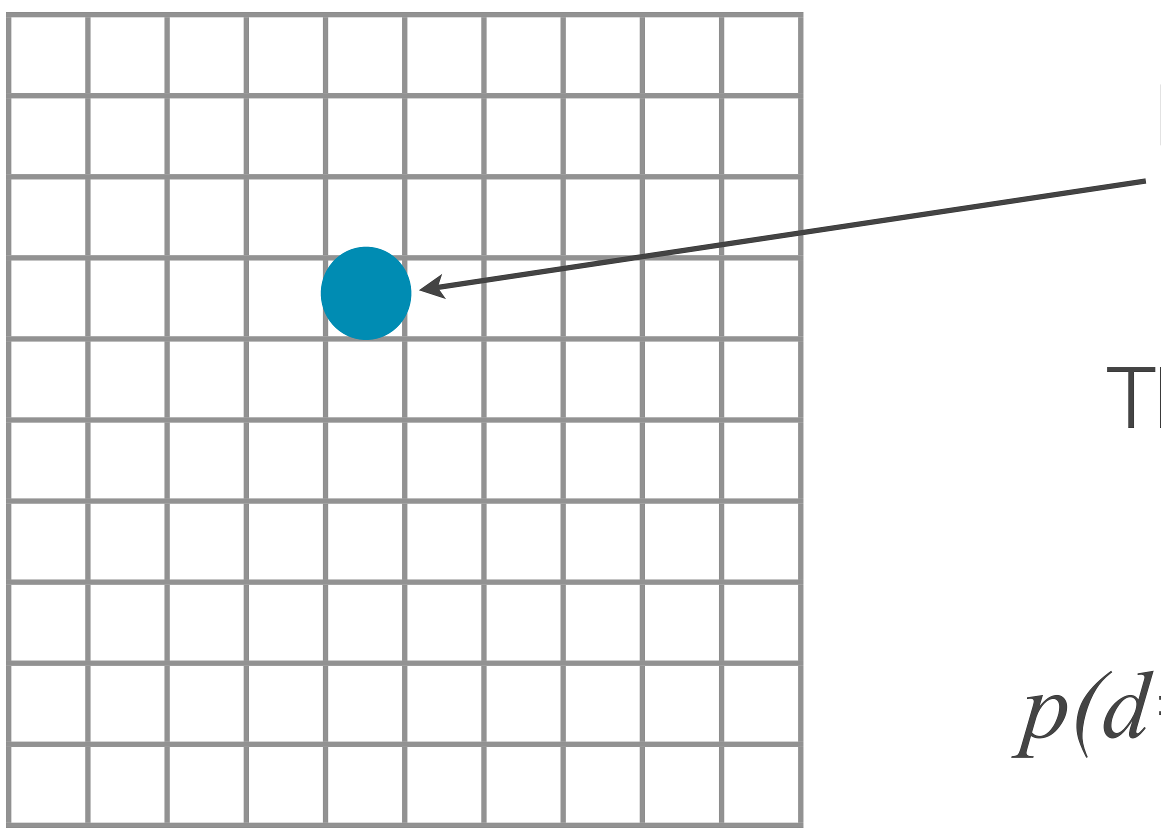




The size principle is not the only way!

- It is sensible, but it follows from certain assumptions about how data were generated (or sampled)

Suppose data could have been generated from the world in general, and only then labelled as belonging to the hypothesis (or not)



Data sampled from the world at random

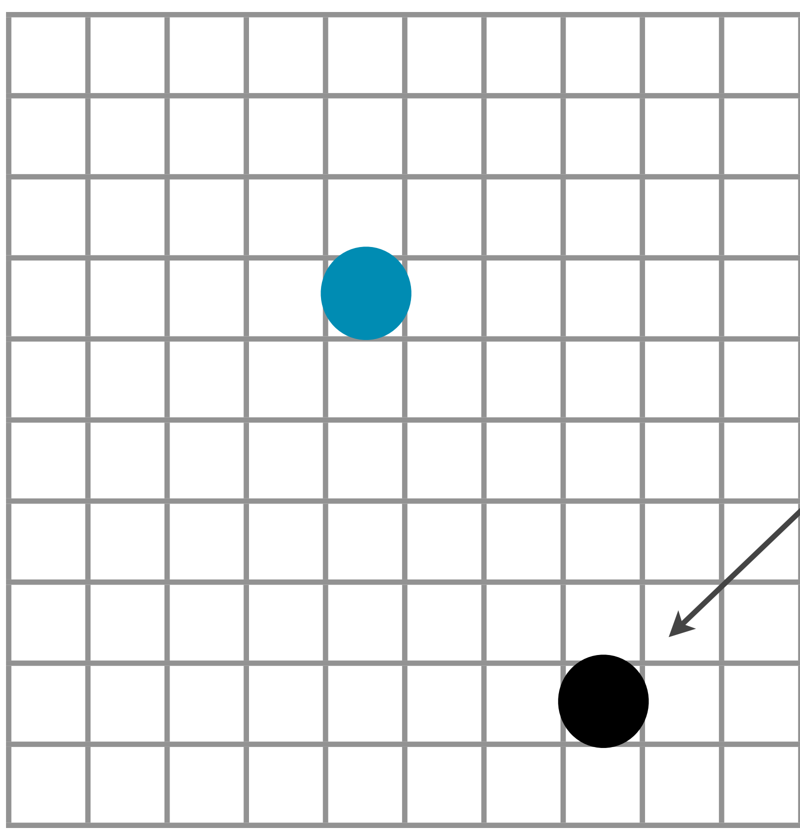
Then labelled as in the hypothesis or not

$$p(d=\bullet|h) = 1 \text{ if in the hypothesis}$$
$$0 \text{ if not}$$

The size principle is not the only way!

- It is sensible, but it follows from certain assumptions about how data were generated (or sampled)

Suppose data could have been generated from the world in general, and only then labelled as belonging to the hypothesis (or not)



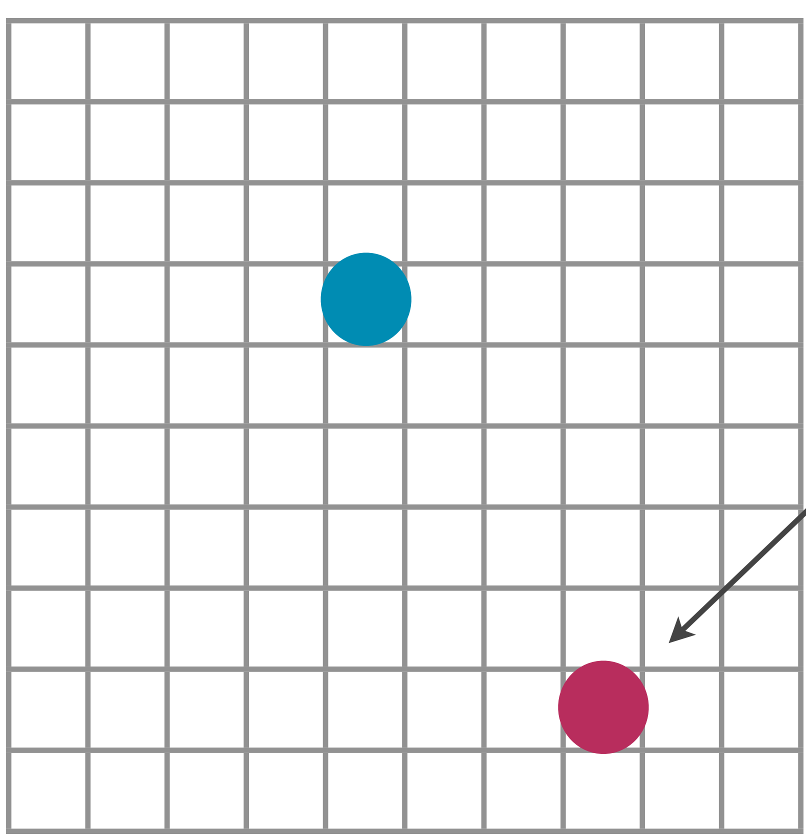
Data sampled from the world at random

$$p(d=\bullet|h) = 1 \text{ if in the hypothesis}$$
$$0 \text{ if not}$$

The size principle is not the only way!

- It is sensible, but it follows from certain assumptions about how data were generated (or sampled)

Suppose data could have been generated from the world in general, and only then labelled as belonging to the hypothesis (or not)



Data sampled from the world at random

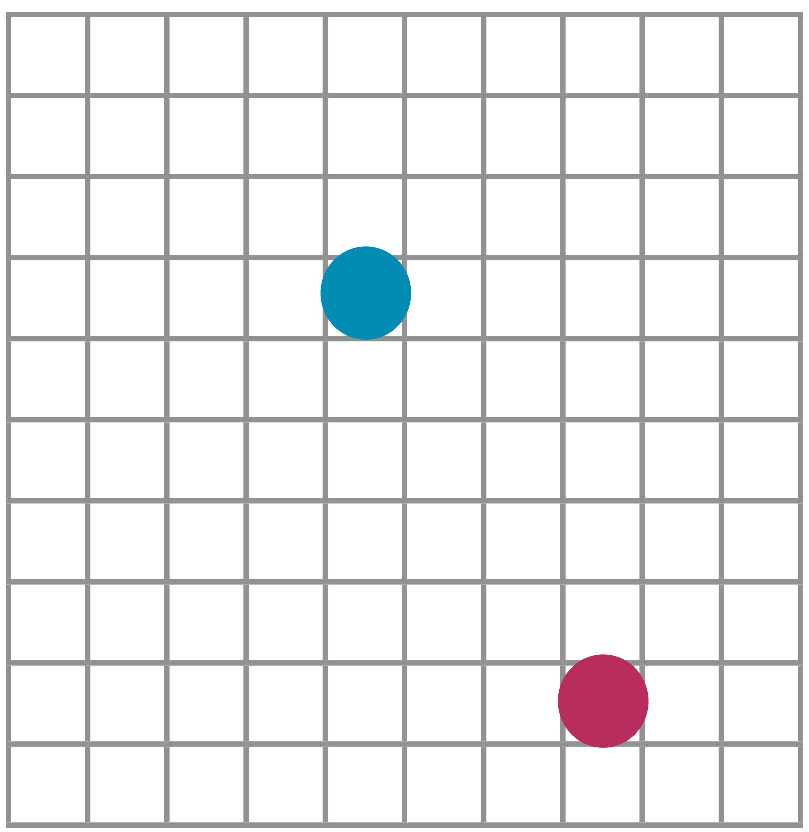
Then labelled as in the hypothesis or not

$$p(d=\bullet|h) = 1 \text{ if in the hypothesis}$$
$$0 \text{ if not}$$

The size principle is not the only way!

- It is sensible, but it follows from certain assumptions about how data were generated (or sampled)

Suppose data could have been generated from the world in general, and only then labelled as belonging to the hypothesis (or not)

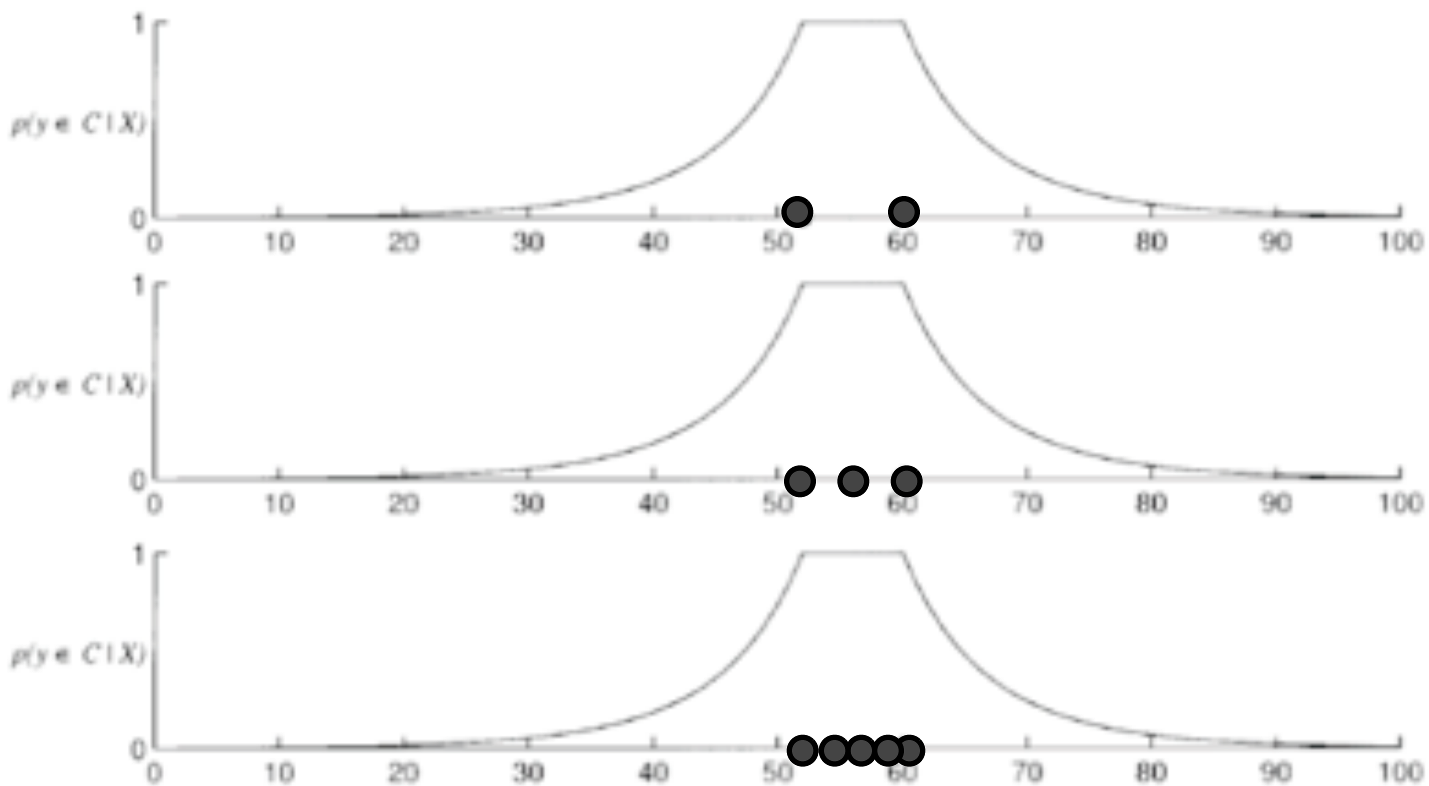


This is called weak sampling

$$p(d=\bullet|h) = \begin{cases} 1 \text{ if in the hypothesis} \\ 0 \text{ if not} \end{cases}$$

The size principle is not the only way!

- If data are weakly sampled, the generalisation curves should not tighten -- there is no suspicious coincidence since the data were generated by the *world,* and not from the hypothesis



Are people sensitive to sampling assumptions?

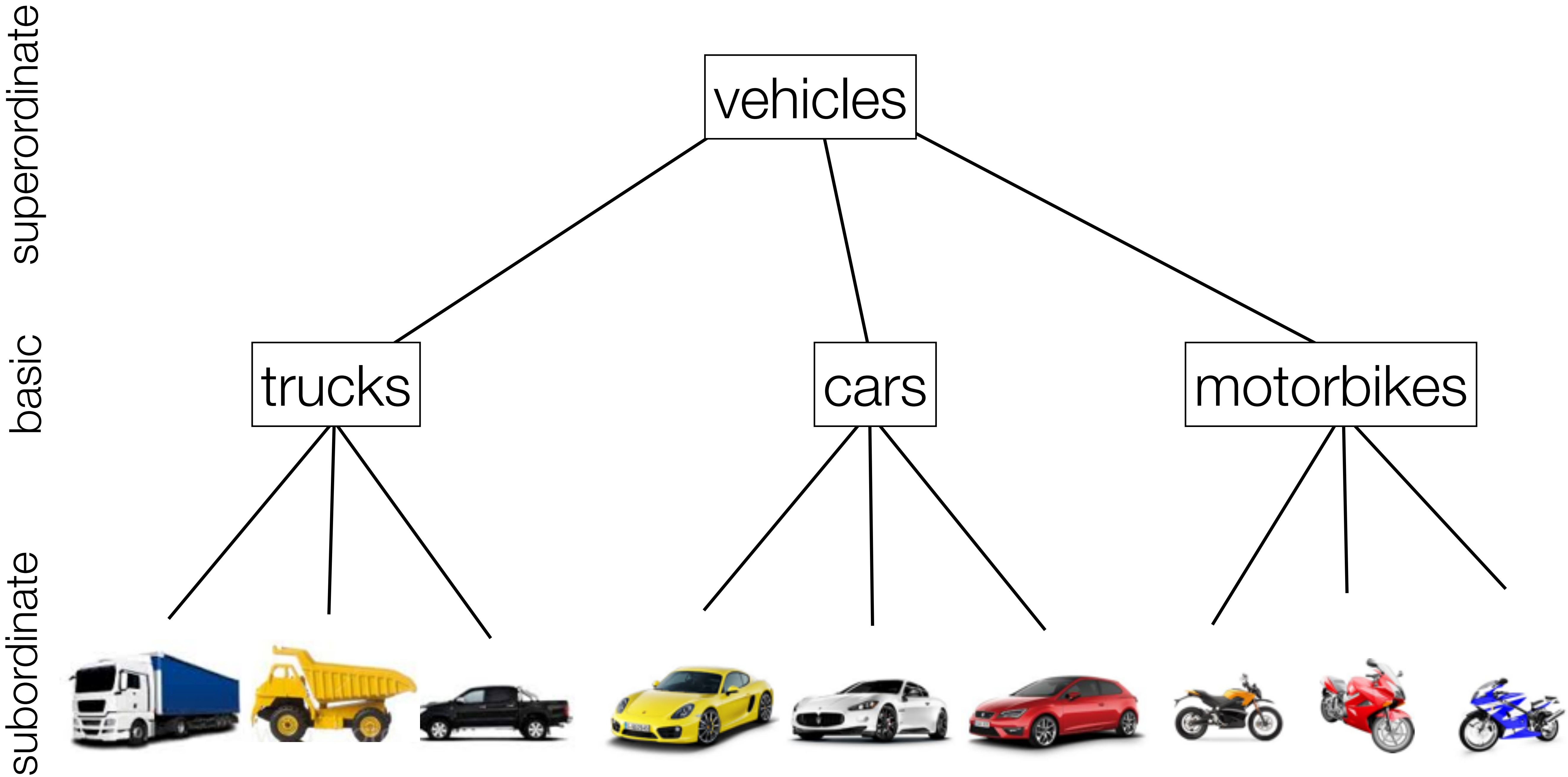
Do people change their generalisations if the data have been sampled differently?

Plan

- ➡ How is the data generated?
  - Strong vs. weak sampling: the idea
  - ➡ People's sensitivity to sampling assumptions
  - Individual differences in sampling sensitivity

Word learning

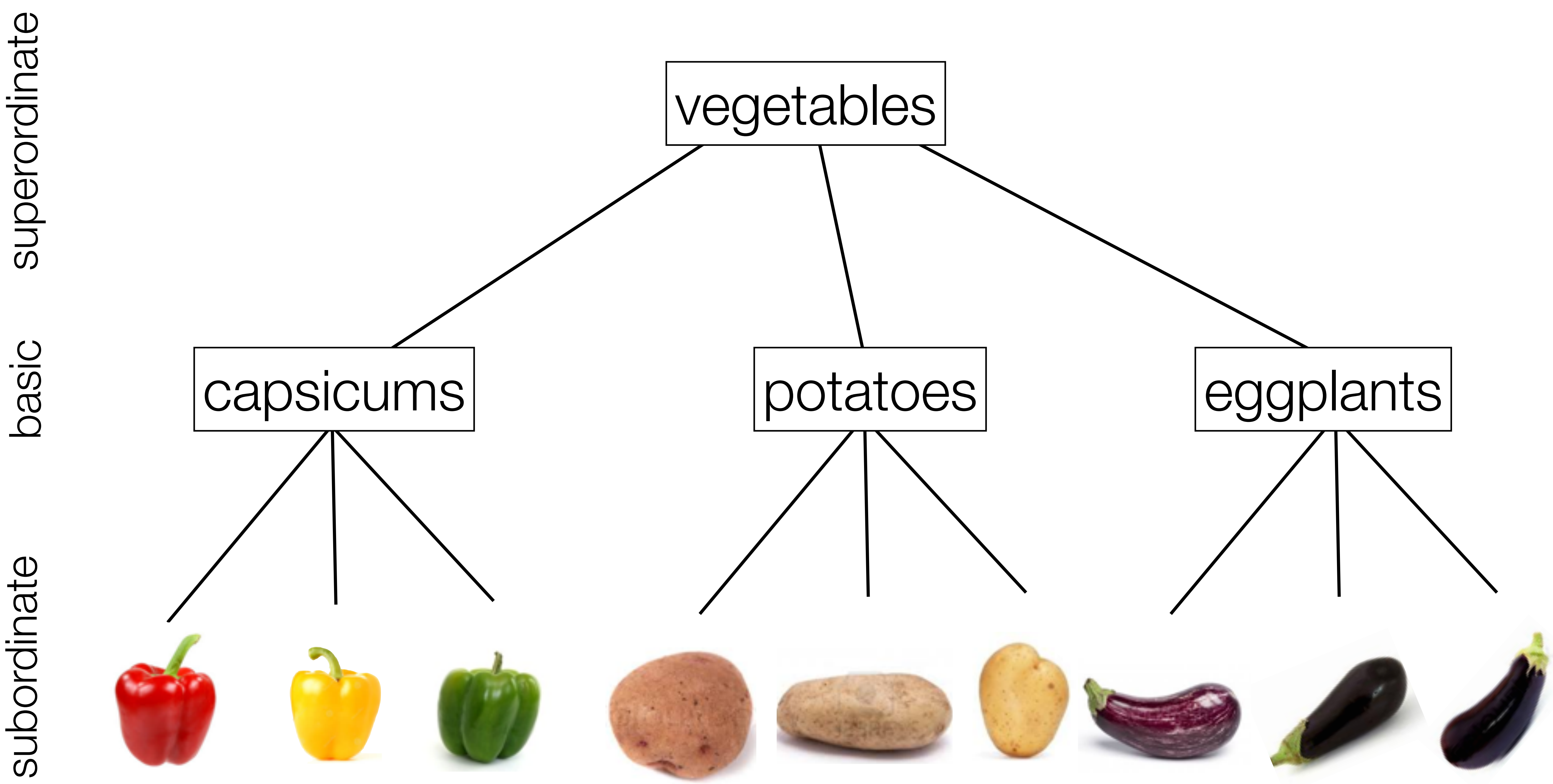
- We've already seen that many domains have a hierarchical or tree-based conceptual structure



Xu & Tenenbaum, 2007

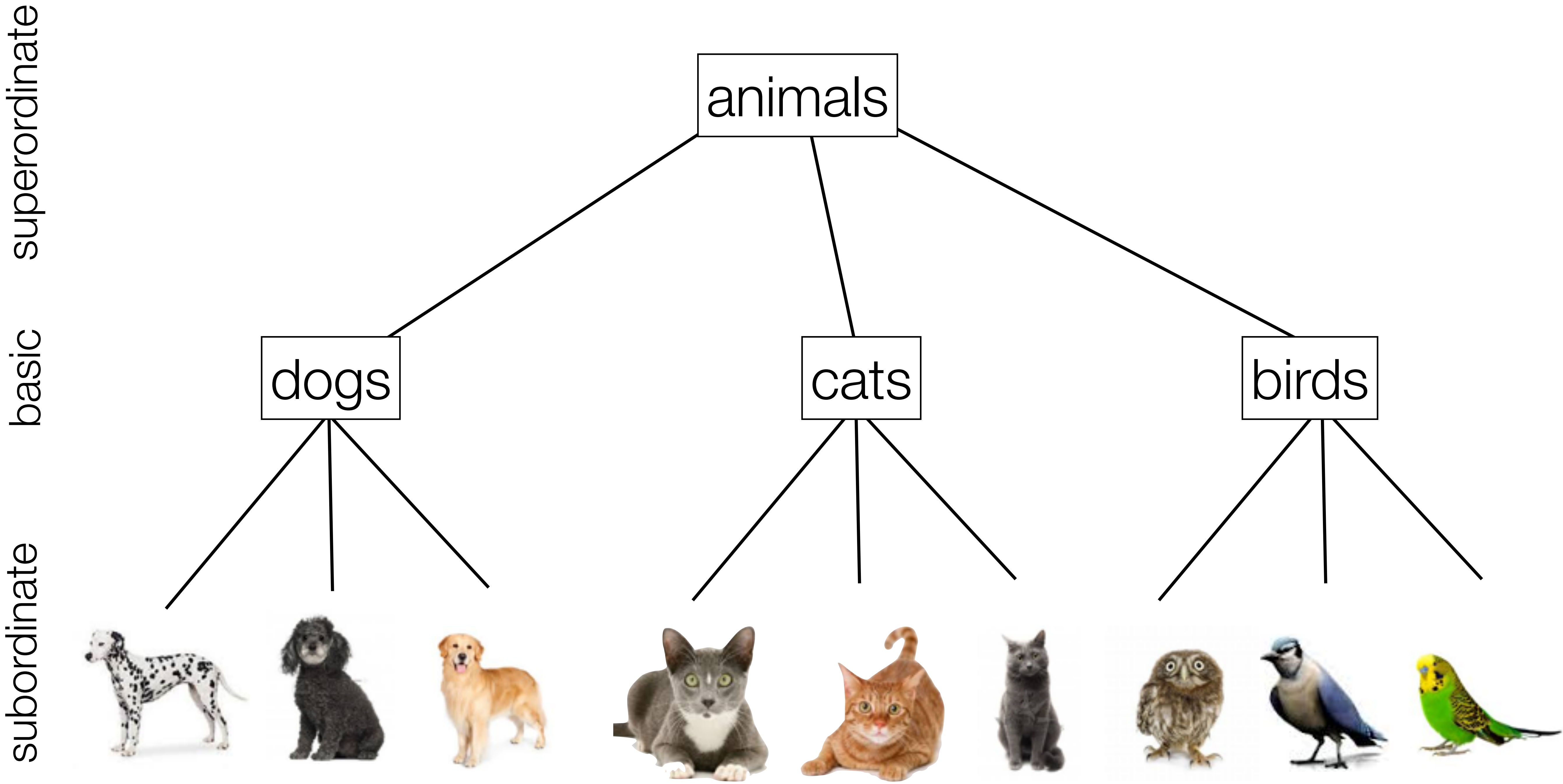
Word learning

- We've already seen that many domains have a hierarchical or tree-based conceptual structure



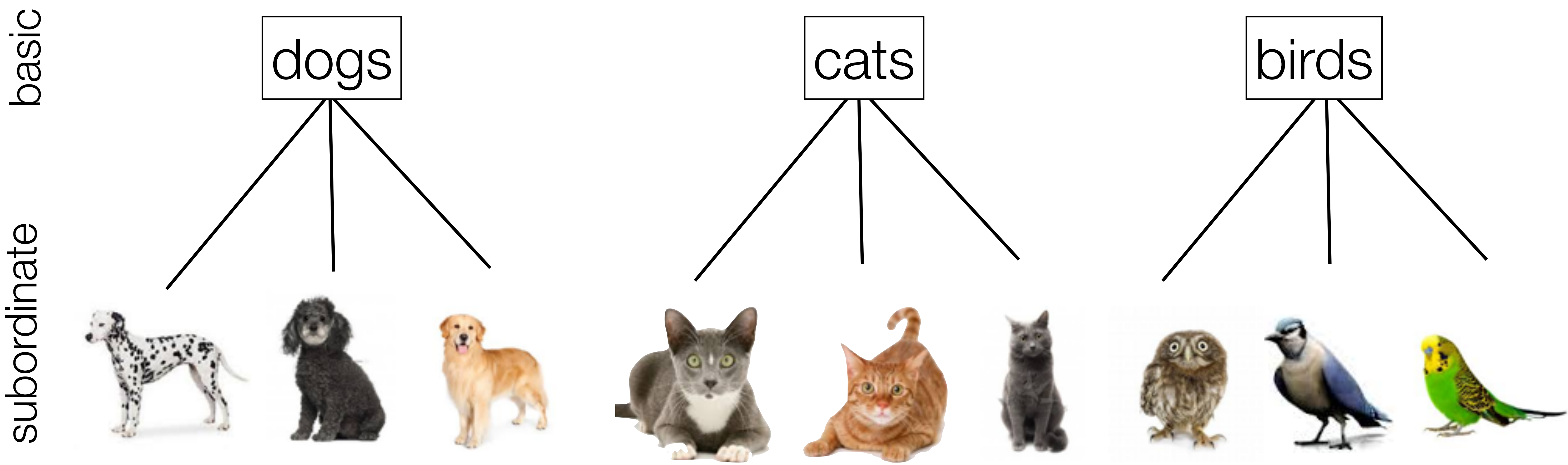
Word learning

- We've already seen that many domains have a hierarchical or tree-based conceptual structure



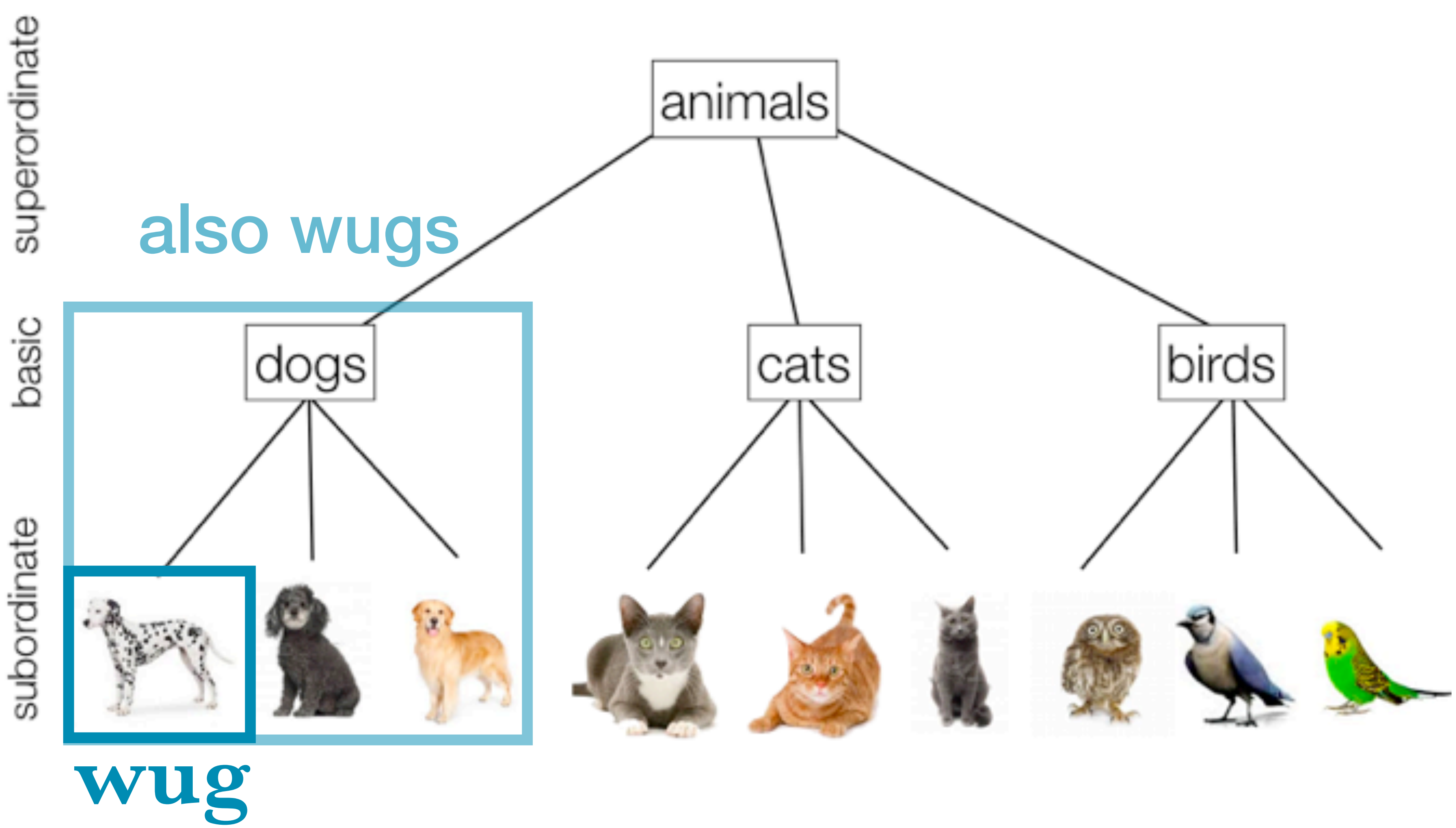
Word learning

- There is lots of independent evidence that the basic level is privileged: it is what people default to when using names, it has the highest inductive power, etc



Word learning

- We would therefore expect that if people were told that *one* item was a wug, people would guess that all other items at the basic level are wugs too

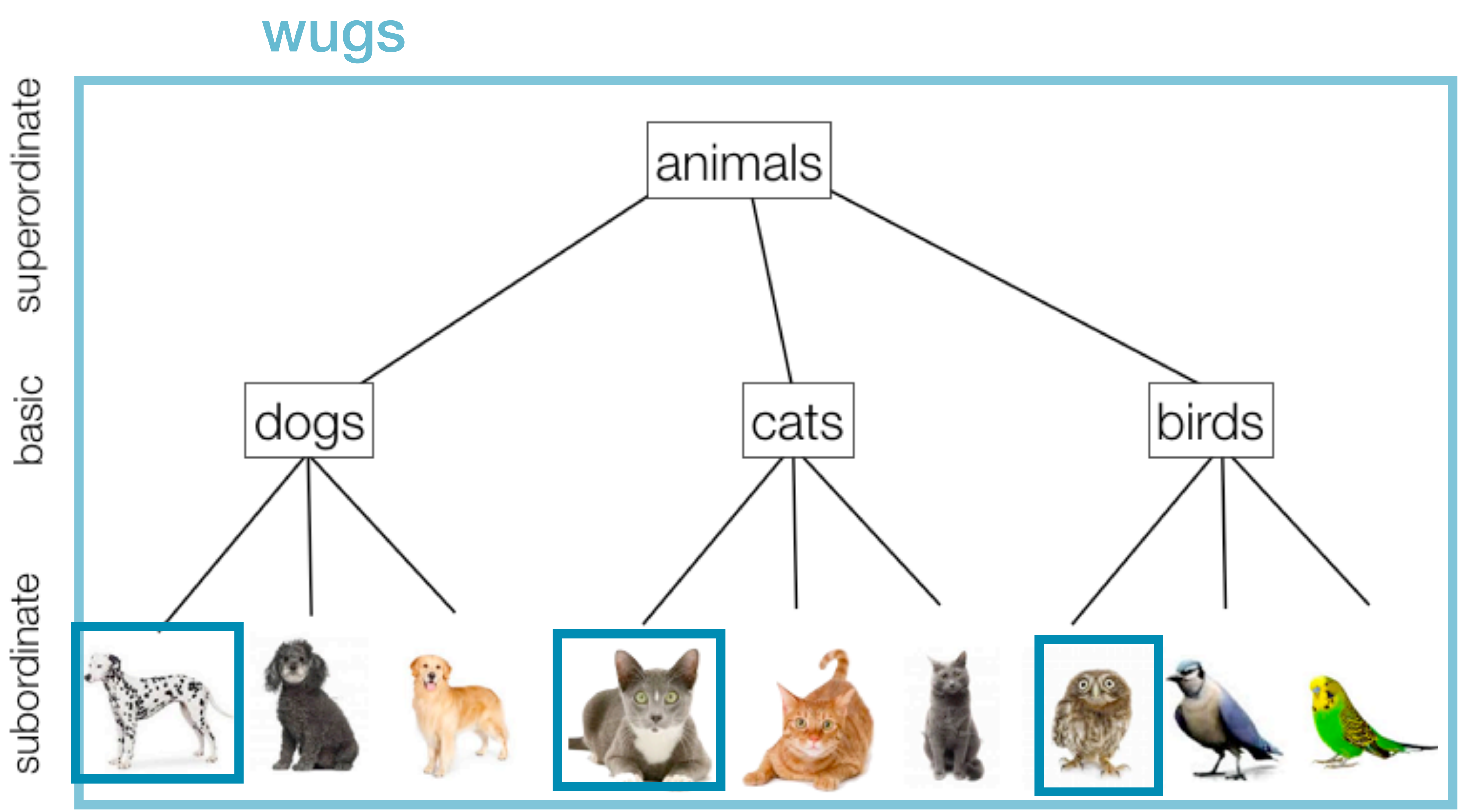


Word learning

- But what if we are given *three* examples of wugs?
- Then it depends on which three examples, and whether people are reasoning based on the size principle...

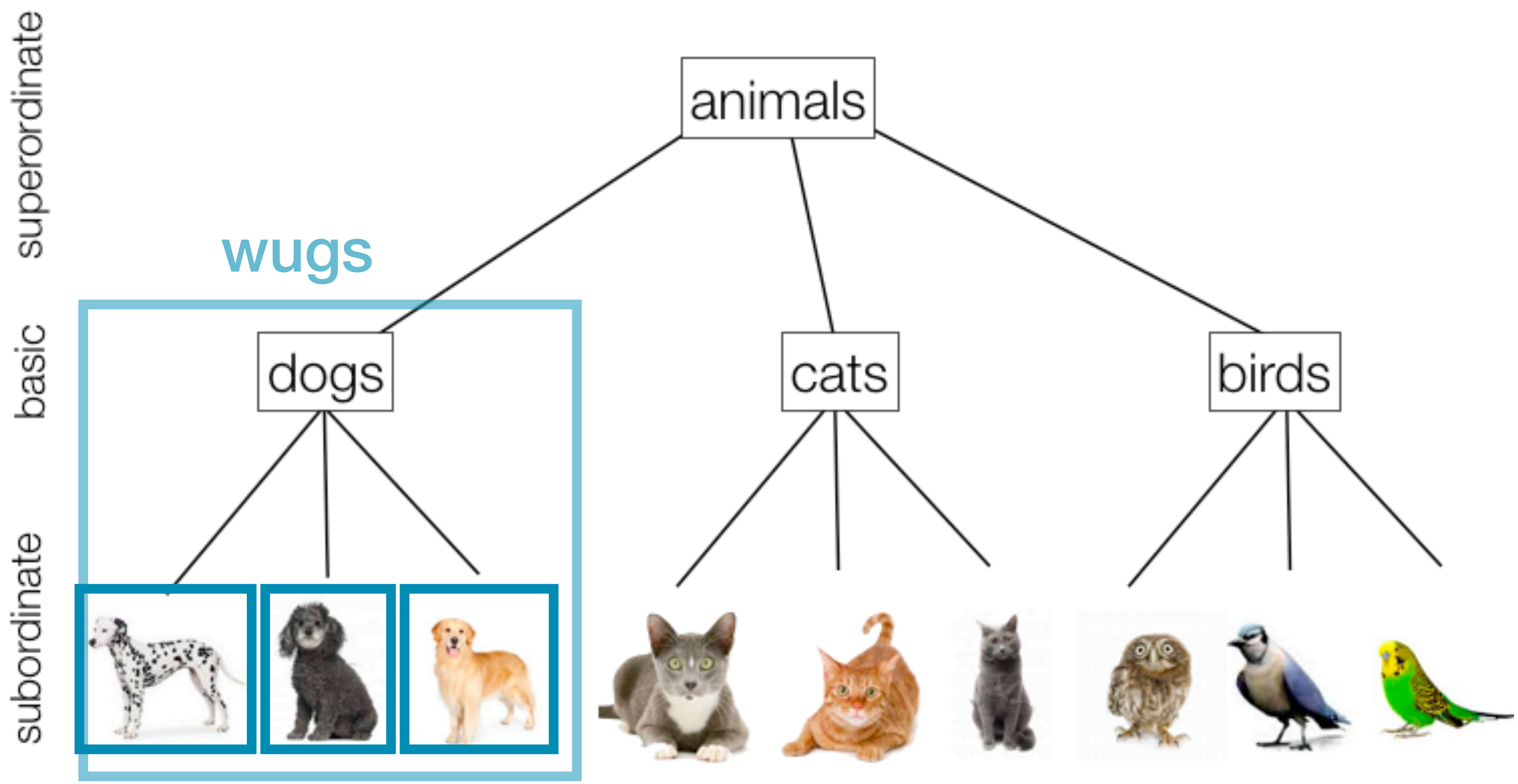
If people use the size principle

- Then they should make the tightest possible generalisation



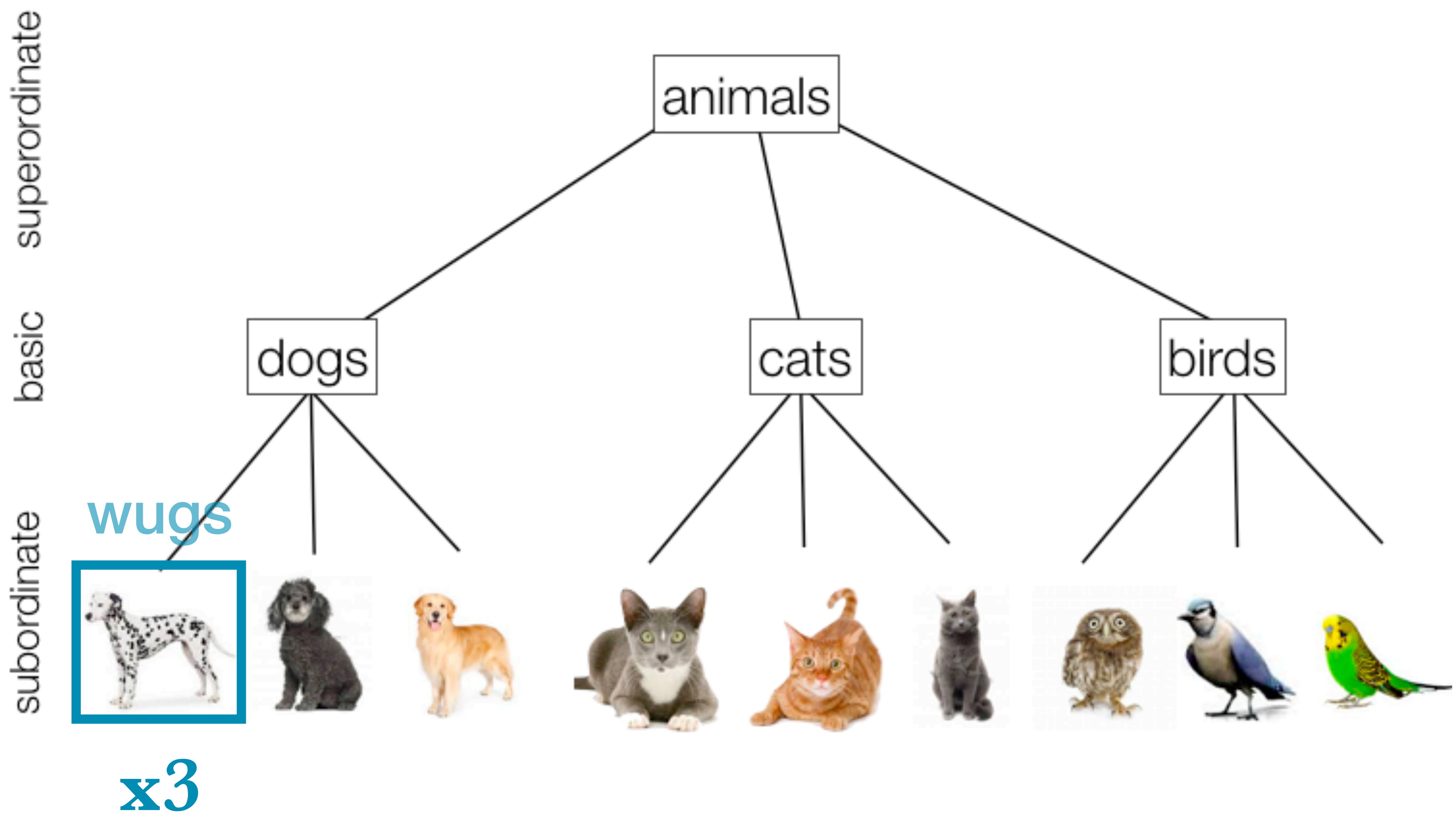
If people use the size principle

- Then they should make the tightest possible generalisation



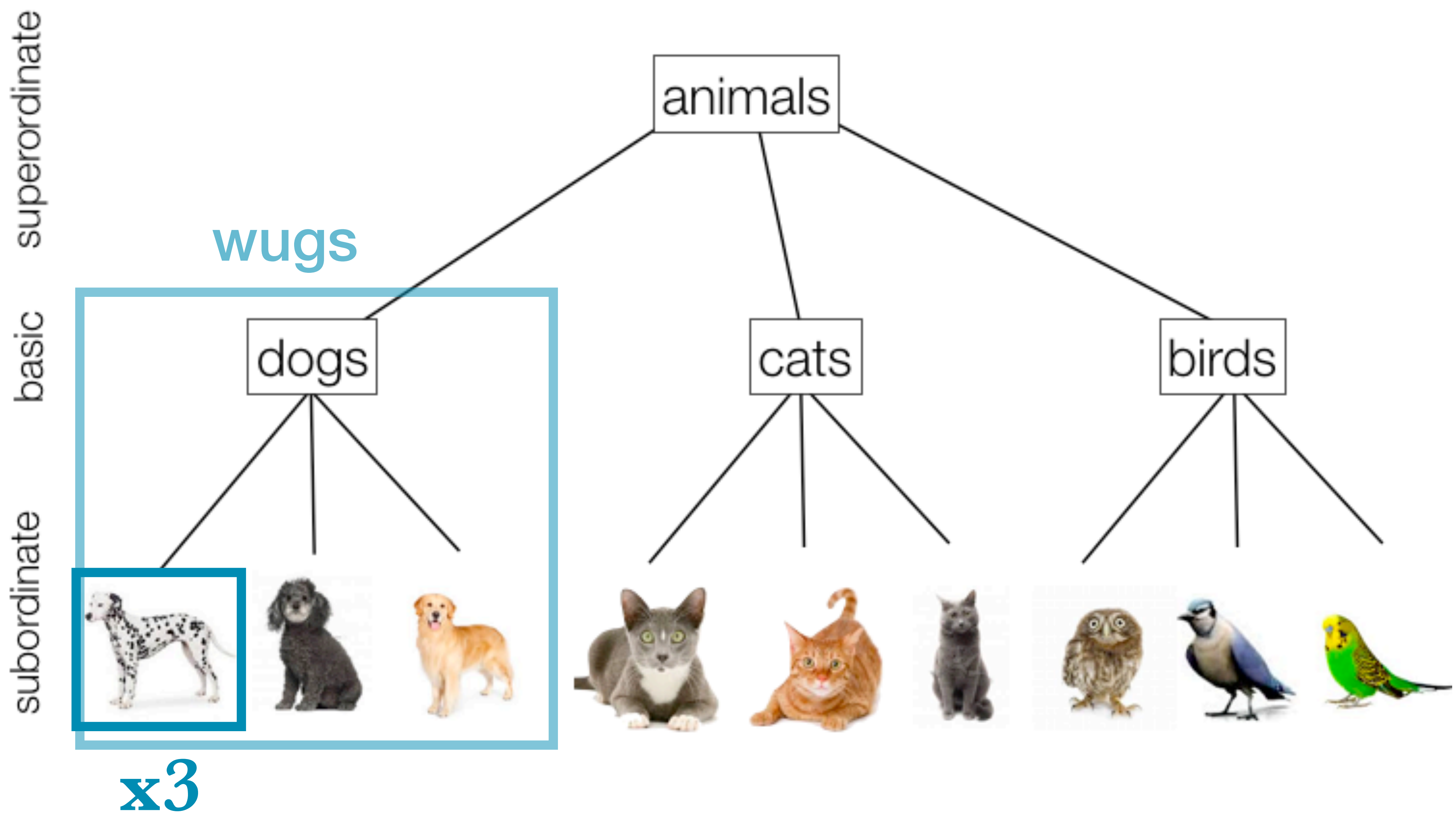
If people use the size principle

- Then they should make the tightest possible generalisation



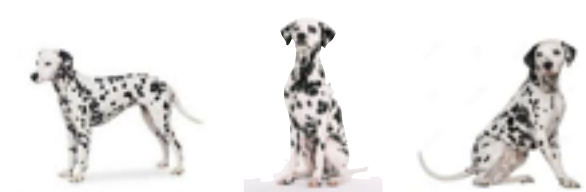
If people DON'T use the size principle

- Then they should not tighten their generalisation when given three of the same item - there is no "suspicious coincidence" to explain



Test

- Four conditions, in each of three domains

	Vegetables	Vehicles	Animals
1 example			
3 subordinate examples			
3 basic-level examples			
3 superordinate examples			

Test

- Four conditions, in each of three domains

Adults

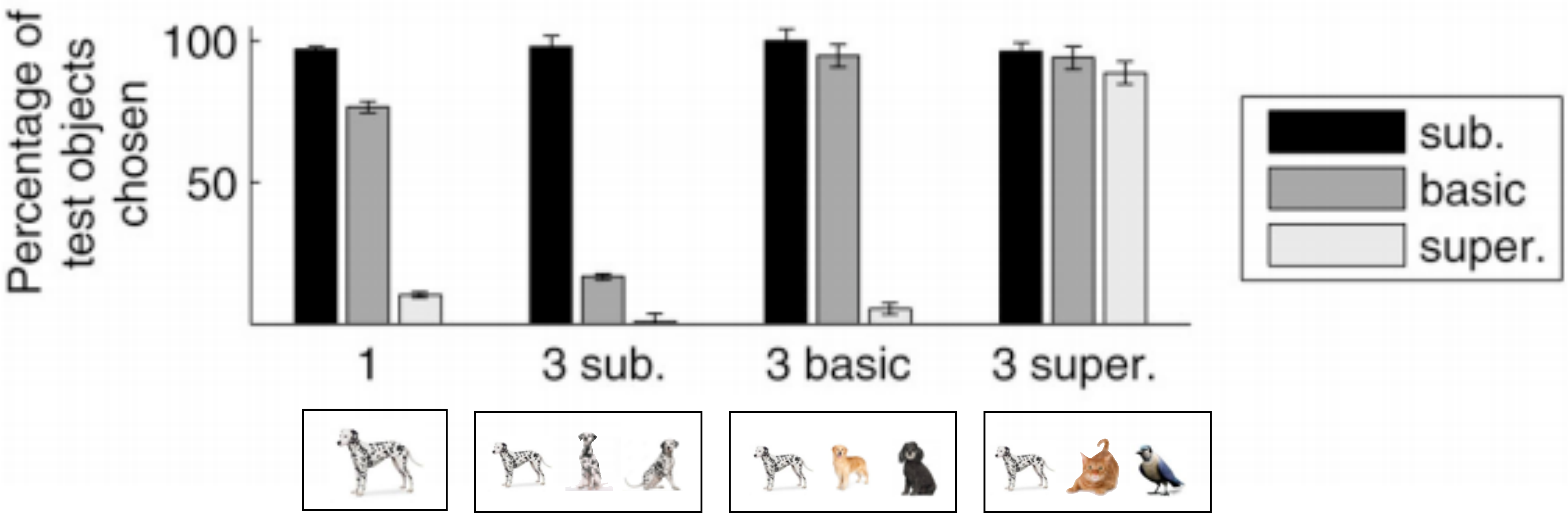
Please select the other objects that this word applies to (i.e., the other wugs).

Four-year-old children

Mr. Frog speaks a different language, and he has different names than we do for his toys. He is going to pick out some of them, and he would like you to help him pick out the others like he has picked out, OK?

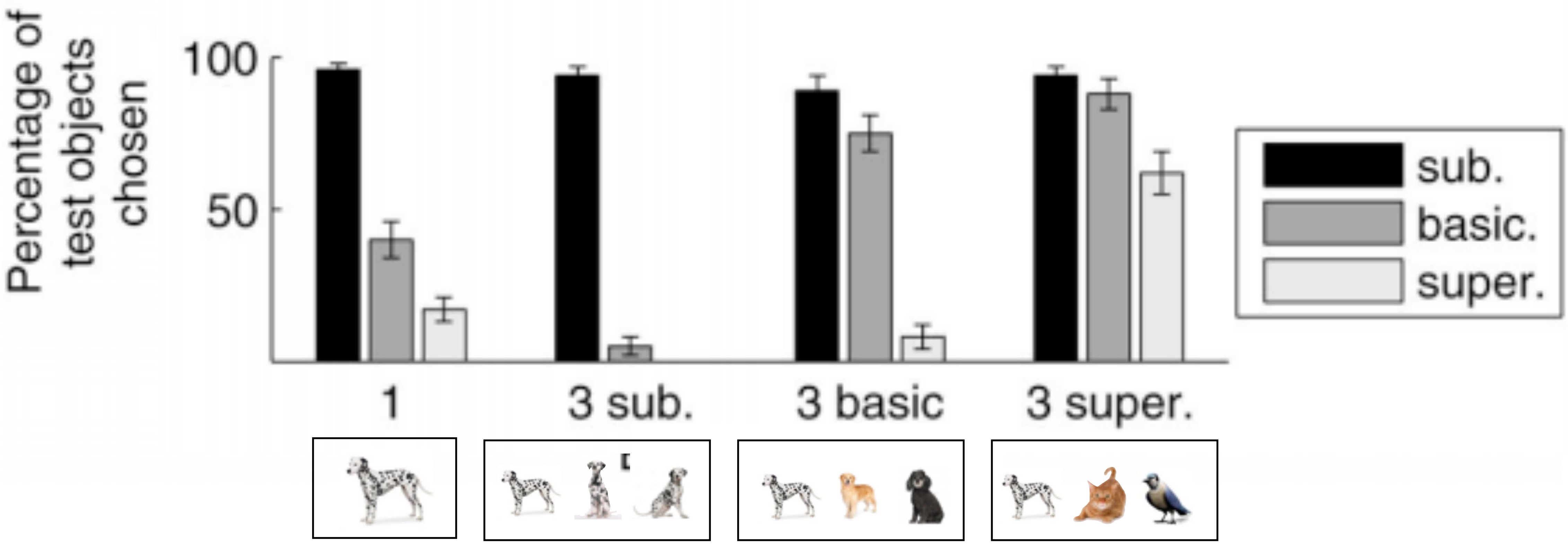
Test

- Adults generalise as predicted by the size principle

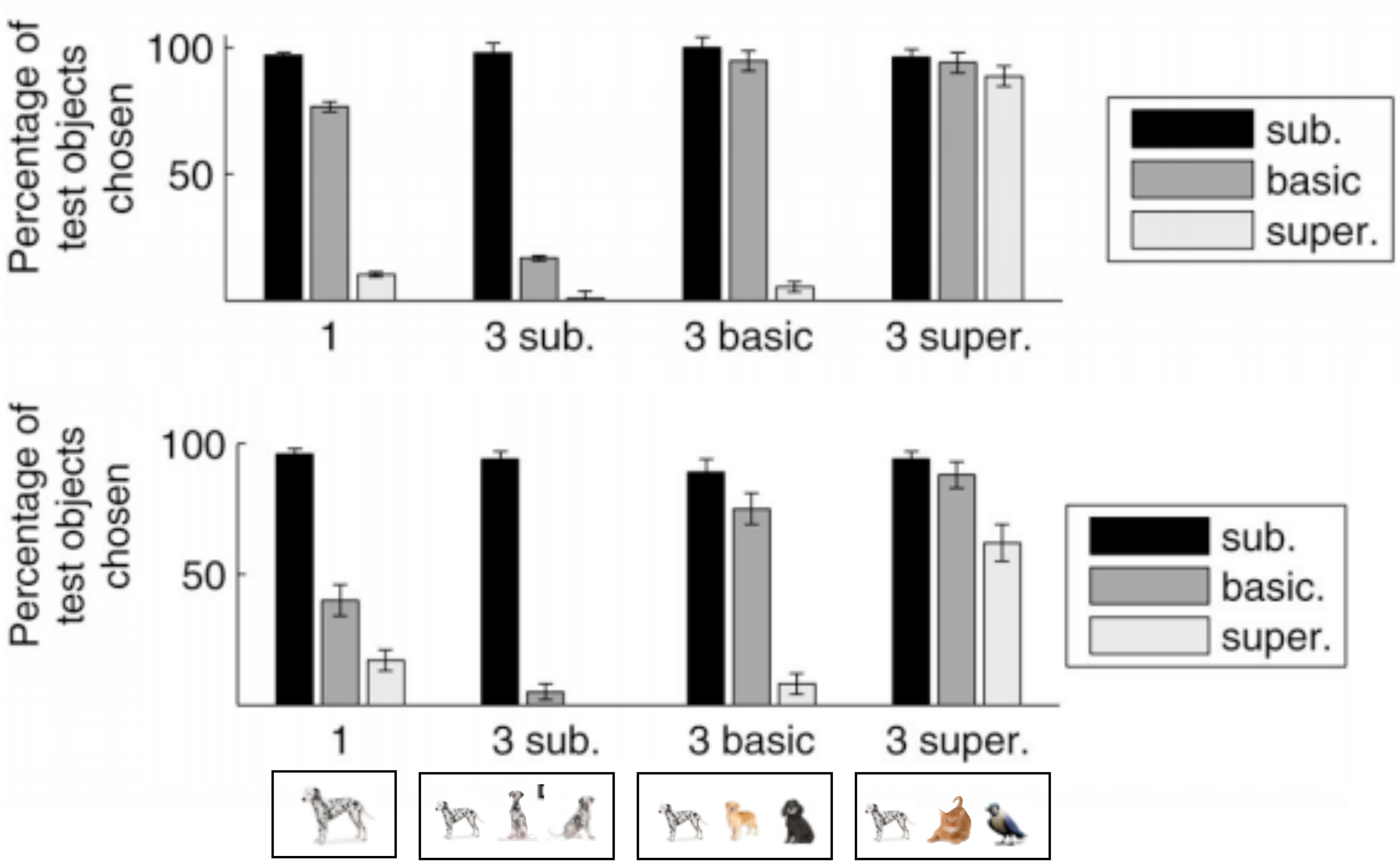


Test

- Four-year old children do the same thing!



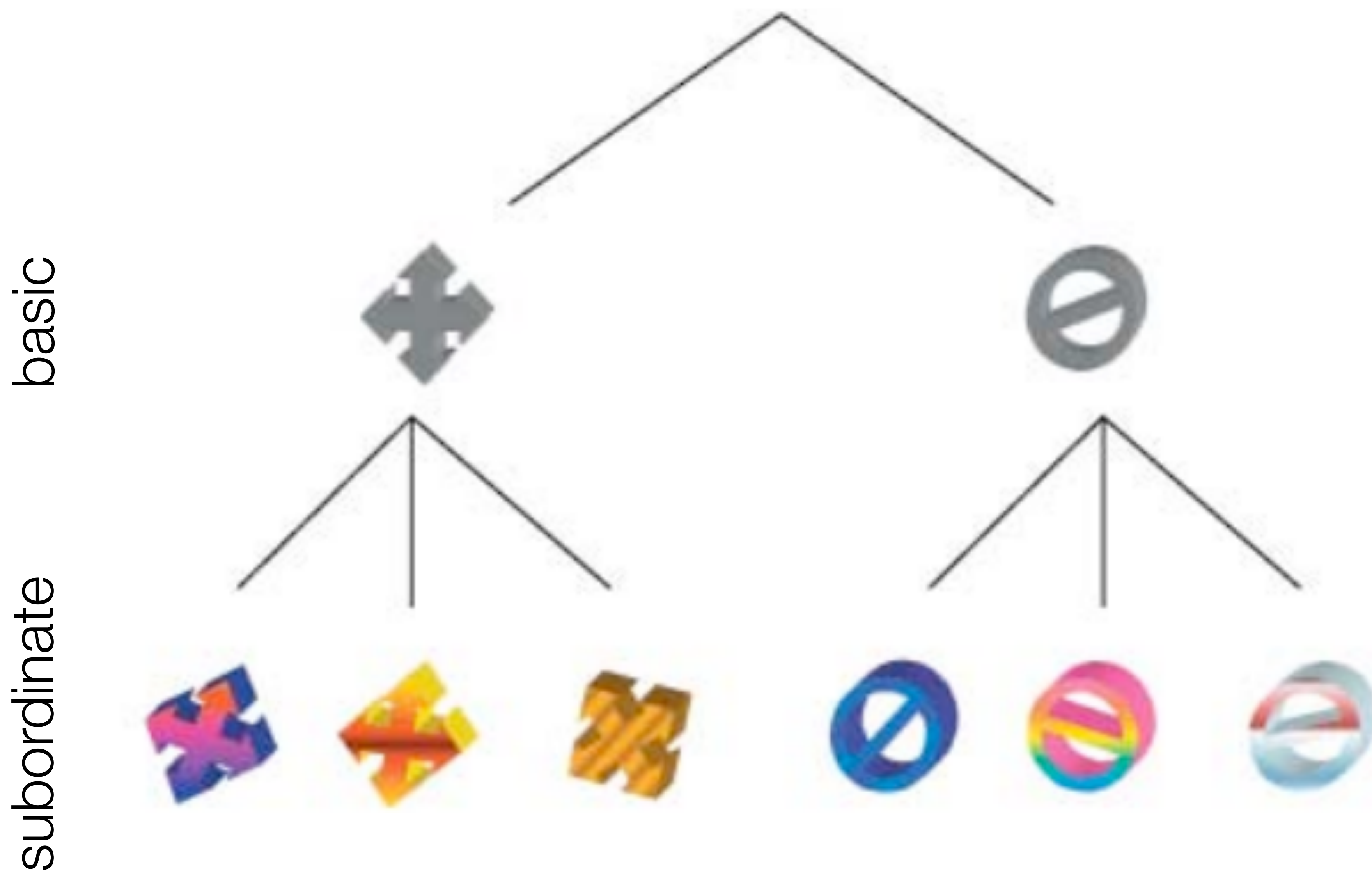
Test



- But so far this just shows that people follow the qualitative pattern predicted by the size principle. It does not imply that they are sensitive to sampling assumptions -- perhaps they would tighten generalisations no matter what

Changing sampling assumptions

- This time we vary how data are sampled (also make the objects novel)

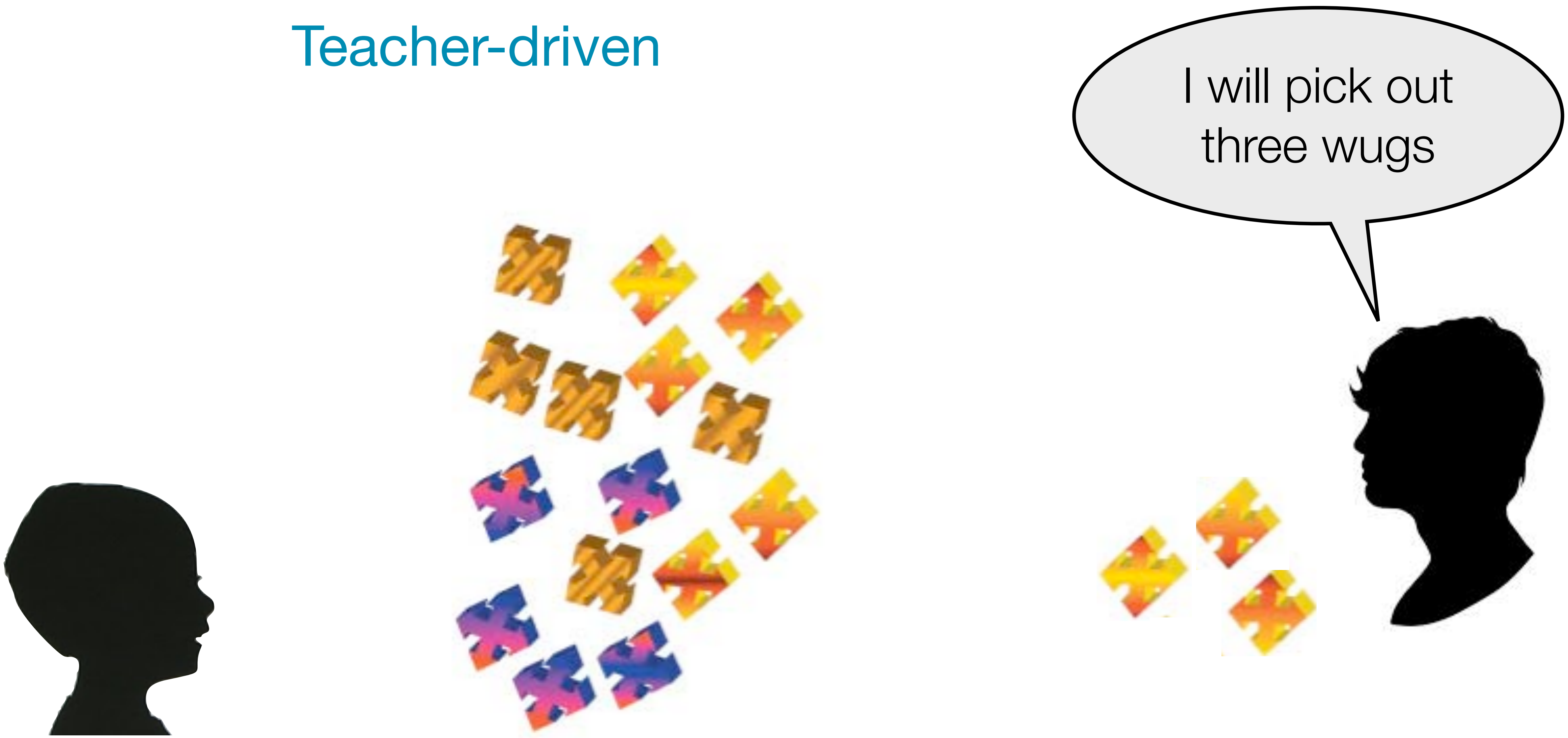


Xu & Tenenbaum, 2007b

Changing sampling assumptions

- This time we vary how data are sampled (also make the objects novel)

Teacher-driven

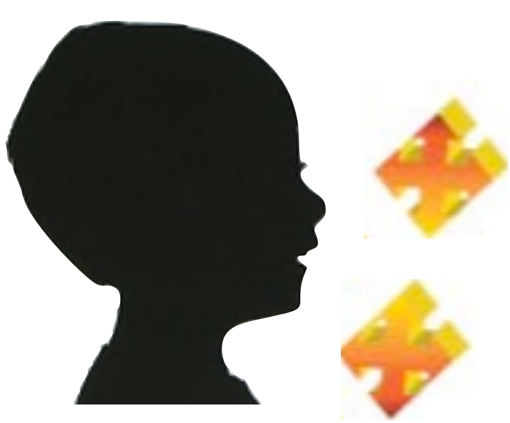


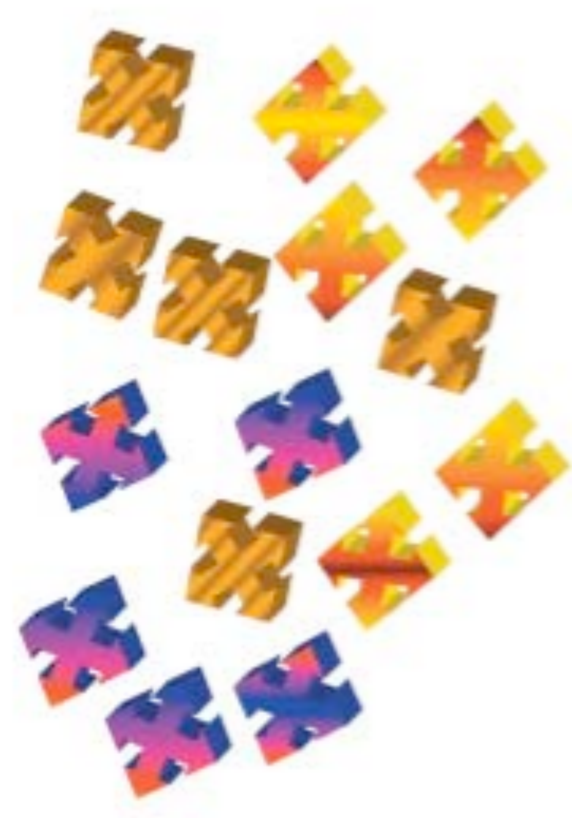
Changing sampling assumptions

- This time we vary how data are sampled (also make the objects novel)

Learner-driven

All participants chose two items from the same subordinate category





Changing sampling assumptions

- This time we vary how data are sampled (also make the objects novel)

Learner-driven

So in this condition people always saw items from the subordinate category, but the 3 items were not chosen by the teacher

Teacher-driven

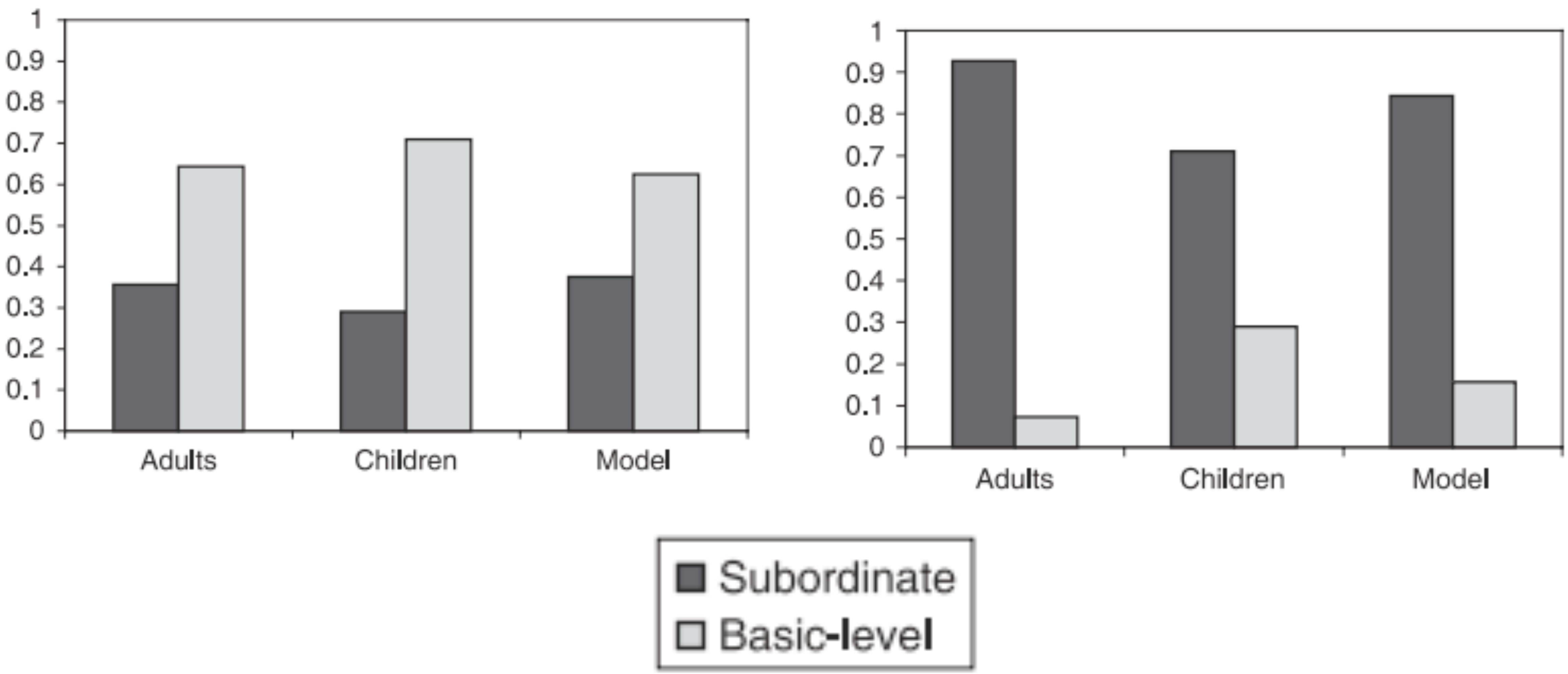
People saw 3 subordinate items, always chosen by the teacher

Changing sampling assumptions

- People generalise tightly only when the teacher sampled the data

Learner-driven

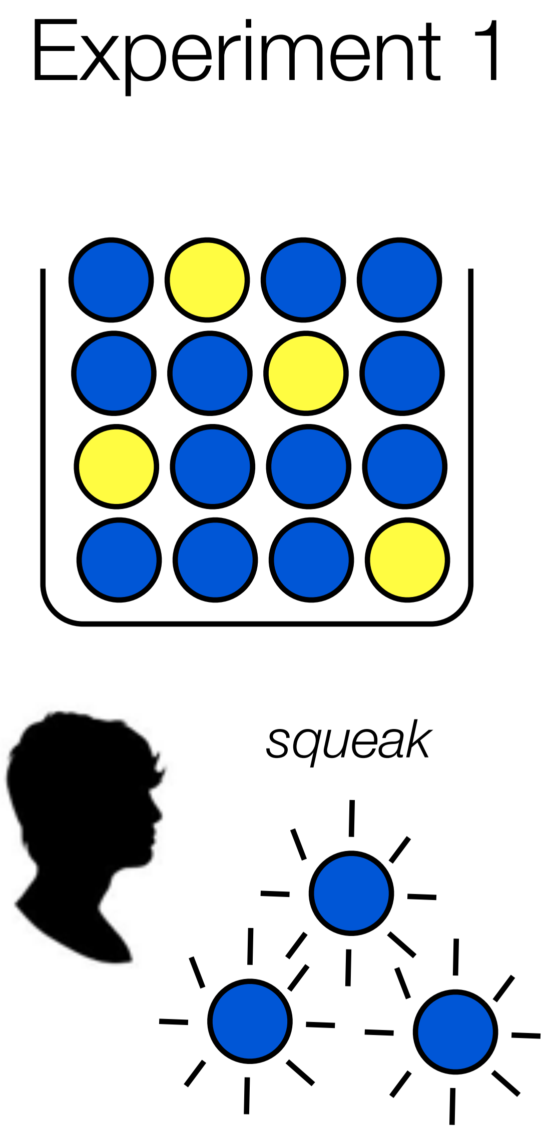
Teacher-driven

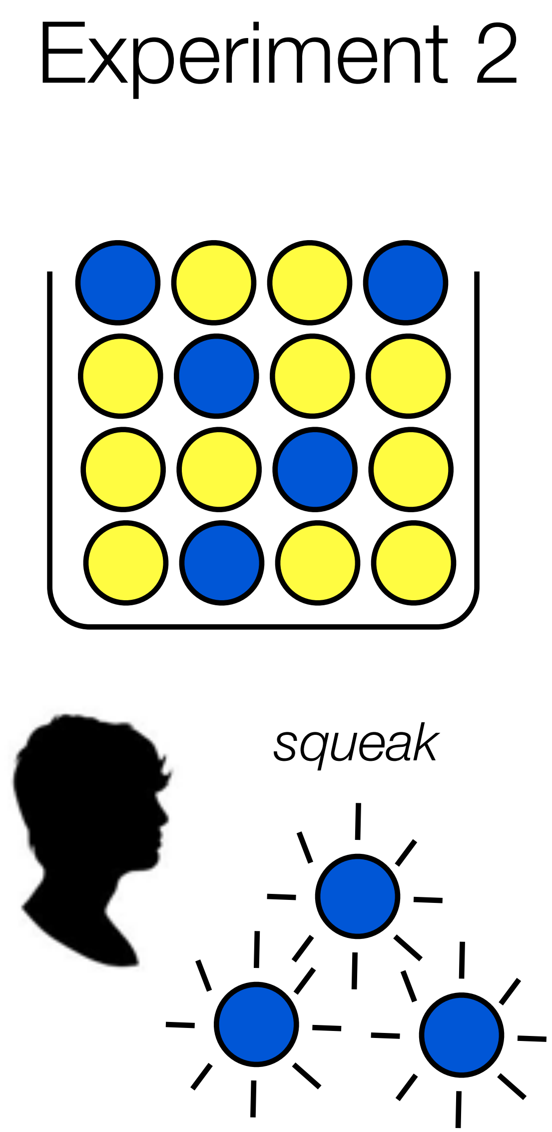


Changing sampling assumptions

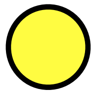
This shows that people generalise word labels differently based on how the data was sampled. How about generalising properties? And what about very young children?

Infants' use of sampling assumptions



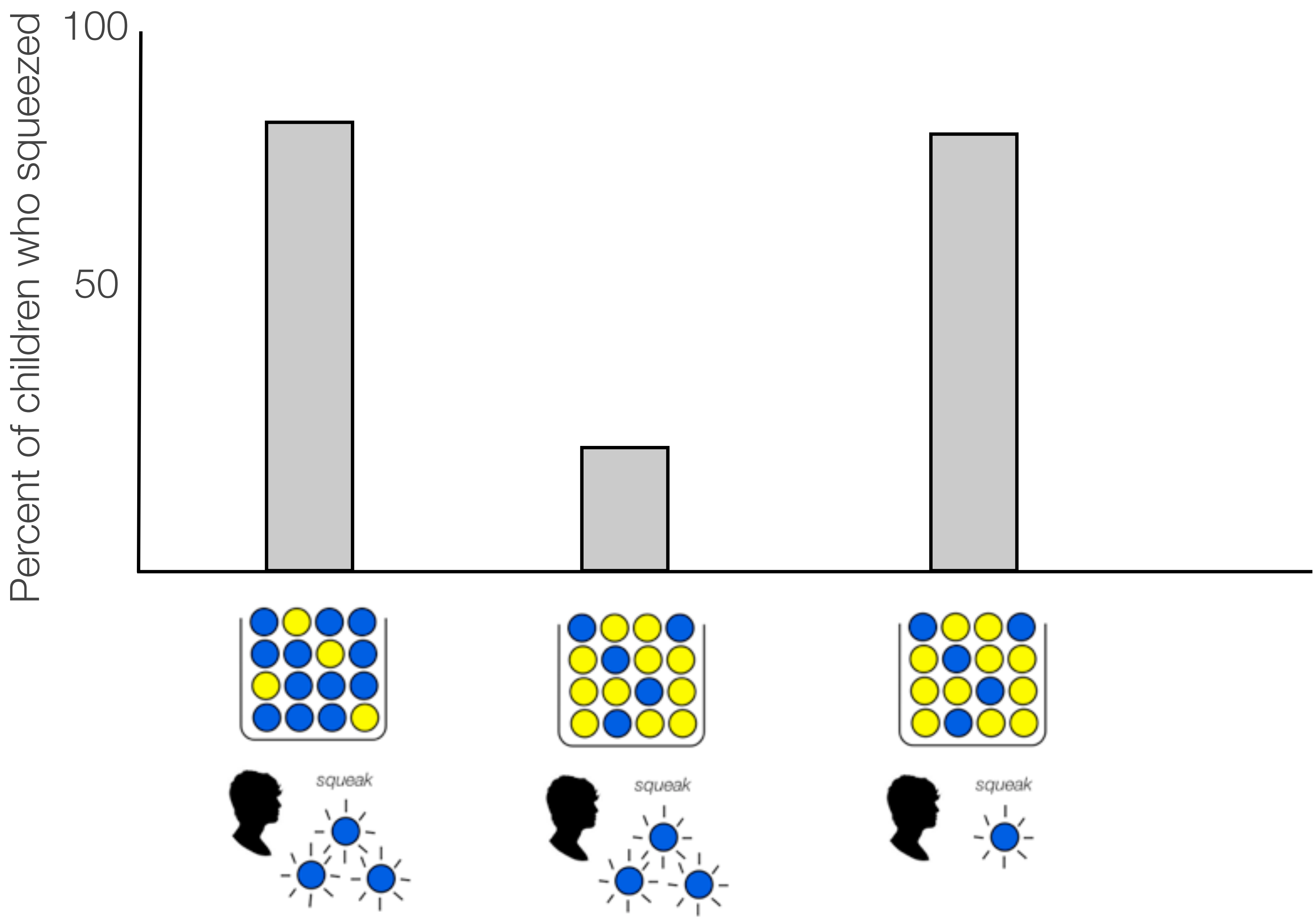


What do 15-month old infants think about this ball? Will it squeak?



Gweon et al., 2010

Infants' use of sampling assumptions

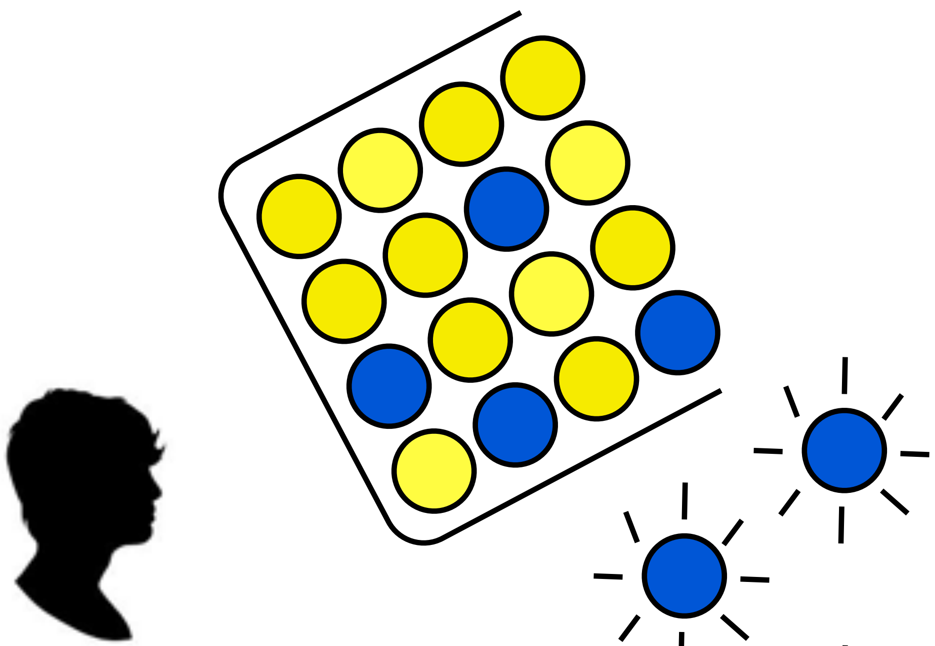


Gweon et al., 2010

Infants' use of sampling assumptions

This is all consistent with the size principle - but what if data is sampled differently?

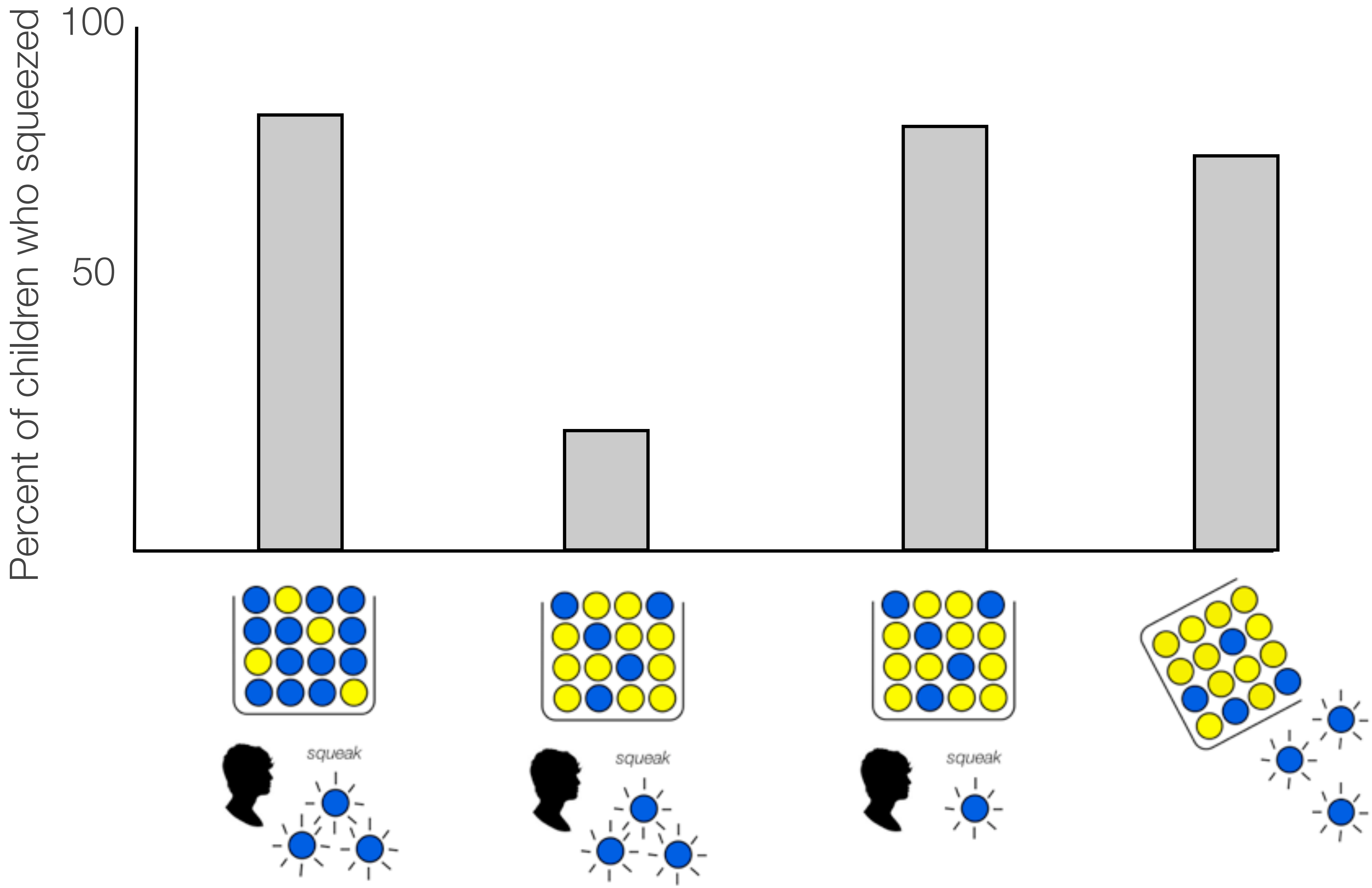
Experiment 4



If infants notice how the data were generated, they should not take this as an indication that yellow balls are not squeaky

Gweon et al., 2010

Infants' use of sampling assumptions



Gweon et al., 2010

Testing the size principle

So far all of this evidence has shown that people (including children) will tighten their generalisations more if they think the examples were generated from the concept/hypothesis directly.

This supports the qualitative ideas, but not necessarily the quantitative ones: people are sensitive to sampling assumptions, but do they tighten their generalisations *as much as* the size principle would predict? Are there individual differences in this?

Testing the size principle

- We can capture the degree to which people assume that any point was strongly sampled (and the size principle should therefore apply)

$$P(x, x \in h \,|\, h, \theta) = \begin{cases} (1-\theta)\dfrac{1}{|\mathcal{X}|} + \theta\dfrac{1}{|h|} & \text{if } x \in h \\ 0 & \text{otherwise,} \end{cases}$$

generalisation probability

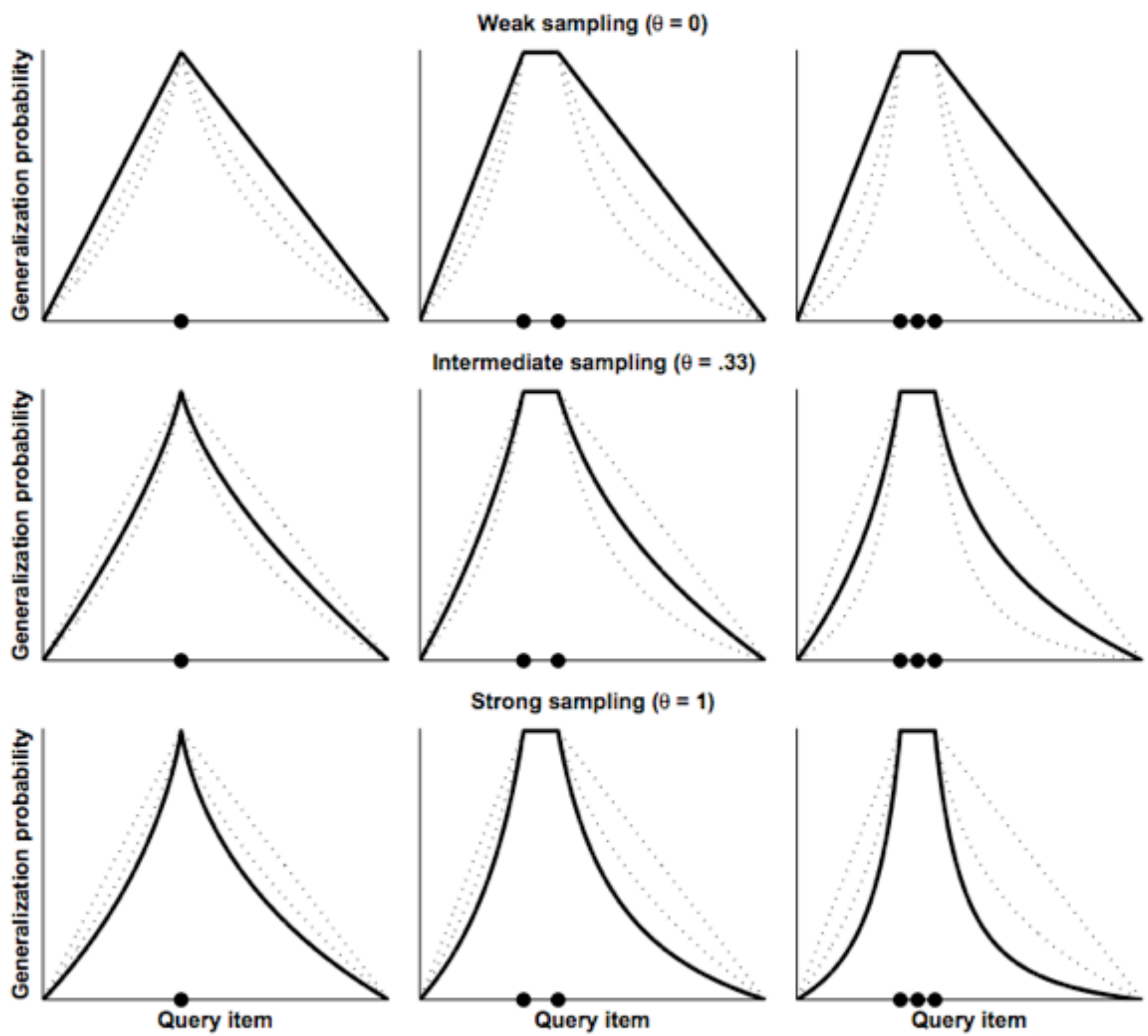
number of total possible items in the world

probability that any observation is strongly sampled

size principle

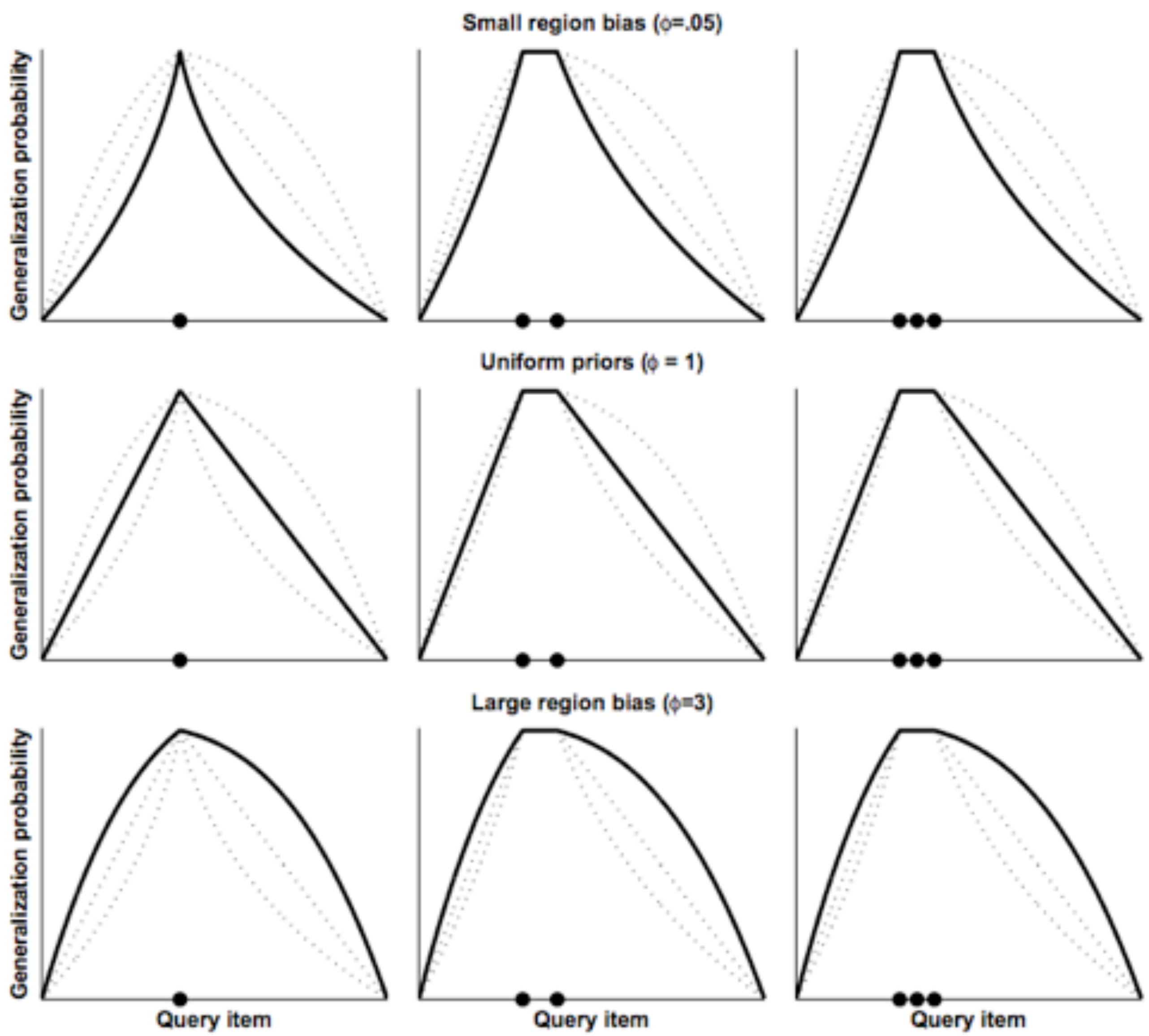
Testing the size principle

- Higher θ leads to tighter generalisations (model)



Testing the size principle

- Note that this is different from a prior; the prior Φ guides how large you think the region is, θ is how much your generalisation tightens with additional data

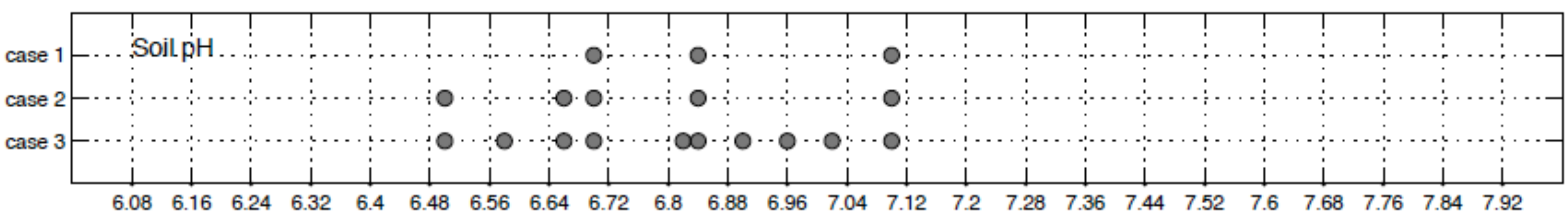


vary prior Φ,
weak sampling
($\theta = 0$)

Testing the size principle

- Task: give people data points that vary on a continuum, and look at how their generalisations change with additional data

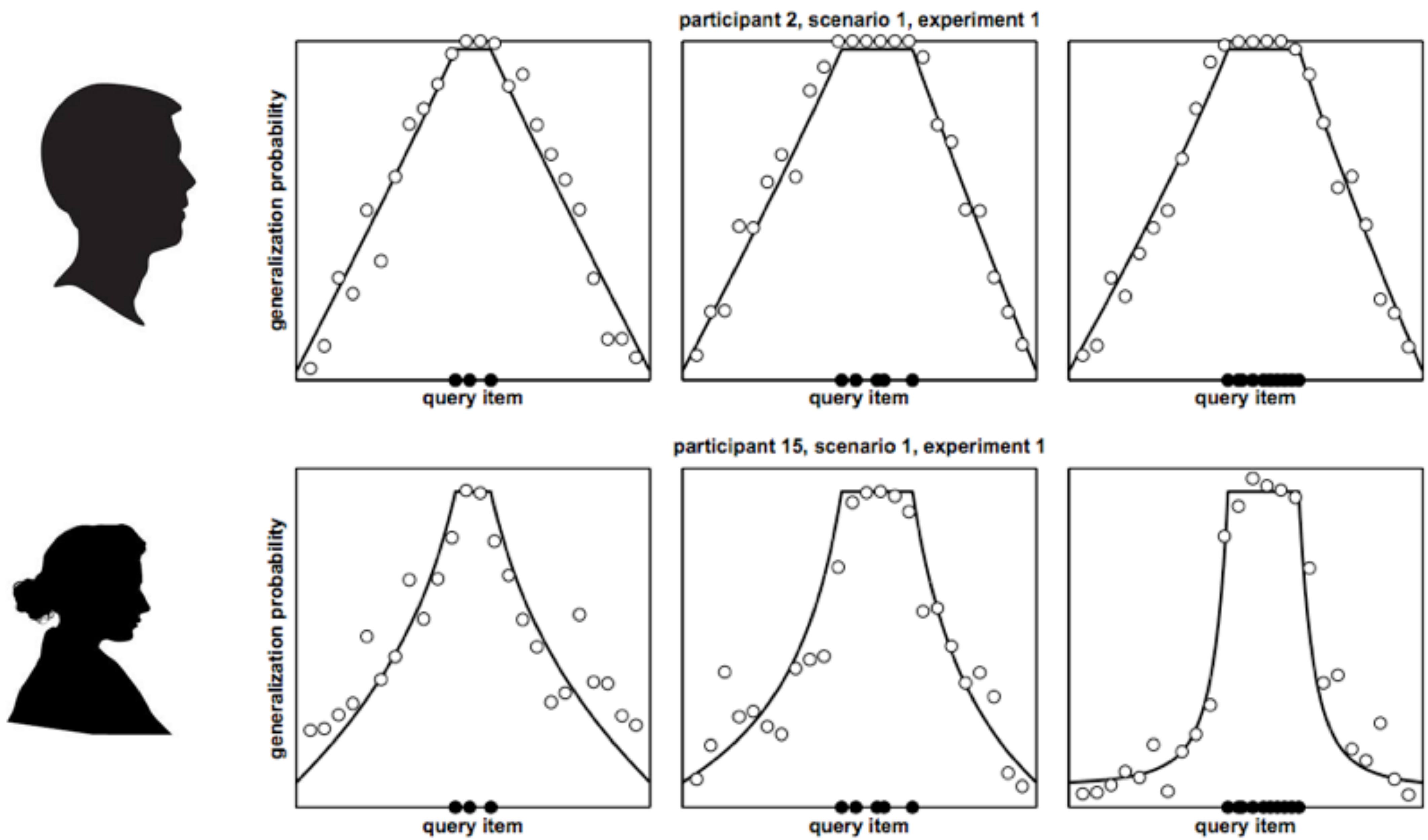
> The colour of the flowers of Hydrangea macrophylla change depending upon soil pH levels. Soils with a pH of less than 6 produce blue flowers, and soils with a pH greater than 8 produce pink flowers. Neutral soils tend to produce very pale cream flowers. Given that cream flowers were produced by Hydrangeas growing in soils with the pH levels shown as black dots below, what is the probability that Hydrangeas would also produce cream flowers if they were grown in soil with the pH level specified by the red question mark?



Three tasks: bandicoot foraging hours, bacteria temperatures, and flowers growing in soils of different pH

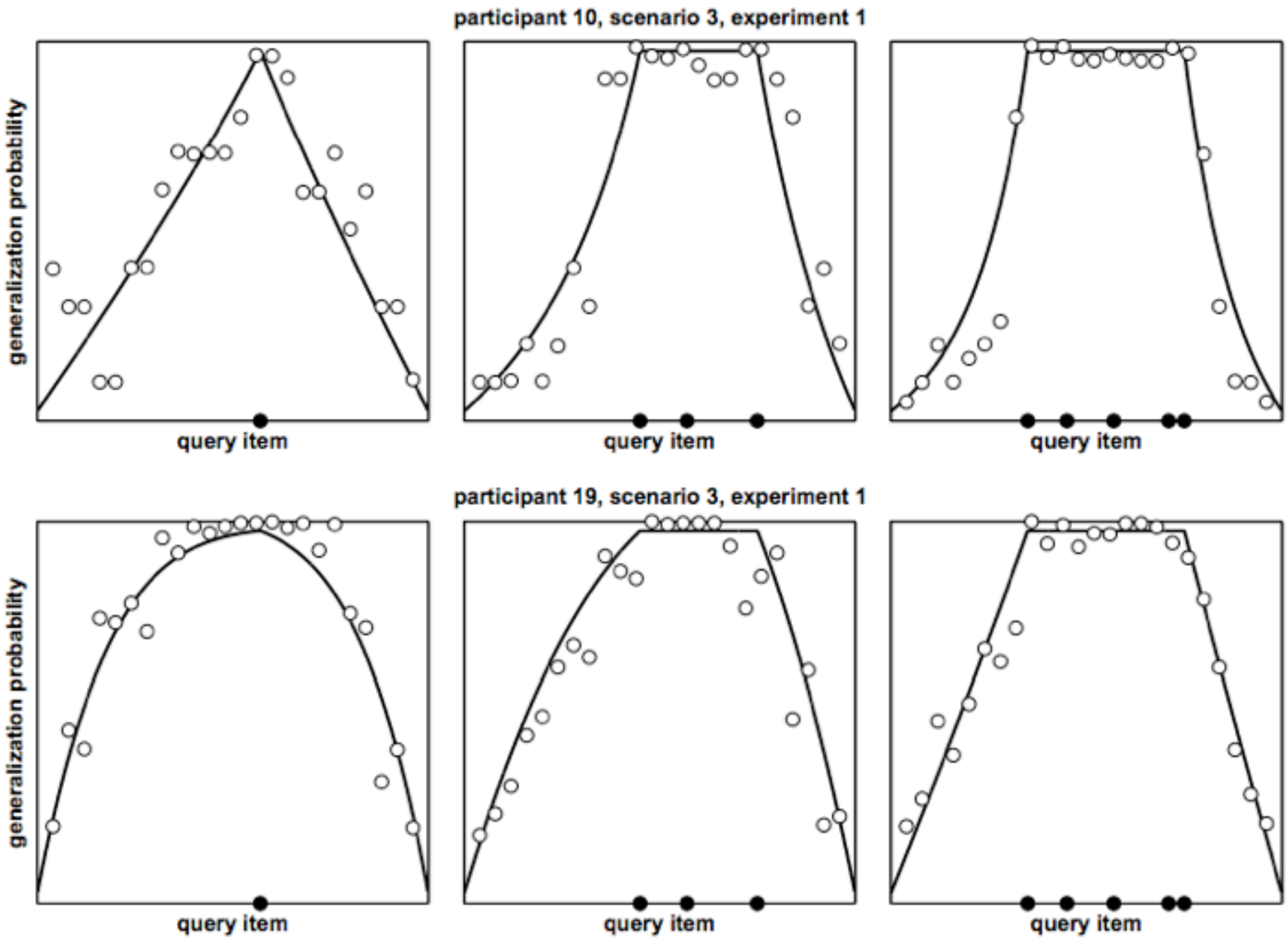
Testing the size principle

- People appear to vary in the degree to which they assume strong sampling



Testing the size principle

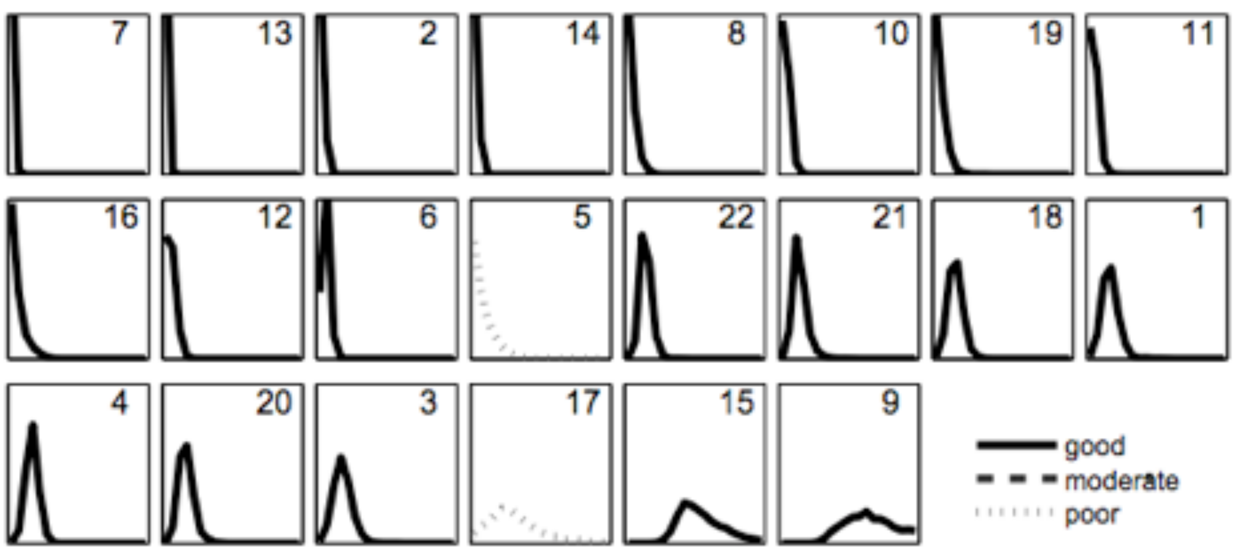
- They also vary in their priors



Testing the size principle

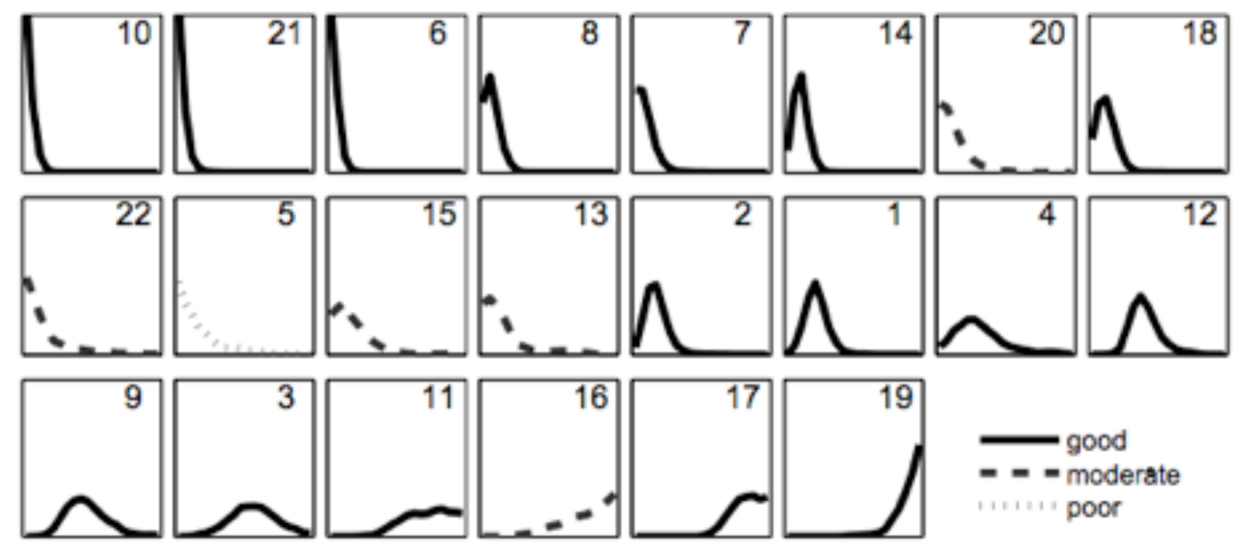
- Individual differences in the degree of sampling assumptions

Bacteria temperatures

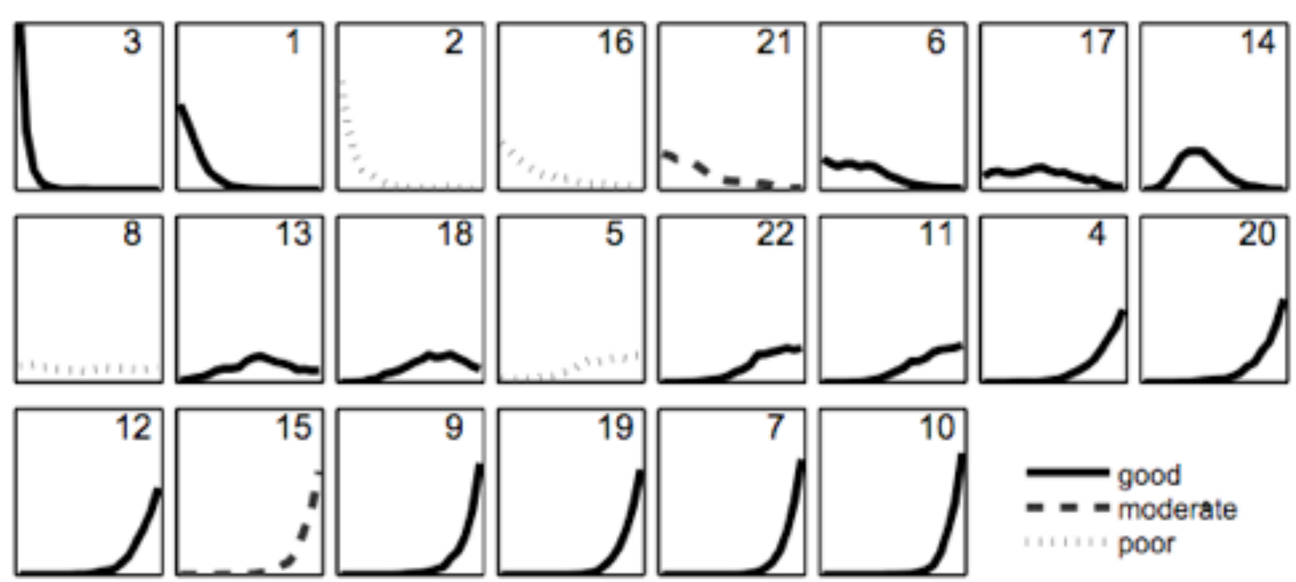


varying θ
(0 to 1)

Soil pH

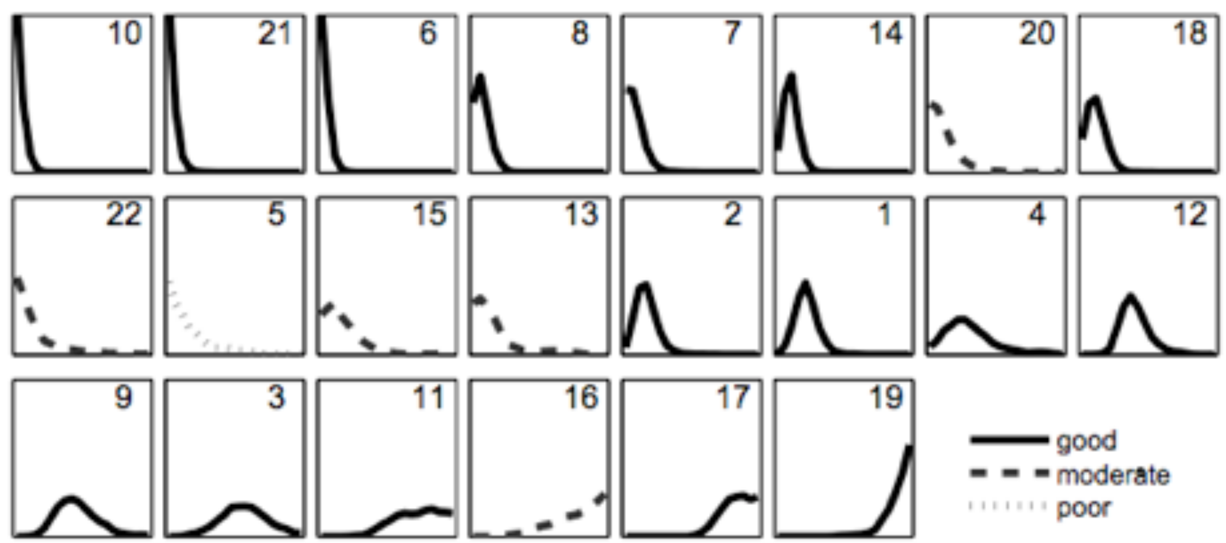


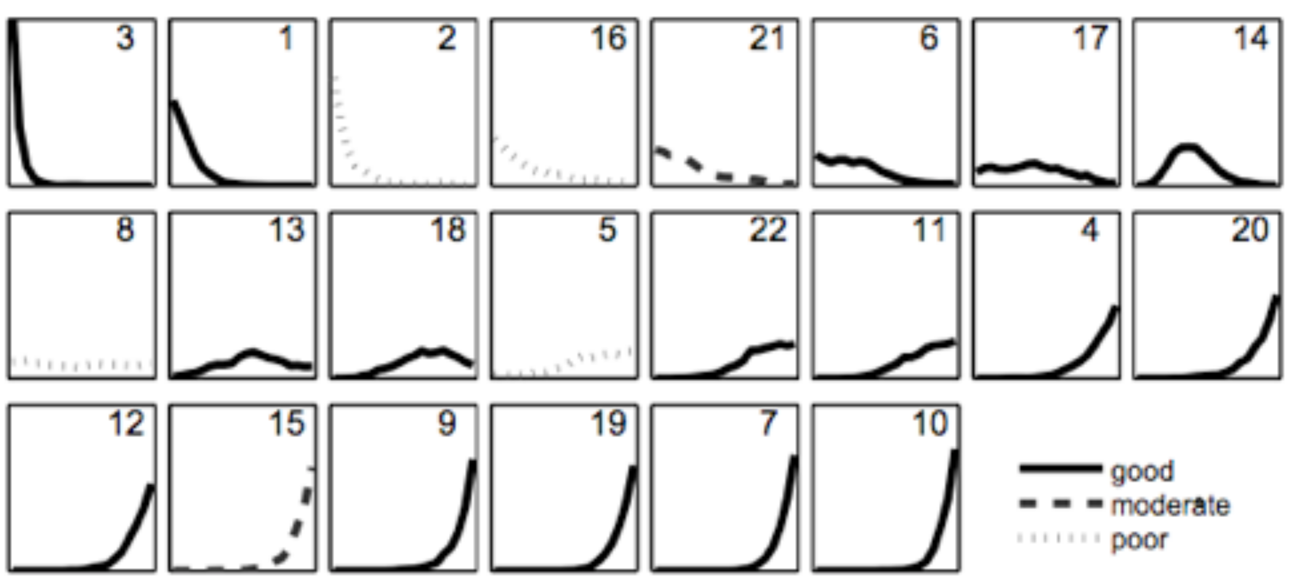
Bandicoot foraging

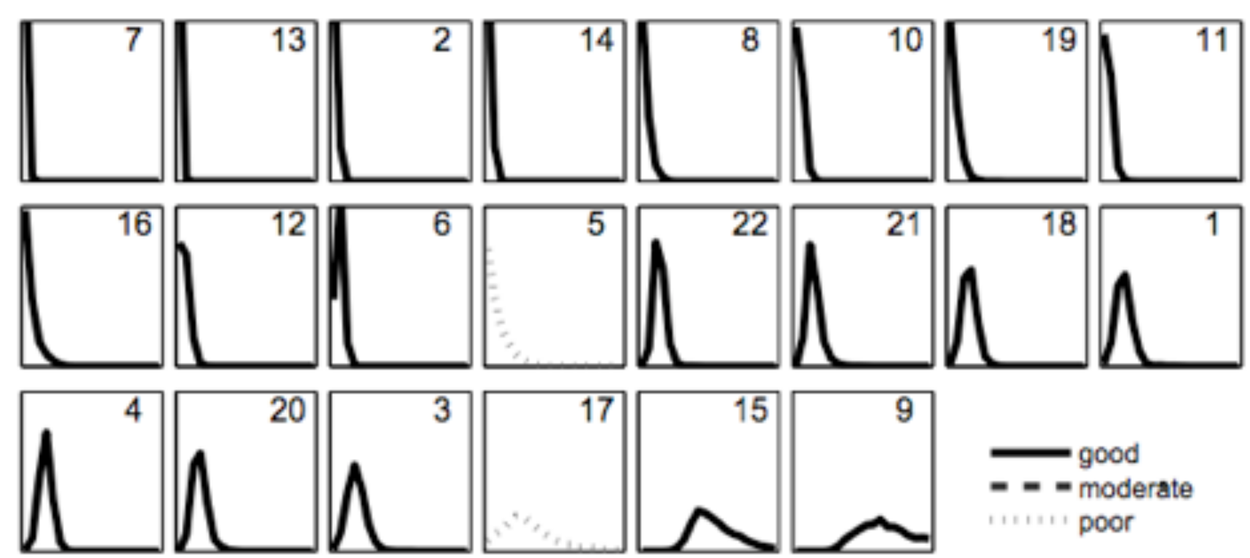


Testing the size principle

- Individual differences in the degree of sampling assumptions
- Note that there is high variance, but very very few don't tighten their generalisations at all (i.e., have $\theta=0$)

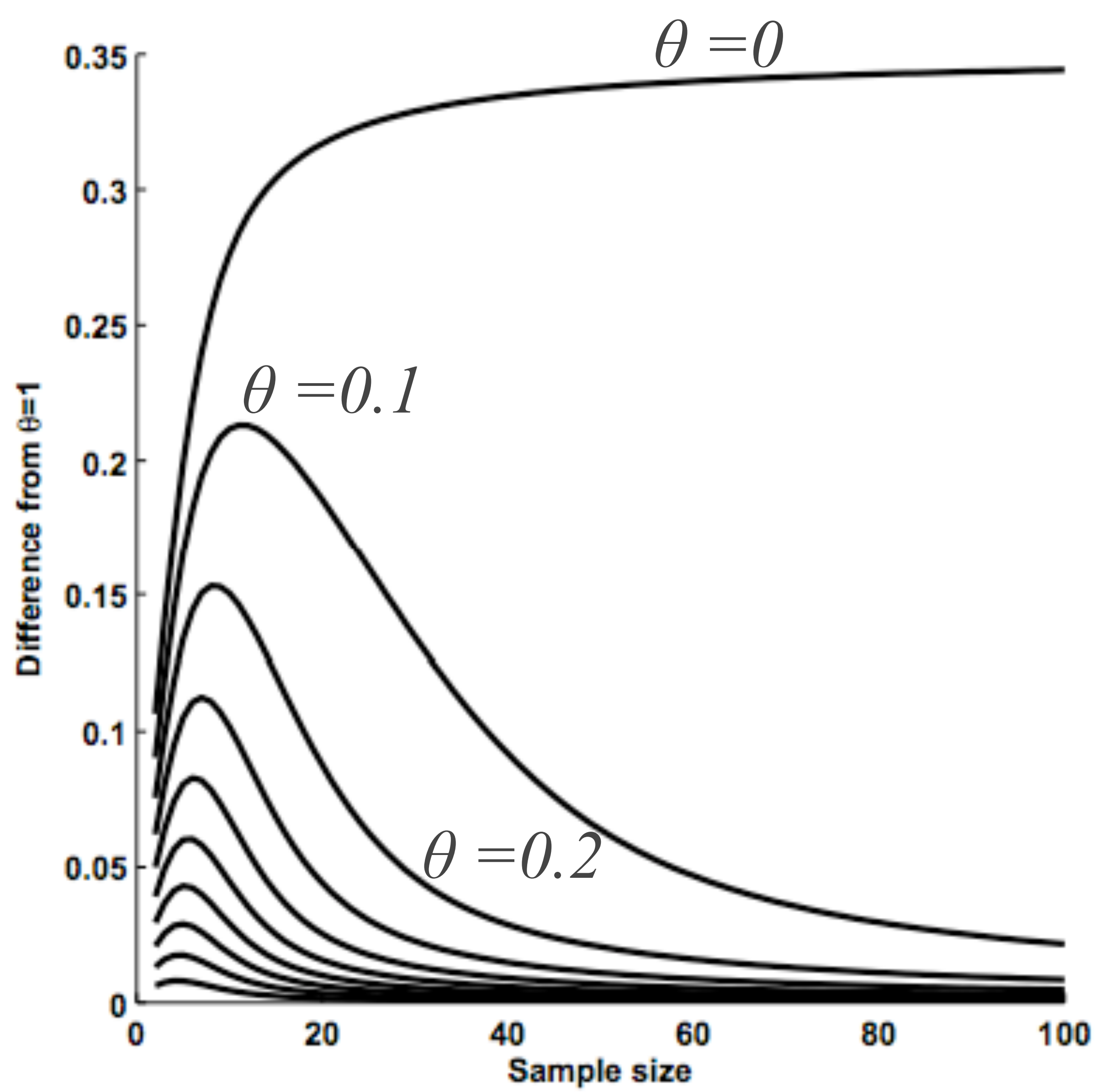






Testing the size principle

- Interestingly, in the long run, *any* θ greater than zero ends up looking the same: you get the same eventual tightening with enough data

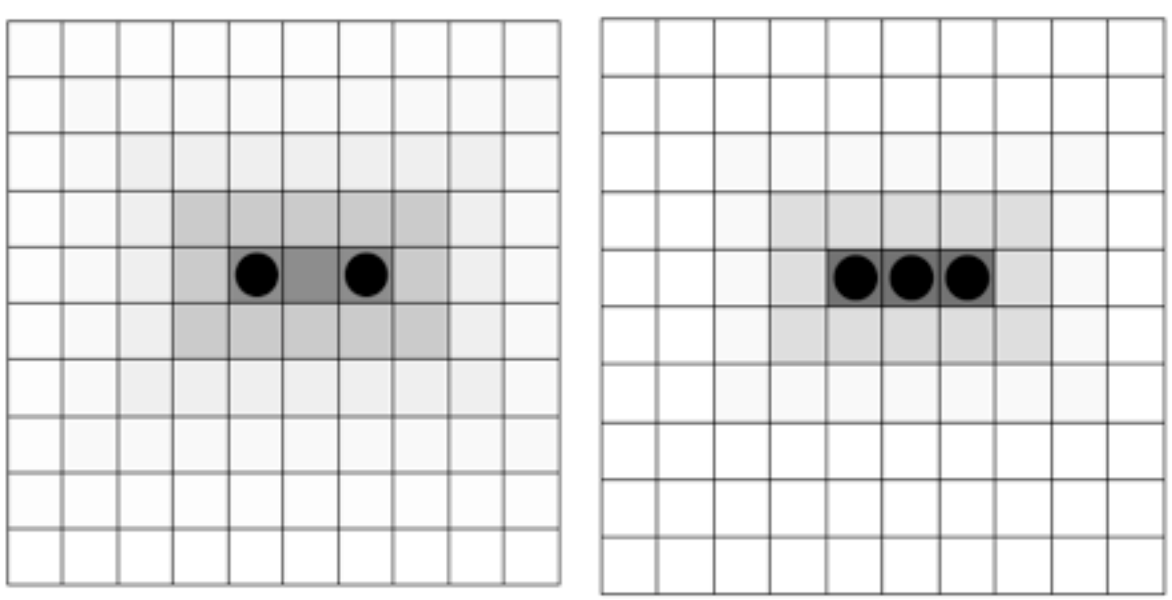


Summary

- Difference between strong and weak sampling

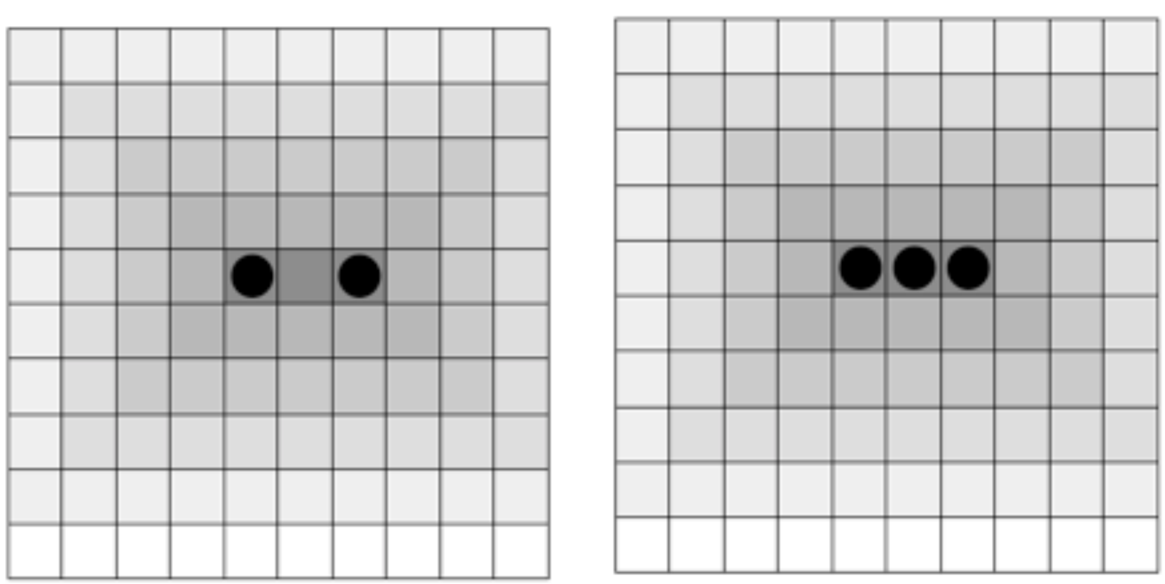
Strong

- Items generated from concept
- Additional items lead to tighter generalisations



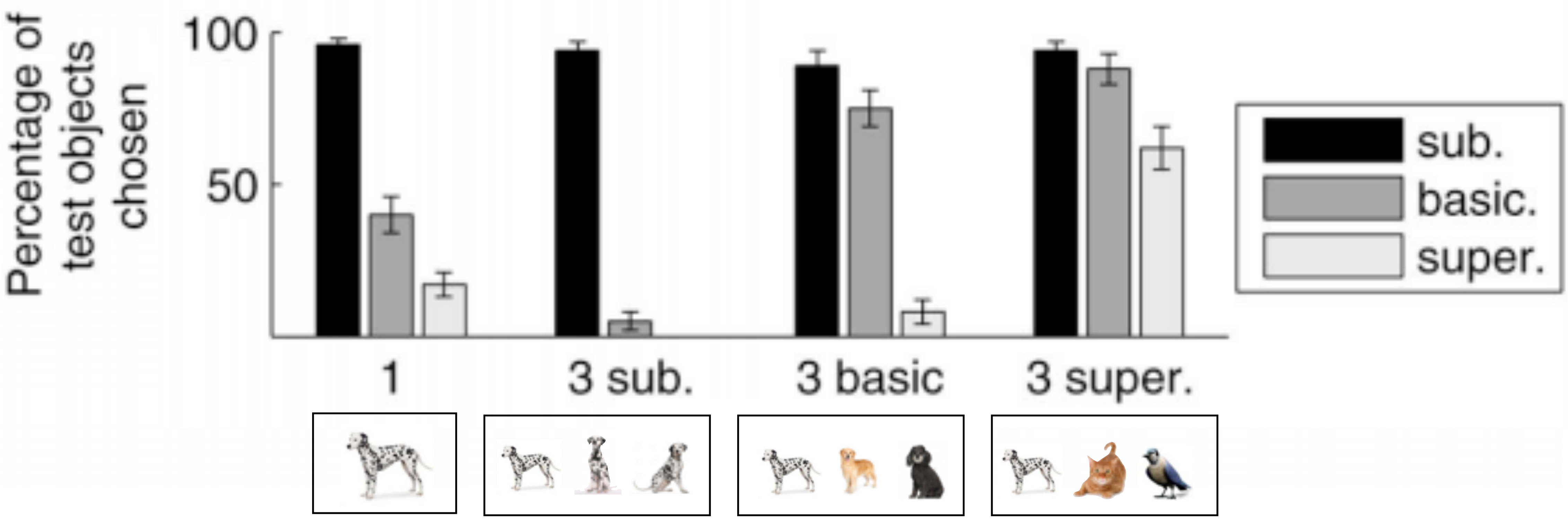
Weak

- Items generated from world and then labelled
- Additional items do not lead to tighter generalisations



Summary

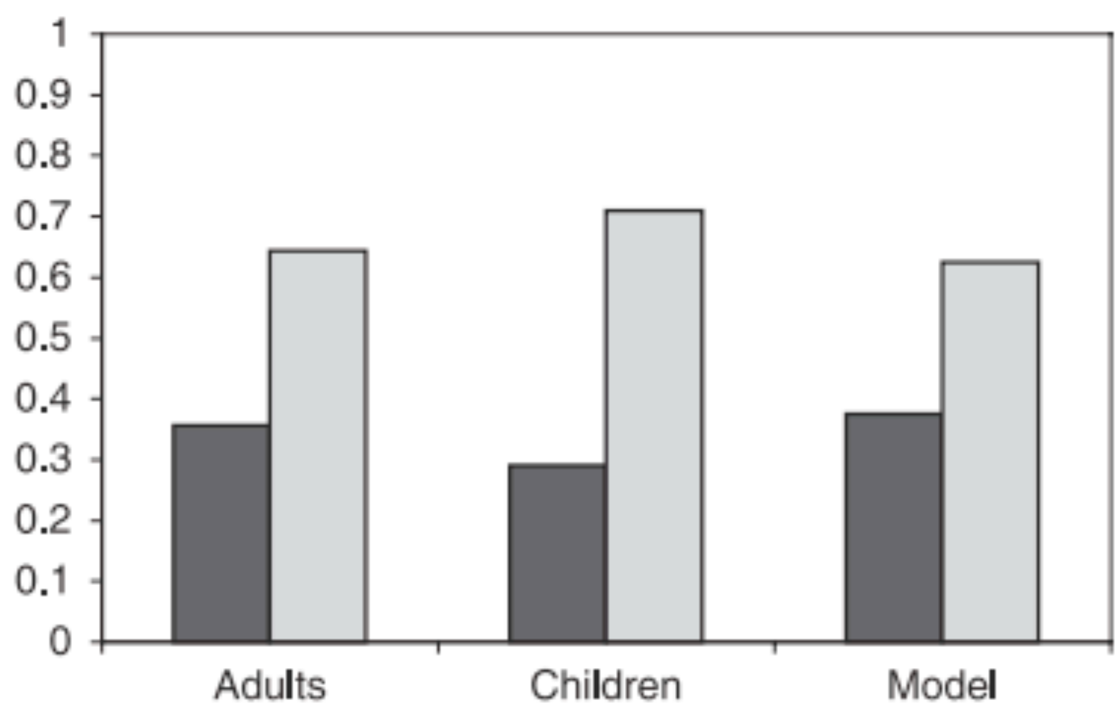
- Difference between strong and weak sampling
- People pay attention to how data were sampled when figuring out how to generalise words



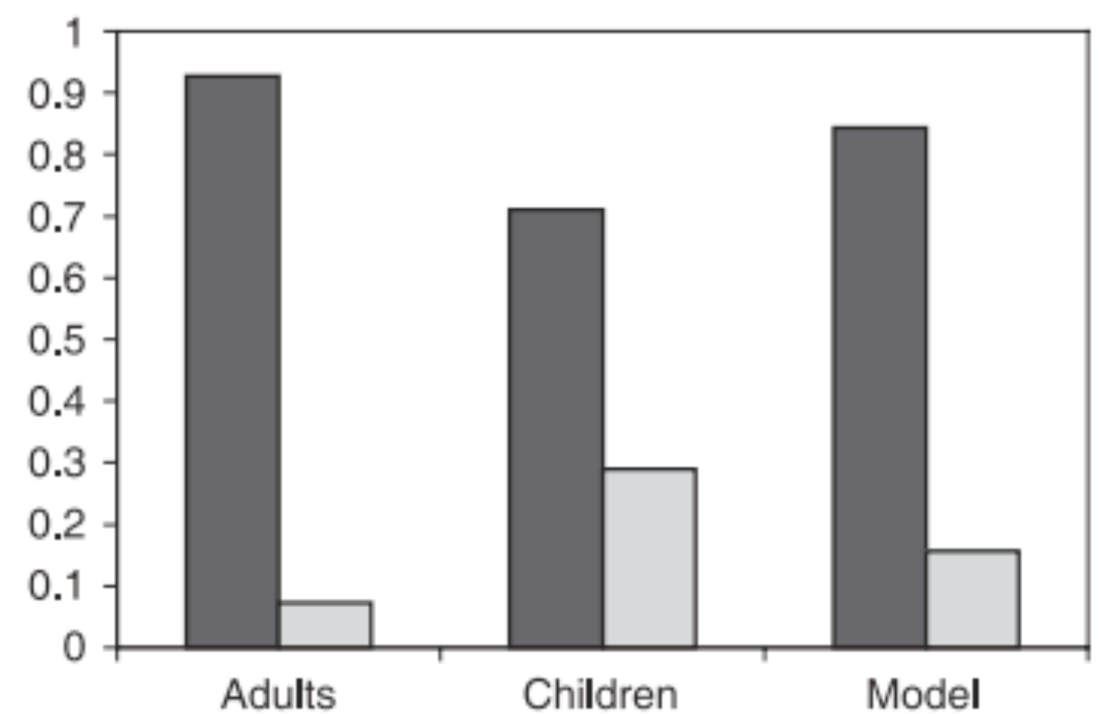
Summary

- Difference between strong and weak sampling
- People pay attention to how data were sampled when figuring out how to generalise words, changing their generalisations if they did not come from the concept

Learner-driven

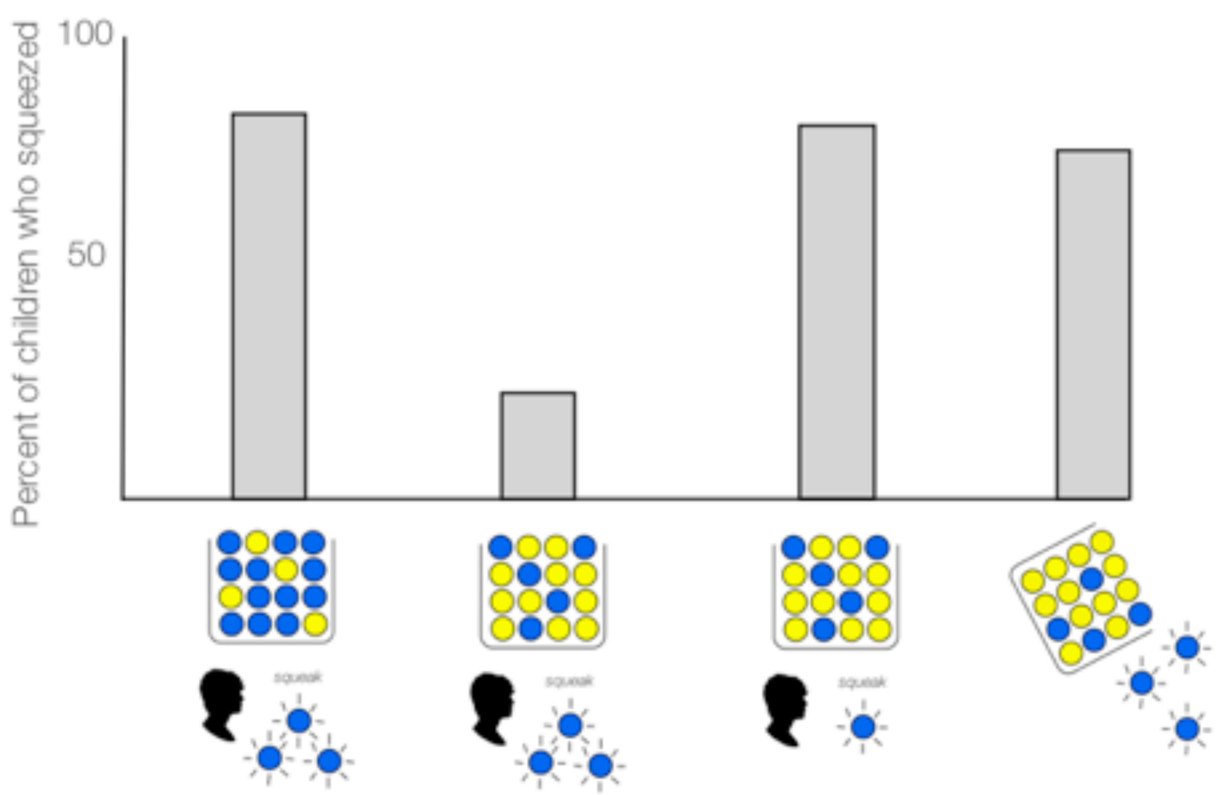


Teacher-driven



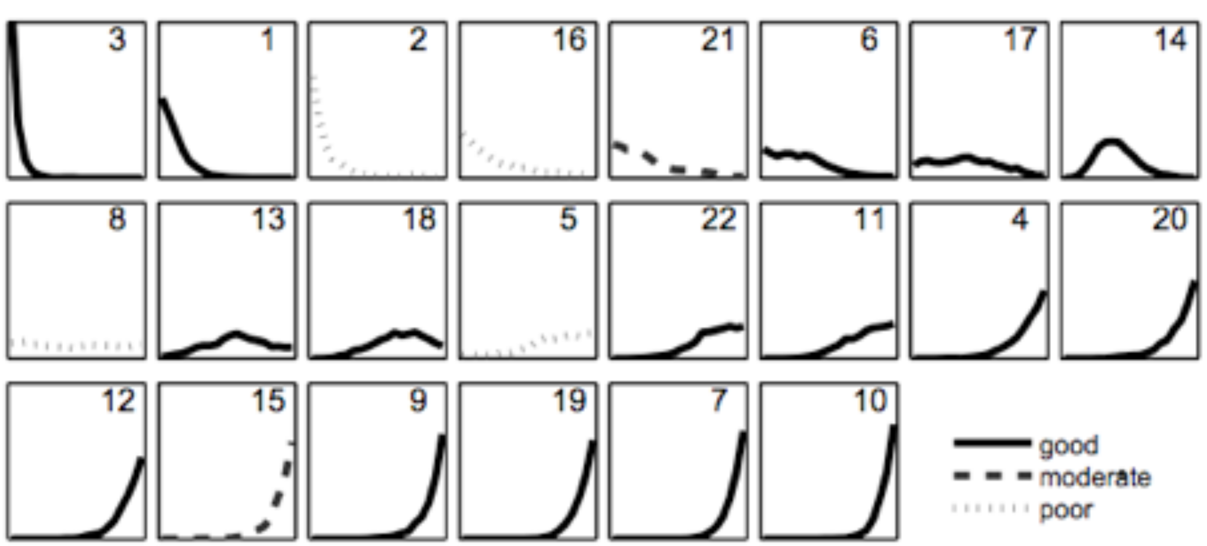
Summary

- Difference between strong and weak sampling
- People pay attention to how data were sampled when figuring out how to generalise words, changing their generalisations if they did not come from the concept
- Even infants do this, and with novel features



Summary

- Difference between strong and weak sampling
- People pay attention to how data were sampled when figuring out how to generalise words, changing their generalisations if they did not come from the concept
- Even infants do this, and with novel features
- People show strong individual differences in the amount to which they assume strong sampling, but almost always they tighten at least somewhat



Additional references (not required)

Weak and strong sampling

- Gweon, H., Tenenbaum, J., & Schulz, L. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences 107*(20): 9066-9071
- Navarro, D., Dry, M., & Lee, M. (2012). Sampling assumptions in inductive generalization. *Cognitive Science 36:* 187-223.
- Tenenbaum, J., & Griffiths, T. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences 24*: 629-641.
- Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review 114*: 245-272.
- Xu, F., & Tenenbaum, J. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science 10*: 288-297.