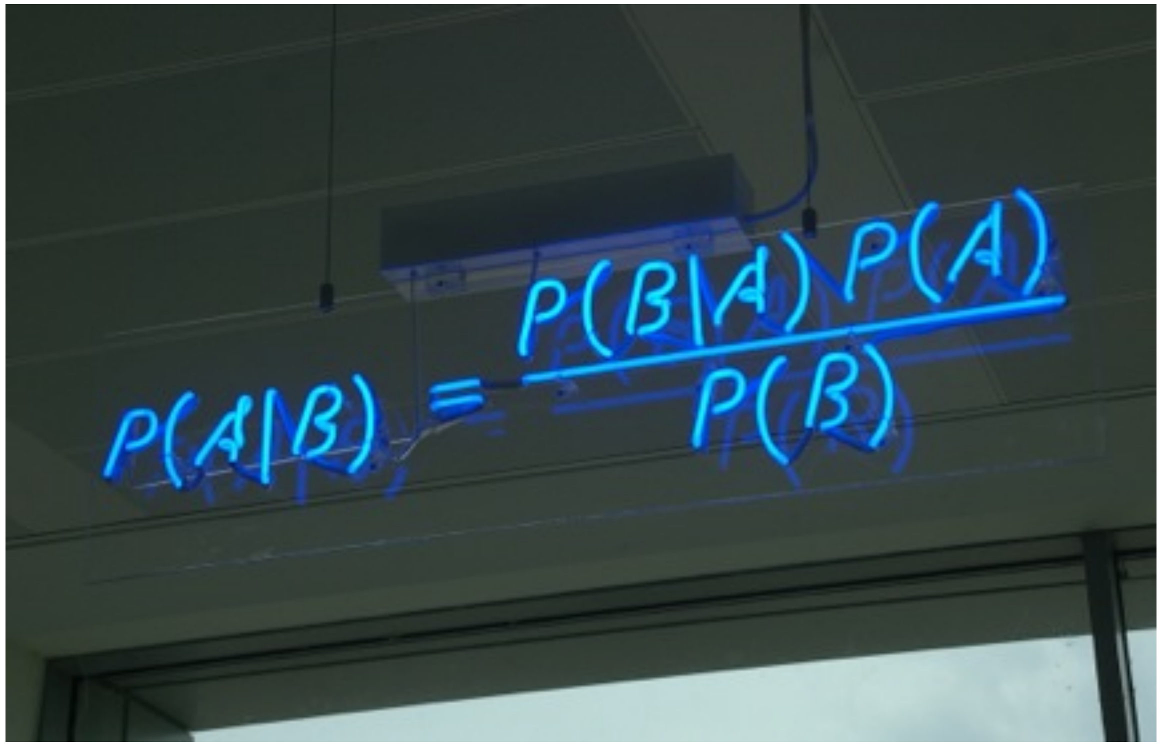
Generalisation and inductive reasoning

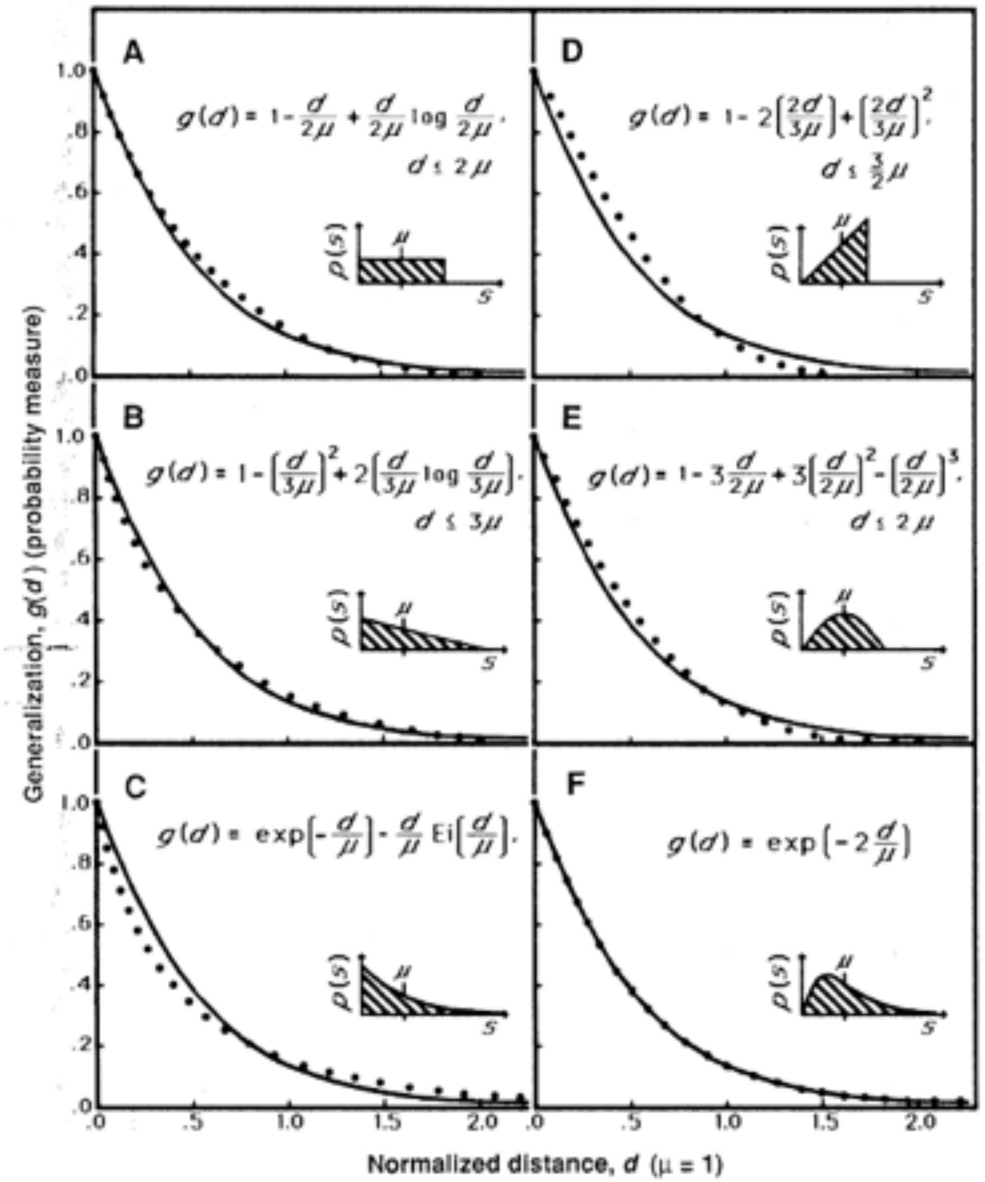
Computational Cognitive Science 2014

Dan Navarro

Where are we at?

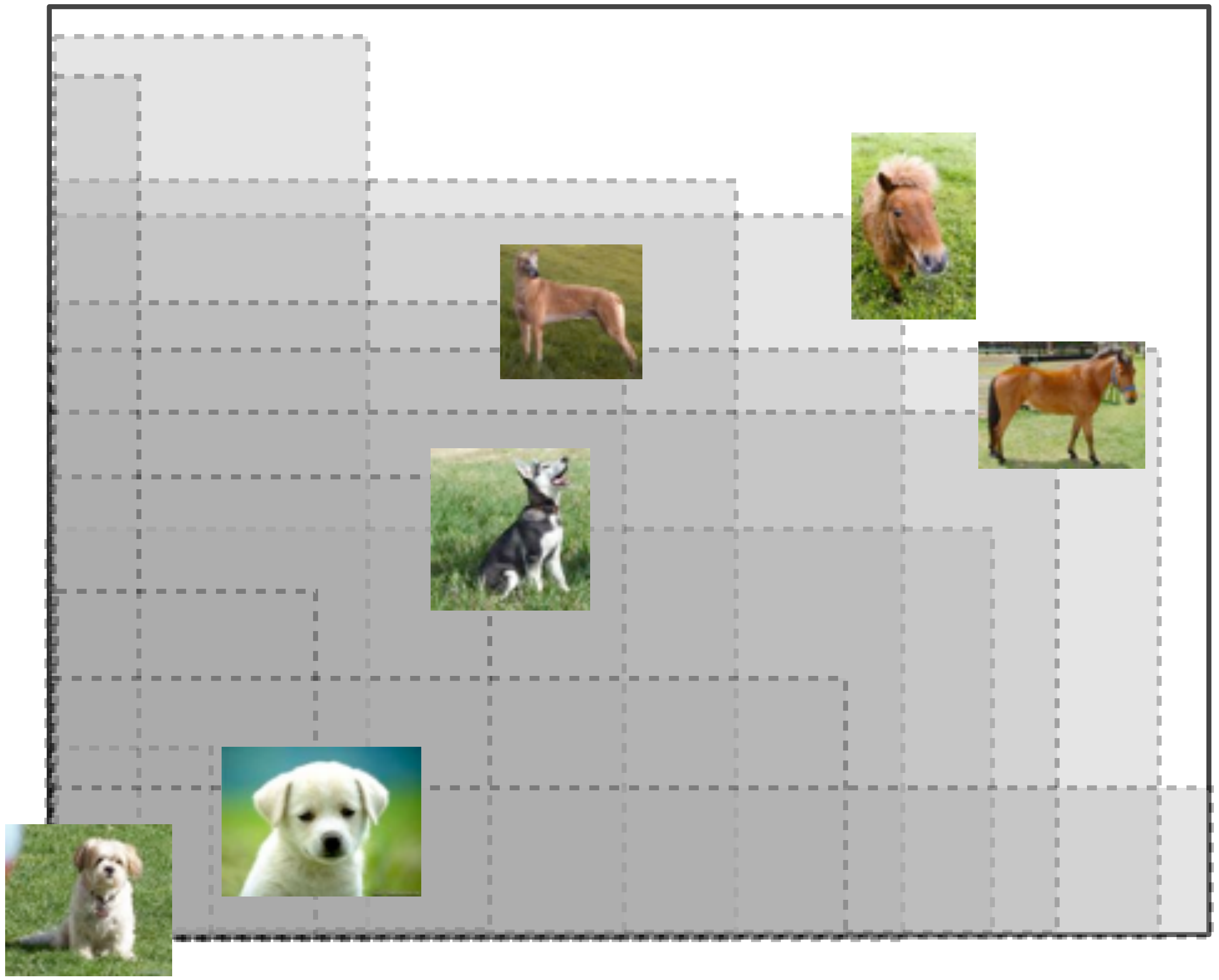


Bayesian statistics as a general purpose tool for doing inductive inference



A specific Bayesian model for human (and non-human) inductive generalisation

Critical insight from last lecture: Structure underpins statistical learning

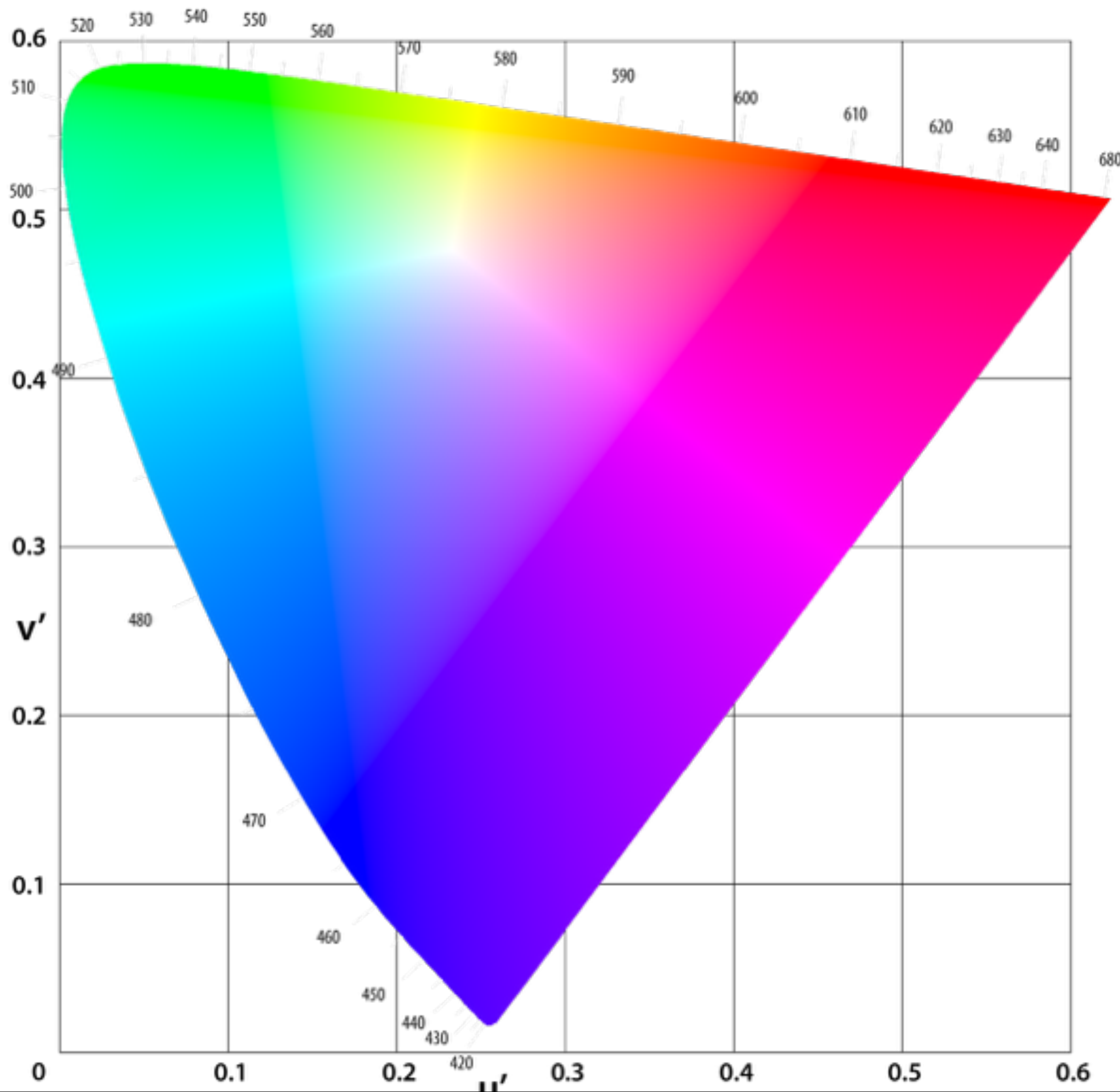


(1) We can organise stimuli into a psychological space (using MDS).

(2) The <u>hypothesis</u> space can be described as a set of possible "regions" in this space

(3) Inductive generalisations are shaped by this structure

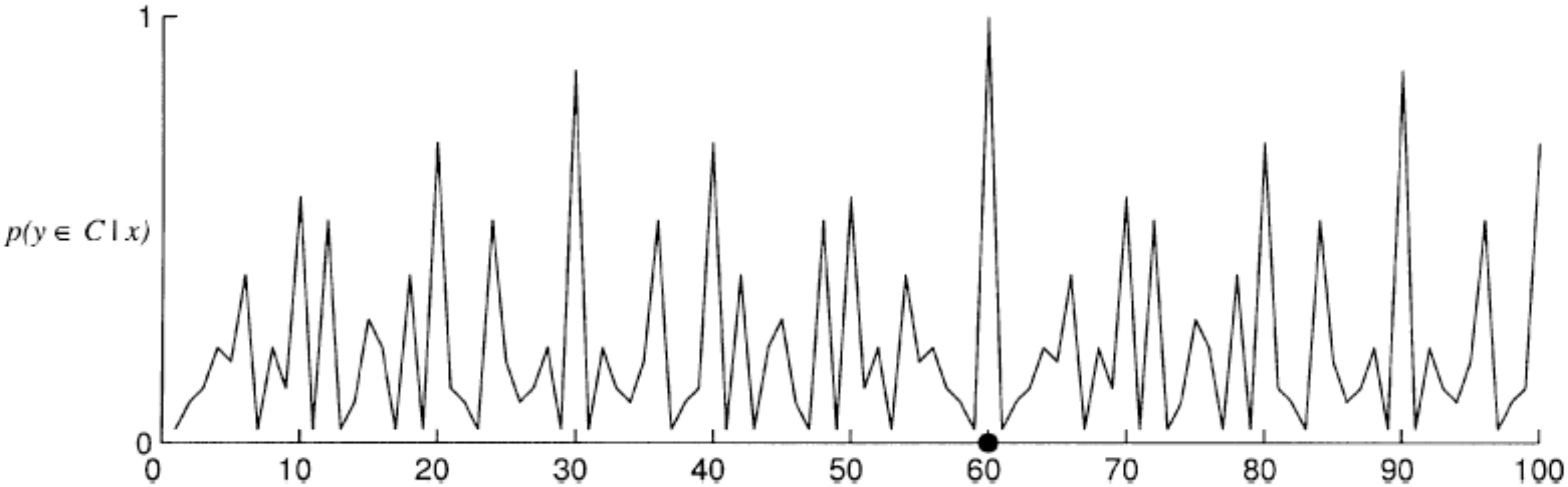
There are some real domains that are truly spatial in structure:



Colour is a genuine “space”.

Generalisations about colours behave as if the hypothesis space consisted of regions in colour space (up to a point)

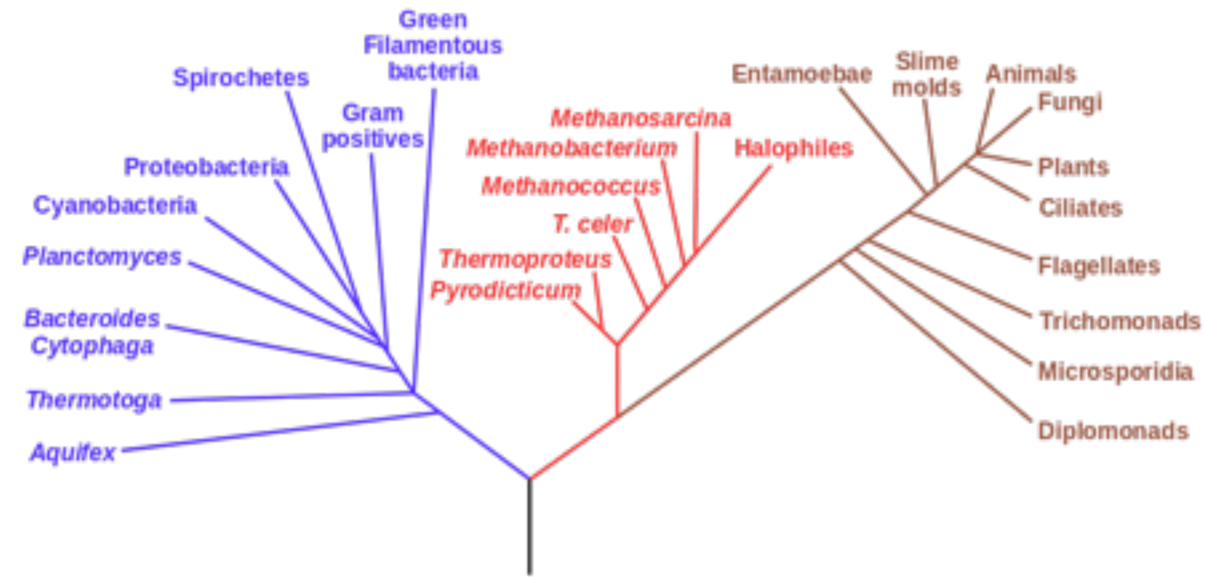
But we also talked about some domains that seem to be less structured

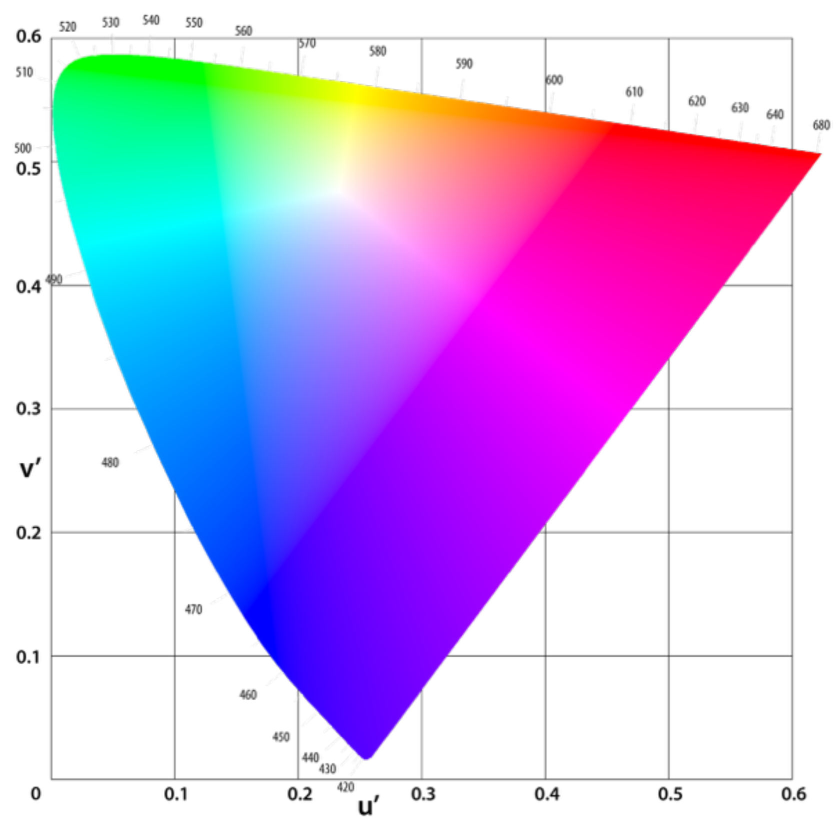


“Number” is only partially spatial: “magnitude” behaves like a psychological space, but “arithmetic properties” have a different structure

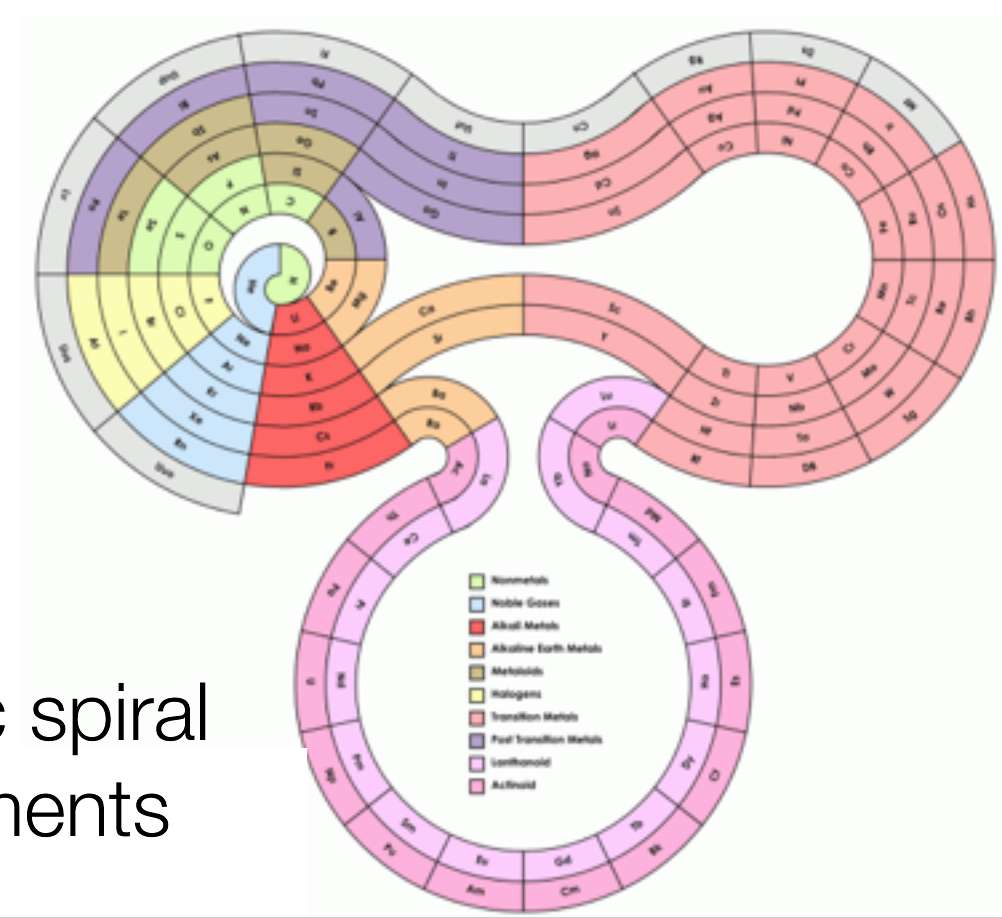
Many different structures exist

evolutionary tree





colour space

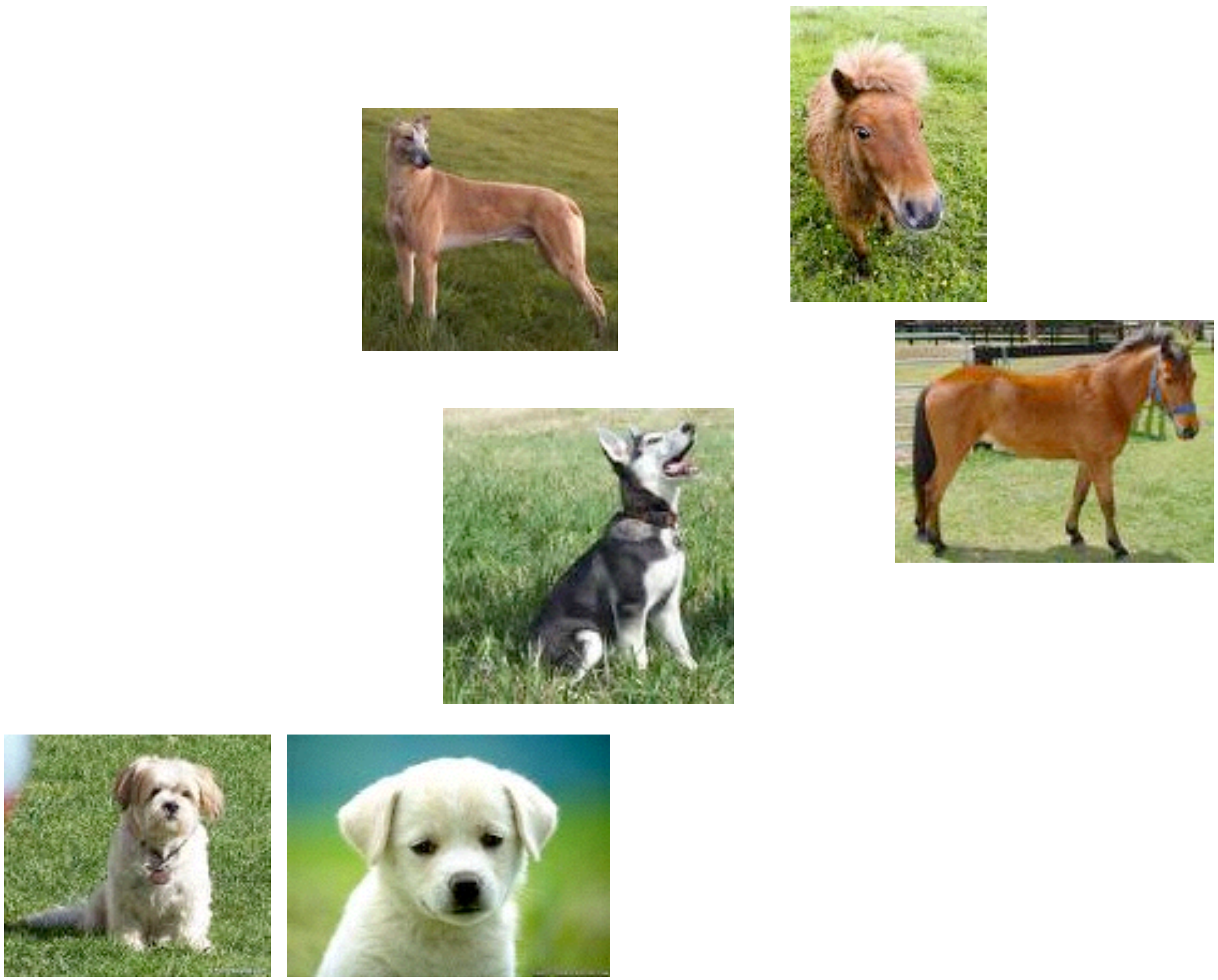


periodic spiral
of elements

Let’s build a structured model of inductive generalisation that is appropriate for biological entities

(based on Sanjana & Tenenbaum 2003)

Our original motivating example used the wrong structure…



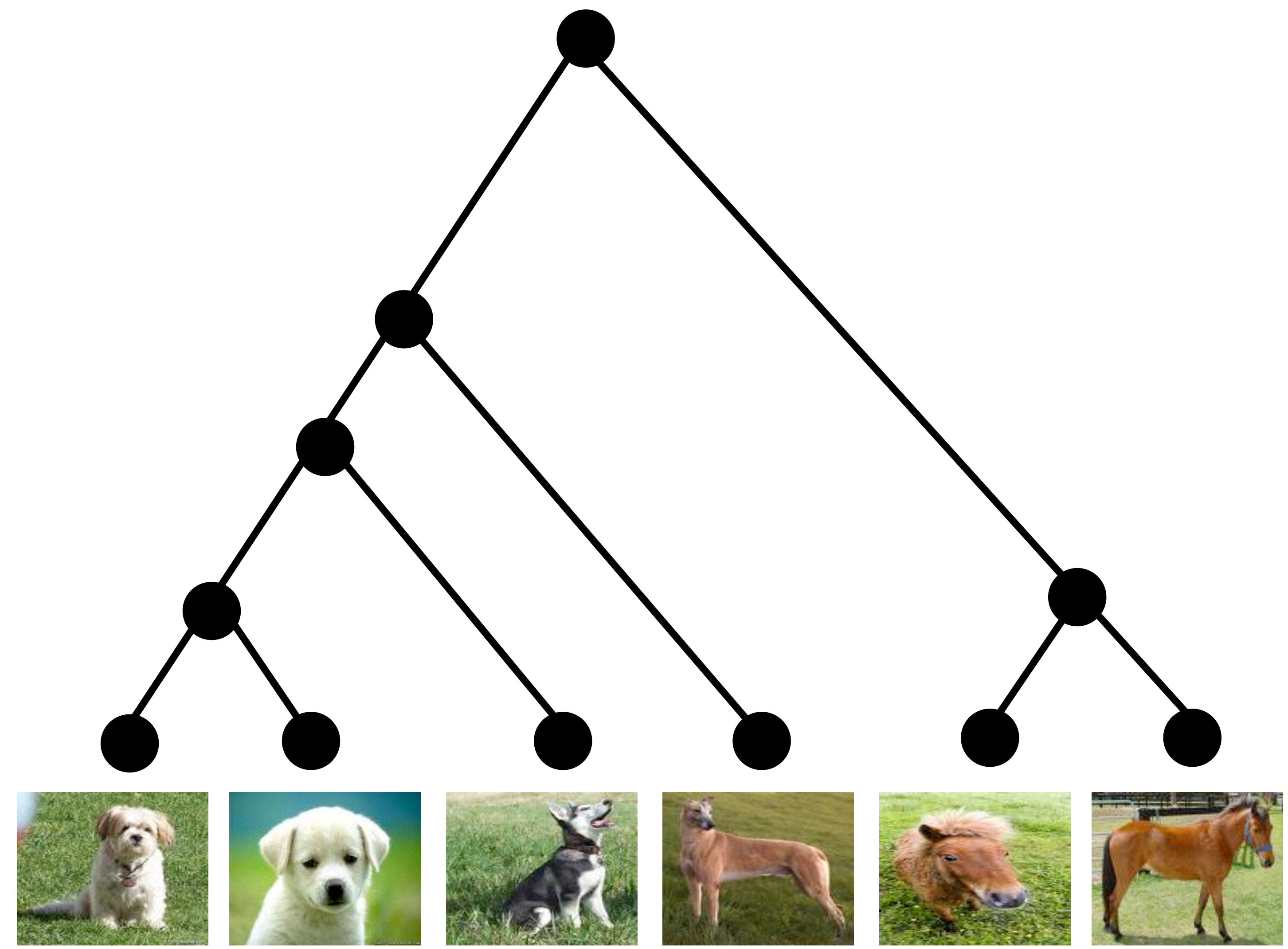
Biological categories probably don't naturally form a "space"

A tree structure might make more sense here

Hierarchical clustering

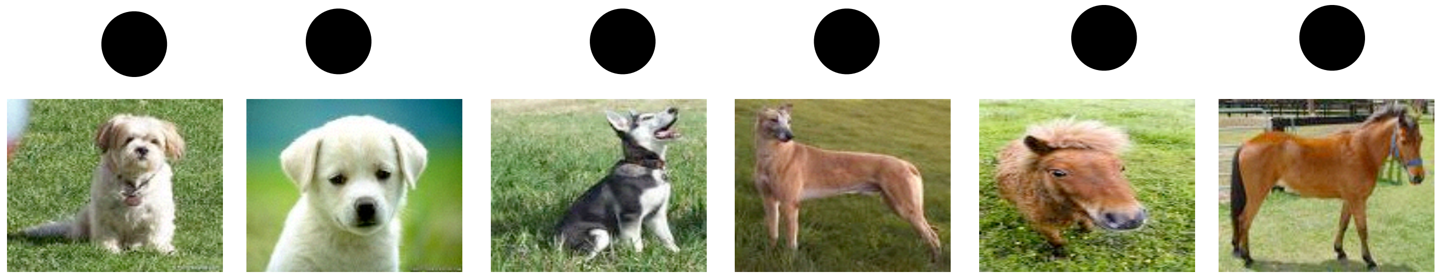
Dissimilarity matrix → Taxonomic tree

	0	1	2	4	5	6
	1	0	1	3	4	6
	2	1	0	1	3	4
	4	3	1	0	2	3
	5	4	3	2	0	1
	6	6	4	3	1	0



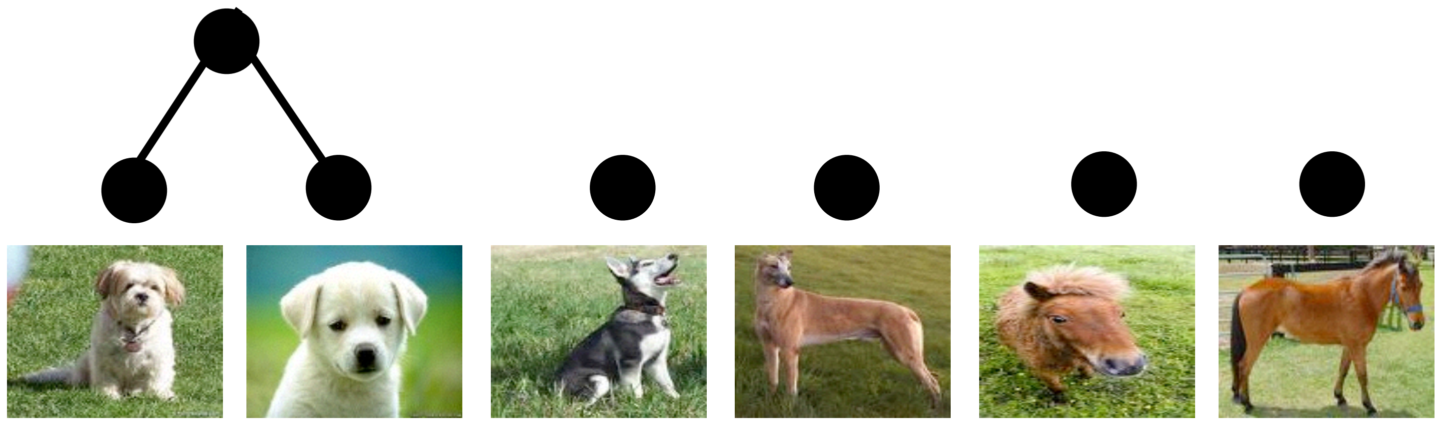
Agglomerative hierarchical clustering

Every object starts in its own cluster



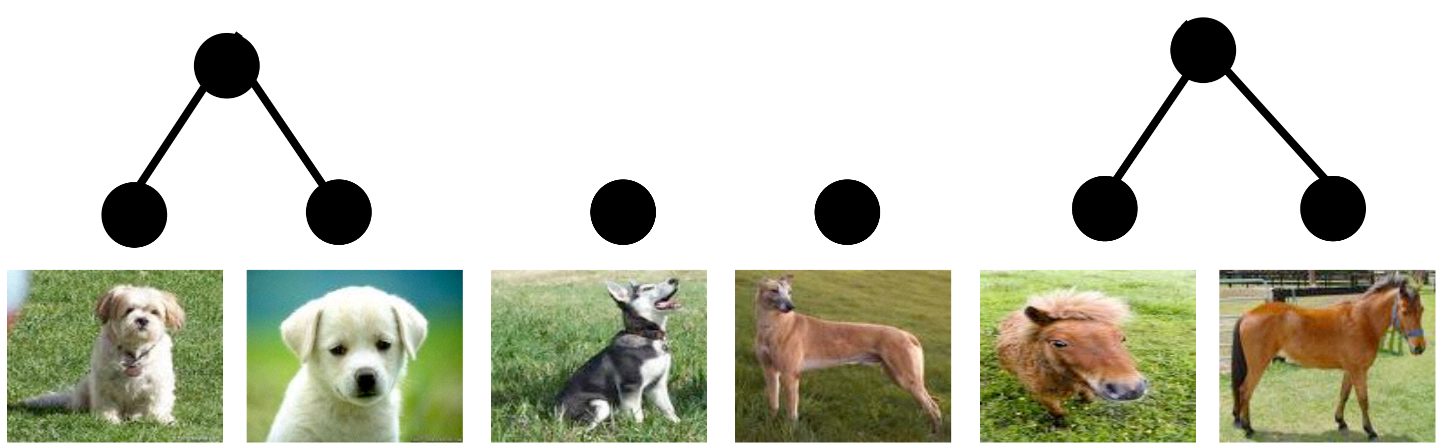
Agglomerative hierarchical clustering

Merge the two “**closest**” clusters

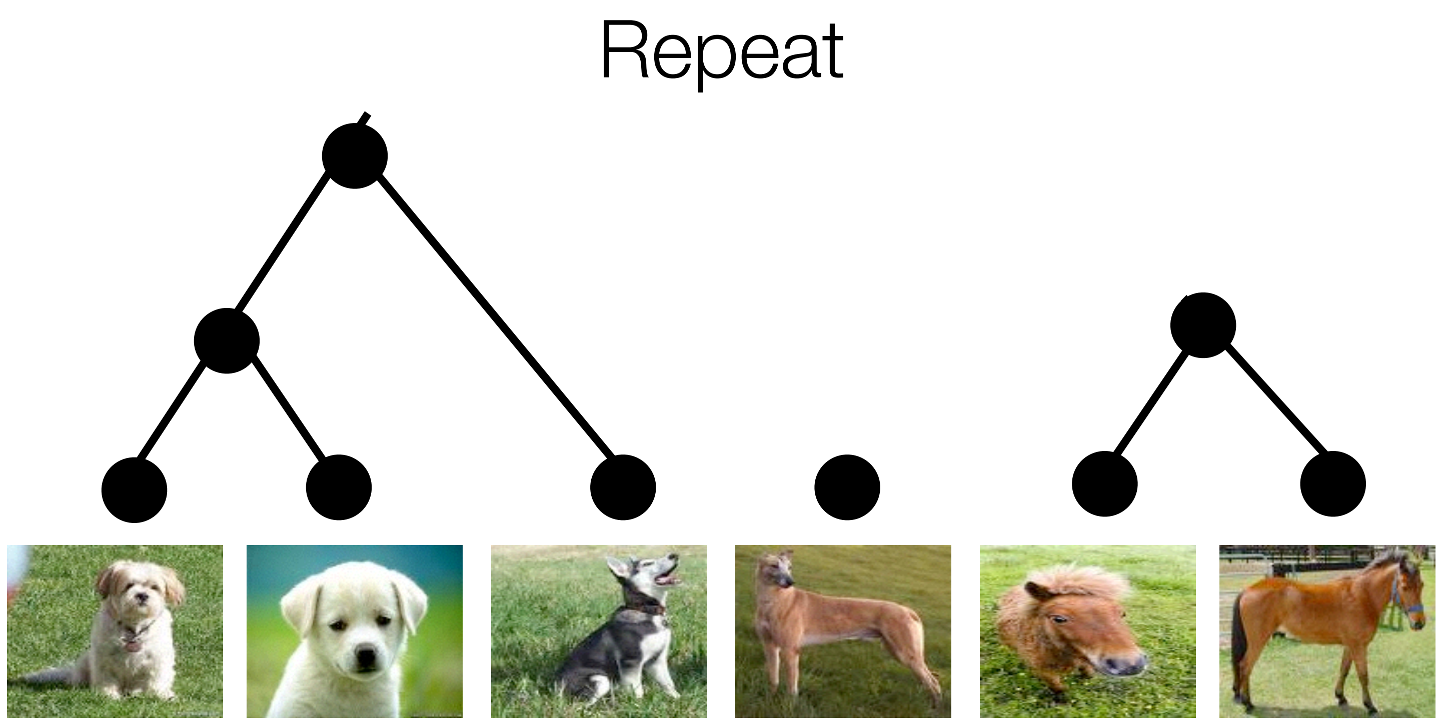


Agglomerative hierarchical clustering

Repeat

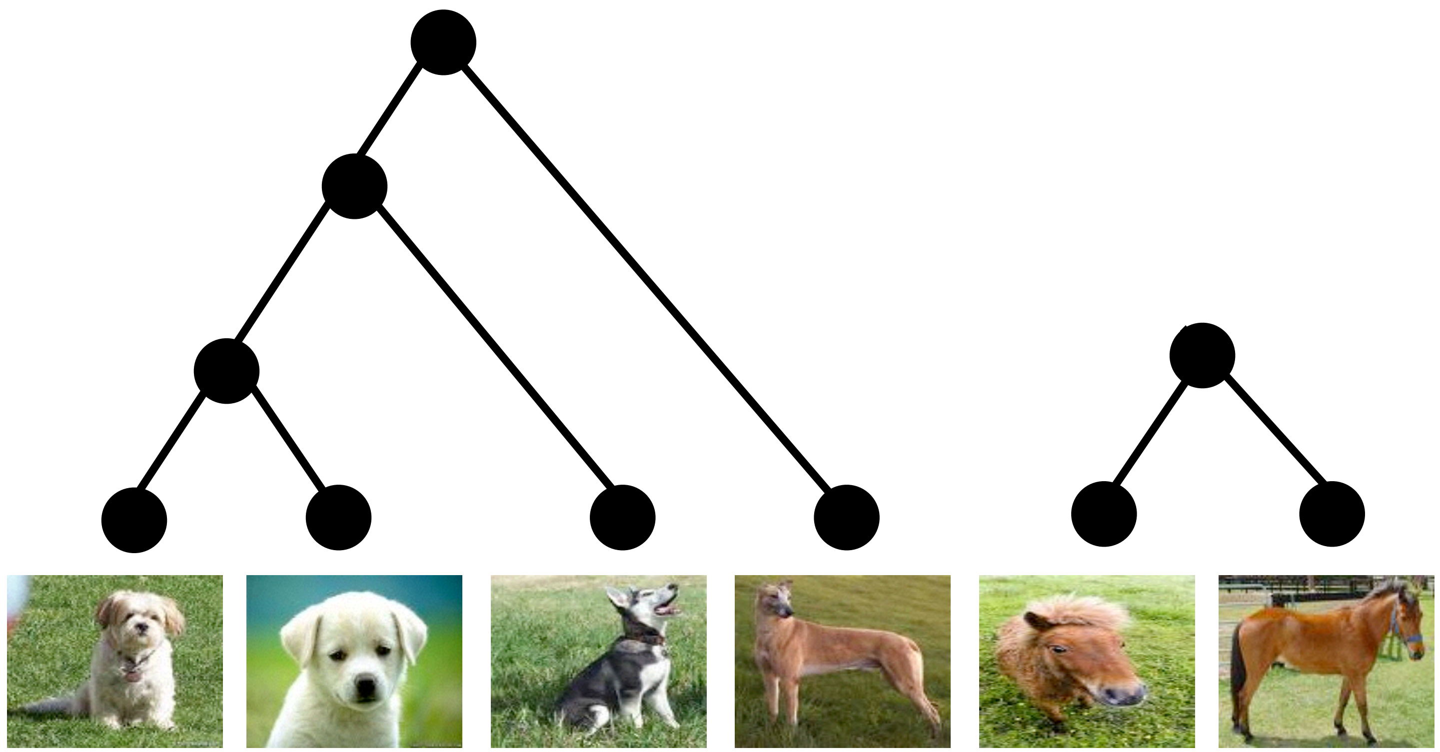


Agglomerative hierarchical clustering



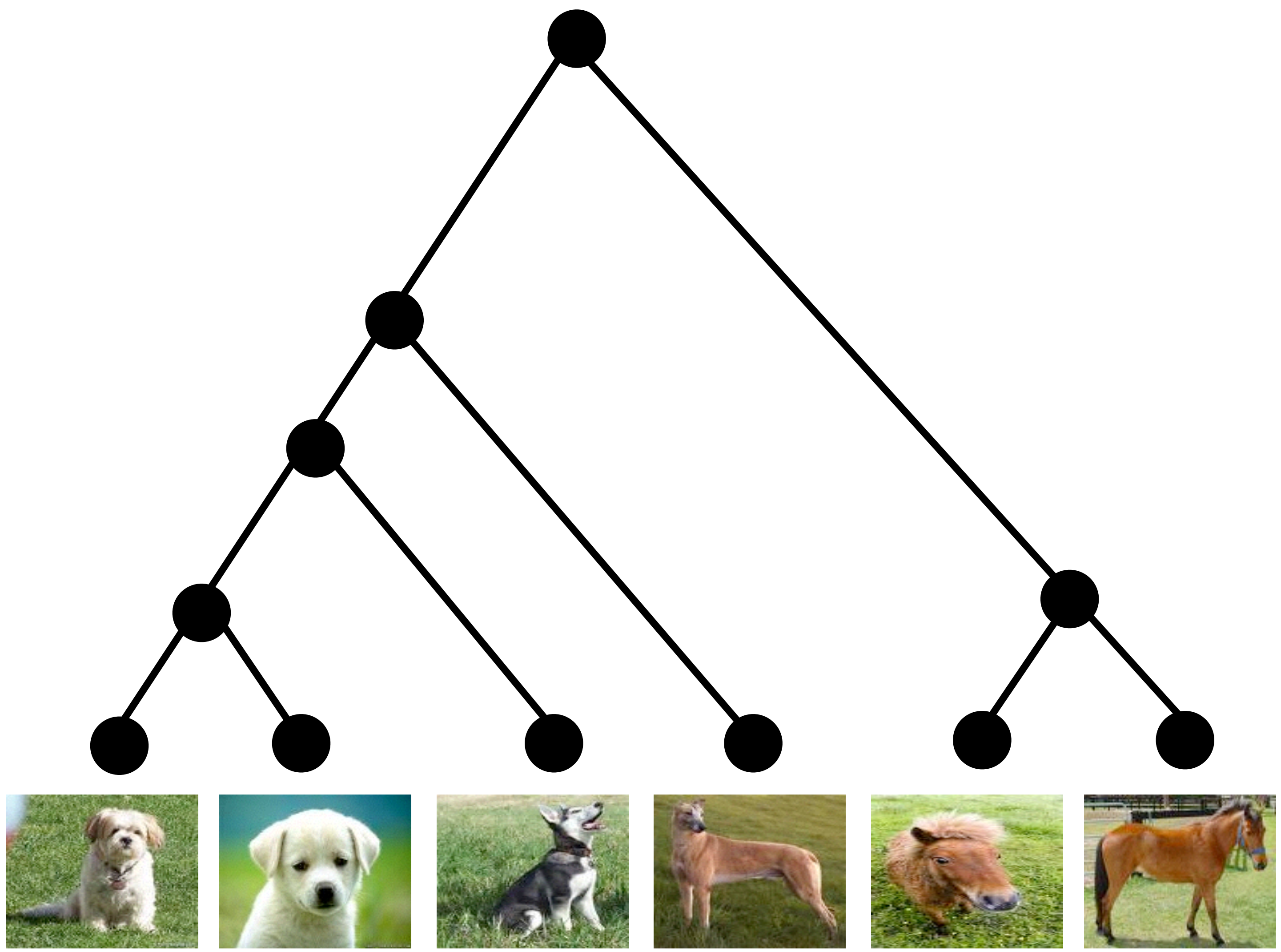
Agglomerative hierarchical clustering

Repeat



Agglomerative hierarchical clustering

Keep repeating until you have the tree



Notation

- Two clusters…
  - Cluster $\mathbf{A}$ contains objects $\mathbf{A} = (a_1, a_2, a_3)$
  - Cluster $\mathbf{B}$ contains objects $\mathbf{B} = (b_1, b_2, b_3)$
  - Let $\mathrm{d}(\mathbf{A}, \mathbf{B})$ be the distance between cluster $\mathbf{A}$ and cluster $\mathbf{B}$
  - Let $\mathrm{d}(a, b)$ be the (known) dissimilarity between item a and item b

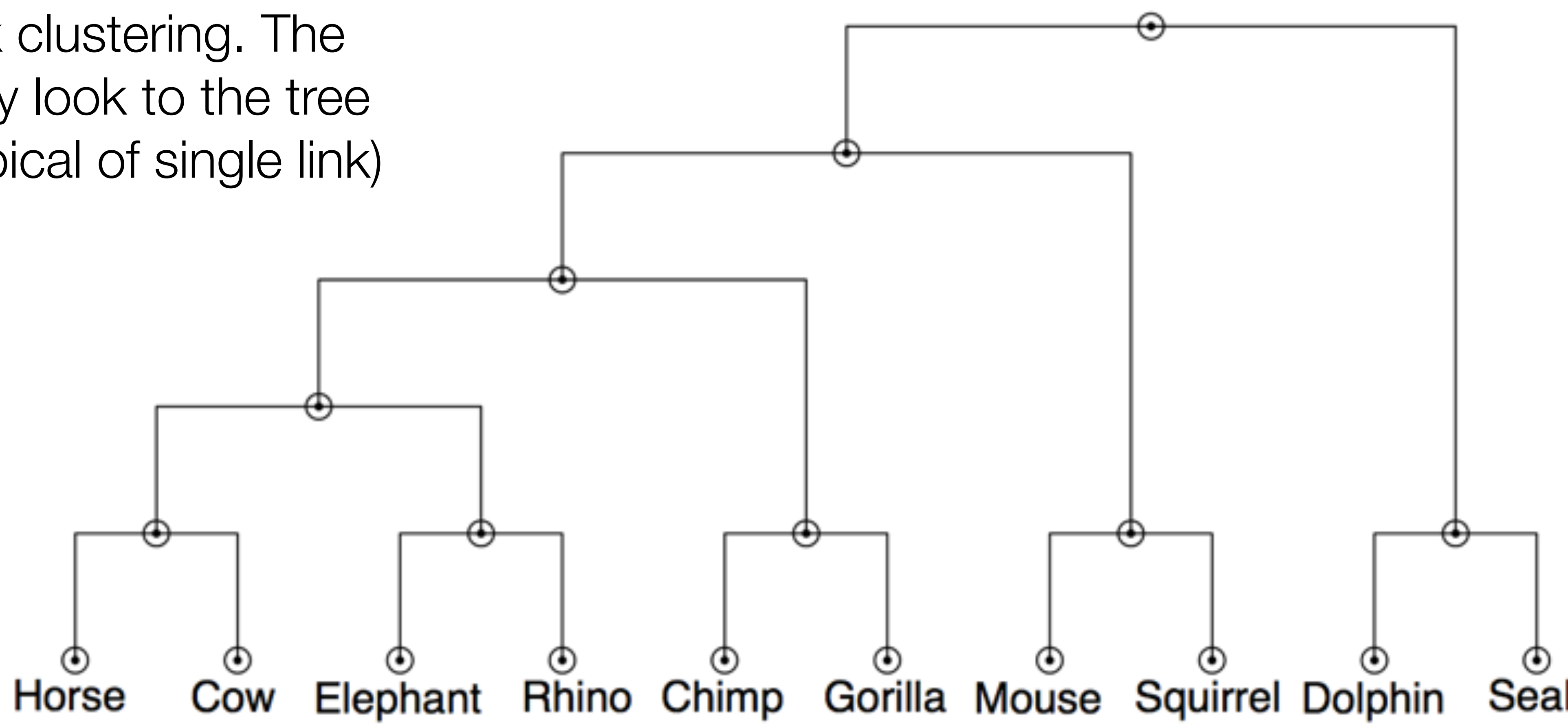
Measuring distance between clusters?

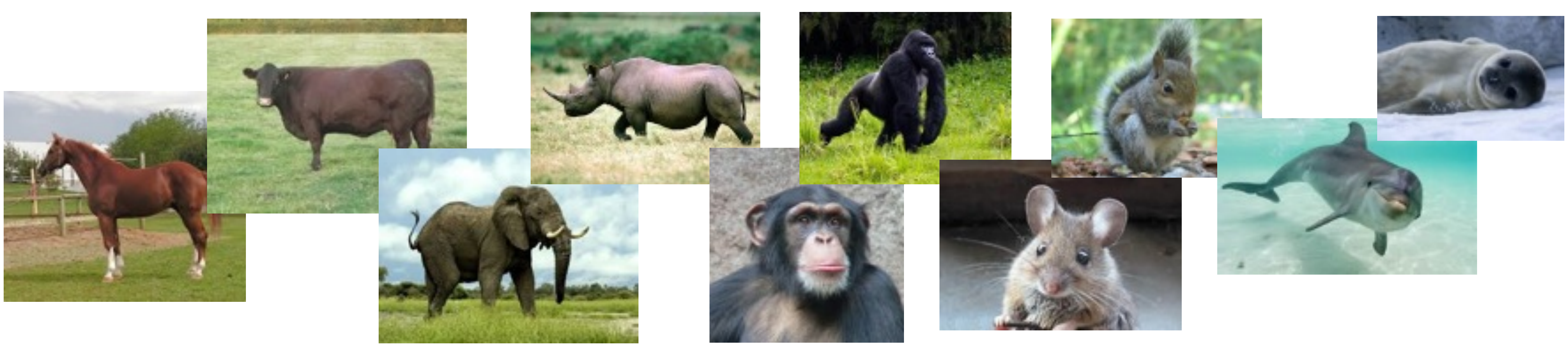
- Two clusters…
    - Cluster $\mathbf{A}$ contains objects $\mathbf{A} = (a_1, a_2, a_3)$
    - Cluster $\mathbf{B}$ contains objects $\mathbf{B} = (b_1, b_2, b_3)$
    - Let $d(\mathbf{A}, \mathbf{B})$ be the distance between cluster $\mathbf{A}$ and cluster $\mathbf{B}$
    - Let $d(a, b)$ be the (known) dissimilarity between item a and item b

- Different "link" functions to define cluster distance $d(\mathbf{A}, \mathbf{B})$
    - Complete link: $d(\mathbf{A}, \mathbf{B}) = \max(\, d(a, b)\,)$ for a in $\mathbf{A}$, b in $\mathbf{B}$
    - Single link: $d(\mathbf{A}, \mathbf{B}) = \min(\, d(a, b)\,)$ for a in $\mathbf{A}$, b in $\mathbf{B}$
    - Average link: $d(\mathbf{A}, \mathbf{B}) = \text{mean}(\, d(a, b)\,)$ for a in $\mathbf{A}$, b in $\mathbf{B}$

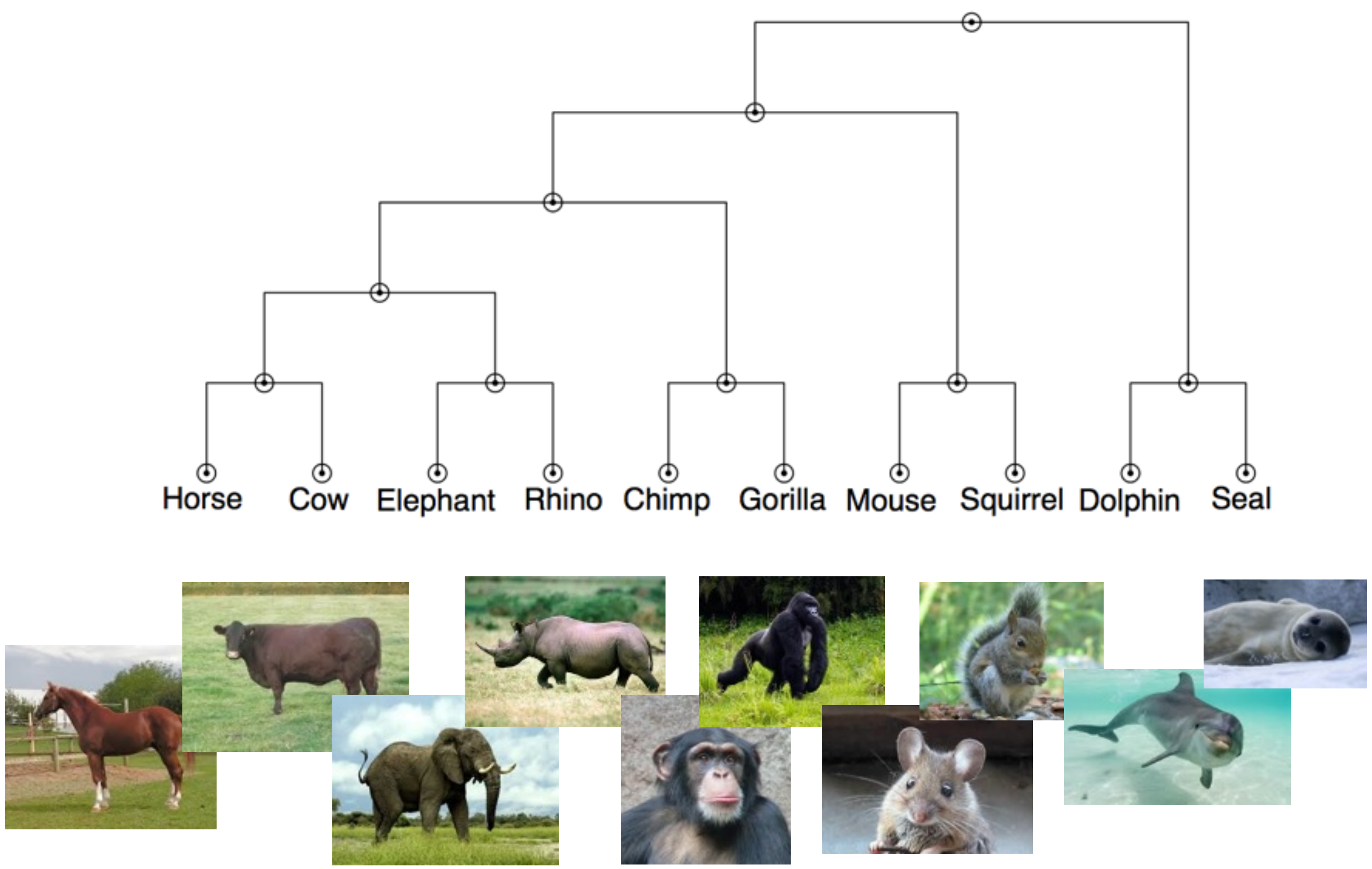
An empirically derived taxonomy

(I think this came from a single link clustering. The long stringy look to the tree is pretty typical of single link)



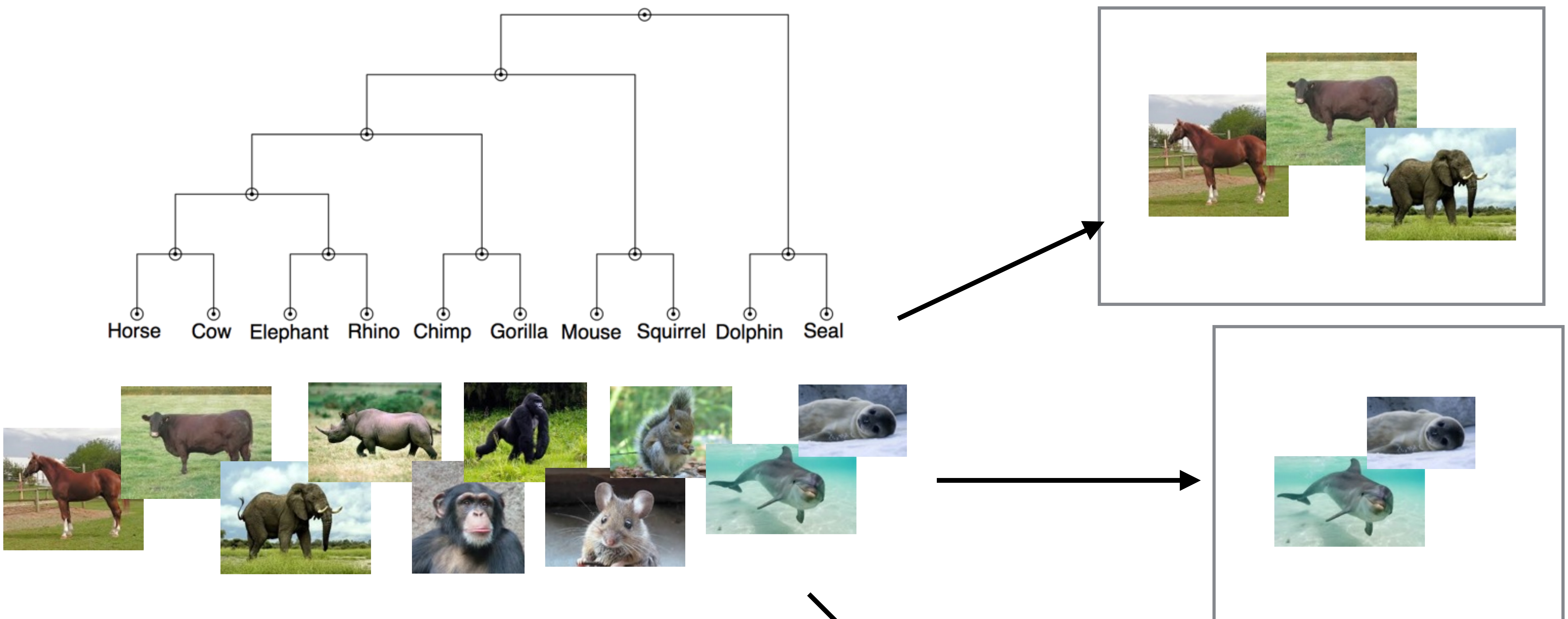


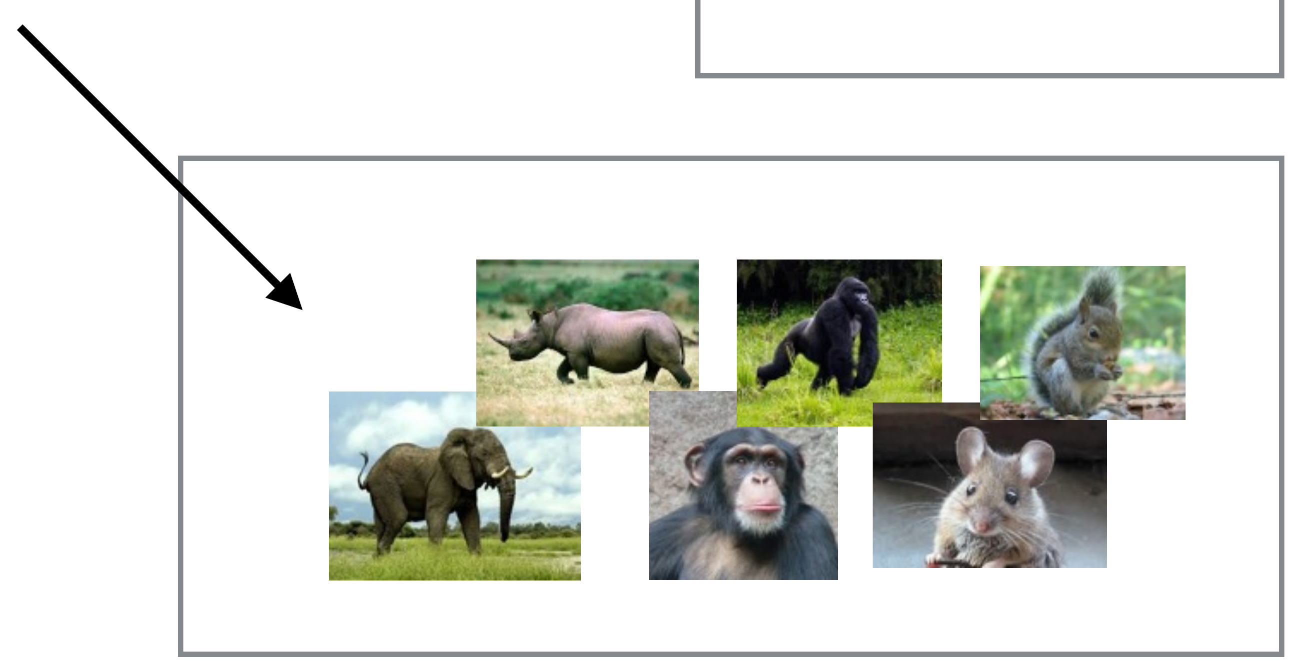
What shall our hypothesis space be?



This is our <u>structure</u>

What shall our hypothesis space be?

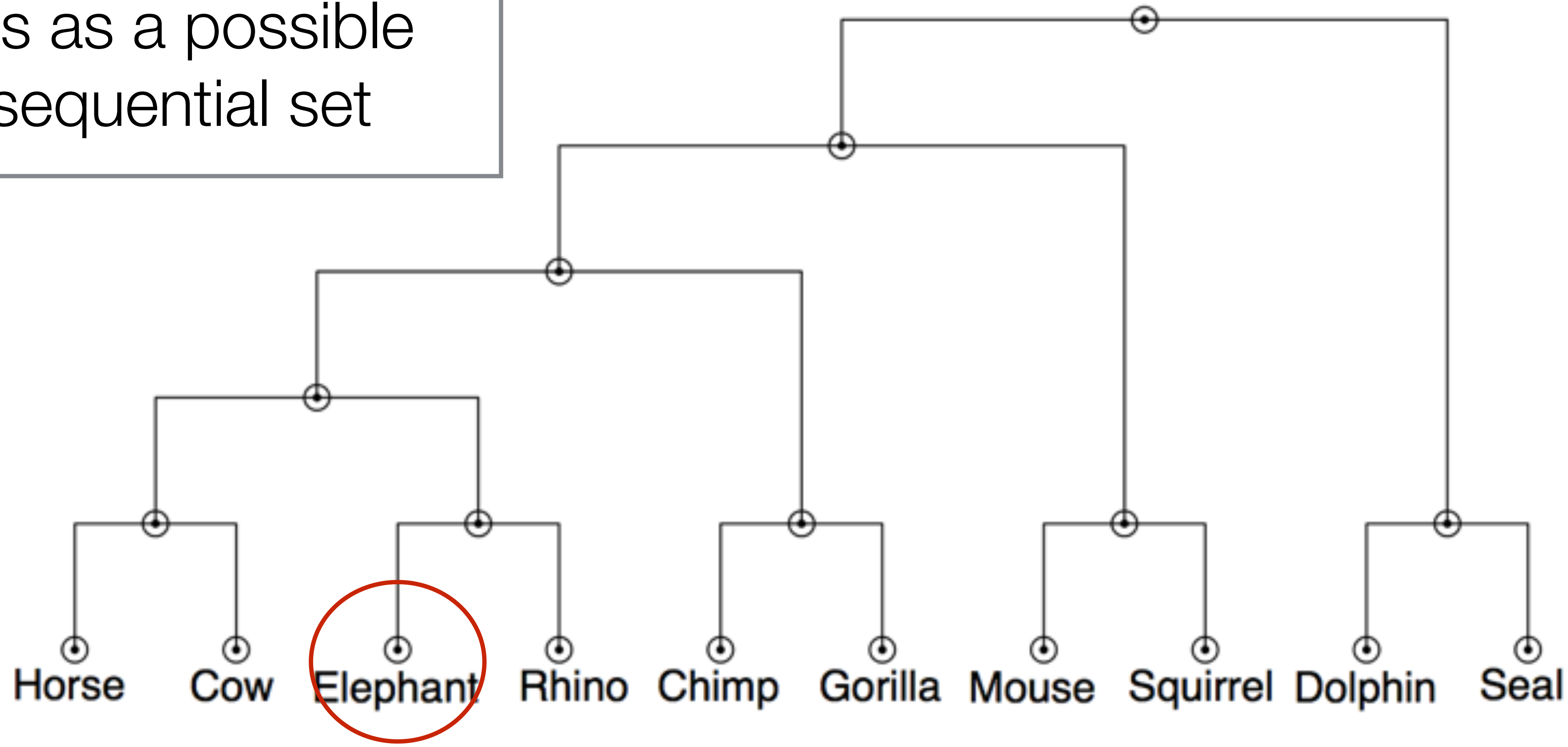




We need to use this structure to define a collection of possible “consequential sets”

First order hypotheses, H_1

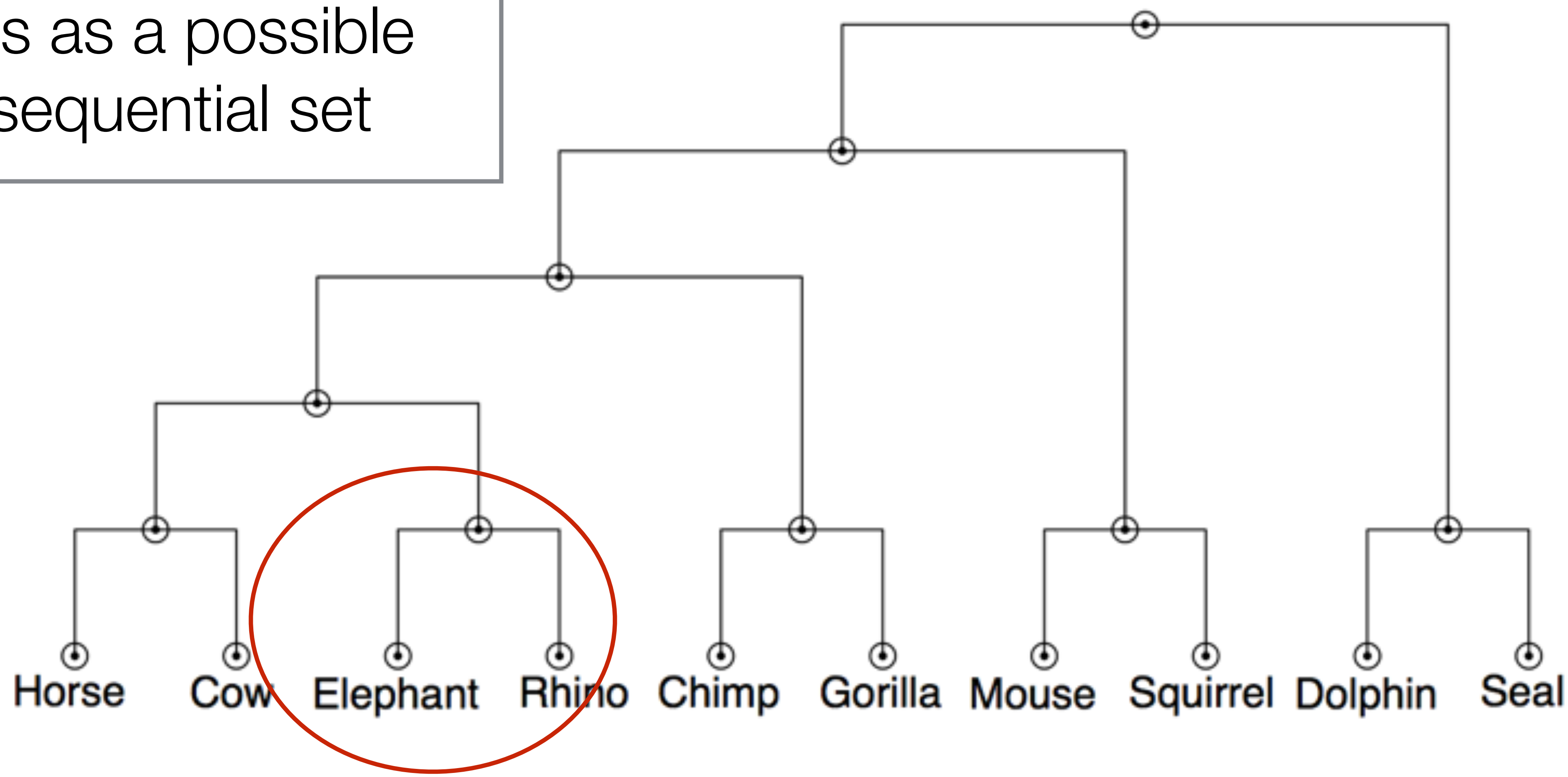
Every cluster in the tree counts as a possible consequential set

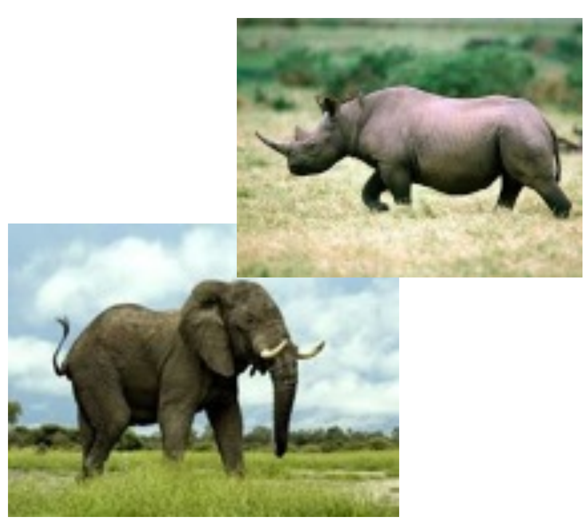


This "set" defines the hypothesis that we have a property that is unique to elephants

First order hypotheses, H_1

Every cluster in the tree counts as a possible consequential set

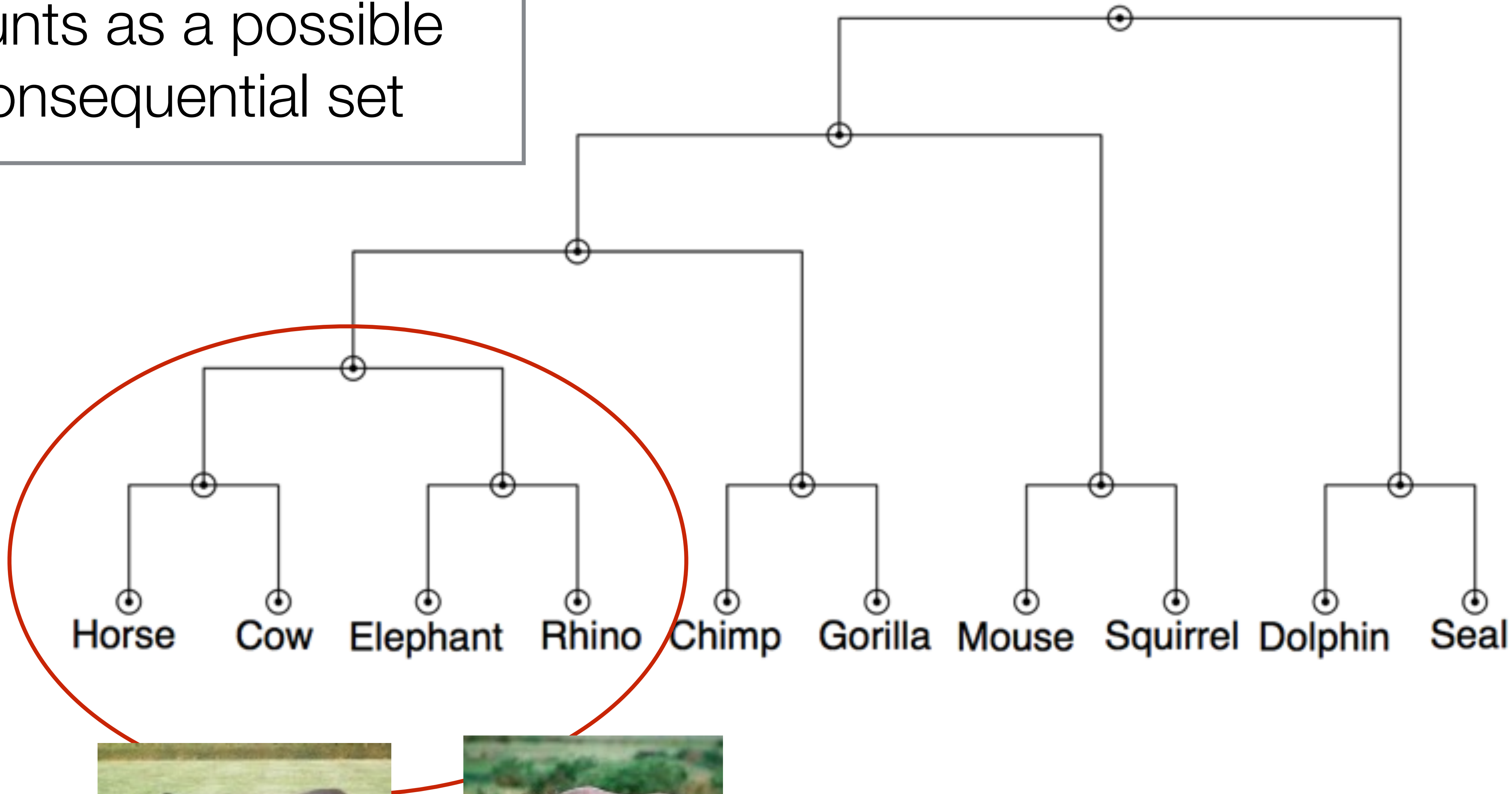


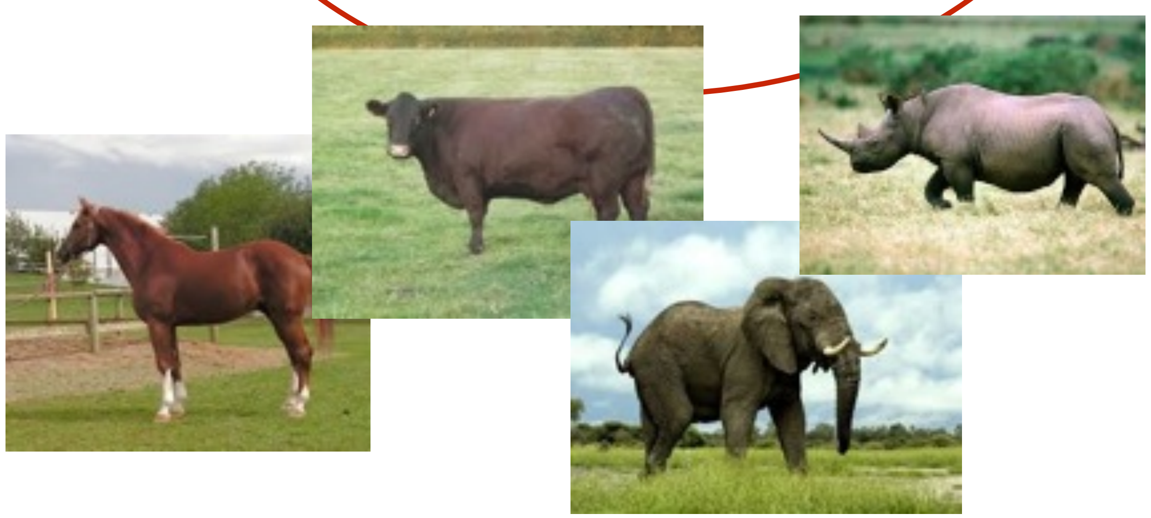


We might also hypothesise that some properties are shared by elephants and rhinos

First order hypotheses, H_1

Every cluster in the tree counts as a possible consequential set

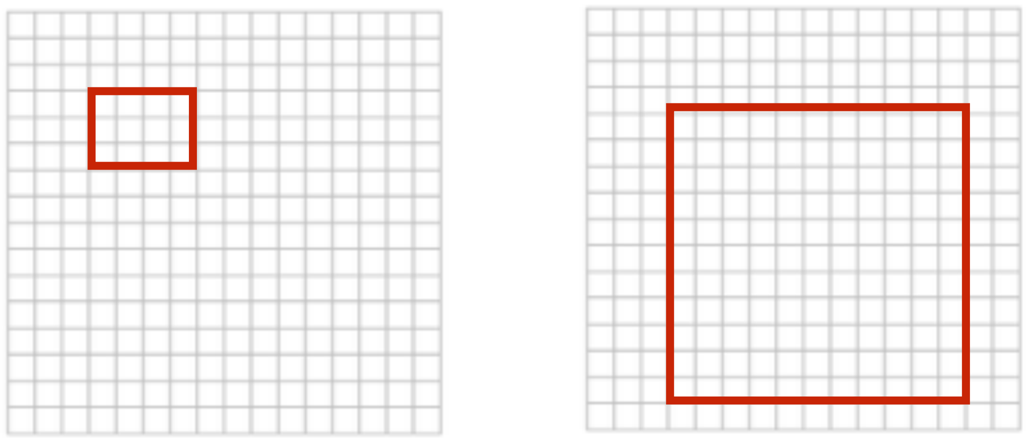




All the large herbivores is also a reasonable consequential set

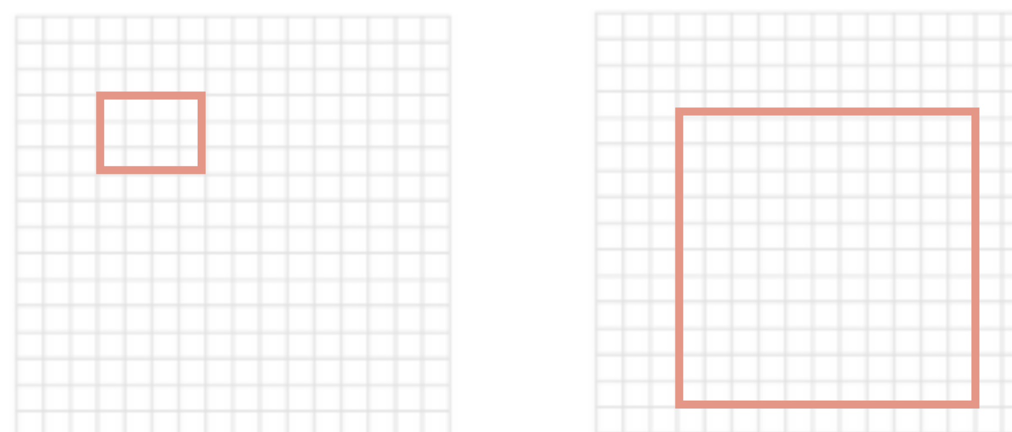
Let’s be a little more precise about what we mean by “first order”

“first order” hypotheses cover a “single” simple entity with respect to the structure

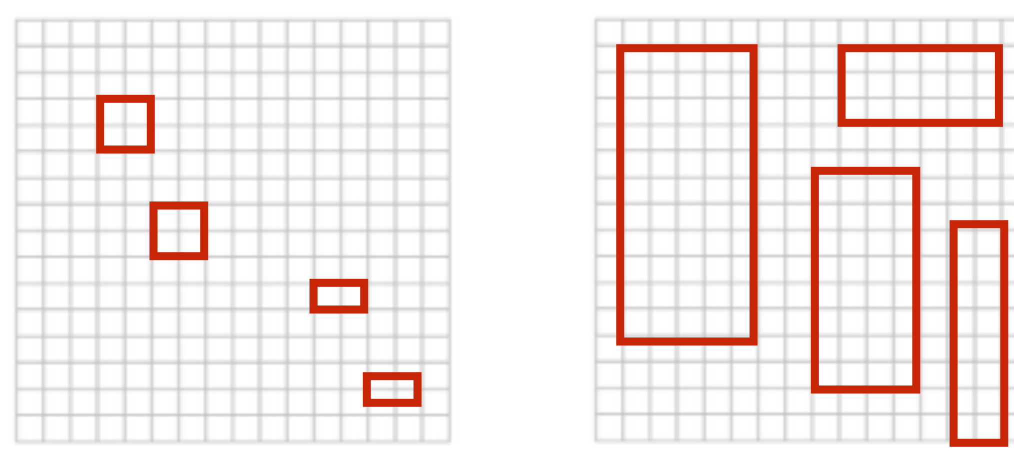


We've encountered hypothesis spaces with higher order hypotheses too

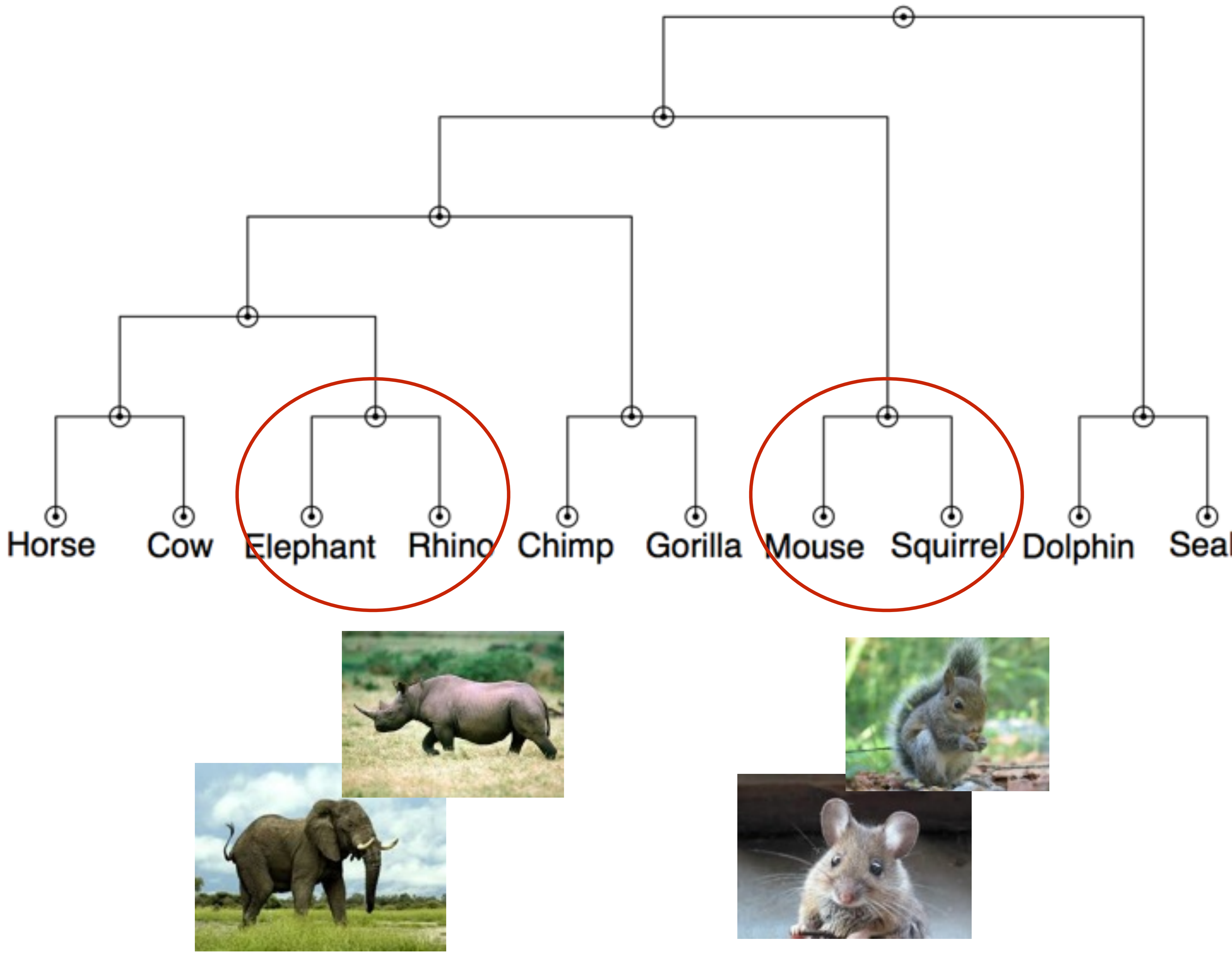
"first order" hypotheses cover a "single" simple entity with respect to the structure



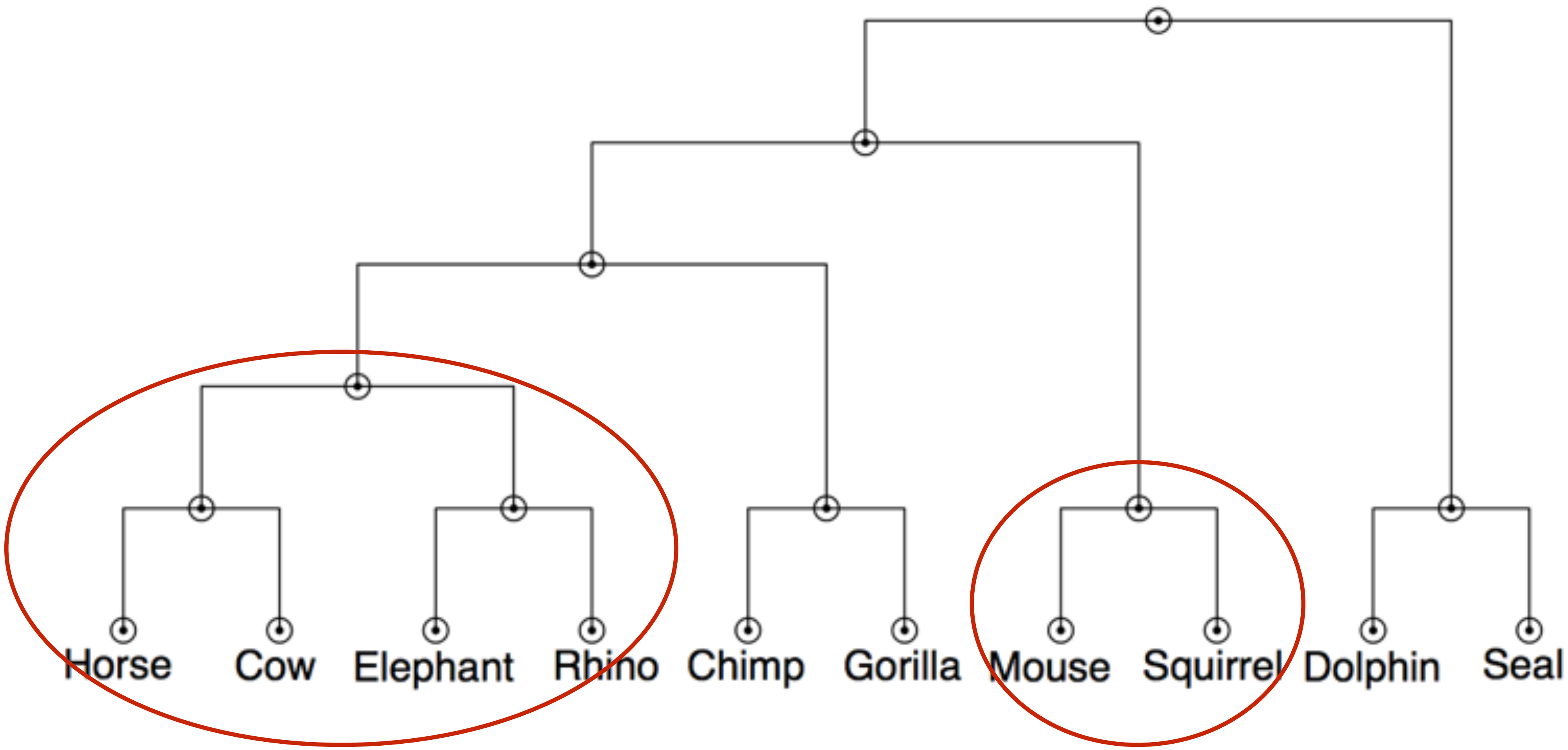
Higher order hypotheses that are built from multiple such entities



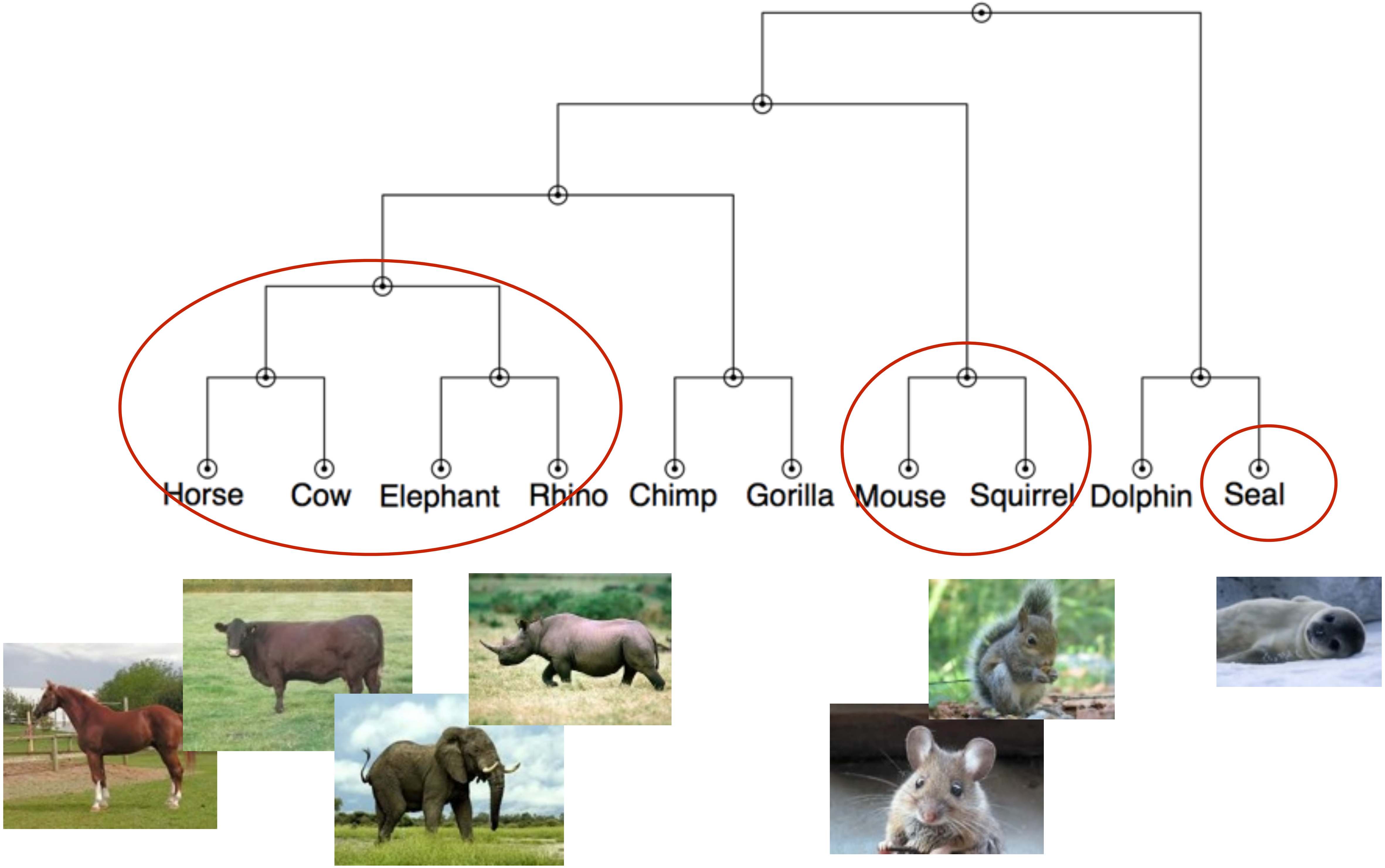
Second order hypotheses, H_2



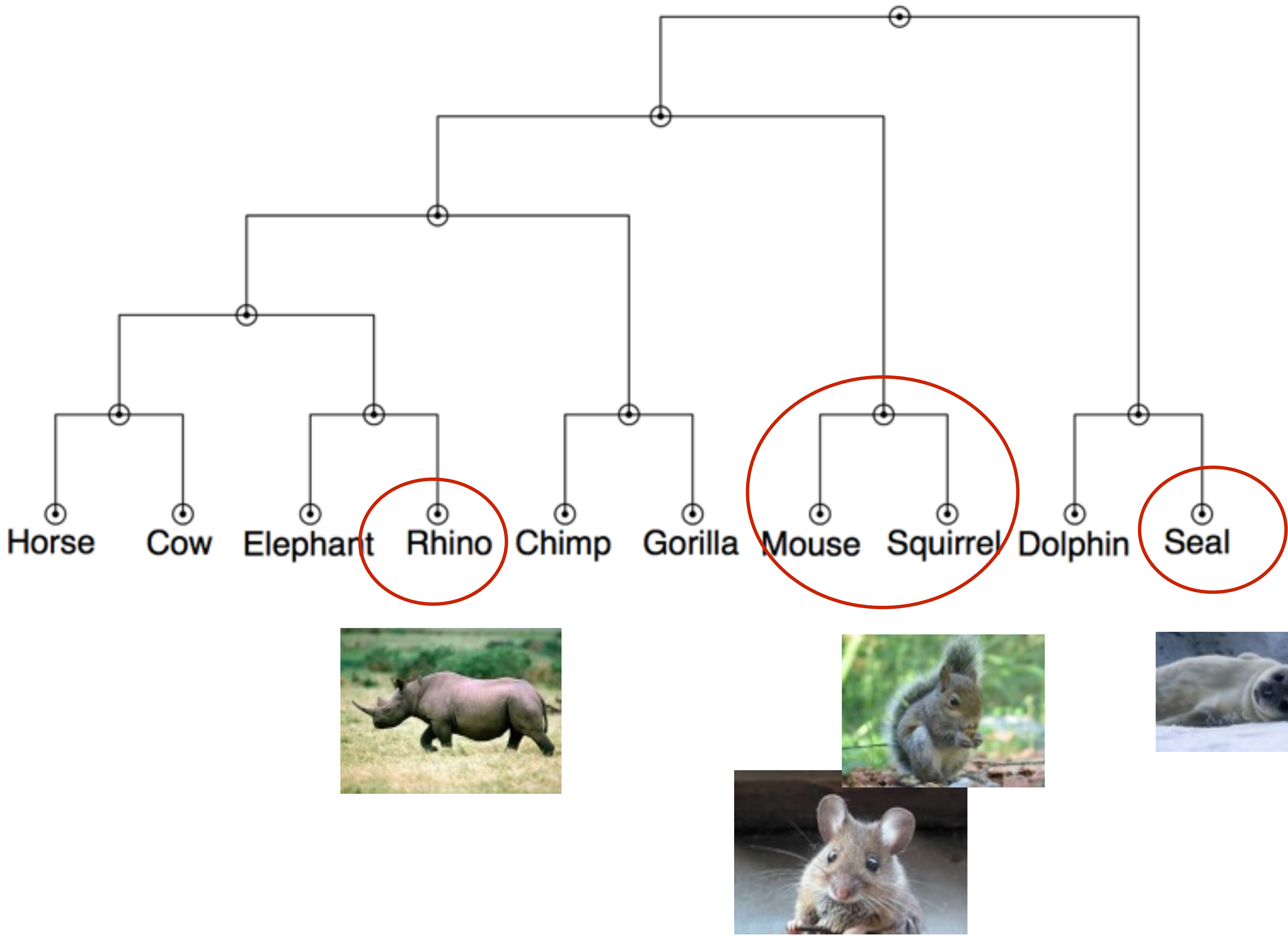
Second order hypotheses, H_2



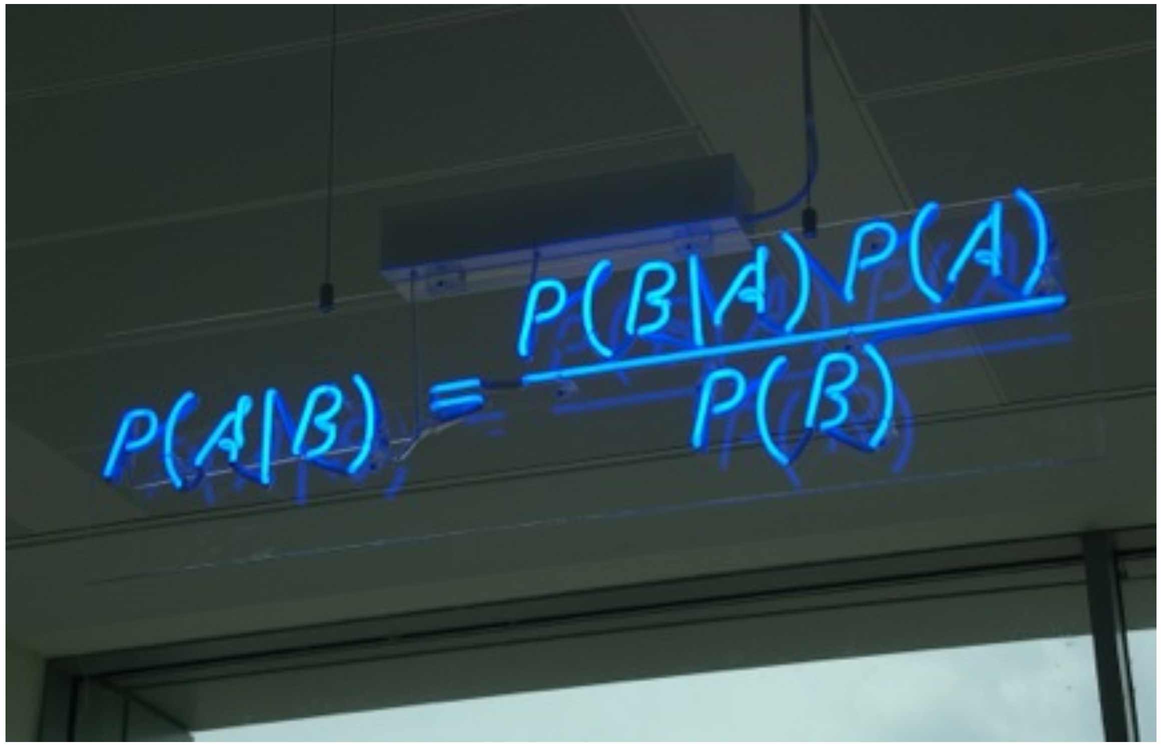
Third order hypotheses, H_3



Third order hypotheses, H_3



Now let's make a Bayesian model



To do this we'll need a prior over hypotheses, P(h), and each hypothesis must define a distribution over possible observations P(x|h), i.e. the likelihood.

Use the prior to enforce simplicity

Prior

$$P(h) \propto \frac{1}{\phi^k}$$

Number of clusters *k* combined in the current hypothesis (i.e., the order of the hypothesis)

Scaling factor (ϕ>1) determining the extent of the simplicity bias (i'll use ϕ=20)

Use the likelihood to enforce data fit

Prior

$$P(h) \propto \frac{1}{\phi^k}$$

Likelihood

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

Our “usual” likelihood: every object within the consequential set is equally likely to be “observed” to have the property

Generalisation probability

Object X_1 has property P

Object X_2 has property P

…

Object X_k has property P

↓

Object C has property P

Compute $Pr(C | X_1 \ldots X_k)$

The probability that the object in the conclusion has property P given that all the objects in the premises do

Generalisation probability

Object X_1 has property P

Object X_2 has property P

…

Object X_k has property P

↓

Object C has property P

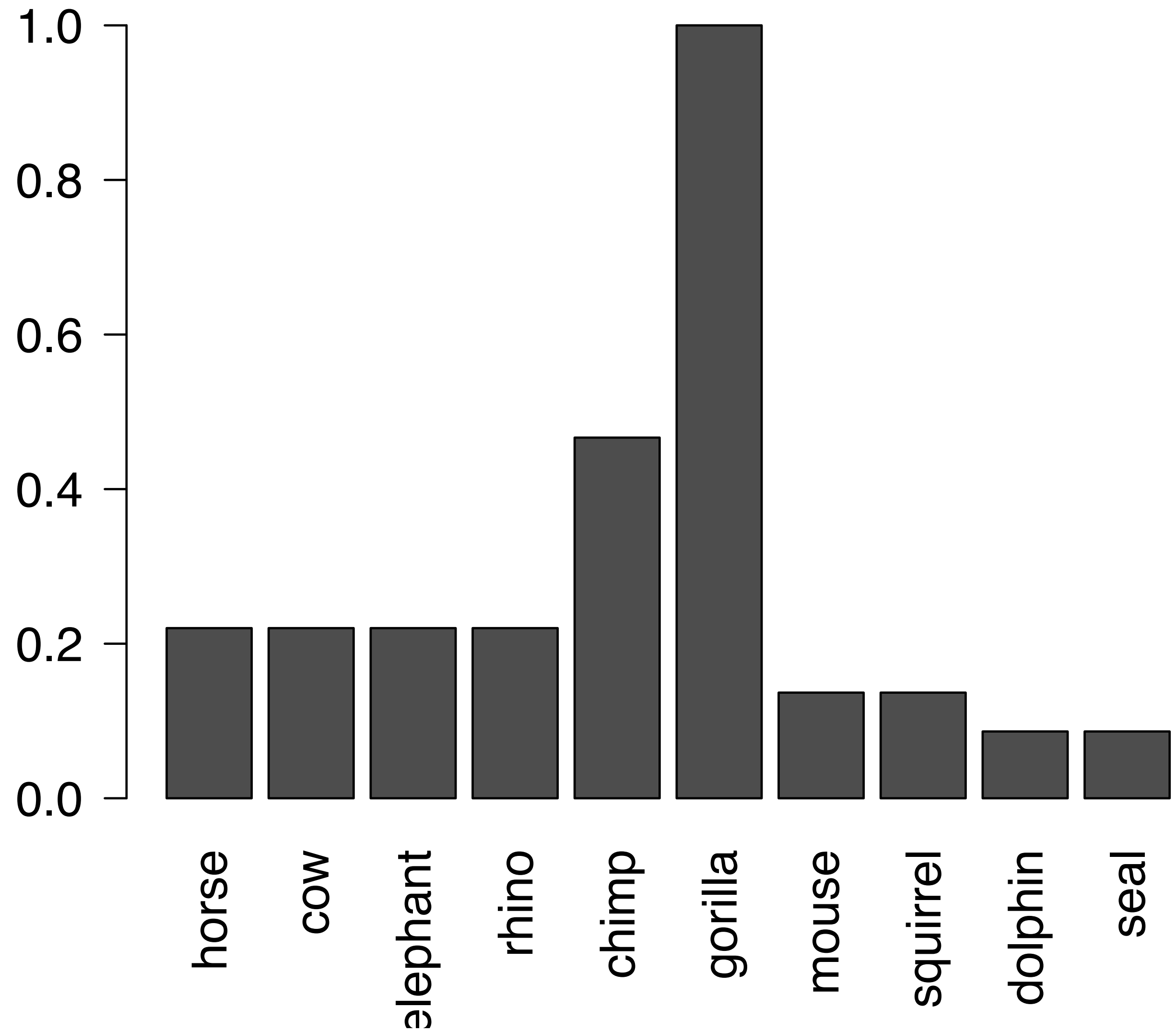
Compute $Pr(C \mid X_1 \ldots X_k)$

MATHS HERE

Code for the model
(animals.R)

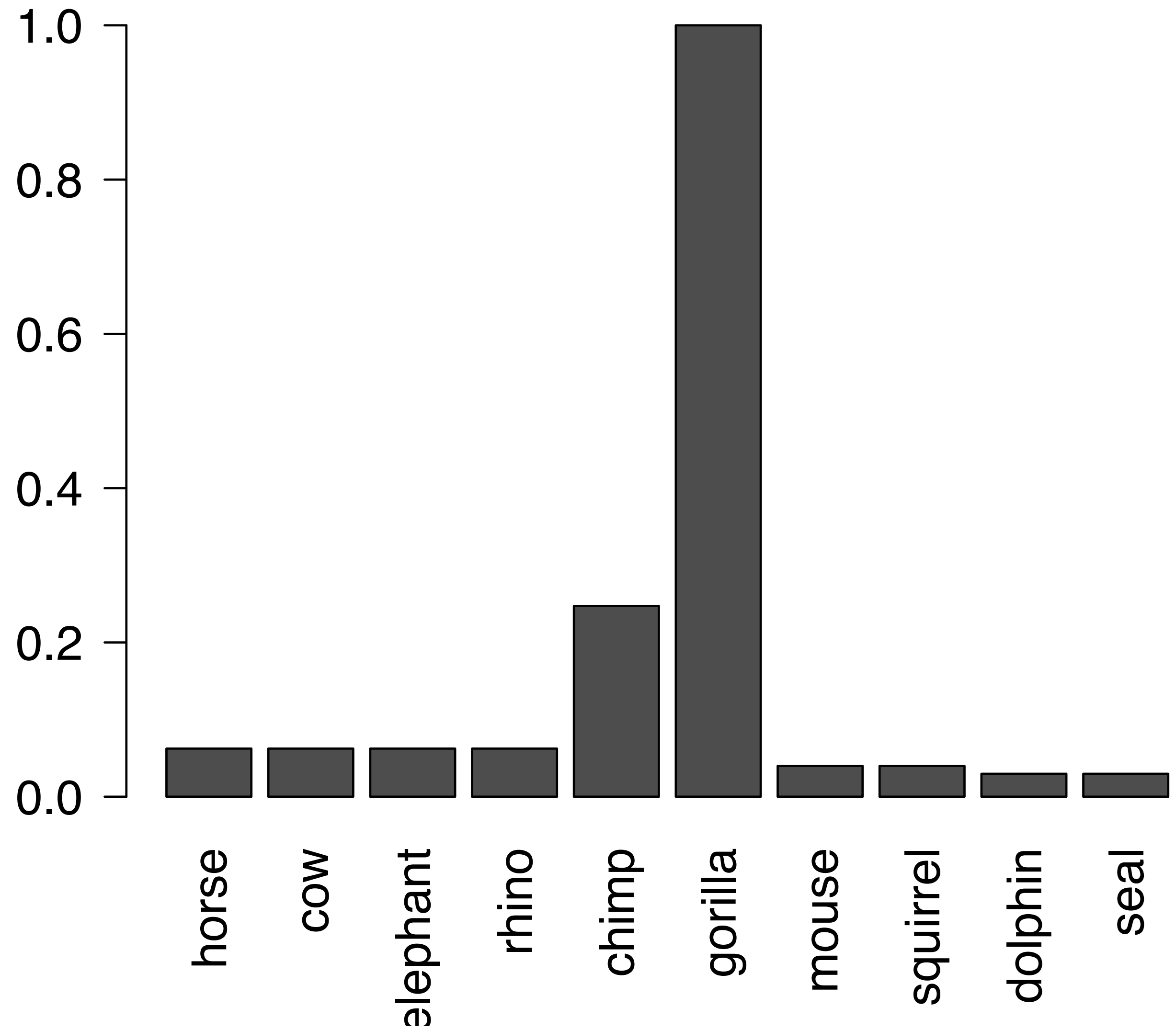
Illustrations of the model at work

Gorilla

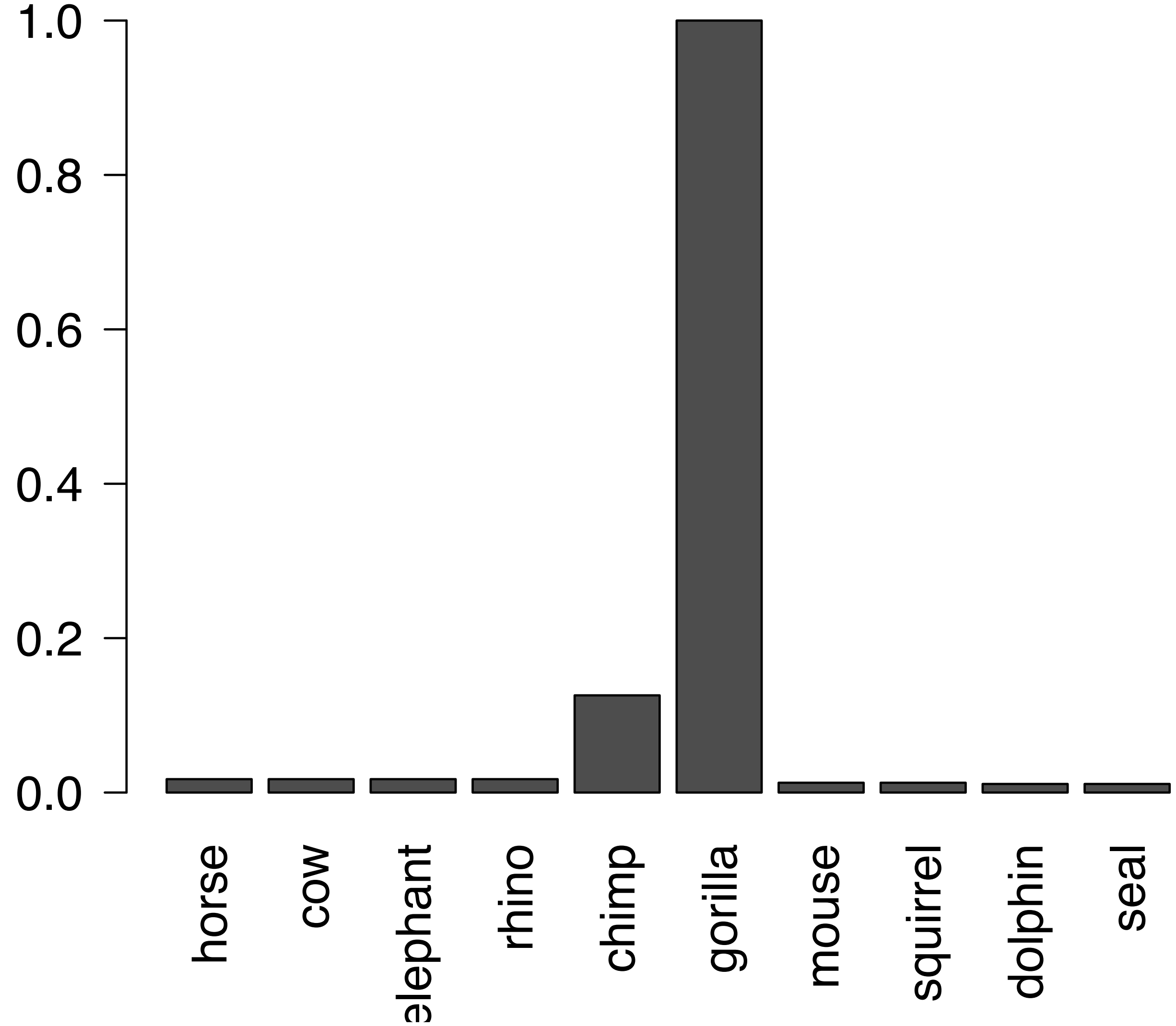


1.0
0.8
0.6
0.4
0.2
0.0
horse
cow
elephant
rhino
chimp
gorilla
mouse
squirrel
dolphin
seal

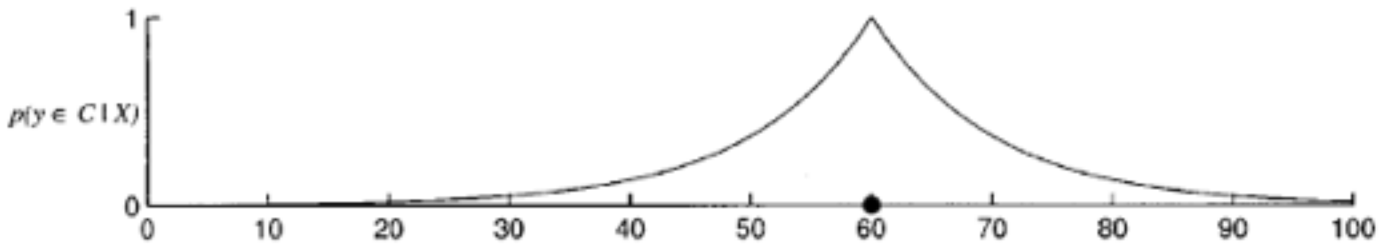
Gorilla, gorilla

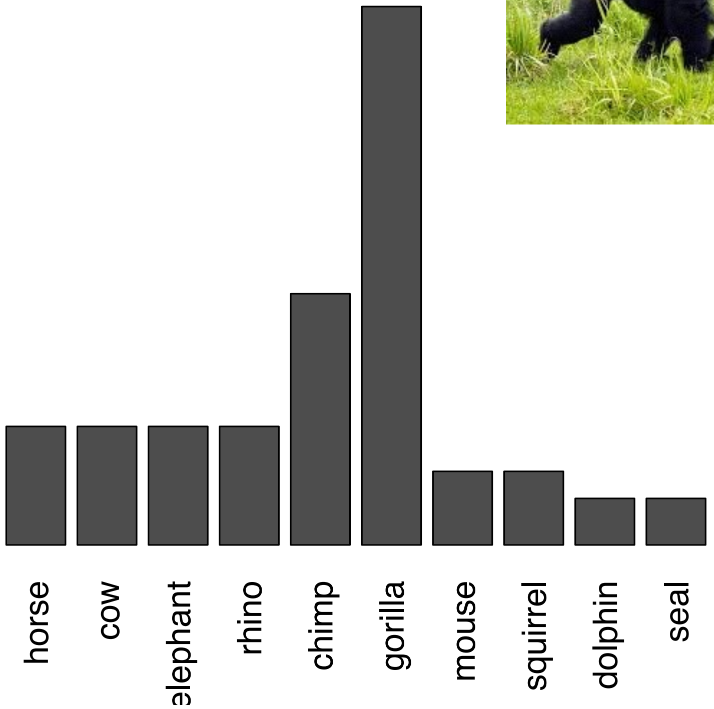


Gorilla, gorilla, gorilla

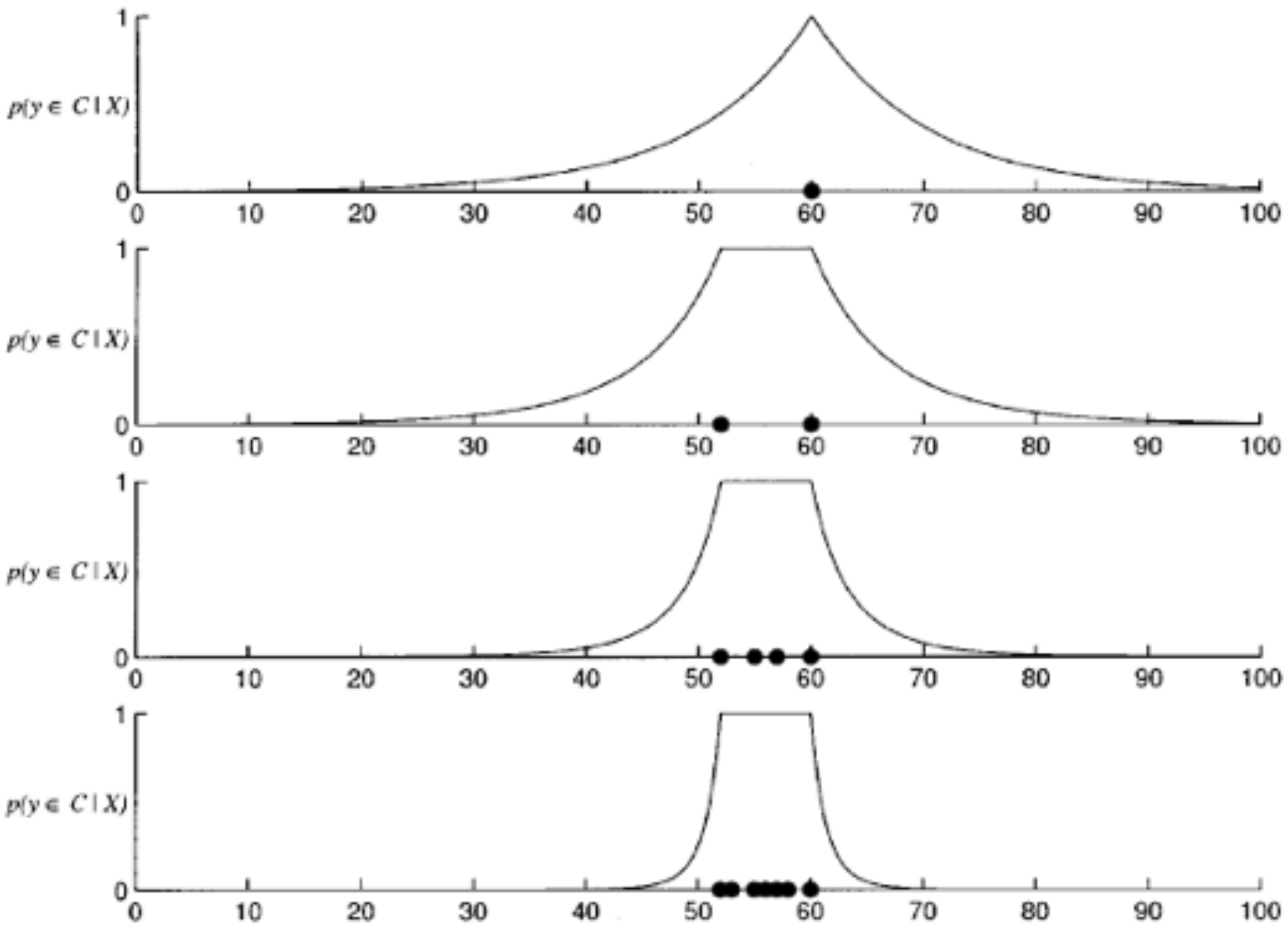


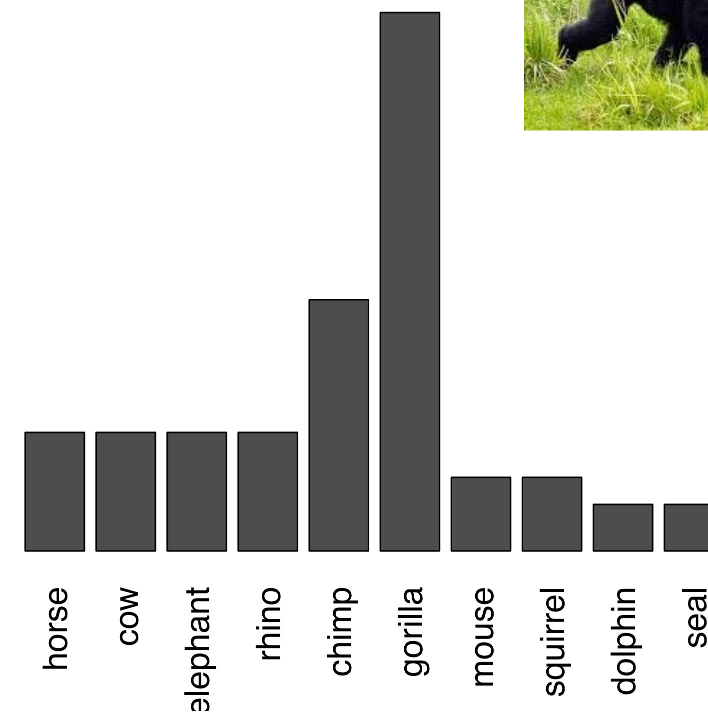
Link to last lecture

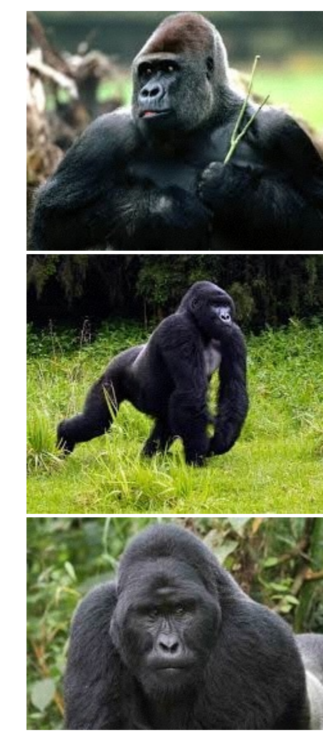


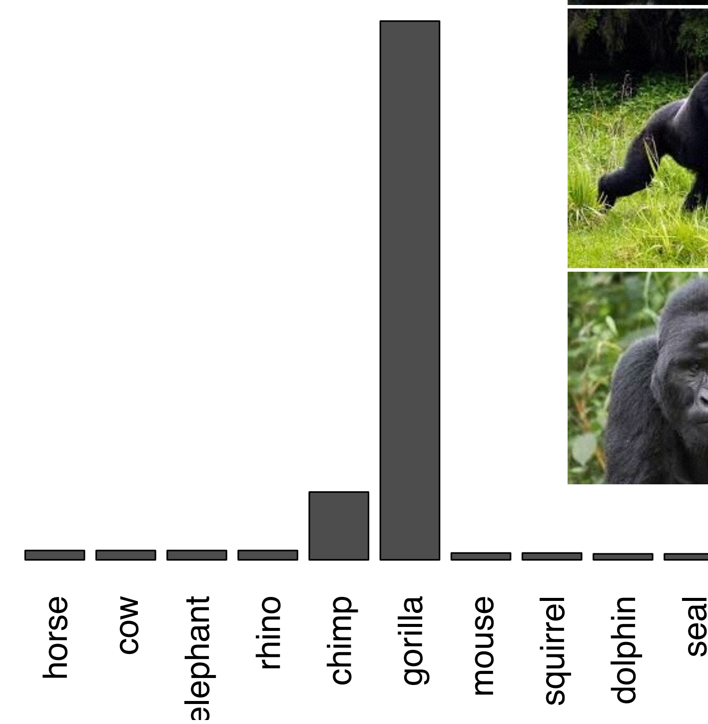


Link to last lecture

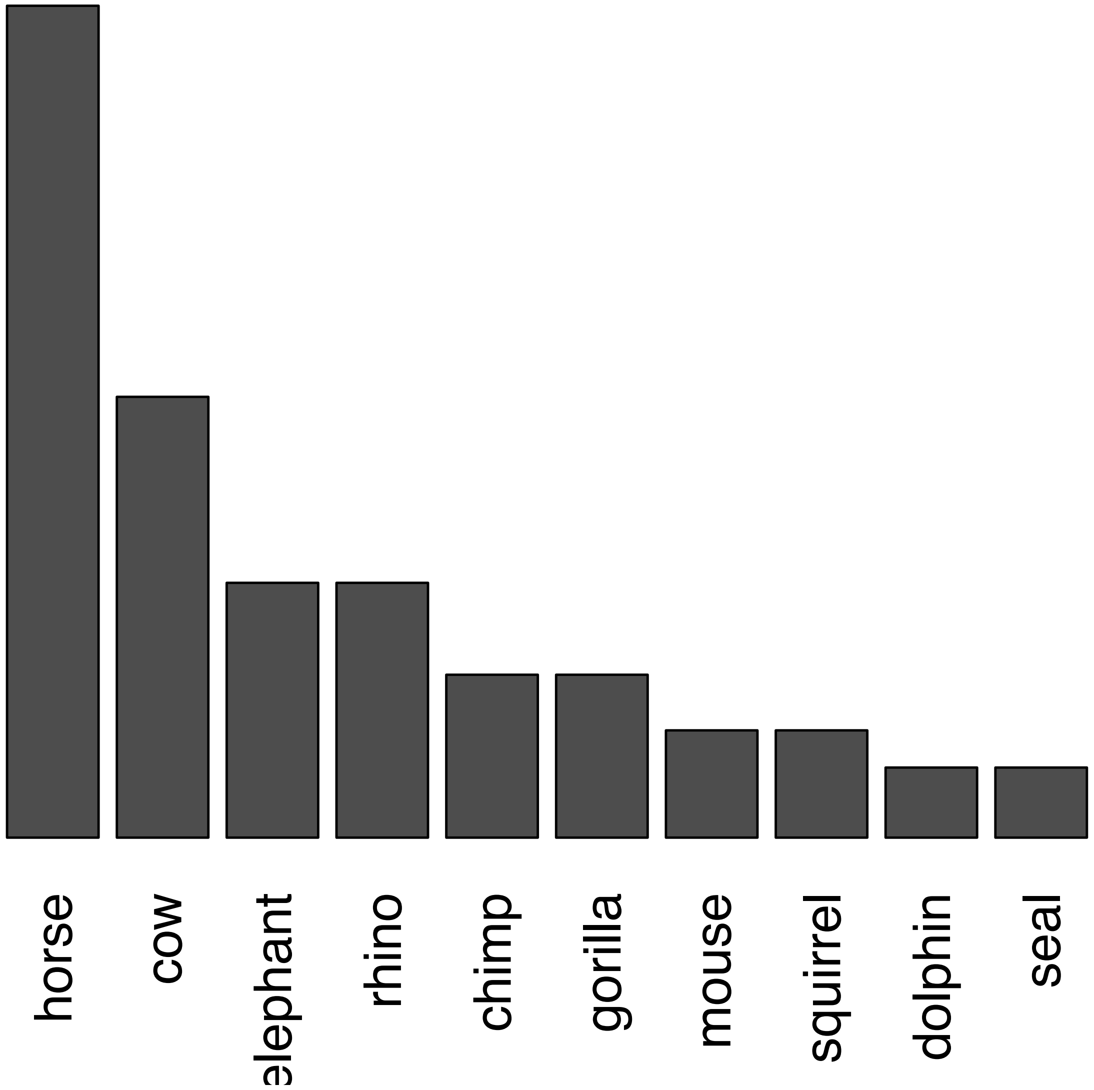






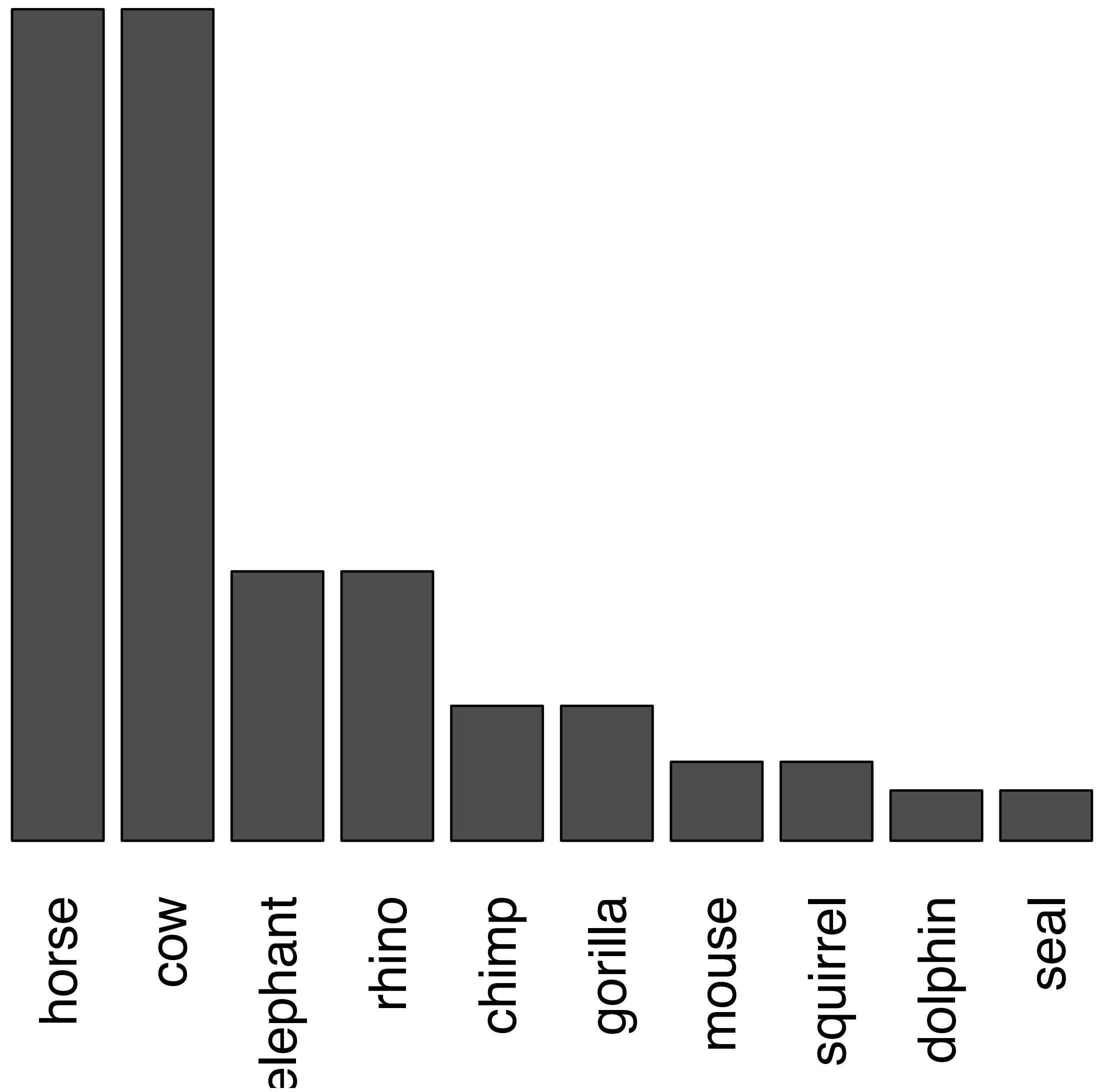


Horse



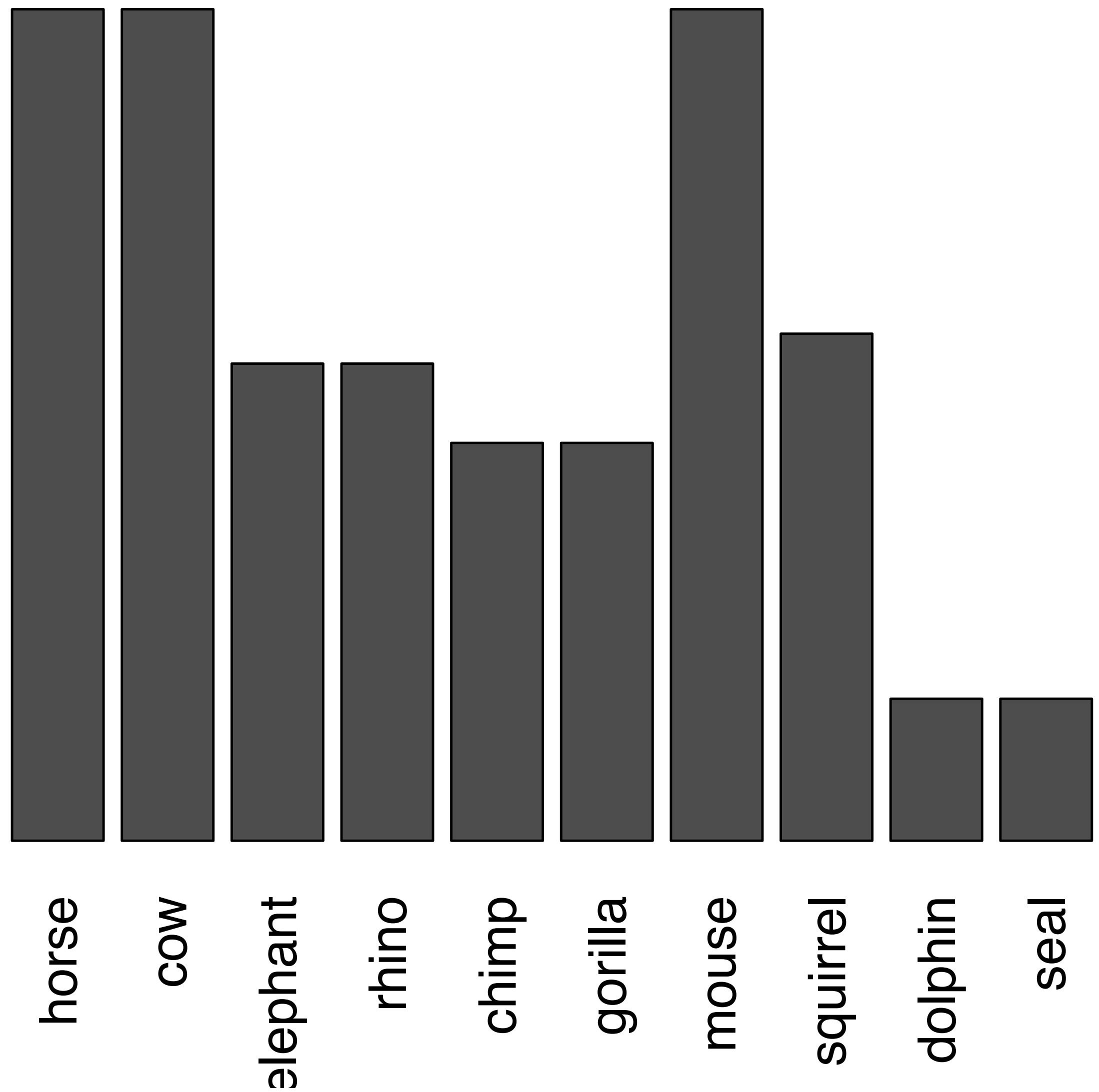
horse
cow
elephant
rhino
chimp
gorilla
mouse
squirrel
dolphin
seal

Horse, cow

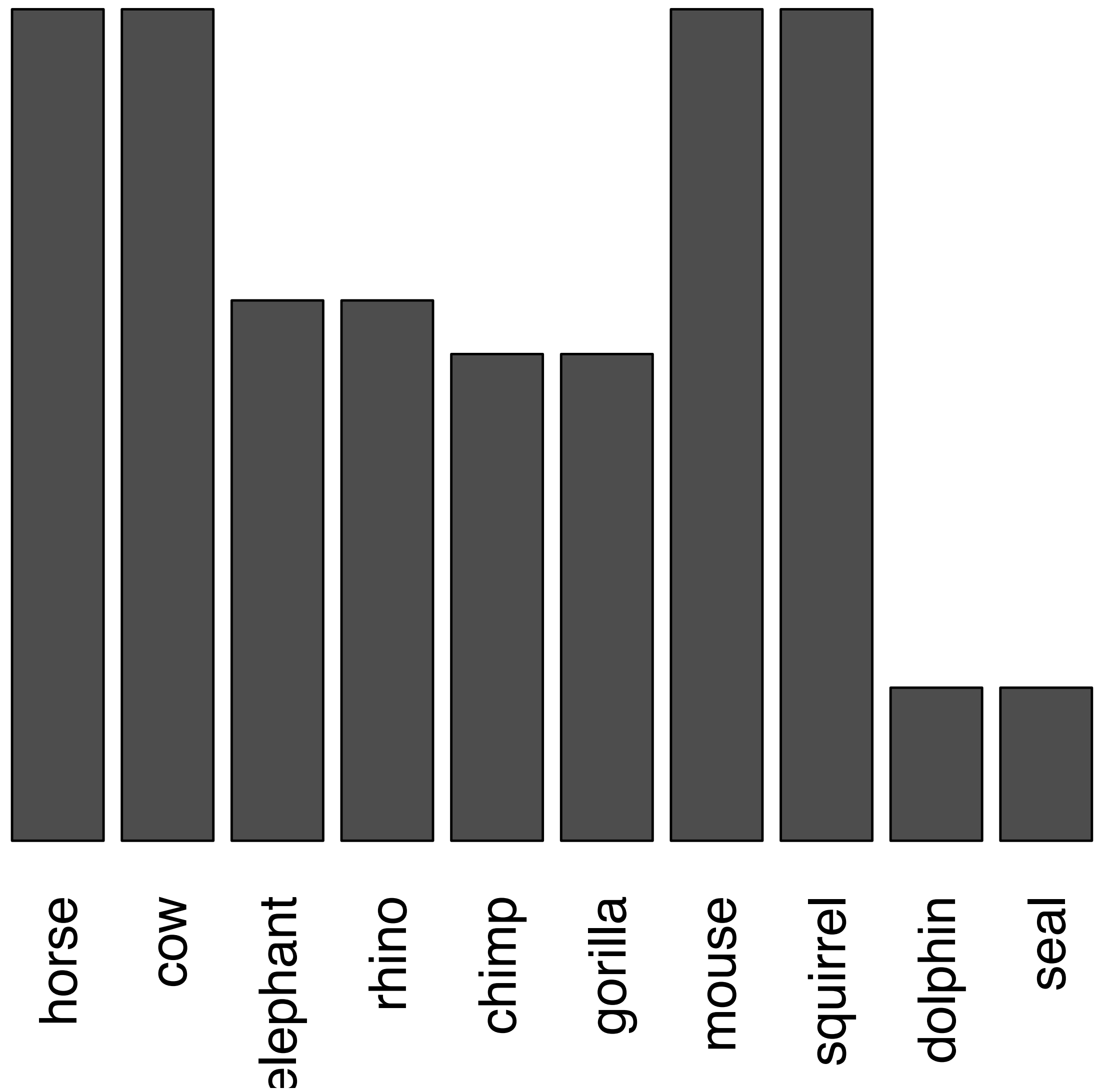


horse
cow
elephant
rhino
chimp
gorilla
mouse
squirrel
dolphin
seal

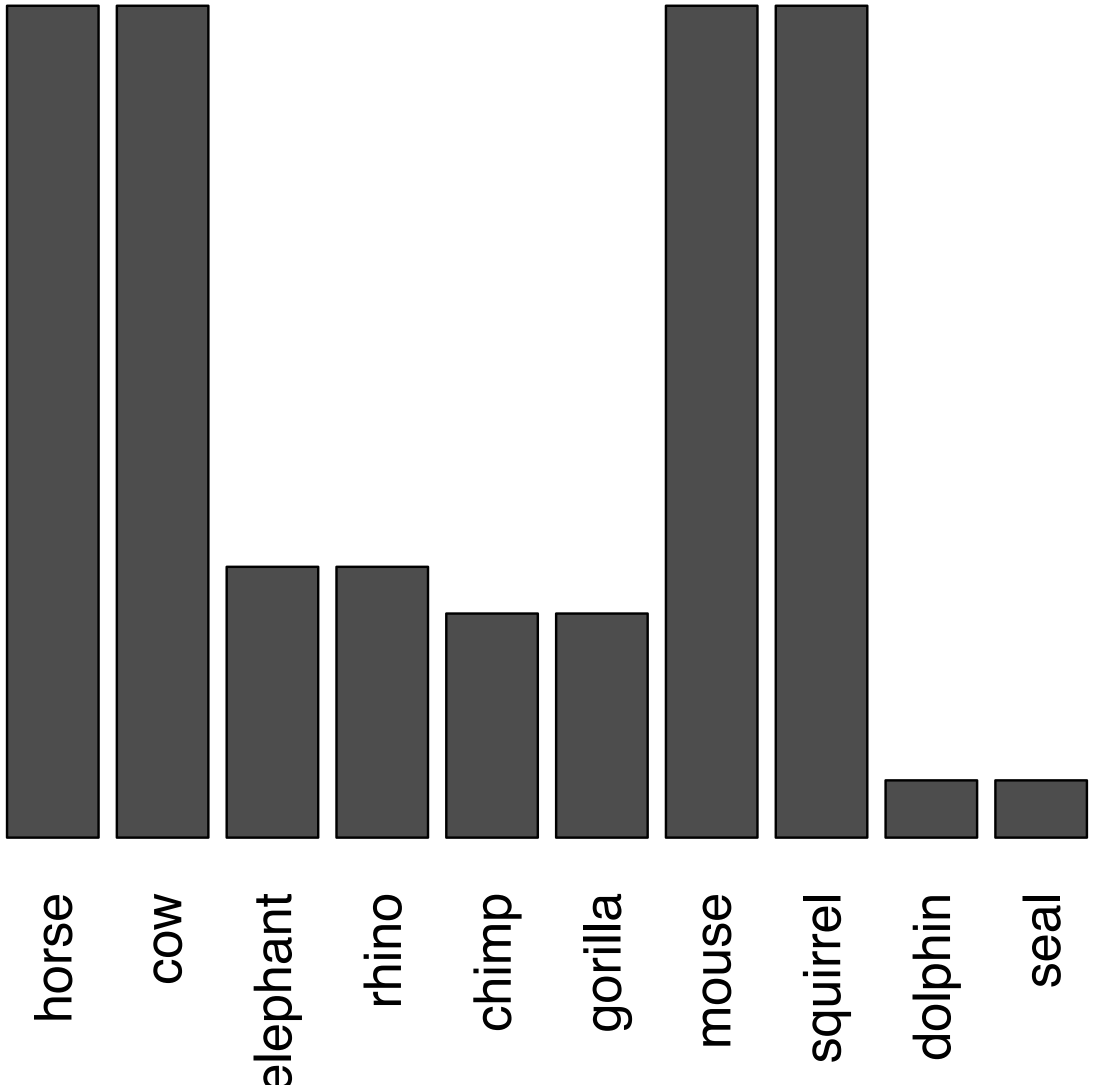
Horse, cow, mouse

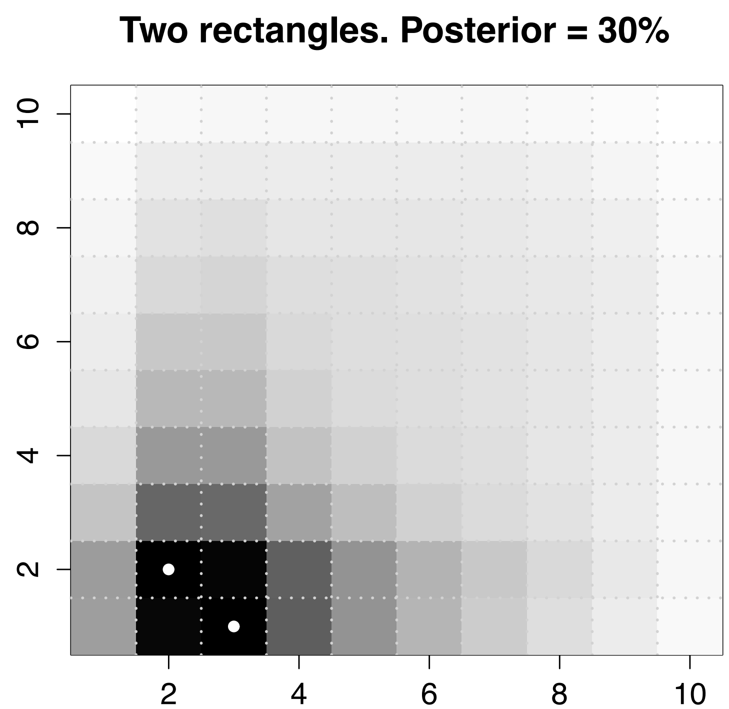


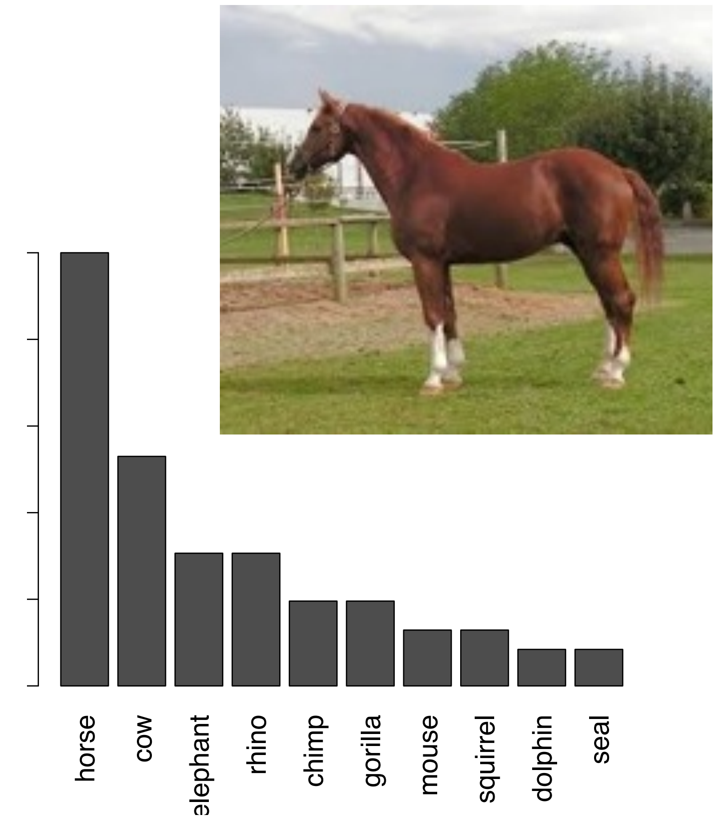
Horse, cow, mouse, squirrel



Horse, cow, mouse, squirrel, horse, squirrel

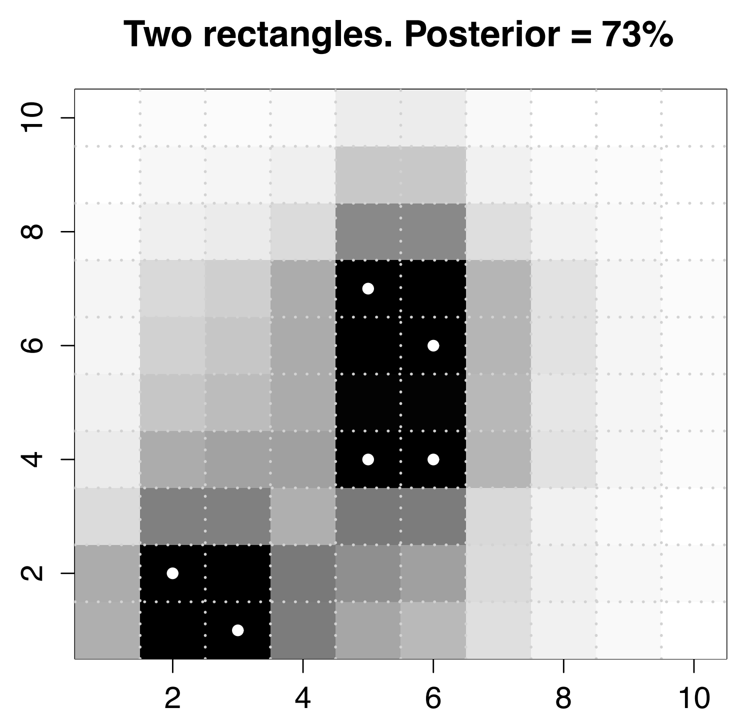


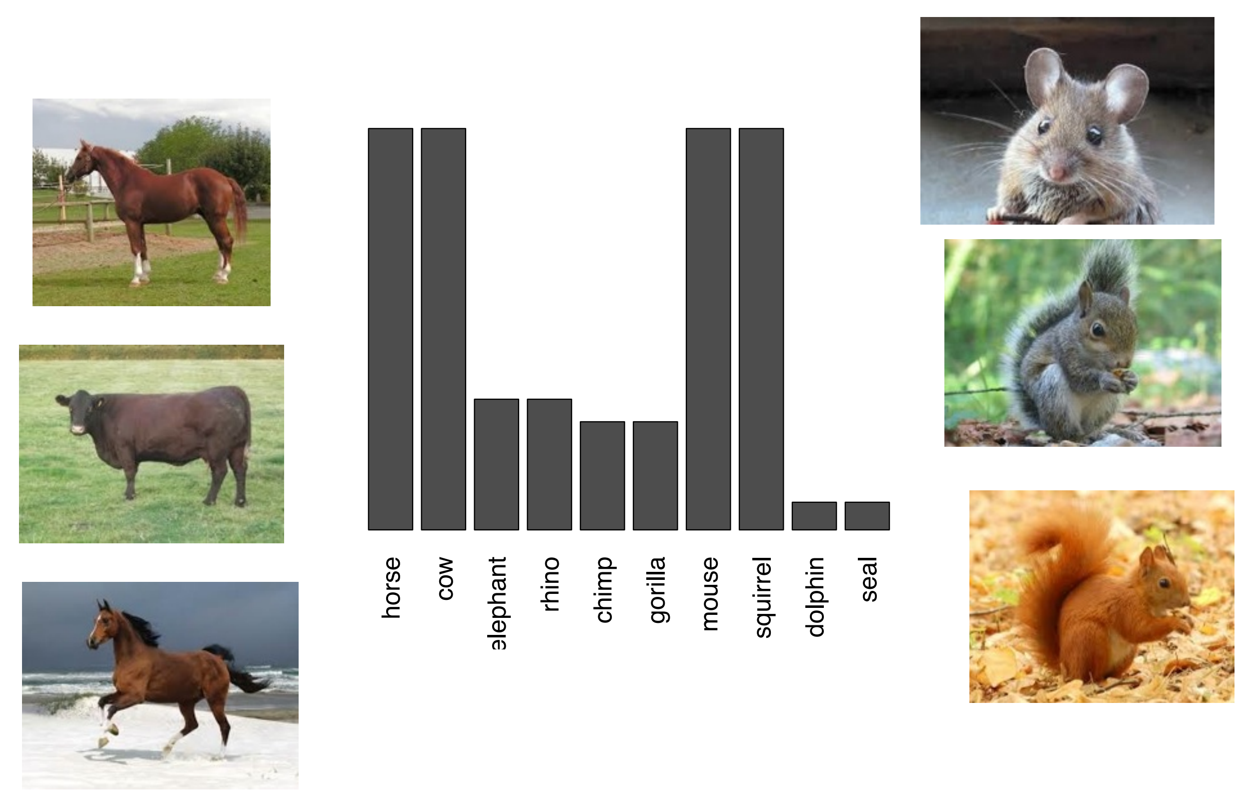




Just like previous examples, a small amount of data leads to to prefer simpler hypotheses

But with more data it becomes quite possible to transfer belief to composite hypotheses





Qualitatively important phenomena in human inductive reasoning

(from Osherson et al 1990)

Deductive reasoning problems

Premises

All humans are mortal

Socrates is human

Therefore, Socrates is mortal

Conclusion

Inductive reasoning problems

Premises →

Cats are mortal

Dogs are mortal

Therefore, Chimpanzees are (probably) mortal

Conclusion →

Premise-conclusion similarity

cow

rhino

strong →

elephant

premises are all similar to the conclusion

Premise-conclusion similarity

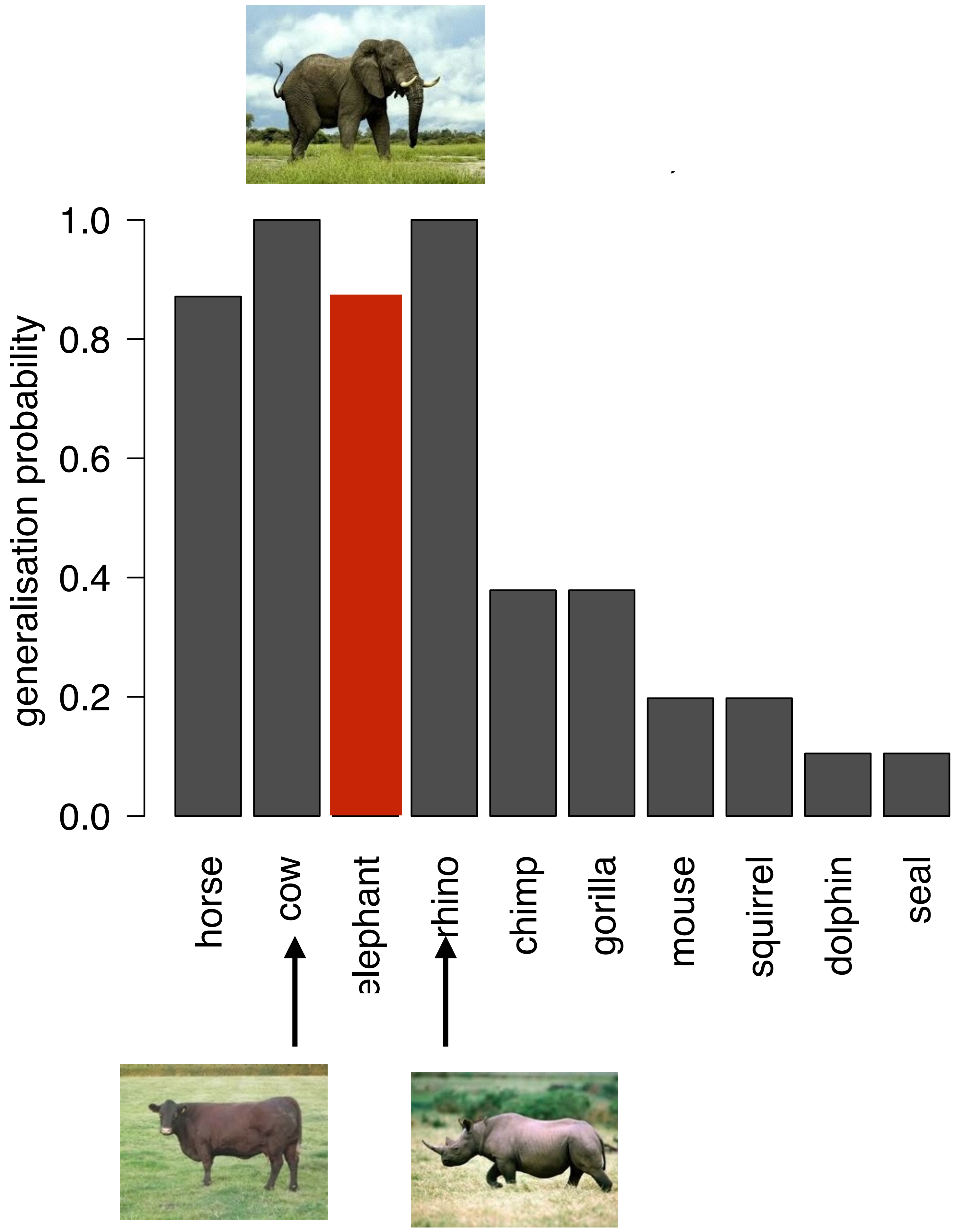
cow

rhino

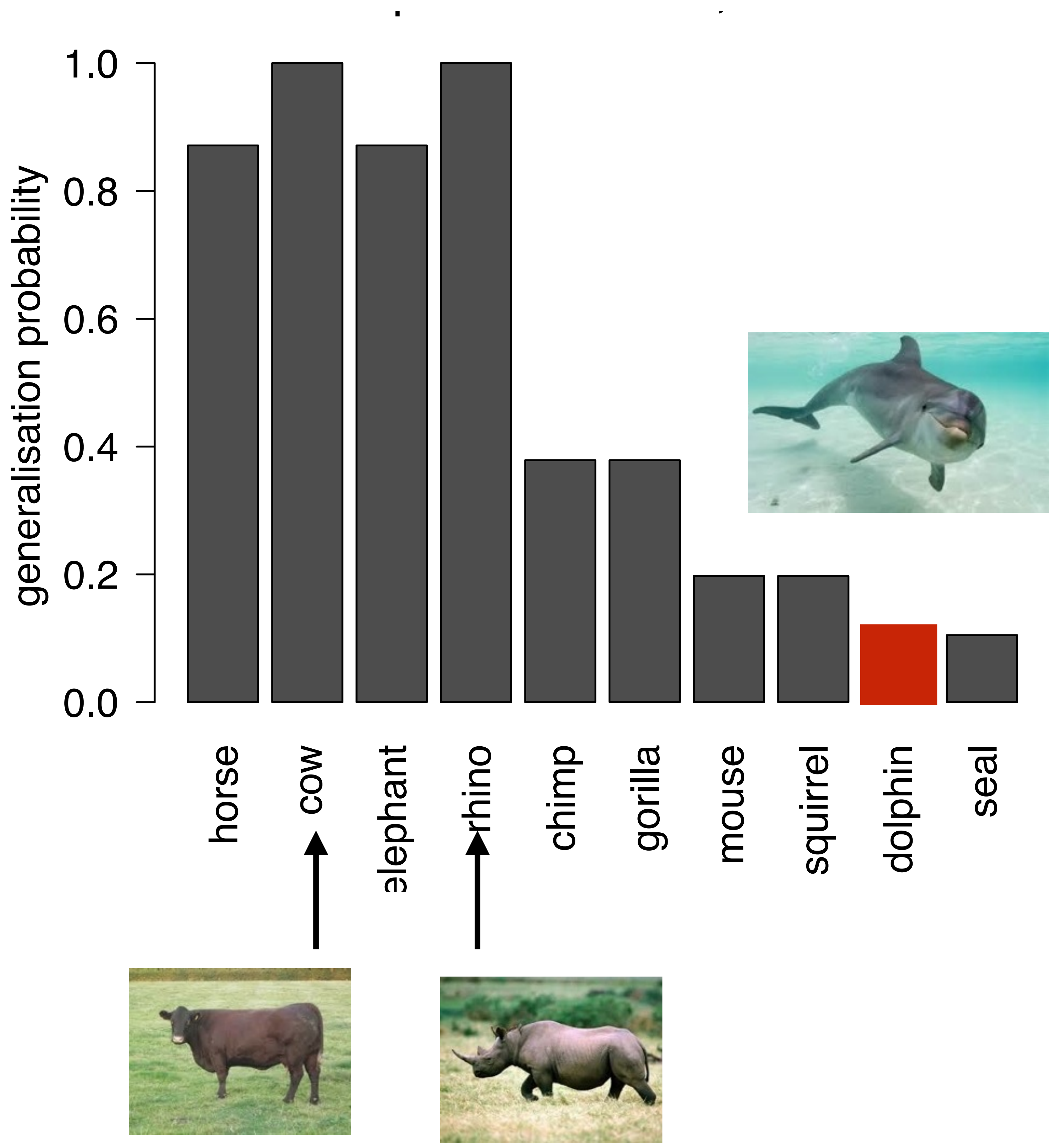
weak →

dolphin

premises are quite dissimilar to the conclusion



generalisation probability
1.0
0.8
0.6
0.4
0.2
0.0
horse
cow
elephant
rhino
chimp
gorilla
mouse
squirrel
dolphin
seal



generalisation probability
1.0
0.8
0.6
0.4
0.2
0.0
horse
cow
elephant
rhino
chimp
gorilla
mouse
squirrel
dolphin
seal

Premise diversity

cow

seal

strong →

chimp

these are dissimilar to each other

neither is especially similar to this

Premise diversity

cow

horse

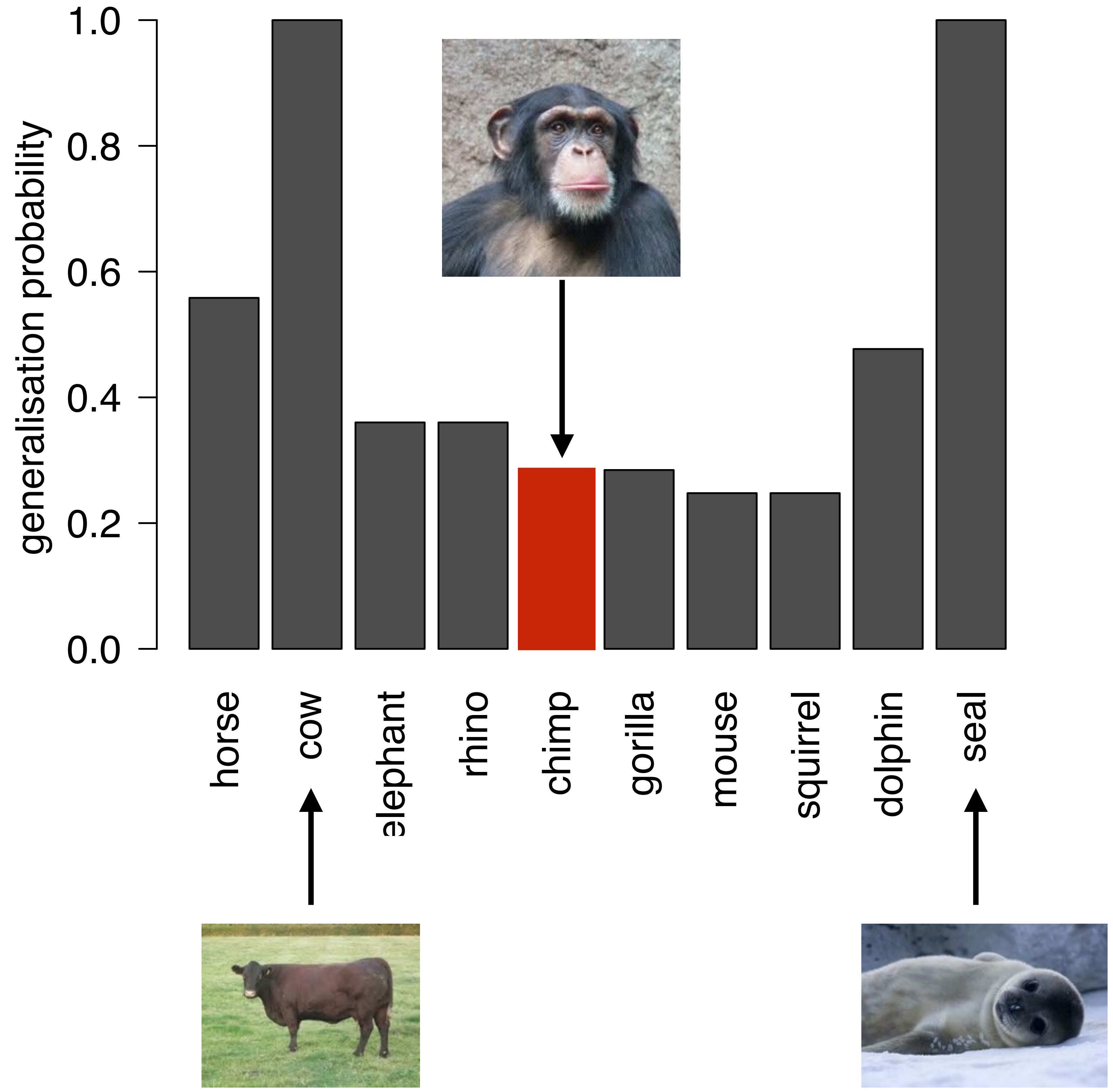
weak

→

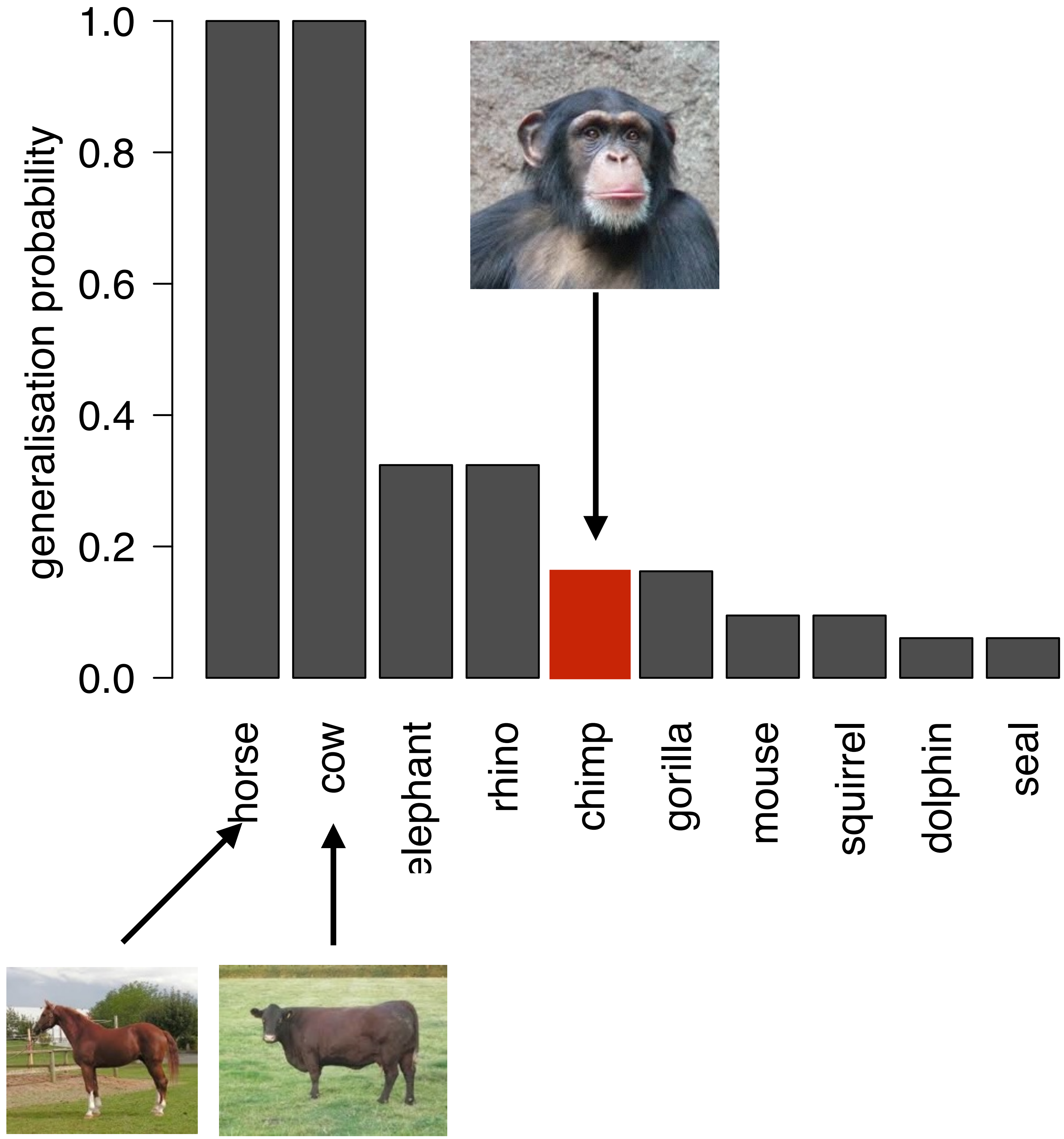
chimp

these are similar to each other

neither is especially
similar to this



generalisation probability
1.0
0.8
0.6
0.4
0.2
0.0
horse
cow
elephant
rhino
chimp
gorilla
mouse
squirrel
dolphin
seal


generalisation probability
1.0
0.8
0.6
0.4
0.2
0.0
horse
cow
elephant
rhino
chimp
gorilla
mouse
squirrel
dolphin
seal

Premise monotonicity*

horse

squirrel

modest →

chimp

Premise monotonicity*

horse

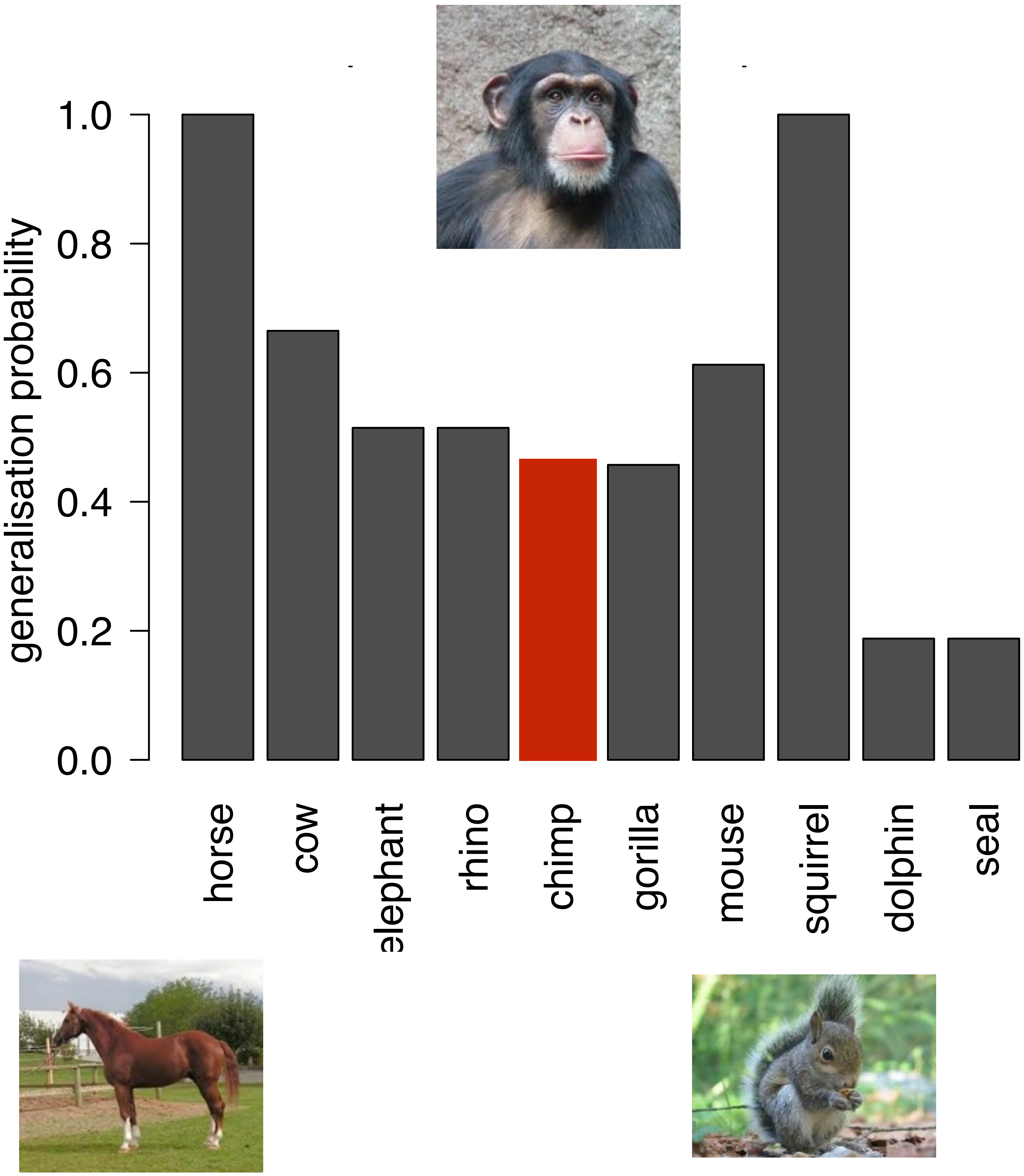
squirrel

elephant

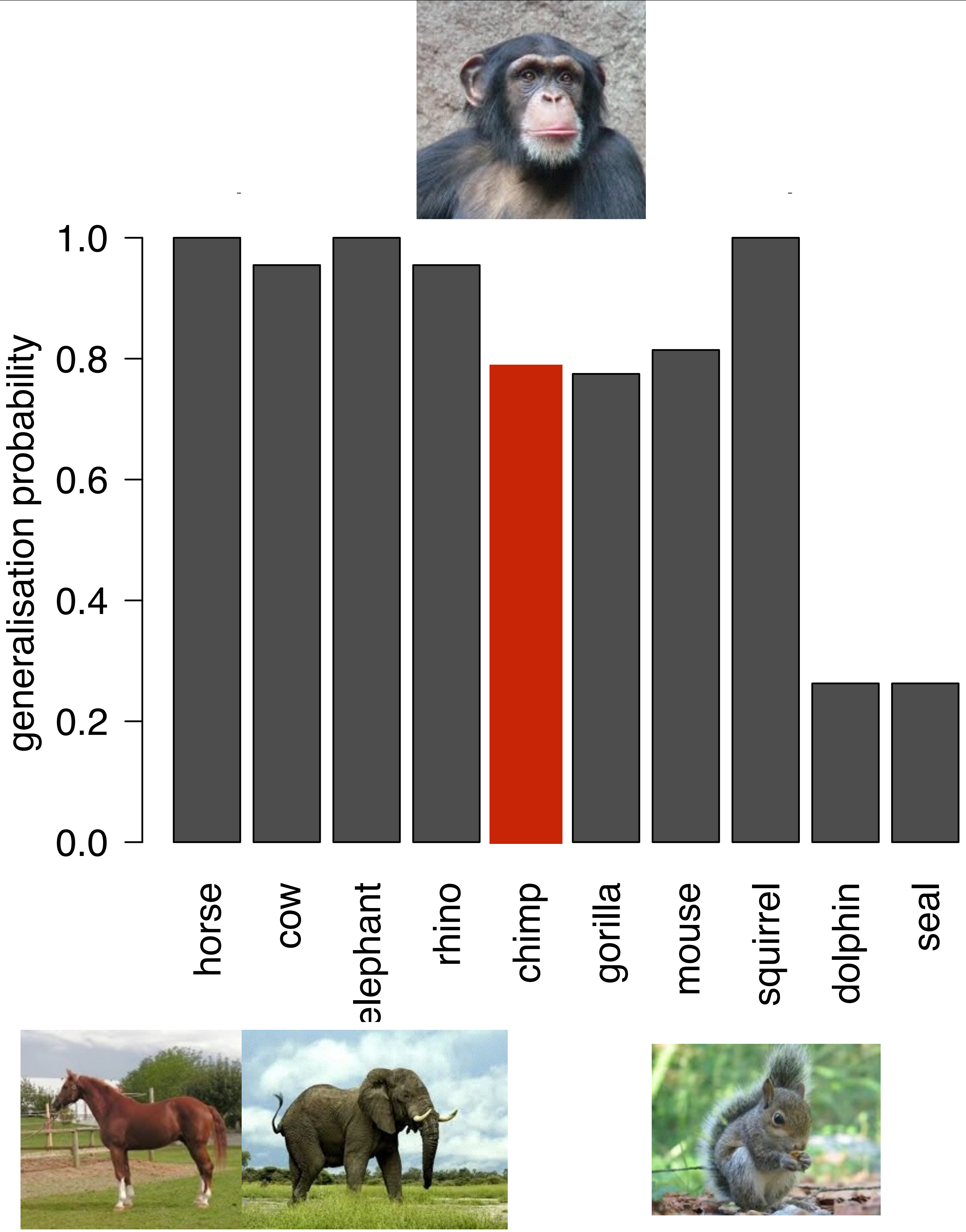
stronger →

chimp

adding premises
usually (not always)
strengthens arguments



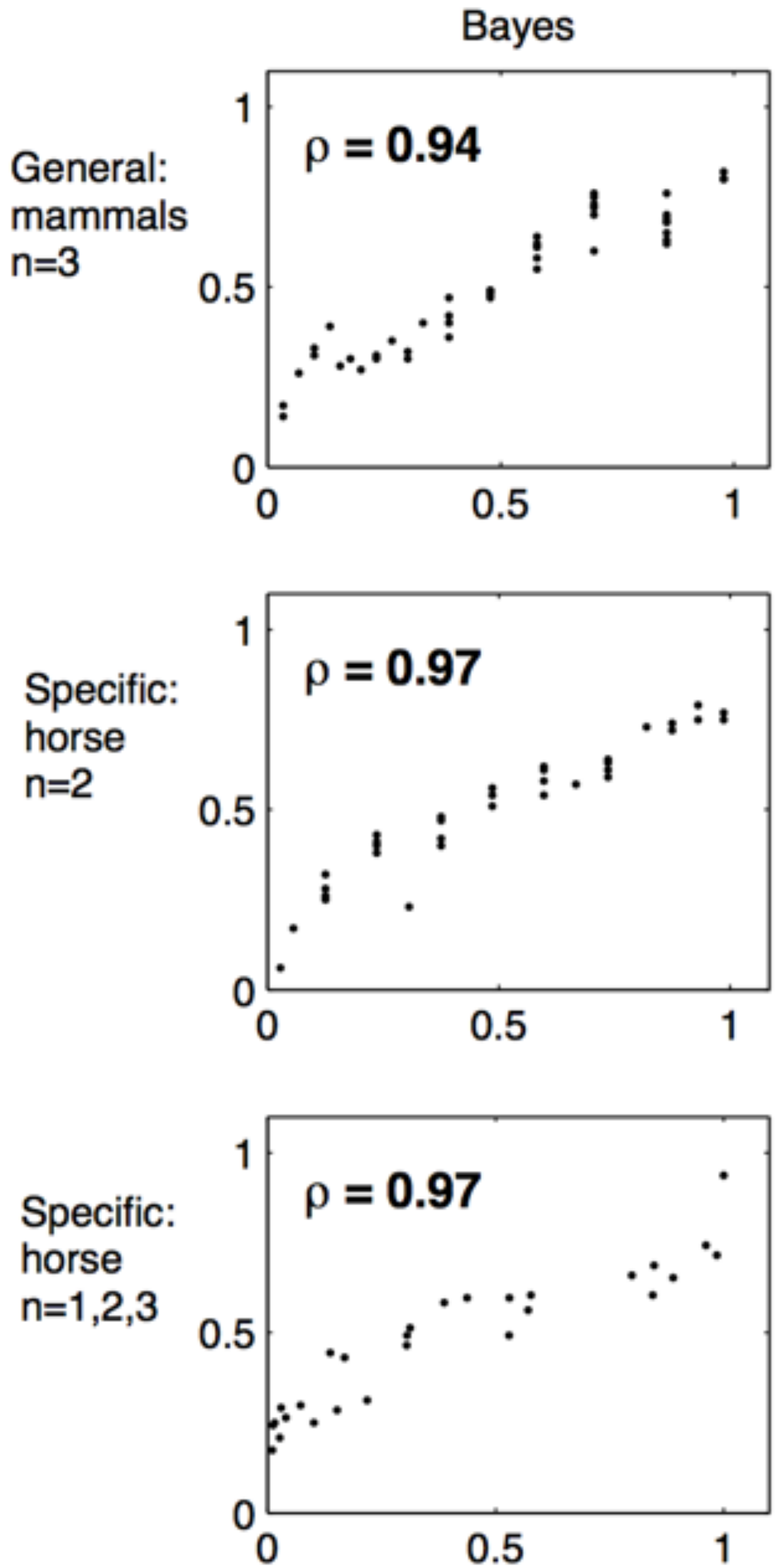
generalisation probability
1.0
0.8
0.6
0.4
0.2
0.0
horse
cow
elephant
rhino
chimp
gorilla
mouse
squirrel
dolphin
seal



generalisation probability
1.0
0.8
0.6
0.4
0.2
0.0
horse
cow
elephant
rhino
chimp
gorilla
mouse
squirrel
dolphin
seal

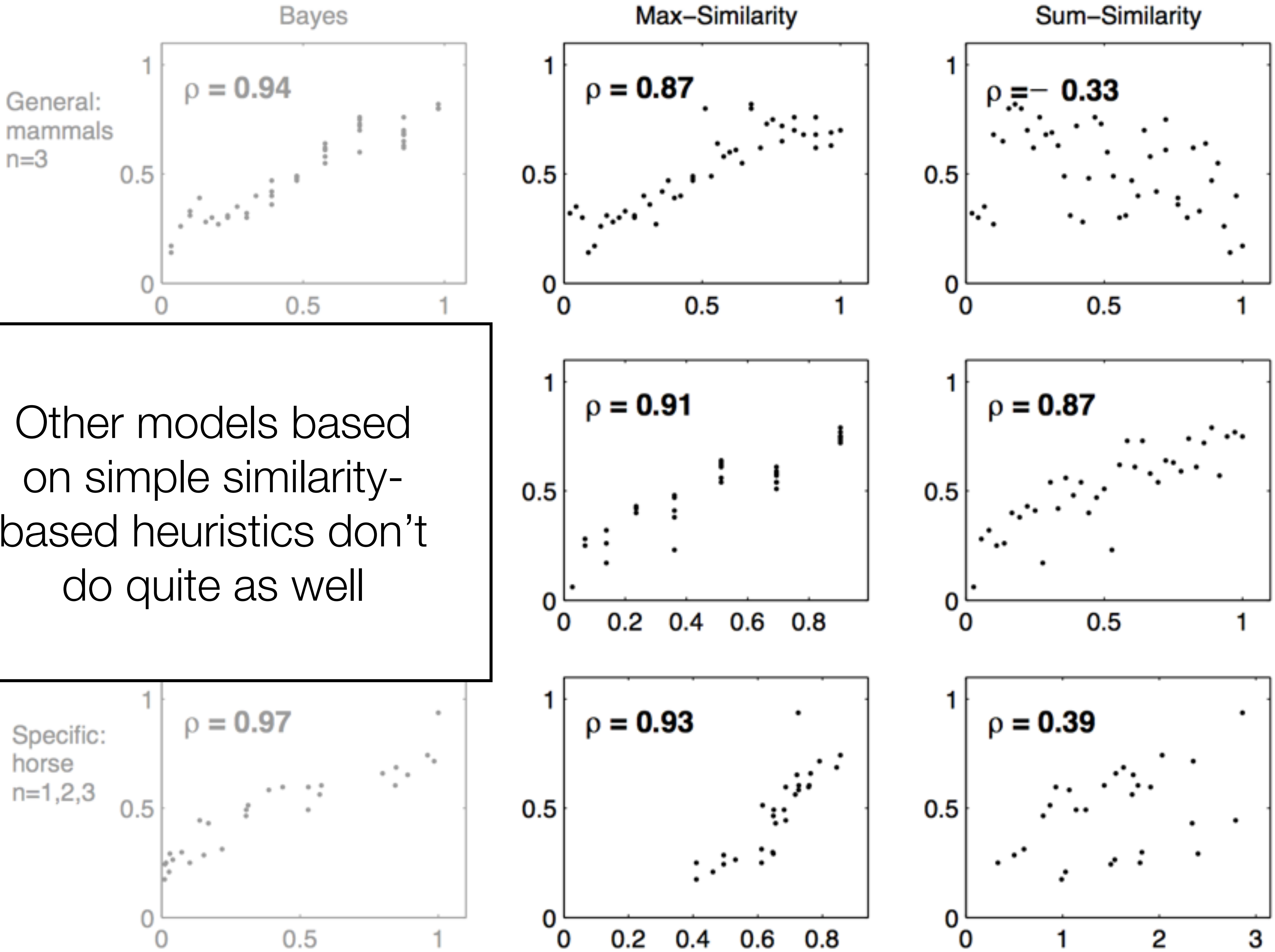
A more quantitative data fitting exercise

(see Sanjana & Tenenbaum 2003)



Model predictions on the x-axis
Human endorsement rates for specific arguments on the y-axis

The Bayesian model does surprisingly well at predicting human judgments



Other models based on simple similarity-based heuristics don't do quite as well

A unified perspective on deductive reasoning and inductive reasoning

(Lassiter & Goodman 2012,
see also Oaksford & Chater 2007)

This is (sensible) inductive reasoning

Cows have X
Horses have X

Therefore, it is plausible that
elephants have X

This is (incorrect) deductive reasoning

Cows have X
Horses have X

Therefore, it is a certainty that
elephants have X

The difference?

Cows have X
Horses have X

Therefore, **it is plausible** that elephants have X

Cows have X
Horses have X

Therefore, **it is a certainty** that elephants have X

In linguistics, this is called an epistemic modal frame

Proposal: the only difference between inductive reasoning and deductive reasoning is that a different standard of proof is required.

It is plausible that elephants have X

“plausible” requires the conclusion to be 55% probable?

It is a certainty that elephants have X

“certain” requires the conclusion to be 95% probable?

Proposal: the only difference between inductive reasoning and deductive reasoning is that a different standard of proof is required.

For any given argument, the endorsement probability ***e*** is some monotonic function of the generalisation probability ***g*** that depends on the frame ***f***

Proposal: the only difference between inductive reasoning and deductive reasoning is that a different standard of proof is required.

For any given argument, the endorsement probability *e* is some monotonic function of the generalisation probability *g* that depends on the frame *f*

endorsement probability for the conclusion in frame f

$$e = g^{\alpha(f)}$$

generalisation probability for the conclusion

standard of proof required in this frame

Ambitious “power law” prediction:

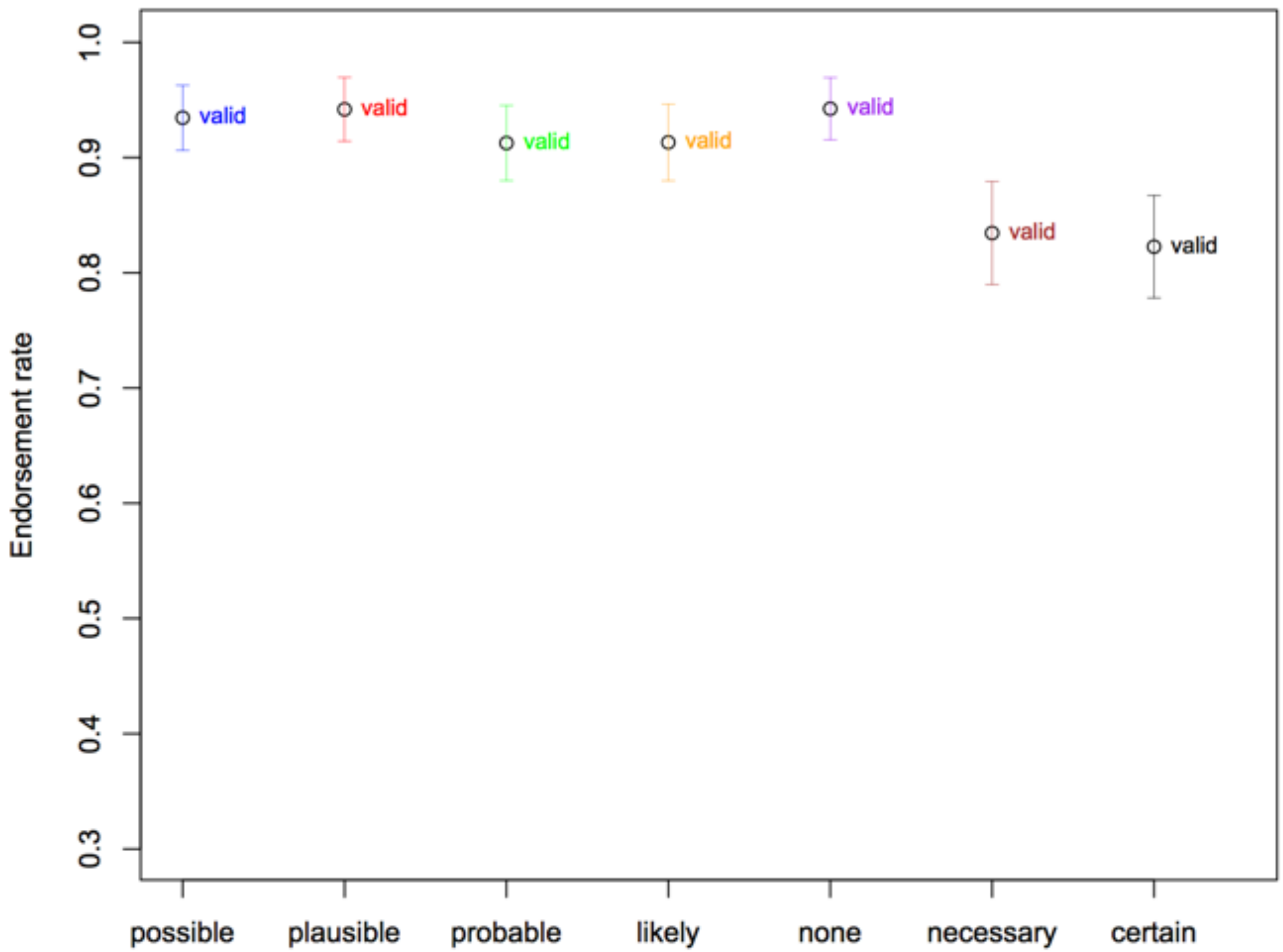
For any argument *A* and any pair of frames *X* and *Y*, the endorsement probabilities for *A* in frames *X* and *Y* are related by some positive power *r*

$$
\begin{aligned}
e(A, X) &= {g_A}^{\alpha_X} \\
&= {g_A}^{r\alpha_Y} \\
&= ({g_A}^{\alpha_Y})^r \\
&= e(A, Y)^r
\end{aligned}
$$

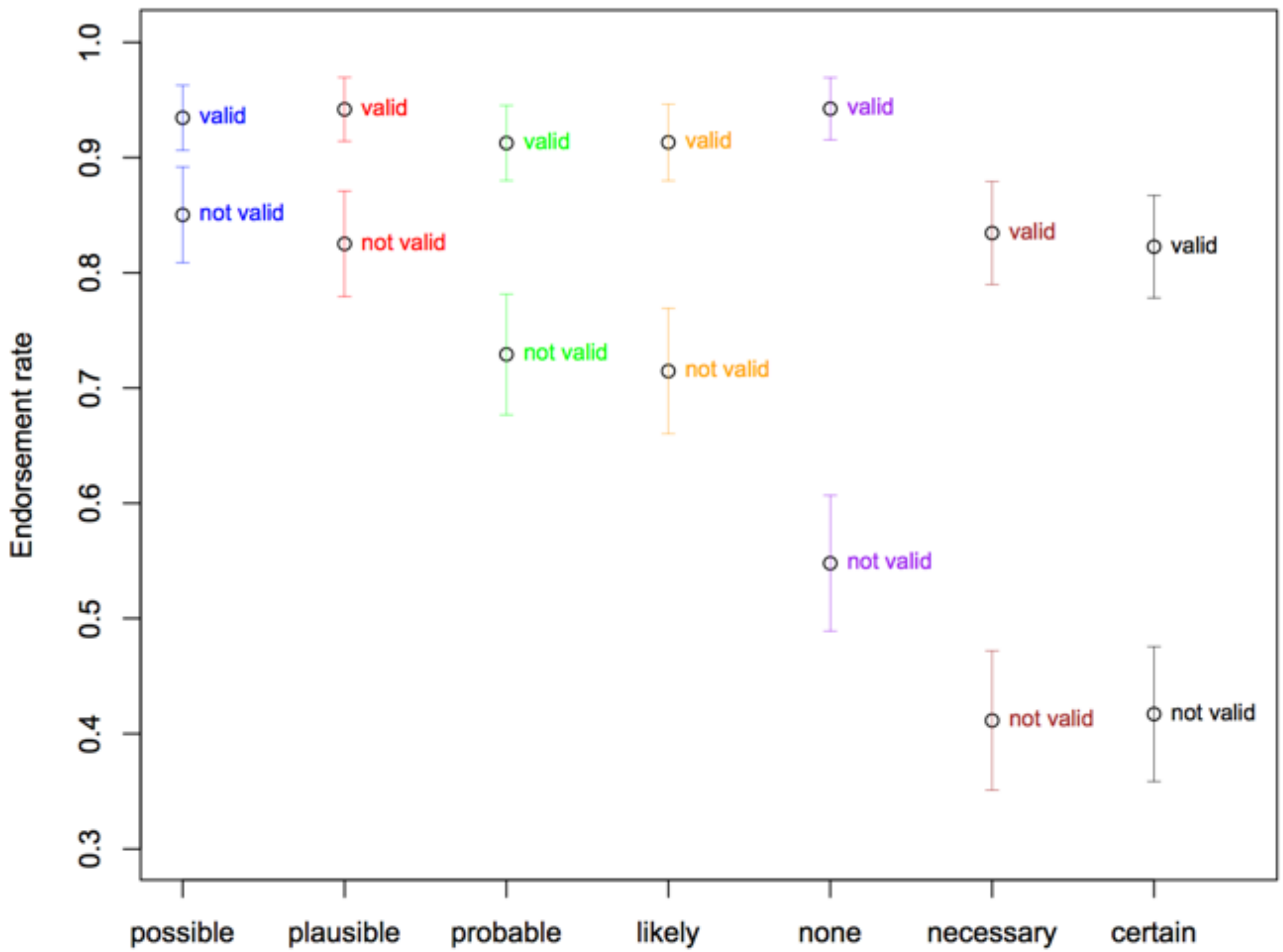
More psychologically critical prediction: frame monotonicity.

If argument *A* is endorsed more than argument *B* in frame *X*, then it must also be endorsed more in frame *Y*

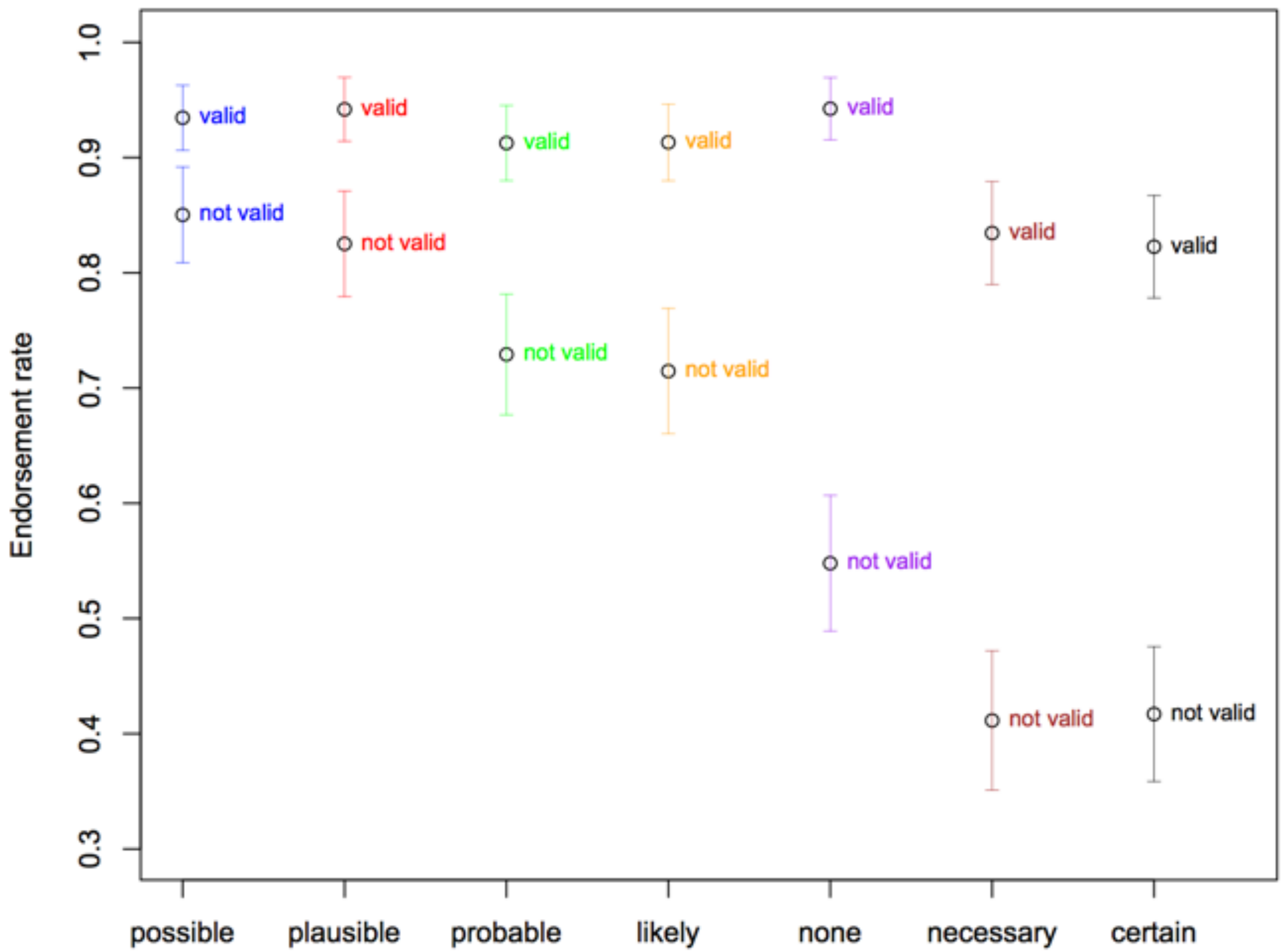
$$
\begin{aligned}
e(A, X) &> e(B, X) \\
\iff \quad {g_A}^{\alpha_X} &> {g_B}^{\alpha_X} \\
\iff \quad {g_A}^{r\alpha_Y} &> {g_B}^{r\alpha_Y} \\
\iff \quad ({g_A}^{\alpha_Y})^r &> ({g_B}^{\alpha_Y})^r \\
\iff \quad e(A, Y)^r &> e(B, Y)^r \\
\iff \quad e(A, Y) &> e(B, Y)
\end{aligned}
$$



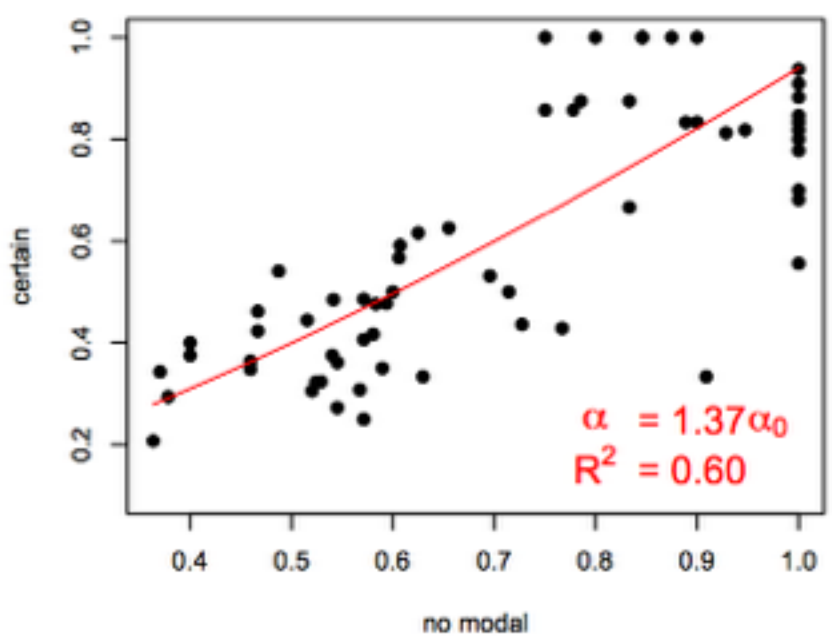
How likely are "deductively valid" arguments to be endorsed in different frames?



How likely are "deductively invalid" arguments to be endorsed in different frames?



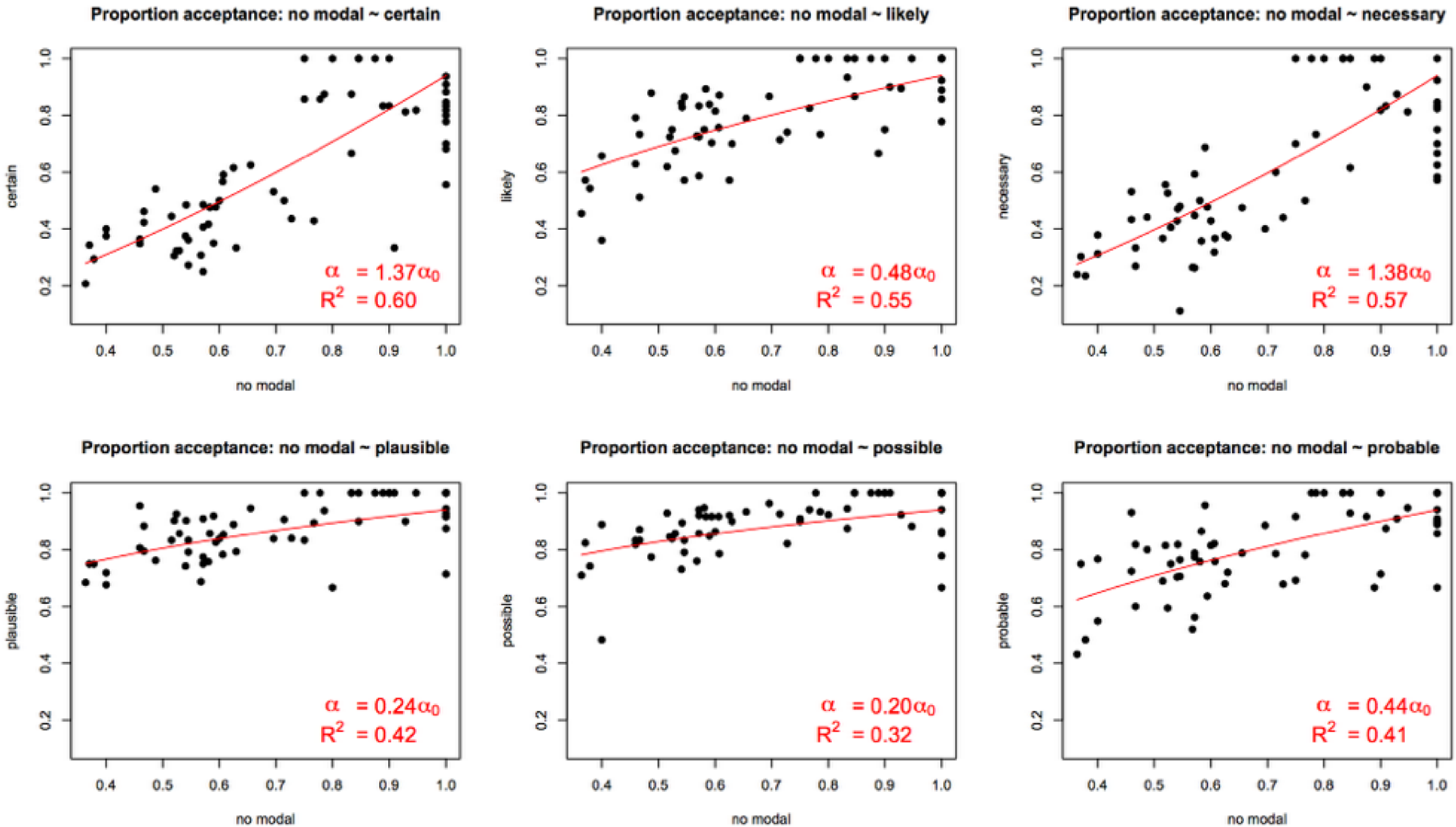
The ordering of the frames looks about the same for both argument types. Suggests frame monotonicity holds. But we can do better…



Endorsement rates plotted for individual arguments, in two separate frames (“certain” and “none provided”)

It’s noisy, but looks pretty linear.

That satisfies monotonicity, and is consistent with a power law (though it’s kind of weak evidence for a power law)



Here it is more generally

Summary

- Extending the inductive generalisation model
  - Hierarchical clustering to obtain tree structures (rather than spaces)
  - Composite hypothesis spaces
- Extended the tasks considered
  - We're not just modelling "simple" stimulus generalisation, we're also talking about inductive reasoning tasks more broadly
  - Presented some evidence that human deductive and inductive reasoning can both be captured within this framework

Readings

- Osherson, Smith, Wilkie, Lopez & Shafir (1990). Category based induction. *Psychological Review 97*, 185-200
- Sanjana & Tenenbaum (2003). Bayesian models of inductive generalisation. *Advances in Neural Information Processing Systems.*
- Lassiter & Goodman (2012). How many types of reasoning? Inference, probability & natural language semantics. *Proceedings of the 34th Annual Conference of the Cognitive Science Society.*