Categories and Concepts - Spring 2019

Category-based induction

Brenden Lake

PSYCH-GA 2207

Reminder: final paper proposals due 11/13

- Final assignment proposal due on Wednesday 11/13 (one half page written).
- Final paper due date is Monday 12/16
- The final paper is written individually (no groups).
- The final paper should address one of the topics covered in the class in more detail. Alternatively, it could investigate a topic that was not covered in class.
- The paper should include a critical review of the literature, along with theoretical conclusions or suggestions for future research. I would expect papers to be about 12 pages long
- If you want to link the paper to your research, that’s encouraged.

Deductive vs. inductive reasoning

Deductive reasoning involves *logical reasoning* from one or more statements (premises) to reach a certain conclusion

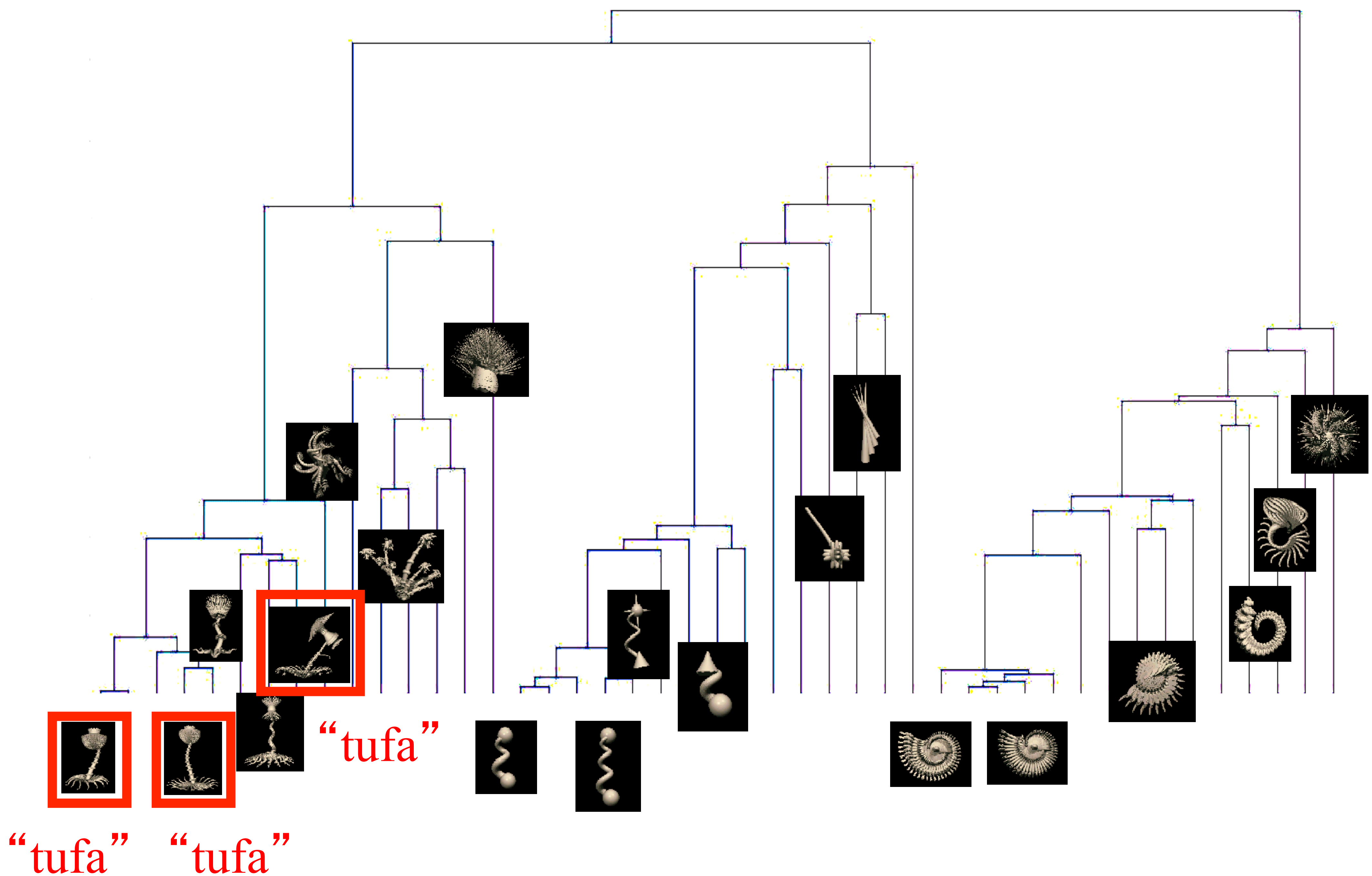
If an angle is between 90° and 180°, then it is obtuse
<u>We know that A = 120°</u>
Therefore, A is an obtuse angle

Inductive reasoning involves *probabilistic reasoning* from premises that supply some evidence for the truth of the conclusion

Foxes have sesamoid bones.
<u>Pigs have sesamoid bones.</u>
Therefore, Gorillas have sesamoid bones.

Category learning is inherently an inductive task, but that's not the type of induction we mean today

Here are some "tufas", where are the others?



Category-based induction (example 1)

Predicting unobserved properties based on a category label

"Can you take care of my dog?"

I have never met your dog, but I can guess
-what food it may like,
-It probably barks, likes chasing squirrels, likes playing catch, etc.
-It will poop,
-It will probably drool,
-etc.

Category-based induction (example 2)

Predicting unobserved properties based on *knowledge of other categories*

“Can you take care of my chinchilla?”

I know next to nothing about chinchillas, but I can guess

-what food it may like,

-It probably squeaks,

-It will poop,

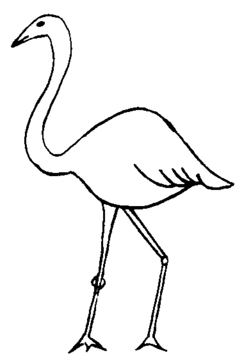
-It doesn’t need to be taken for a walk,

-etc.

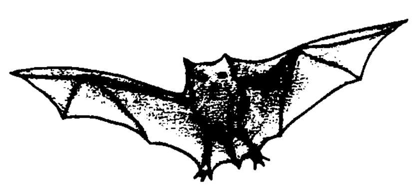
Category-based induction (example 3)

(Gelman & Markman, 1986)

Provided



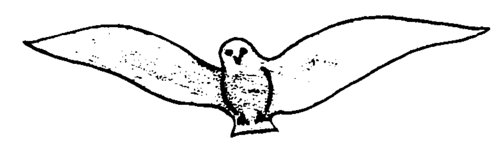
“This bird’s heart has a right aortic arch only”



“This bat’s heart has a left aortic arch only”

Query

“What does this bird’s heart have?”



Results: 4 year olds generalize based on category membership ~68% of time, overriding a distractor chosen for strong perceptual similarity

Category-based induction is very common in communication

"Why didn't you turn in your homework?"

- "My dog ate my homework" is straightforward
- "My dad ate my homework" requires an explanation

"Where did you get those shoes?"

- "I picked these shoes up at the mall" is straightforward
- "I picked these shoes off my neighbor's porch" requires an explanation

First study of category-based induction

JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR **14**, 665–681 (1975)

Inductive Judgments about Natural Categories

LANCE J. RIPS
University of Chicago

The present study examined the effects of semantic structure on simple inductive judgments about category members. For a particular category (e.g., *mammals*), subjects were told that one of the species (e.g., *horses*) had a given property (an unknown disease) and were asked to estimate the proportion of instances in the other species that possessed the property. The results indicated that category structure—in particular, the typicality of the species—influenced subjects' judgments. These results were interpreted by models based on the following assumption: When little is known about the underlying distribution of a property, subjects assume that the distribution mirrors that of better-known properties. For this reason, if subjects learn that an unknown property is possessed by a typical species (i.e., one that shares many of its properties with other category members), they are more likely to generalize than if the same fact had been learned about an atypical species.

Gaps in our knowledge of facts force us to rely on inductive methods in determining the truth or probability of certain statements. One, by now traditional, way of studying inductive strategies experimentally is through concept attainment tasks, which have been claimed to provide a direct analogue of inductive reasoning (Hunt, Marin, & Stone, 1966; Trabasso, Rollins, & Shaughnessy, 1971). The basis of the analogy is that in concept formation paradigms, as in inductive reasoning, tentative hypotheses are advanced on the basis of preliminary evidence. These hypotheses are strengthened by confirming evidence or are revised in the light of contradictory making, and tachistoscopic recognition. Nevertheless, it is concept attainment that is most often cited as the counterpart of inductive reasoning.

However, Rosch (1975) has noted that concept attainment paradigms may differ critically from other inductive situations. Most concept attainment studies employ logical combinations of binary attributes so that the resulting concept has well-defined boundaries. A concept so defined has an all-or-none structure, in the sense that no instance is a better exemplar of the concept than any other. Natural language concepts, on the other hand, do possess internal structure, and

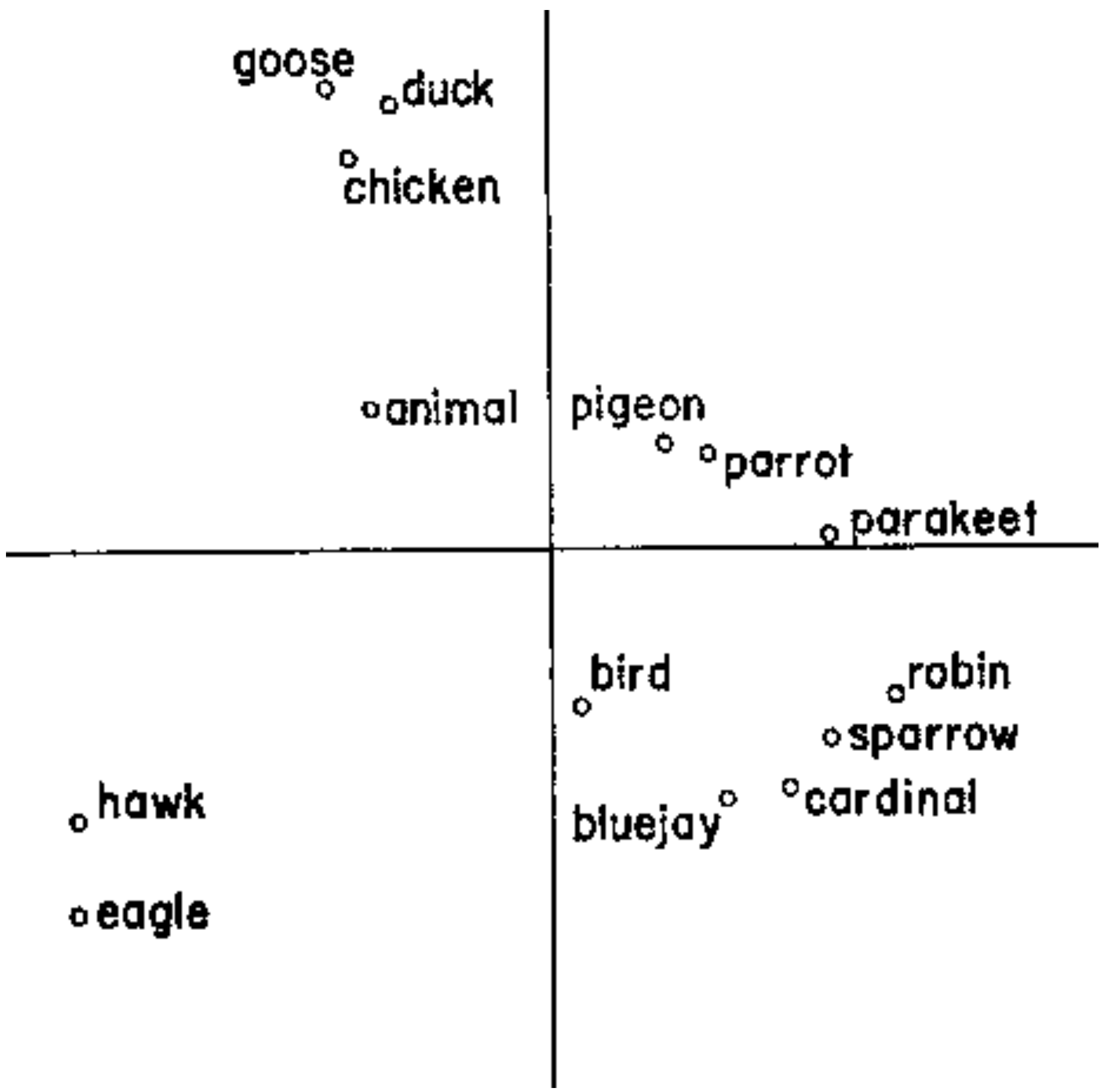
Rips' (1975) category-based induction task

- Used mammals and birds
- Used *blank predicates*, which hopefully do not have any specific effect on induction
  - e.g., a new type of contagious disease
  - (however, as you know from Kemp & Tenenbaum reading, disease isn't a good choice for taxonomic reasoning!)

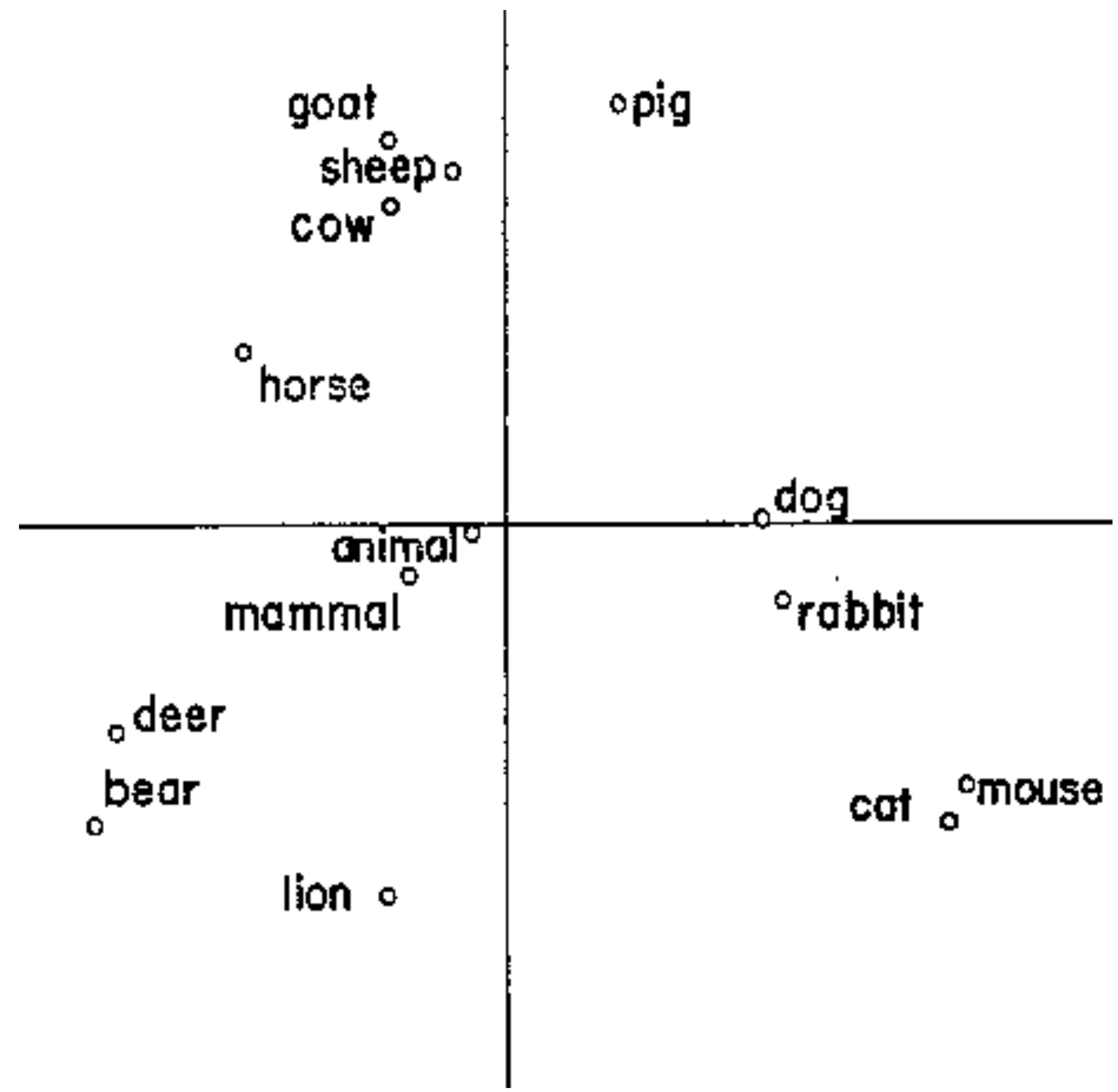
- Example trial: "If <u>pigs</u> have a disease, what proportion of <u>deer</u> would be likely to get the disease?"
  - Only one premise category
  - Answer given as a proportion
  - Two basic variables: can substitute "pigs" (premise) or "deer" (conclusion) with any mammal, including "dogs", "rabbits," etc.

Rips's (1975) semantic spaces

Birds



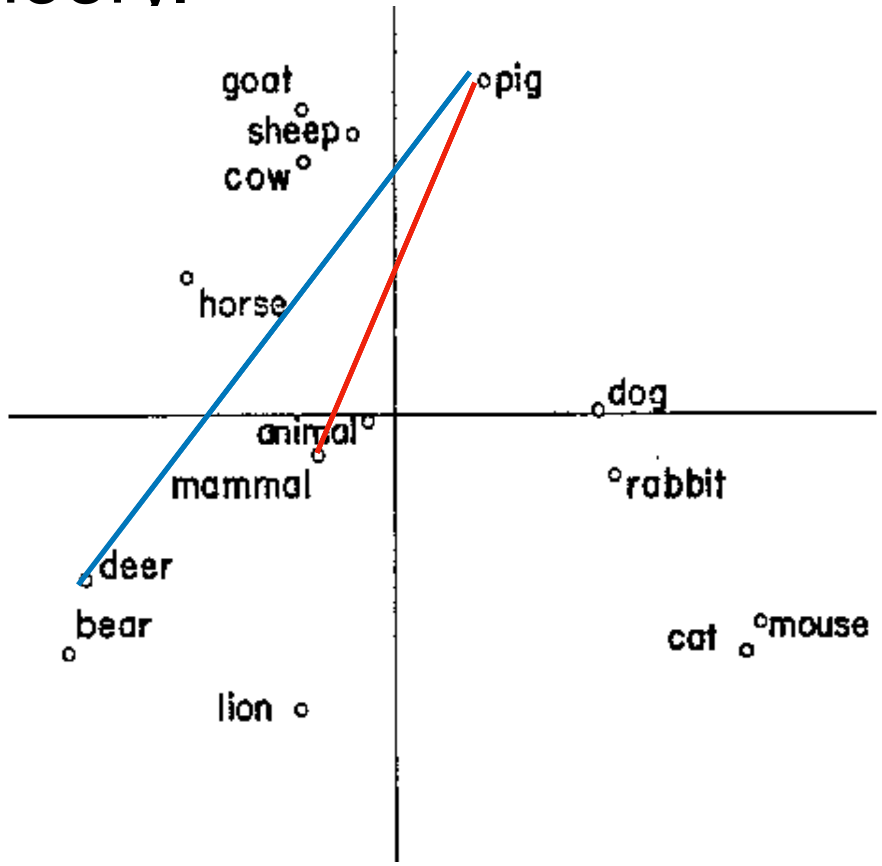
Mammals



Rips' (1975) results

- Subjects' answers seem to rely on two factors:
  - Similarity of the premise to conclusion category (smaller distance is stronger argument)
  - Typicality of premise category (smaller distance in MDS space from premise to superordinate category leads to strong argument)
- Typicality of the conclusion category had no effect (but it may be redundant with the first two).
- Note that there is a categorical component and a noncategorical component to this theory.

"If <u>pigs</u> have a disease, what proportion of <u>deer</u> would be likely to get the disease?"



Psychological Review
1990, Vol. 97, No. 2, 185–200

0033-295X/90/$00.75

Category-Based Induction

Daniel N. Osherson
Massachusetts Institute of Technology

Edward E. Smith
University of Michigan

Ormond Wilkie
Massachusetts Institute of Technology

Alejandro López
University of Michigan

Eldar Shafir
Princeton University

An argument is categorical if its premises and conclusion are of the form *All members of C have property P,* where C is a natural category like FALCON or BIRD, and P remains the same across premises and conclusion. An example is *Grizzly bears love onions. Therefore, all bears love onions.* Such an argument is psychologically strong to the extent that belief in its premises engenders belief in its conclusion. A subclass of categorical arguments is examined, and the following hypothesis is advanced: The strength of a categorical argument increases with (a) the degree to which the premise categories are similar to the conclusion category and (b) the degree to which the premise categories are similar to members of the lowest level category that includes both the premise and the conclusion categories. A model based on this hypothesis accounts for 13 qualitative phenomena and the quantitative results of several experiments.

The Problem of Argument Strength

Fundamental to human thought is the confirmation relation, joining sentences $P_1 \ldots P_n$ to another sentence C just in case belief in the former leads to belief in the latter. Theories of confirmation may be cast in the terminology of argument strength, because $P_1 \ldots P_n$ confirm C only to the extent that $P_1 \ldots P_n/C$ is a strong argument. We here advance a partial theory of argument strength, hence of confirmation.

To begin, it will be useful to review the terminology of argument strength. By an *argument* is meant a finite list of sentences, the last of which is called the *conclusion* and the others its *premises*. Schematic arguments are written in the form P_1

belief in the conclusion of an argument (independently of its premises) is not sufficient for argument strength. For this reason, Argument 1 is stronger than Argument 2 for most people, even though the conclusion of Argument 2 is usually considered more probable than that of Argument 1. An extended discussion of the concept of argument strength is provided in Osherson, Smith, and Shafir (1986). It will be convenient to qualify an argument as strong, without reference to a particular person S, whenever the argument is strong for most people in a target population (e.g., American college students). We also say that $P_1 \ldots P_n$ confirm C if $P_1 \ldots P_n/C$ is strong.

An illuminating characterization of argument strength

Examples of Osherson et al. induction problems

Specific Argument (all categories at same level)

Mosquitoes use the neurotransmitter dihedron.
Ants use the neurotransmitter dihedron.

Bees use the neurotransmitter dihedron.

General Argument (conclusion at more general level)

Grizzly bears love onions.
Polar bears love onions.

All bears love onions

Two key Osherson et al. variables for inductive strength

Similarity-coverage model:

- **Similarity** of premises to conclusion
  - *maximum* of each pair of premise-conclusion categories

- **Coverage**: How well the premise categories cover the superordinate category that includes all the categories mentioned?
  - *average* of similarity, as computed above, between premise set and each member of higher-level category

- Note analogy to Rips: also a categorical and noncategorical (similarity) component

Two key Osherson et al. variables

- **Similarity** of premises to conclusion
  - *maximum* of each pair of premise-conclusion categories

Example**:**

Flies use the neurotransmitter dihedron.
<u>Ants use the neurotransmitter dihedron.</u>
Bees use the neurotransmitter dihedron.

Similarity({Flies, Ants}, Bees)
= Max[(Similarity(Flies, Bees), Similarity(Ants, Bees)]
= Similarity(Flies, Bees)

- **Coverage**: How well the premise categories cover the superordinate category that includes all the categories mentioned?
  - *average* of similarity, as computed above, between premise set and each member of higher-level category

Example**:**

Grizzly bears love onions.
<u>Polar bears love onions.</u>
All bears love onions

AVERAGE OF…
Similarity({Grizzly, Polar}, Black bears) = Similarity(Grizzly, Black bears), …
Similarity({Grizzly, Polar}, Grizzly bears) = Similarity(Grizzly, Grizzly), …
Similarity({Grizzly, Polar}, Panda bears) = Similarity(Polar, Panda),…

Summary of 13 phenomena

Table 1
Summary of the 13 Phenomena

Phenomenon	Stronger argument (Version a)	Weaker argument (Version b)
General arguments		
1. Premise Typicality	ROBIN/BIRD [73]	PENGUIN/BIRD [7]
2. Premise Diversity	HIPPO, HAMSTER/ MAMMAL [59]	HIPPO, RHINO/MAMMAL [21]
3. Conclusion Specificity	BLUEJAY, FALCON/ BIRD [75]	BLUEJAY, FALCON/ANIMAL [5]
4. Premise Monotonicity	HAWK, SPARROW EAGLE/BIRD [75]	SPARROW, EAGLE/BIRD [5]
Specific arguments		
5. Premise–Conclusion Similarity	ROBIN, BLUEJAY/ SPARROW [76]	ROBIN, BLUEJAY/GOOSE [4]
6. Premise Diversity	LION, GIRAFFE/ RABBIT [52]	LION, TIGER/RABBIT [28]
7. Premise Monotonicity	FOX, PIG WOLF/GORILLA [66]	PIG, WOLF/GORILLA [14]
8. Premise–Conclusion Asymmetry	MICE/BAT [41] (40)	BAT/MICE [39] (20)
Mixed arguments		
9. Nonmonotonicity–General	CROW, PEACOCK/ BIRD [68]	CROW, PEACOCK RABBIT/BIRD [12]
10. Nonmonotonicity–Specific	FLY/BEE [51]	FLY, ORANGUTAN/BEE [29]
General and specific arguments		
11. Inclusion Fallacy	ROBIN/BIRD [52]	ROBIN/OSTRICH [28]
Limiting-case arguments		
12. Premise–Conclusion Identity	PELICAN/PELICAN	
13. Premise–Conclusion Inclusion	ANIMAL/BIRD	

Note. Number of subjects in Study 1 preferring each argument is given in brackets. Entries in parentheses are results of Study 2.

Premise Typicality

In general, the more typical the premise categories, the stronger the argument (via coverage).
So, (3) is stronger than (4).

(3) **Robins** have a high potassium in their blood.
All birds have a high potassium in their blood.

(4) **Penguins** have a high potassium in their blood.
All birds have a high potassium in their blood.

Premise Diversity

The more variable the premise categories, the stronger the argument (via coverage). So, (6) is stronger than (5):

(5) Hippopotamuses require Vitamin K.
Rhinoceroses require Vitamin K.
Humans require Vitamin K.

(6) Hippopotamuses require Vitamin K.
Bats require Vitamin K.
Humans require Vitamin K.

Premise Monotonicity

If you add more categories (all of them being at the same level) to the premises, the argument gets stronger (via similarity and/or coverage). So, (8) is stronger than (7):

(7) **Foxes** have sesamoid bones.
<u>**Pigs** have sesamoid bones.</u>
Gorillas have sesamoid bones.

(8) **Foxes** have sesamoid bones.
Pigs have sesamoid bones.
<u>**Wolves** have sesamoid bones.</u>
Gorillas have sesamoid bones.

Inclusion fallacy

Argument (9) is felt to be stronger than argument (10).

(9) <u>Robins have an ulnar artery.</u>
All Birds have an ulnar artery.

(10) <u>Robins have an ulnar artery.</u>
Ostriches have an ulnar artery.

Why is this a fallacy? Argument 10 is logically entitled by argument 9

Testing the model: “Osherson horse” dataset

Experiment: a set of arguments were written on cards and ranked for strength

Argument template:

X requires biotin for hemoglobin synthesis.
Y requires biotin for hemoglobin synthesis.
―――――――――――――――――――――――――――
Horses require biotin for hemoglobin synthesis.

Confirmation Scores for Two-Premise Specific Arguments (Horse, Experiment 4)

Mammals		Score	Mammals		Score
		strong			
COW	CHIMP	.79	GORILLA	SEAL	.41
COW	GORILLA	.75	GORILLA	ELEPHANT	.61
COW	MOUSE	.74	GORILLA	RHINO	.63
COW	SQUIRREL	.72	MOUSE	SQUIRREL	.17
COW	DOLPHIN	.73	MOUSE	DOLPHIN	.28
COW	SEAL	.73	MOUSE	SEAL	.25
COW	ELEPHANT	.75	MOUSE	ELEPHANT	.58
COW	RHINO	.77	MOUSE	RHINO	.62
CHIMP	GORILLA	.23	SQUIRREL	DOLPHIN	.32
CHIMP	MOUSE	.42	SQUIRREL	SEAL	.26
CHIMP	SQUIRREL	.40	SQUIRREL	ELEPHANT	.54
CHIMP	DOLPHIN	.40	SQUIRREL	RHINO	.61
CHIMP	SEAL	.43	DOLPHIN	SEAL	.06 weak
CHIMP	ELEPHANT	.59	DOLPHIN	ELEPHANT	.54
CHIMP	RHINO	.64	DOLPHIN	RHINO	.54
GORILLA	MOUSE	.48	SEAL	ELEPHANT	.51
GORILLA	SQUIRREL	.47	SEAL	RHINO	.56
GORILLA	DOLPHIN	.38	ELEPHANT	RHINO	.57

Results: Similarity-coverage model correlates r=0.96 with confirmation scores

Testing the model: “Osherson mammals” dataset

Experiment: a set of arguments were written on cards and ranked for strength

Argument template:

X requires biotin for hemoglobin synthesis.
Y requires biotin for hemoglobin synthesis.
Z requires biotin for hemoglobin synthesis.
___
All mammals require biotin for hemoglobin synthesis.

Confirmation Scores for Three-Premise, General Arguments

Mammals	Score	Mammals	Score	
HORSE COW MOUSE	.33	COW SEAL ELEPHANT	.47	
HORSE COW SEAL	.39	COW ELEPHANT RHINO	.14	weak
HORSE COW RHINO	.17	CHIMP GORILLA SQUIRREL	.30	
HORSE CHIMP SQUIRREL	.55	CHIMP GORILLA DOLPHIN	.31	
HORSE CHIMP SEAL	.75	CHIMP GORILLA SEAL	.30	
HORSE GORILLA SQUIRREL	.64	CHIMP SQUIRREL DOLPHIN	.80	
HORSE GORILLA DOLPHIN	.73	CHIMP SQUIRREL ELEPHANT	.62	
HORSE MOUSE SQUIRREL	.28	CHIMP SQUIRREL RHINO	.61	
HORSE MOUSE SEAL	.69	CHIMP DOLPHIN ELEPHANT	.72	
HORSE MOUSE RHINO	.42	GORILLA MOUSE SEAL	.82	strong
HORSE SQUIRREL SEAL	.63	GORILLA MOUSE ELEPHANT	.58	
HORSE SQUIRREL ELEPHANT	.47	GORILLA SQUIRREL DOLPHIN	.80	
HORSE DOLPHIN SEAL	.27	GORILLA SEAL ELEPHANT	.60	
HORSE DOLPHIN ELEPHANT	.49	GORILLA ELEPHANT RHINO	.26	
COW CHIMP DOLPHIN	.76	MOUSE SQUIRREL SEAL	.35	
COW CHIMP SEAL	.70	MOUSE DOLPHIN SEAL	.32	
COW CHIMP ELEPHANT	.40	MOUSE SEAL ELEPHANT	.70	
COW MOUSE SEAL	.68	MOUSE SEAL RHINO	.65	
COW MOUSE RHINO	.40	MOUSE ELEPHANT RHINO	.31	
COW SQUIRREL DOLPHIN	.76	SQUIRREL DOLPHIN SEAL	.30	
COW SQUIRREL RHINO	.36	SQUIRREL DOLPHIN RHINO	.68	
COW DOLPHIN ELEPHANT	.48	SQUIRREL SEAL RHINO	.62	
COW DOLPHIN RHINO	.49			

Results: Similarity-coverage model correlates r=0.87 with confirmation scores

Limitations of Osherson et al. approach

- Model is very successful at predicting phenomena in category-based induction (this paper is always cited)
- But the model doesn't do well when predicting meaningful predicates (non-blank; Heit & Rubinstein, 1994)

(11) Given that tuna/rabbits have **blood that contains between 2% and 3% potassium**, how likely are whales to have blood that contains between 2% and 3% potassium?"
(12) Given that tuna/rabbits **usually gather a large amount of food at once**, how likely are whales to usually gather a large amount of food at once?"

Induction is greater for rabbit-whale for biological properties, and tuna-whale in behavioral properties, so the predicate makes a difference. Oops!

Psychological Review
2009, Vol. 116, No. 1, 20–58

0033-295X/09/$12.00 DOI: 10.1037/a0014282

Structured Statistical Models of Inductive Reasoning

Charles Kemp
Carnegie Mellon University

Joshua B. Tenenbaum
Massachusetts Institute of Technology

Everyday inductive inferences are often guided by rich background knowledge. Formal models of induction should aim to incorporate this knowledge and should explain how different kinds of knowledge lead to the distinctive patterns of reasoning found in different inductive contexts. This article presents a Bayesian framework that attempts to meet both goals and describe 4 applications of the framework: a taxonomic model, a spatial model, a threshold model, and a causal model. Each model makes probabilistic inferences about the extensions of novel properties, but the priors for the 4 models are defined over different kinds of structures that capture different relationships between the categories in a domain. The framework therefore shows how statistical inference can operate over structured background knowledge, and the authors argue that this interaction between structure and statistics is critical for explaining the power and flexibility of human reasoning.

Keywords: inductive reasoning, property induction, knowledge representation, Bayesian inference

Humans are adept at making inferences that take them beyond the limits of their direct experience. Even young children can learn the meaning of a novel word from a single labeled example (Heibeck & Markman, 1987), predict the trajectory of a moving object when it passes behind an occluder (Spelke, 1990), and choose a gait that allows them to walk over terrain they have never before encountered. Inferences like these may differ in many respects, but common to them all is the need to go beyond the information given (Bruner, 1973).

Two different ways of going beyond the available information can be distinguished. Deductive inferences draw out conclusions that may have been previously unstated but were implicit in the data provided. Inductive inferences go beyond the available data in a more fundamental way and arrive at conclusions that are likely but not certain given the available evidence. Both kinds of infer-

This article describes a formal approach to inductive inference that should apply to many different problems, but we focus on the problem of property induction (Sloman & Lagnado, 2005). In particular, we consider cases where one or more categories in a domain are observed to have a novel property and the inductive task is to predict how the property is distributed over the remaining categories in the domain. For instance, given that bears have sesamoid bones, which species is more likely to share this property: moose or salmon (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Rips, 1975)? Moose may seem like the better choice because they are more similar biologically to bears, but different properties can lead to different patterns of inference. For example, given that a certain disease is found in bears, it may seem more likely that the disease is found in salmon than in moose—perhaps the bears picked up the disease from something they ate.

Structured statistical models of inductive reasoning

- Everyday inductive inferences are guided by rich background knowledge
- Different kinds of knowledge leads to distinct patterns of reasoning
- Theory-based approaches and “the knowledge view” explain how knowledge matters, but rarely attempt to formalize the content of intuitive theories
- Kemp and Tenenbaum present a “Bayesian property induction” framework that can capture different types of knowledge as different structures and stochastic processes operating over these structures

Review: Examples of Intuitive Theories or General Knowledge Relevant to Concept Learning

- biological knowledge
- naive physics
- naive psychology
    - beliefs, desires, goals
    - psychological effects of different events
    - different personality types
- causal mechanics of various machines and artifacts

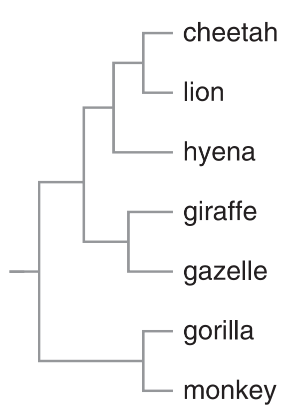
Most relevant to Kemp and Tenenbaum…

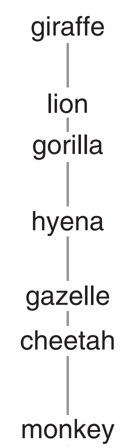
- biological knowledge
- geographical knowledge
- commonsense knowledge, physical knowledge
- causal, ecological knowledge

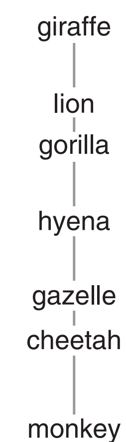
Four different intuitive theories to support inductive reasoning, unified as Bayesian property induction

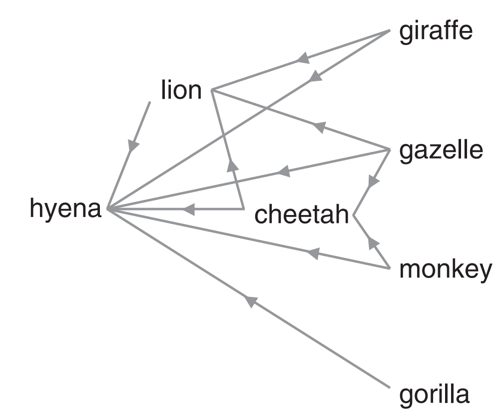
	Taxonomic model	Spatial model	Threshold model	Causal model
Example property	*has enzyme X132 in its bloodstream*	*ideally consumes around 15 g of sodium per week in the wild*	*is heavy enough to trigger most pit traps*	*carries leptospirosis*
Structural Form (F)	tree	low dimensional space	low dimensional space	directed graph

Structure (S)









	Taxonomic model	Spatial model	Threshold model	Causal model
Stochastic Process (T)	diffusion	diffusion	drift	transmission

Features $\{f^i\}$

Taxonomic model

	f^1	f^2	f^3	f^4	f^5	f^{new}
cheetah	●	○	○	●	●	●
lion	●	○	○	●	●	?
hyena	○	●	○	●	●	?
giraffe	○	●	●	○	●	?
gazelle	○	●	●	○	●	?
gorilla	○	●	○	○	○	?
monkey	○	●	○	○	○	?

Spatial model

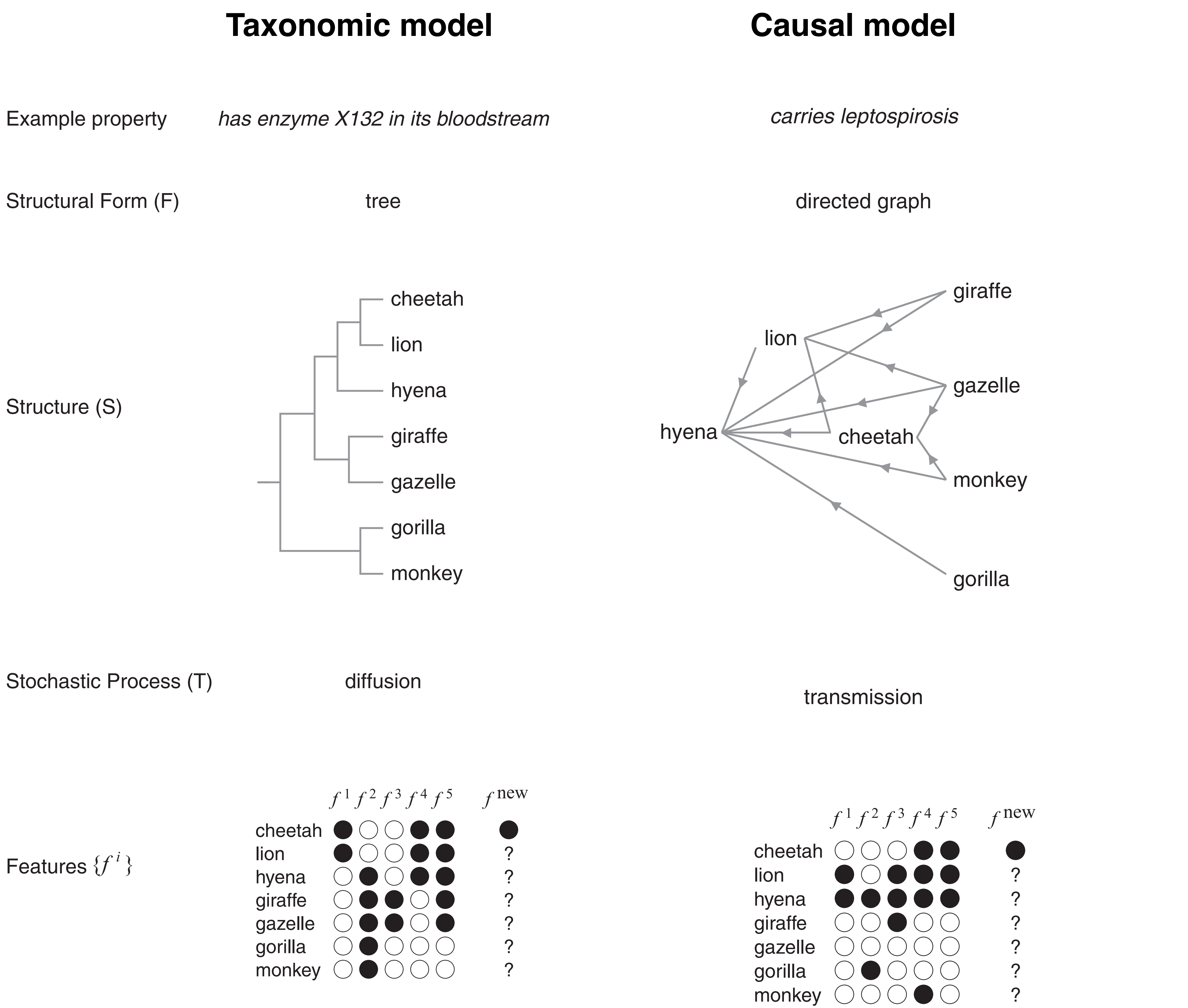
	f^1	f^2	f^3	f^4	f^5	f^{new}
cheetah	●	●	●	○	○	●
lion	○	○	○	●	○	?
hyena	○	○	●	●	○	?
giraffe	○	○	○	○	●	?
gazelle	●	●	●	○	○	?
gorilla	○	○	○	●	○	?
monkey	●	○	○	○	○	?

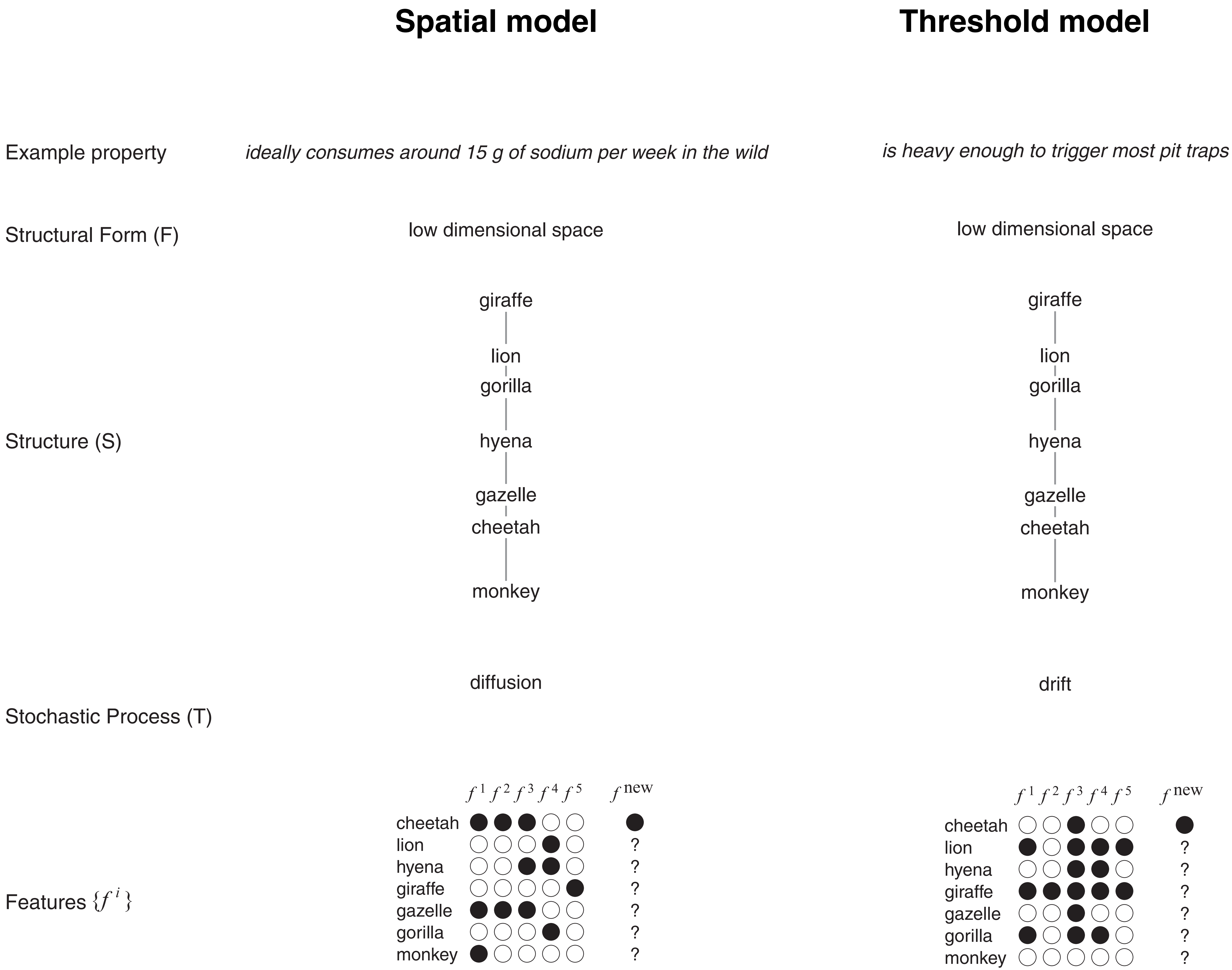
Threshold model

	f^1	f^2	f^3	f^4	f^5	f^{new}
cheetah	○	○	●	○	○	●
lion	●	○	●	●	●	?
hyena	○	○	●	●	○	?
giraffe	●	●	●	●	●	?
gazelle	○	○	●	○	○	?
gorilla	●	○	●	●	○	?
monkey	○	○	○	○	○	?

Causal model

	f^1	f^2	f^3	f^4	f^5	f^{new}
cheetah	○	○	○	●	●	●
lion	●	○	●	●	●	?
hyena	●	●	●	●	●	?
giraffe	○	○	●	○	○	?
gazelle	○	○	○	○	○	?
gorilla	○	●	○	○	○	?
monkey	○	○	○	●	○	?


Taxonomic model
Causal model
Example property
has enzyme X132 in its bloodstream
carries leptospirosis
Structural Form (F)
tree
directed graph
Structure (S)
cheetah
lion
hyena
giraffe
gazelle
gorilla
monkey
Stochastic Process (T)
diffusion
transmission
Features $\{f^i\}$
f^1 f^2 f^3 f^4 f^5 f^{new}
?



Spatial model
Threshold model
Example property
ideally consumes around 15 g of sodium per week in the wild
is heavy enough to trigger most pit traps
Structural Form (F)
low dimensional space
low dimensional space
Structure (S)
giraffe
lion
gorilla
hyena
gazelle
cheetah
monkey
giraffe
lion
gorilla
hyena
gazelle
cheetah
monkey
Stochastic Process (T)
diffusion
drift
Features $\{f^i\}$
f^1 f^2 f^3 f^4 f^5 f^{new}
cheetah
lion
hyena
giraffe
gazelle
gorilla
monkey
f^1 f^2 f^3 f^4 f^5 f^{new}
cheetah
lion
hyena
giraffe
gazelle
gorilla
monkey

Bayesian property induction

$f \in F$: **feature in set of all possible features**

$X = \{$**cheetah, monkey**$\}$: **set of premise categories**

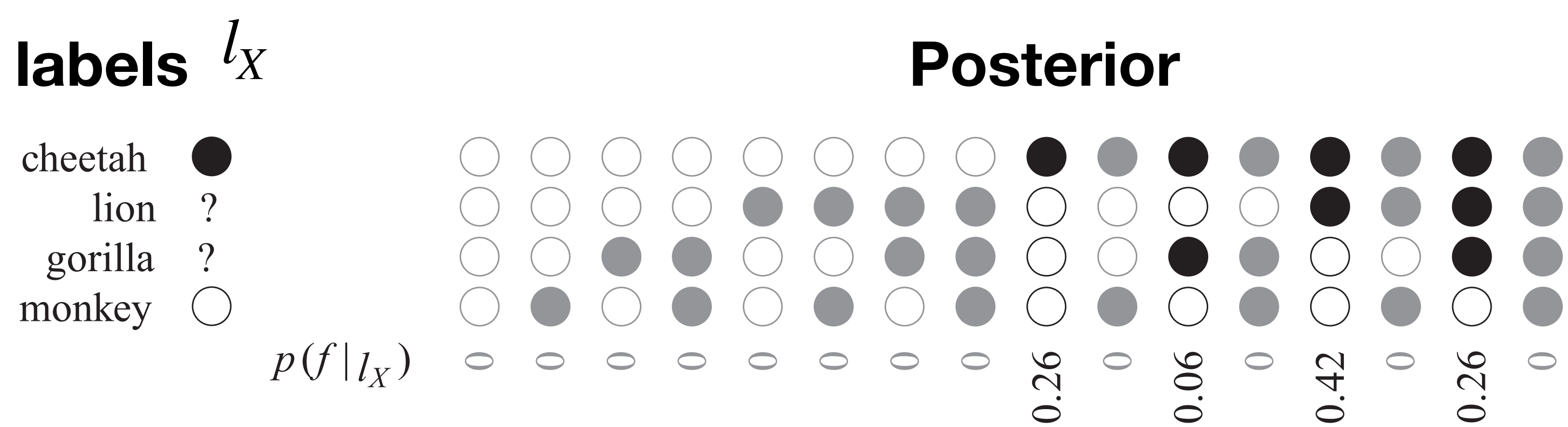
$Y = \{$**lion, gorilla**$\}$: **conclusion categories**

$l_X = \{1,0\}$: **feature labels for premise categories**

cheetahs have sesamoid bones.
<u>monkeys DO NOT have sesamoid bones.</u>
Do lions have sesamoid bones?
Do gorilla have sesamoid bones?

Posterior over features given evidence

$$P(f \mid l_X) = \frac{P(l_X|f)P(f)}{\sum_{f'} P(l_X|f')P(f')}$$



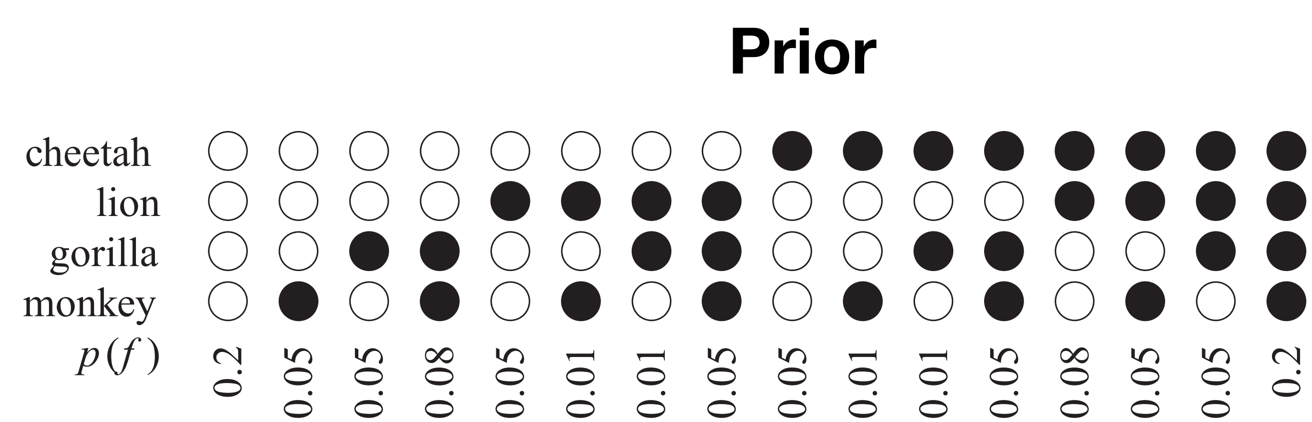
Likelihood

$$P(l_X|f) \propto \begin{cases} 1, & \text{if } f_X = l_X \\ 0, & \text{otherwise} \end{cases}$$

simple likelihood that just checks for consistency

Prior over features

$P(f)$ process that generates features / assigns them prior probability as an "intuitive theory"



Making Bayesian predictions

$f \in F$: **feature in set of all possible features**

$X = \{$**cheetah, monkey**$\}$: **set of premise categories**

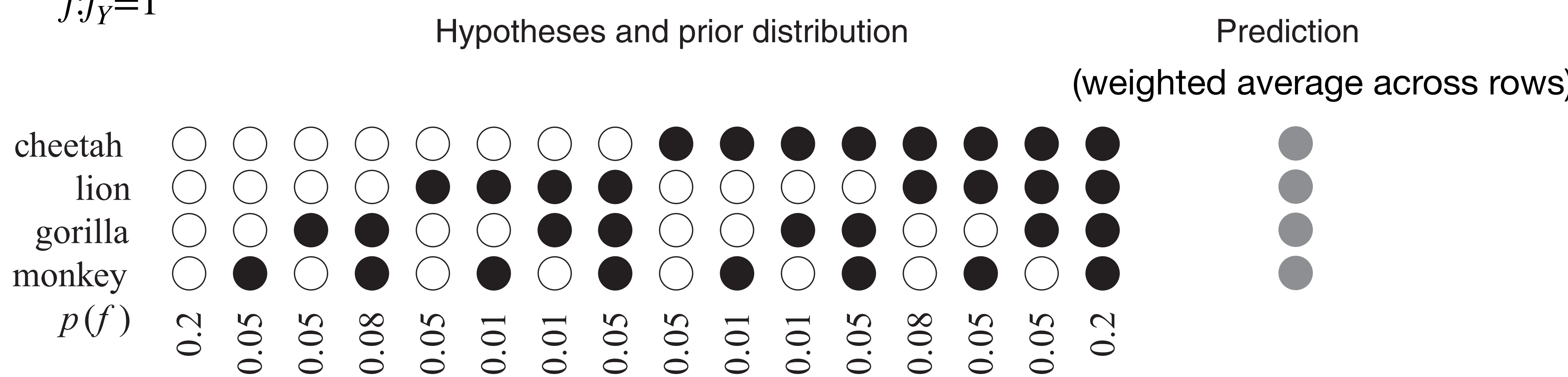
$Y = \{$**lion, gorilla**$\}$: **conclusion categories**

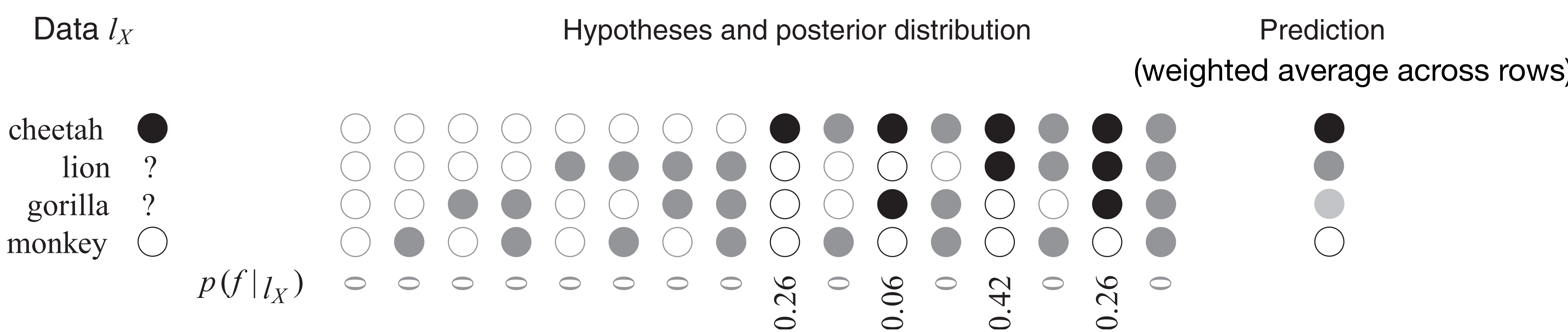
$l_X = \{1,0\}$: **feature labels for premise categories**

cheetahs have sesamoid bones.
<u>monkeys DO NOT have sesamoid bones.</u>
Do lions have sesamoid bones?
Do gorilla have sesamoid bones?

Posterior predictive distribution

$$P(f_Y = 1 \mid l_X) = \sum_{f: f_Y = 1} P(f \mid l_X)$$

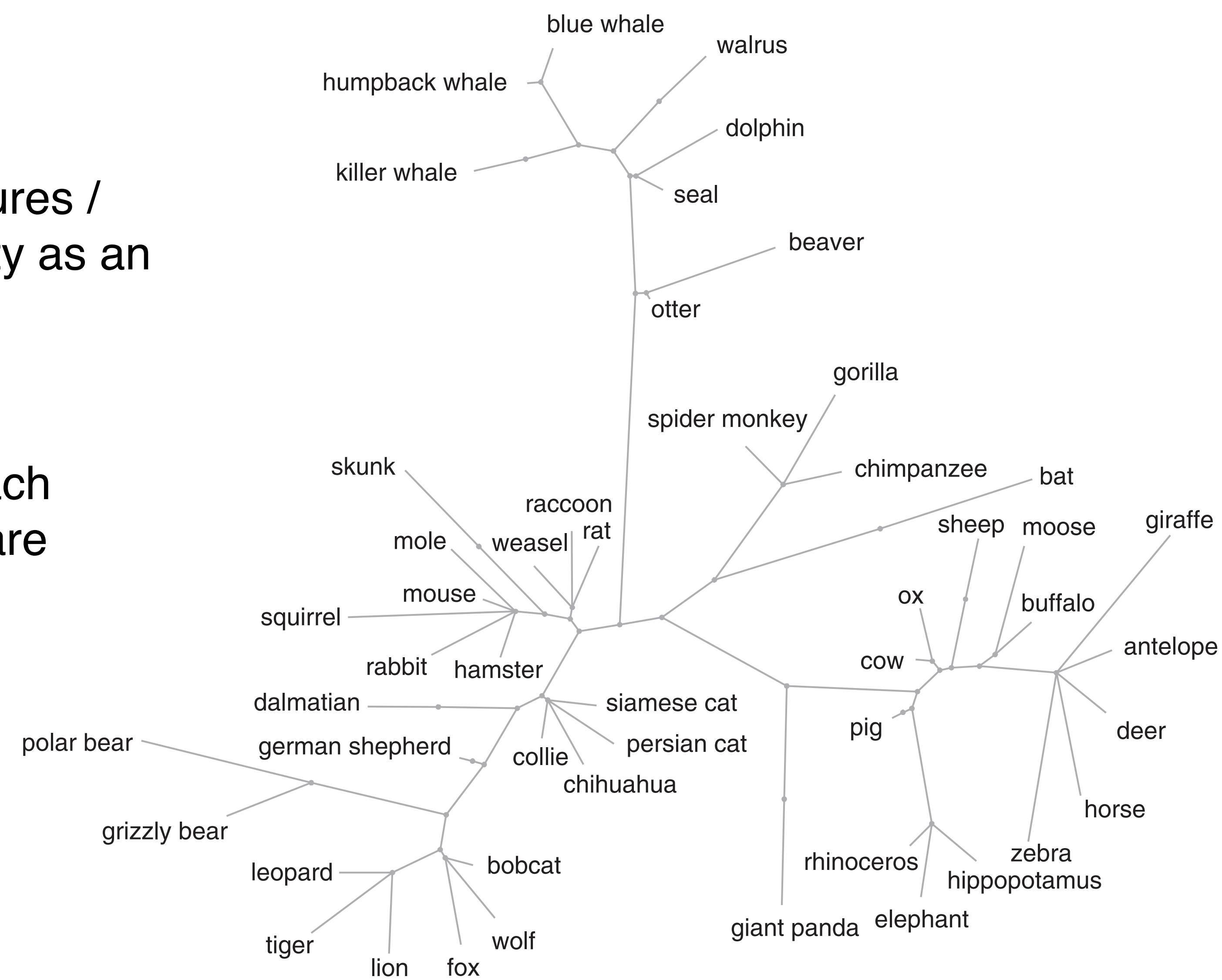




Taxonomic intuitive theory

Prior over features

$P(f)$ process that generates features / assigns them prior probability as an "intuitive theory"



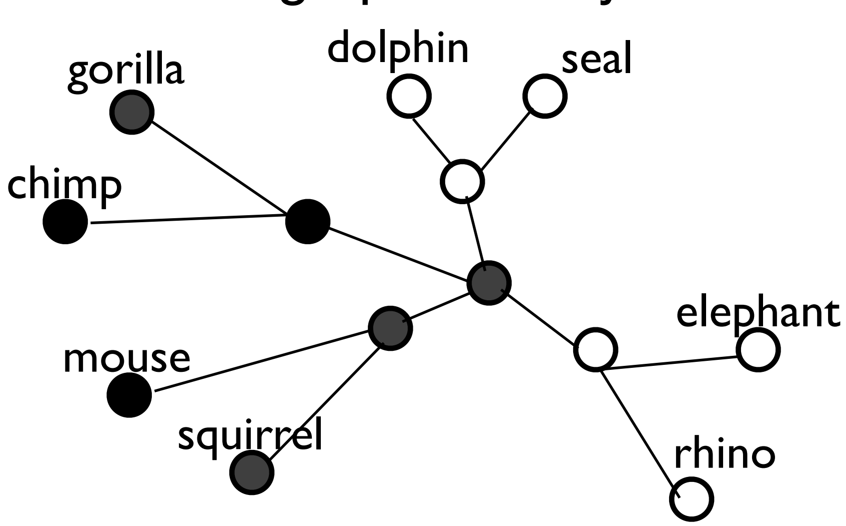
Naive enumeration of prior probability of each feature would require 2^{50} numbers if there are 50 mammals

$f \in F$: **feature**

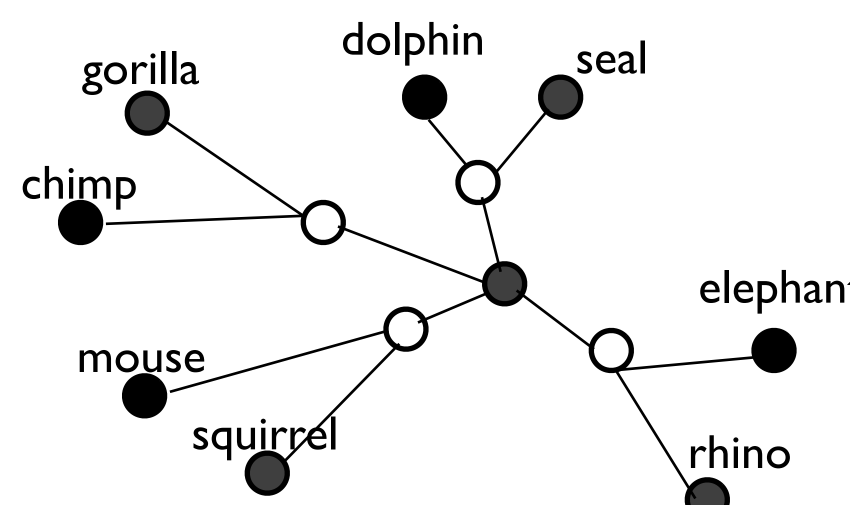
● on
○ off

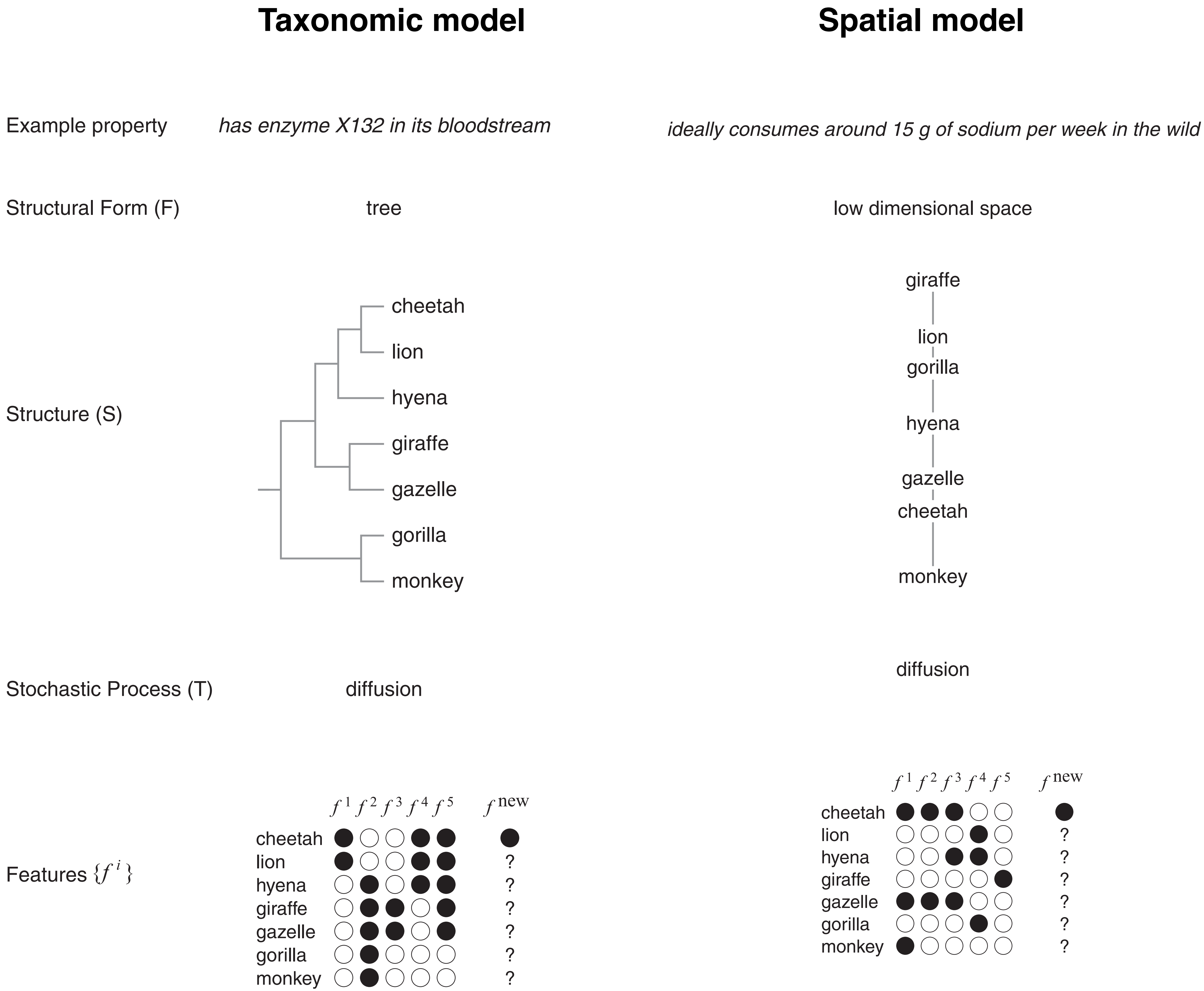
Graph with "diffusion process" encourages smooth features

High probability



Low probability



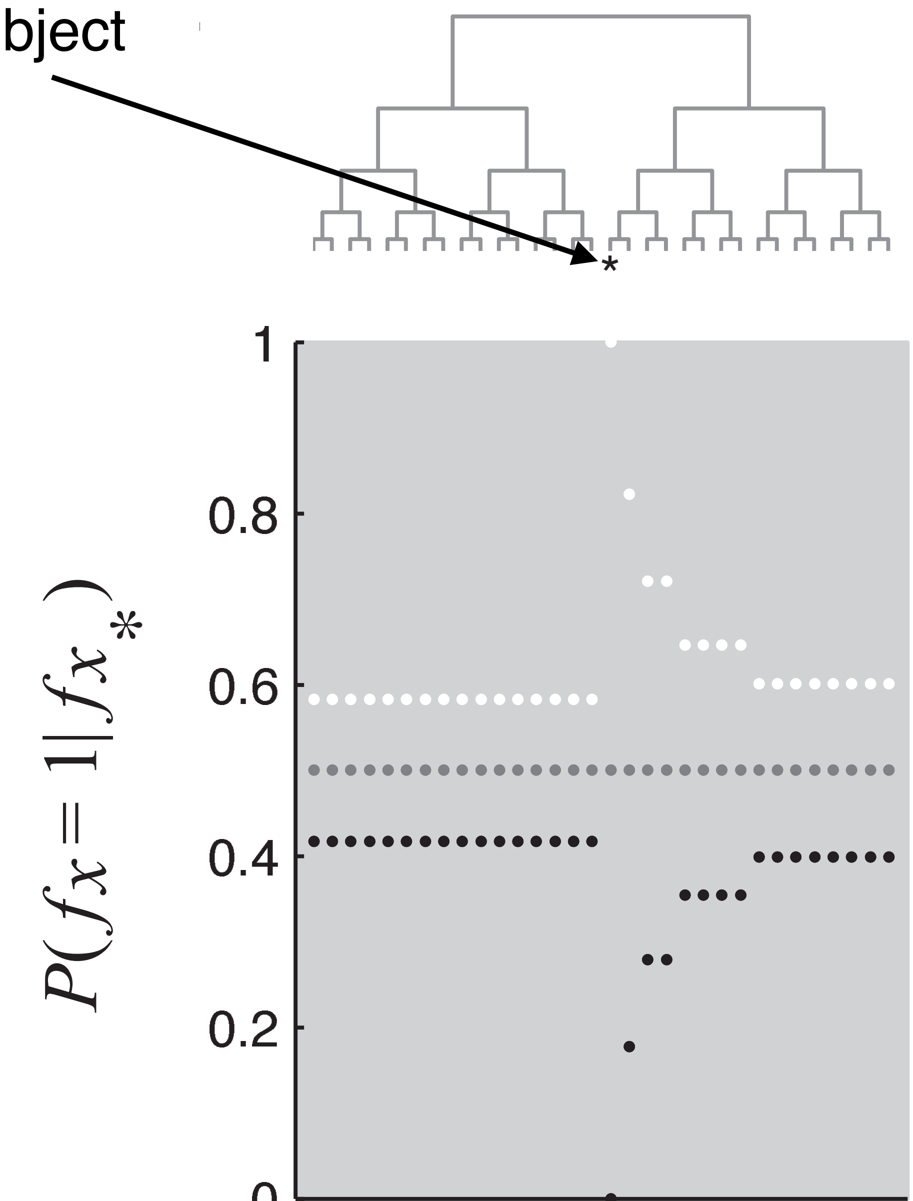

Taxonomic model
Spatial model
Example property
has enzyme X132 in its bloodstream
ideally consumes around 15 g of sodium per week in the wild
Structural Form (F)
tree
low dimensional space
Structure (S)
cheetah
lion
hyena
giraffe
gazelle
gorilla
monkey
giraffe
lion
gorilla
hyena
gazelle
cheetah
monkey
Stochastic Process (T)
diffusion
diffusion
Features $\{f^i\}$
f^1 f^2 f^3 f^4 f^5 f^{new}
cheetah
lion
hyena
giraffe
gazelle
gorilla
monkey
?
f^1 f^2 f^3 f^4 f^5 f^{new}
cheetah
lion
hyena
giraffe
gazelle
gorilla
monkey
?

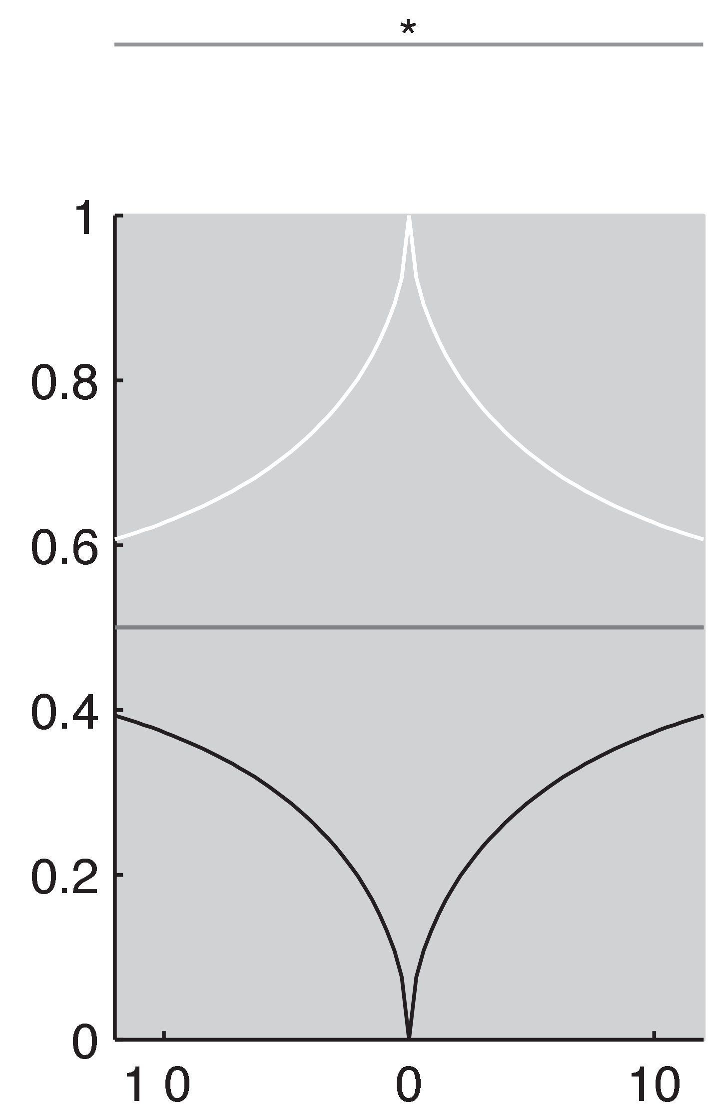
Making Bayesian predictions

$$P(f_Y = 1 \mid l_X) = \sum_{f:f_Y=1} P(f \mid l_X)$$

	Taxonomic model	**Spatial model**
Example property	*has enzyme X132 in its bloodstream*	*ideally consumes around 15 g of sodium per week in the wild*
Structural Form (F)	tree	low dimensional space

Observed object





white: if * is observed to be true
black: if * is observed to be false

Biological reasoning about animals

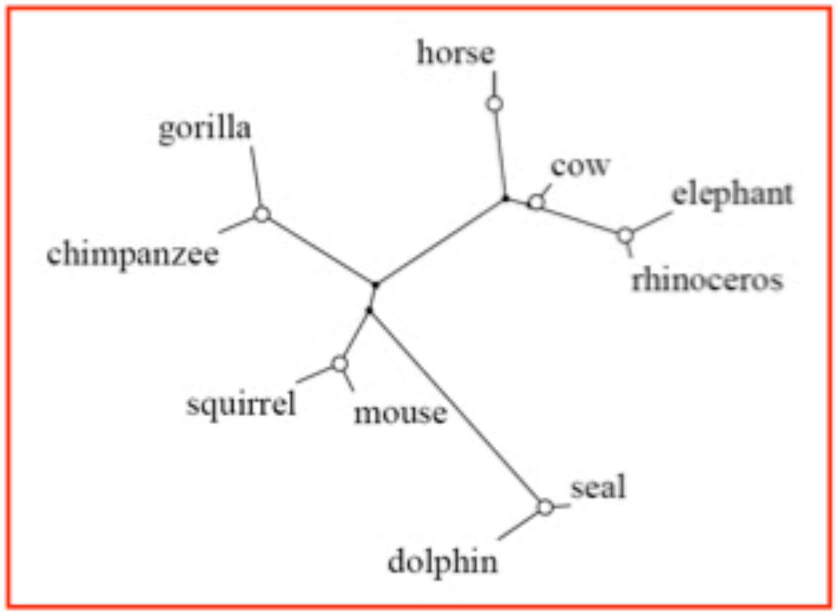
A tree fits better than a 2D space

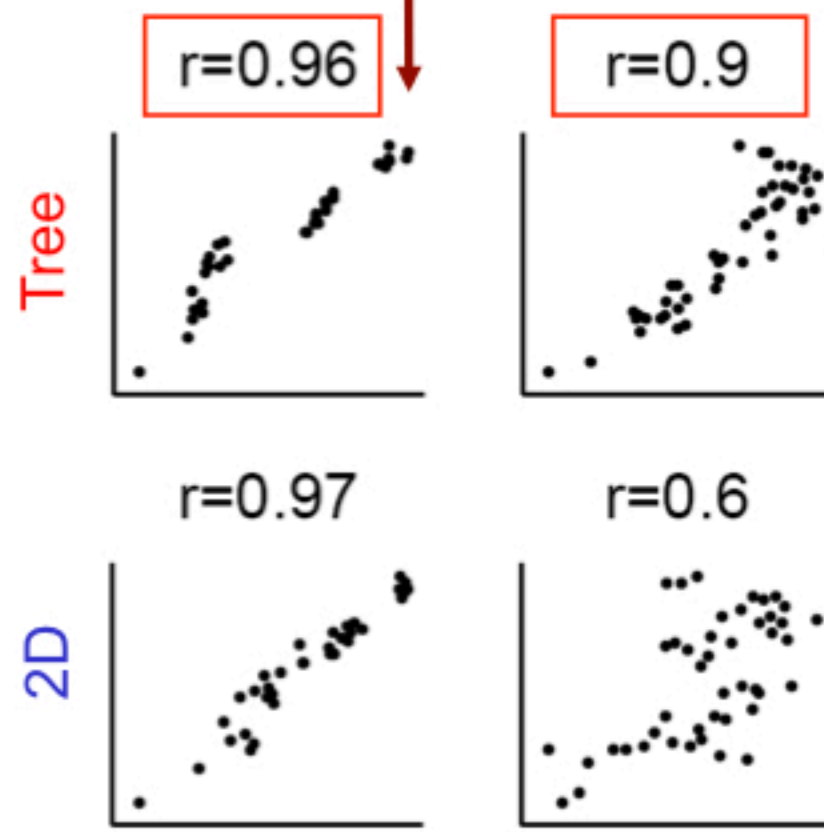
“Osherson horse” dataset

Cows have property P.
Elephants have property P.

Horses have property P.

Correlation of human participant judgments with model judgements





Evaluated across a range of difference premise/conclusion combinations .

“Osherson mammals” dataset

Gorillas have property P.
Mice have property P.
Seals have property P.

All mammals have property P.

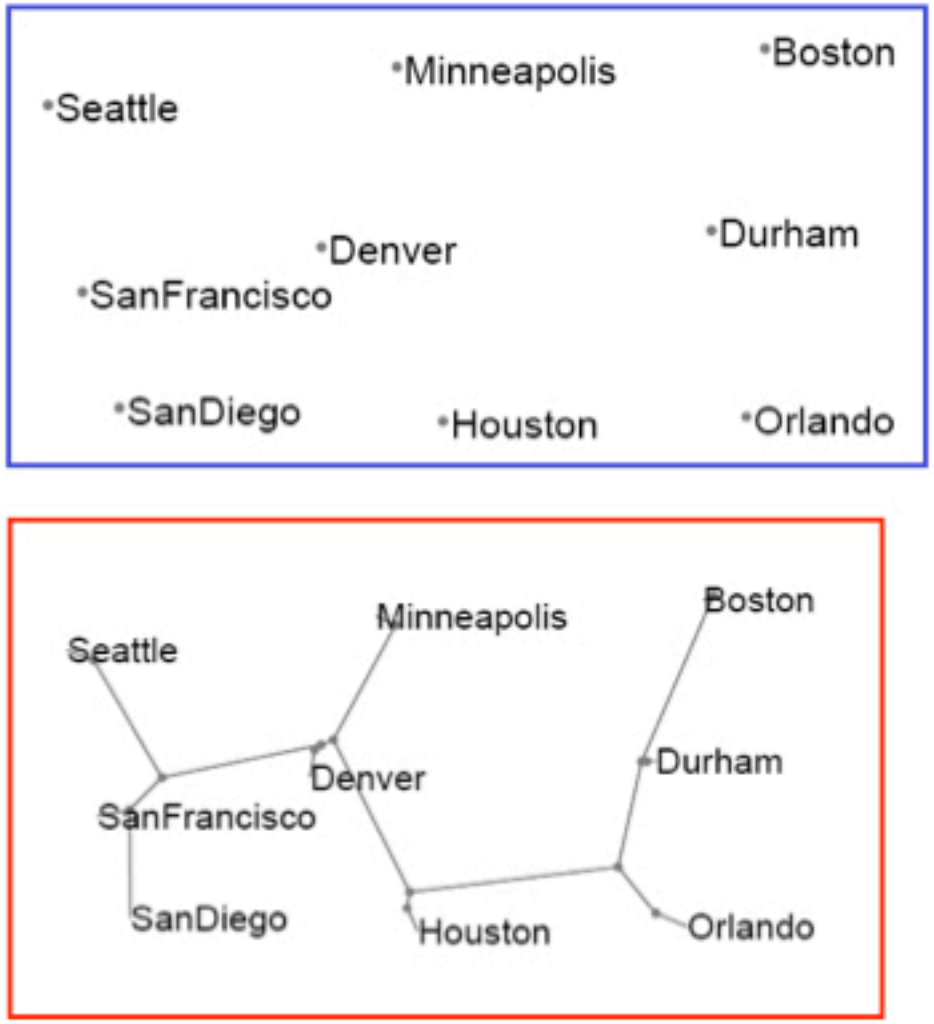
Reminder: Osheron’s Similiarity-coverage model achieved r=0.96 and r=0.87, respectively

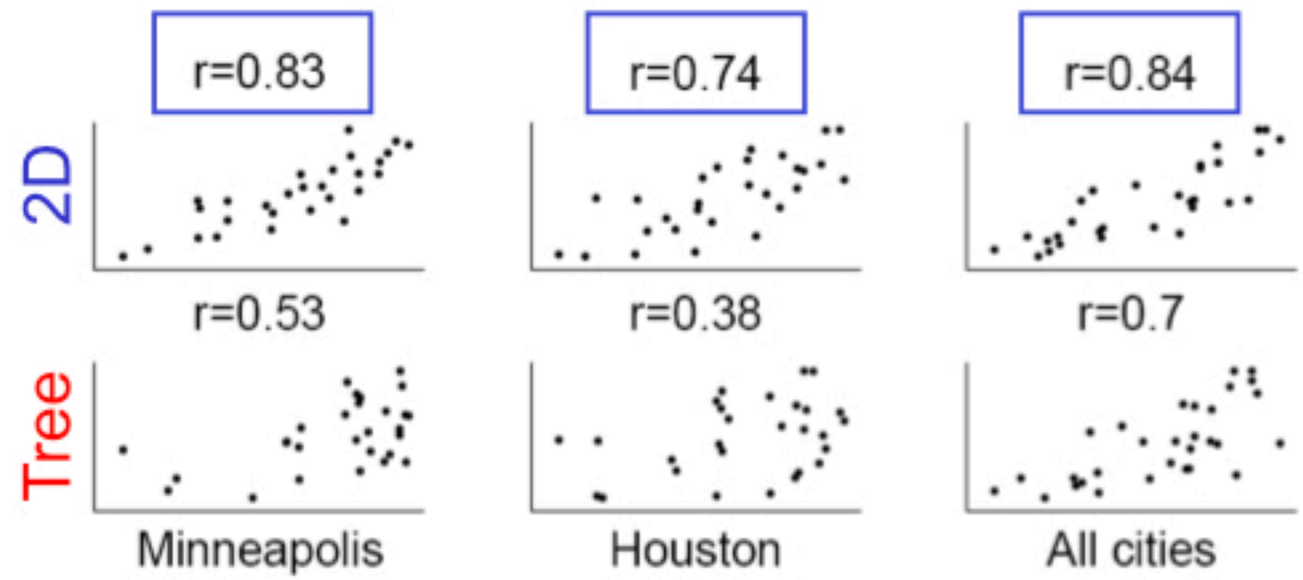
slide credit: Josh Tenenbaum

Spatial reasoning about cities

A 2D space fits better than a tree

"Given that a certain kind of native American artifact has been found in sites near city X, how likely is the same artifact to be found near city Y?"



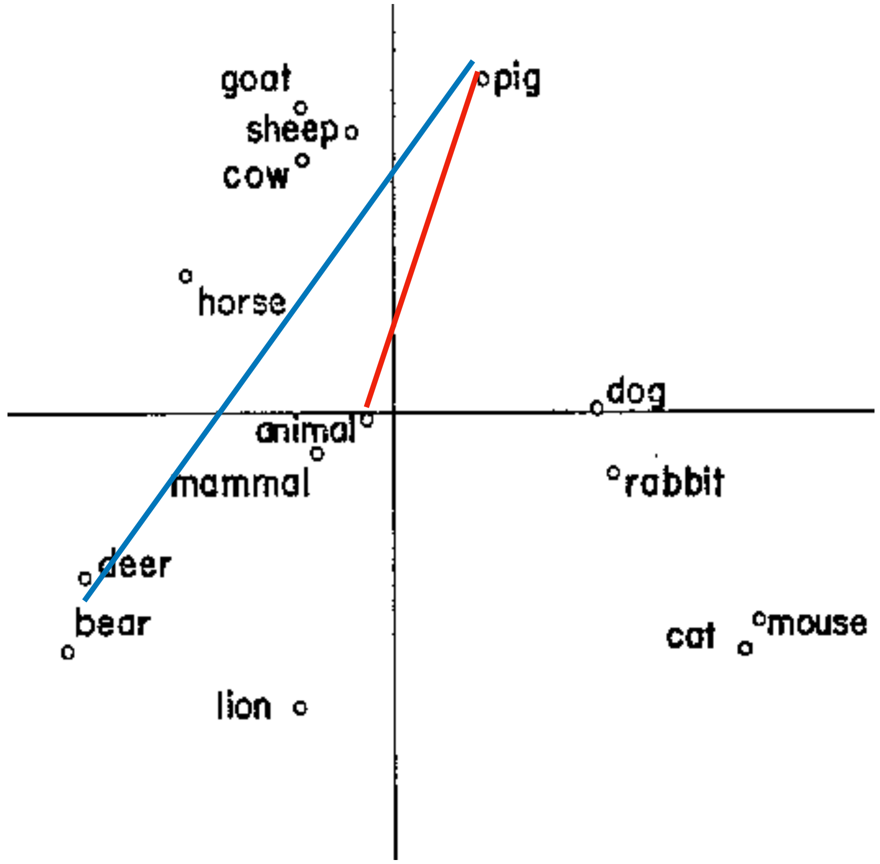


slide credit: Josh Tenenbaum

Reminder: Rips's (1975) model was spatial

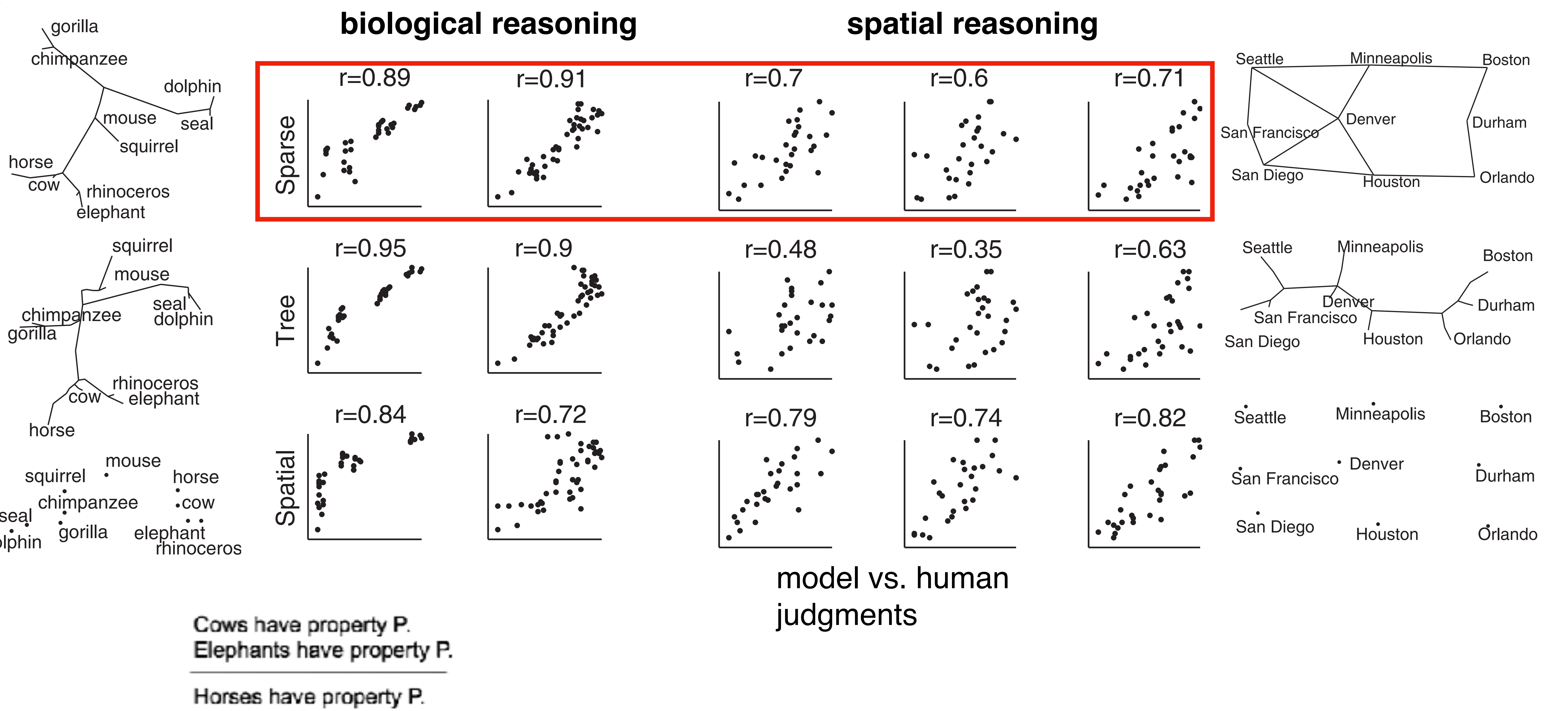
"If pigs have a disease, what proportion of deer would be likely to get the disease?"

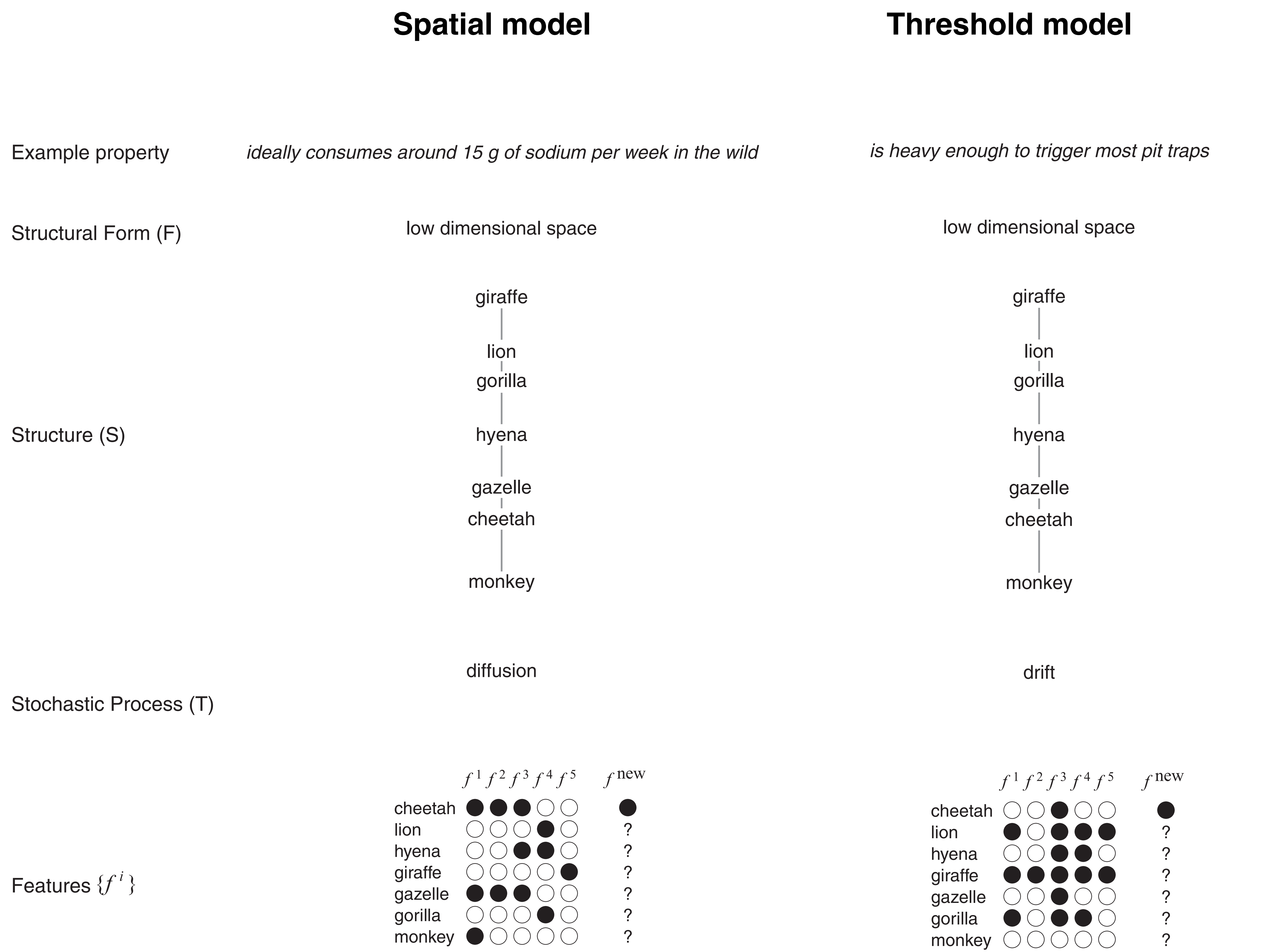
- Subjects' answers seem to rely on two factors:
  - Similarity of the premise to conclusion category (smaller distance is stronger argument)
  - Typicality of premise category (smaller distance in MDS space from premise to superordinate category leads to strong argument)



Important caveat: Do we need special purpose knowledge structures to explain these inferences?

Lake et al. (2018; *Cognitive Science*) show that learning more generic, sparse structures can account for both taxonomic and spatial reasoning





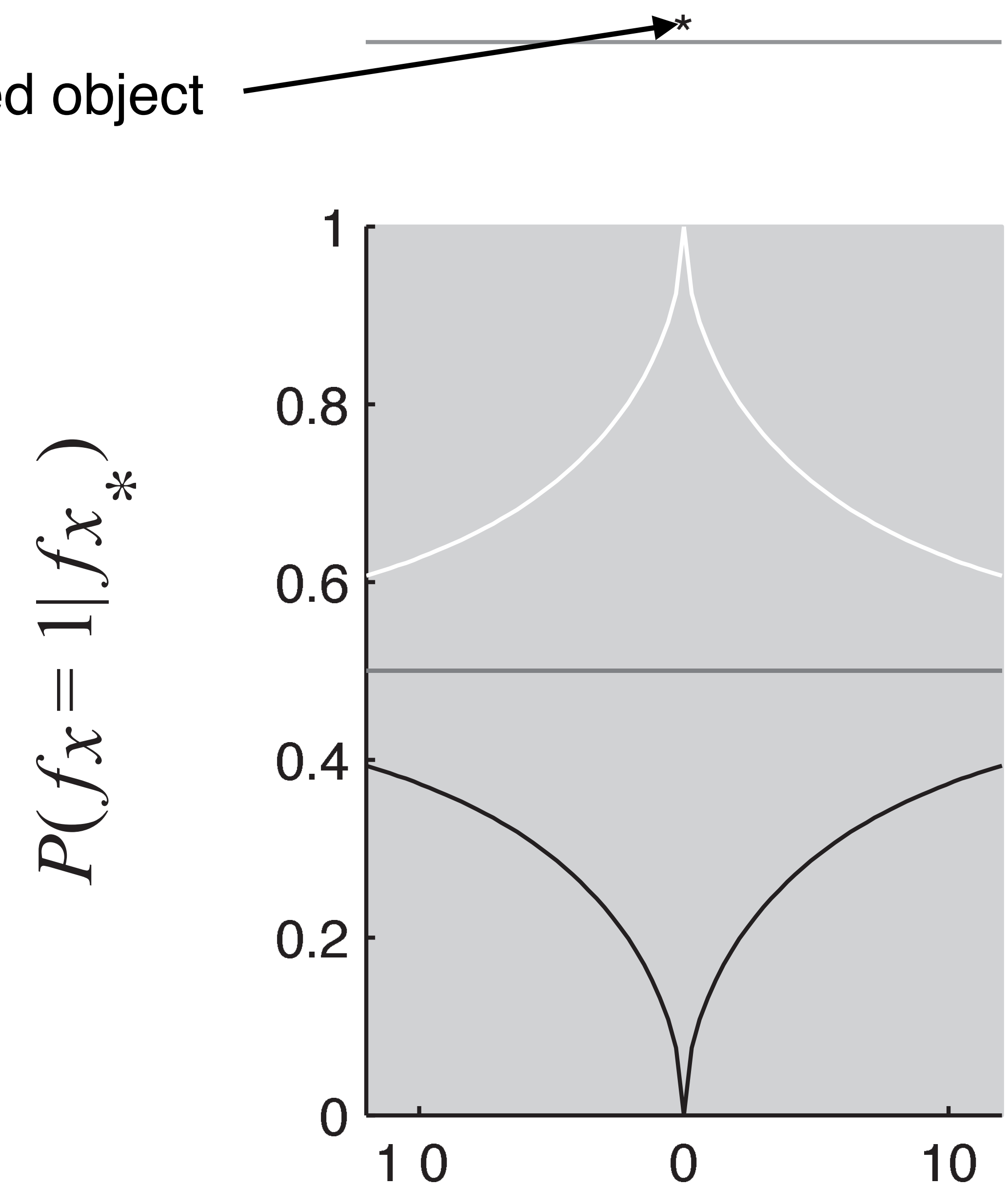
Spatial model
Threshold model
Example property
ideally consumes around 15 g of sodium per week in the wild
is heavy enough to trigger most pit traps
Structural Form (F)
low dimensional space
low dimensional space
Structure (S)
giraffe
lion
gorilla
hyena
gazelle
cheetah
monkey
giraffe
lion
gorilla
hyena
gazelle
cheetah
monkey
Stochastic Process (T)
diffusion
drift
Features $\{f^i\}$
f^1 f^2 f^3 f^4 f^5 f^{new}
cheetah
lion
hyena
giraffe
gazelle
gorilla
monkey
f^1 f^2 f^3 f^4 f^5 f^{new}
cheetah
lion
hyena
giraffe
gazelle
gorilla
monkey

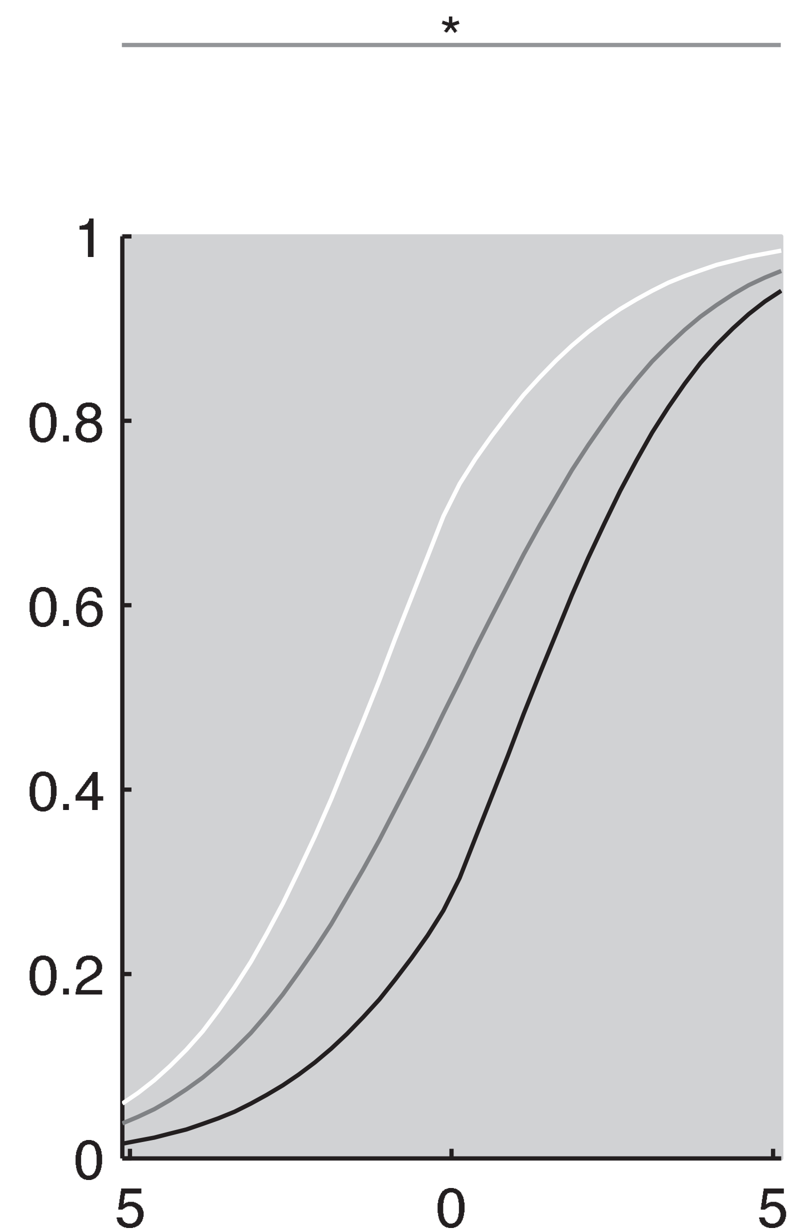
Making Bayesian predictions

$$P(f_Y = 1 \mid l_X) = \sum_{f:f_Y=1} P(f \mid l_X)$$

	Spatial model	**Threshold model**
Example property	*ideally consumes around 15 g of sodium per week in the wild*	*is heavy enough to trigger most pit traps*
process	**diffusion**	**drift**

Observed object



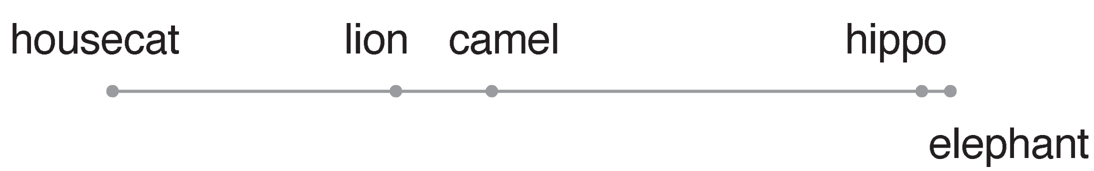


white: if * is observed to be true
black: if * is observed to be false

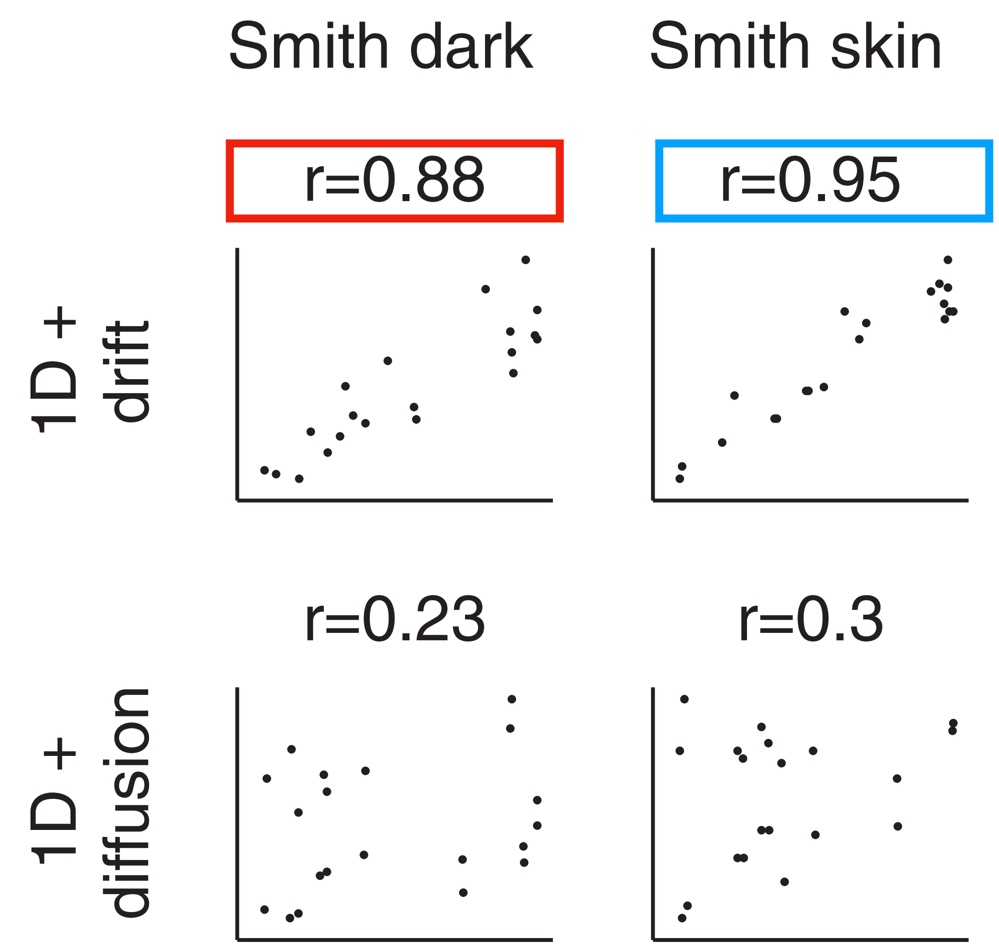
Reasoning about threshold properties

A drift process fits better than diffusion

"Given that animal X has a visual system that fully adapts to darkness in less than 5 minutes, how likely does Y?"



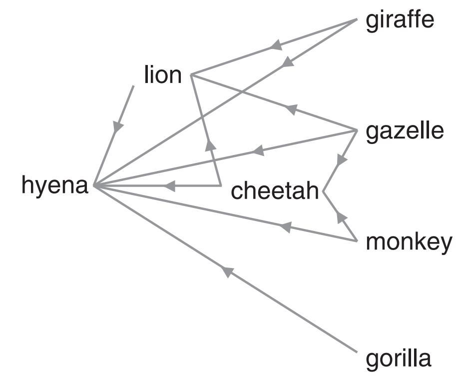
"Given that animal X has skin that is more resistant to penetration than most synthetic fibers, how likely does Y?"



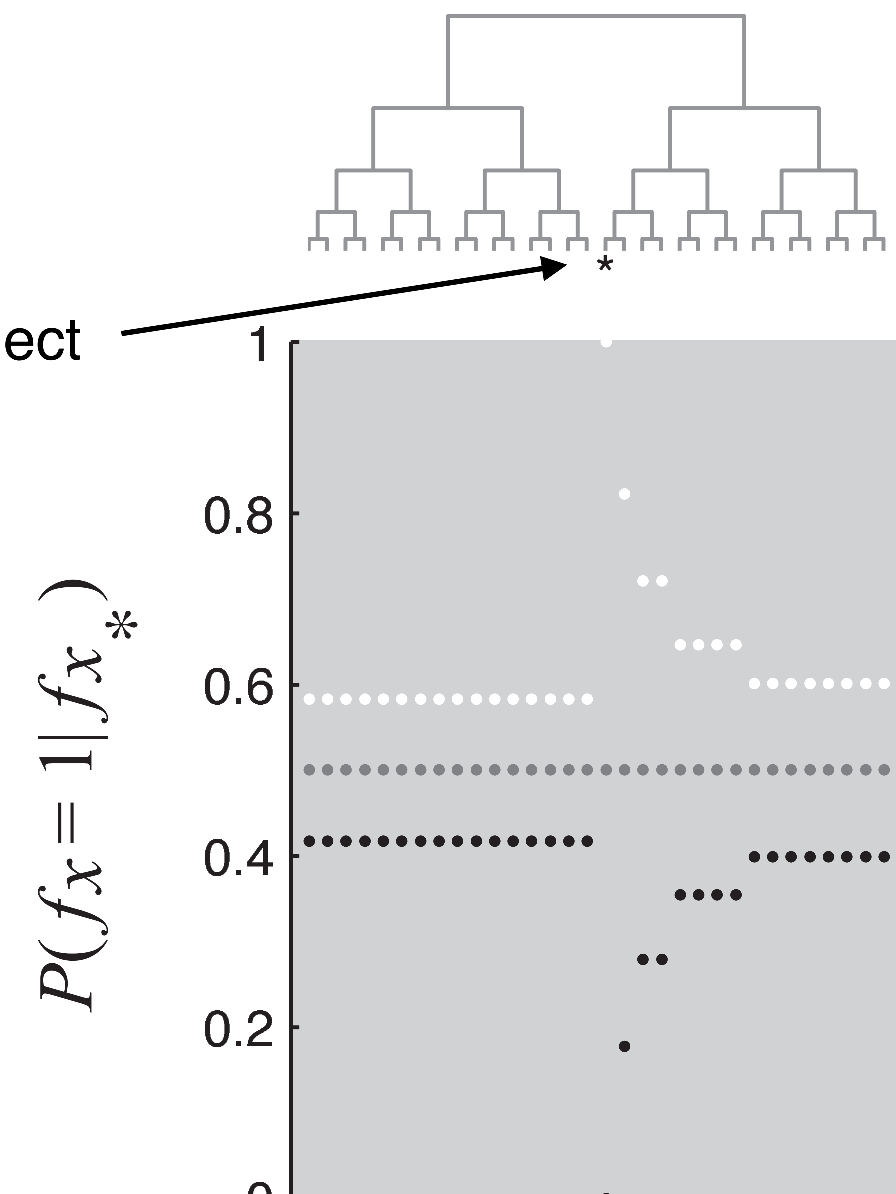
Making Bayesian predictions

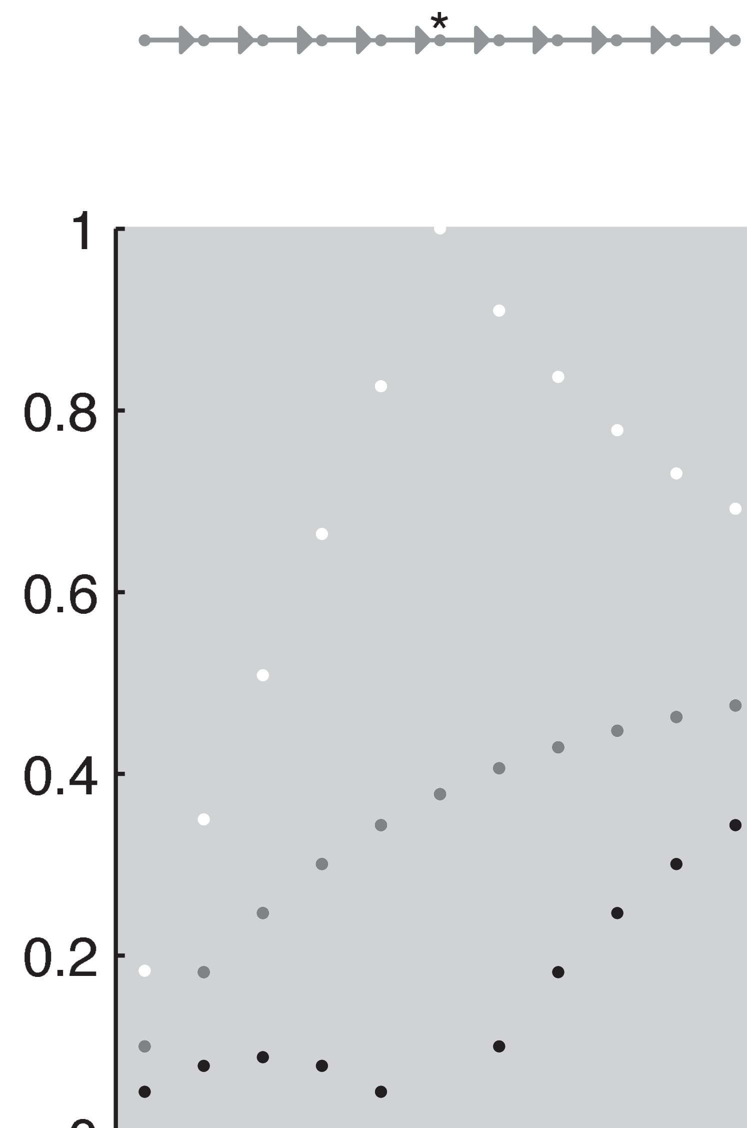
$$P(f_Y = 1 \mid l_X) = \sum_{f:f_Y=1} P(f \mid l_X)$$

	Taxonomic model	**Causal model**
Example property	*has enzyme X132 in its bloodstream*	*carries leptospirosis*
Structural Form (F)	tree	directed graph



Observed object





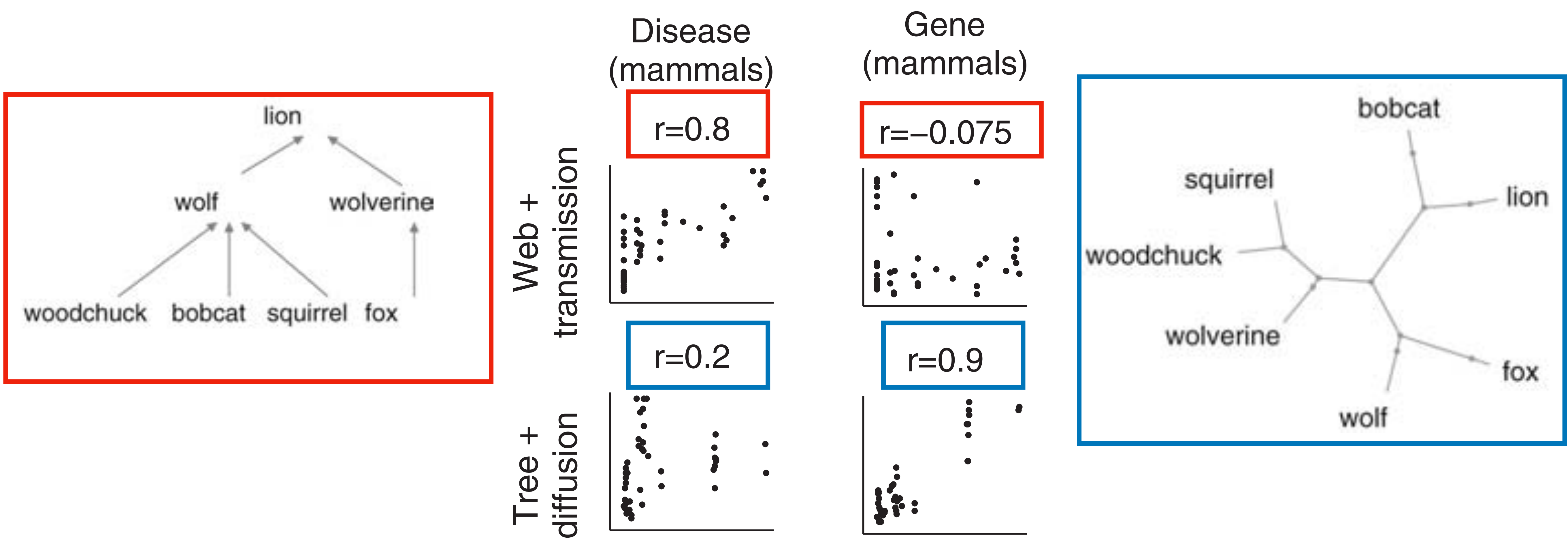
white: if * is observed to be true
black: if * is observed to be false

Reason about causal transmission

Another double dissociation: A causal web fits better than a tree for diseases, and vice versa for biological properties

"Assuming that animal X has disease D, how likely is animal Y?"

"Assuming that animal X has Gene XR-23, how likely is animal Y?"



Summary: Structured statistical models of inductive reasoning

- Everyday inductive inferences are guided by rich background knowledge and intuitive theories, but rarely do models attempt to formalize the content of intuitive theories
- Kemp and Tenenbaum present a “Bayesian property induction” framework that can capture different types of knowledge as different structures and stochastic processes operating over these structures
- Having the right structure AND the right process are crucial to make good inferences

Anderson's (1991) Bayesian Proposal for Predictions with Uncertain Categorization

Psychological Review
1991, Vol. 98, No. 3, 409–429

0033-295X/91/$3.00

The Adaptive Nature of Human Categorization

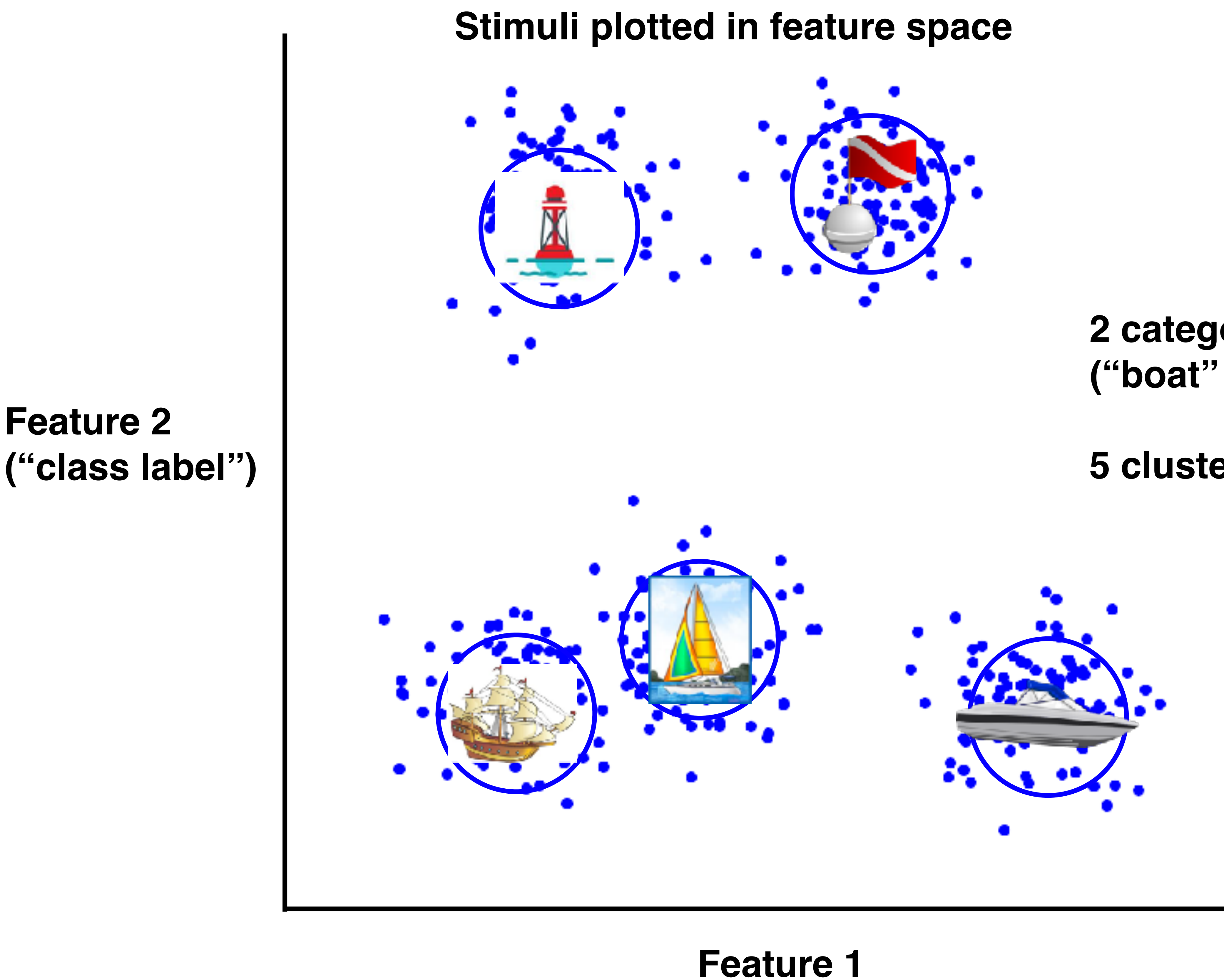
John R. Anderson
Carnegie Mellon University

A rational model of human categorization behavior is presented that assumes that categorization reflects the derivation of optimal estimates of the probability of unseen features of objects. A Bayesian analysis is performed of what optimal estimations would be if categories formed a disjoint partitioning of the object space and if features were independently displayed within a category. This Bayesian analysis is placed within an incremental categorization algorithm. The resulting rational model accounts for effects of central tendency of categories, effects of specific instances, learning of linearly nonseparable categories, effects of category labels, extraction of basic level categories, base-rate effects, probability matching in categorization, and trial-by-trial learning functions. Although the rational model considers just 1 level of categorization, it is shown how predictions can be enhanced by considering higher and lower levels. Considering prediction at the lower, individual level allows integration of this rational analysis of categorization with the earlier rational analysis of memory (Anderson & Milson, 1989).

Anderson (1990) presented a rational analysis of human cognition. The term *rational* derives from similar "rational-man" analyses in economics. Rational analyses in other fields are sometimes called *adaptationist analyses.* Basically, they are efforts to explain the behavior in some domain on the assumption that the behavior is optimized with respect to some criteria of adaptive importance. This article begins with a general characterization of how one develops a rational theory of a particular cognitive phenomenon. Then I present the basic theory of categorization developed in Anderson (1990) and review the steps involved in a research program that attempts to understand cognition in terms of its adaptation to the environment:

1. The first task is to specify what the system is trying to optimize. Perhaps such models are ultimately to be justified in terms of maximizing some evolutionary criterion like number of surviving offspring. However, this is not a very workable criterion in most applications. Thus, economics uses wealth as the variable to be optimized; optimal foraging theory (Stephens & Krebs, 1986) often uses caloric intake; and the rational theory of memory (Anderson & Milson, 1989) uses retrieval of

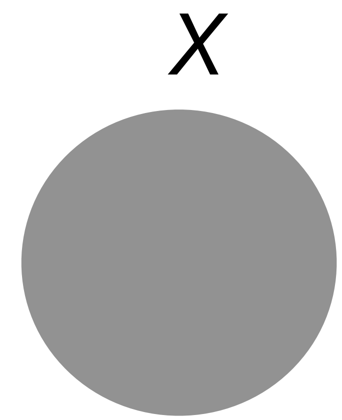
Reminder: Anderson’s Rational model



2 categories
(“boat” vs. “buoy”)

5 clusters

The Rational model's formula for predictions with uncertain categorization



Suppose X is an unknown fruit (with some observed properties), and you want to predict whether or not it is sweet:

$$P(X\ \textbf{is sweet}) = \sum_{C} P(\textbf{sweet} \mid C) P(C \mid X)$$

is the proper Bayesian thing to do, if you don't know the right category C

Also, note the connection with how predictions are made in Kemp & Tenenbaum

$$P(f_Y = 1 \mid l_X) = \sum_{f : f_Y = 1} P(f \mid l_X)$$

$f \in F$: **feature in set of all possible features**

$Y = \{\textbf{lion}, \textbf{gorilla}\}$: **conclusion categories**

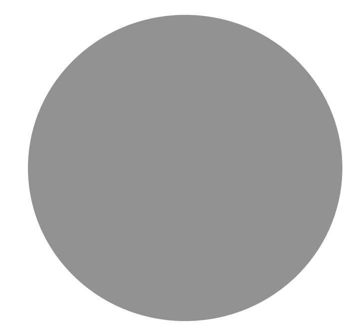
$l_X = \{1,0\}$: **feature labels for premise categories**

The Rational model's formula for predictions with uncertain categorization

conditional probabilities

P(apple|X)=.70 P(sweet|apple)=.50
P(pear|X)=.30 P(sweet|pear)=.90

X



making a Bayesian prediction

$$P(X\ \textbf{is sweet}) = \sum_{C} P(\textbf{sweet} \,|\, C)P(C \,|\, X)$$

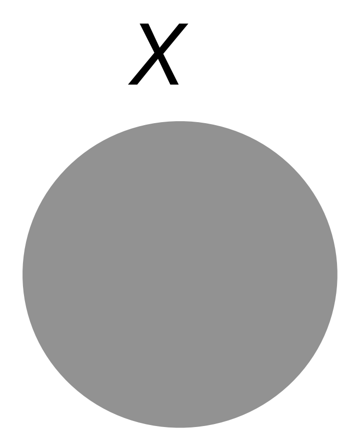
$$= P(\textbf{sweet} \,|\, \textbf{apple})P(\textbf{apple} \,|\, X) + P(\textbf{sweet} \,|\, \textbf{pear})P(\textbf{pear} \,|\, X)$$

$$= .50(.70) + .90(.30) = .62$$

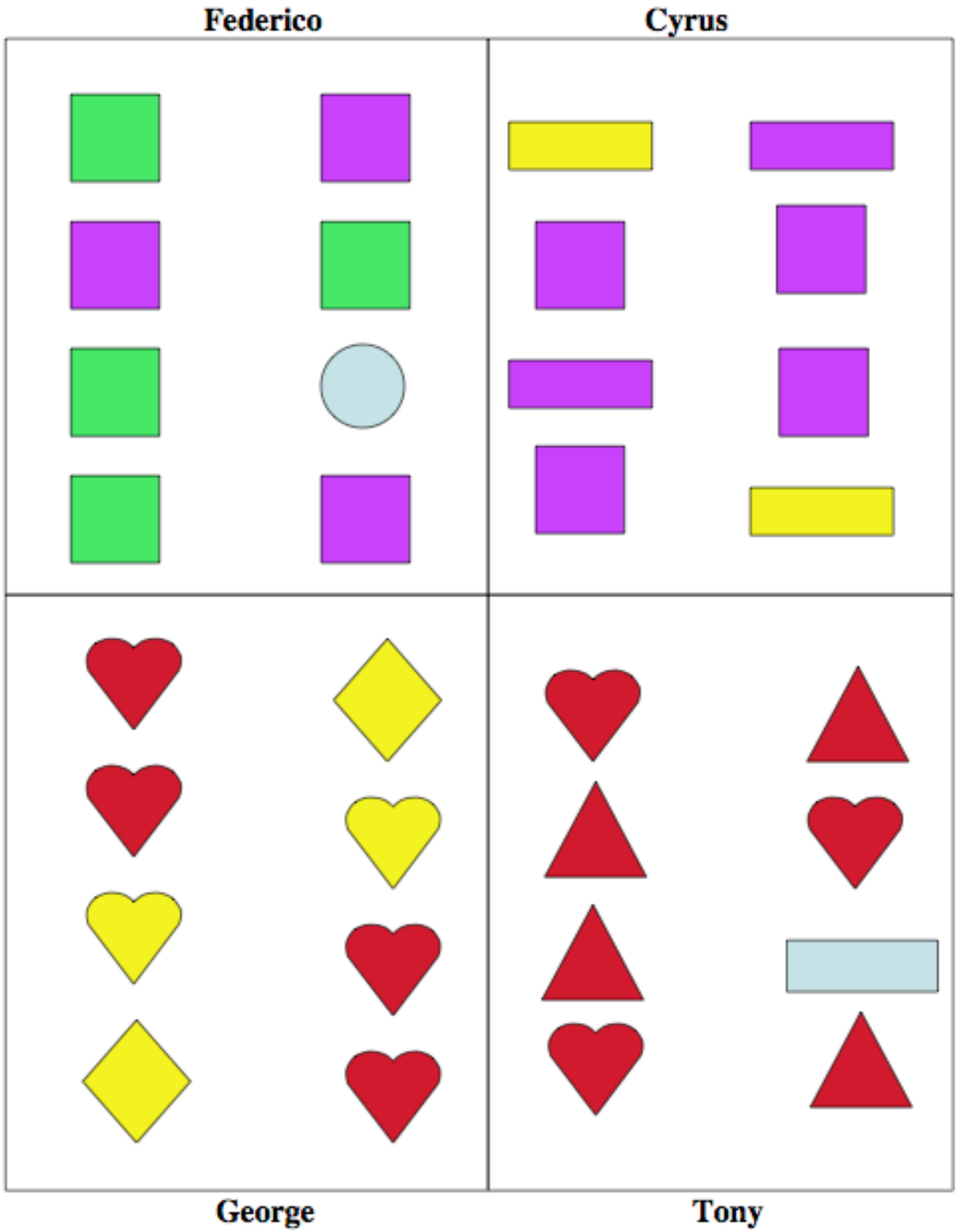
Note that if you had only used the figure for apples, ignoring that uncertainty of categorization, you would have estimated a .50 probability that X is sweet.

Questions arising from Anderson's model

- Will people actually use multiple categories in making these inductive predictions?
- Murphy & Ross (1994) were skeptical that people are doing this… and a similar experiment from Murphy & Ross (2010) is on the next slide



$$P(X\ \textbf{is sweet}) = \sum_{C} P(\textbf{sweet} \mid C)P(C \mid X)$$



(Murphy & Ross, 2010)

Consider a new red drawing.
Who most likely drew it?
Probability?
What shape is it likely to have?

Normative answer for: What shape is it likely to have?

$$P(\textbf{shape} = \textbf{heart} \mid \textbf{color} = \textbf{red}) =$$

$$\sum_C P(\textbf{shape} = \textbf{heart} \mid C) P(C \mid \textbf{color} = \textbf{red})$$

$$= P(\textbf{shape} = \textbf{heart} \mid \textbf{George}) P(\textbf{George} \mid \textbf{color} = \textbf{red})$$
$$+ P(\textbf{shape} = \textbf{heart} \mid \textbf{Tony}) P(\textbf{Tony} \mid \textbf{color} = \textbf{red})$$

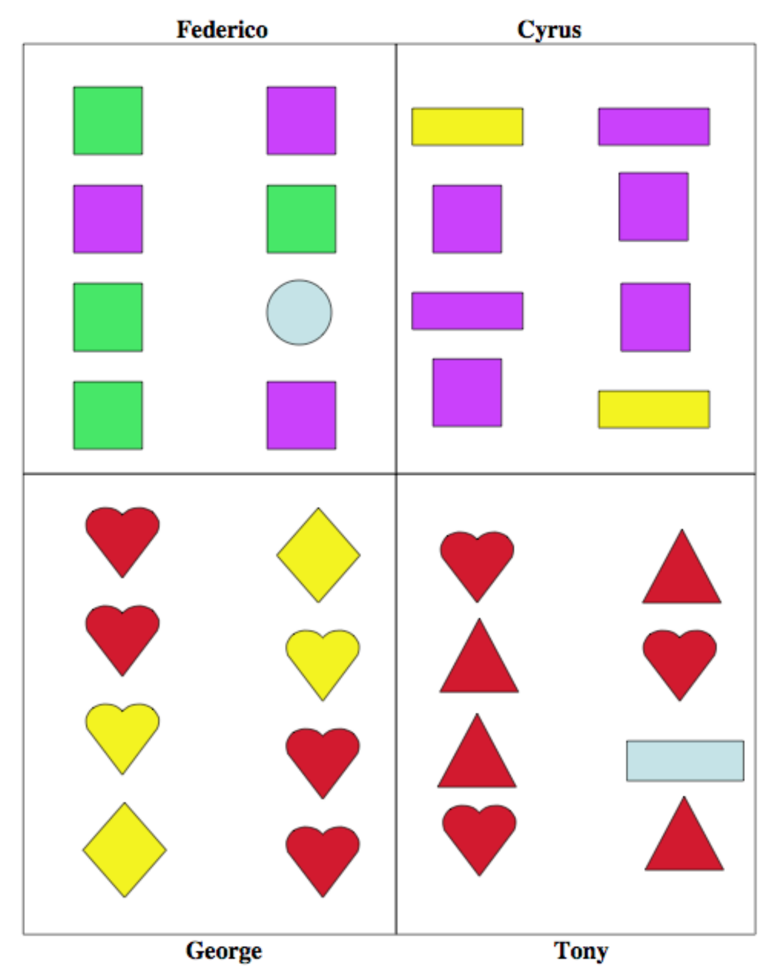
Murphy and Ross (2010) Results

- We can classify individual subjects by strategy in this design
  - Only 30% of responses used multiple categories
  - 22 subjects consistently focused on single category (answer: triangle)
  - 7 always used multiple categories, as the rational model predicts (answer: heart)
  - Others were mixed

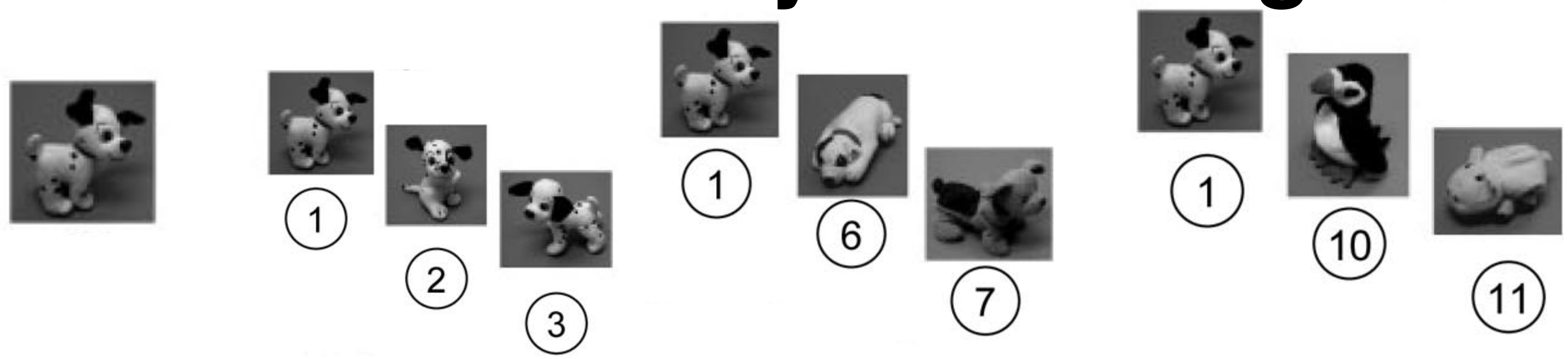
Normative answer for: What shape is it likely to have?

$$P(\textbf{shape} = \textbf{heart} \mid \textbf{color} = \textbf{red}) =$$

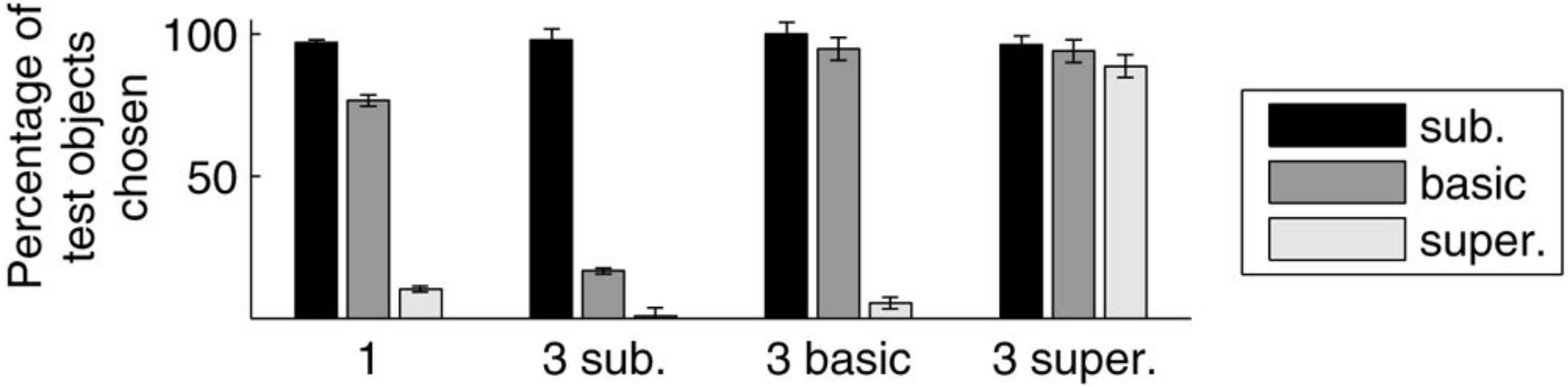
$$\sum_C P(\textbf{shape} = \textbf{heart} \mid C) P(C \mid \textbf{color} = \textbf{red})$$



Murphy and Ross find that people aren't Bayesian. But if people only consider the most likely category, how do we explain success of other Bayesian categorization models?

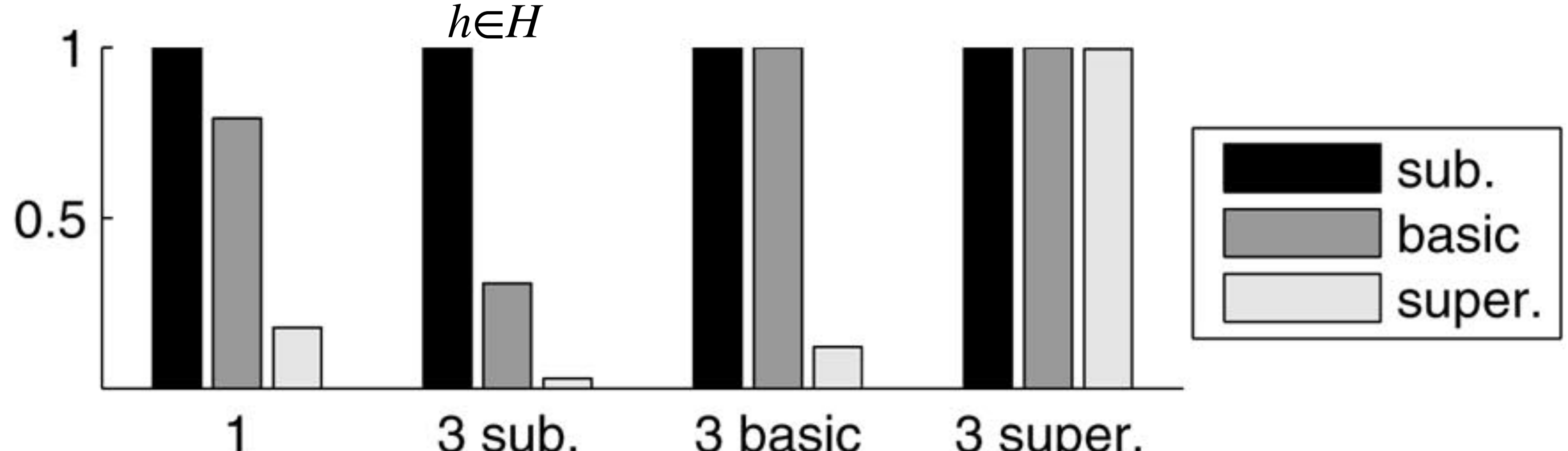


Adults



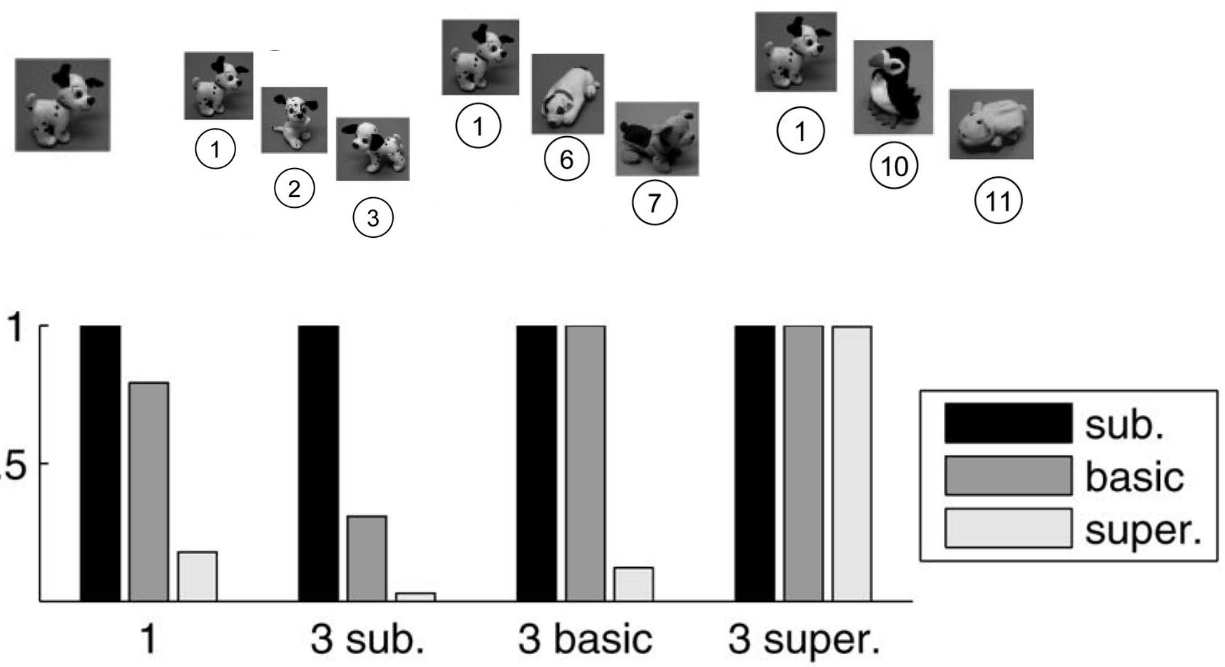
Bayesian model

$$p(y \in C | X) = \sum_{h \in H} P(y \in C | h) p(h | X)$$



Review: Most likely hypothesis doesn't cut it

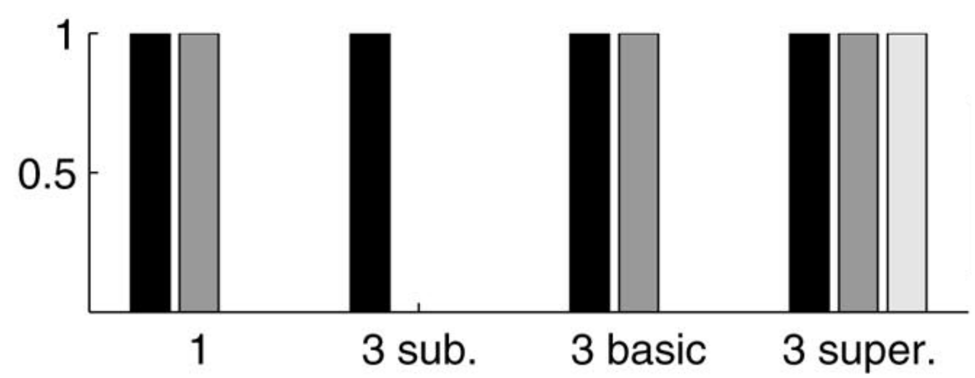
Full Bayesian model



Generalizing to a new example *y*

$$p(y \in C \mid X) = \sum_{h \in H} P(y \in C \mid h) p(h \mid X)$$

Most likely hypothesis only



Generalizing to a new example *y*

$$h^* = \mathbf{argmax}_{h \in H} \, P(h \mid X)$$

$$p(y \in C \mid X) = P(y \in C \mid h^*)$$

generalization is too sharp

Review: Making Bayesian predictions

$f \in F$: **feature in set of all possible features**

$X = \{$**cheetah**, **monkey**$\}$: **set of premise categories**

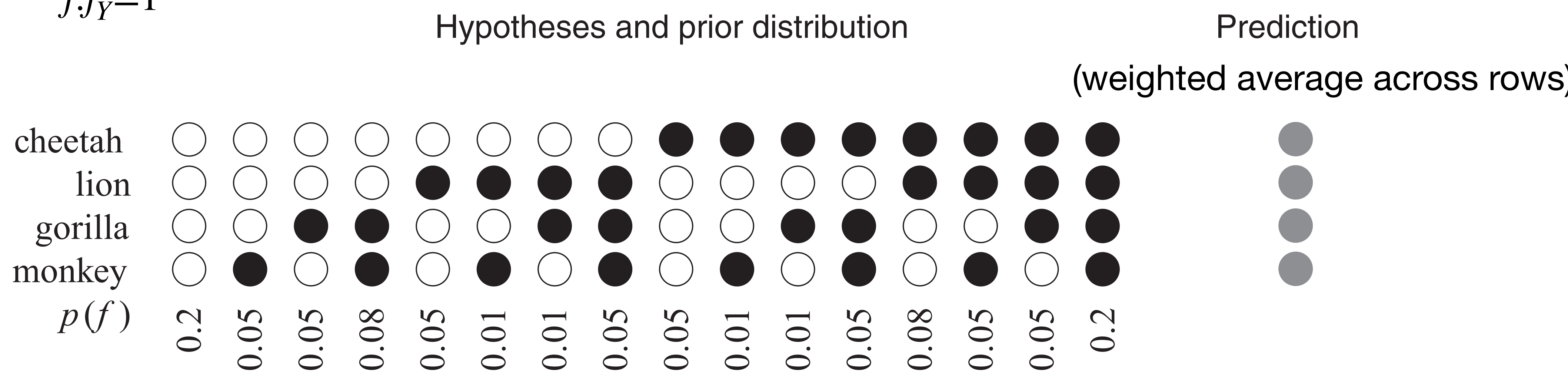
$Y = \{$**lion**, **gorilla**$\}$: **conclusion categories**

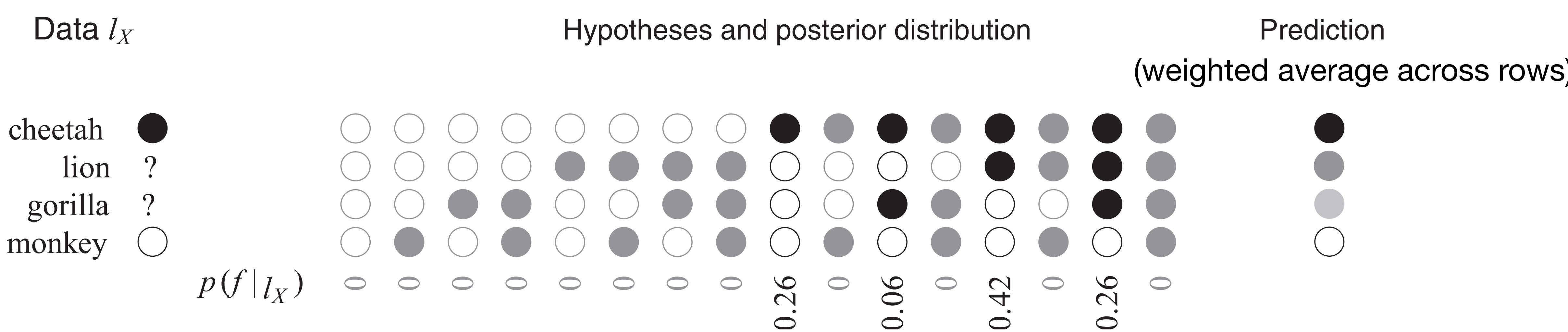
$l_X = \{1{,}0\}$: **feature labels for premise categories**

cheetahs have sesamoid bones.
<u>monkeys DO NOT have sesamoid bones.</u>
Do lions have sesamoid bones?
Do gorilla have sesamoid bones?

Posterior predictive distribution

$$P(f_Y = 1 \mid l_X) = \sum_{f : f_Y = 1} P(f \mid l_X)$$





Conclusions

Categories are important in induction:

- Even when the question is about a specific category, typicality to a general category is important (Rips; Osherson)
- When people choose a category, they base their induction on it even if they aren't sure it's the right category (Murphy & Ross)

Knowledge is also very important in induction, and Kemp and Tenenbaum provide a model that shows how knowledge and statistics can combine to make inferences

Acknowledgements

Thanks Greg Murphy for the first version of many of these slides