DAY 1: EXPERIMENT

Not just coding it up, but all workflow stuff up to running it

1. Background: replicability and proper procedure
2. Workflow and organisation
3. Experiment design
4. Coding experiment
5. Ethics and pre-registration
6. Hosting experiment on a server
7. Downloading data

ORGANISATION MATTERS!

- Folder structure and documentation ensures replicability and will save you *so much time* in the long run
- People can differ but the main things are:
  - Clear organisation that is stable from project to project
  - Shared structure among people on the same project
  - Folder organisation that parallels and supports your workflow

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

docs

experiment

analysis

models

Please make this directory structure for yourself right now in your summer school folder. We will be assuming throughout this week that you are following it

(tomorrow we'll show you git but we'll keep it simple for now)

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

experiment

expt1

expt2

analysis

expt1

expt2

Very modulariseable if you add additional experiments or analyses to the same project

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

Any documents you make for yourself, e.g. describing design decisions, or figures (that are not data or model outputs)

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

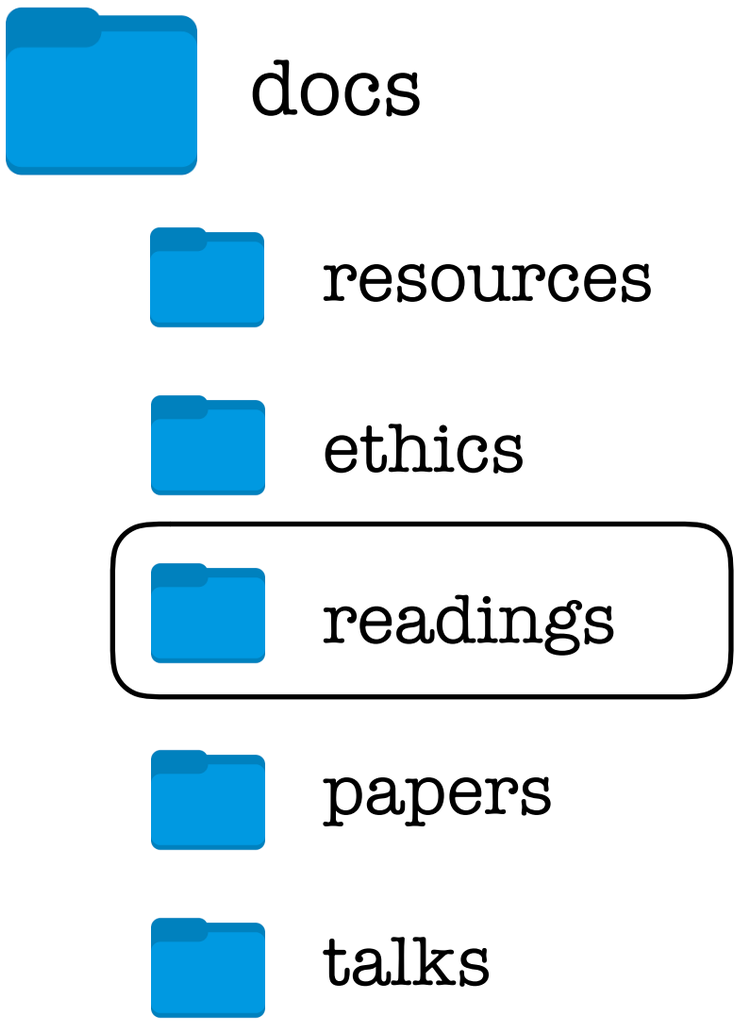
- docs
  - resources
  - ethics
  - readings
  - papers
  - talks

All ethics and ethics related documents

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)



Relevant papers and your notes on them — your future self will thank you!
(if you use LaTeX or a bibliography manager, integrate this as appropriate)

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

- docs
  - resources
  - ethics
  - readings
  - papers
    - cogsci18
    - cognition
  - talks

This is what you write - also expandable

Suggestion: LaTex and overleaf

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

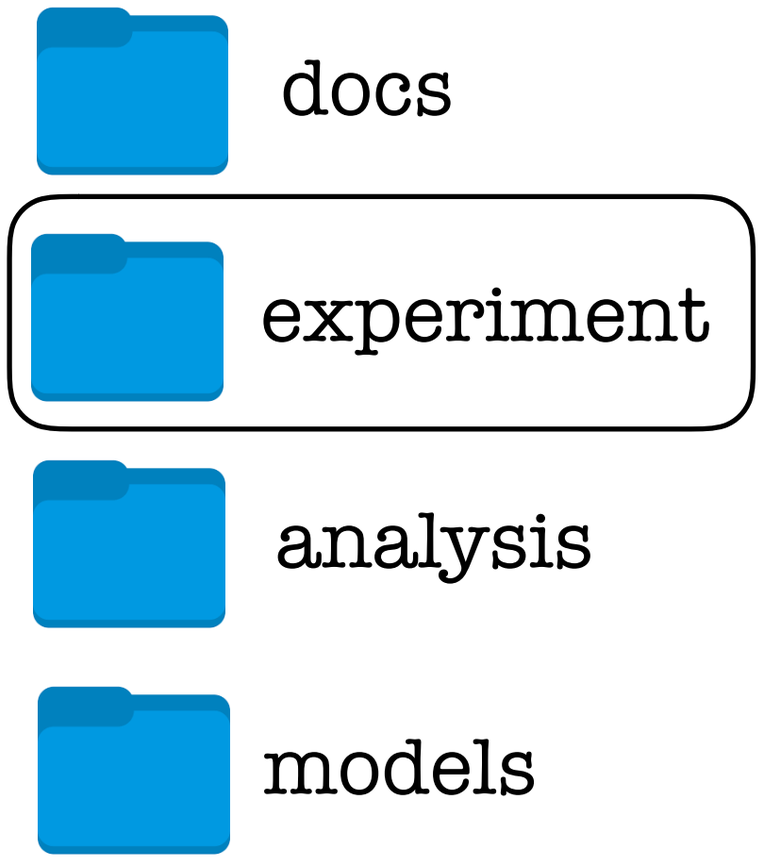
- docs
- resources
- ethics
- readings
- papers
- talks

Any talks or presentations (including posters) about this

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)



SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

experiment

resources

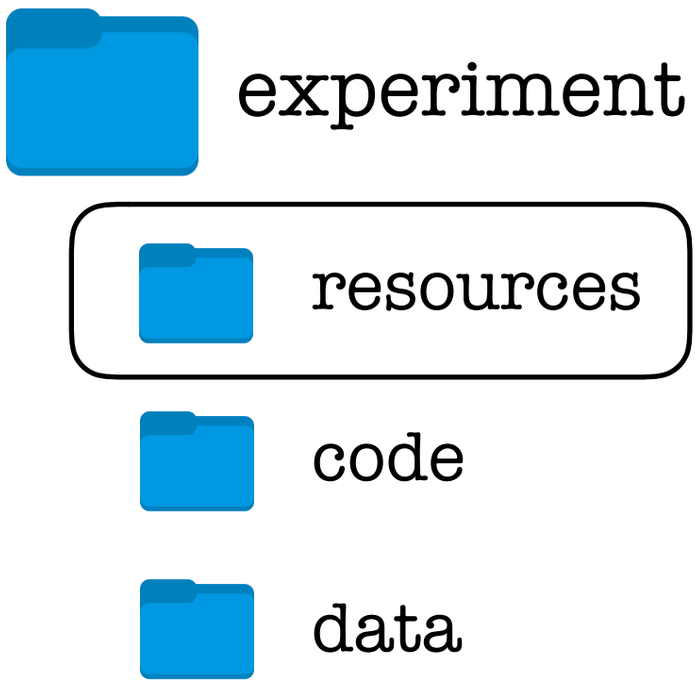
code

data

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

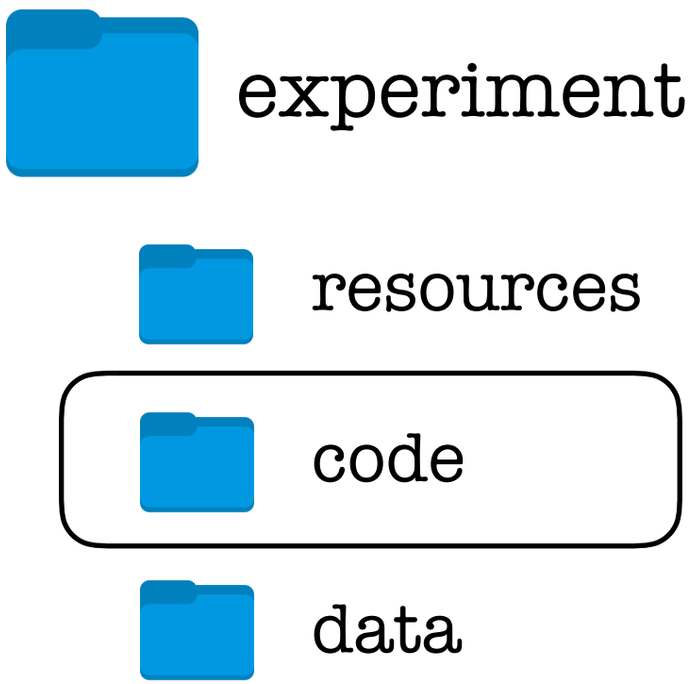


Experiment-specific resources like stimuli (through the process of development)

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

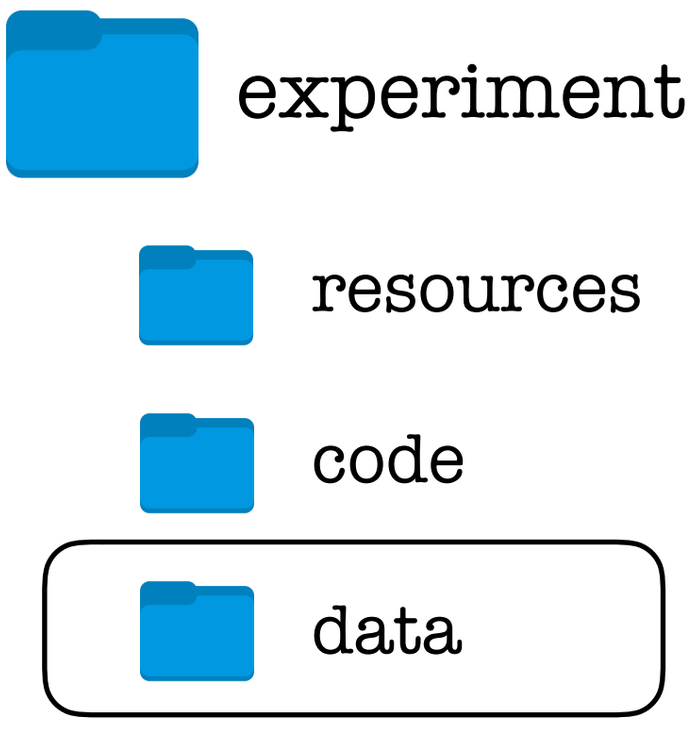


Code for actually implementing the experiment (e.g. Javascript, qualtrics, etc). If it's not on a computer then you do not need this

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)



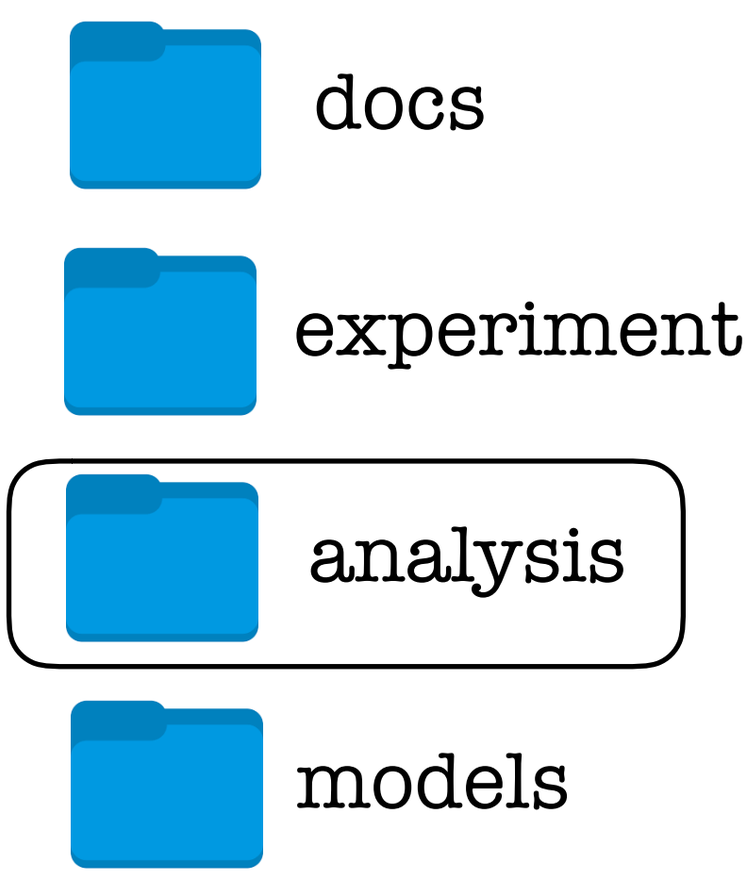
Raw data in a easy to read file (e.g., csv) along with a description of the file (definition of variables, etc)

It is very important to keep the raw data separate from any analyses, so you can always go back to it if you need

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)



Everything associated with analysing your experimental data except for the raw data

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

analysis

resources

code

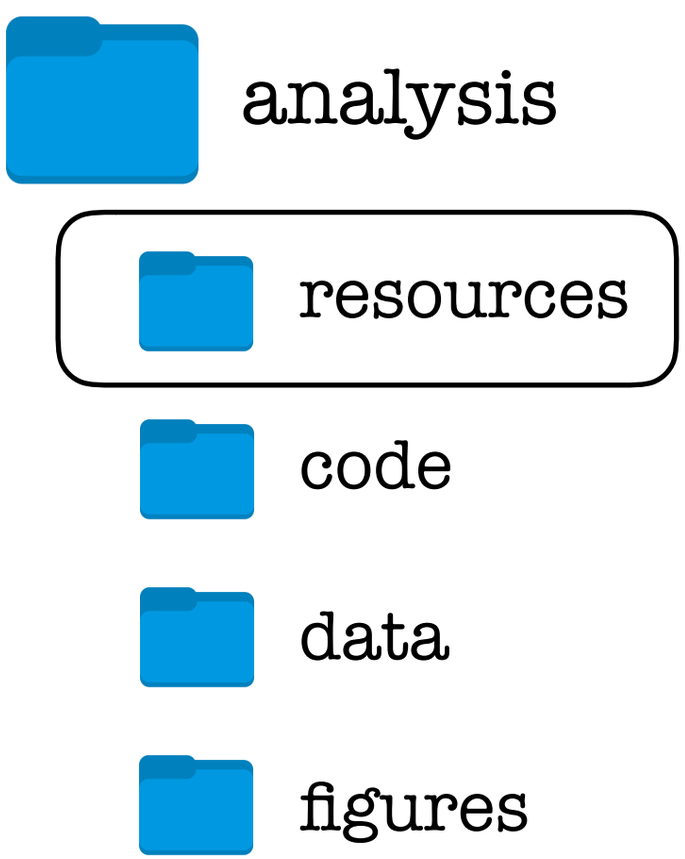
data

figures

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)



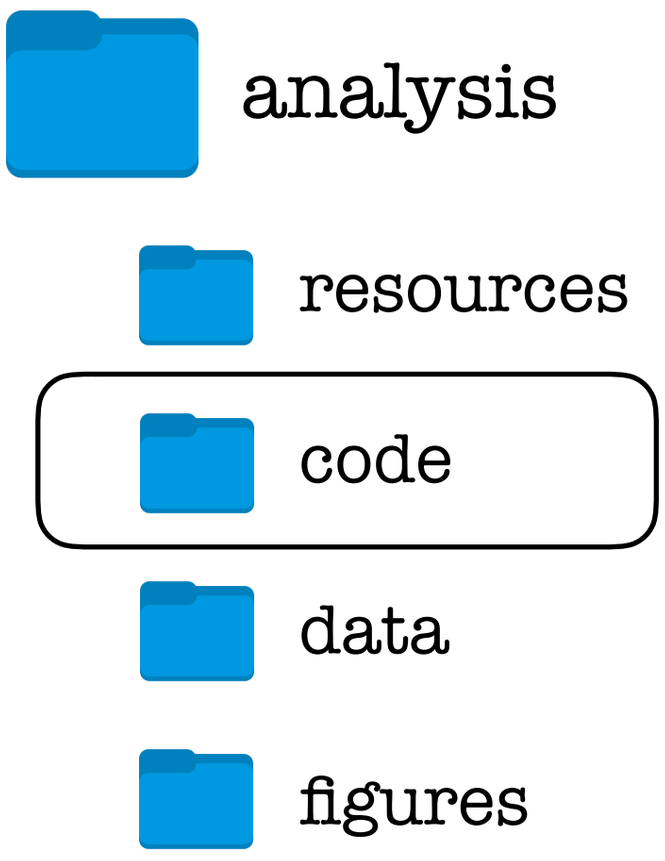
Documents explaining analysis choices, etc

(you may not have this)

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

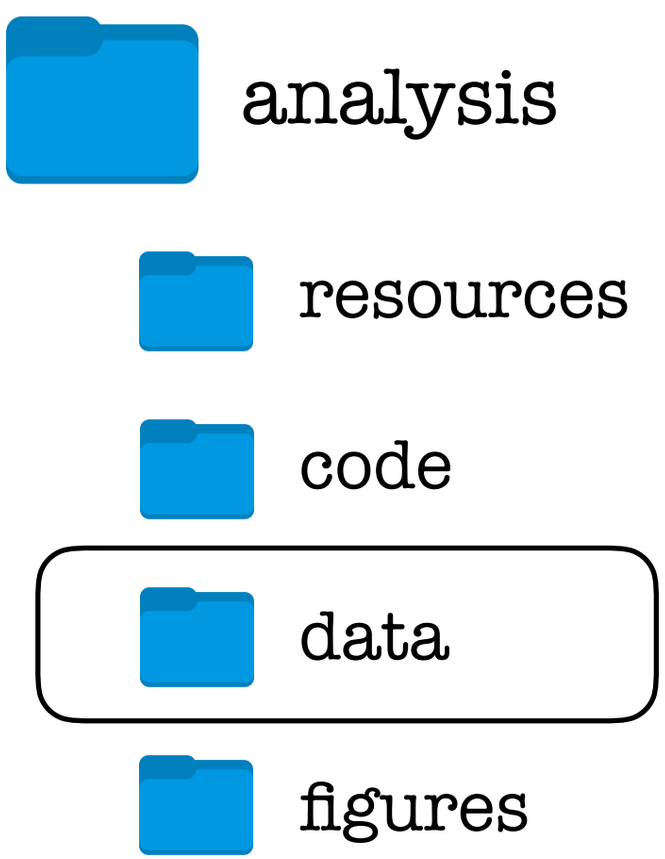


All of the analysis code — Dani will cover this tomorrow.

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

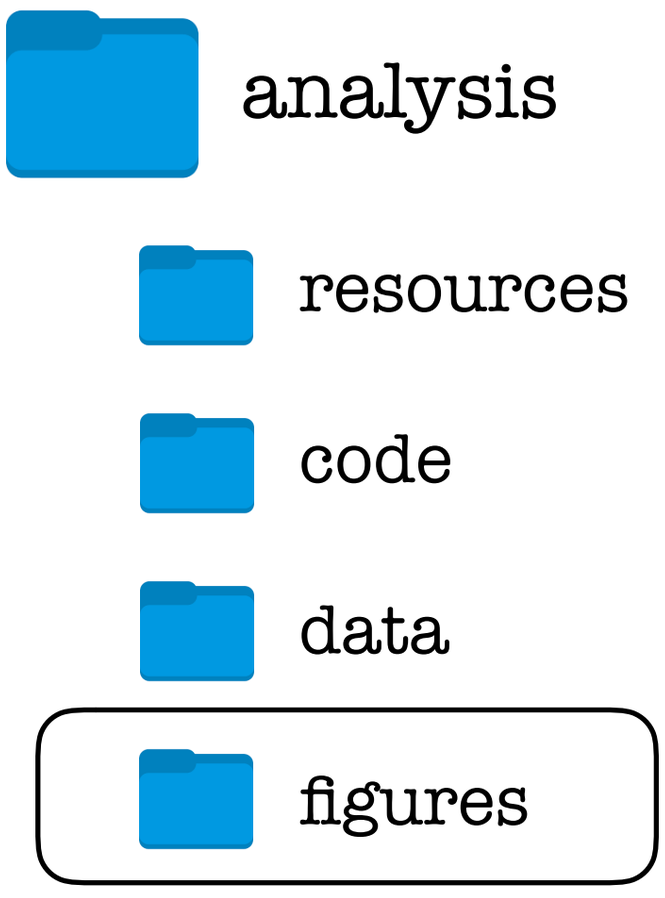


Output datafiles from your analysis code (e.g., subsets, cleaned files, etc)

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

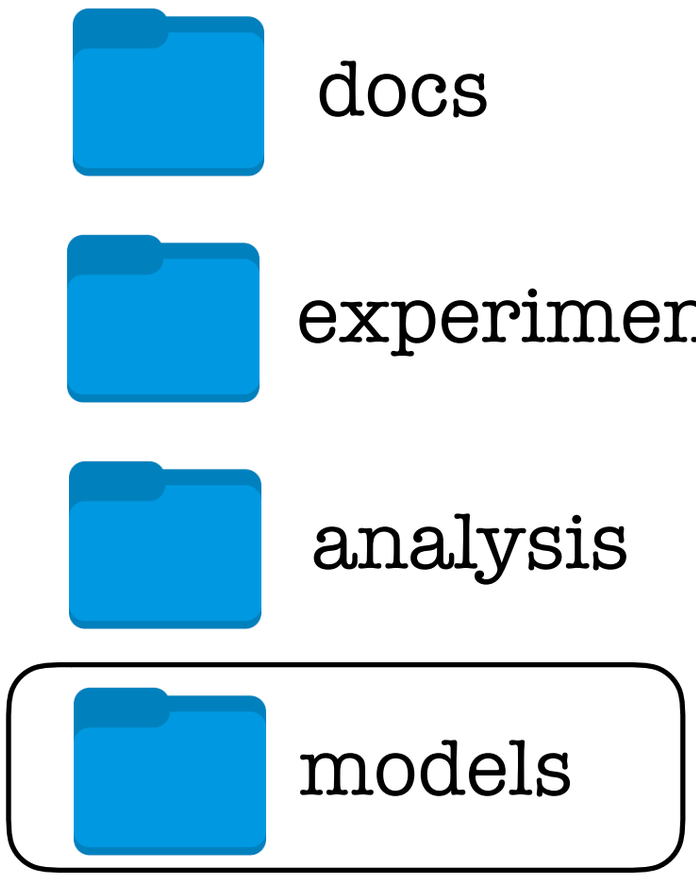


Figures generated from doing the data analysis

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)



Everything associated with any of your computational models (*not* for analysing the data — for theorising about it).

Charles will talk about this on Day 3

SUGGESTION

Every single project should have the following internal directory structure

(no spaces, ever)

models

resources

code

data

figures

Same basic breakdown, however

DAY 1: EXPERIMENT

Not just coding it up, but all workflow stuff up to running it

1. Background: replicability and proper procedure
2. Workflow and organisation
3. Experiment design
4. Coding experiment
5. Ethics and pre-registration
6. Hosting experiment on a server
7. Downloading data

DESIGN ISSUES

Split into groups of two or three. 10-15 minutes:

1. What's your typical process for finding an interesting question and figuring out how to address it? (or, if you haven't done it, thoughts about ideal process)

   www.menti.com, code **14 46 10**

2. What aspects of this process seem the most difficult to you, or you don't know how to approach?

   www.menti.com, code **33 86 39**

A SAMPLE EXPERIMENT

The scientific problem: how do people generalise from individual category examples?

These are edible…

A SAMPLE EXPERIMENT

The scientific problem: how do people generalise from individual category examples?

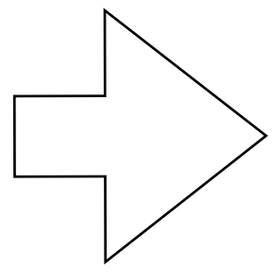
Can I eat this…?

A SAMPLE EXPERIMENT

This is well studied, often in a framework called a category induction task

Premise: EAGLES have more than one fovea per eye

Conclusion: HAWKS have more than one fovea per eye



Osherson et al., 1990; Medin et al., 2003

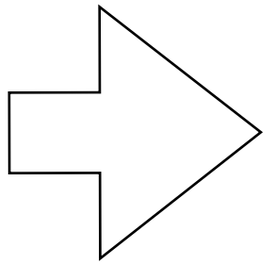
GENERALISING FROM A FEW EXAMPLES

Premise monotonicity: Adding premises to an argument typically strengthens it

EAGLES have more than one fovea per eye

FALCONS have more than one fovea per eye

HAWKS have more than one fovea per eye



GENERALISING FROM A FEW EXAMPLES

Premise monotonicity: Adding premises to an argument typically strengthens it

more likely that hawks have multiple fovaea

GENERALISING FROM A FEW EXAMPLES

Premise non-monotonicity: Occurs, but more rarely
(when adding premises to an argument weakens it)

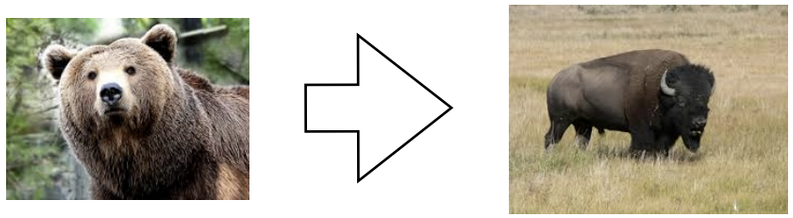
GENERALISING FROM A FEW EXAMPLES

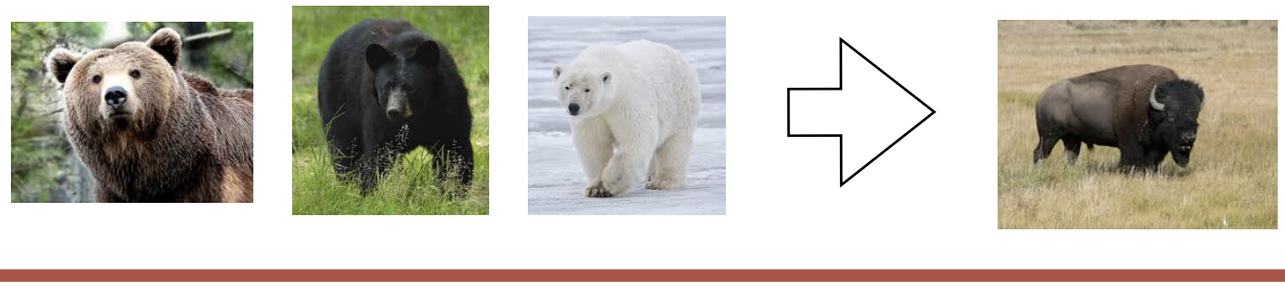
Premise non-monotonicity: Occurs, but more rarely (when adding premises to an argument weakens it)

less likely for buffalo to have the property

GENERALISING FROM A FEW EXAMPLES

Premise non-monotonicity





Explained with the relevance theory of induction: adding premises should weaken an argument if the added categories reinforce a property shared by all of the premises but not the conclusion

Wilson & Sperber, 2004

GENERALISING FROM A FEW EXAMPLES

Seems sensible, but why? If nothing can be assumed about how the premises are sampled (which is what most models of category-based induction implicitly assume) then this reasoning is “irrational” (i.e., not statistically optimal)

Explained with the relevance theory of induction: adding premises should weaken an argument if the added categories reinforce a property shared by all of the premises but not the conclusion

GENERALISING FROM A FEW EXAMPLES

Seems sensible, but why? If nothing can be assumed about how the premises are sampled (which is what most models of category-based induction implicitly assume) then this reasoning is “irrational” (i.e., not statistically optimal)

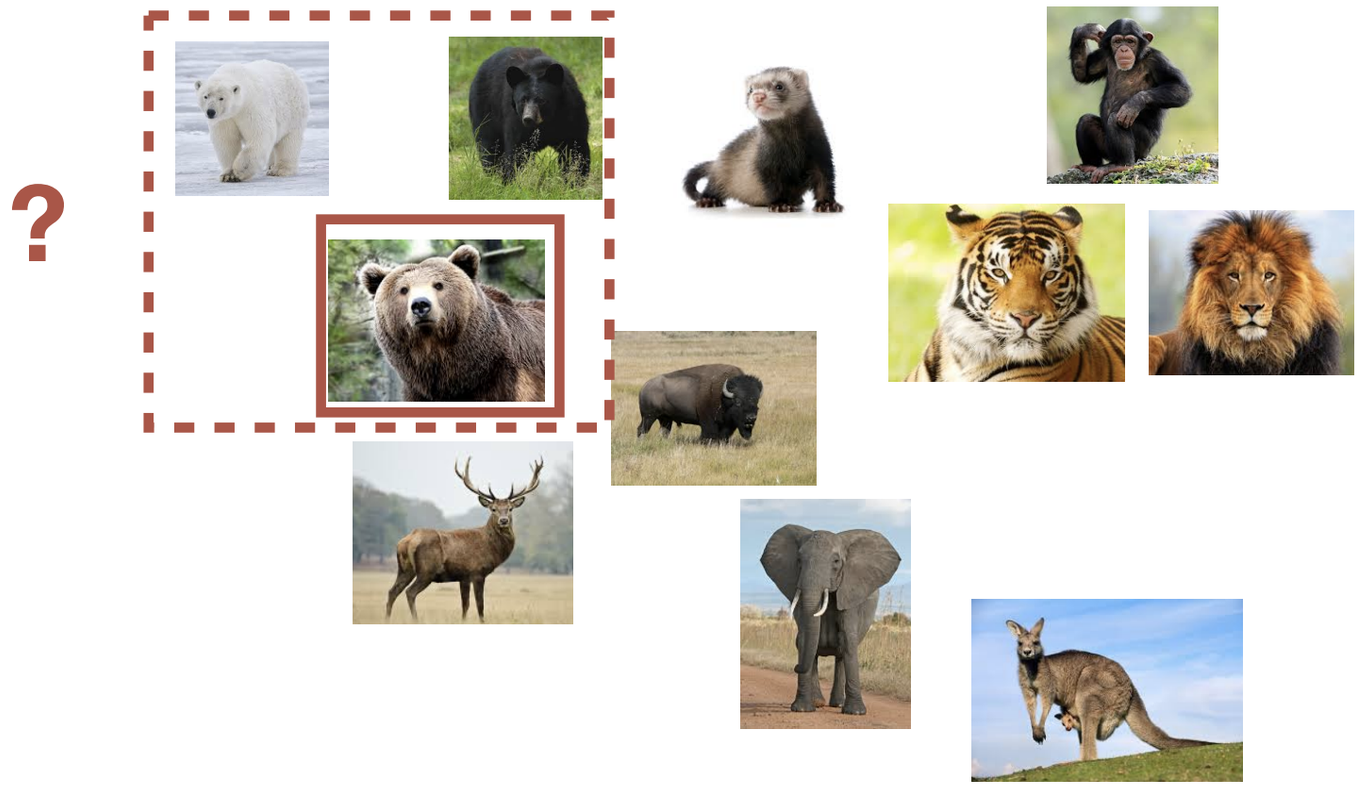
In the real world, arguments are constructed for social purposes. Accounting for this in our statistical assumptions can explain non-monotonicity

A MODEL OF CATEGORY-BASED INDUCTION

The world consists of a set of things which may or may not have some property *P*

related to Heit, 1998; Sanjana & Tenenbaum, 2003

A MODEL OF CATEGORY-BASED INDUCTION



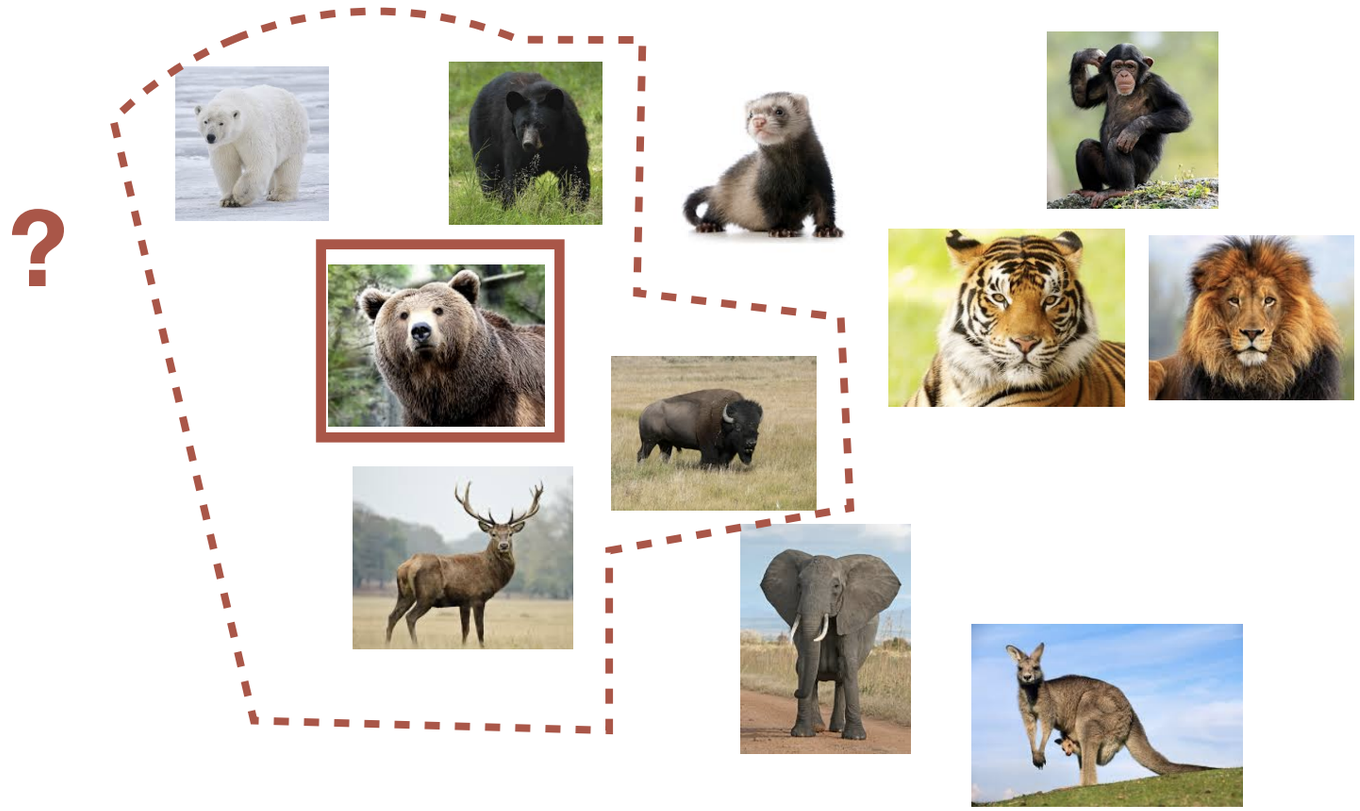
Each hypothesis *h* captures how far a property should be extended

A MODEL OF CATEGORY-BASED INDUCTION

?

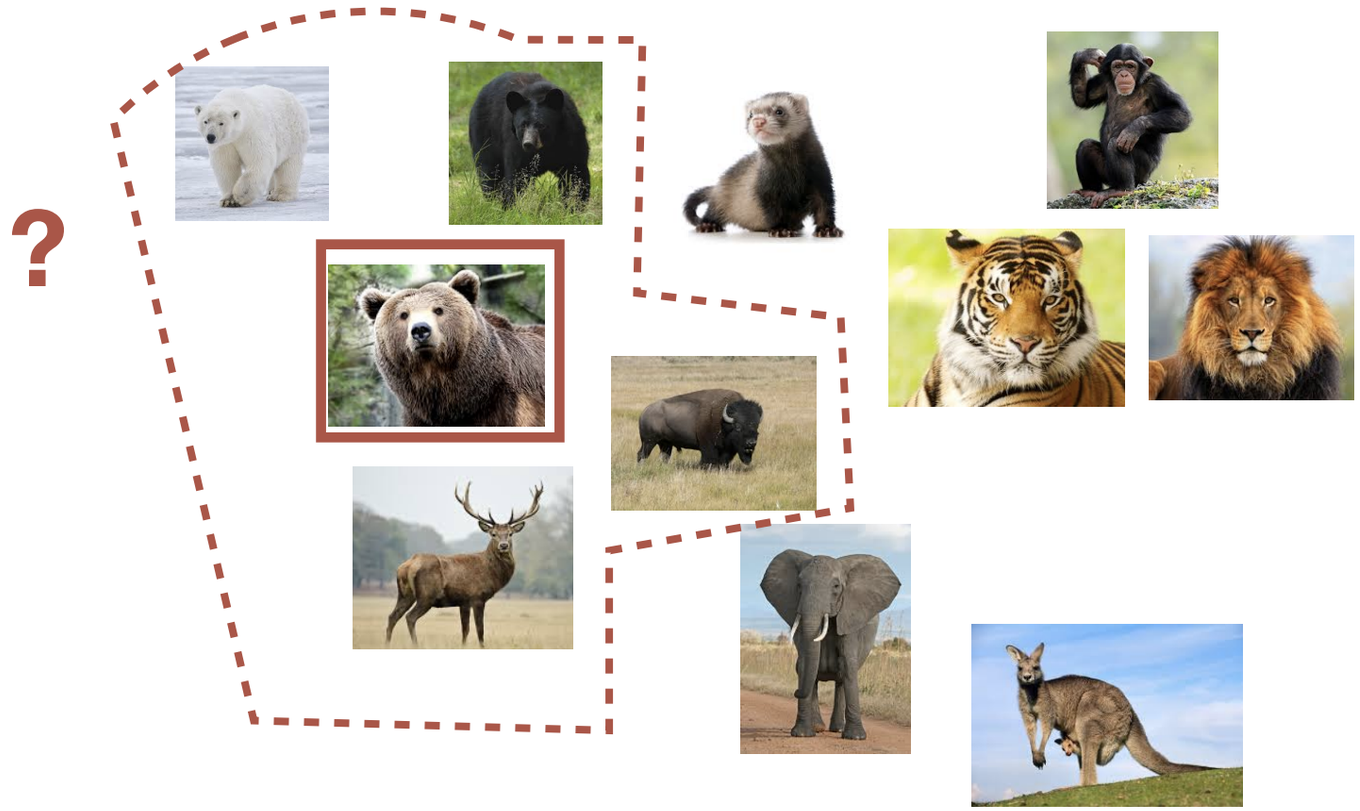
Each hypothesis *h* captures how far a property should be extended

A MODEL OF CATEGORY-BASED INDUCTION



Each hypothesis *h* captures how far a property should be extended

A MODEL OF CATEGORY-BASED INDUCTION



Belief in *h* after having seen data *x* is given by Bayes' Rule

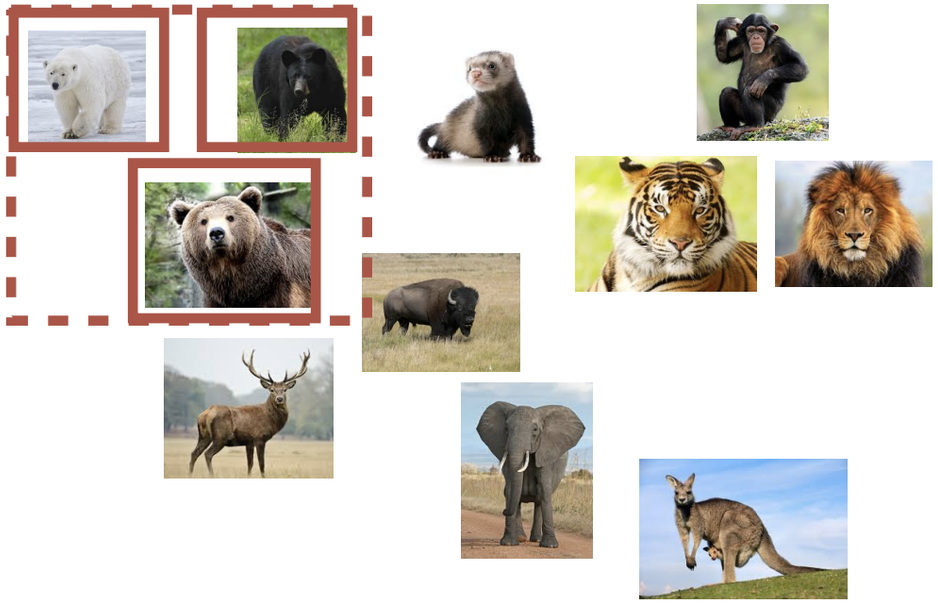
$$P(h \mid x) = \frac{P(x \mid h)P(h)}{\sum_{h'} P(x \mid h')P(h')}.$$

… but how far to generalise depends on the assumptions about how the data were generated

A MODEL OF CATEGORY-BASED INDUCTION

Strong sampling: Picking instances from the concept (having *P*), as one would in order to communicate about it

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



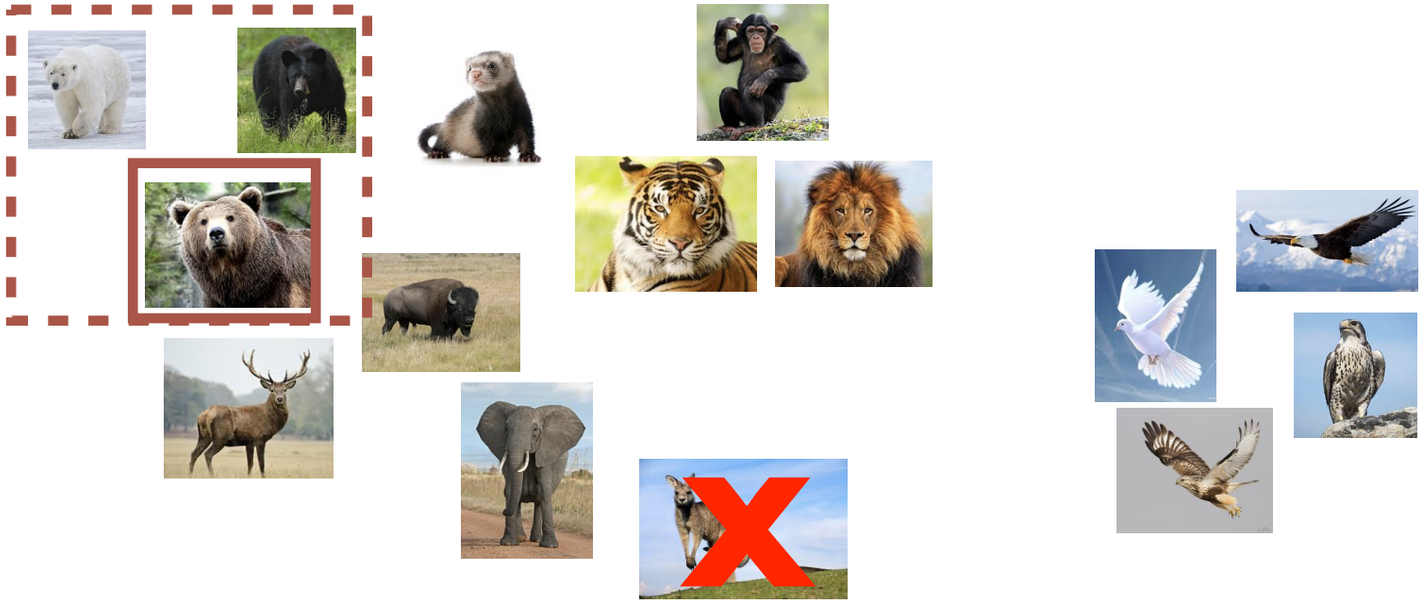
Licenses non-monotonic reasoning: otherwise, poor communication

… but how far to generalise depends on the assumptions about how the data were generated

A MODEL OF CATEGORY-BASED INDUCTION

Weak sampling: Picking instances from the world at random, and then labeling them as having property *P* or not

$$P(x \mid h) \propto \begin{cases} 1 & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases}$$

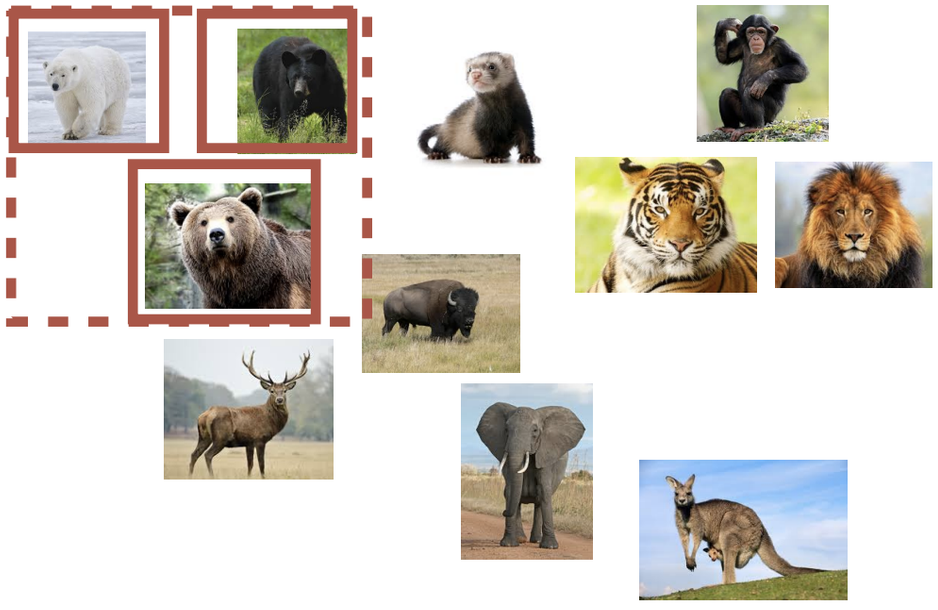


… but how far to generalise depends on the assumptions about how the data were generated

A MODEL OF CATEGORY-BASED INDUCTION

Weak sampling: Picking instances from the world at random, and then labeling them as having property *P* or not

$$P(x \mid h) \propto \begin{cases} 1 & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases}$$

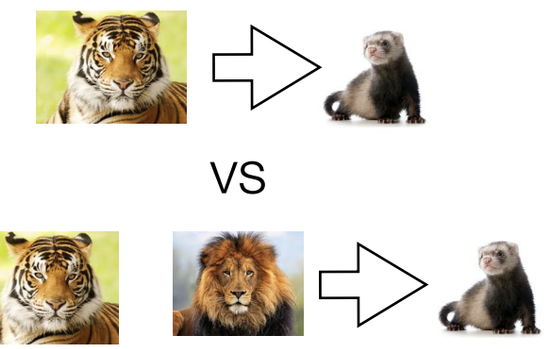


Does not license non-monotonic reasoning: just happened to be that way (i.e., the selection of items is not meaningful)

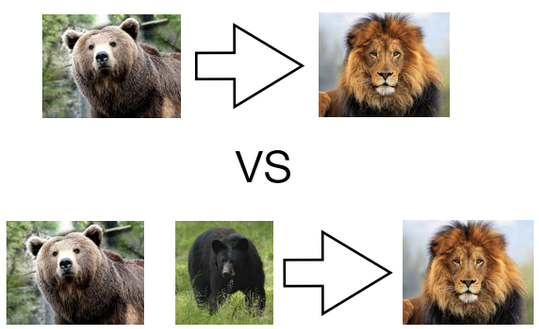
… but how far to generalise depends on the assumptions about how the data were generated

DIFFERENT SAMPLING ASSUMPTIONS YIELD DIFFERENT PREDICTIONS

TARGET 1

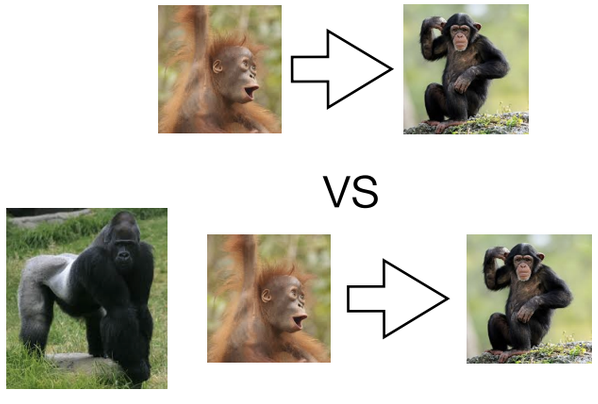


TARGET 2



Non-monotonic: Additional argument should make conclusion weaker (if strong sampling, not if weak)

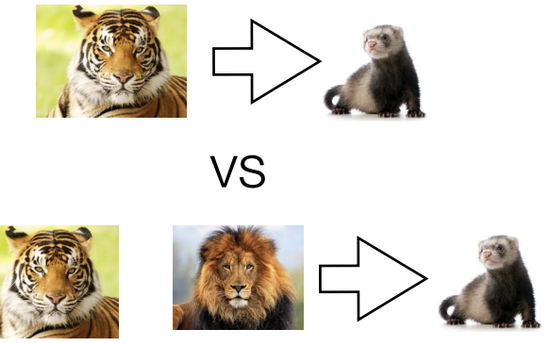
CONTROL



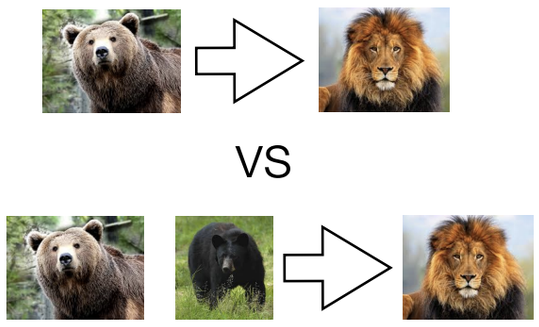
Monotonic: Additional argument should make conclusion stronger (if strong sampling, not if weak)

DIFFERENT SAMPLING ASSUMPTIONS YIELD DIFFERENT PREDICTIONS

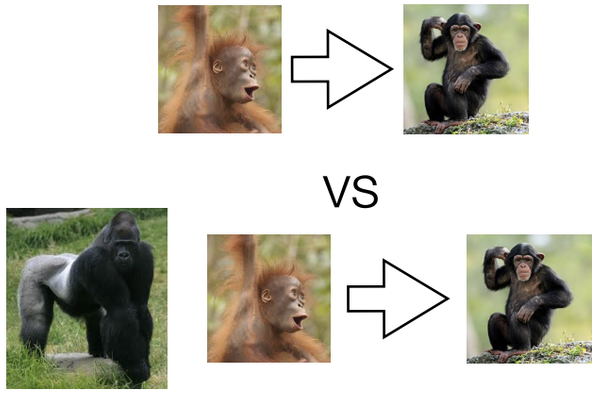
TARGET 1

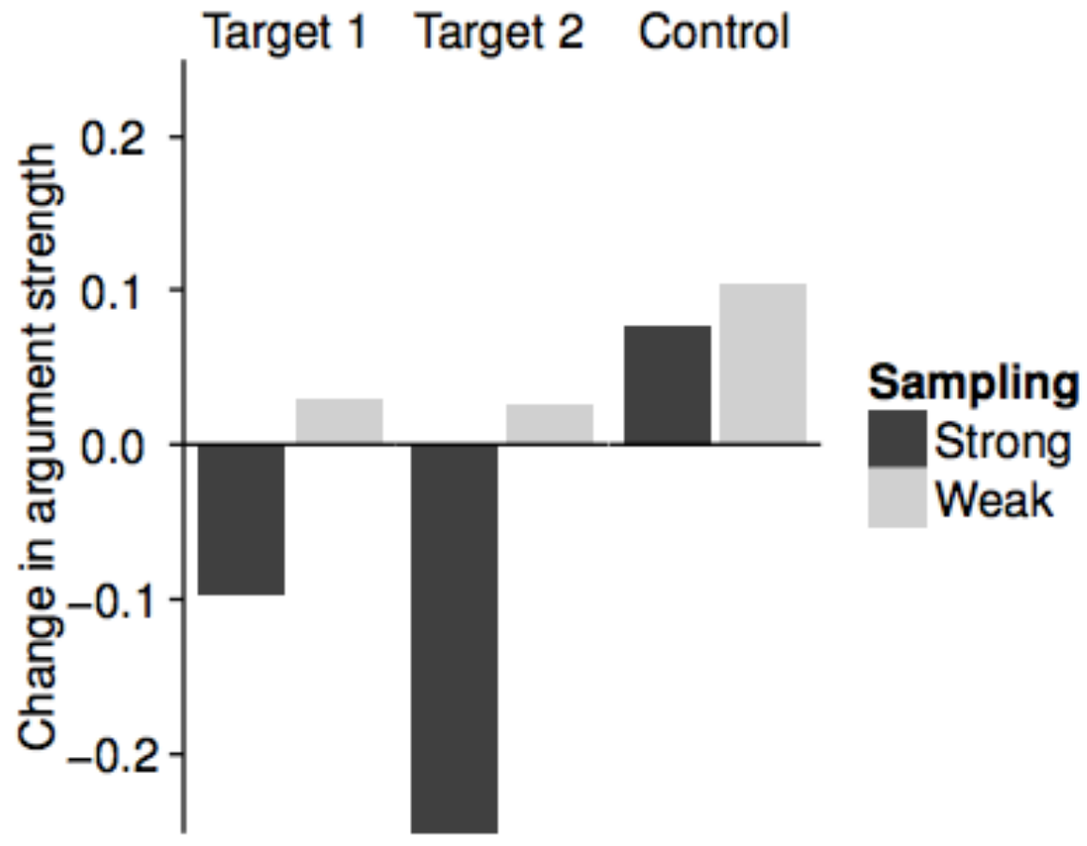


TARGET 2



CONTROL





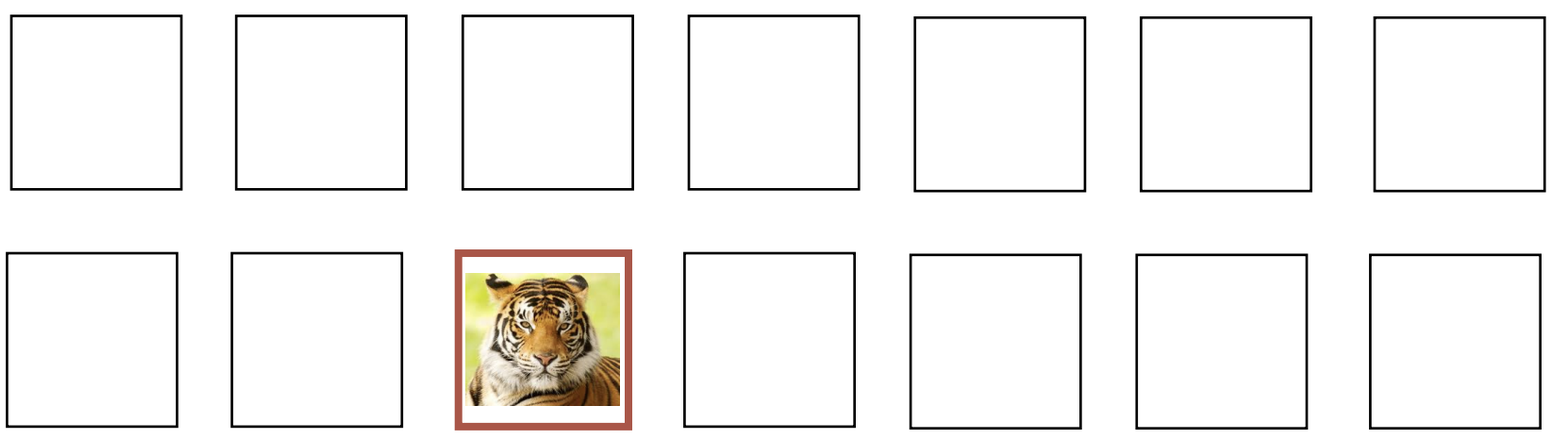
DIFFERENT SAMPLING ASSUMPTIONS YIELD DIFFERENT PREDICTIONS

Do people change their pattern of reasoning based on mainpulating the cover story about how the data were generated (socially, or not)?

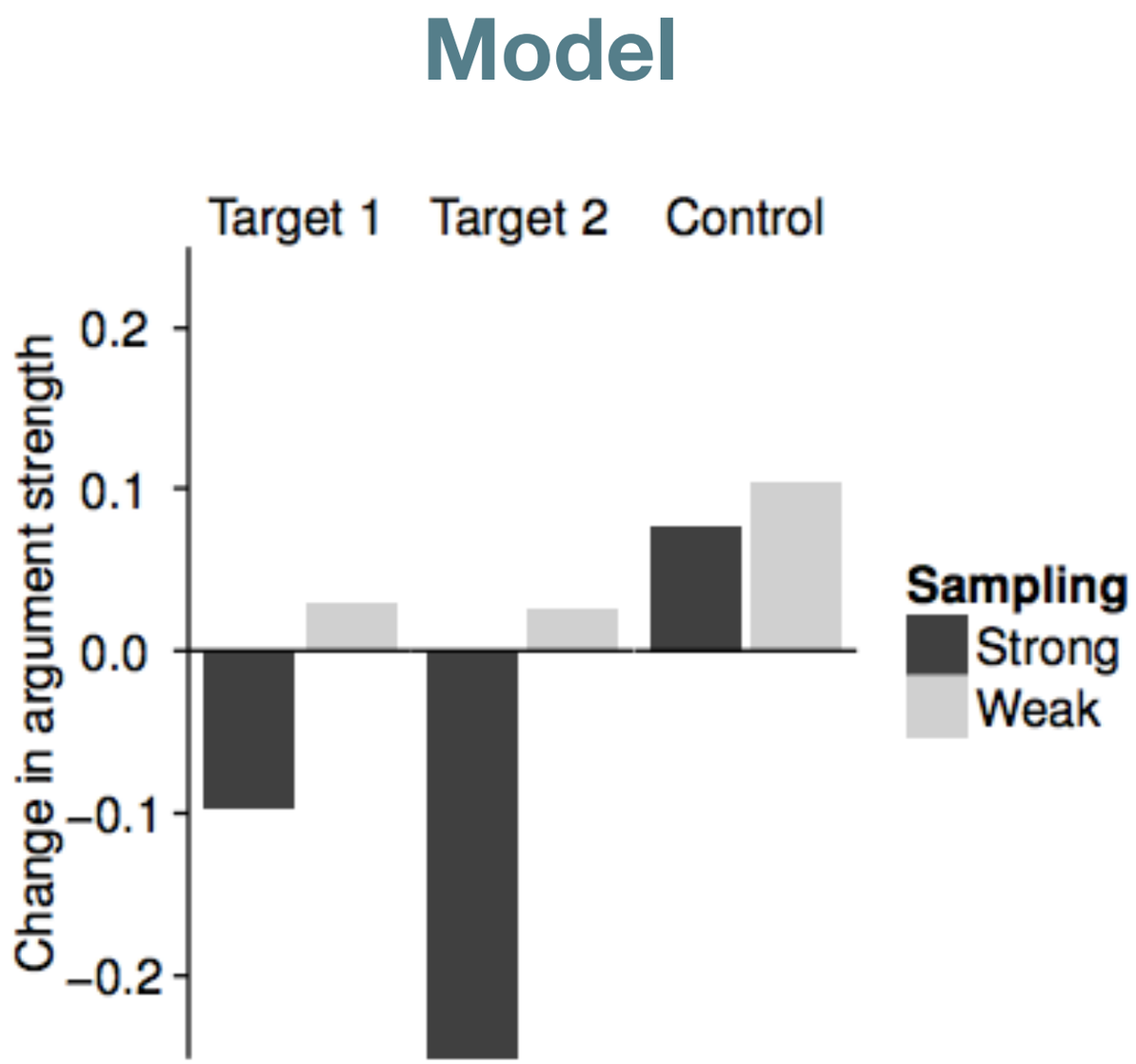
COVER STORY MANIPULATION

HELPFUL: People were told that the second fact in each trial was generated by a past player of the game who was trying to be helpful

RANDOM: People "drew" the second fact randomly from a set of cards drawn on the screen, one for each animal

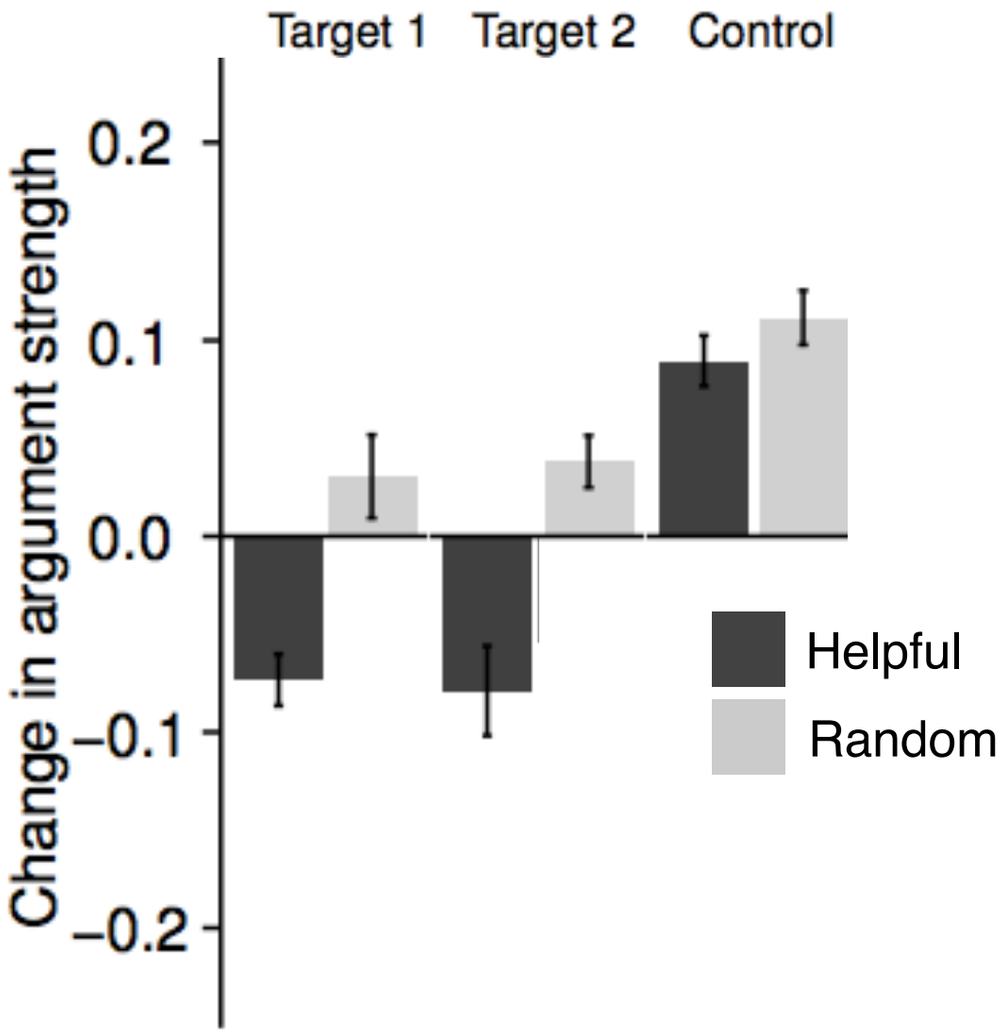


CHANGING THE SOCIAL STORY CHANGES THE PATTERN OF PEOPLE'S REASONING

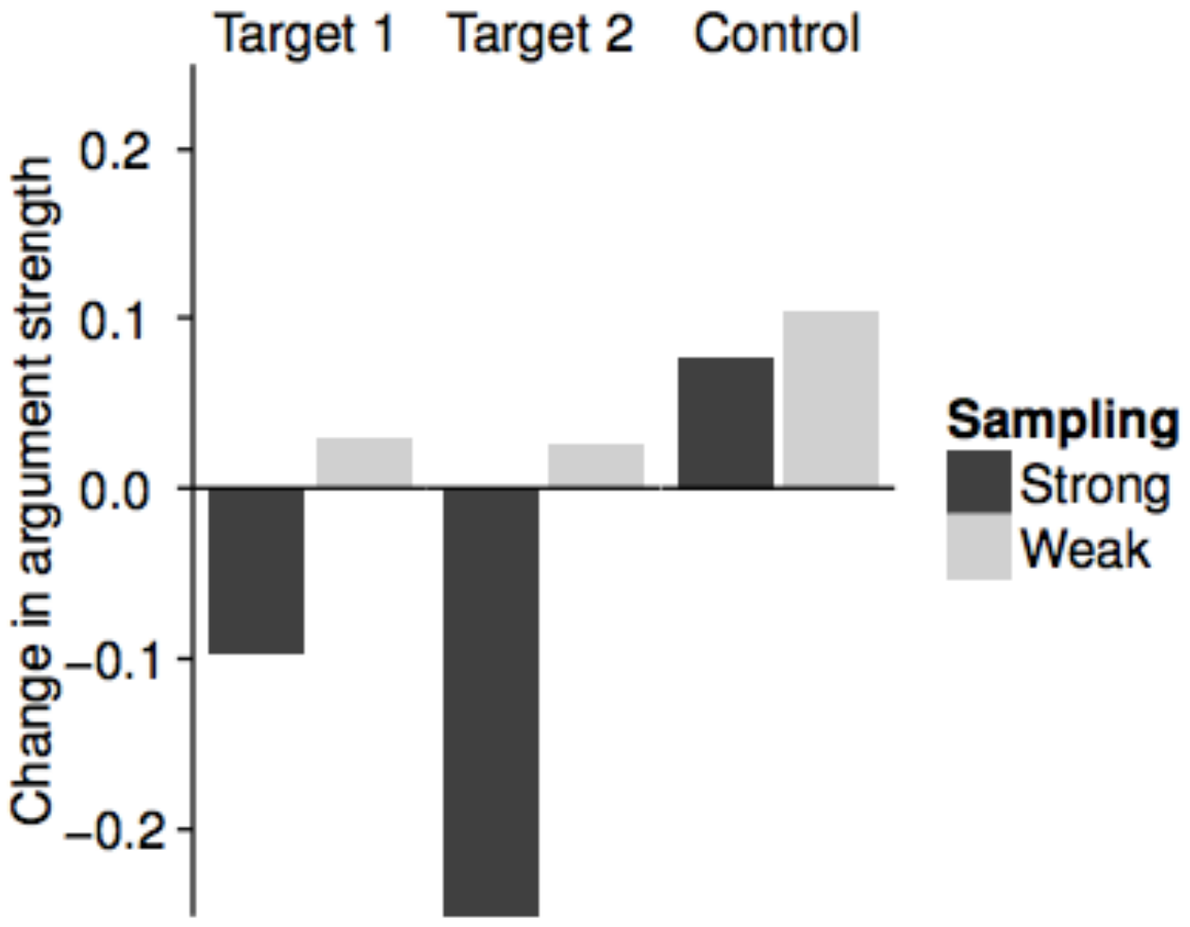


CHANGING THE SOCIAL STORY CHANGES THE PATTERN OF PEOPLE'S REASONING

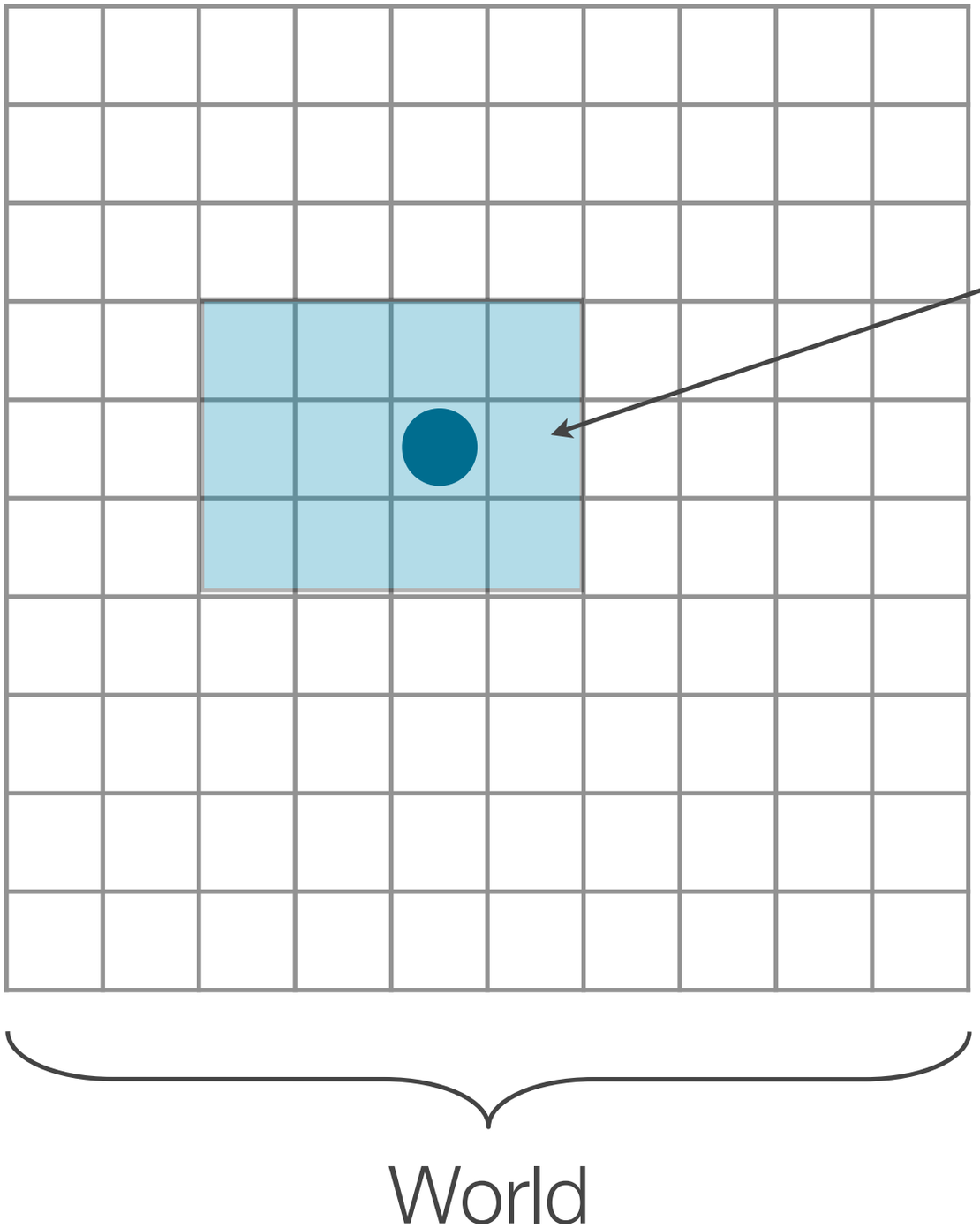
People



Model



SAMPLING ALSO AFFECTS HOW YOU SHOULD RESPOND TO ADDITIONAL DATAPOINTS



Strong sampling

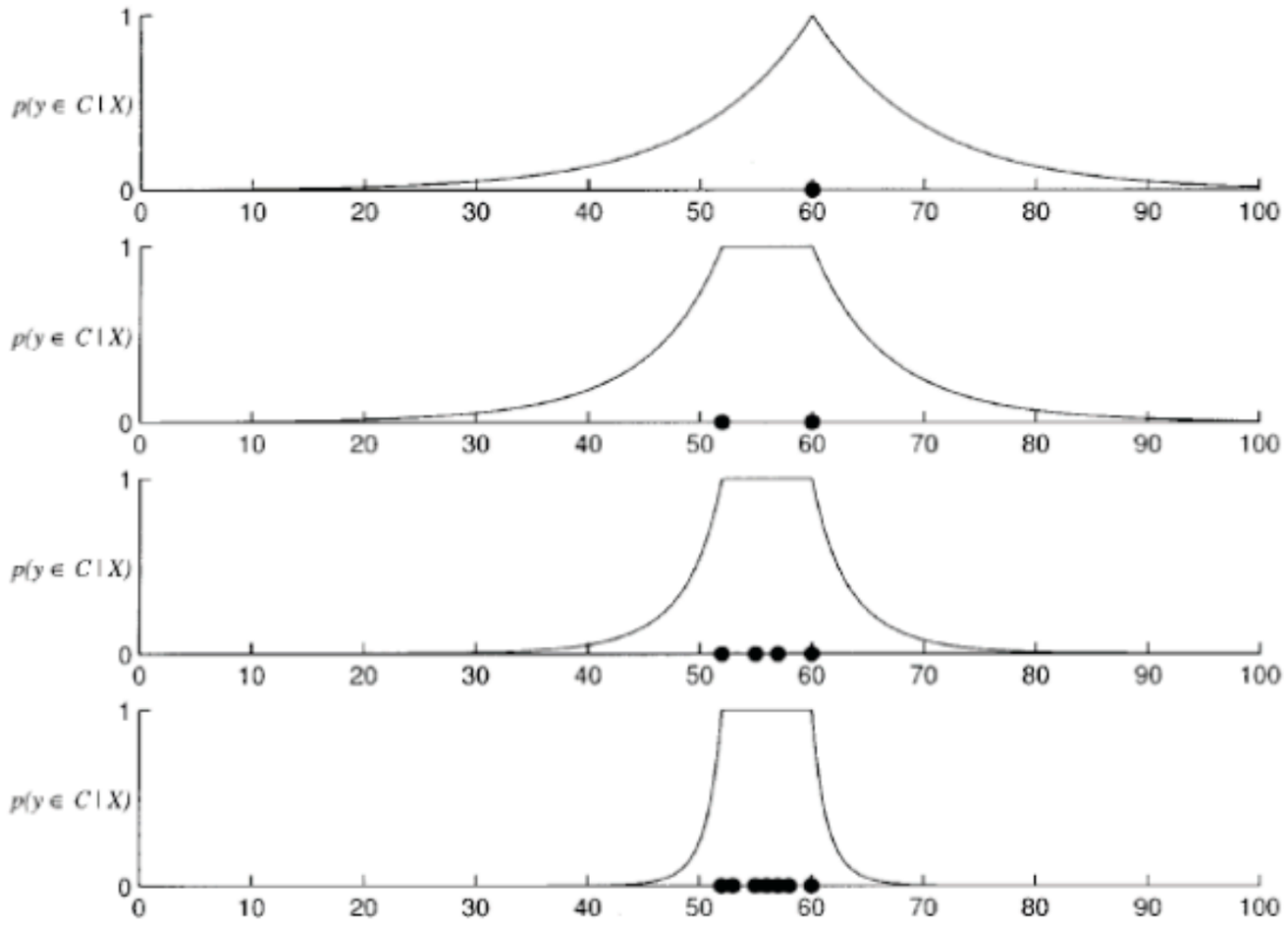
Hypothesis of size n

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

This is known as the size principle

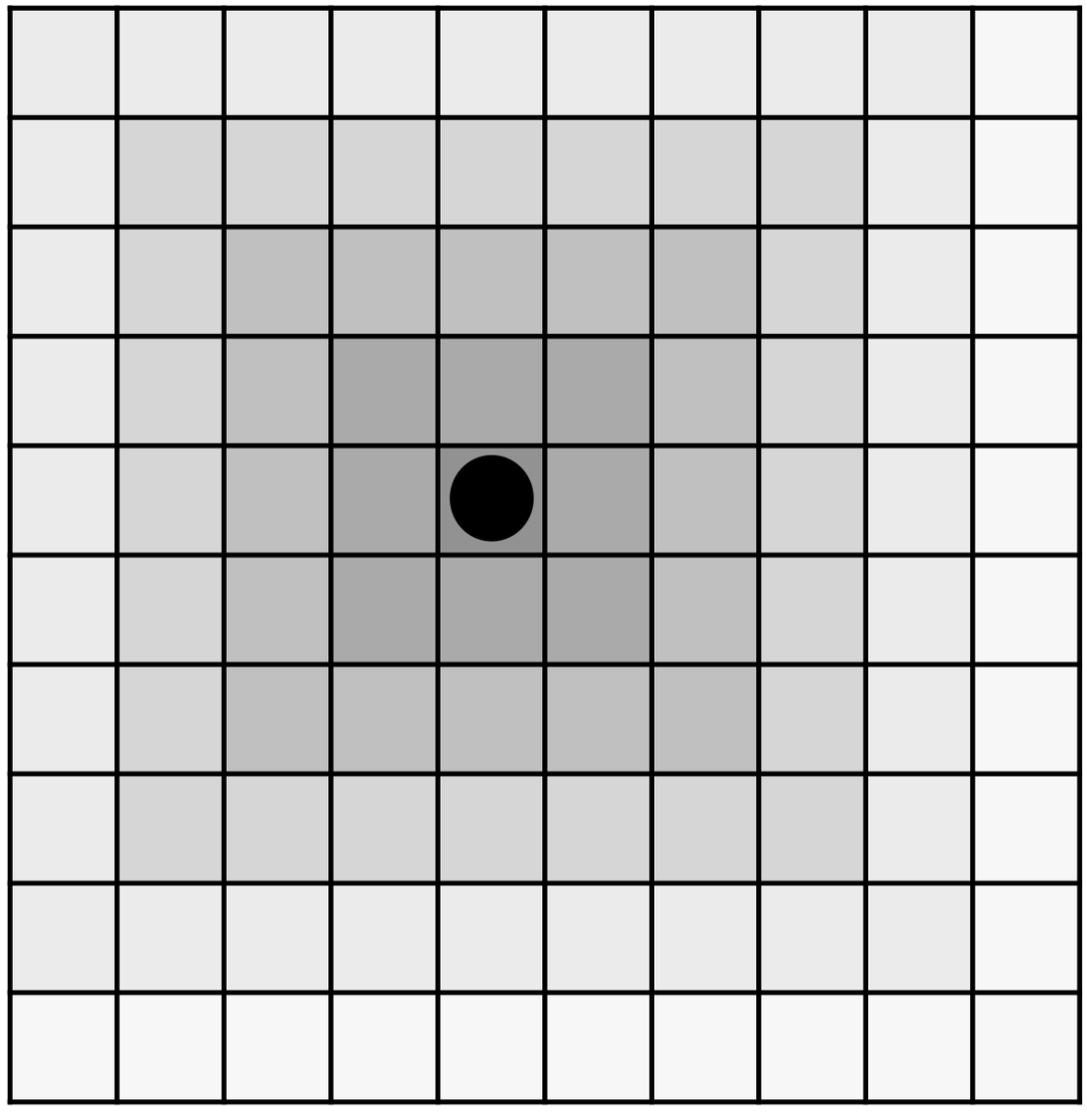
CONSEQUENCE OF SIZE PRINCIPLE

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



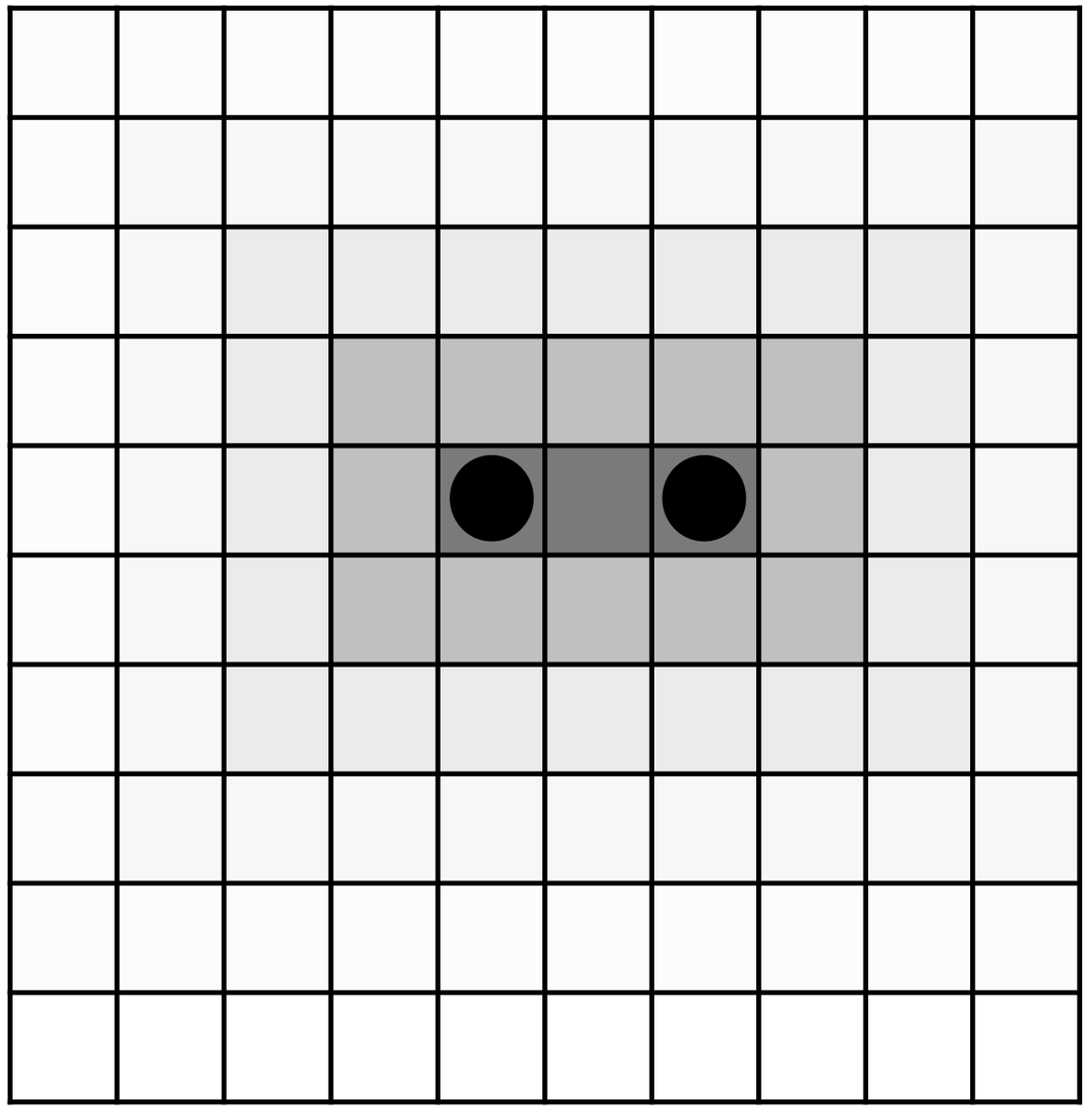
CONSEQUENCE OF SIZE PRINCIPLE

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



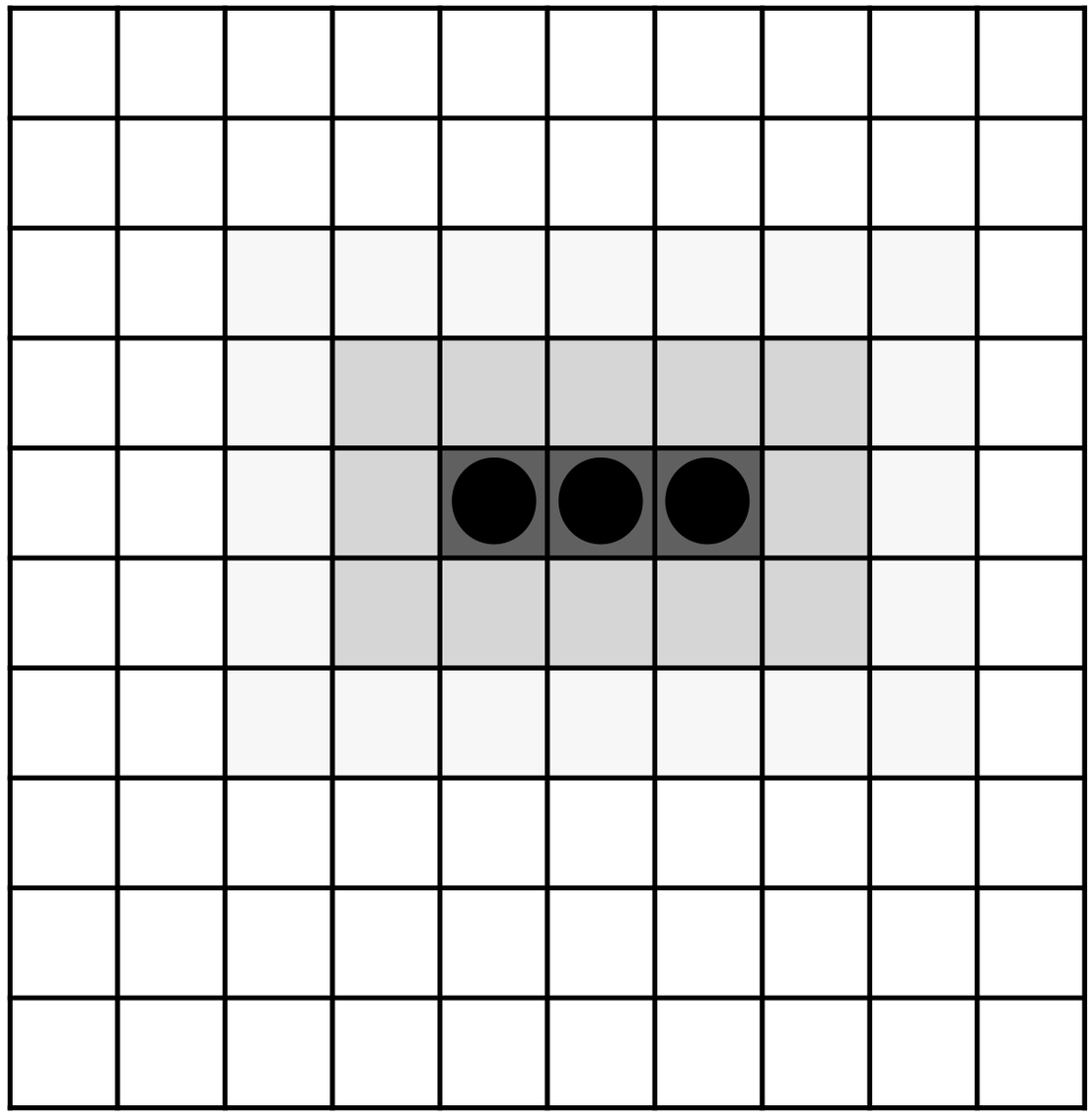
CONSEQUENCE OF SIZE PRINCIPLE

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



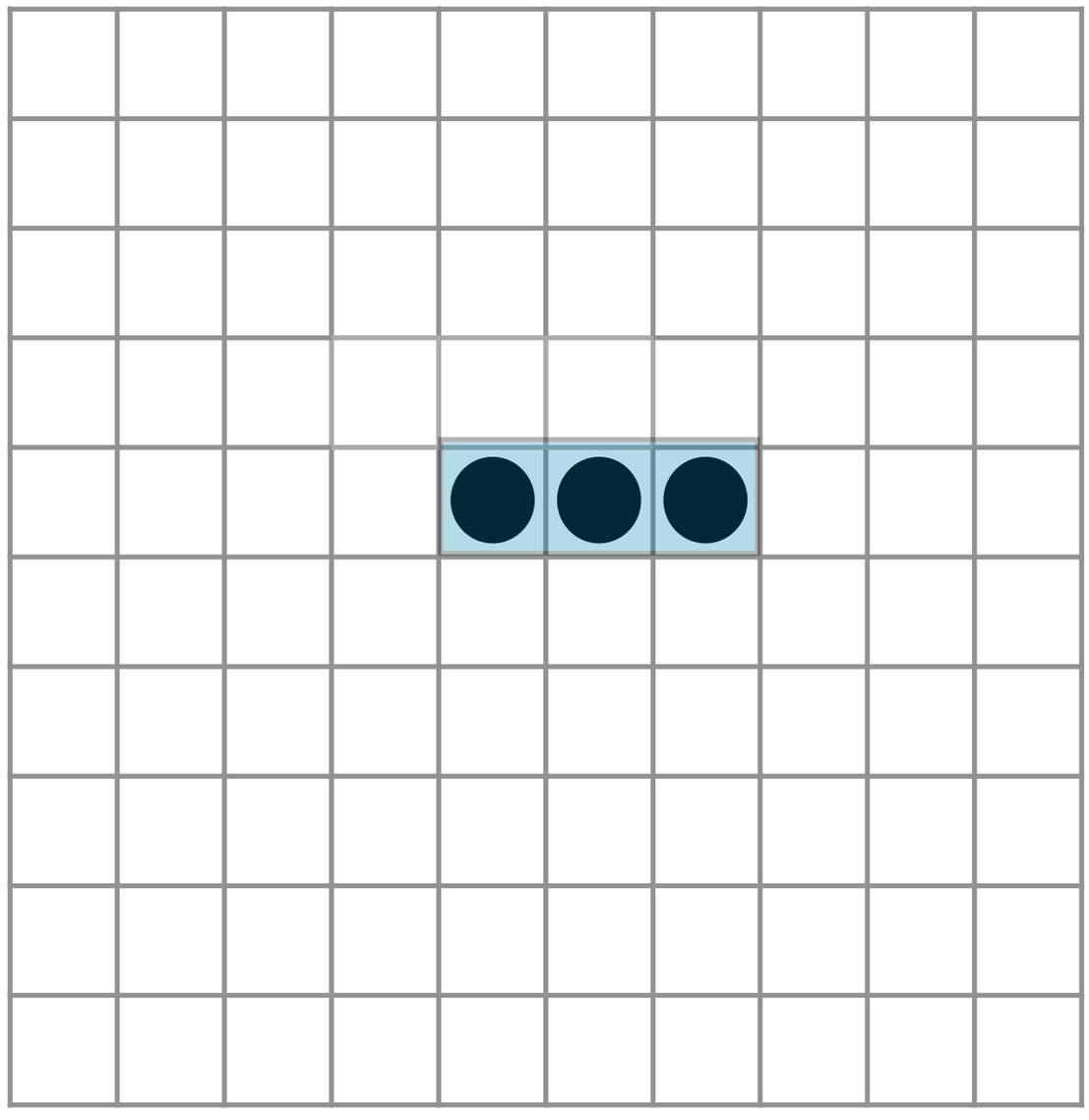
CONSEQUENCE OF SIZE PRINCIPLE

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



CONSEQUENCE OF SIZE PRINCIPLE

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



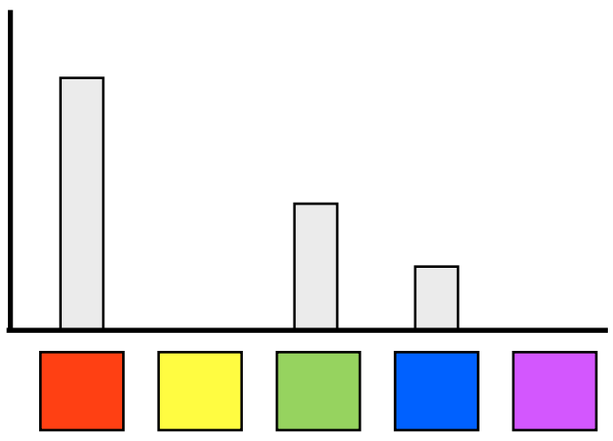
This is because it's quite a suspicious coincidence for these data points to have been generated if the true hypothesis is *not* h

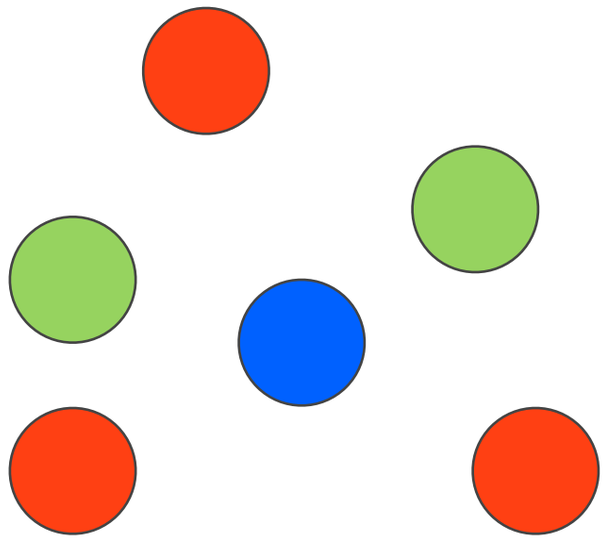
WEAK SAMPLING IS DIFFERENT!

WEAK SAMPLING IS DIFFERENT!

- The size principle follows from strong sampling assumptions about how data were generated

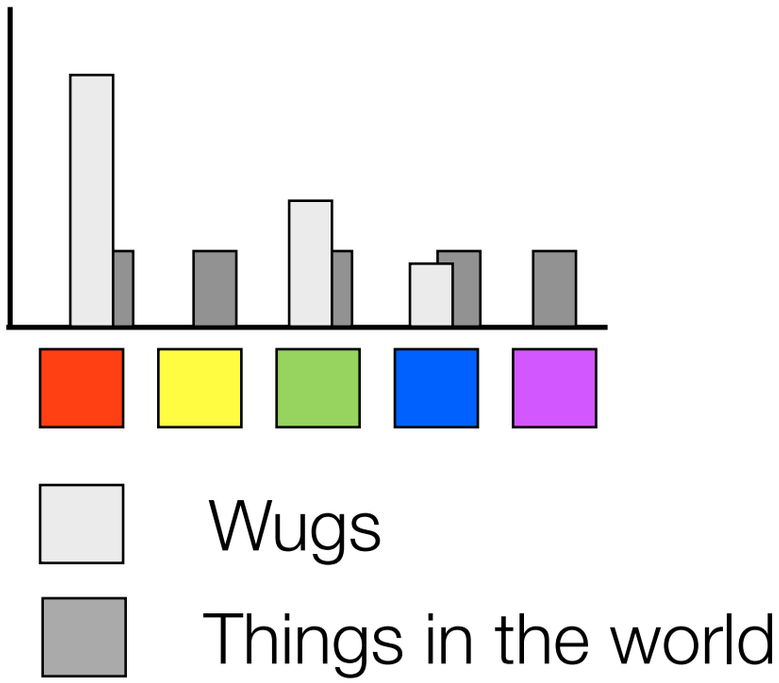
Each point drawn independently and at random from the hypothesis

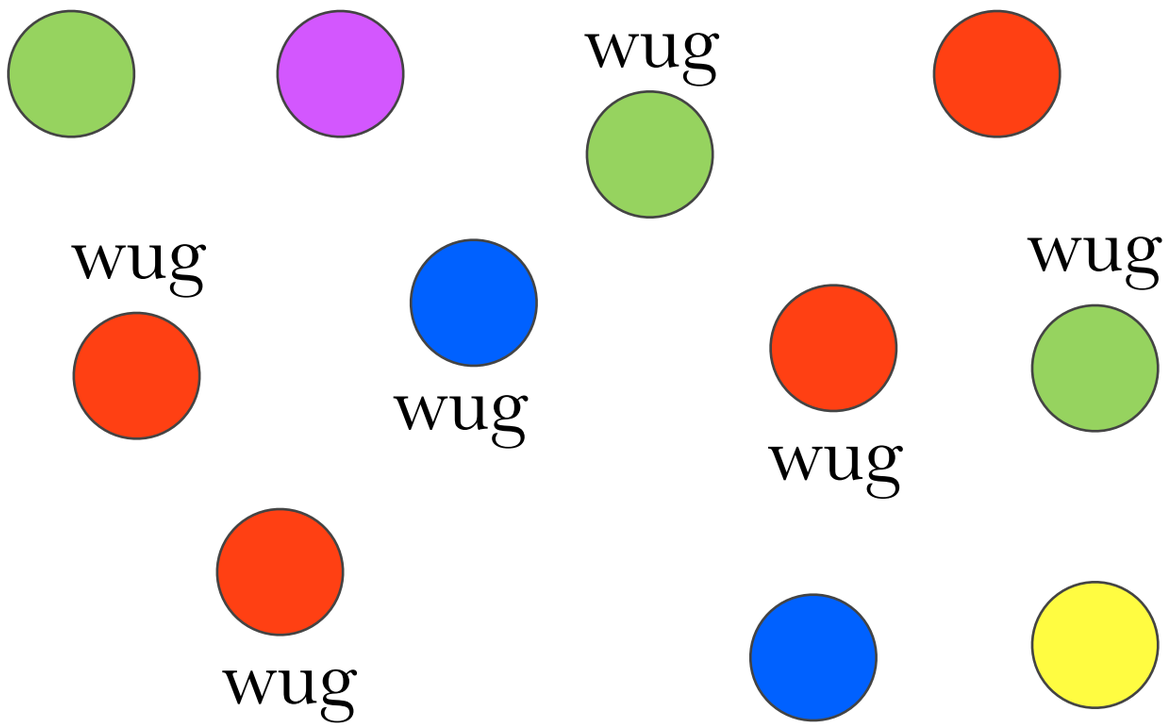




WEAK SAMPLING IS DIFFERENT!

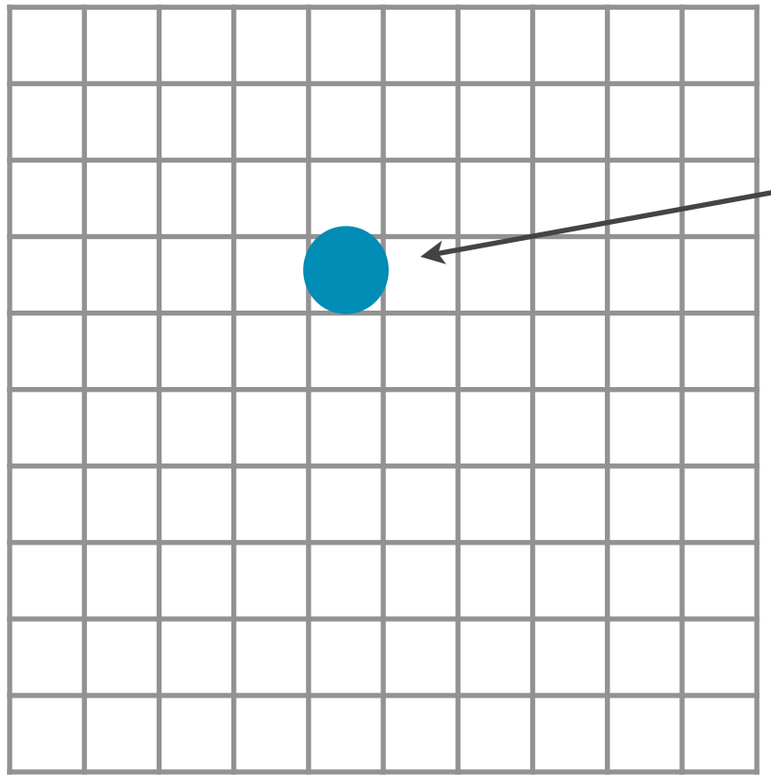
- Weak sampling suggests that data were generated from the world in general, and then only labelled as belonging to the hypothesis (or not)





WEAK SAMPLING IS DIFFERENT!

- Weak sampling suggests that data were generated from the world in general, and then only labelled as belonging to the hypothesis (or not)



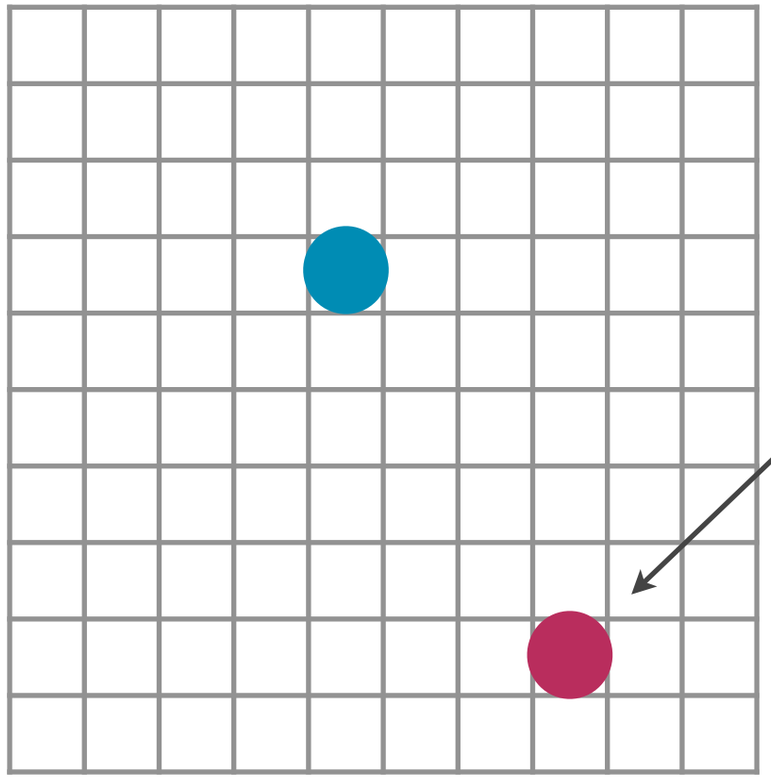
Data sampled from the world at random

Then labelled as in the hypothesis or not

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

WEAK SAMPLING IS DIFFERENT!

- Weak sampling suggests that data were generated from the world in general, and then only labelled as belonging to the hypothesis (or not)



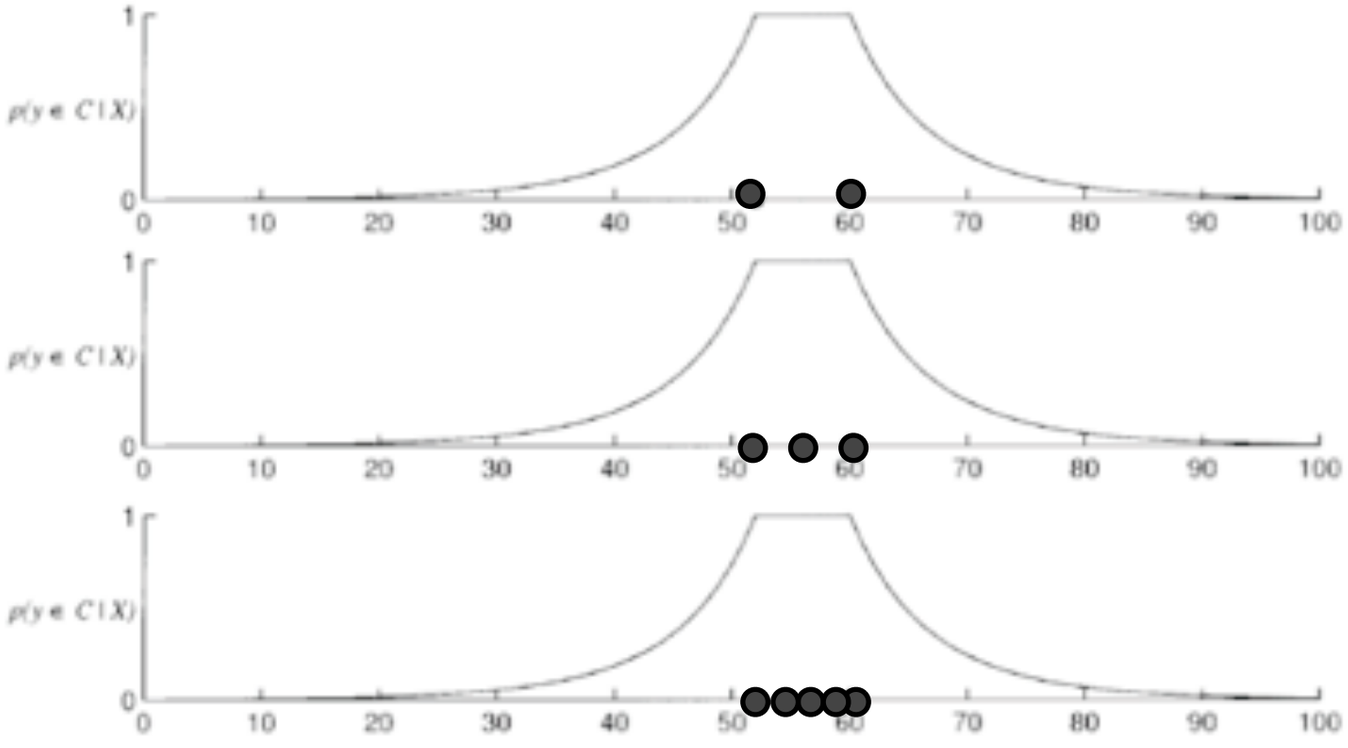
Data sampled from the world at random

Then labelled as in the hypothesis or not

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

WEAK SAMPLING IS DIFFERENT!

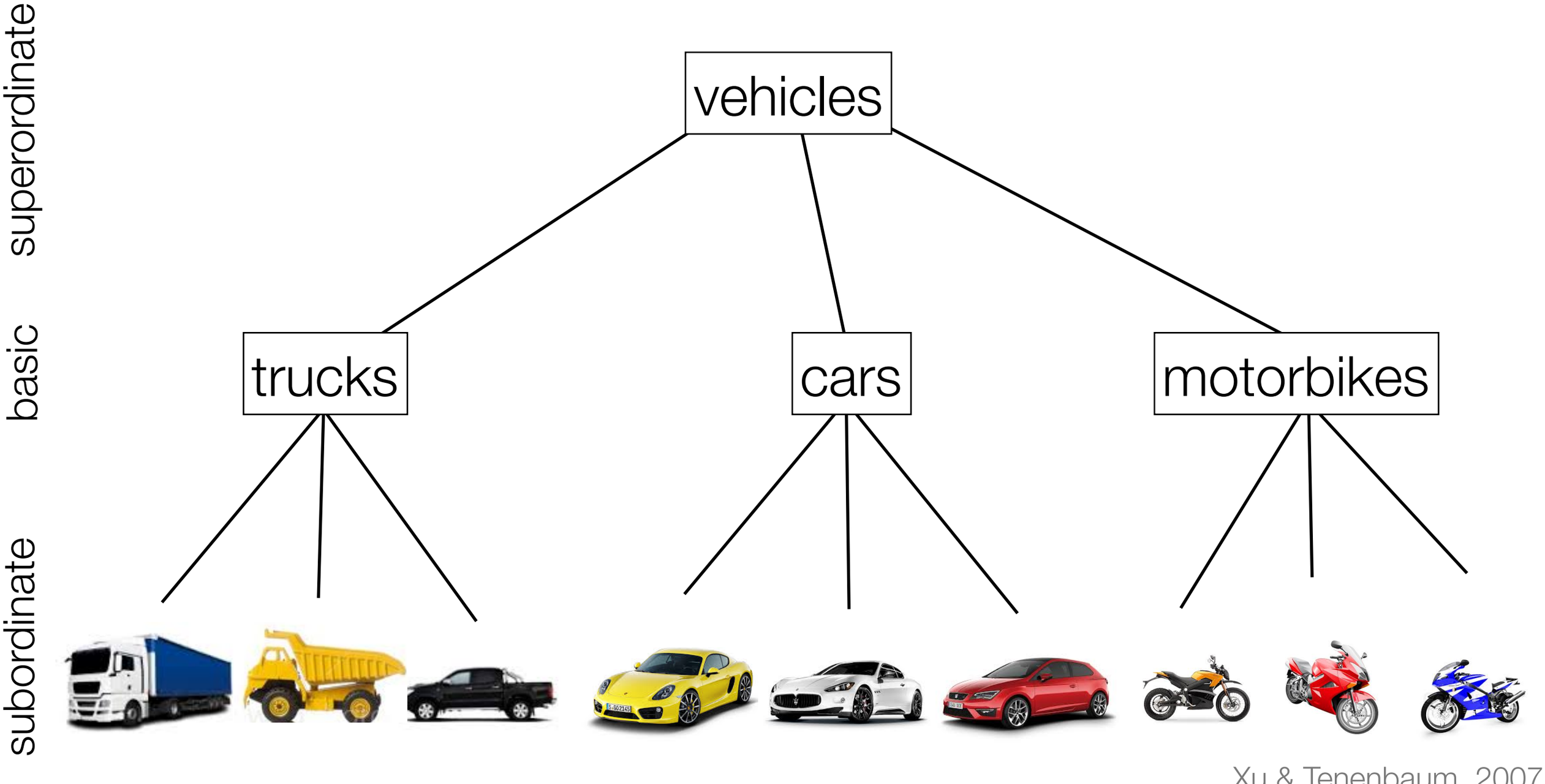
- If data are weakly sampled, the generalisation curves should not tighten -- there is no suspicious coincidence since the data were generated by the *world,* and not from the hypothesis



ARE PEOPLE SENSITIVE TO SAMPLING ASSUMPTIONS WHEN REASONING ABOUT ADDITIONAL DATA?

WORD LEARNING

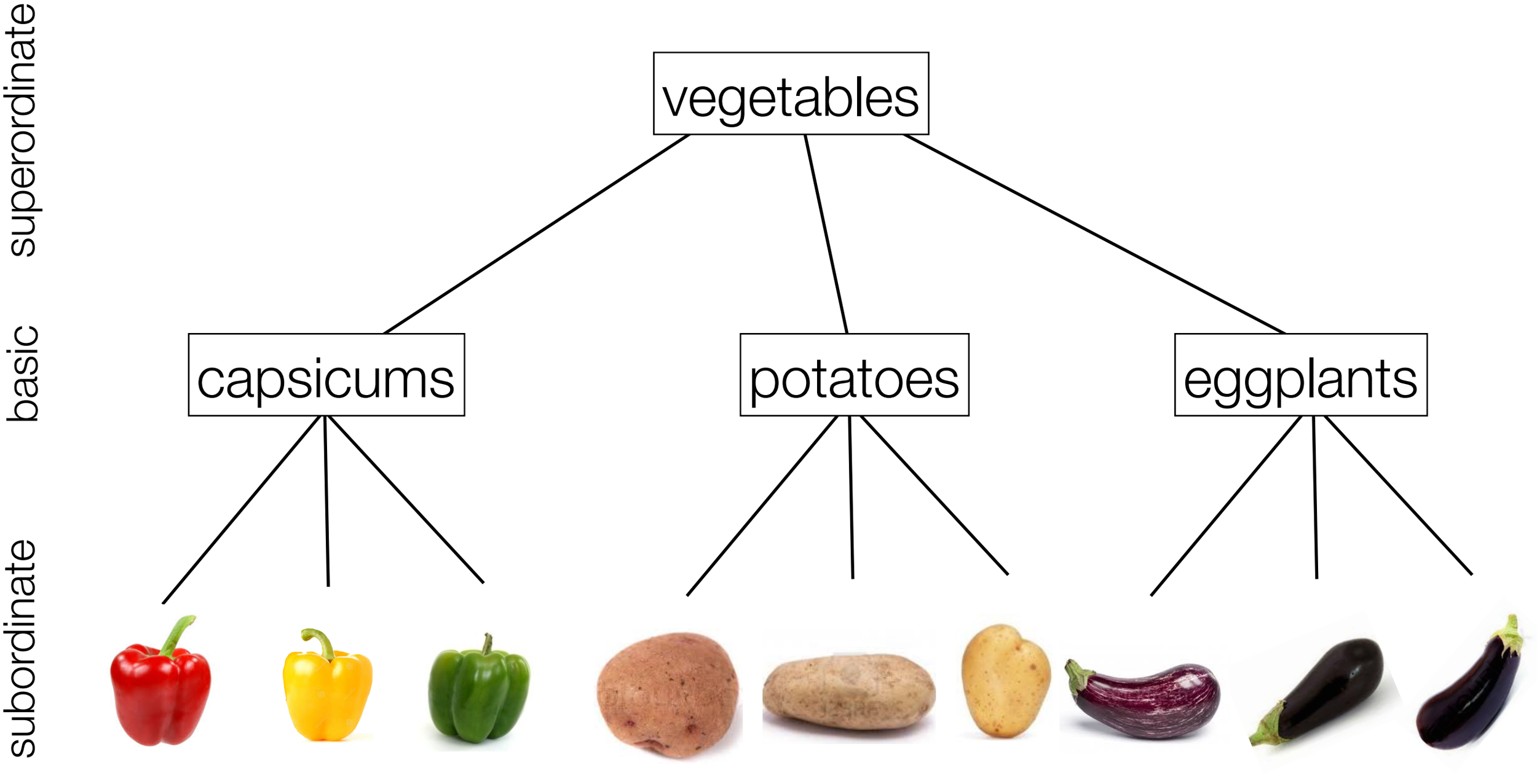
- Many domains have a hierarchical or tree-based conceptual structure



Xu & Tenenbaum, 2007

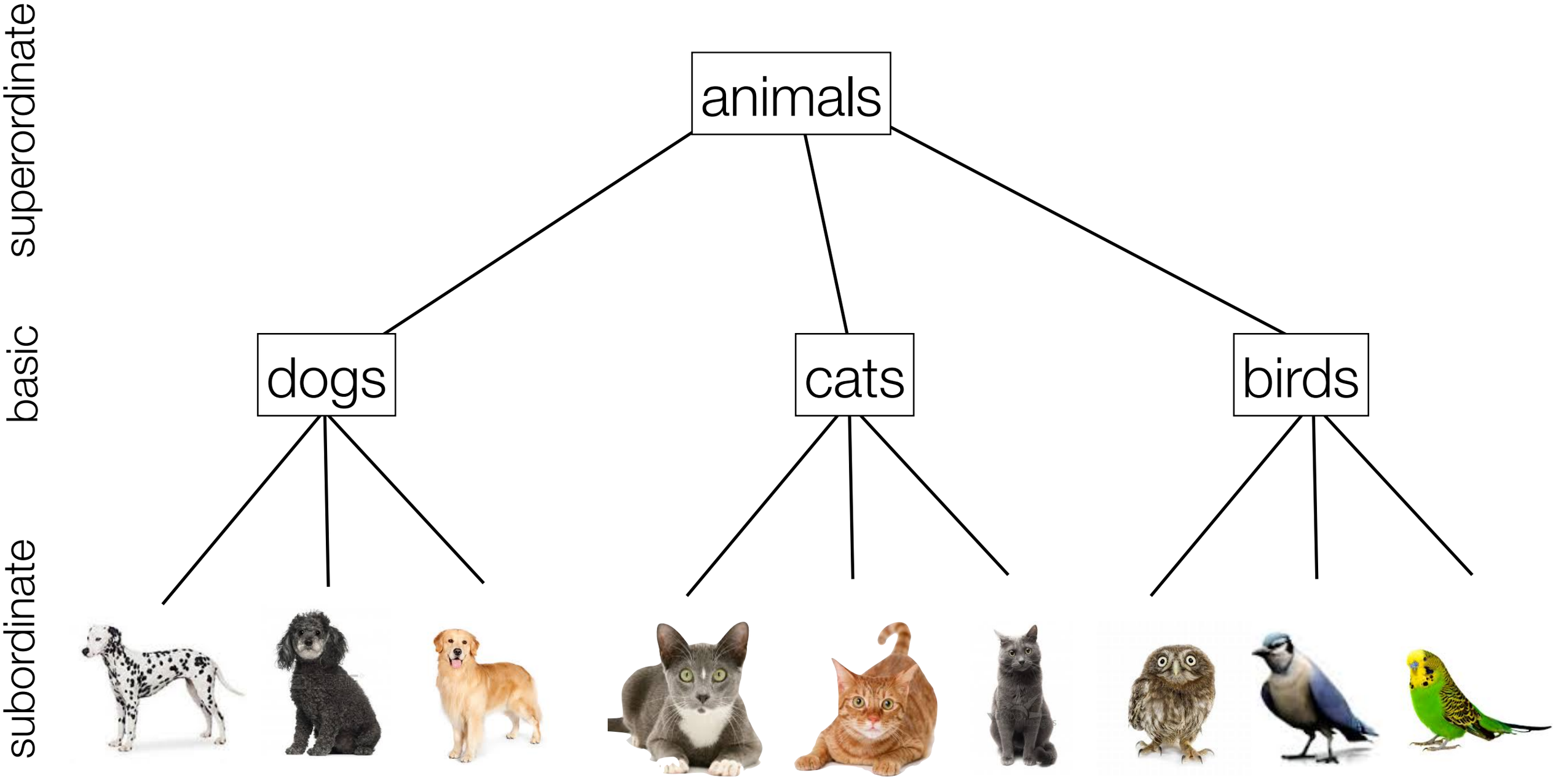
WORD LEARNING

- Many domains have a hierarchical or tree-based conceptual structure



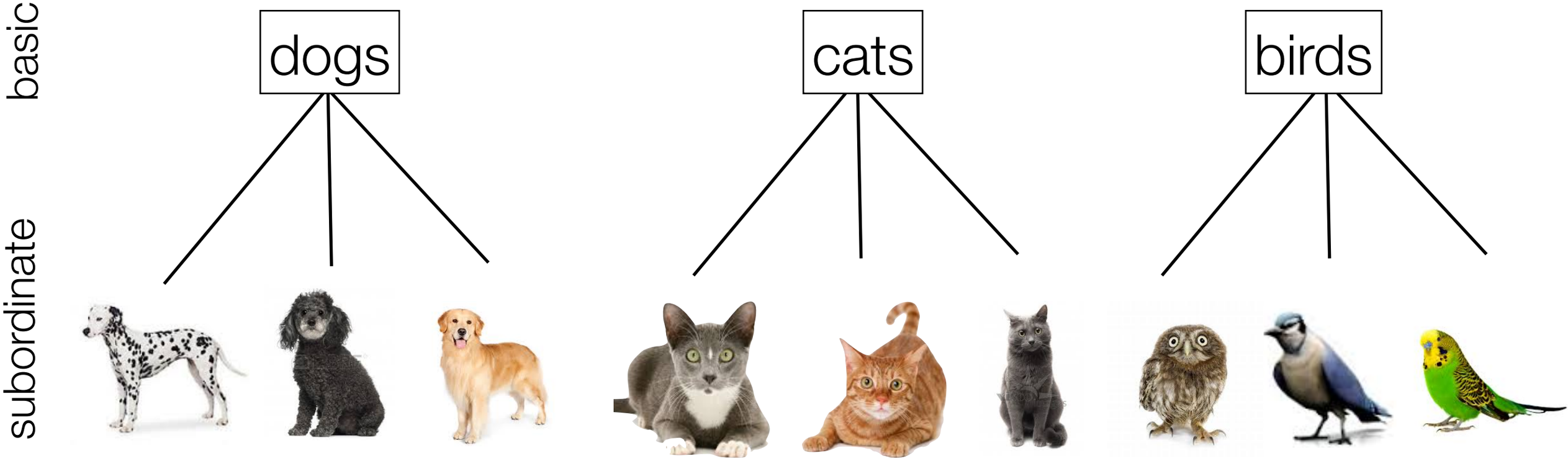
WORD LEARNING

- Many domains have a hierarchical or tree-based conceptual structure



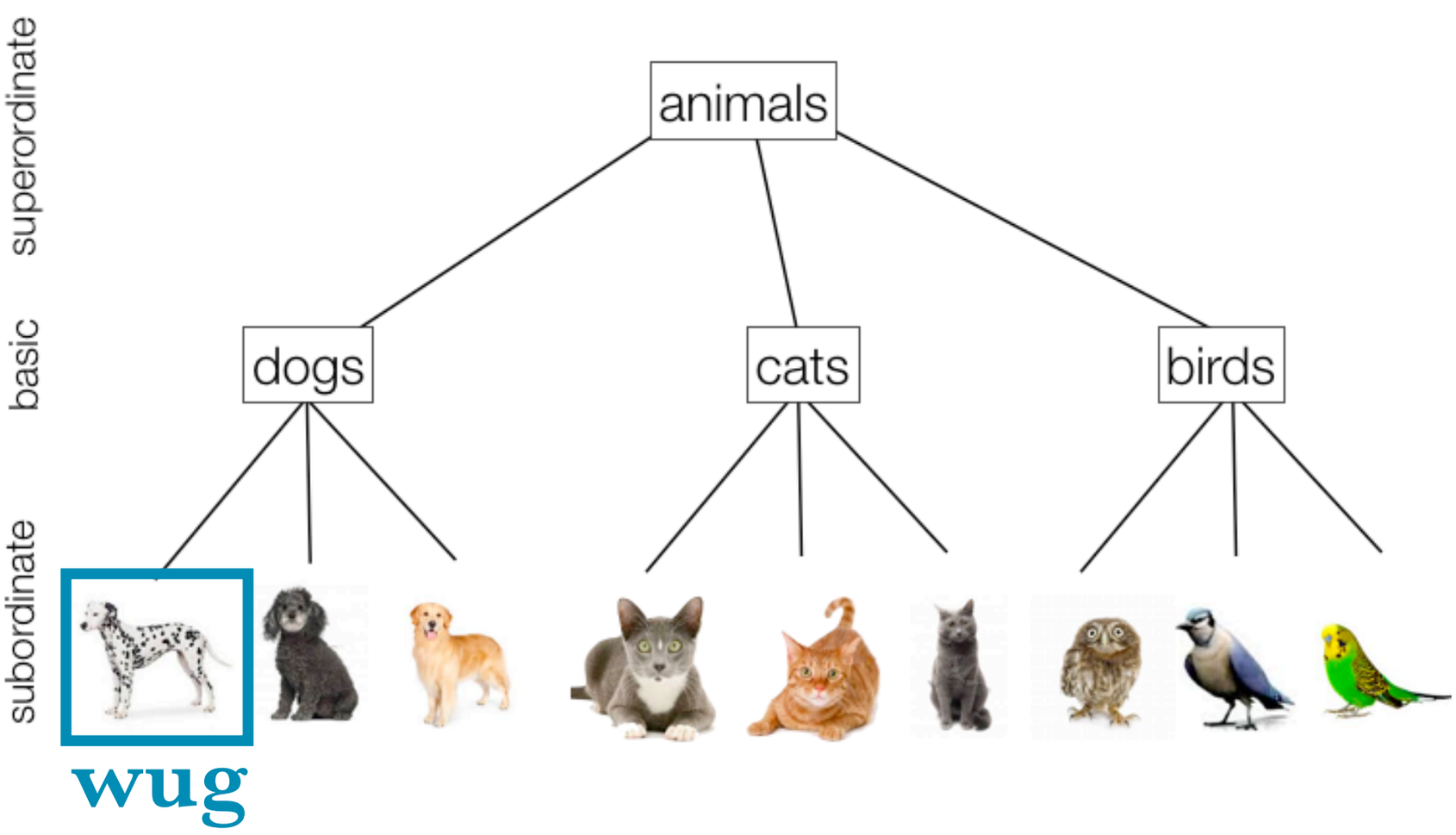
WORD LEARNING

- 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



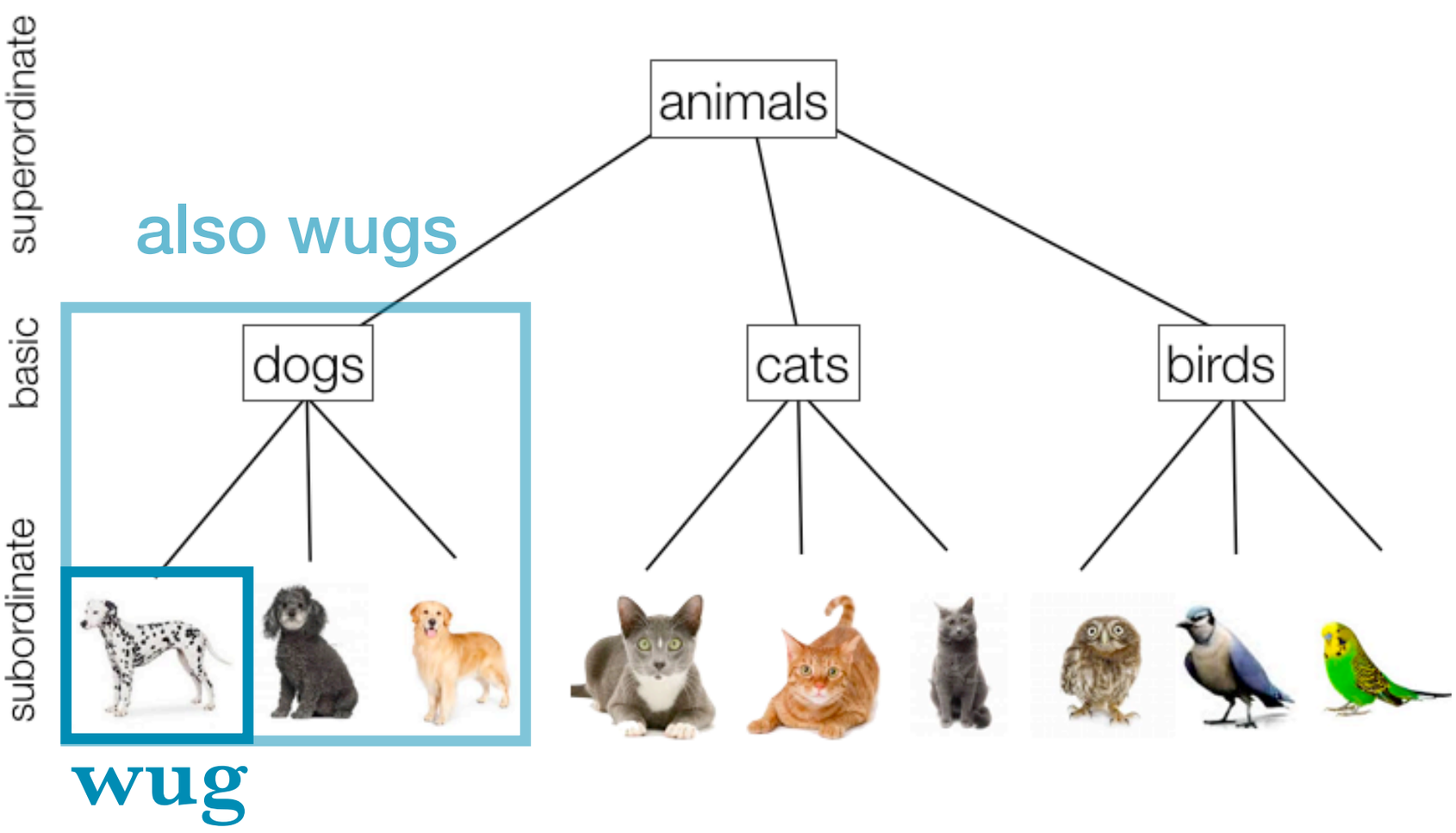
WORD LEARNING

- 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

- 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

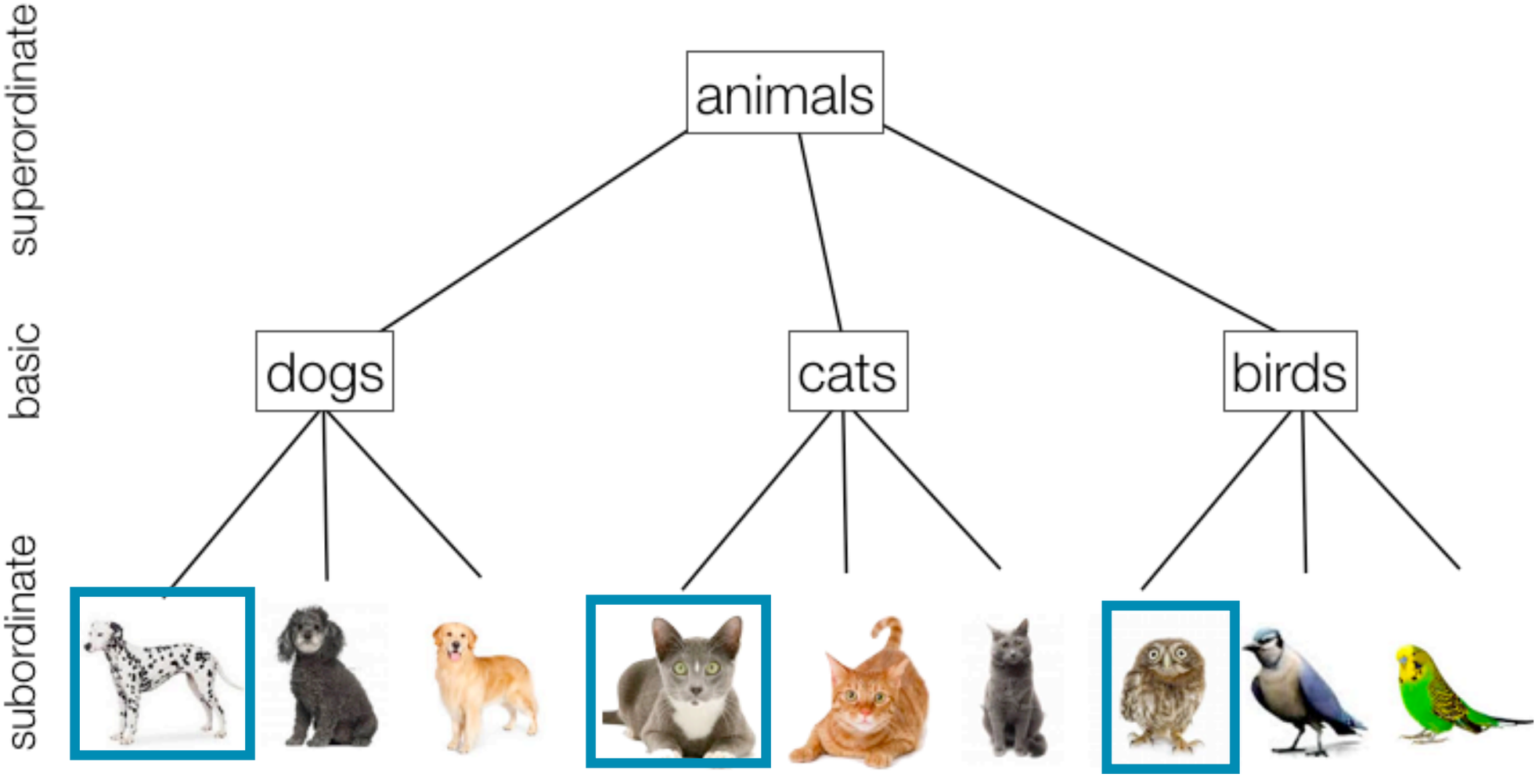
- But what if we are given *three* examples of wugs?
- Then it depends on which three examples, and whether people are reasoning based on the size principle...

IF PEOPLE ARE ASSUMING STRONG SAMPLING…

- Then they should make the tightest possible generalisation

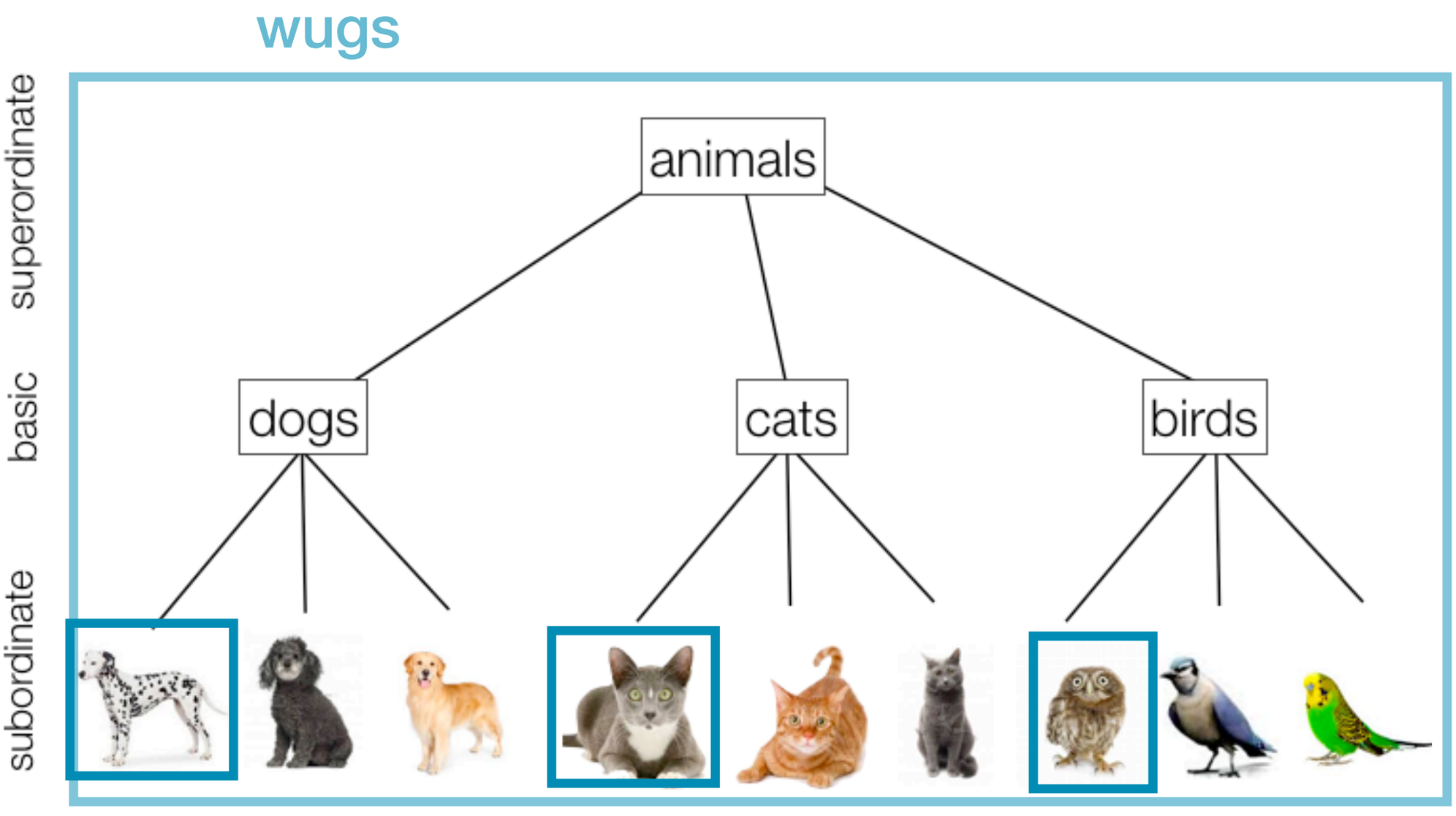
IF PEOPLE ARE ASSUMING STRONG SAMPLING…

- Then they should make the tightest possible generalisation



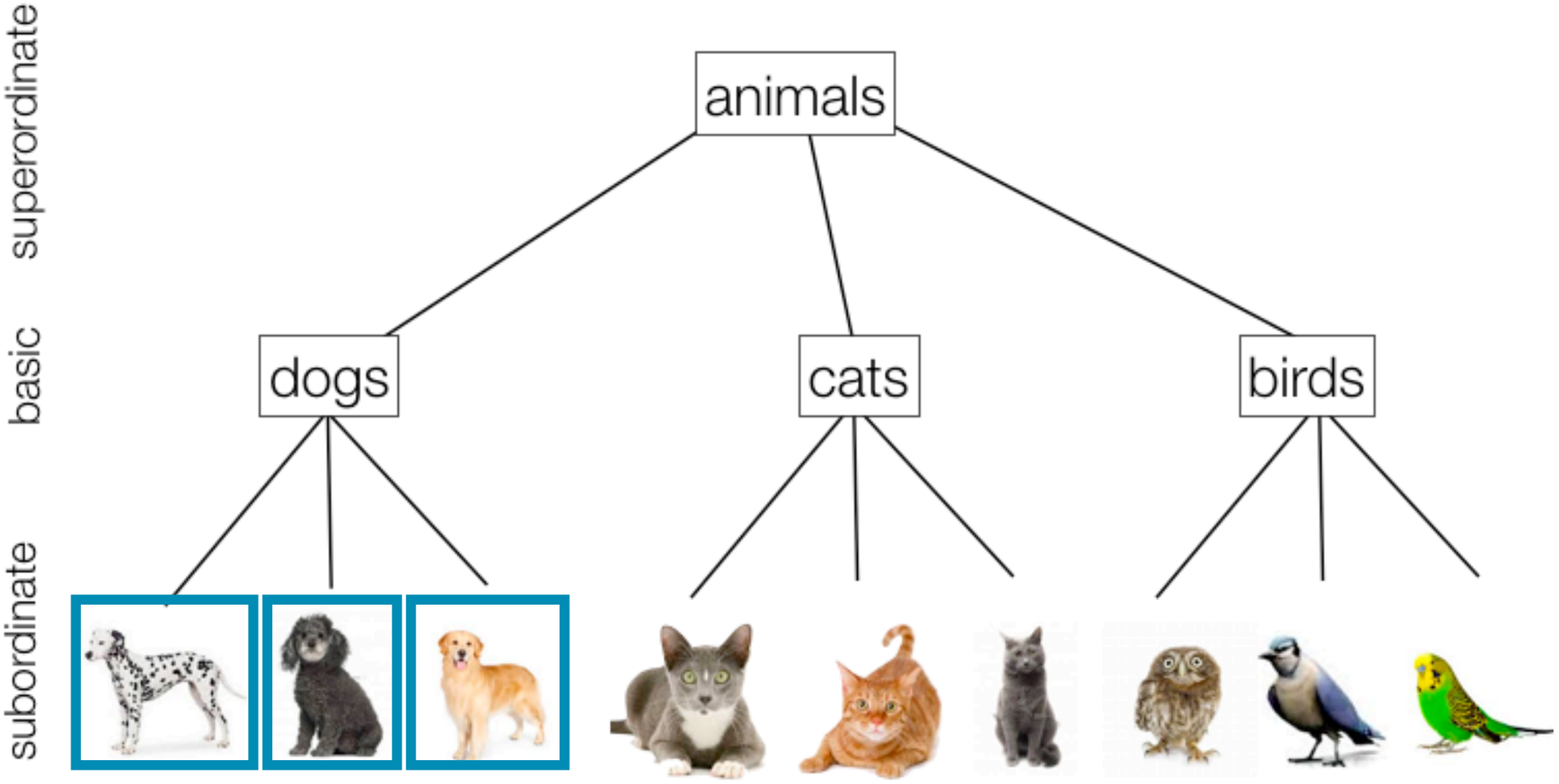
IF PEOPLE ARE ASSUMING STRONG SAMPLING…

- Then they should make the tightest possible generalisation



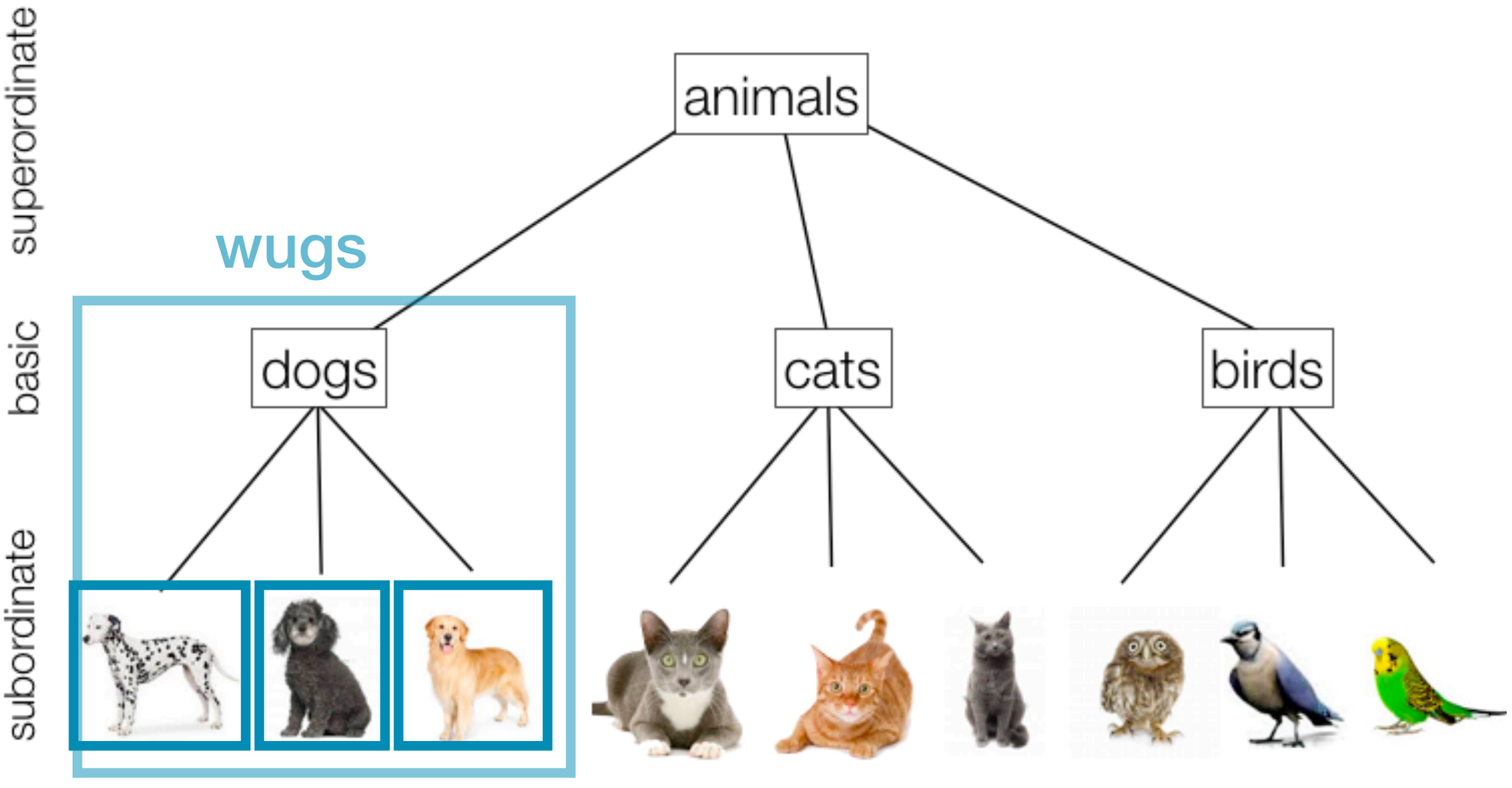
IF PEOPLE ARE ASSUMING STRONG SAMPLING…

- Then they should make the tightest possible generalisation



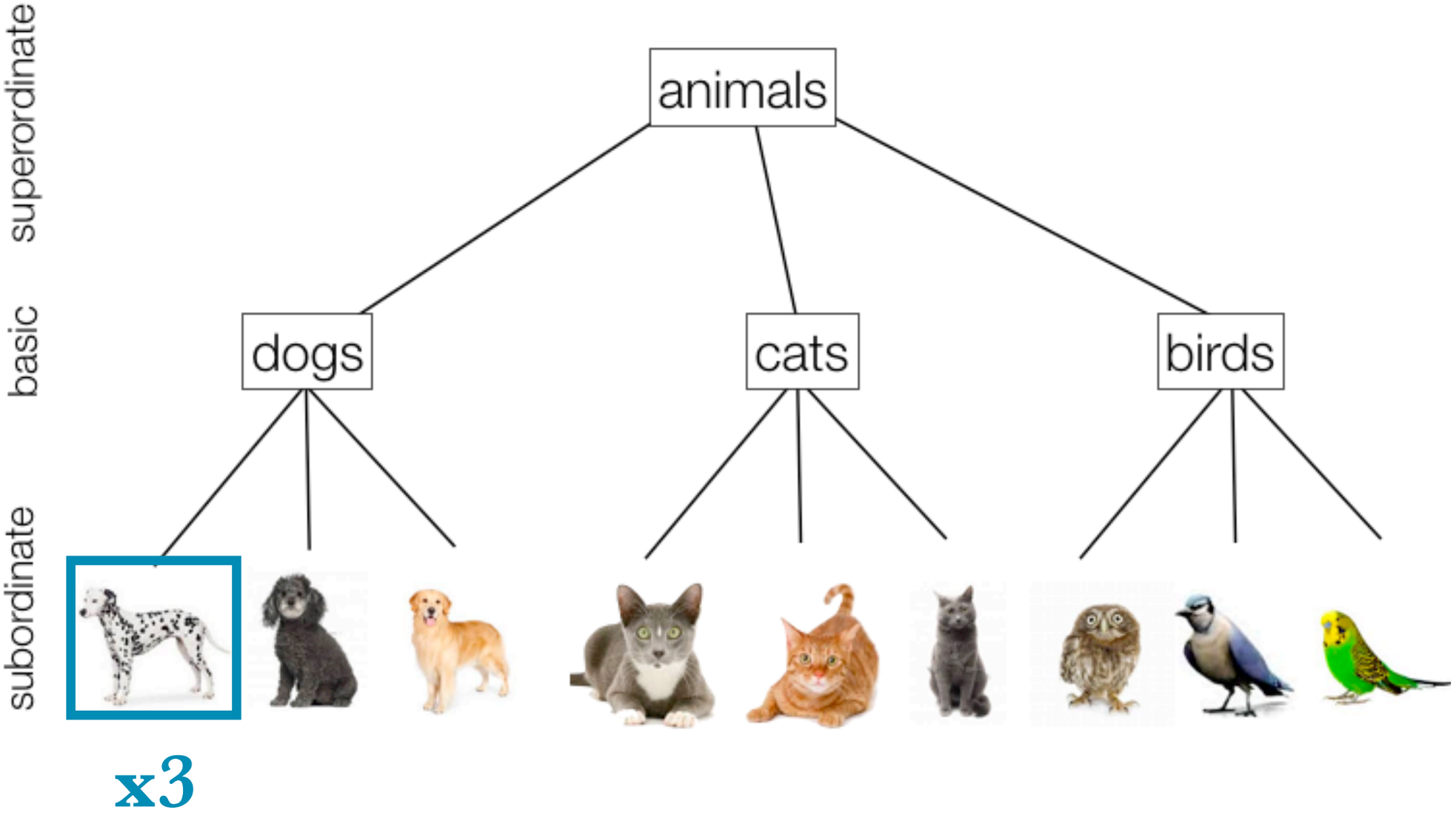
IF PEOPLE ARE ASSUMING STRONG SAMPLING…

- Then they should make the tightest possible generalisation



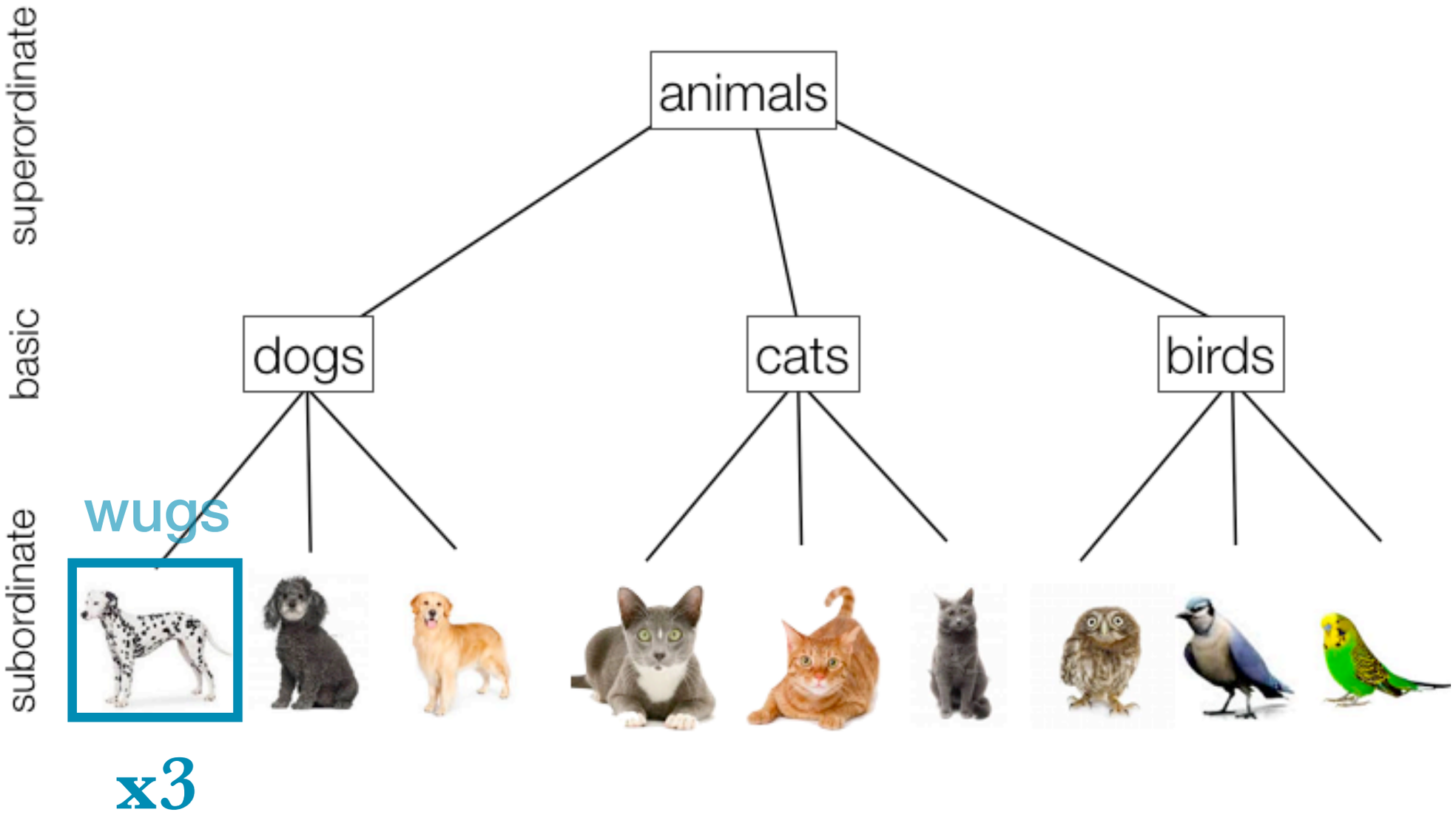
IF PEOPLE ARE ASSUMING STRONG SAMPLING…

- Then they should make the tightest possible generalisation

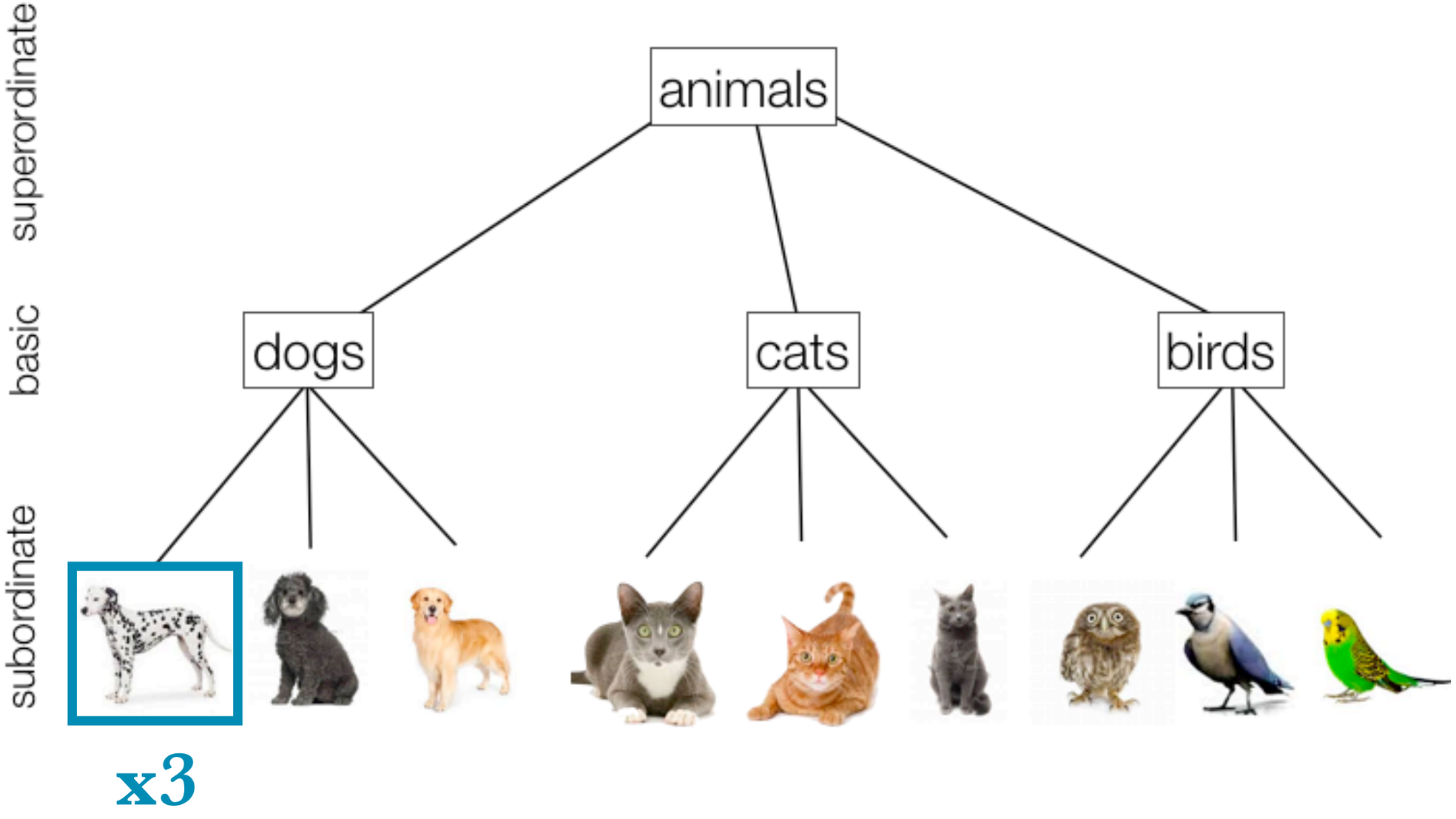


IF PEOPLE ARE ASSUMING STRONG SAMPLING…

- Then they should make the tightest possible generalisation

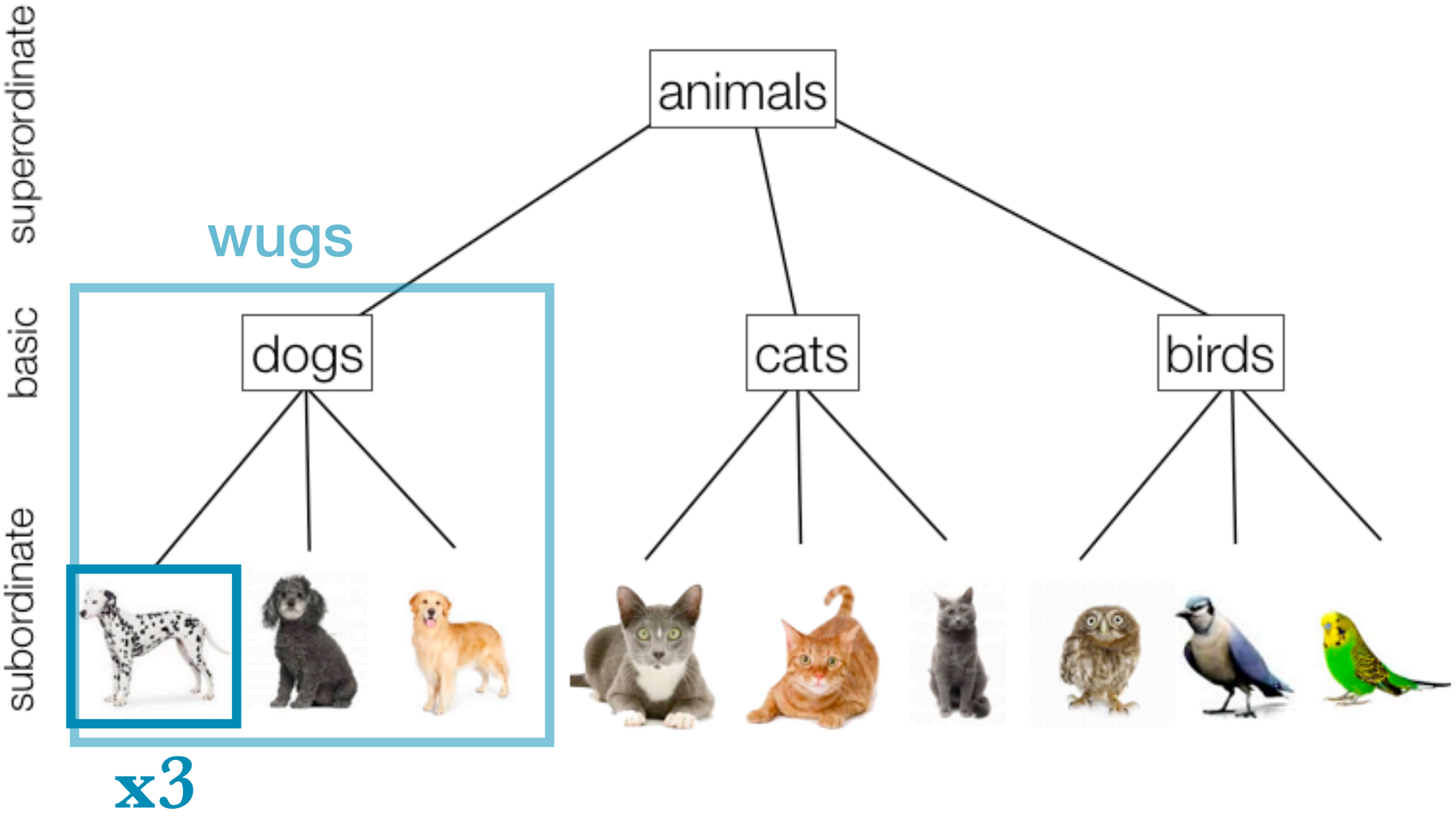


IF PEOPLE ARE ASSUMING WEAK SAMPLING…



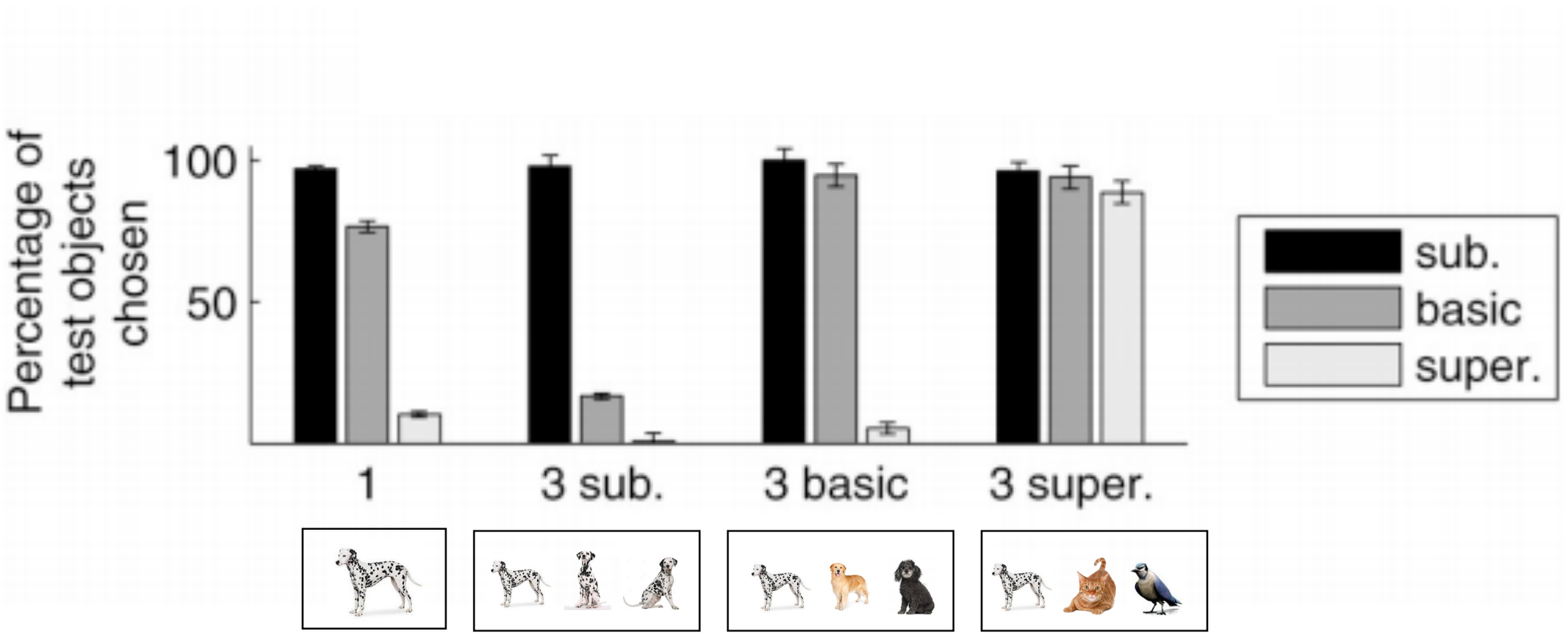
IF PEOPLE ARE ASSUMING WEAK SAMPLING…

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



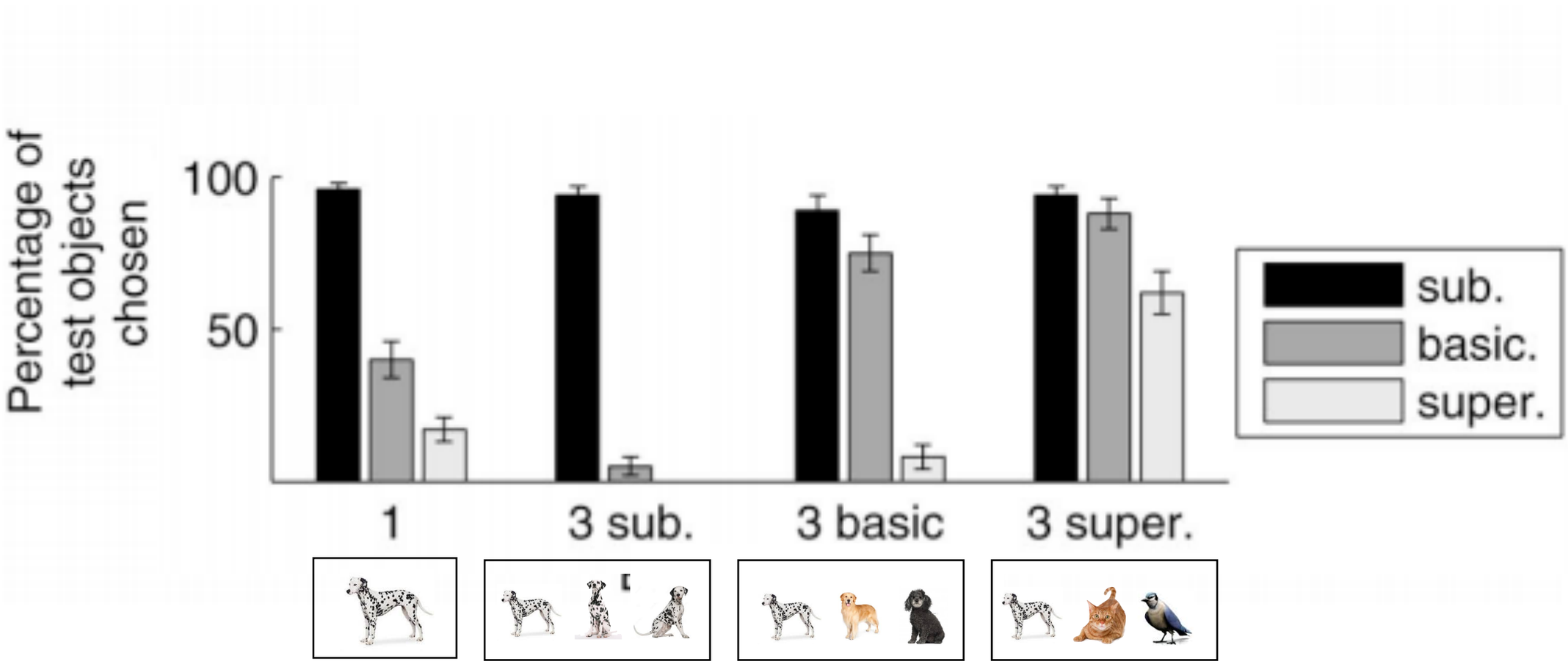
EXPERIMENTAL TEST

- Adults generalise as predicted by the size principle

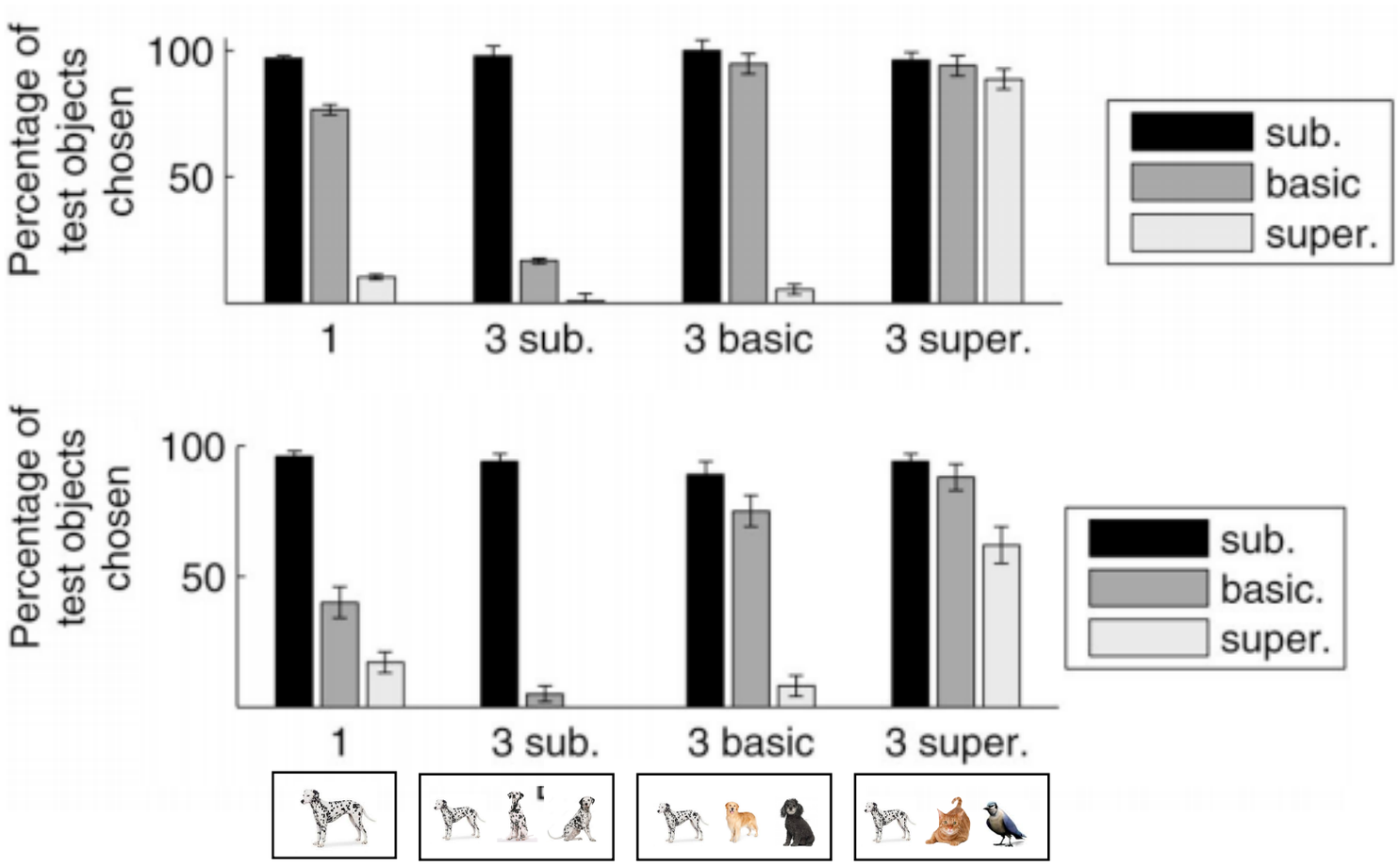


EXPERIMENTAL TEST

- Four-year old children do the same thing!



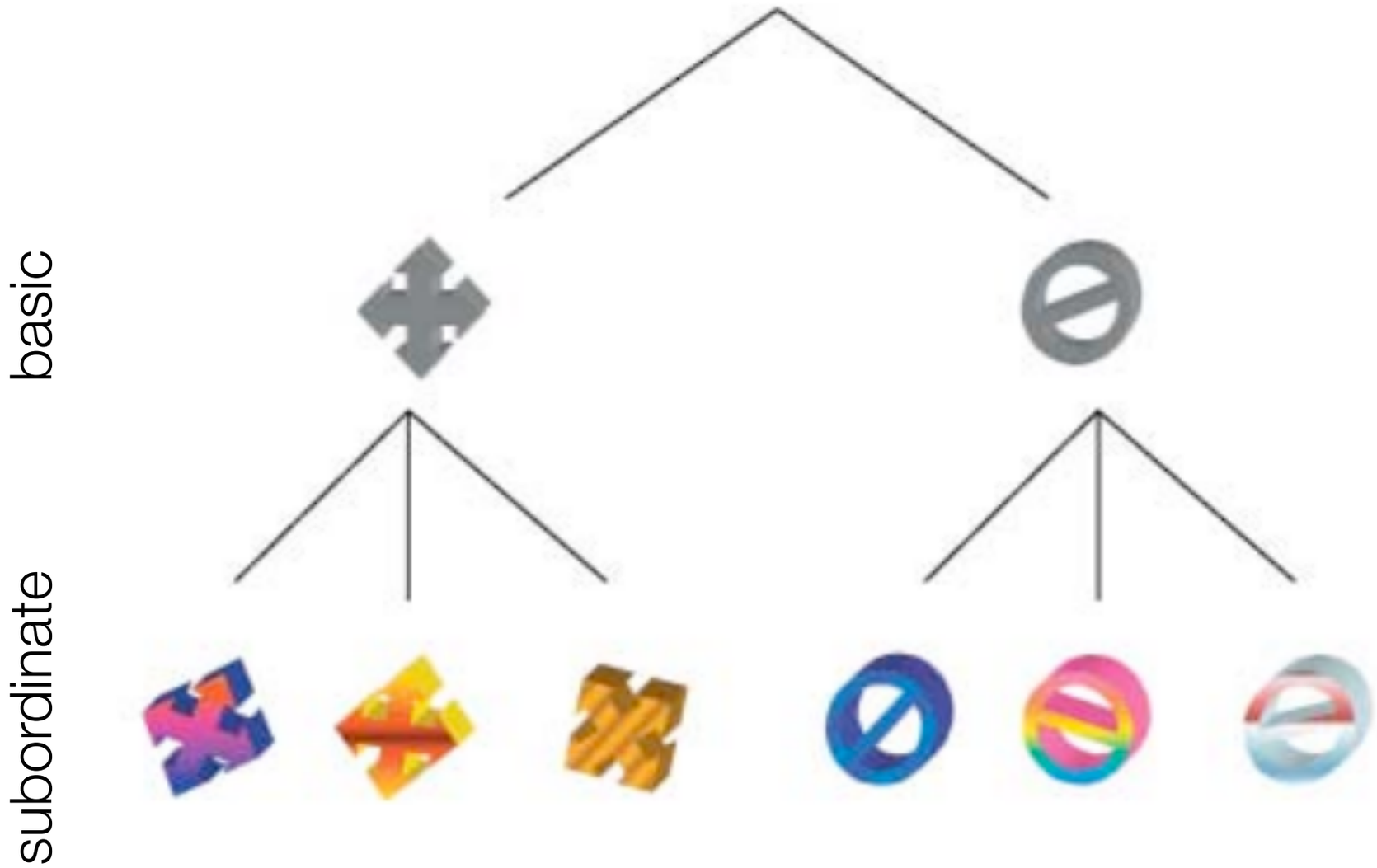
EXPERIMENTAL TEST



- 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)

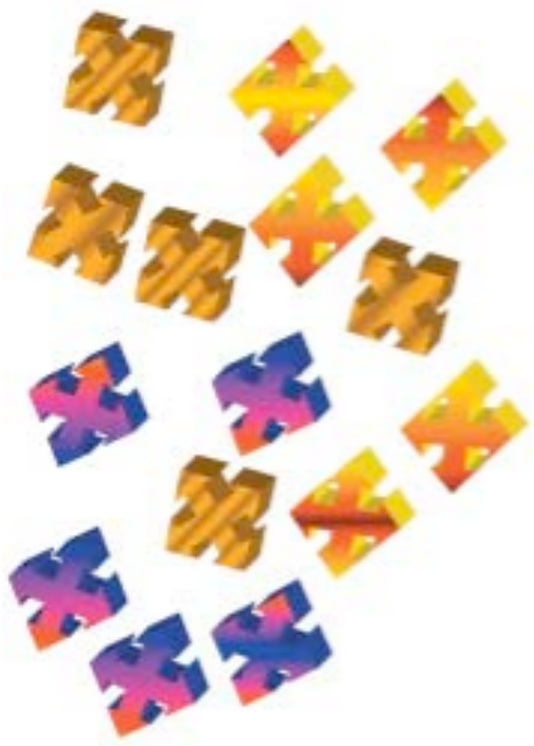


Xu & Tenenbaum, 2007b

CHANGING SAMPLING ASSUMPTIONS

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

Teacher-driven

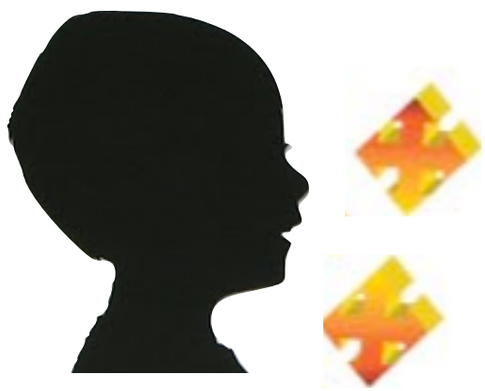


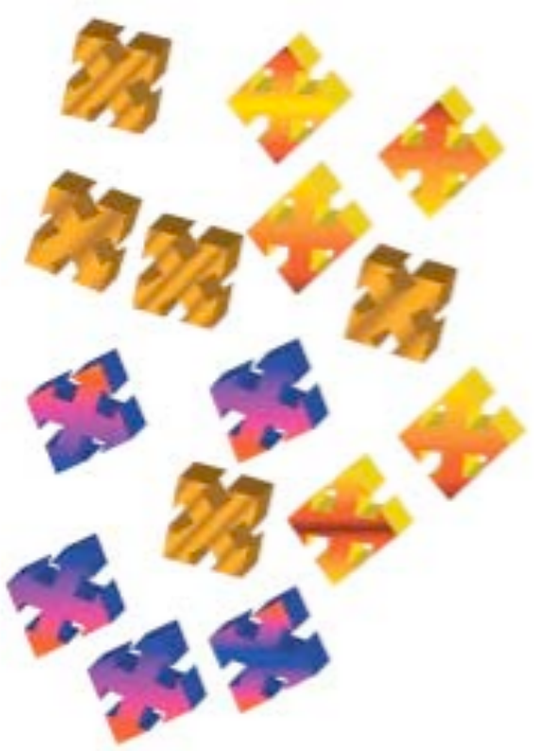
CHANGING SAMPLING ASSUMPTIONS

- 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

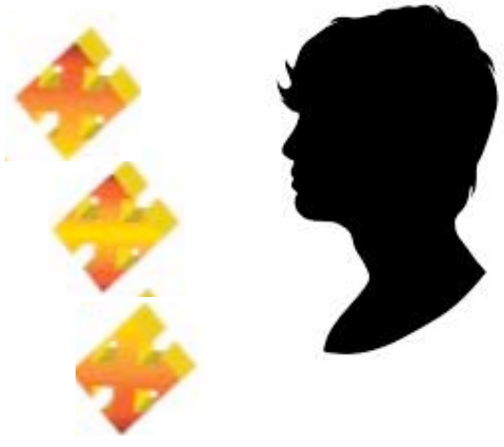
- This time we vary how data are sampled (also make the objects novel)

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

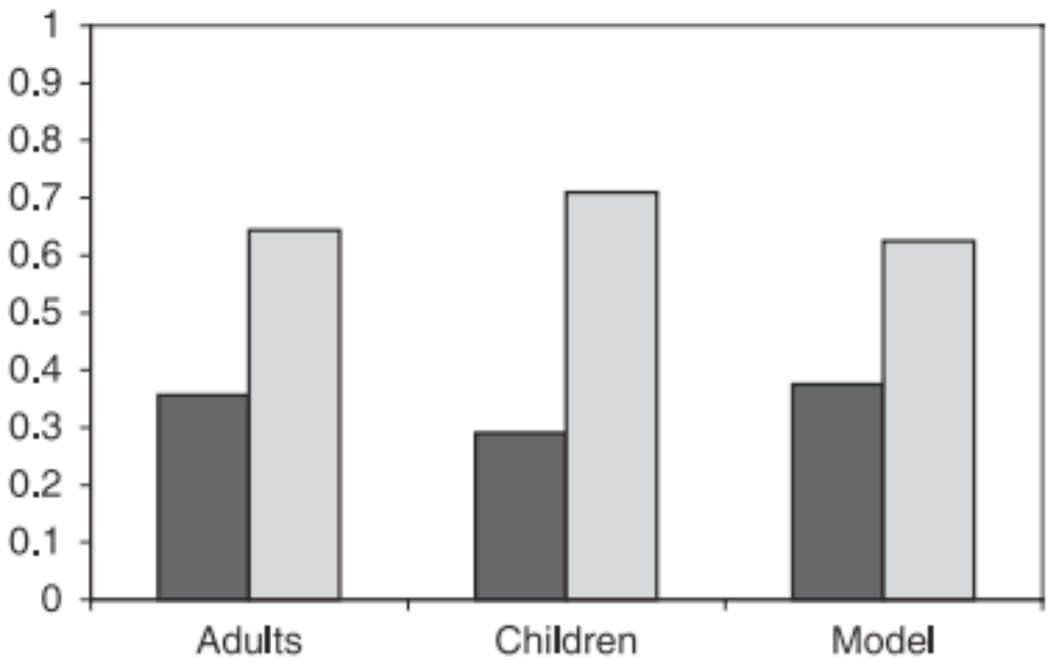
People saw 3 subordinate items, always chosen by the teacher



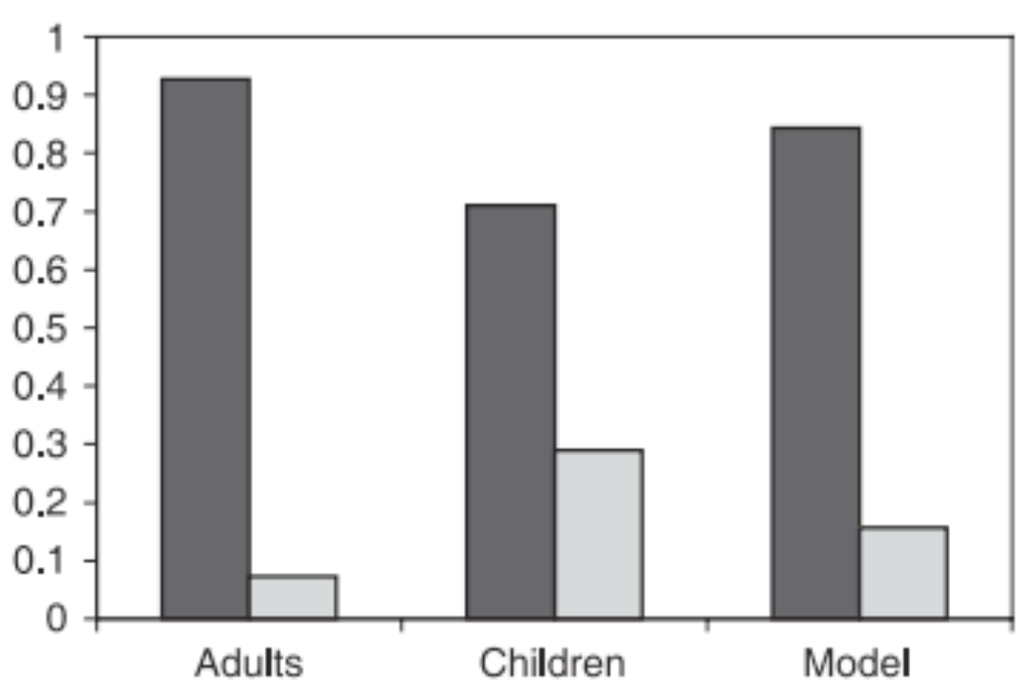
CHANGING SAMPLING ASSUMPTIONS

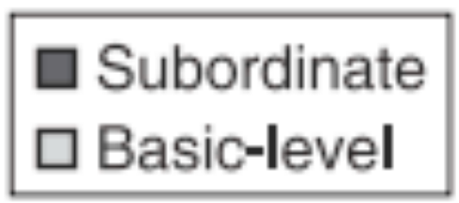
- People generalise tightly only when the teacher sampled the data

Learner-driven



Teacher-driven





SAMPLING ASSUMPTIONS

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.

But we've considered only two different ways data might be generated: strong (helpful) or weak.

In real life, data can be **censored** in many ways that should affect generalisation

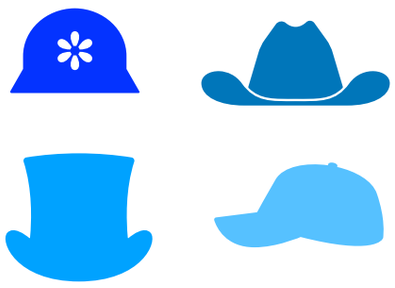
CENSORED DATA

Suppose I have a box of clothing accessories, but you don't know what's in it. I like to play a game where I pick examples and you need to predict what colour they will be.

Category sampling:
I choose only hats

Small N

Large N



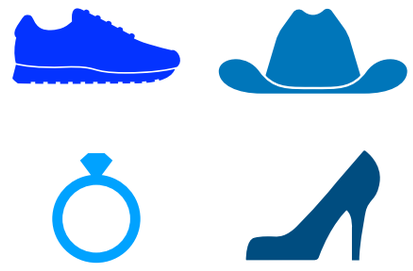
What is the probability that a non-hat is blue?

No size principle: similar with both large and small N

Property sampling:
I choose only blue things

Small N

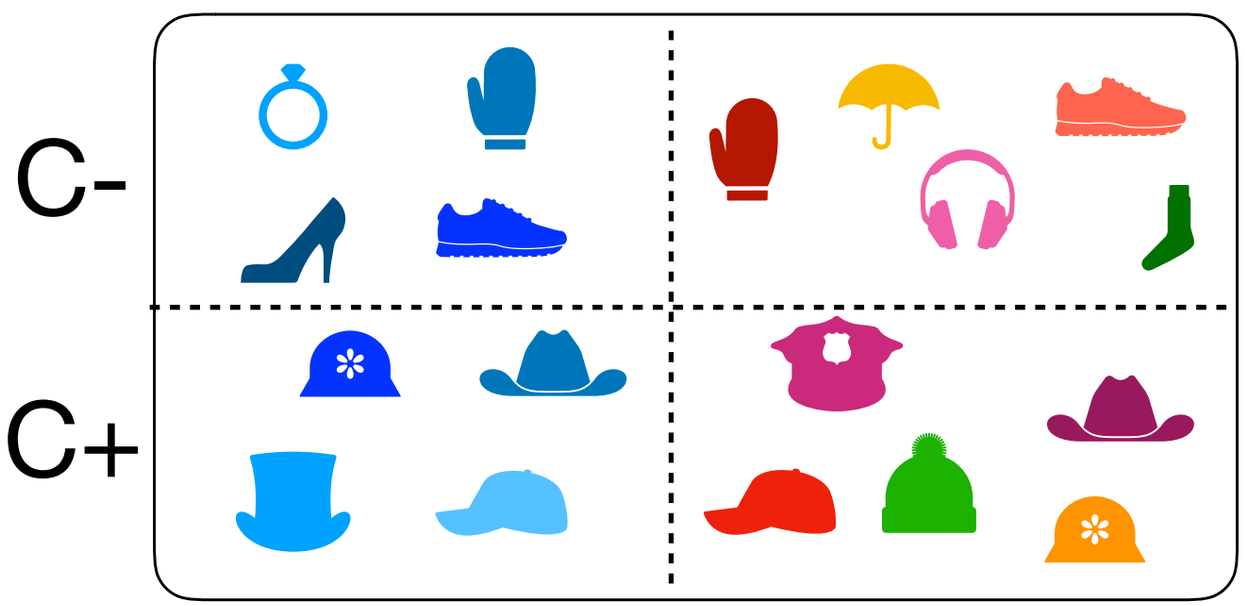
Large N



What is the probability that a non-hat is blue?

Intuitively less with large N

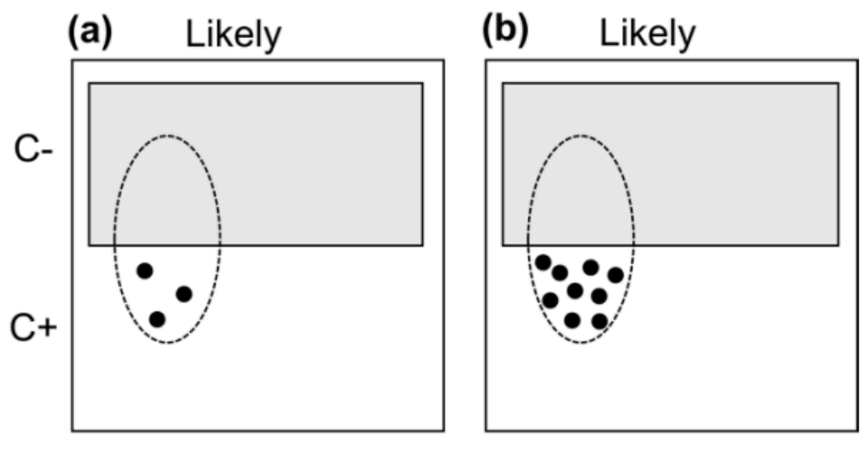
CENSORED DATA



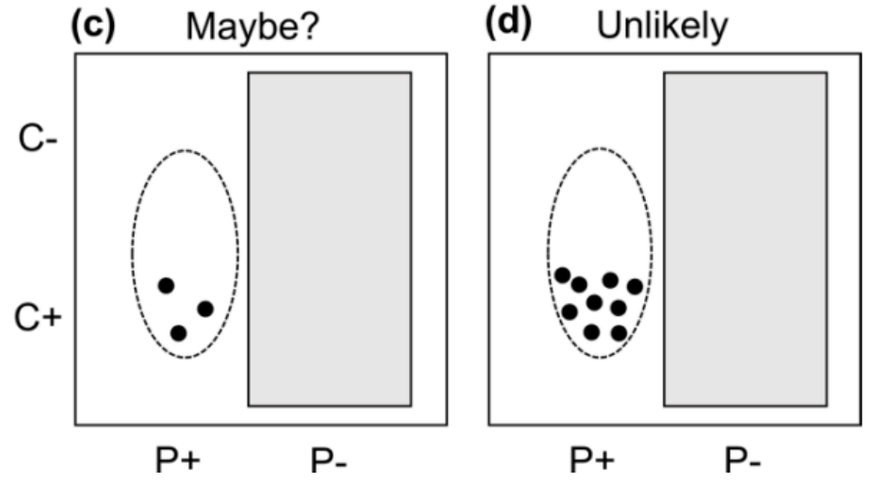
What is the probability that a non-hat is blue?

What is the probability of C-P+?

Prediction of category sampling with increasing N

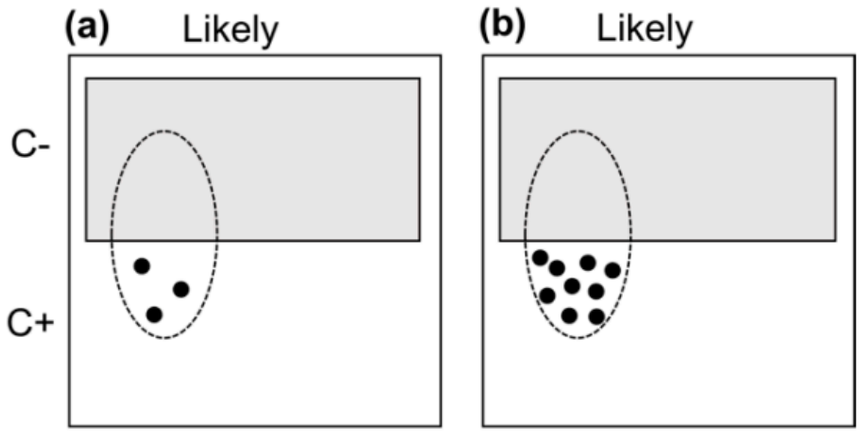


Prediction of property sampling with increasing N

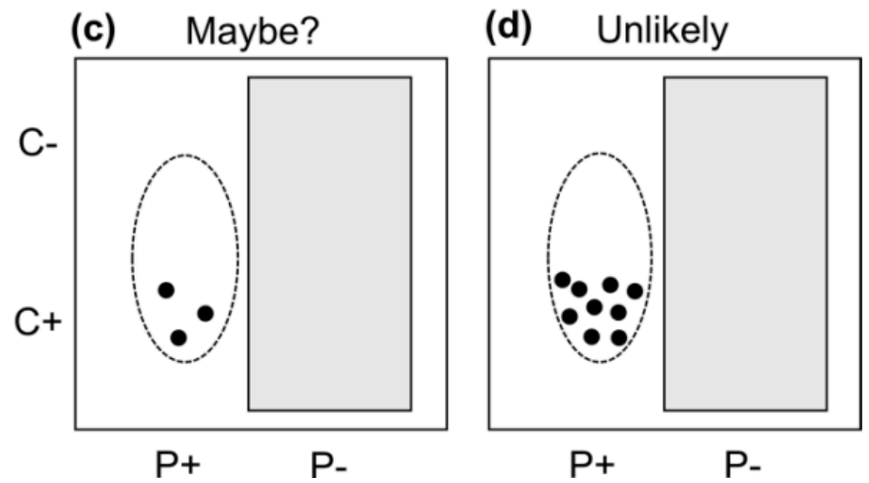


OUR TASK: DESIGN AN EXPERIMENT TO TEST THIS HYPOTHESIS

Prediction of category sampling with increasing N



Prediction of property sampling with increasing N



What is the probability of C-P+?

- Conditions / manipulation?
- Task?
- Instructions?