Applying a Bayesian Measure of Representativeness to Sets of Images

Joshua T. Abbott, Katherine A. Heller,
Zoubin Ghahramani, and Thomas L. Griffiths

How do people determine which elements of a set are most representative of that set?









Representativeness

- Judgment and Decision Making: *Representativeness Heuristic*

(Kahneman and Tversky, 1972)

- Categorization: *Typicality*

(Mervis and Rosch, 1981)

Proposals

data *d* is representative of a hypothesized process or concept, *h*, if it is similar to the observations *h* typically generates

- Similarity
- Likelihood
- Bayesian

Bayesian measure

Good example of a concept - one that best provides evidence for the concept relative to possible alternatives

$$R(d, h_i) = \log \frac{P(d \mid h_i)}{\sum_{i \neq j} P(d \mid h_j) P'(h_j)}$$

where $P'(h_j) = \dfrac{P(h_j)}{1 - P(h_i)}$

(Tenenbaum and Griffiths, 2001)

Bayesian measure

d =

h_1 = *"fair coin"*, $P(h_1) = 0.9$

h_2 = *"two-headed coin"*, $P(h_2) = 0.05$

h_3 = *"weighted coin – heads 3/5"*, $P(h_3) = 0.05$

$$R(d, h_i) = \log \frac{P(d \mid h_i)}{\sum_{i \neq j} P(d \mid h_j) P'(h_j)}$$

Bayesian measure

d =

h_1 = "fair coin", $P(h_1)$ = 0.9
h_2 = "two-headed coin", $P(h_2)$ = 0.05
h_3 = "weighted coin – heads 3/5", $P(h_3)$ = 0.05

$$R(d, h_i) = \log \frac{P(d \mid h_i)}{\sum_{i \neq j} P(d \mid h_j) P'(h_j)}$$

$R(HHTHT, h_1)$ = 0.59

Bayesian measure

d =

h_1 = "fair coin", $P(h_1) = 0.9$
h_2 = "two-headed coin", $P(h_2) = 0.05$
h_3 = "weighted coin – heads 3/5", $P(h_3) = 0.05$

$$R(d, h_i) = \log \frac{P(d \mid h_i)}{\sum_{i \neq j} P(d \mid h_j) P'(h_j)}$$

$R(HHTHT, h_1) = 0.59$
$R(HHHHH, h_1) = -2.85$

HHTHT is more representative of a fair coin than HHHHH

Limitations

- Requires pre-defined concept/hypotheses
- Simple, artificial stimuli

Limitations

- Requires pre-defined concept/hypotheses
  Extend measure to sets of objects - with concepts generated on the fly

- Simple, artificial stimuli
  Evaluate on large database of naturalistic stimuli

Outline

- Representativeness and Bayesian Sets
- Application to a large image database
- Empirical Evaluation

Outline

- Representativeness and Bayesian Sets
- Application to a large image database
- Empirical Evaluation

Representativeness with Sets

$$R(d, h_i) = \log \frac{P(d \mid h_i)}{\sum_{i \neq j} P(d \mid h_j) P'(h_j)}$$ where $$P'(h_j) = \frac{P(h_j)}{1 - P(h_i)}$$

Representativeness with Sets

$$R(d, D_s) = \log \frac{P(d \mid D_s)}{\sum_{s \neq t} P(d \mid D_t) P'(D_t)} \quad \text{where} \quad P'(D_t) = \frac{P(D_t)}{1 - P(D_s)}$$

for a set of items $D_s = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset D$

Representativeness with Sets

$$R(d, D_s) = \log \frac{P(d \mid D_s)}{\sum_{s \neq t} P(d \mid D_t) P'(D_t)} \quad \text{where} \quad P'(D_t) = \frac{P(D_t)}{1 - P(D_s)}$$

for a set of items $D_s = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset D$

… but how do we compute this efficiently?


SWEET BABY RAY'S
The Sauce is the Boss!
SQUEEZABLE
SWEET BABY RAY'S
GOURMET SAUCES
Award Winning
Barbecue Sauce
NET WT. 18 OZ. (1 LB. 2 OZ.) 510g
STAY CLEAN CAP
French's
Classic
YELLOW
"BRING OUT THE BEST"
HELLMANN'S
REAL
MAYONNAISE
?
?
?

SWEET BABY RAY'S
The Sauce is the Boss!
A1
STEAK SAUCE
STAY CLEAN CAP
French's
Classic YELLOW
meijer
HONEY
SQUEEZABLE
SWEET BABY RAY'S
GOURMET SAUCES
Award Winning
Barbecue Sauce
HEINZ 57
HEINZ
TOMATO KETCHUP
"BRING OUT THE BEST"
HELLMANN'S
REAL
MAYONNAISE

“condiments”

Bayesian Sets

Given a data collection D and a subset of items $D_s = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ representing a concept, rank an item $\mathbf{x}^* \in \{ D \backslash D_s \}$

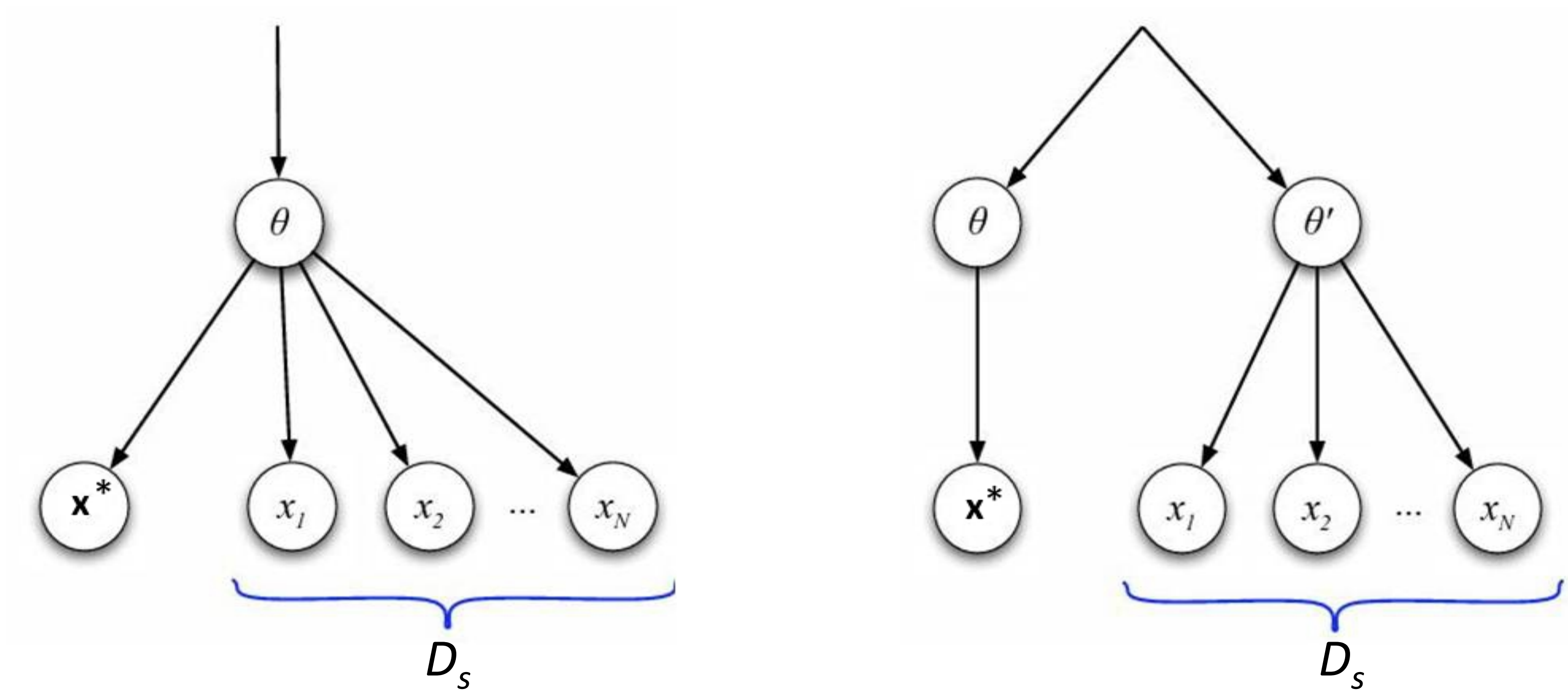
$$Bscore(\mathbf{x}^*) = \frac{p(\mathbf{x}^*, D_s)}{p(\mathbf{x}^*)\,p(D_s)}$$

(Ghahramani and Heller, 2005)

Bayesian Sets

Given a data collection D and a subset of items $D_s = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ representing a concept, rank an item $\mathbf{x}^* \in \{ D \backslash D_s \}$

$$Bscore(\mathbf{x}^*) = \frac{p(\mathbf{x}^*, D_s)}{p(\mathbf{x}^*)\,p(D_s)}$$



Bayesian Sets

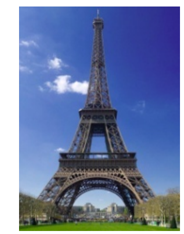
Given a data collection *D* and a subset of items $D_s = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ representing a concept, rank an item $\mathbf{x}^* \in \{ D \backslash D_s \}$

$$Bscore(\mathbf{x}^*) = \frac{p(\mathbf{x}^*, D_s)}{p(\mathbf{x}^*)\, p(D_s)}$$

for sparse binary data, can be computed efficiently as a single matrix-vector multiplication

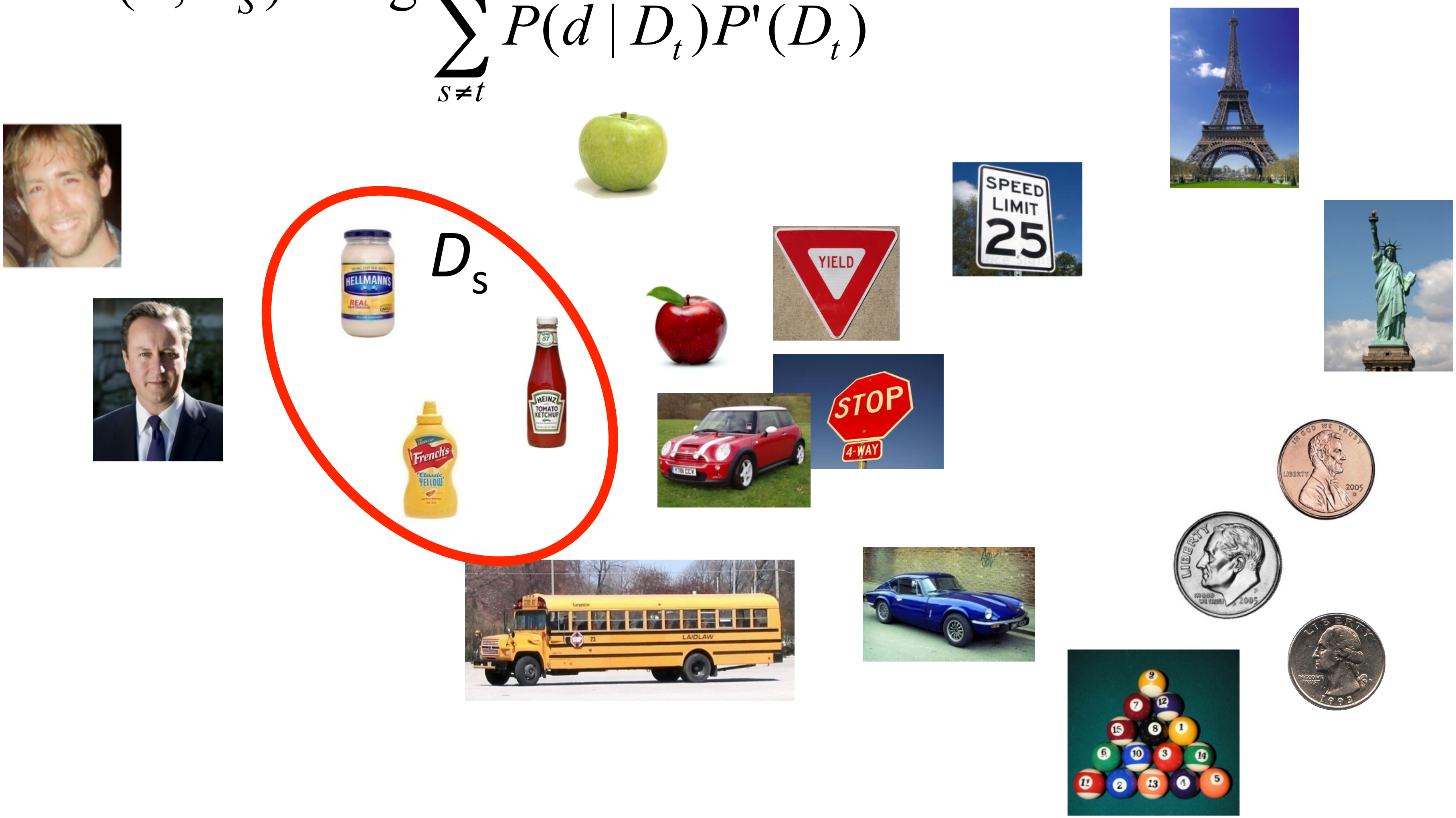
Representativeness and Bayesian Sets

$$R(d, D_s) = \log \frac{P(d \mid D_s)}{\sum_{s \neq t} P(d \mid D_t) P'(D_t)}$$



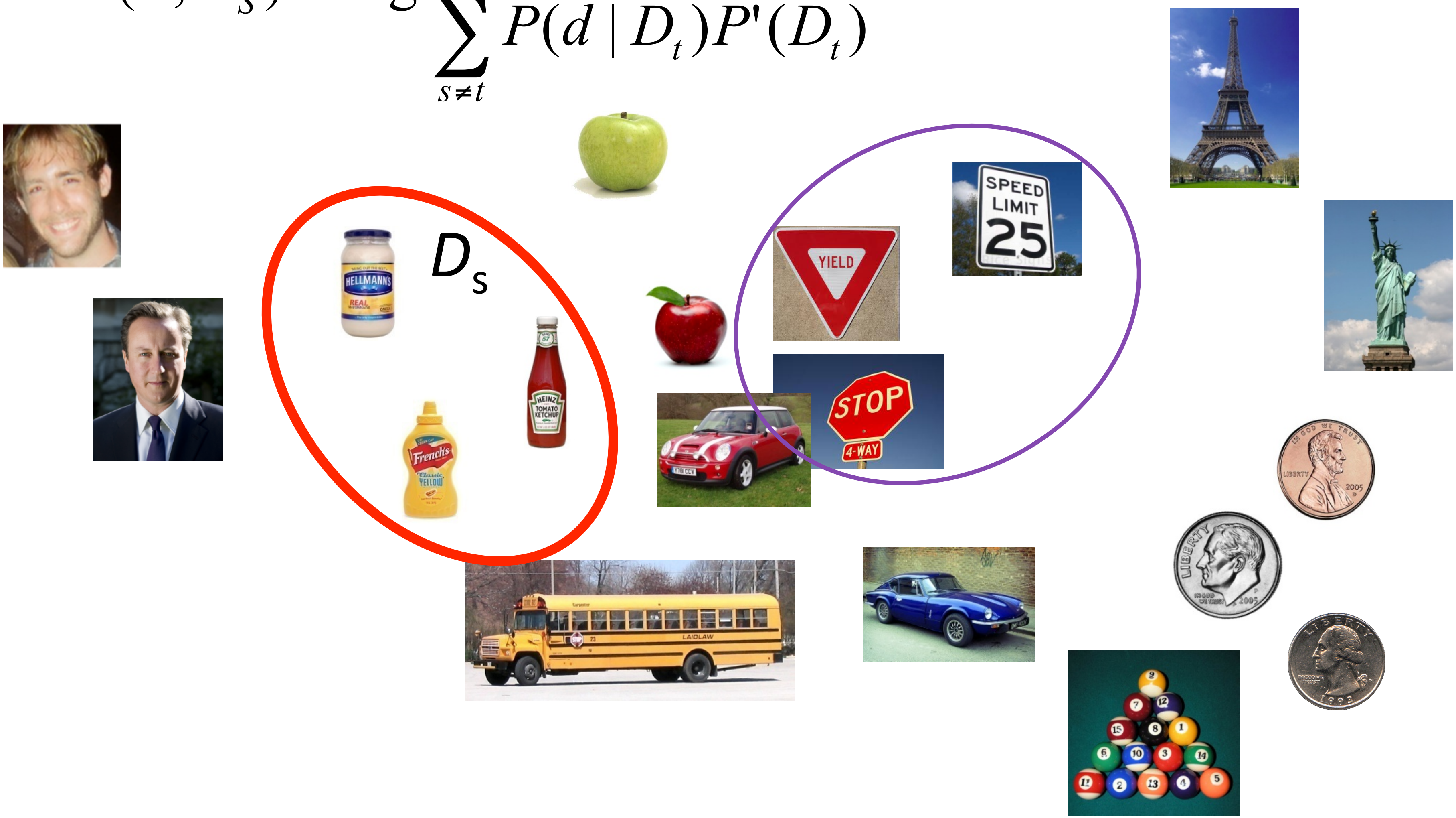
Representativeness and Bayesian Sets

$$R(d, D_s) = \log \frac{P(d \mid D_s)}{\sum_{s \neq t} P(d \mid D_t) P'(D_t)}$$



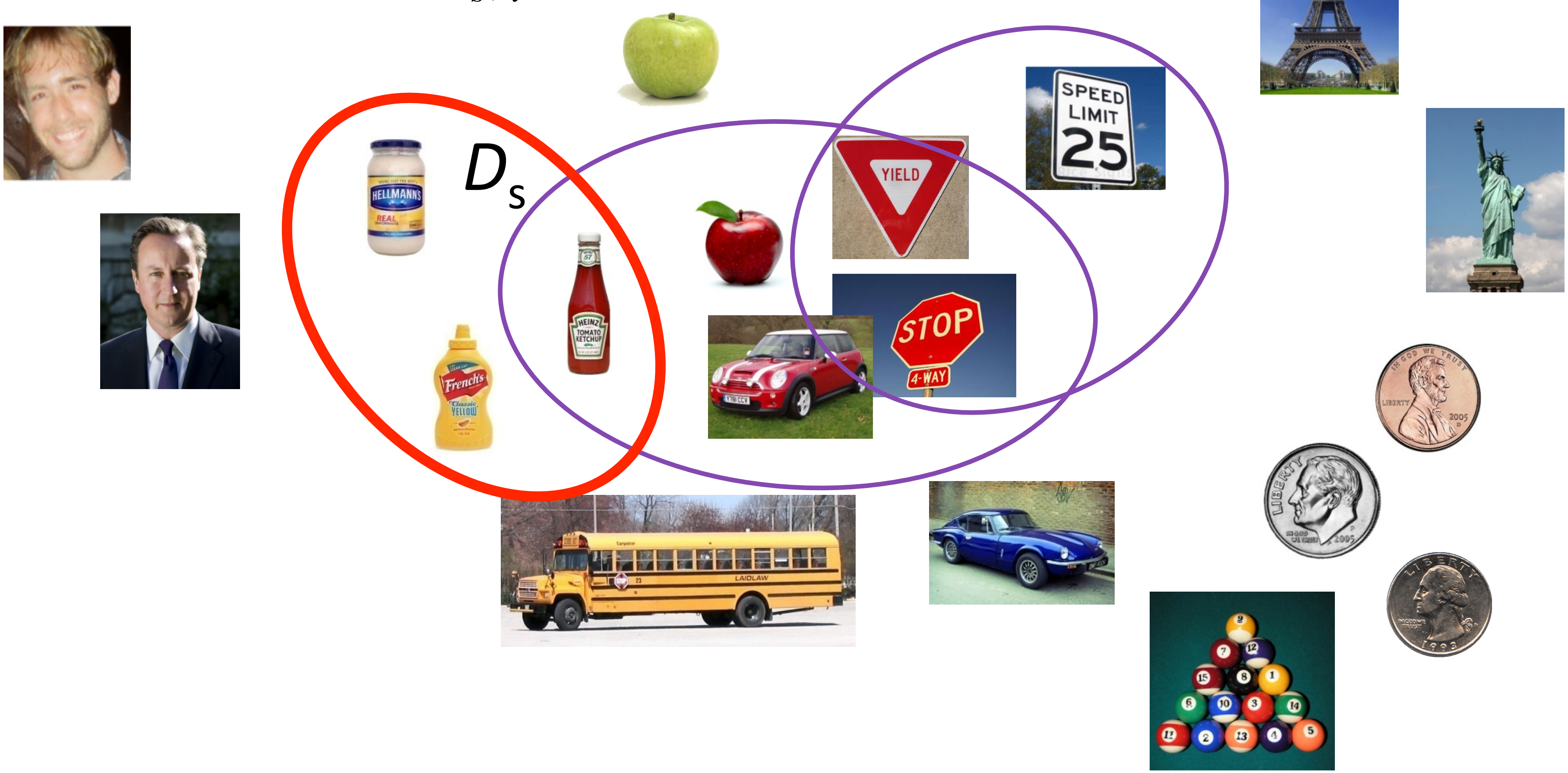
Representativeness and Bayesian Sets

$$R(d, D_S) = \log \frac{P(d \mid D_s)}{\sum_{s \neq t} P(d \mid D_t) P'(D_t)}$$



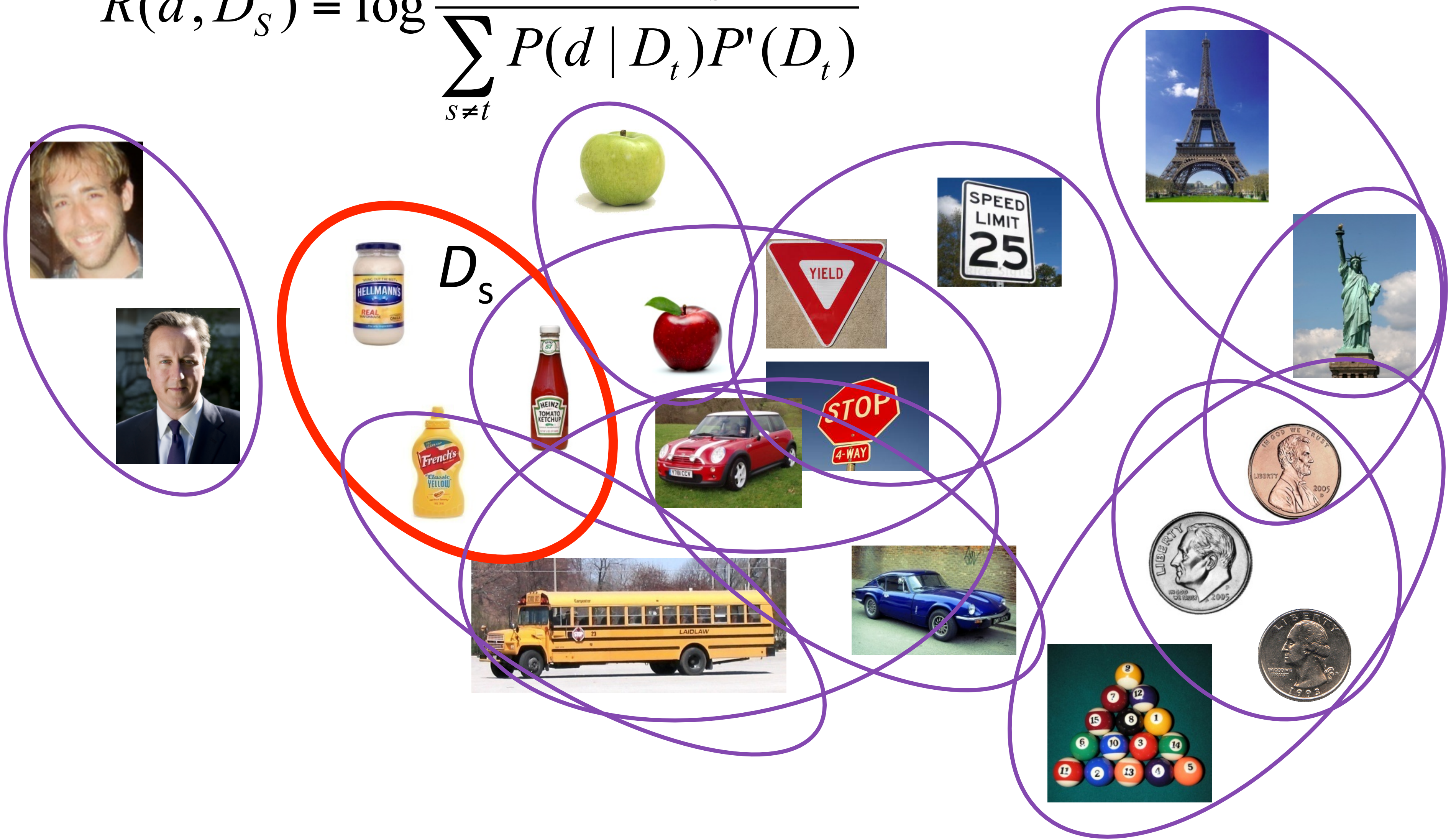
Representativeness and Bayesian Sets

$$R(d, D_S) = \log \frac{P(d \mid D_s)}{\sum_{s \neq t} P(d \mid D_t) P'(D_t)}$$



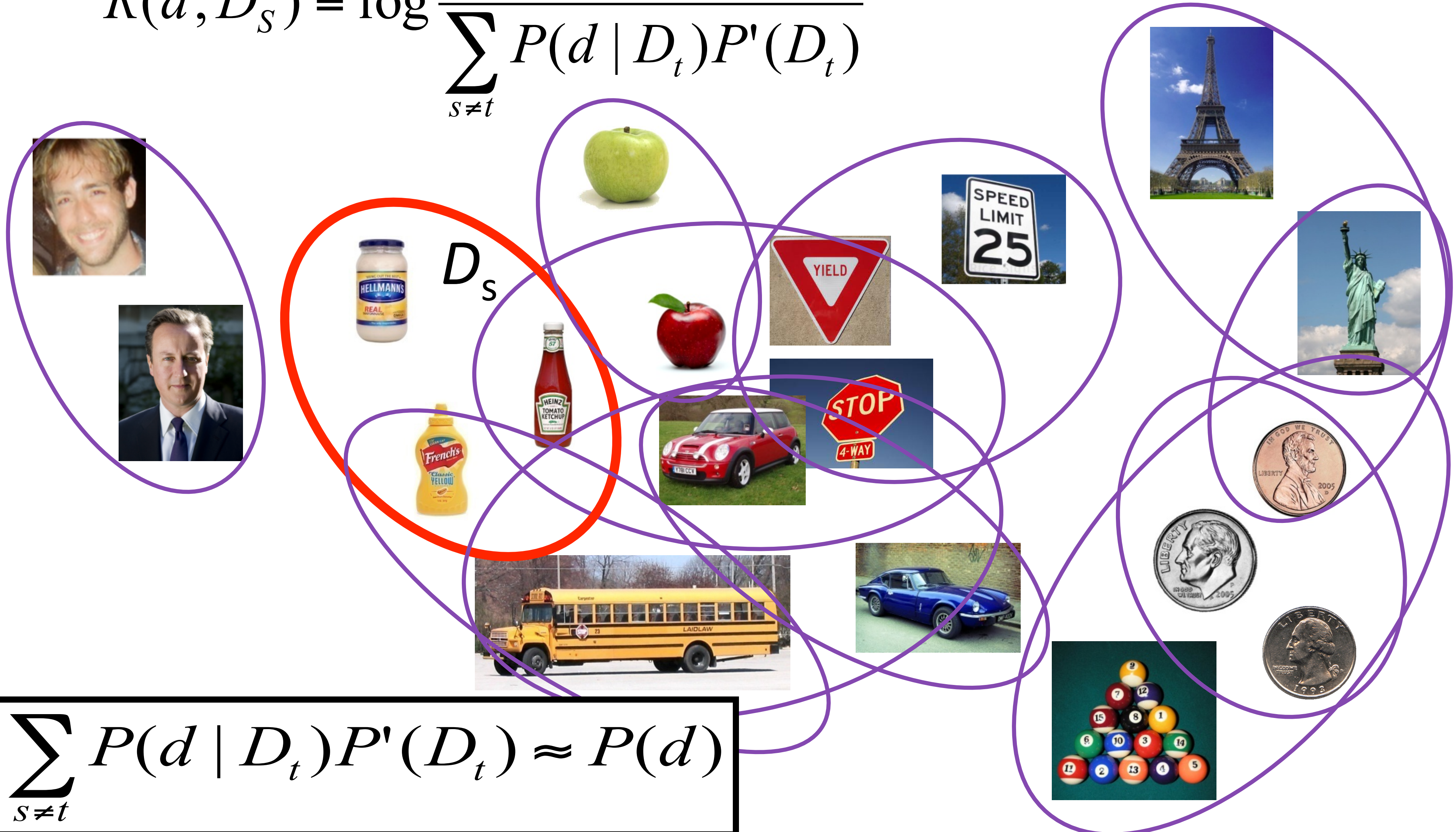
Representativeness and Bayesian Sets

$$R(d, D_s) = \log \frac{P(d \mid D_s)}{\sum_{s \neq t} P(d \mid D_t) P'(D_t)}$$



Representativeness and Bayesian Sets

$$R(d, D_S) = \log \frac{P(d \mid D_s)}{\sum_{s \neq t} P(d \mid D_t) P'(D_t)}$$



$$\boxed{\sum_{s \neq t} P(d \mid D_t) P'(D_t) \approx P(d)}$$

Representativeness and Bayesian Sets

$$R(d, D_S) = \log \frac{P(d \mid D_s)}{\sum_{s \neq t} P(d \mid D_t) P'(D_t)}$$

$$\approx \log \frac{P(d \mid D_s)}{P(d)}$$

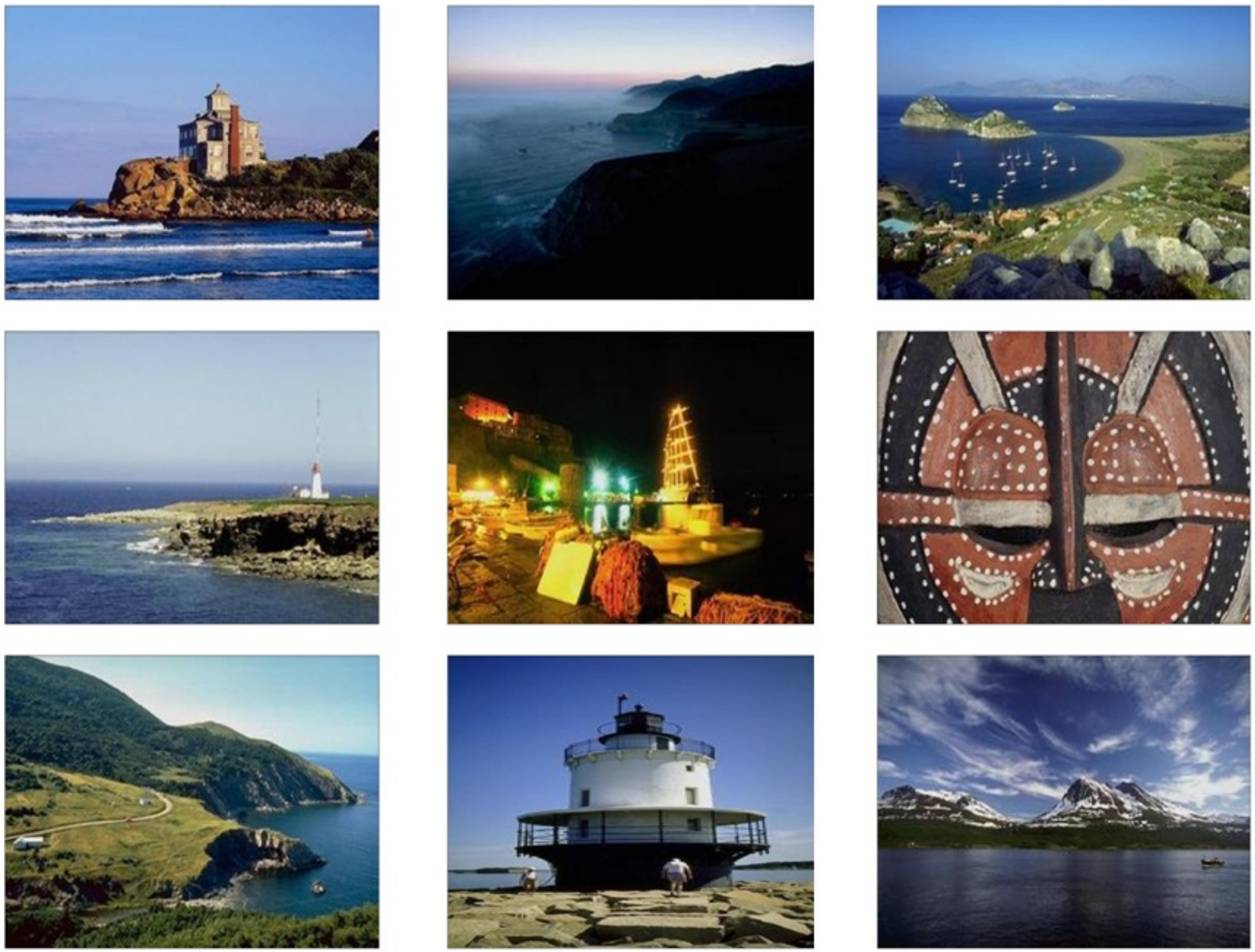
$$= \log \frac{P(d, D_s)}{P(d) P(D_s)}$$

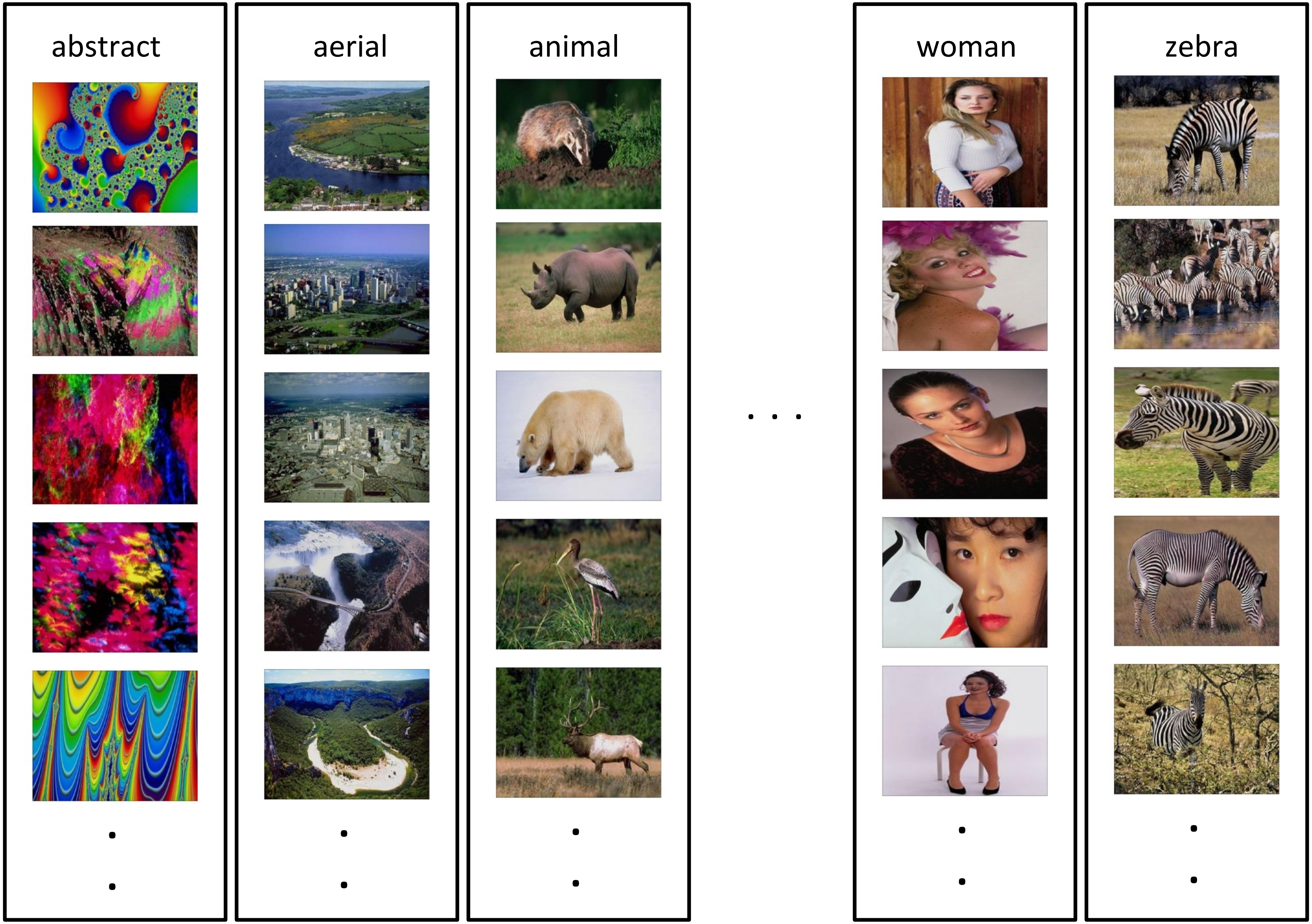
$$= \log Bscore(d)$$

Outline

- Representativeness and Bayesian Sets
- Application to a large image database
- Empirical Evaluation

How do people determine which images of a labeled set are most representative of that set?





(Heller and Ghahramani, 2006)

50 labeled sets depicting unique categories, with varying numbers of images per set (mean=264)

Images are represented as 240-D feature vectors:

- 48 Gabor texture features
- 27 Tamura texture features
- 165 color histogram features

Post-processed through binarization stage

(Heller and Ghahramani, 2006)

Representativeness framework

input: a set of items, D_w, for a particular
category label w

for each item $\mathbf{x}_i \in \mathrm{D}_w$ **do**

let $\mathrm{D}_{wi} = \{ \mathrm{D}_w \setminus \mathbf{x}_i \}$

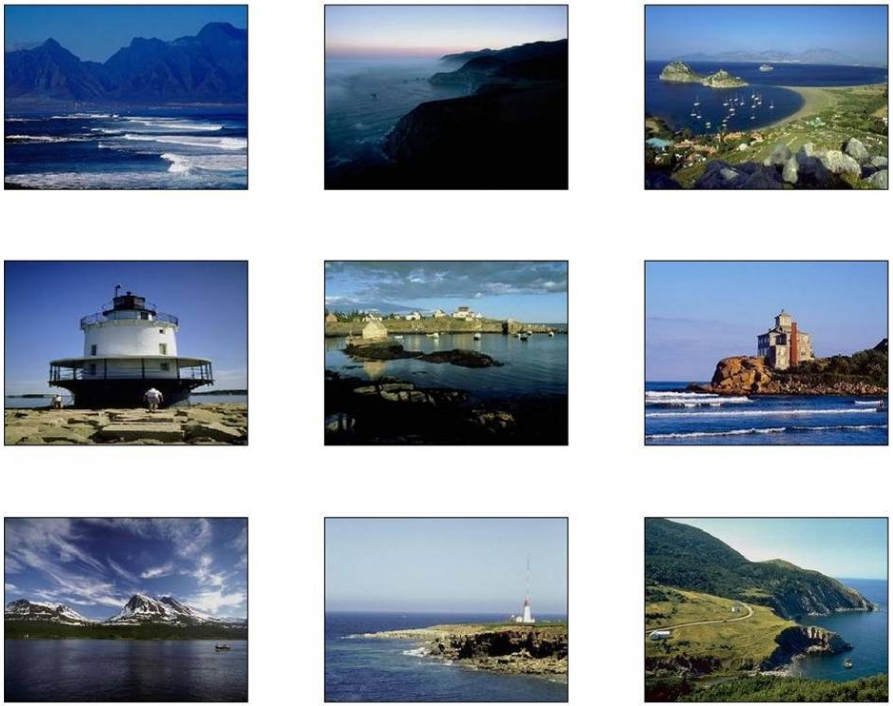
compute *score*($\mathbf{x}_i$, D_{wi})

end for

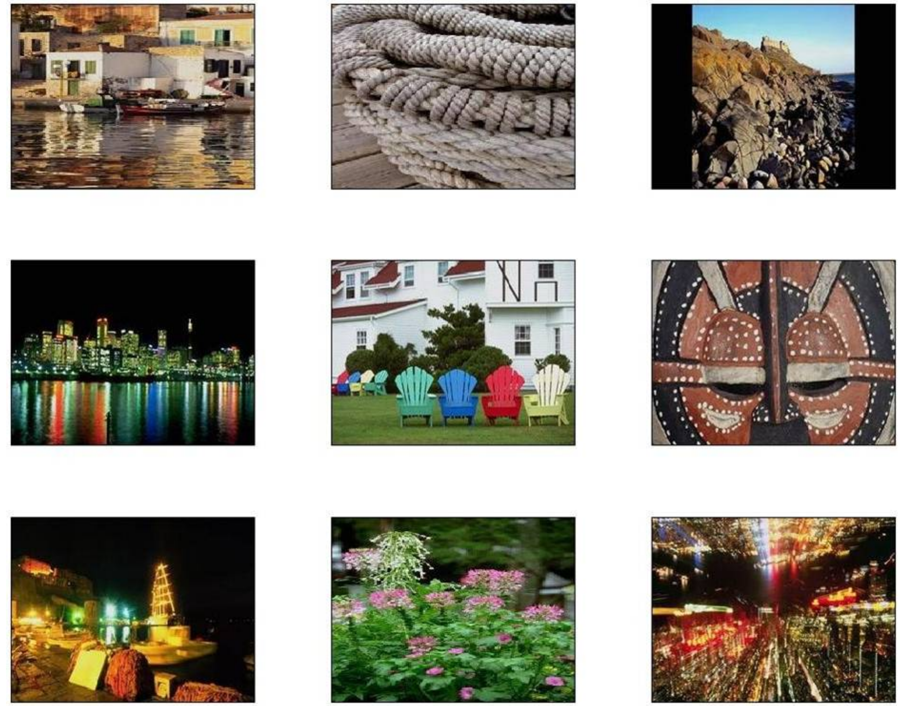
rank items in D_w by this score

output: ranked list of items in D_w

Top 9

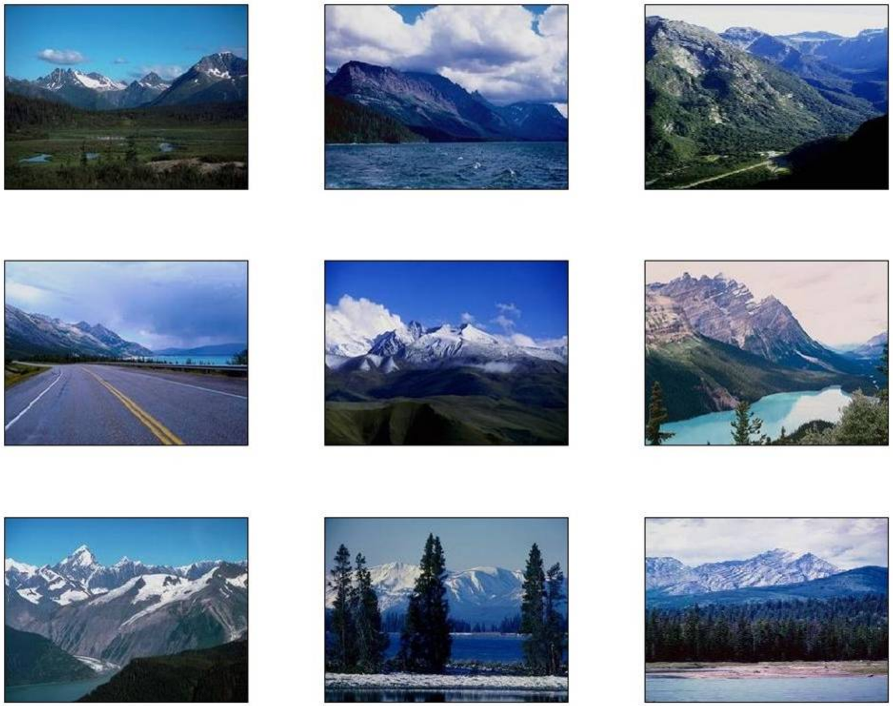


Bottom 9



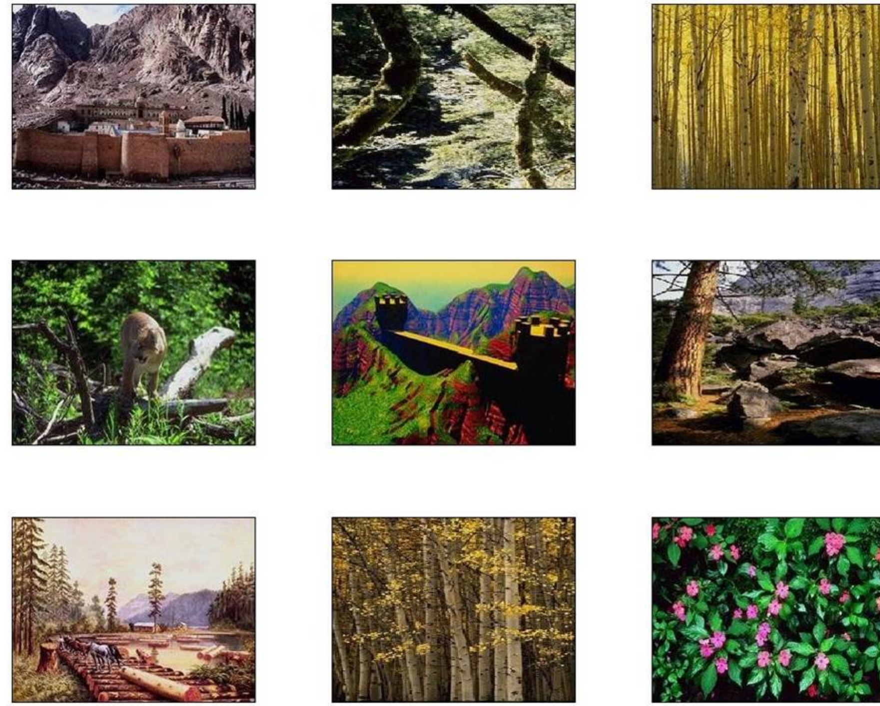
“coast”

Top 9



“mountains”

Bottom 9



Outline

- Representativeness and Bayesian Sets
- Application to a large image database
- Empirical Evaluation

Models of Representativeness

Bayesian Model

$$Bscore(\mathbf{x}^*) = \frac{p(\mathbf{x}^*, D_s)}{p(\mathbf{x}^*)p(D_s)}$$

Likelihood Model

$$Lscore(\mathbf{x}^*) = p(\mathbf{x}^* \mid D_s)$$

Prototype Model

$$Pscore(\mathbf{x}^*) = \exp\{-\lambda dist(\mathbf{x}^*, \mathbf{x}_{proto})\}$$

Exemplar Model

$$Escore(\mathbf{x}^*) = \sum_{x_j \in D_s} \exp\{-\lambda dist(\mathbf{x}^*, \mathbf{x}_j)\}$$

Method

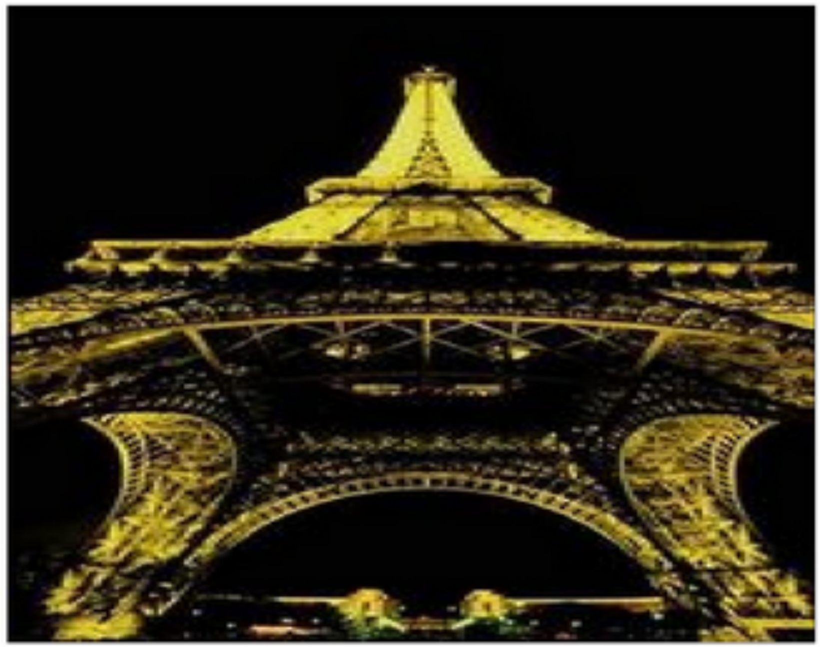
Participants: 500 (10 per category) via Amazon MT

Stimuli: Union of top 10 and bottom 10 ranked images per category, for each model*

*(excluding Exemplar model)

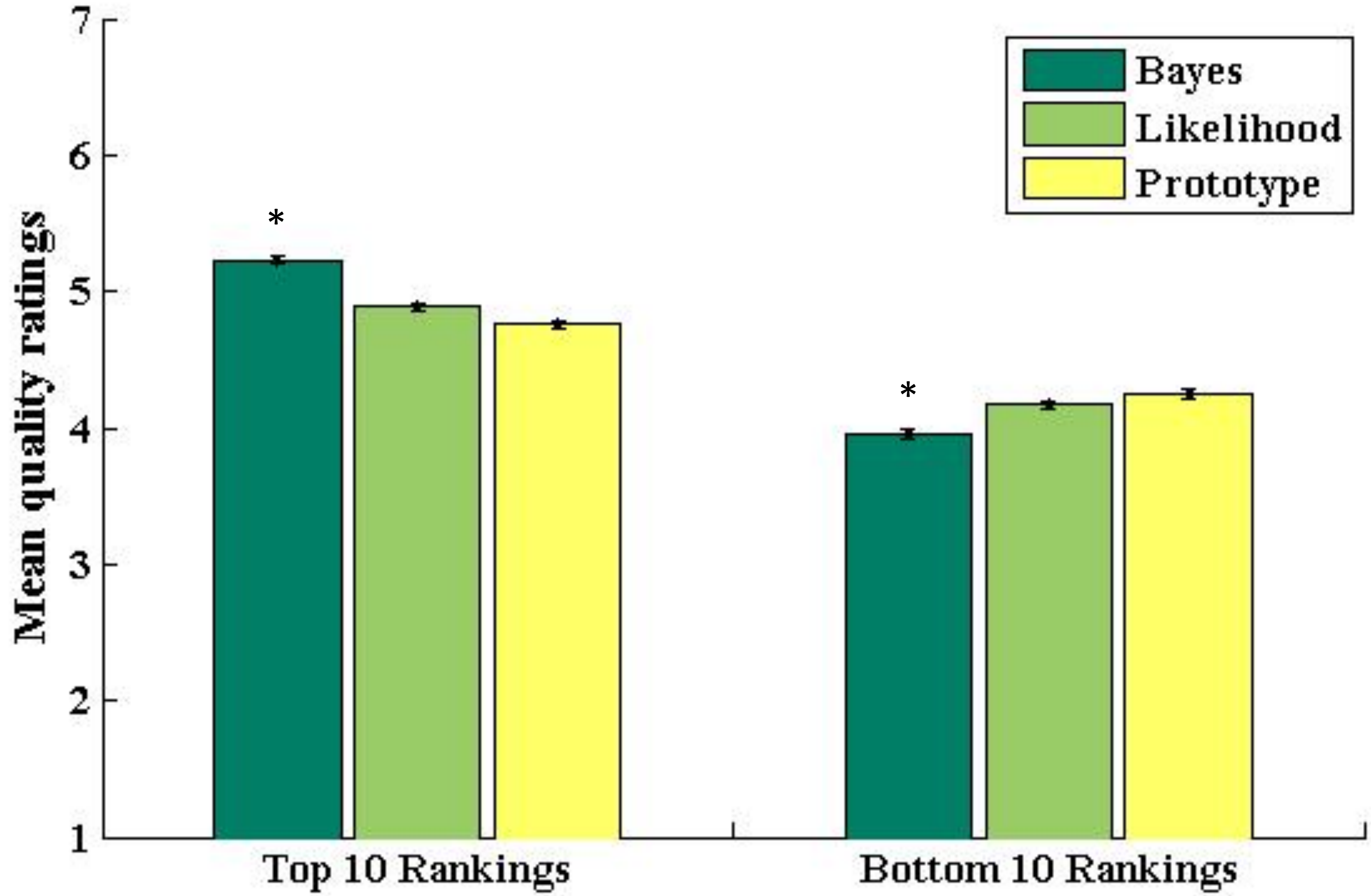
Is the image below a good example of the concept "eiffel" ?

Instructions: On a scale from **1** to **7**, please rate how good an example the image below is of the concept "eiffel", with a rating of **1** meaning the image is not a very good example and a rating of **7** meaning the image is a very good example.



1	2	3	4	5	6	7
○	○	○	○	○	○	○

Not Very Good Example　　　　Very Good Example



Bayes
Likelihood
Prototype
Mean quality ratings
7
6
5
4
3
2
1
*
*
Top 10 Rankings
Bottom 10 Rankings

Spearman rank-order correlation

how well the actual scores from the models fit with the entire set of human judgments

Bayesian model	($\rho = 0.352$)*
Likelihood model	($\rho = 0.220$)
Prototype model	($\rho = 0.160$)
Exemplar model	($\rho = 0.212$)

Summary

- Extended an existing Bayesian model of representativeness to handle sets of items
- Showed relationship to Bayesian Sets and exploited this to evaluate on a large database of naturalistic images
- Results provide strong evidence for this characterization of representativeness

Summary

Closer integration of methods from cognitive science and machine learning

- ⇨ first quantitative comparison of Bayesian Sets algorithm to human judgments
- ⇨ first evaluation of Bayesian measure of representativeness in context of a real applied problem

Questions?

CoCoSci @ Berkeley

joshua.abbott@berkeley.edu

Bayesian Representativeness:

J.B. Tenenbaum and T. L. Griffiths. The rational basis of representativeness. *Proceedings of 23rd CogSci (2001)*

Bayesian Sets:

Z. Ghahramani and K. A. Heller. Bayesian Sets. *NIPS (2005)*

K. A. Heller and Z. Ghahramani. A simple Bayesian framework for content-based image retrieval. IEEE CVPR (2006)

Extra Slides

Finding Outliers in Sets

Take an image from one category and inject it into all other categories, run algorithms and see where it ranks

Model	**Avg. Pos.**	**S.E.**
Bayesian	0.805	∓ 0.014
Likelihood	0.779	∓ 0.013
Prototype	0.734	∓ 0.015
Exemplar	0.734	∓ 0.016

“democratic”, “US President”

“current”, “world leader”

Bayesian Sets

Given a data collection *D* and a subset of items $D_s = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ representing a concept, rank an item $\mathbf{x}^* \in \{ D \backslash D_s \}$

$$Bscore(\mathbf{x}^*) = \frac{p(\mathbf{x}^*, D_s)}{p(\mathbf{x}^*)p(D_s)}$$

$$p(\mathbf{x}^*) = \int p(\mathbf{x}^* \mid \theta) p(\theta) d\theta$$

$$p(D_s) = \int \left[\prod_{n=1}^{N} p(\mathbf{x}_n \mid \theta) \right] p(\theta) d\theta$$

$$p(\mathbf{x}^*, D_s) = \int \left[\prod_{n=1}^{N} p(\mathbf{x}_n \mid \theta) \right] p(\mathbf{x}^* \mid \theta) p(\theta) d\theta$$

Bayesian Sets

Given a data collection *D* and a subset of items $D_s = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ representing a concept, rank an item $\mathbf{x}^* \in \{ D \backslash D_s \}$

$$Bscore(\mathbf{x}^*) = \frac{p(\mathbf{x}^*, D_s)}{p(\mathbf{x}^*)\, p(D_s)}$$

Assume each item $\mathbf{x}_i \in D$ is represented as a binary feature vector $\mathbf{x}_i = (x_{i1}, \ldots, x_{iJ})$ where $x_{ij} \in \{0,1\}$ and defined under a model in which each element of $\mathbf{x}_i$ has an independent Bernoulli distribution

$$p(\mathbf{x}_i \mid \theta) = \prod_j \theta_j^{x_{ij}} (1-\theta_j)^{1-x_{ij}}$$

and conjugate Beta prior

$$p(\theta \mid \alpha, \beta) = \prod_j \frac{\Gamma(\alpha_j + \beta_j)}{\Gamma(\alpha_j)\Gamma(\beta_j)} \theta_j^{\alpha_j - 1} (1-\theta_j)^{\beta_j - 1}$$

Bayesian Sets

Given a data collection D and a subset of items $D_s = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ representing a concept, rank an item $\mathbf{x}^* \in \{ D \backslash D_s \}$

$$Bscore(\mathbf{x}^*) = \frac{p(\mathbf{x}^*, D_s)}{p(\mathbf{x}^*)p(D_s)}$$

$$= \prod_j \frac{\alpha_j + \beta_j}{\alpha_j + \beta_j + N} \left(\frac{\widetilde{\alpha}_j}{\alpha_j} \right)^{x_{*j}} \left(\frac{\widetilde{\beta}_j}{\beta_j} \right)^{1 - x_{*j}}$$

where $\widetilde{\alpha}_j = \alpha_j + \sum_{n=1}^{N} x_{nj}$

$$\widetilde{\beta}_j = \beta_j + N - \sum_{n=1}^{N} x_{nj}$$

Bayesian Sets

$$\log Bscore(\mathbf{x}^*) = c + \sum_j s_j x_{*j}$$

where

$$c = \sum_j \log(\alpha_j + \beta_j) - \log(\alpha_j + \beta_j + N) + \log \tilde{\beta}_j - \log \beta_j$$

$$s_j = \log \tilde{\alpha}_j - \log \alpha_j - \log \tilde{\beta}_j + \log \beta_j$$

and x_{*j} is the j^{th} component of $\mathbf{x}^*$

Image features

Texture features (75):
We represent images using two types of texture features, 48 Gabor texture features and 27 Tamura texture features. We computed coarseness, contrast and directionality Tamura features, for each of 9 (3x3) tiles. We applied 6 scale sensitive and 4 orientation sensitive Gabor filters to each image point and compute the mean and standard deviation of the resulting distribution of filter responses.

Color features (165):
Computed HSV 3D histogram with 8 bins for H and 5 each for value and saturation. The lowest value bin was not partitioned into hues since these are hard to distinguish.

Binarization:
Each feature was binarized by computing the skewness of the distribution of that feature and giving a binary value of 1 to images falling in the 20 percentile of the heavier tail of the feature distribution.