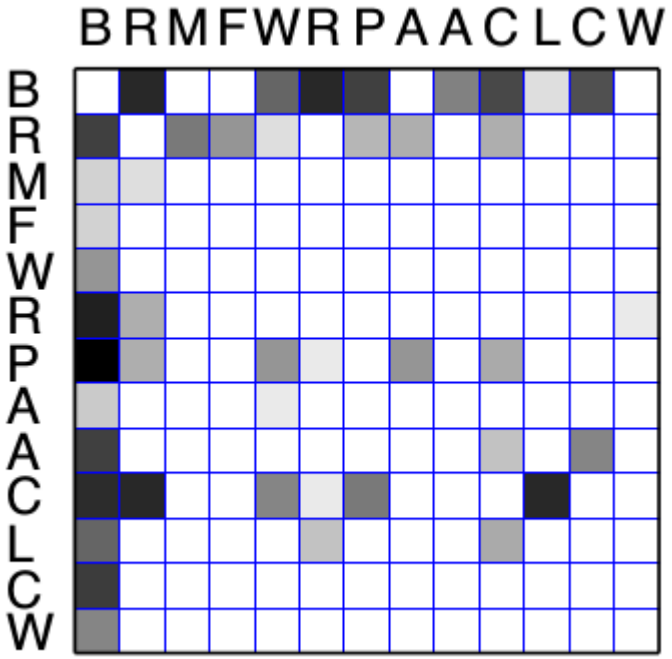
Similarity

Cogsci 131

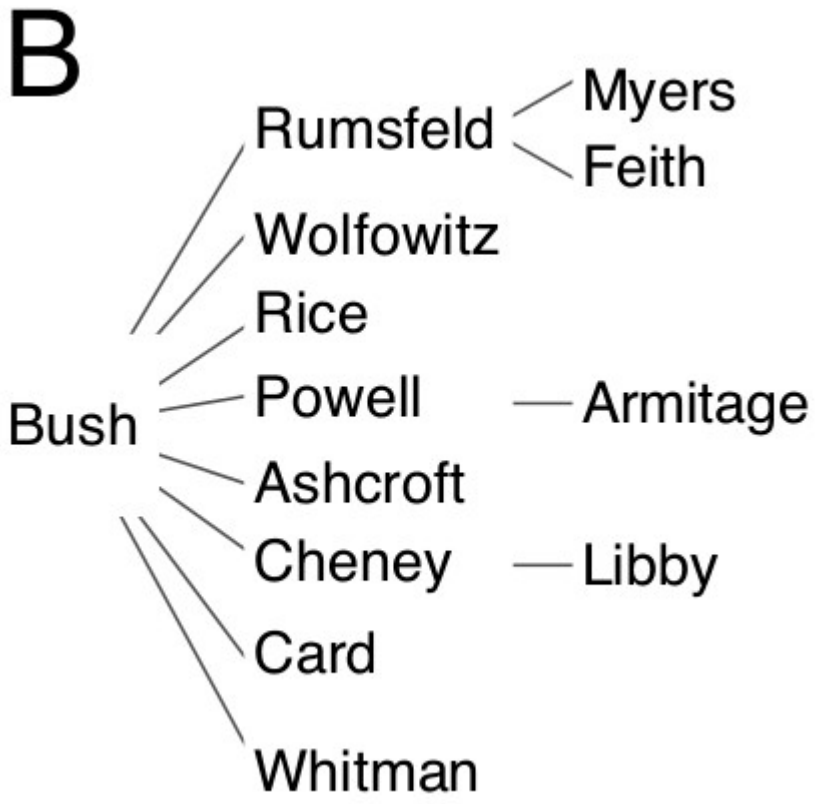
Kemp & Tenenbaum (2008)

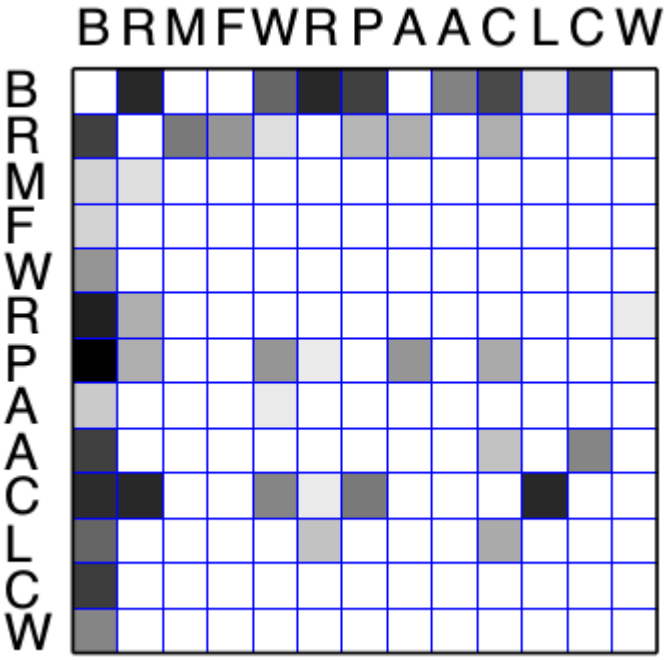
- **In many cases, psychological space has interesting, nontrivial structure.**
- **How do you learn these domains?**

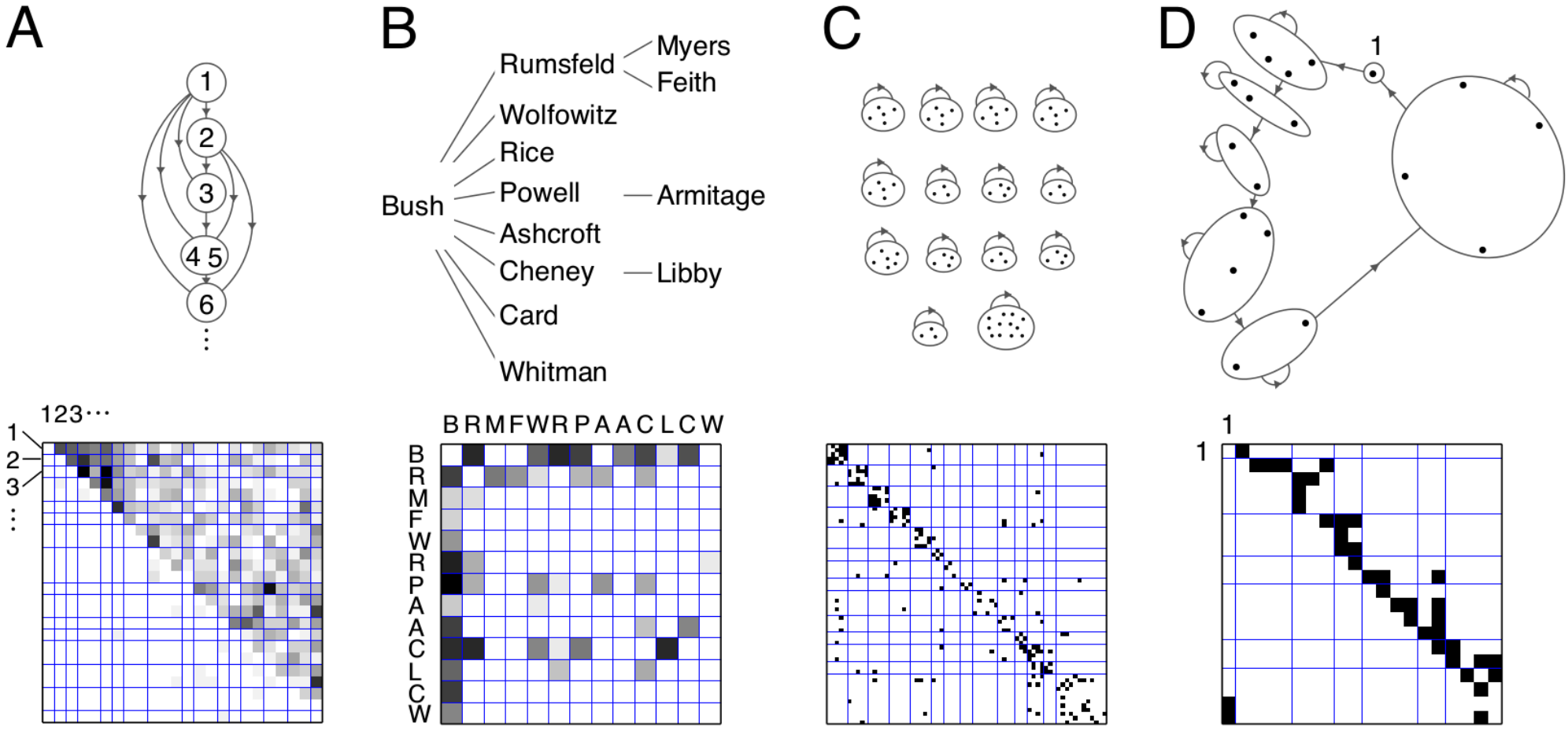
+ relational data (or similarity, or feature-based)



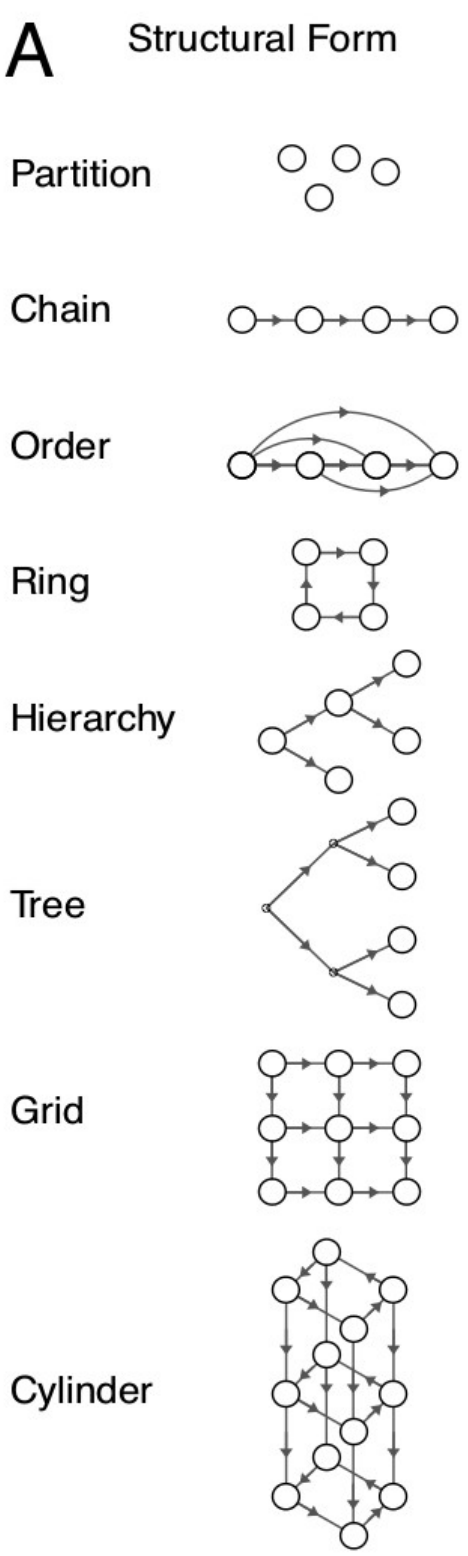
The model infers the structure behind the observed relational data

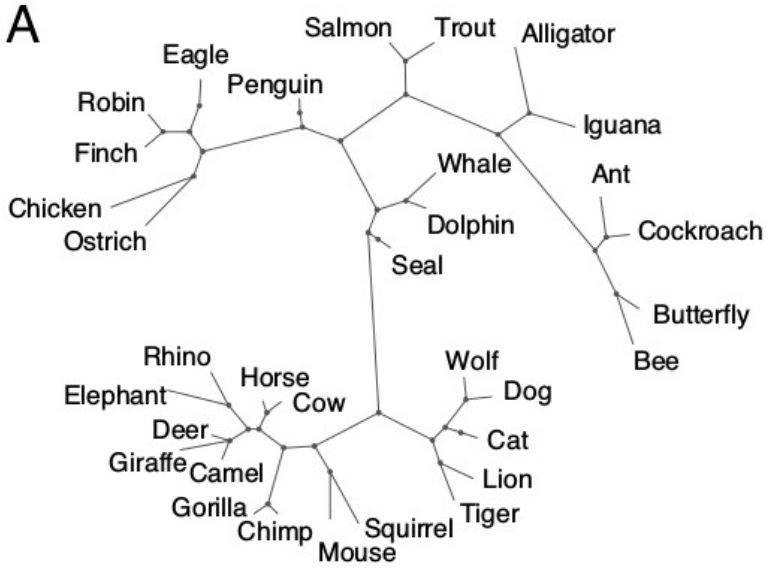


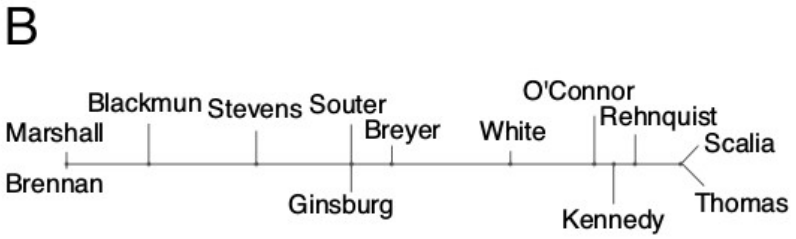


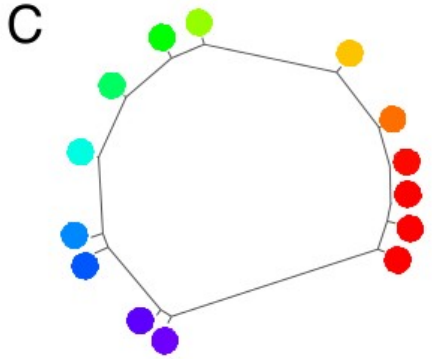

A
1
2
3
4 5
6
123⋯
1
2
3
B
Rumsfeld
Myers
Feith
Wolfowitz
Rice
Bush
Powell
Armitage
Ashcroft
Cheney
Libby
Card
Whitman
B R M F W R P A A C L C W
B R M F W R P A A C L C W
C
D
1
1
1

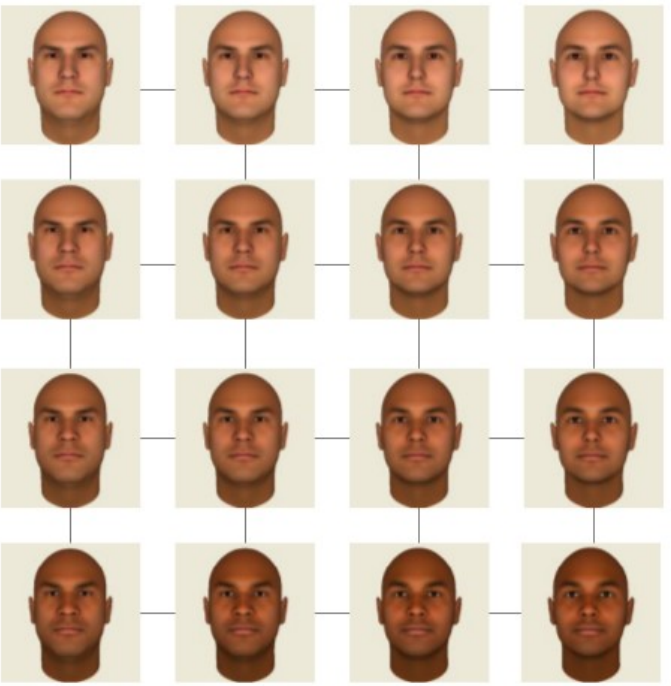
A rich set of hypotheses:

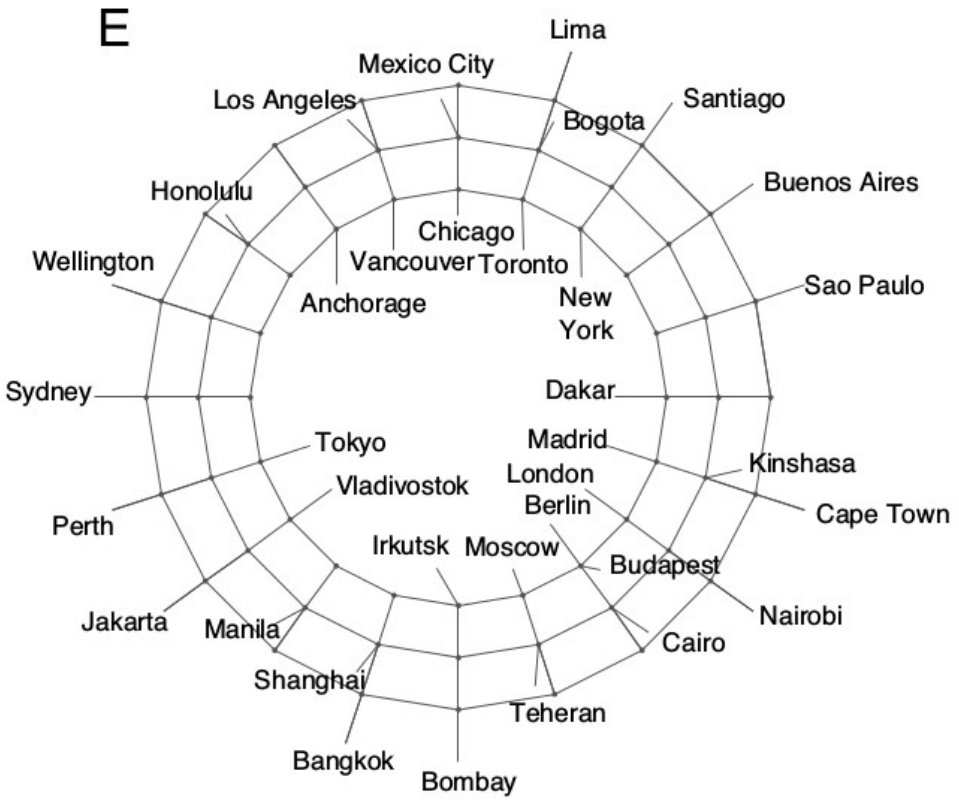



A
Salmon
Trout
Alligator
Eagle
Penguin
Robin
Iguana
Finch
Whale
Ant
Chicken
Dolphin
Cockroach
Ostrich
Seal
Butterfly
Bee
Rhino
Wolf
Horse
Elephant
Cow
Dog
Deer
Cat
Giraffe
Camel
Lion
Gorilla
Tiger
Chimp
Squirrel
Mouse

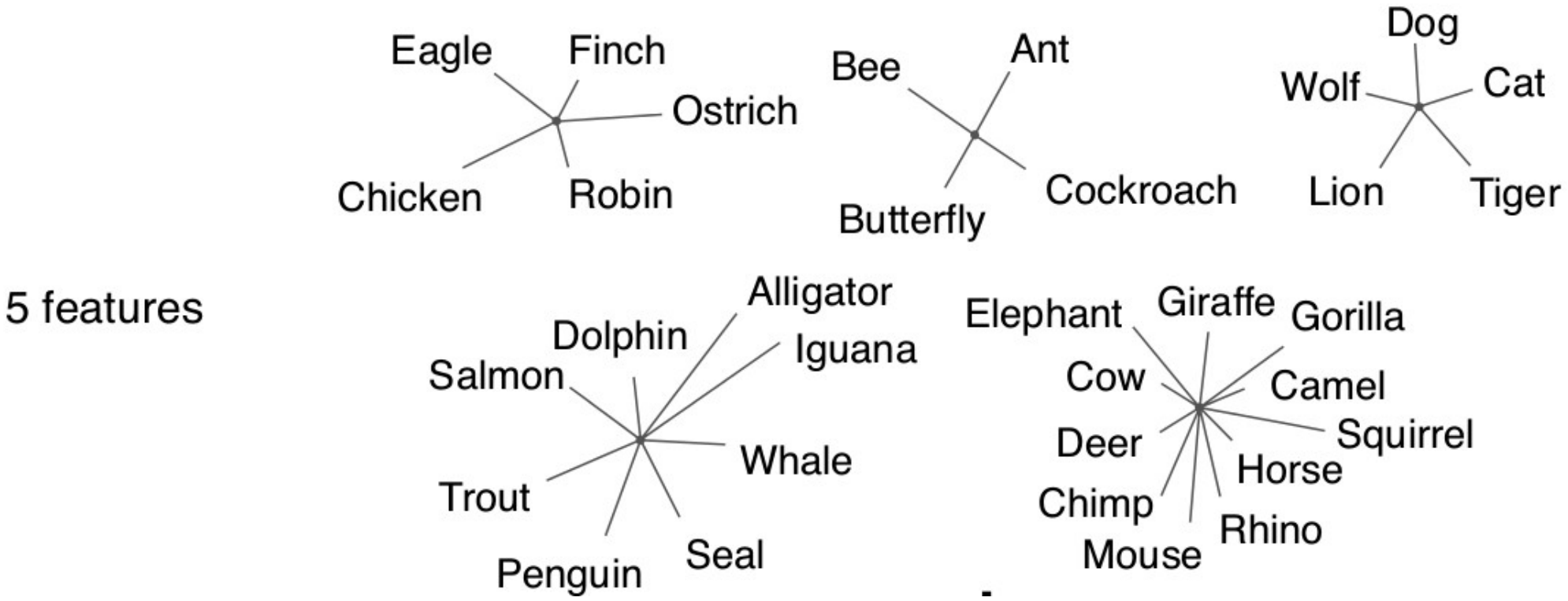

B
Blackmun
Stevens
Souter
O'Connor
Marshall
Breyer
White
Rehnquist
Scalia
Brennan
Ginsburg
Kennedy
Thomas


C

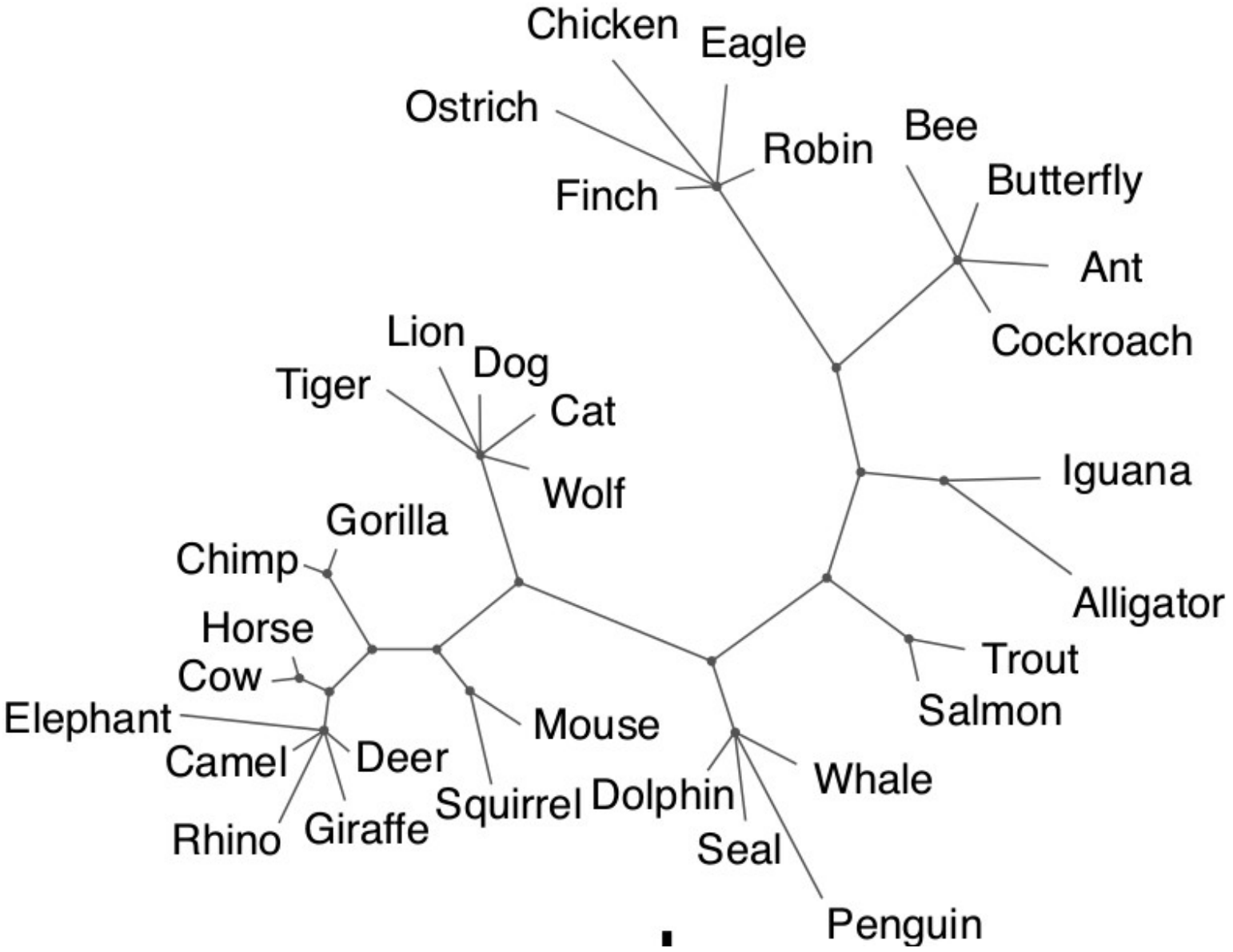

D


E
Lima
Mexico City
Los Angeles
Santiago
Bogota
Buenos Aires
Honolulu
Chicago
Wellington
Vancouver
Toronto
Sao Paulo
Anchorage
New York
Sydney
Dakar
Tokyo
Madrid
Kinshasa
London
Vladivostok
Berlin
Cape Town
Perth
Irkutsk
Moscow
Budapest
Nairobi
Jakarta
Manila
Cairo
Shanghai
Teheran
Bangkok
Bombay

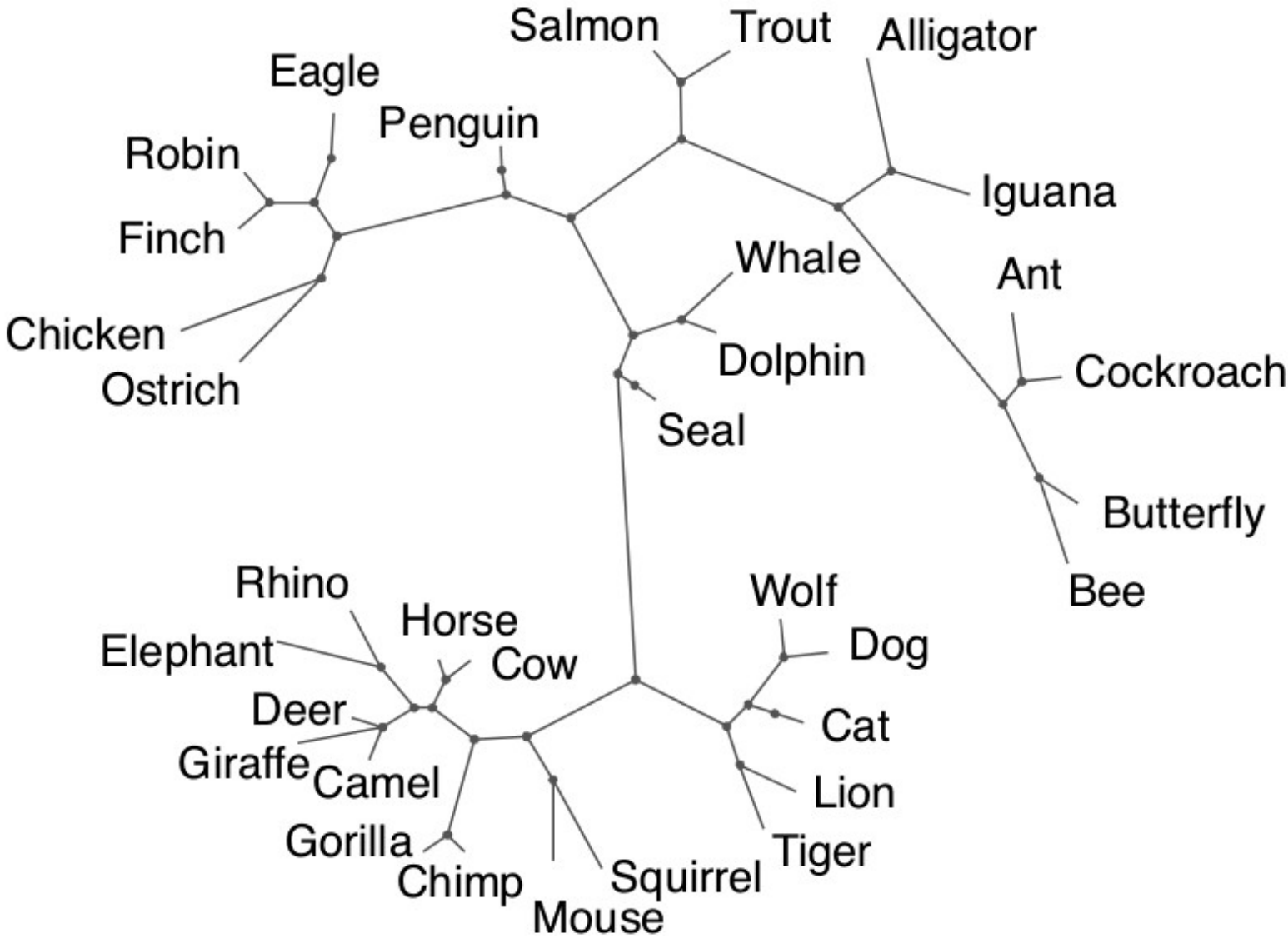
Also can model developmental trends



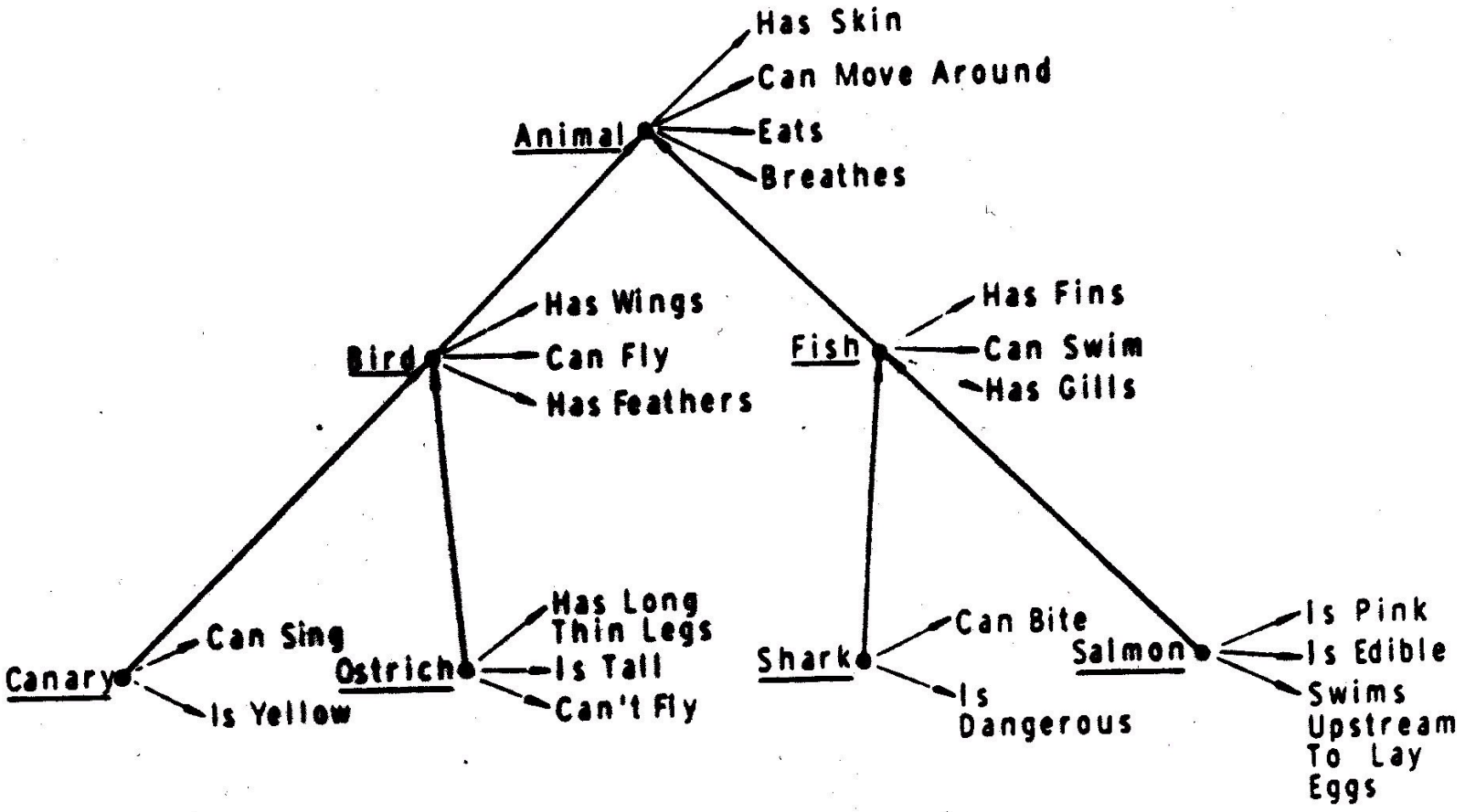
20 features

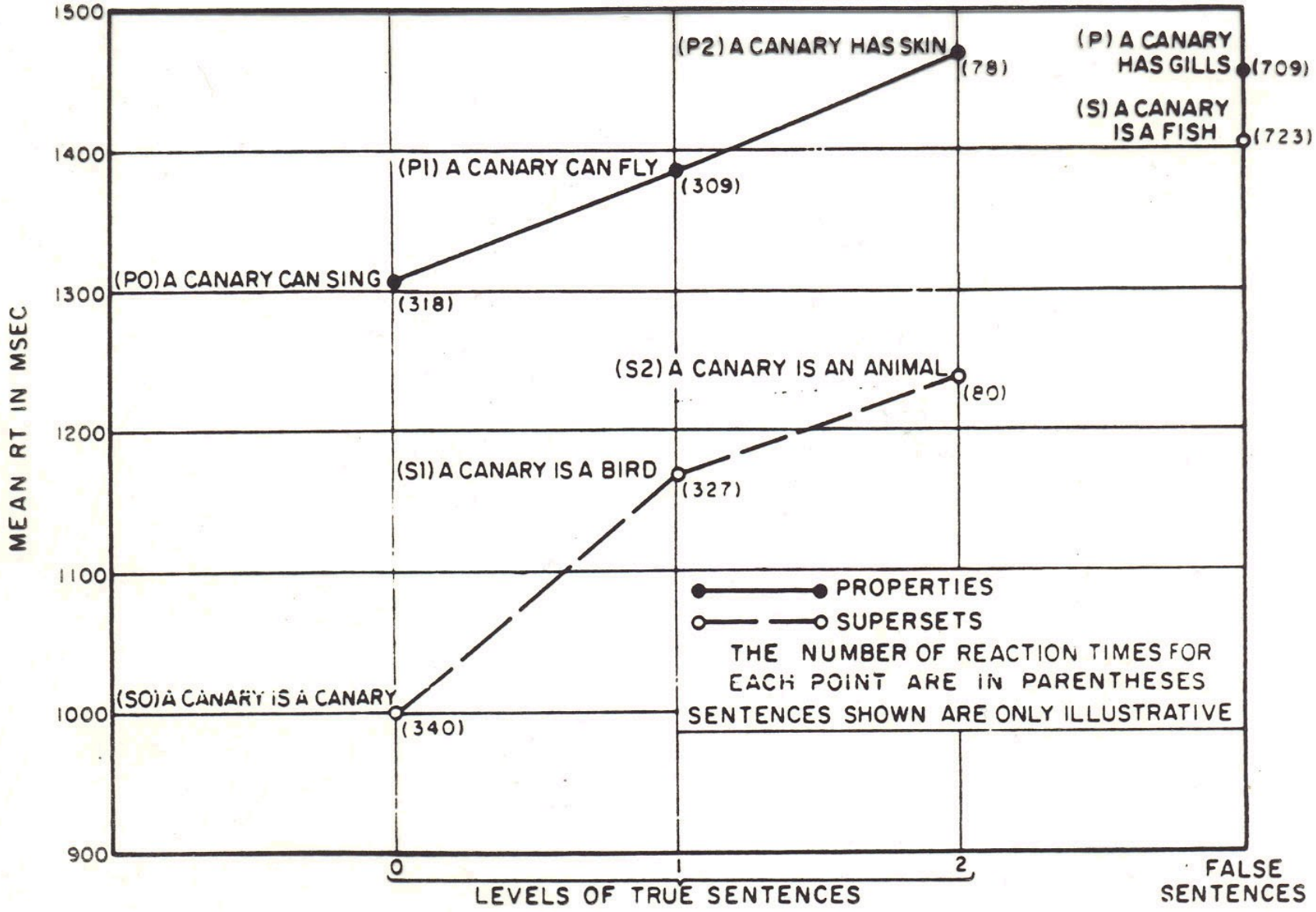

Chicken
Eagle
Ostrich
Robin
Bee
Finch
Butterfly
Ant
Cockroach
Lion
Dog
Tiger
Cat
Wolf
Iguana
Gorilla
Chimp
Alligator
Horse
Trout
Cow
Salmon
Elephant
Mouse
Deer
Camel
Whale
Squirrel
Dolphin
Rhino
Giraffe
Seal
Penguin

110 features



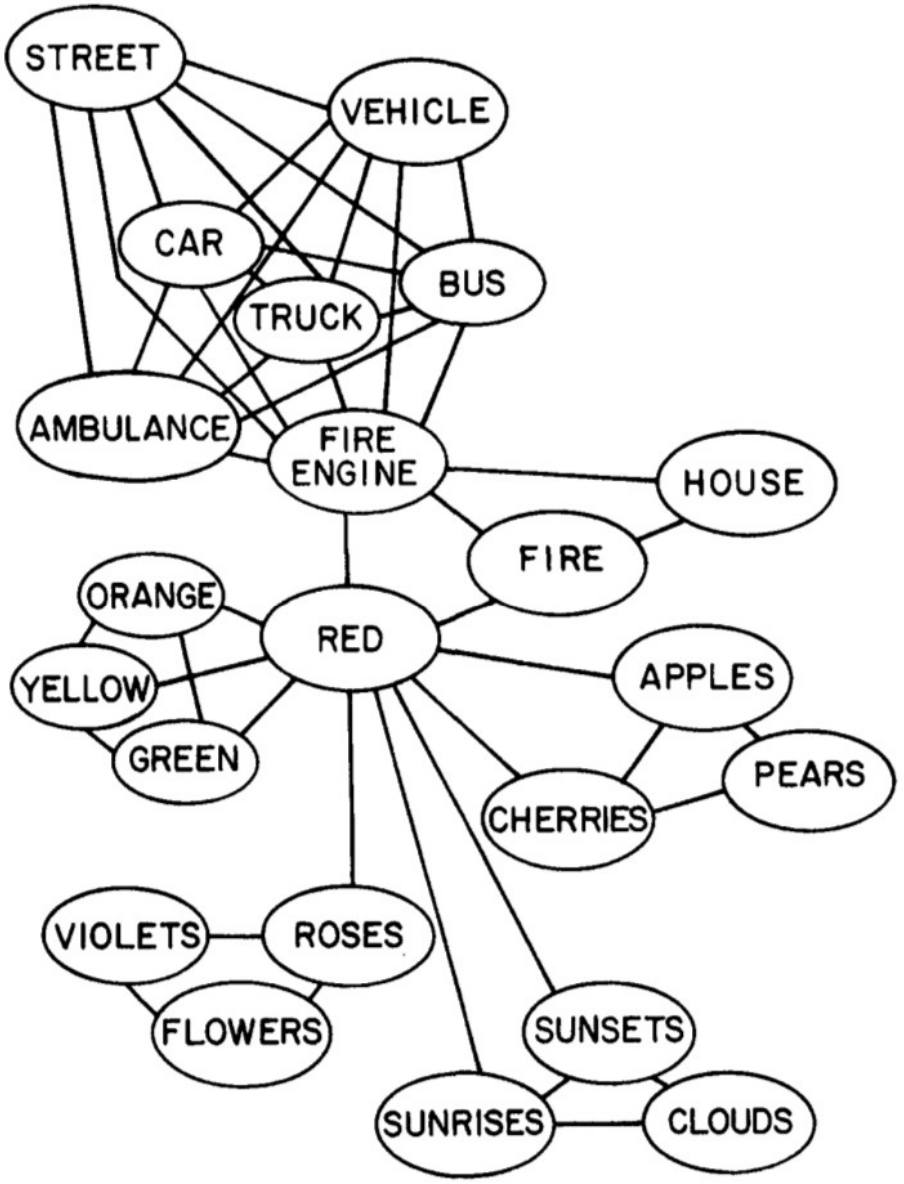
Quillian 1968




1500
1400
1300
1200
1100
1000
900
MEAN RT IN MSEC
(P2) A CANARY HAS SKIN
(78)
(P) A CANARY HAS GILLS
(709)
(S) A CANARY IS A FISH
(723)
(P1) A CANARY CAN FLY
(309)
(PO) A CANARY CAN SING
(318)
(S2) A CANARY IS AN ANIMAL
(80)
(S1) A CANARY IS A BIRD
(327)
(SO) A CANARY IS A CANARY
(340)
PROPERTIES
SUPERSETS
THE NUMBER OF REACTION TIMES FOR EACH POINT ARE IN PARENTHESES
SENTENCES SHOWN ARE ONLY ILLUSTRATIVE
0
1
2
LEVELS OF TRUE SENTENCES
FALSE SENTENCES

Semantic networks

(Collins & Loftus 1975)



Similarity?

- **So what is similarity exactly?**
- **Can it function as the “core” of conceptual representations?**

Properties of distance

- **A lot of work on similarity is motivated by thinking about properties of distance (in mathematics, called a metric)**

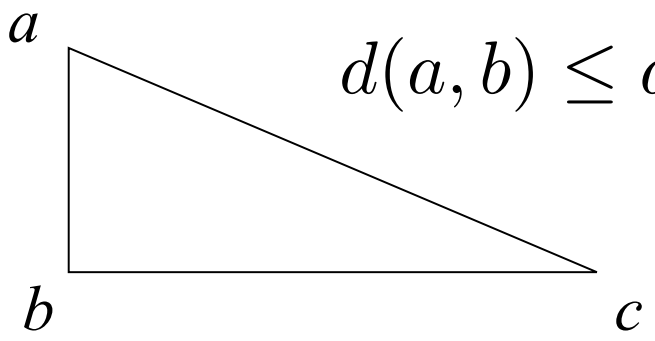
- **Distances are symmetric:**

$$d(a, b) = d(b, a)$$

- **Distances obey the triangle inequality:**

$$d(a, b) \leq d(a, c) + d(c, b)$$

Violations of triangle inequality



$$d(a,b) \leq d(a,c) + d(c,b)$$

Can find similarity judgments that violate this:

a

c

b

(Tversky, 1977)

Some complications of similarity

- **Some quirks of psychological similarity:**
  - Similarity and distance are not always inversely related (across studies)
  - Similarity is not always symmetric.
  - Similarity depends on context.

Similarity is not symmetric

- **Examples:**
  - 1007 is similar to 1000, but 1000 is not similar to 1007
  - An ellipse is similar to a circle, but a circle is not similar to an ellipse.

- Why is this a problem for MDS? Are distance measures in MDS necessarily symmetric?

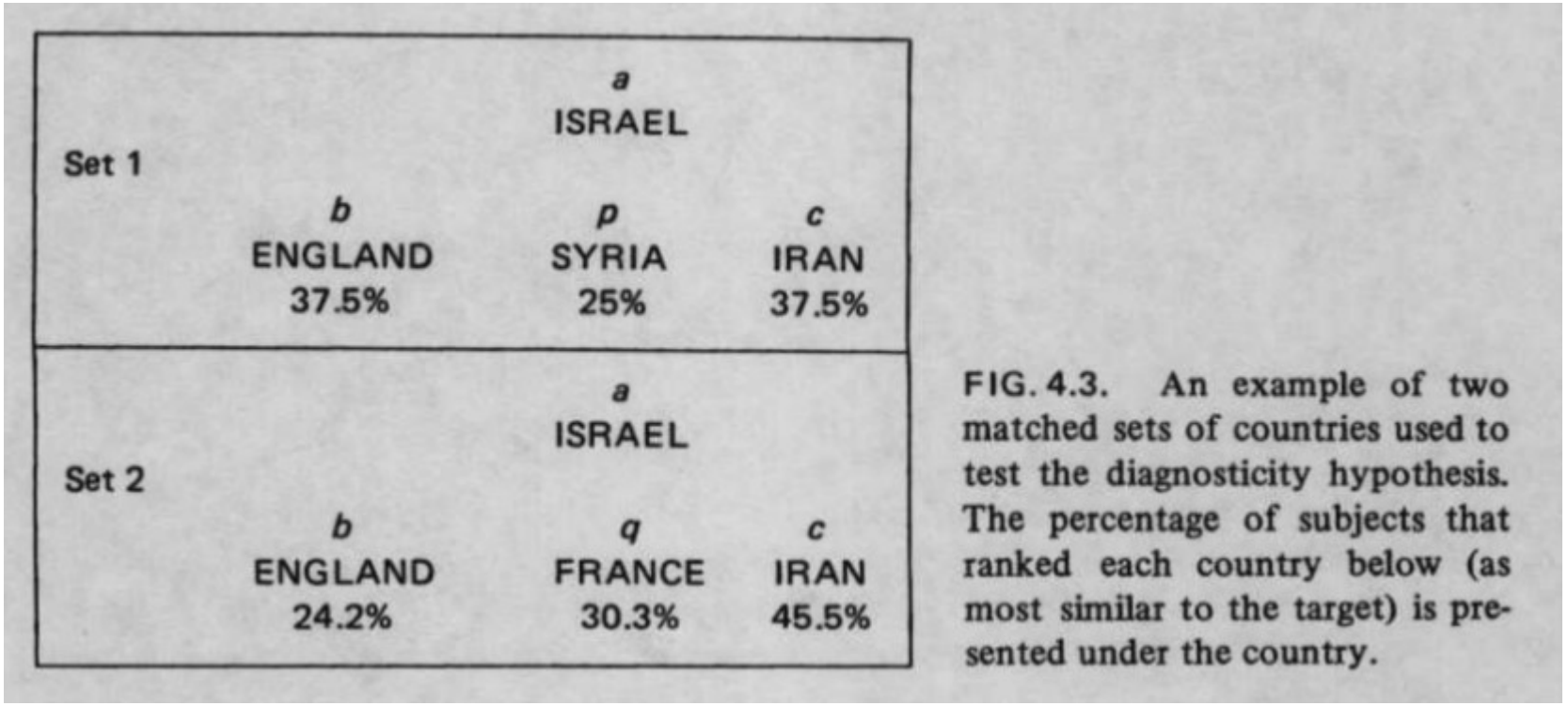
Tversky & Gati

TABLE 4.1
Percentage of Subjects That Selected the Prominent Pair in the Similarity Group (Π_s) and in the Difference Group (Π_d)

	Prominent Pairs	*Nonprominent Pairs*	Π_s	Π_d	$\Pi_s + \Pi_d$
1	W. Germany–E. Germany	Ceylon–Nepal	66.7	70.0	136.7
2	Lebanon–Jordan	Upper Volta–Tanzania	69.0	43.3	112.3
3	Canada–U.S.A.	Bulgaria–Albania	80.0	16.7	96.7
4	Belgium–Holland	Peru–Costa Rica	78.6	21.4	100.0
5	Switzerland–Denmark	Pakistan–Mongolia	55.2	28.6	83.8
6	Syria–Iraq	Liberia–Kenya	63.3	28.6	91.9
7	U.S.S.R.–U.S.A.	Paraguay–Ecuador	20.0	100.0	120.0
8	Sweden–Norway	Thailand–Burma	69.0	40.7	109.7
9	Turkey–Greece	Bolivia–Honduras	51.7	86.7	138.4
10	Austria–Switzerland	Zaire–Madagascar	79.3	24.1	103.4
11	Italy–France	Bahrain–Yemen	44.8	70.0	114.8
12	China–Japan	Guatemala–Costa Rica	40.0	93.1	133.1
13	S. Korea–N. Korea	Nigeria–Zaire	63.3	60.0	123.3
14	Uganda–Libya	Paraguay–Ecuador	23.3	65.5	88.8
15	Australia–S. Africa	Iceland–New Zealand	57.1	60.0	117.1
16	Poland–Czechoslovakia	Colombia–Honduras	82.8	37.0	119.8
17	Portugal–Spain	Tunis–Morocco	55.2	73.3	128.5
18	Vatican–Luxembourg	Andorra–San Marino	50.0	85.7	135.7
19	England–Ireland	Pakistan–Mongolia	80.0	58.6	138.6
20	Norway–Denmark	Indonesia–Philippines	51.7	25.0	76.7
	Average		59.1	54.4	113.5

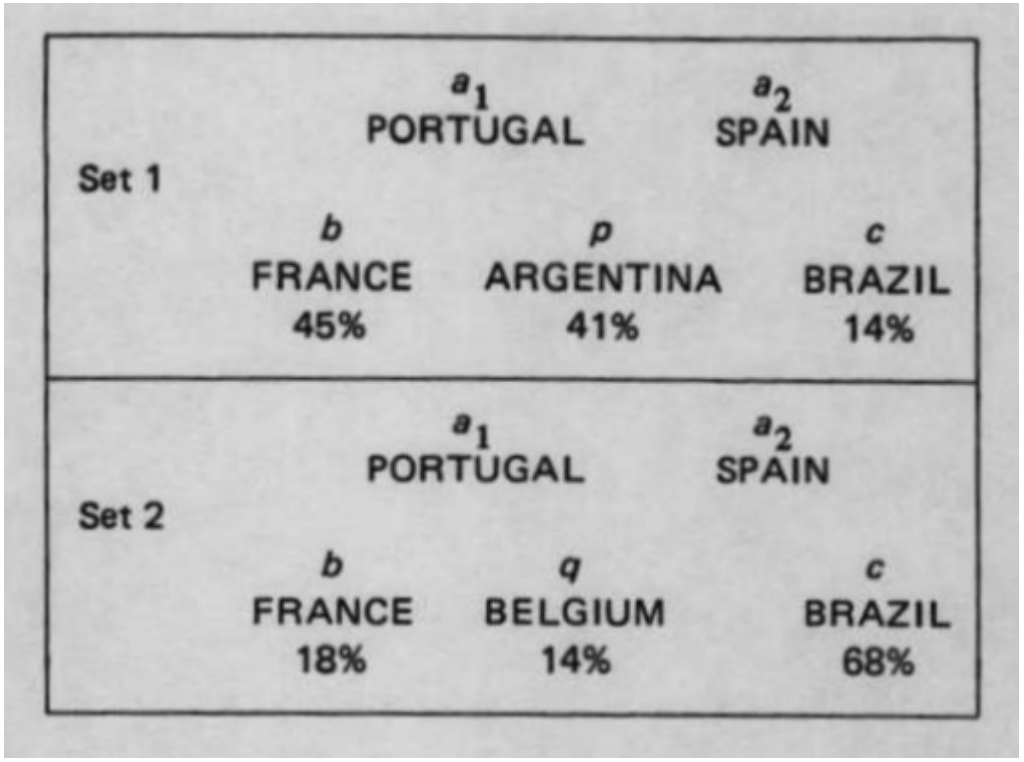
	p	q	Π
1	U.S.A.	Mexico	91.1
2	U.S.S.R.	Poland	98.6
3	China	Albania	94.1
4	U.S.A.	Israel	95.6
5	Japan	Philippines	94.2
6	U.S.A.	Canada	97.1
7	U.S.S.R.	Israel	91.1
8	England	Ireland	97.1
9	W. Germany	Austria	87.0
10	U.S.S.R.	France	82.4
11	Belgium	Luxembourg	95.6
12	U.S.A.	U.S.S.R.	65.7
13	China	N. Korea	95.6
14	India	Ceylon	97.1
15	U.S.A.	France	86.8
16	U.S.S.R.	Cuba	91.1
17	England	Jordan	98.5
18	France	Israel	86.8
19	U.S.A.	W. Germany	94.1
20	U.S.S.R.	Syria	98.5
21	France	Algeria	95.6

Similarity depends on context



Set 1		*a* ISRAEL	
	b ENGLAND 37.5%	*p* SYRIA 25%	*c* IRAN 37.5%
Set 2		*a* ISRAEL	
	b ENGLAND 24.2%	*q* FRANCE 30.3%	*c* IRAN 45.5%

FIG. 4.3. An example of two matched sets of countries used to test the diagnosticity hypothesis. The percentage of subjects that ranked each country below (as most similar to the target) is presented under the country.


Set 1
a_1
PORTUGAL
a_2
SPAIN
b
FRANCE
45%
p
ARGENTINA
41%
c
BRAZIL
14%
Set 2
a_1
PORTUGAL
a_2
SPAIN
b
FRANCE
18%
q
BELGIUM
14%
c
BRAZIL
68%

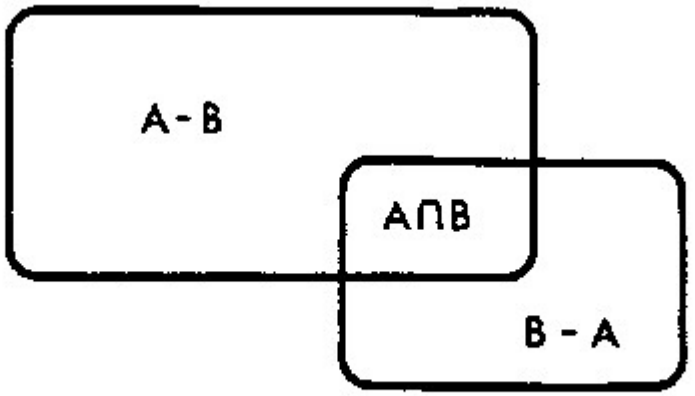
Similarity depends on context

TABLE 4.3
Average Similarities of Countries in Homogeneous (s_1) and Heterogeneous (s_2) Contexts

	Countries	$s_0(a, b)$	$s_e(a, b)$
American countries	Panama–Costa Rica	12.30	13.29
	Argentina–Chile	13.17	14.36
	Canada–U.S.A.	16.10	15.86
	Paraguay–Bolivia	13.48	14.43
	Mexico–Guatemala	11.36	12.81
	Venezuela–Colombia	12.06	13.06
	Brazil–Uruguay	13.03	14.64
	Peru–Ecuador	13.52	14.61
European countries	England–Ireland	13.88	13.37
	Spain–Portugal	15.44	14.45
	Bulgaria–Greece	11.44	11.00
	Sweden–Norway	17.09	15.03
	France–W. Germany	10.88	11.81
	Yugoslavia–Austria	8.47	9.86
	Italy–Switzerland	10.03	11.14
	Belgium–Holland	15.39	17.06

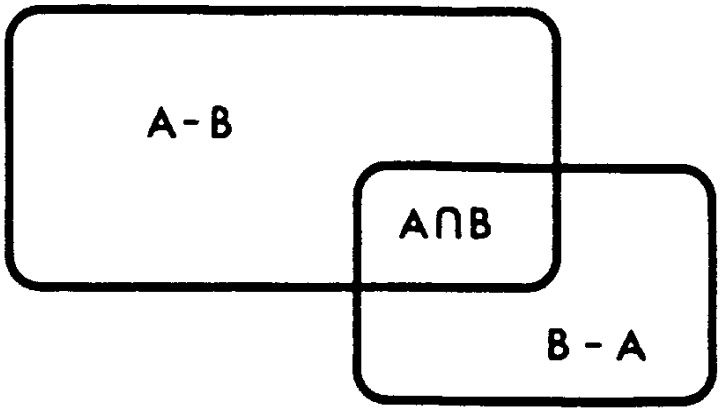
$$s(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

where $\theta, \alpha, \beta \geq 0$



$$s(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

where $\theta, \alpha, \beta \geq 0$



Tversky's contrast model

$$s(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

$$\text{where } \theta, \alpha, \beta \geq 0$$