Categories and Concepts - Spring 2019
Prototype and exemplar theories

Brenden Lake

PSYCH-GA 2207

What is a chair?

What is a chair?







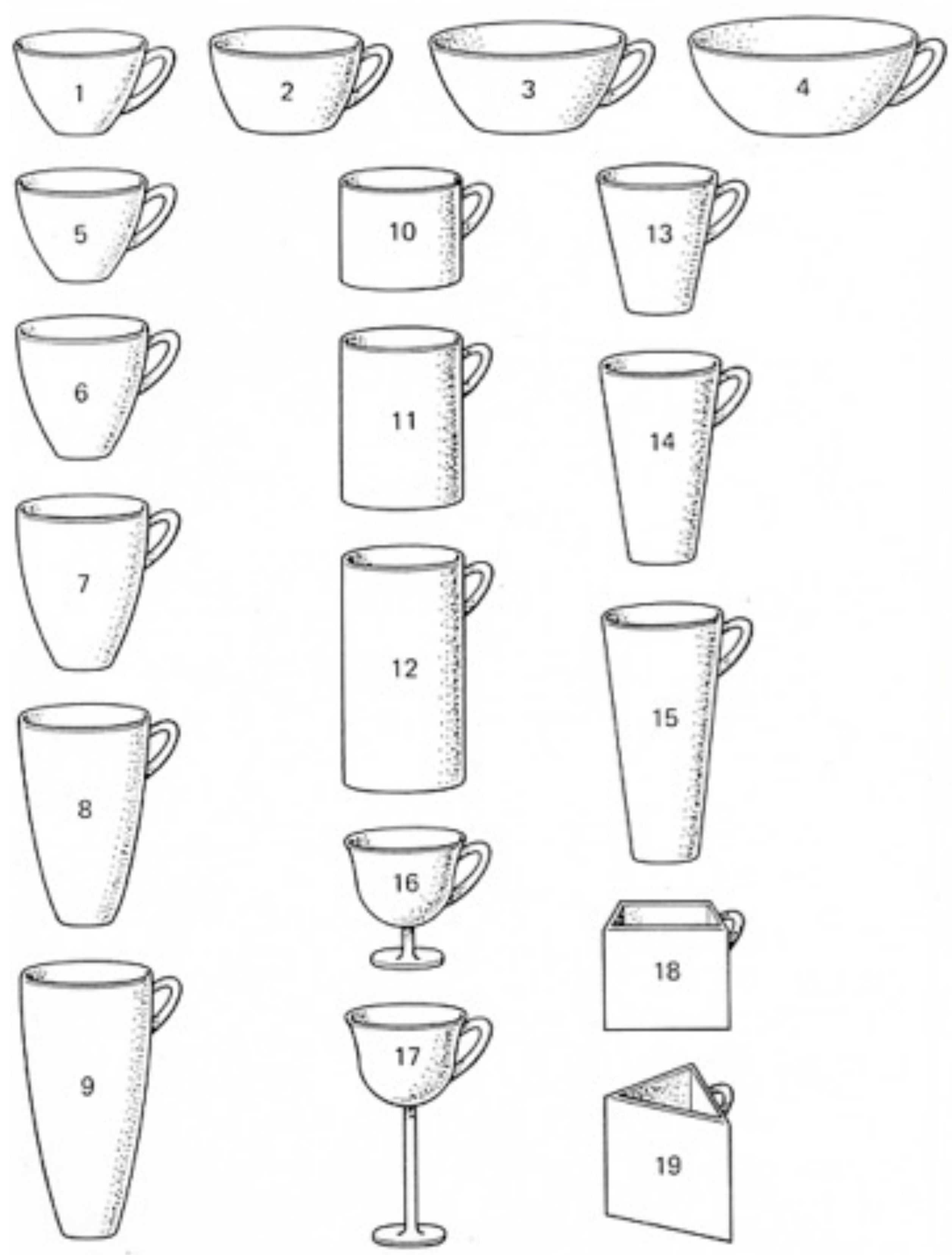




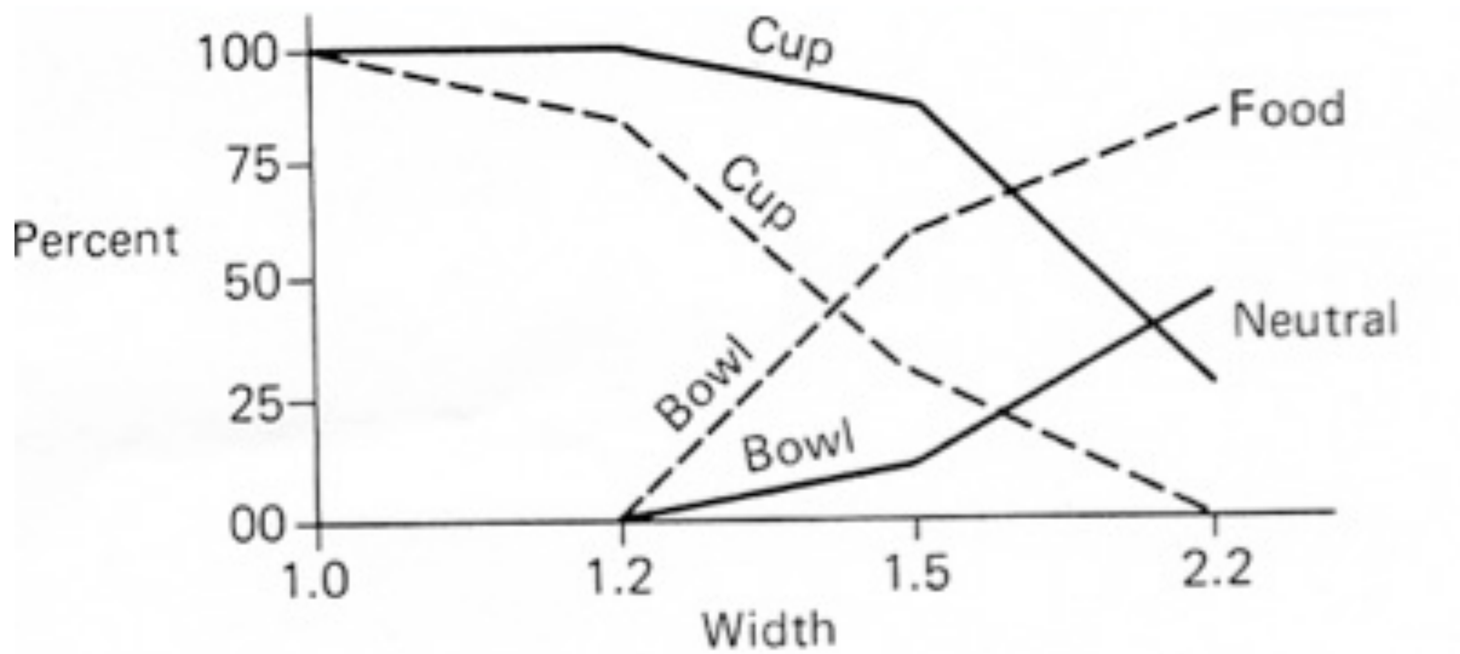

What is a game?

Are concepts well-defined?

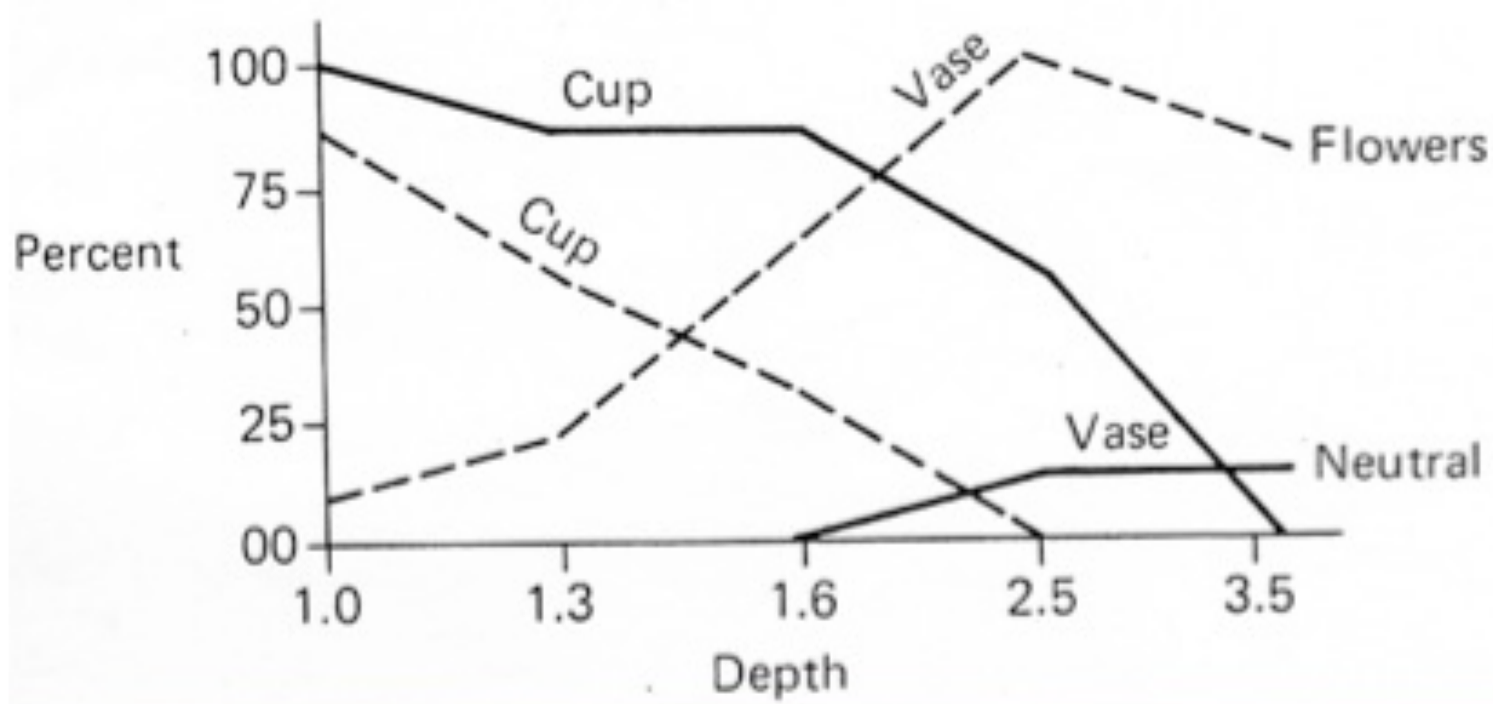
(Labov, 1973)



Are concepts well-defined? Some of Labov's results



(a) Use of names *cup* and *bowl* in Food and Neutral contexts

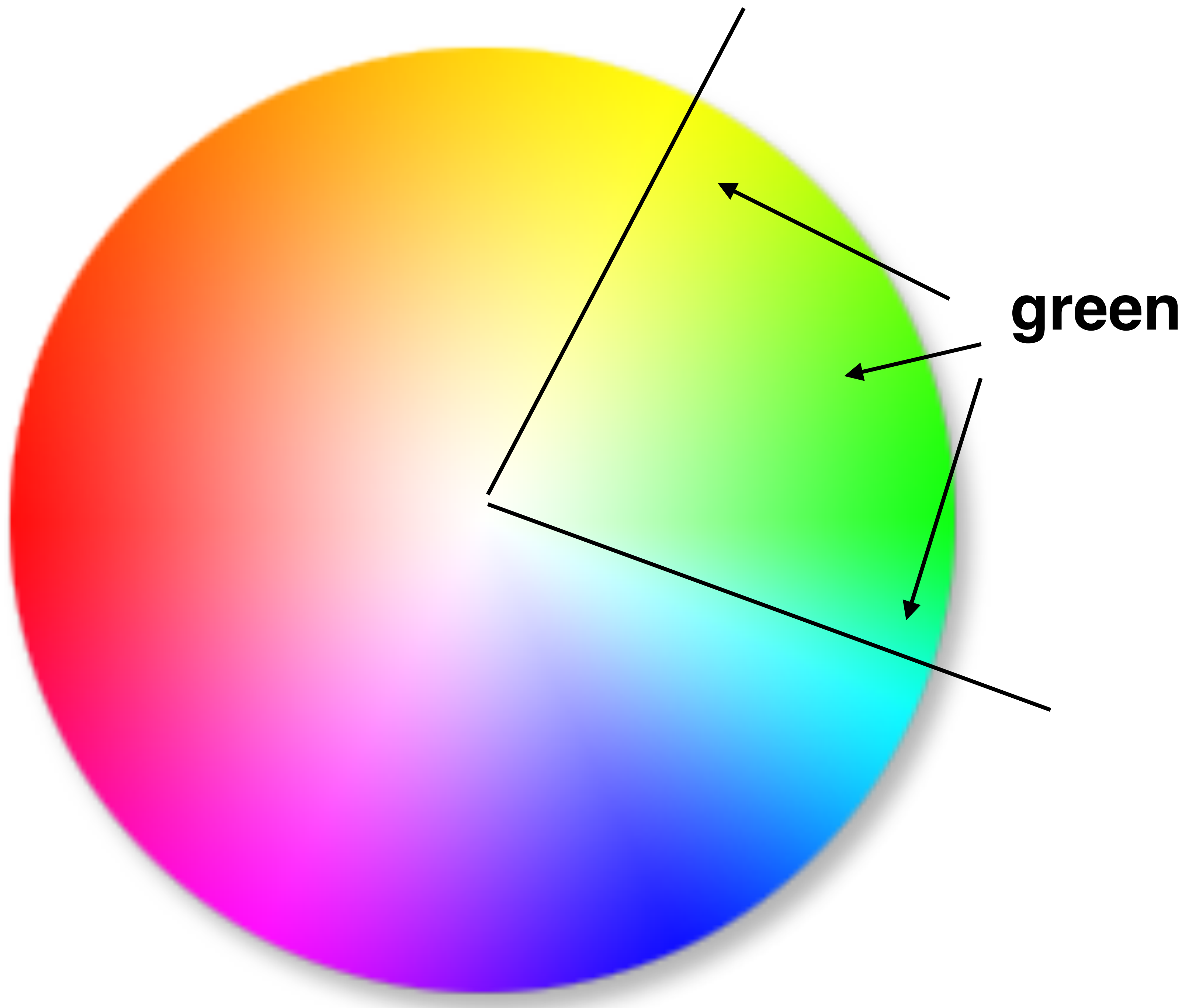


The Classical View

- ~~Concepts can be defined: there are necessary and sufficient conditions for category membership~~

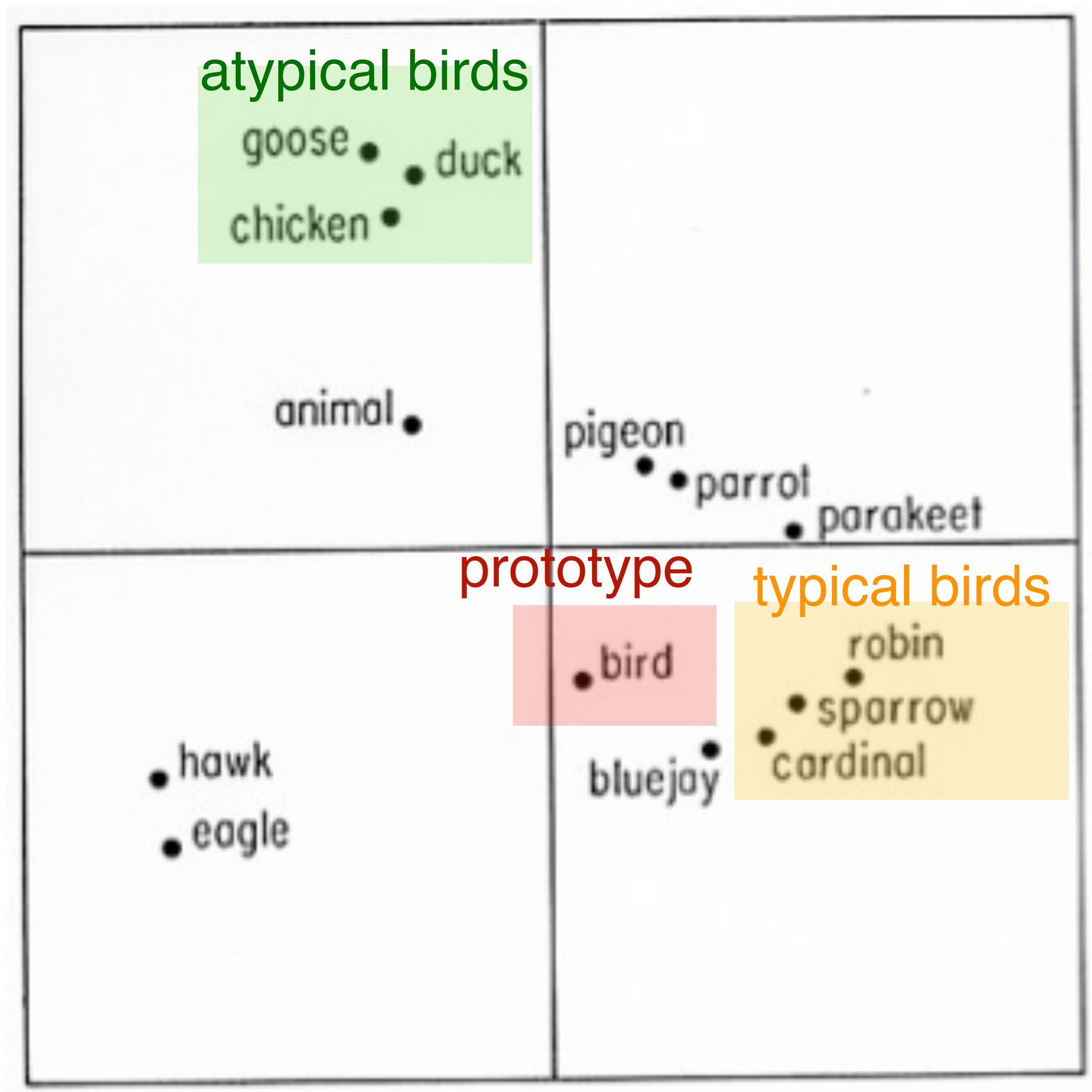
?
- **Membership is all-or-nothing. All members of a category are equally good members.**

Are all members of a category equally good?



Are all members of a category equally good?

Multi-dimensional scaling solution
derived from people's similarity ratings



(Rips, Shoben, & Smith, 1973)

Typicality ratings are highly reliable and predict many aspects of categorization.

Category: furniture

Member	Rank	Score (1 [typical]... 7 [not typical])
chair	1.5	1.04
sofa	1.5	1.04
couch	3.5	1.10
table	3.5	1.10
easy chair	5	1.33
dresser	6.5	1.37
rocking chair	6.5	1.37
coffee table	8	1.38
rocker	9	1.42
love seat	10	1.44
...		
stove	50	5.40
counter	51	5.44
clock	52	5.48
drapes	53	5.67
refrigerator	54	5.70
picture	55	5.75
closet	56	5.95
vase	57	6.23
ashtray	58	6.35
fan	59	6.49
telephone	60	6.68

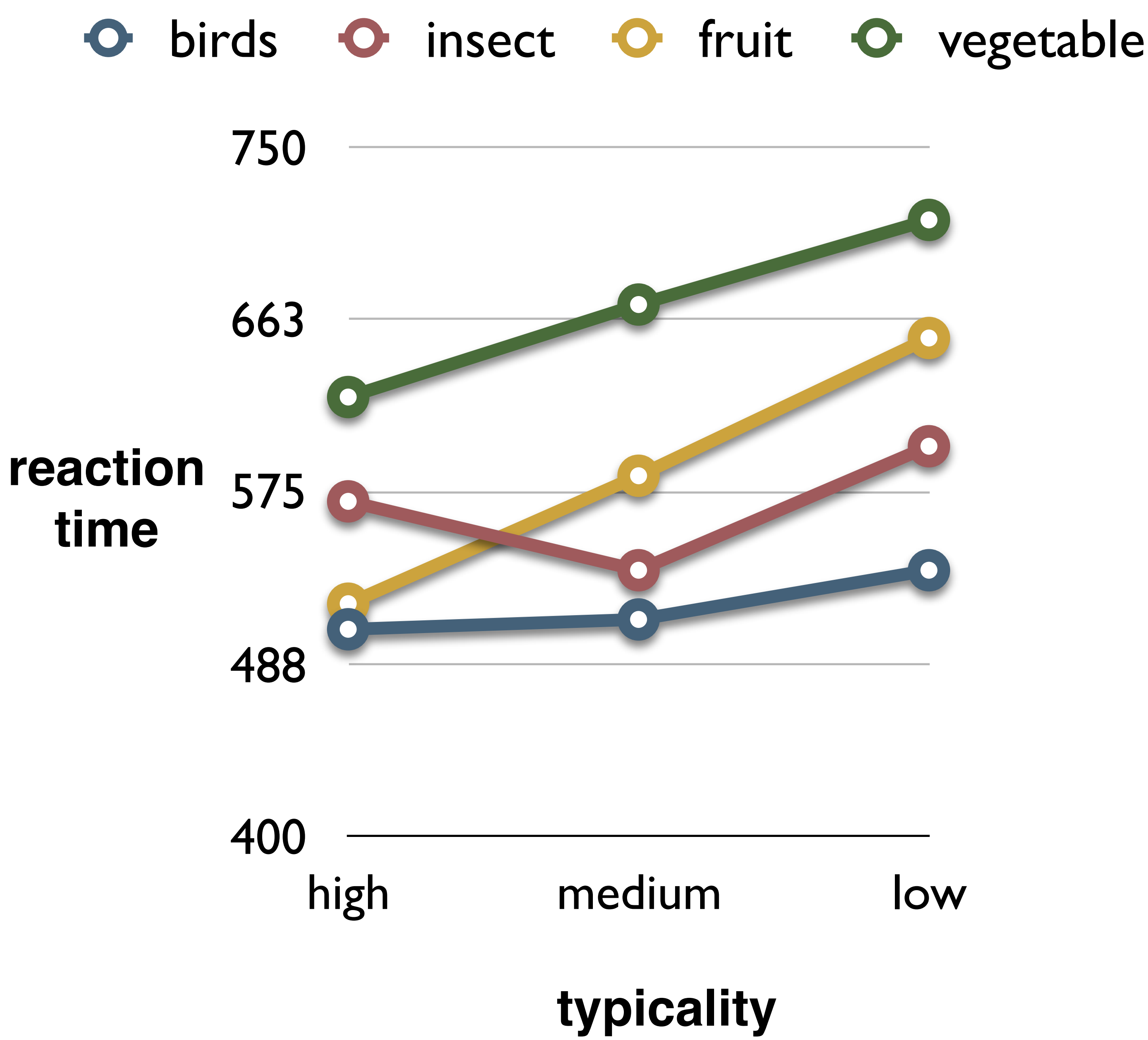
(Rosch, 1975)

Instructions:

...'Which breeds of dogs are real "doggy dogs"? To me, a German Shepard is very doggy, but a Pekinese is not. "Rate the extent to which each instance represents your idea or image of the meaning of the category"' (1-7 scale)

Goodness ratings are **highly reliable** across participant

Typicality predicts sentence verification



Prompt (True or False):
“A X [item] is a member of category Y”

(Smith, Rips, & Shoben, 1974)

Typicality predicts category production

Task: "Name as many examples of Category X as you can..."

Result: The earlier the object is named, the more "typical" it is

Example task, "Name as many birds as you can":

robin,
sparrow, typical birds
canary,
….,
ostrich,
emu, atypical birds
etc.

(e.g., Rosch, Simpson, & Miller, 1976)

Typicality affects…

- RT in sentence verification tasks
  - “A robin is a bird”
- Picture identification RT (robin vs. duck)
- Learning
  - People learn typical items first
  - Learning is better if you teach with typical items
- Order of production in a list
- Order in speech production (“apples and lemons” more likely than “lemons and apples”)
- Category-based induction

- **Actually, typicality affects every task that uses categories**

Should we throw out the classical view? Cautionary note: Even **definitional** concepts show typicality effects

odd number	typicality
3	1.6
7	1.9
23	2.4
57	2.6
501	3.5
447	3.7

plane geometry figure	
square	1.3
triangle	1.5
rectangle	1.9
circle	2.1
trapezoid	3.1
ellipse	3.4

These typicality ratings also affect sentence verification tasks, just like graded concepts.

(Armstrong, Gleitman, & Gleitman, 1983)

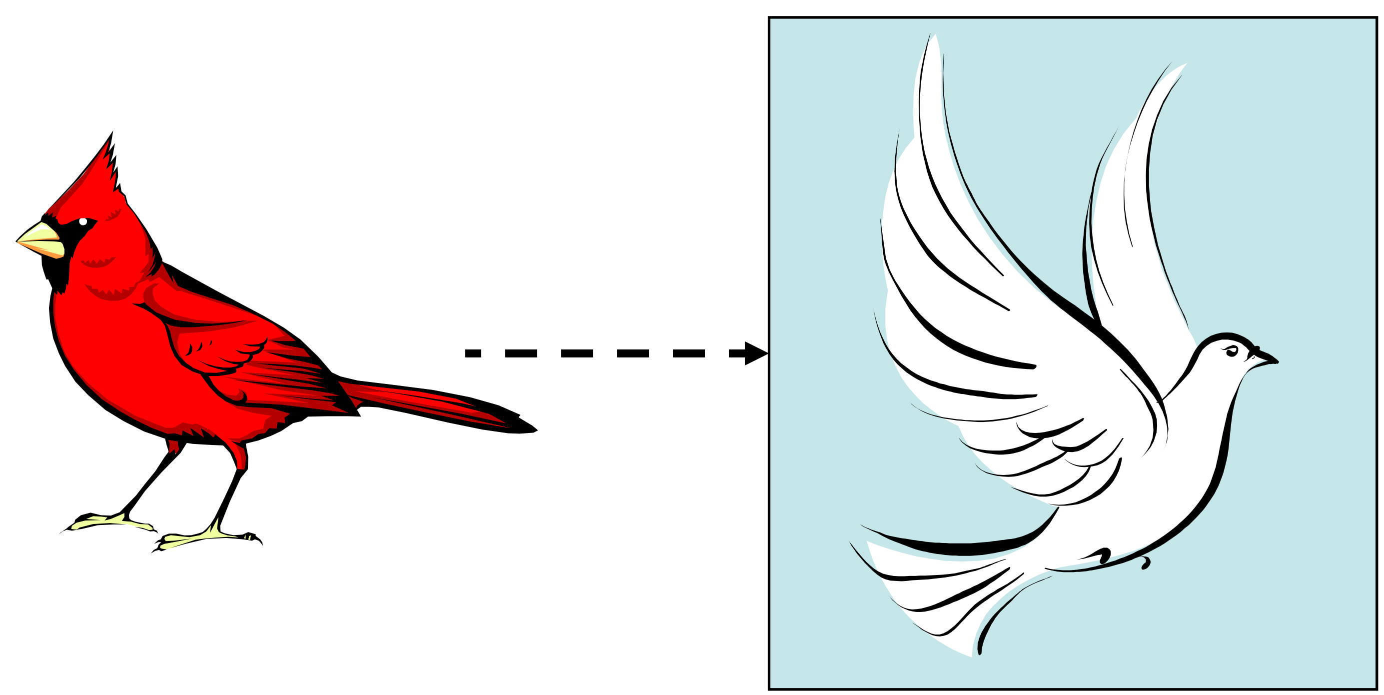
The Classical View

- Concepts can be defined: there are necessary and sufficient conditions for category membership

- Membership is all-or-nothing. All members of a category are equally good members.

If not “classical,” then what?
Prototype theory

- There are different versions of the theory, which go like this: **Concepts are a summary representations based on typical properties or central tendency of a category, or an ideal image**
- Earliest alternative, but now not the only, or the most popular theory



Bird?

Prototypical Bird

An example artificial category learning task, supporting the notion of learning a prototype or "ideal image"

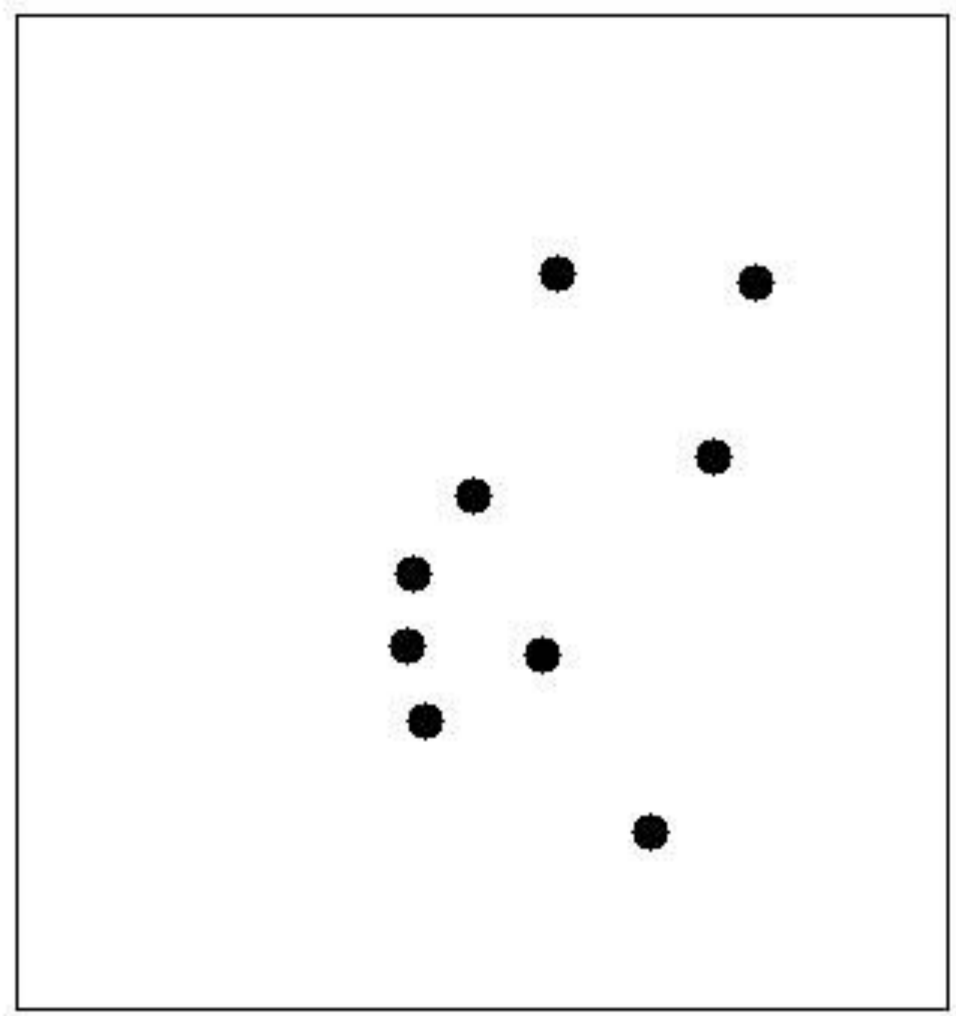
(Posner & Keele, 1968)

1. **Instructions**: You will see stimuli from Category A or B. Please indicate which category you think is correct.

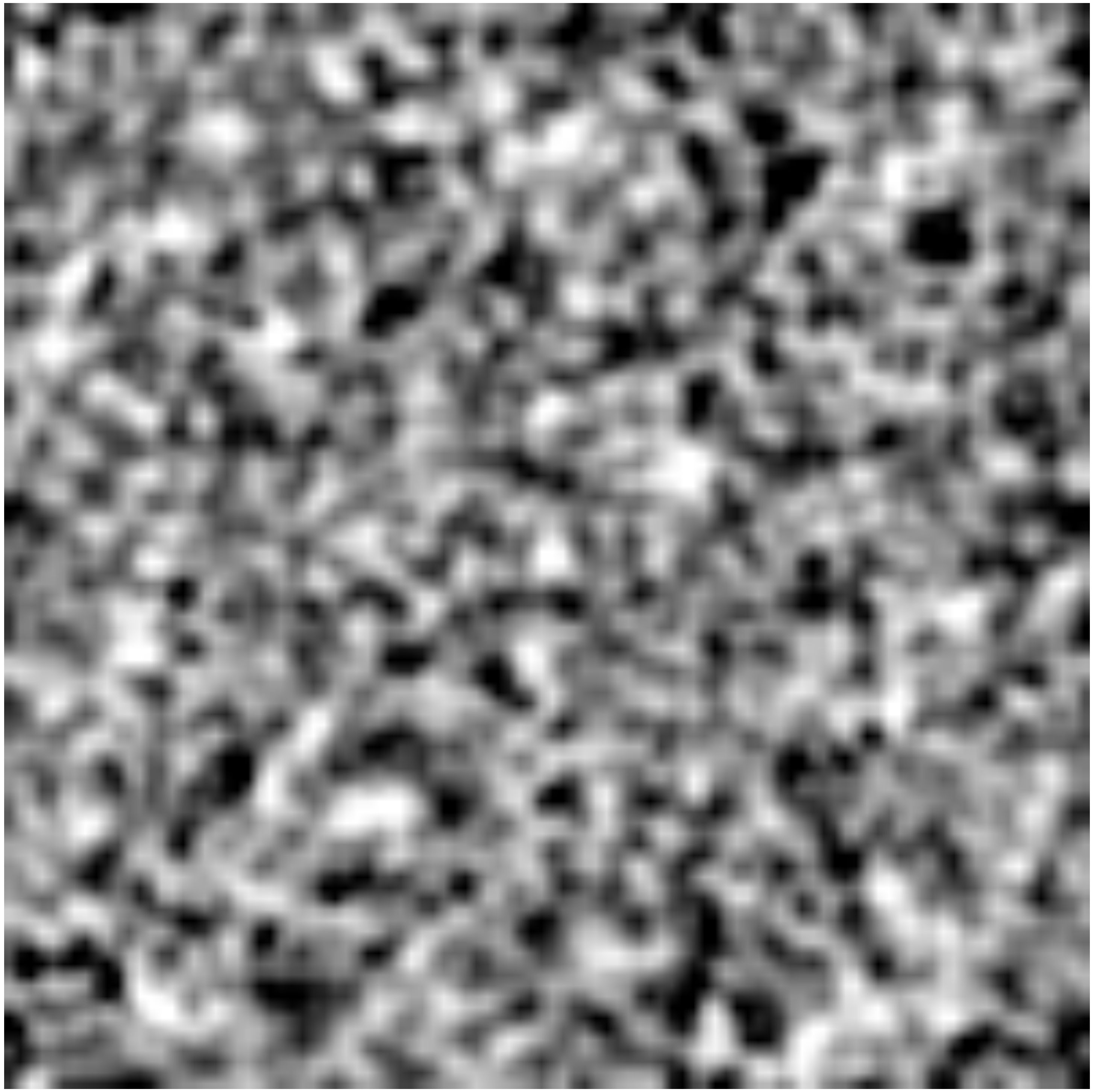
2. **Training phase**: Participants see stimuli one at a time. For each item, they respond "A" or "B". Usually, feedback (the correct answer) is received during training.

3. **Test phase (optional)**: Participants may respond to additional stimuli. No feedback is given.

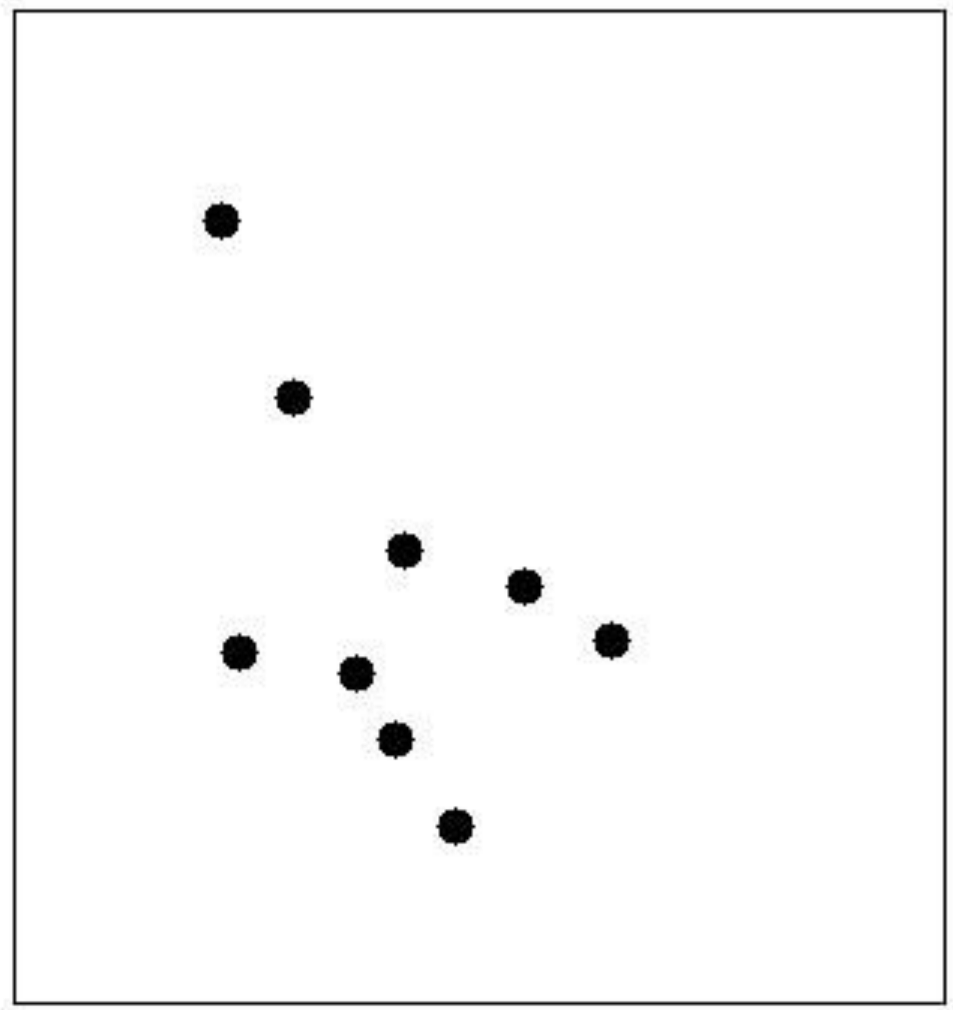
Category A or B?



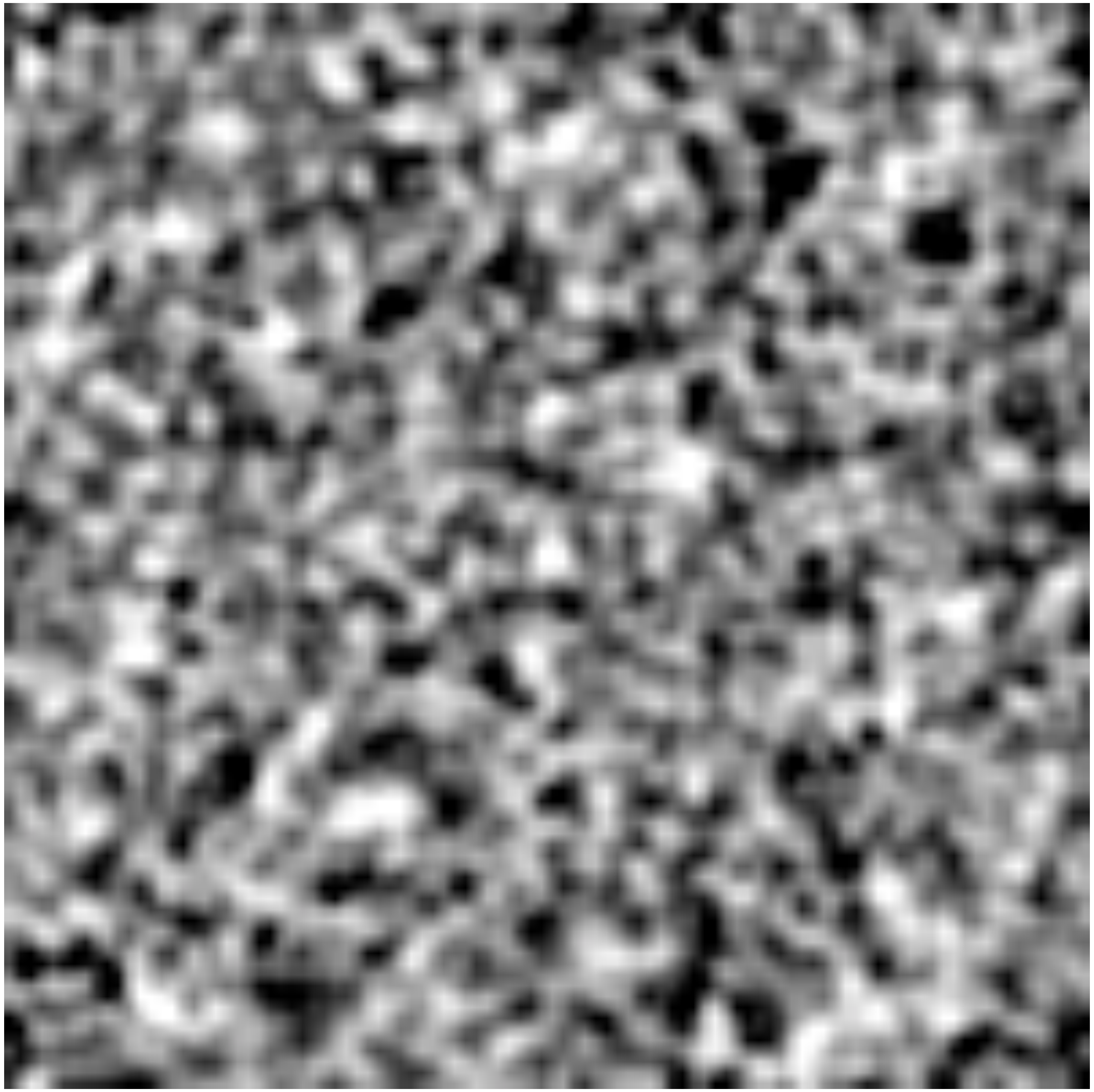
A



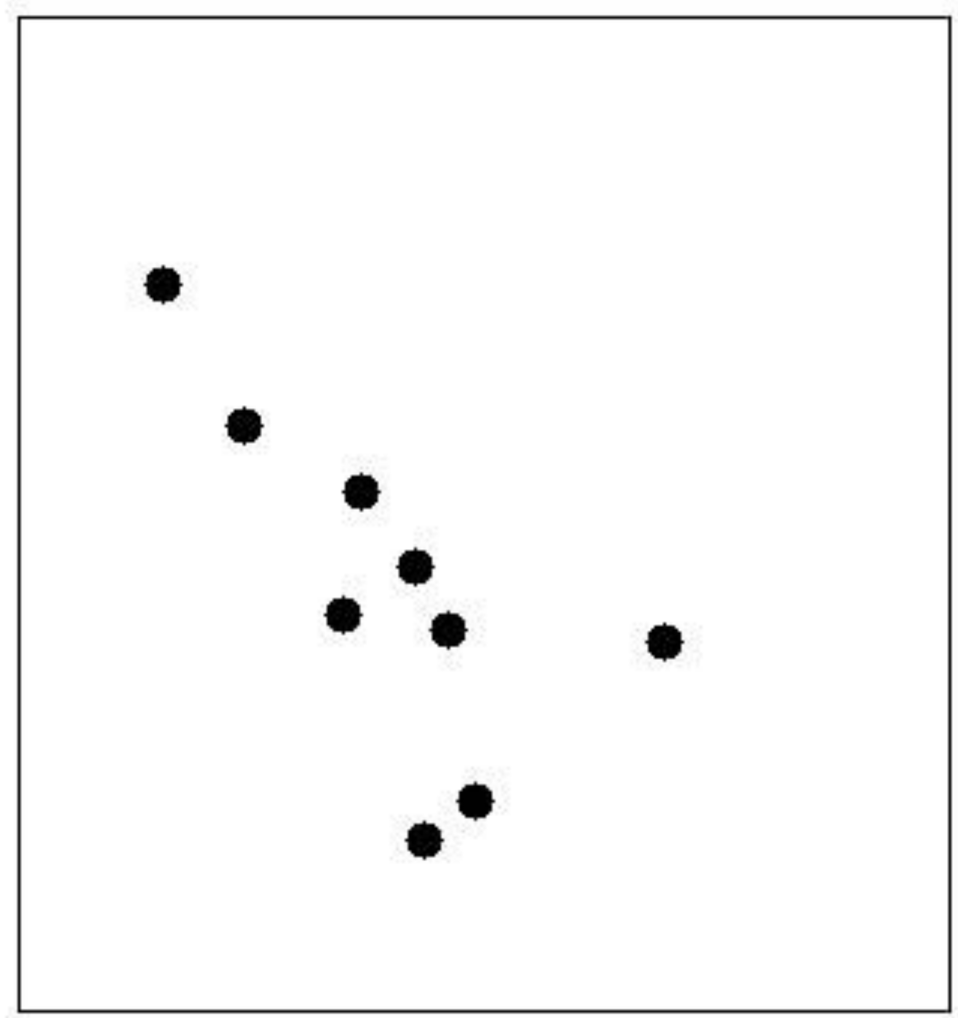
Category A or B?



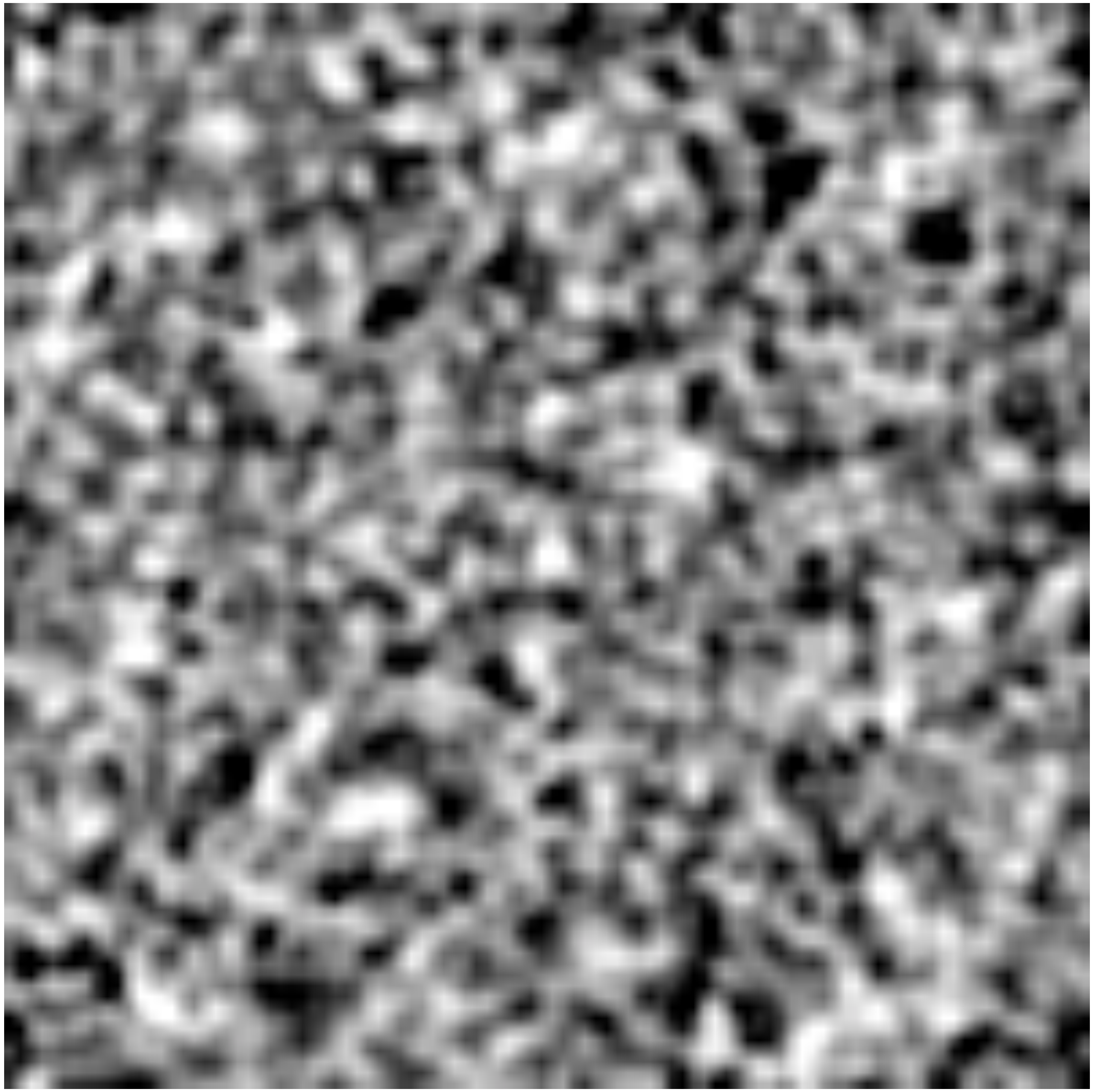
B



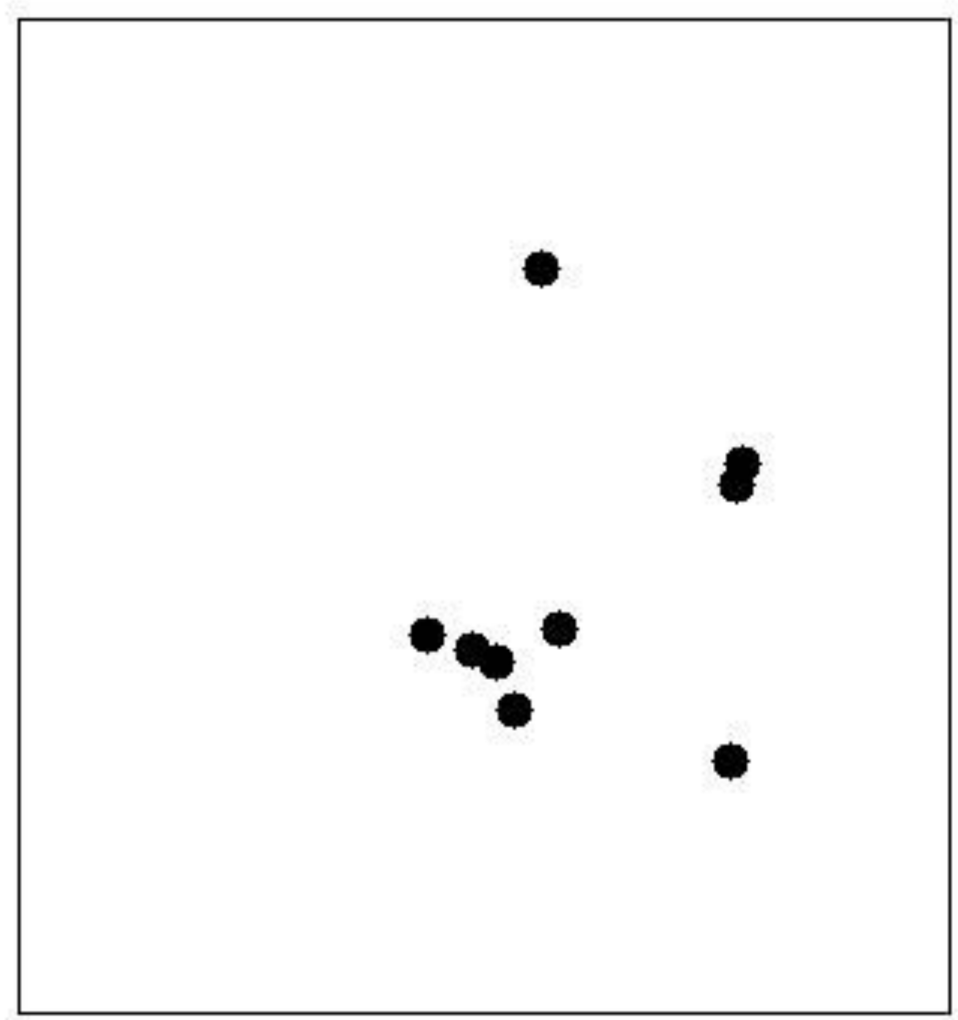
Category A or B?



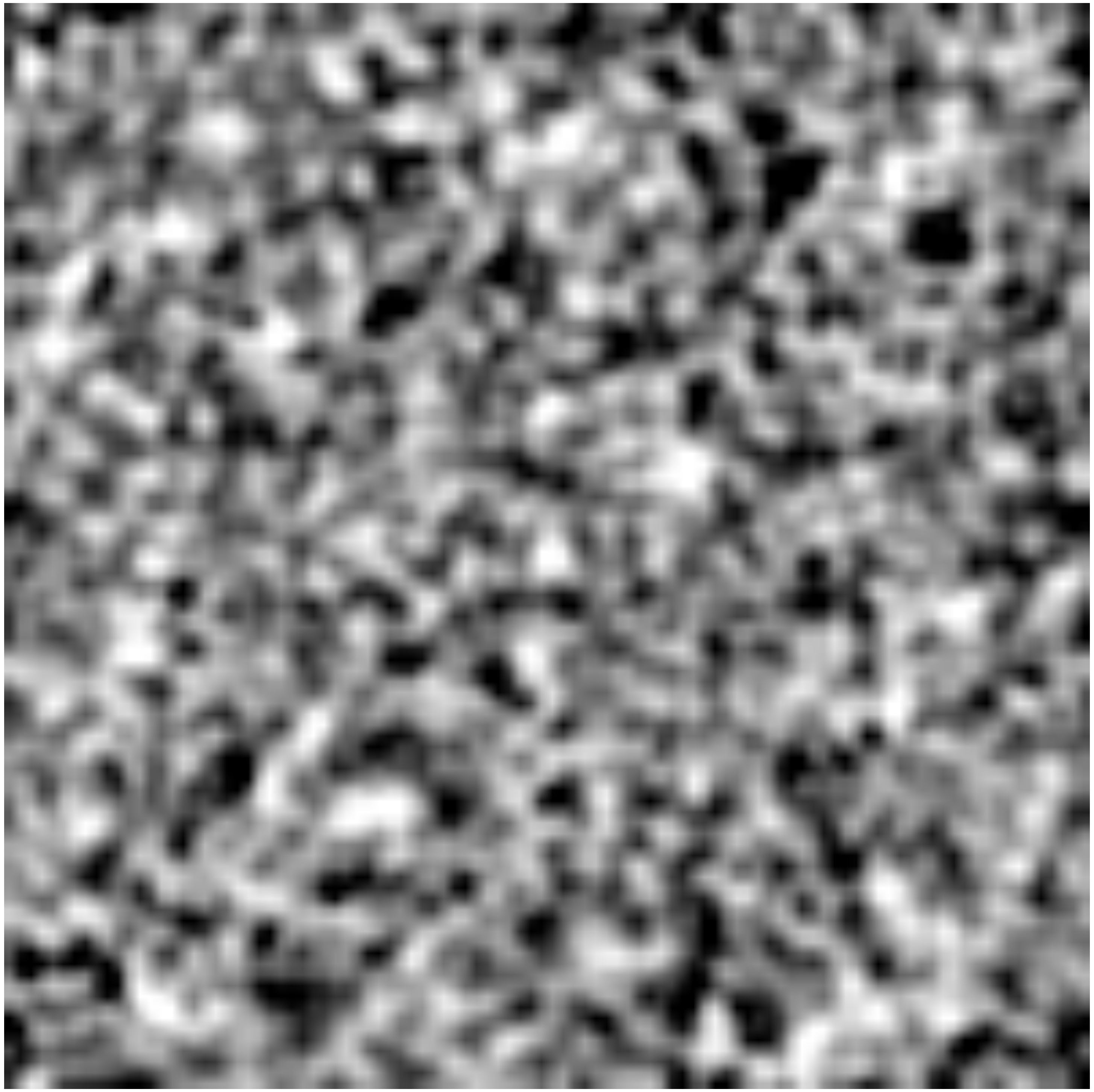
B



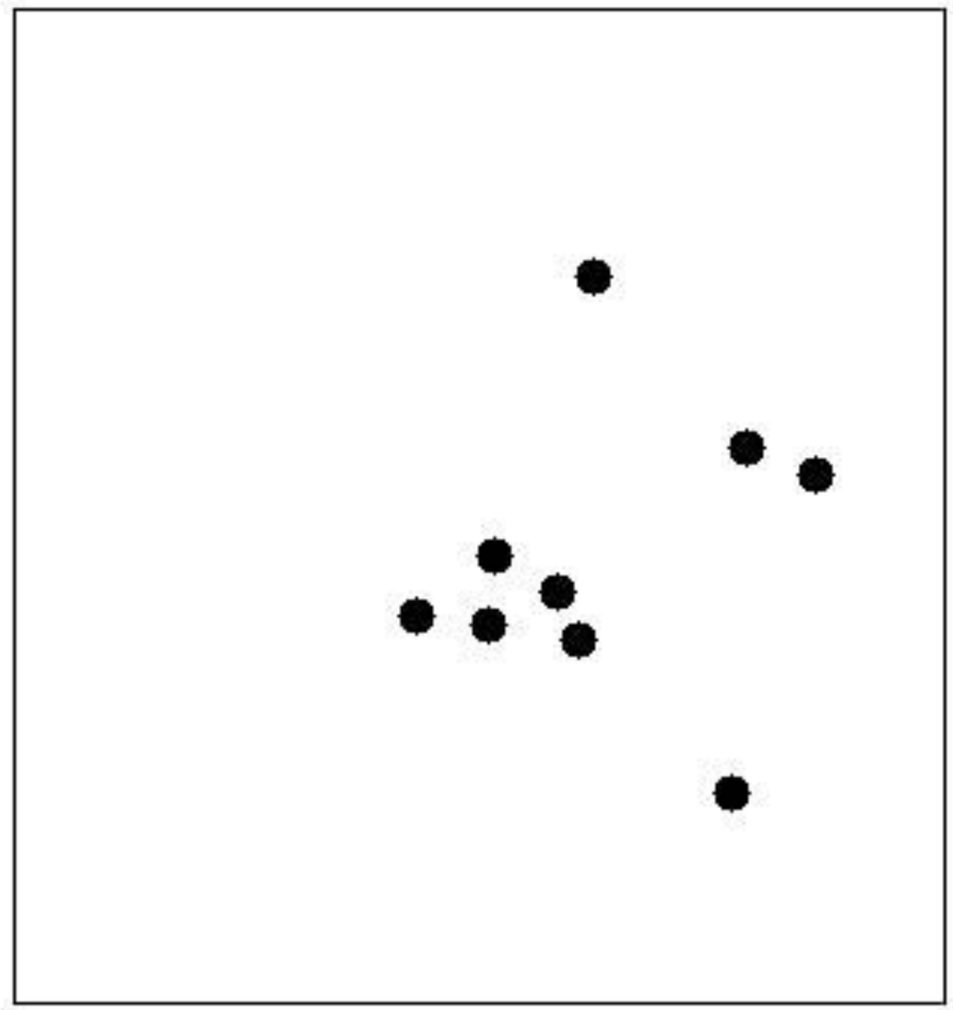
Category A or B?



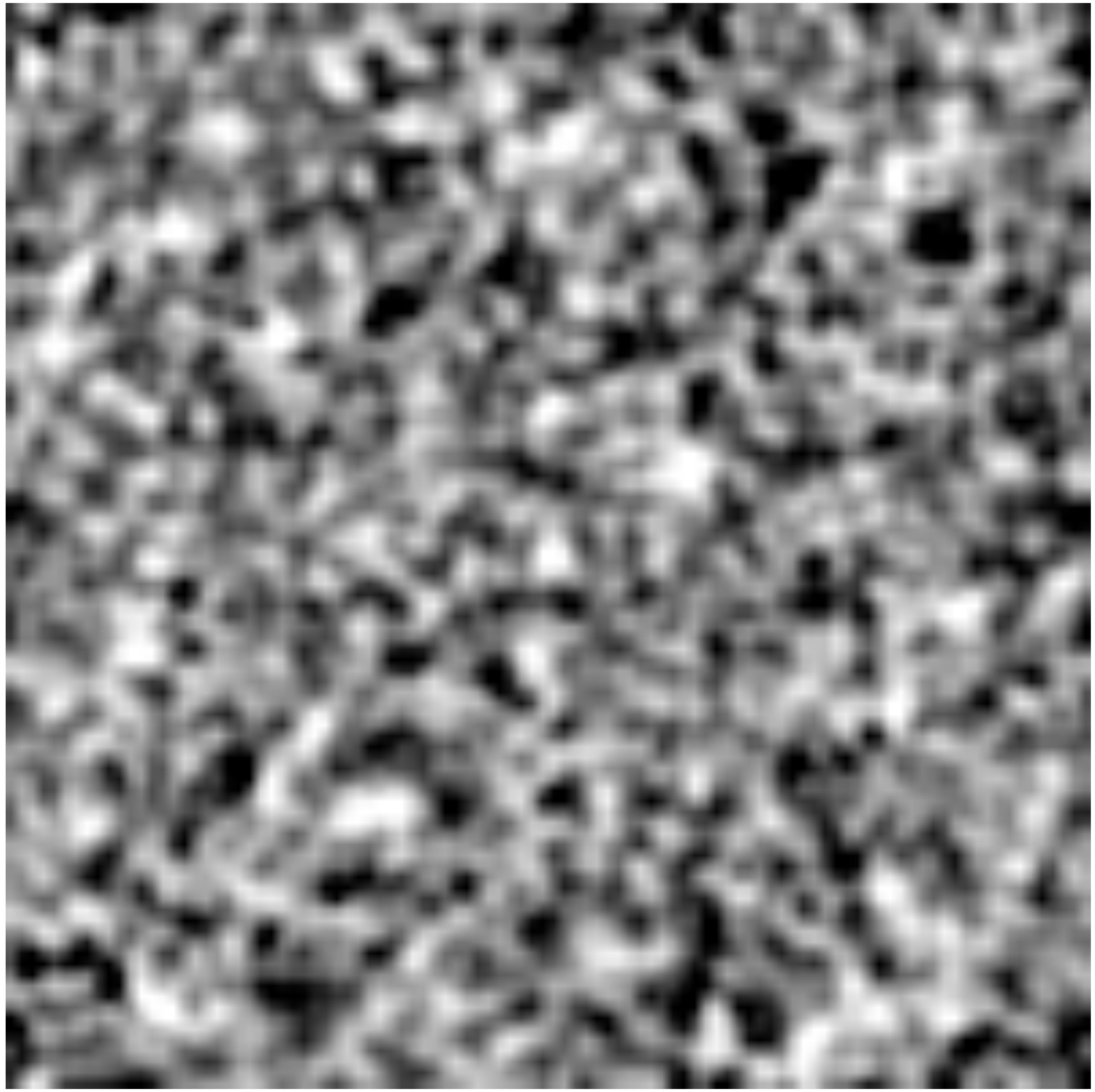
A



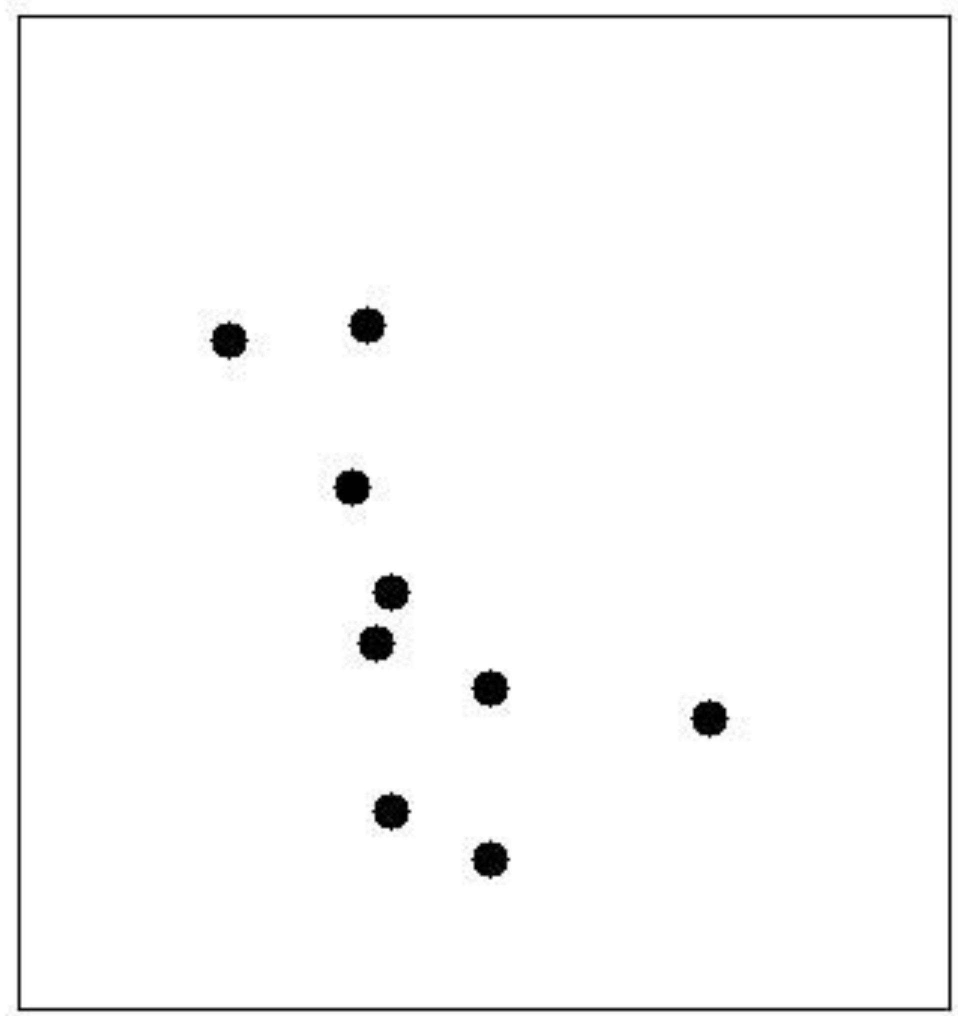
Category A or B?



A



Category A or B?



B

Training period is done. Now for testing…

Items seen during test period (after training)

(Posner & Keele, 1968)

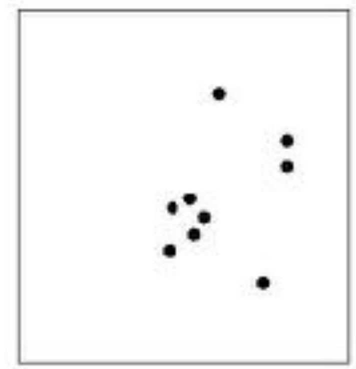
After training, participants were tested on:
-- the prototypes (new)
-- some pattern distortions (old)
-- some pattern distortions (new)

Result:
(Accuracy for prototype = Accuracy for old distortions)

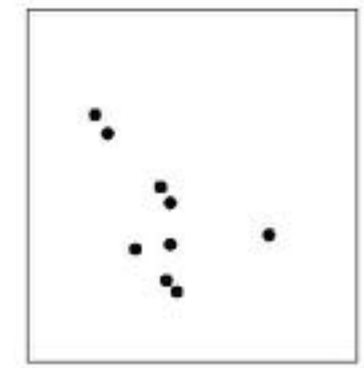
> Accuracy for new distortions

Suggests that some form of abstract representation is learned, like an “ideal image” or prototype

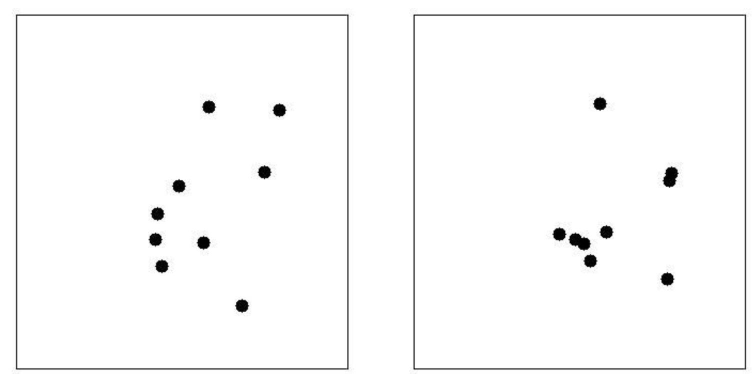
Prototype A: (not seen)



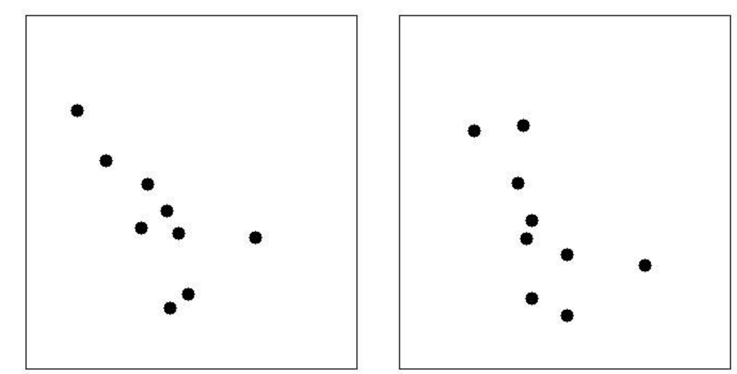
Prototype B: (not seen)



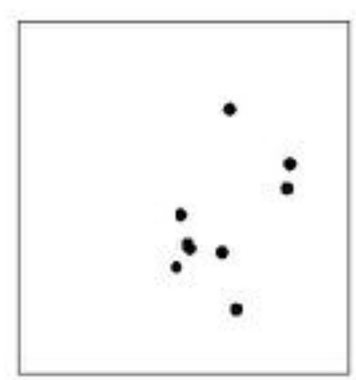
Distortions of A: Old



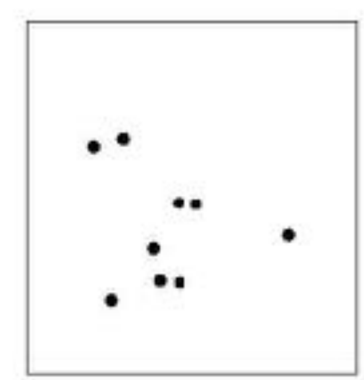
Distortions of B: Old



Distortions of A: New



Distortions of B: New



Prototype theory and the Rosch and Mervis family resemblance view

“Many features in common, but no feature is shared by everyone”

(image from Armstrong, Gleitman, & Gleitman, 1983)

Rosch & Mervis's (1975) family resemblance view

Typicality rests on two main factors:

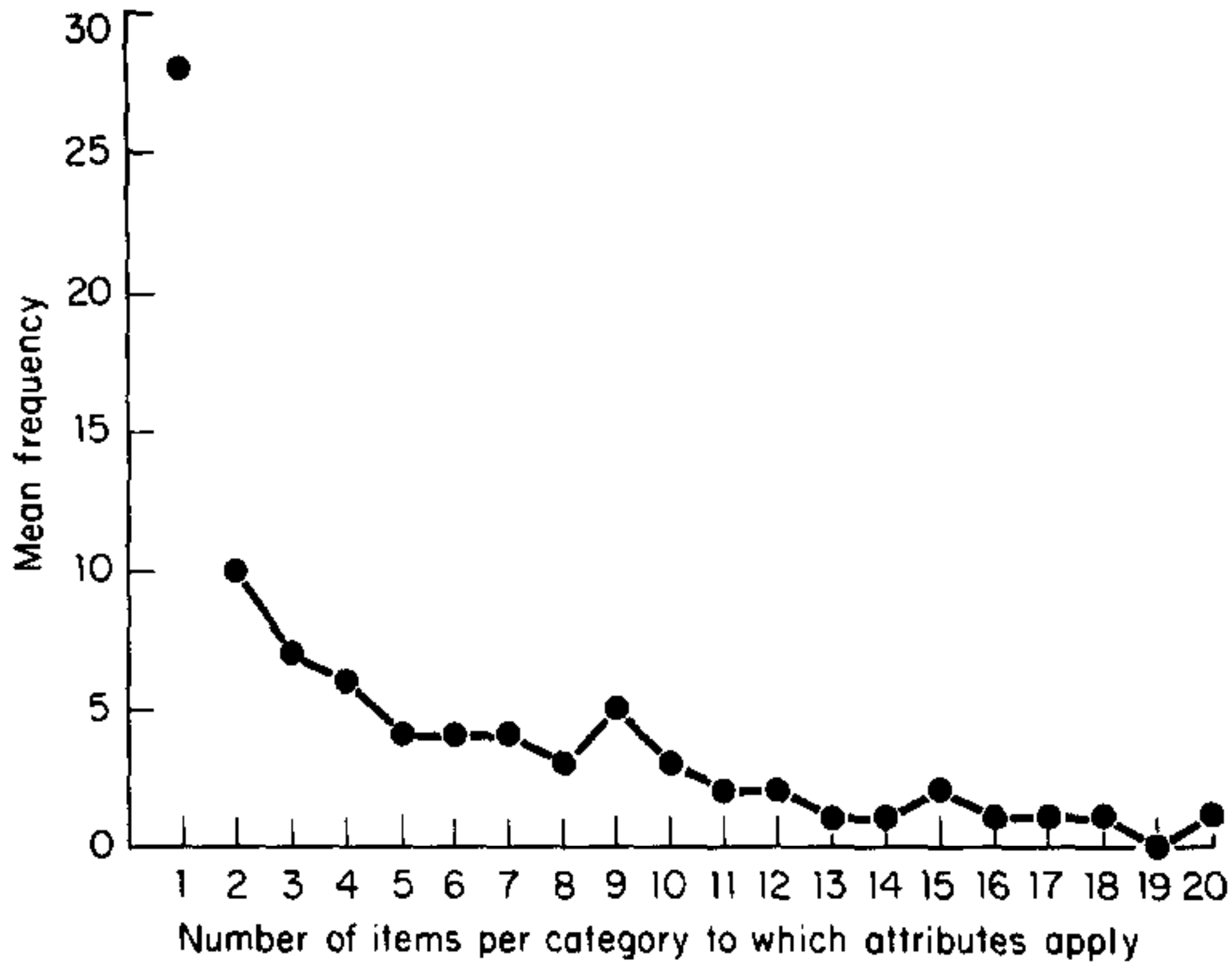
- having features frequently found in the category; and
- not having features frequently found in <u>other</u> categories

Attribute listing experiment

TABLE 1
SUPERORDINATE CATEGORIES AND ITEMS USED IN EXPERIMENTS 1 AND 2

	Category					
Item	Furniture	Vehicle	Fruit	Weapon	Vegetable	Clothing
1	Chair	Car	Orange	Gun	Peas	Pants
2	Sofa	Truck	Apple	Knife	Carrots	Shirt
3	Table	Bus	Banana	Sword	String beans	Dress
4	Dresser	Motorcycle	Peach	Bomb	Spinach	Skirt
5	Desk	Train	Pear	Hand grenade	Broccoli	Jacket
6	Bed	Trolley car	Apricot	Spear	Asparagus	Coat
7	Bookcase	Bicycle	Plum	Cannon	Corn	Sweater
8	Footstool	Airplane	Grapes	Bow and arrow	Cauliflower	Underpants
9	Lamp	Boat	Strawberry	Club	Brussel sprouts	Socks
10	Piano	Tractor	Grapefruit	Tank	Lettuce	Pajamas
11	Cushion	Cart	Pineapple	Teargas	Beets	Bathing suit
12	Mirror	Wheelchair	Blueberry	Whip	Tomato	Shoes
13	Rug	Tank	Lemon	Icepick	Lima beans	Vest
14	Radio	Raft	Watermelon	Fists	Eggplant	Tie
15	Stove	Sled	Honeydew	Rocket	Onion	Mittens
16	Clock	Horse	Pomegranate	Poison	Potato	Hat
17	Picture	Blimp	Date	Scissors	Yam	Apron
18	Closet	Skates	Coconut	Words	Mushroom	Purse
19	Vase	Wheelbarrow	Tomato	Foot	Pumpkin	Wristwatch
20	Telephone	Elevator	Olive	Screwdriver	Rice	Necklace

Family resemblance



Most attributes are unique to just one category item, and very few apply across all items (max of 20)

Family resemblance

The most typical category members share more attributes than the least typical members.

TABLE 2

NUMBER OF ATTRIBUTES IN COMMON TO FIVE MOST AND FIVE LEAST PROTOTYPICAL MEMBERS OF SIX CATEGORIES

Category	Most typical members	Least typical members
Furniture	13	2
Vehicle	36	2
Fruit	16	0
Weapon	9	0
Vegetable	3	0
Clothing	21	0

Rosch & Mervis category structure - Exp 5

TABLE 3
ARTIFICIAL CATEGORY STRUCTURES USED IN EXPERIMENTS 5 AND 6

		Type of category structure			
		Control set		Symmetric experimental set	
Use of the category	Item in category	Letter string	Family resemblance score	Letter string	Family resemblance score
Basic category structure	1	HPNWD	12	JXPHM	15
	2	HPNSJ	12	XPHMQ	19
	3	GKNTJ	12	PHMQB	21
	4	4KCTG	12	HMQBL	21
	5	4KC6D	12	MQBLF	19
	6	HPC6B	12	QBLFS	15
Nonoverlap contrast category (Experiment 5)	1	R7QUM	12	CTRVG	15
	2	R7QXV	12	TRVGZ	19
	3	Z5Q2V	12	RVGZK	21
	4	L5F27	12	VGZKD	21
	5	L5F1M	12	GZKDW	19
	6	R7F19	12	ZKDWN	15

Artificial category learning results (Exp 5) - non-overal contrast

High family resemblance items have fewer errors, faster RT, and higher typicality ratings.

TABLE 4
EFFECT OF DEGREE OF FAMILY RESEMBLANCE ON RESPONSE MEASURES

Stimulus type	Response measures								
	Number of errors			Reaction time (msec)			Prototypicality rating		
	Hi[a]	Med	Lo	Hi	Med	Lo	Hi	Med	Lo
Symmetric experimental	2.8	4.4	5.5	560	617	692	5.0	3.4	2.1
Control	6.5	6.4	6.7	670	651	644	3.7	3.4	3.4

[a] Hi, Med, and Lo refer to family resemblance scores.

Rosch & Mervis category structure - Exp 6

TABLE 3
ARTIFICIAL CATEGORY STRUCTURES USED IN EXPERIMENTS 5 AND 6

Use of the category	Item in category	Control set: Letter string	Control set: Family resemblance score	Control set: Overlap score	Symmetric experimental set: Letter string	Symmetric experimental set: Family resemblance score	Symmetric experimental set: Overlap score
Basic category structure	1	HPNWD	12	0	JXPHM	15	0
	2	HPNSJ	12	2	XPHMQ	19	1
	3	GKNTJ	12	4	PHMQB	21	2
	4	4KCTG	12	5	HMQBL	21	3
	5	4KC6D	12	3	MQBLF	19	4
	6	HPC6B	12	1	QBLFS	15	5
Overlapped contrast category[a] (Experiment 6)	1	8SJKT		4[b]	GVRTC		0
	2	8SJ3G		3	VRTCS		1
	3	9UJCG		3	RTCSF		2
	4	4UZC9		2	TCSFL		3
	5	4UZRT		3	CSFLB		4
	6	MSZR5		3	SFLBQ		5

[a] Overlap is with the basic category structure not the nonoverlap contrast category.
[b] Contrast strings in control do not have same structure as initial category strings and were not analyzed in Experiment 6.

Artificial category learning results (Exp 6) - overlap contrast category

Low overlap items have fewer errors, faster RT, and higher typicality ratings.

TABLE 5

EFFECT OF DEGREE OF OVERLAP ON RESPONSE MEASURES FOR CONTROL SET

Response measure	Degree of overlap		
	Low	Medium	High
Number of errors	7.1	9.4	12.6
Reaction time (msec)	909	986	1125
Prototypicality rating	5.3	3.4	1.8

Prototypicality can be induced purely by overlap with contrast categories

Overall conclusion, Rosch & Mervis

Typicality is based on both

- An item’s similarity to other category members (increasing)
- An item’s similarity to members of other categories (decreasing)

For some reason, many later accounts ignore #2, in spite of correlational and experimental evidence for it in this paper

Basic assumptions of Prototype Theory

- A conceptual representation is a summary representation of the category

- It represents the typical properties or central tendency of the category

- Items that differ in their closeness to the representation vary in typicality

Is a prototype enough?

dog prototype

my dog

Consider the feature, “Needs to be brushed occasionally, not every day”

This feature would be true for most dogs, but certainly not my dog.

I may have a separate exemplar-based representation for my dog

Is a prototype enough?

“apple”

summary representation, where typicality is the sum weights of the present features:

- “round” (weight 1.0)
- “edible” (weight 0.8)
- “sweet” (weight 0.7)
- ….
- “green” (weight 0.5)?
- “red” (weight 0.5)?

Would an apple be best if it is both green and red at the same time? what if features are contradictory?

what would the prototype look like?












What is an alternative? Exemplars

- No summary representation

- Concepts are represented by remembered category members: “exemplars”
  - exemplars are labeled by their category membership

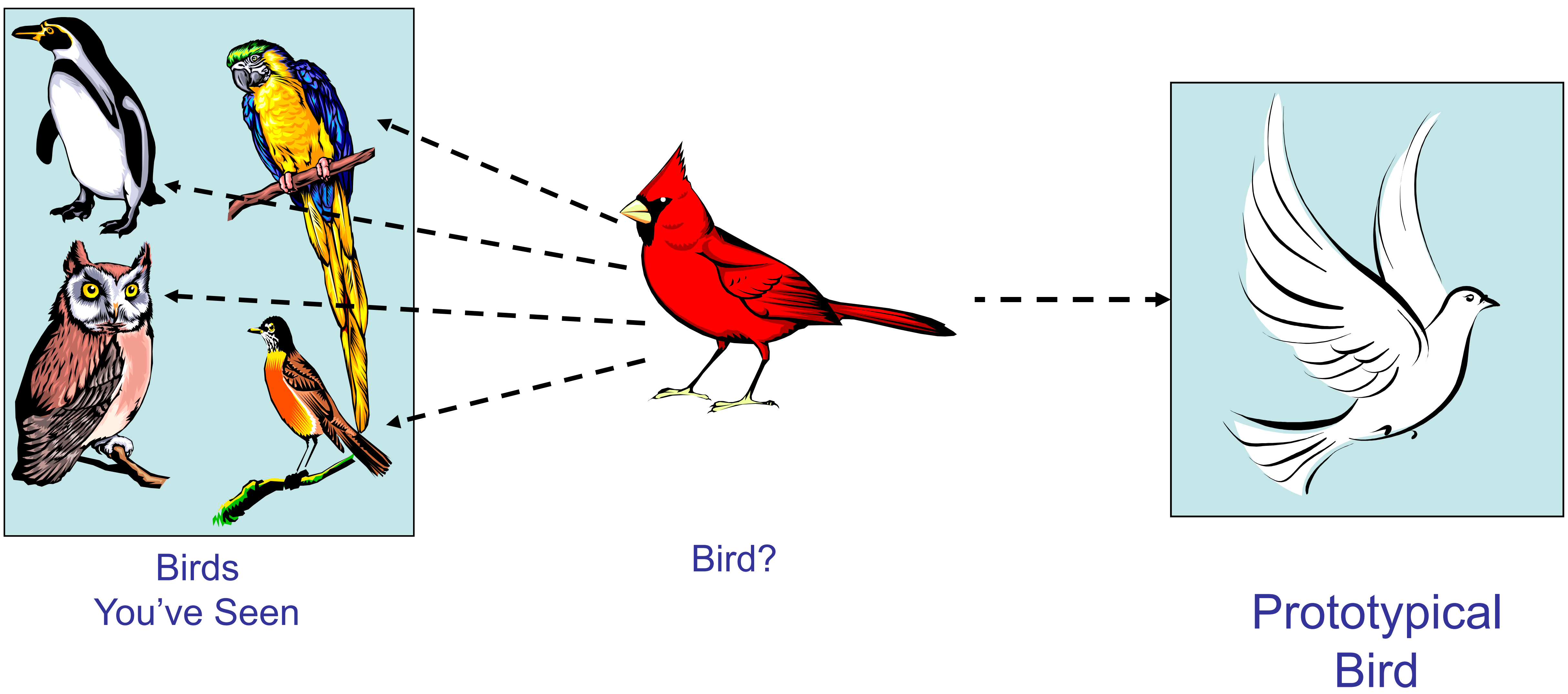
- Categorization is done by retrieving similar exemplars, and noting their category membership (simplifying greatly)

- Can account for good performance on unseen prototypes, since they are similar to many exemplars

Exemplar vs. prototype theories

exemplar theory

prototype theory



What is a chair? a set like this…

Medin & Schaffer (1978) Example

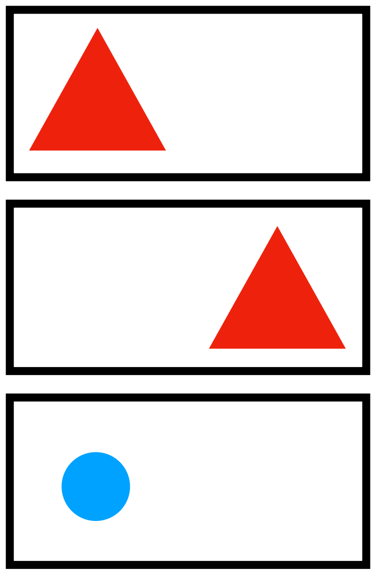
(what do these feature table mean?)

Category A

(abstract table)

D1 (shape)	D2 (size)	D3 (color)	D4 (position)
1	1	1	1
1	1	1	0
0	0	0	1

(what people actually see)



Medin & Schaffer Example

Category A

D1	D2	D3	D4
1	1	1	1
1	1	1	0
0	0	0	1

Category B

D1	D2	D3	D4
0	0	0	0
0	0	1	1
1	1	0	0

Transfer item

0	1	0	1

Medin and Schaffer's Context Model
Similarity between two items *x* and *y*

Category A

	D_1	D_2	D_3	D_4
x	1	1	1	1
	1	1	1	0
	0	0	0	1

sim(y, x)

0.09

Category B

D_1	D_2	D_3	D_4
0	0	0	0
0	0	1	1
1	1	0	0

Transfer item

y	0	1	0	1

$$\textbf{sim}(y, x) = \prod_{D_i} m^{1[x_i \neq y_i]} = m \cdot 1 \cdot m \cdot 1 = 0.09$$

"mismatch" free parameter such that $m = 0.3$

Similarity between item y and category C

Category A

D1	D2	D3	D4	$\text{sim}(y,x)$
1	1	1	1	**0.09**
1	1	1	0	**0.027**
0	0	0	1	**0.3**

$$\text{sim}(y, A) = 0.417$$

Category B

D1	D2	D3	D4	$\text{sim}(y,x)$
0	0	0	0	**0.09**
0	0	1	1	**0.09**
1	1	0	0	**0.09**

$$\text{sim}(y, B) = 0.27$$

Transfer item

y	0	1	0	1

$$\text{sim}(y, C) = \sum_{x \in C} \text{sim}(y, x)$$

for one of the classes C

Probability of response

Category A

D1	D2	D3	D4	sim(y, x)
1	1	1	1	0.09
1	1	1	0	0.027
0	0	0	1	0.3

$$\text{sim}(y, A) = 0.417$$

Category B

D1	D2	D3	D4	sim(y, x)
0	0	0	0	0.09
0	0	1	1	0.09
1	1	0	0	0.09

$$\text{sim}(y, B) = 0.27$$

Transfer item

y	0	1	0	1

$$P(y \in A) = \frac{\text{sim}(y, A)}{\text{sim}(y, A) + \text{sim}(y, B)} = \frac{0.417}{0.417 + 0.27} = 0.61$$

$$P(y \in B) = \frac{\text{sim}(y, B)}{\text{sim}(y, A) + \text{sim}(y, B)} = \frac{0.27}{0.417 + 0.27} = 0.39$$

Context model of classification

"classification is based on similarity to all exemplars in a class"
(Medin & Schaffer, 1978)

Item similarity

$$\textbf{sim}(y, x) = \prod_{D_i} m^{1[x_i \neq y_i]}$$

"mismatch" free parameter *m*

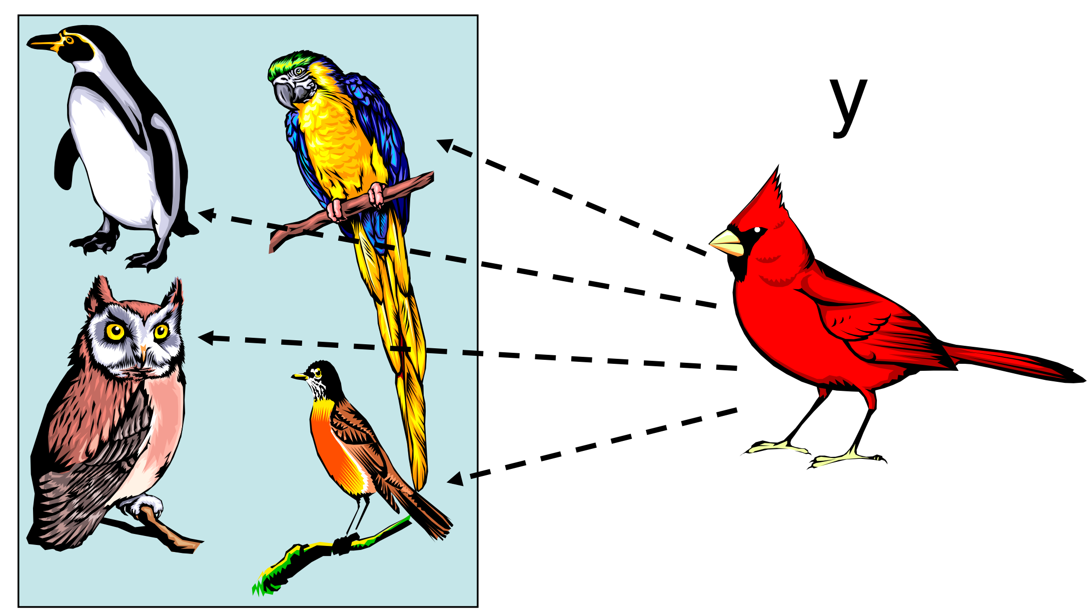
Category similarity

$$\textbf{sim}(y, C) = \sum_{x \in C} \textbf{sim}(y, x)$$

Probability of classification

$$P(y \in C) = \frac{\textbf{sim}(y, C)}{\sum_{C'} \textbf{sim}(y, C')}$$

$x \in C$



Birds
You've Seen

Bird?

Close similarity is critical

Compare two items (m = .3)

- Item Y overlaps with 2 features for each of 3 items
- Item Z perfectly matches 1 item, but for 2 items not at all
- Similarity of Y to category = .09 + .09 + .09 = .27
- Similarity of Z to category = 1 + 0 + 0 = 1

- Z is much more similar to the category, even though it has only 4 matching features compared to 6 for Y
  - critically, Y would likely be favored in prototype models
- <u>Configural</u> similarity is important

The mismatch parameter *m*

A bit of a "catch all" for various factors:

- The intrinsic mismatch between the stimulus feature and the exemplar's feature;
  - red vs. green could have lower *m* than red vs. maroon
- Attention to the dimension (due to learning)
- (M&S also suggest forgetting; if you don't remember the feature, mismatch will have little effect)

Another worry is that the context model is only defined for discrete features

- Other models have separated some of these into different variables, and work for continuous features (Nosofsky's Generalized Context Model, and ALCOVE)

Medin & Schaffer Experiment 2

famous “5-4” category structure

TRAINING STIMULI

"A" STIMULI

STIMULUS NUMBER	DIMENSION VALUES C	F	S	N	FE	RATING
4	1	1	1	0	4.9	4.8
7	1	0	1	0	3.3	5.4
15	1	0	1	1	3.2	5.1
13	1	1	0	1	4.8	5.2
5	0	1	1	1	4.5	5.2

Prototype: 1 1 1 1

"B" STIMULI

STIMULUS NUMBER	DIMENSION VALUES C	F	S	N	FE	RATING
12	1	1	0	0	5.5	5.0
2	0	1	1	0	5.2	5.1
14	0	0	0	1	3.9	5.2
10	0	0	0	0	3.1	5.5

Prototype: 0 0 0 0

Key comparison is stimulus 4 vs stimulus 7

- Prototype model would predict stimulus 4 is easier to learn
  - It’s more similar to the prototype
- Exemplar model would predict stimulus 7 is easier to learn
  - Has two near neighbors, 15 and 4
- **Results favor exemplar model**: stimulus 7 had fewer error (FE) and higher confidence rating

What is an exemplar?

- Nosofsky (1998, JEP:LMC) asks what *IS* an exemplar? Is it an item or is it an experience?
  - e.g., is your dog an exemplar?
  - or is each experience of encountering your dog?

- He presented items with varying frequency.
  - If each <u>item</u> is an exemplar, then the frequency shouldn't have any effect
  - If each <u>experience</u> is an exemplar, then typicality and classification accuracy should be slanted towards more frequent exemplars

What is an exemplar?

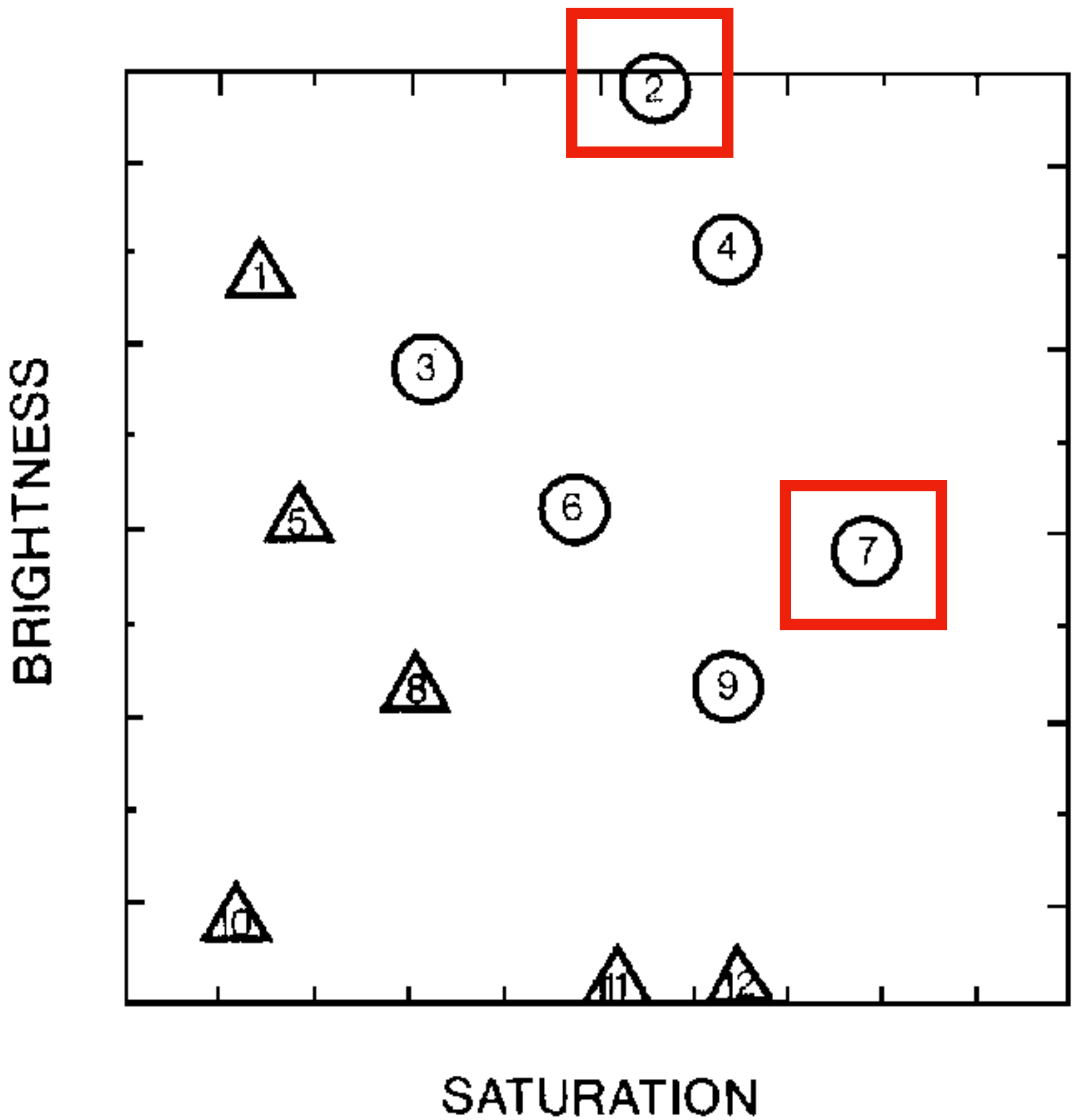


Figure 1. Category structure tested in Experiments 1 and 2. (Stimuli enclosed by triangles = members of Category 1; Stimuli enclosed by circles = members of Category 2)

Either item 2, or item 7 (depending on condition), was manipulated to appear 5x more often

What is an exemplar? It seems to be an experience

- Nosofsky found each experience was in fact “an exemplar.” That is, the frequency that individual exemplars occurred strongly influenced typicality

- Classification accuracy and typicality ratings increased for high-frequency exemplars, and also increased category memory that were similar to high-frequency exemplars

Issues with exemplar experiments

Artificial category structures tested are unrealistic

- e.g., example from Medin & Schaffer

People are smart — if you give them categories that aren't captured well by prototypes, they may do something else…

Category A

D1	D2	D3	D4
1	1	1	1
1	1	1	0
0	0	0	1

Category B

D1	D2	D3	D4
0	0	0	0
0	0	1	1
1	1	0	0

How much is learned? Medin & Schaffer's experiments

- 6 items, 20 blocks of learning
  - 16% didn't learn categories

- 9 items (famous "5-4" structure) 16 blocks
  - 44% didn't learn!!

- Same as Exp. 2, but with faces
  - no one learns a damn thing

- 11 items, 16 blocks
  - 50% didn't learn

- Exercise: Can you think of any real-world, pair of categories that you need more than 1 feature to discriminate?

Acknowledgements

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