Rational Approximations to the Rational Model of Categorization

Adam Sanborn

NIPS Workshop

Machine Learning Meets Human Learning

Collaborators

Tom Griffiths

- Dan Navarro

A Rational Model of Categorization

- Rational models predict behavior as an optimal solution to a computational problem
- Anderson developed a rational model of categorization by considering the goal of categorization, the structure of the environment, and the costs of processing

Goal of Categorization

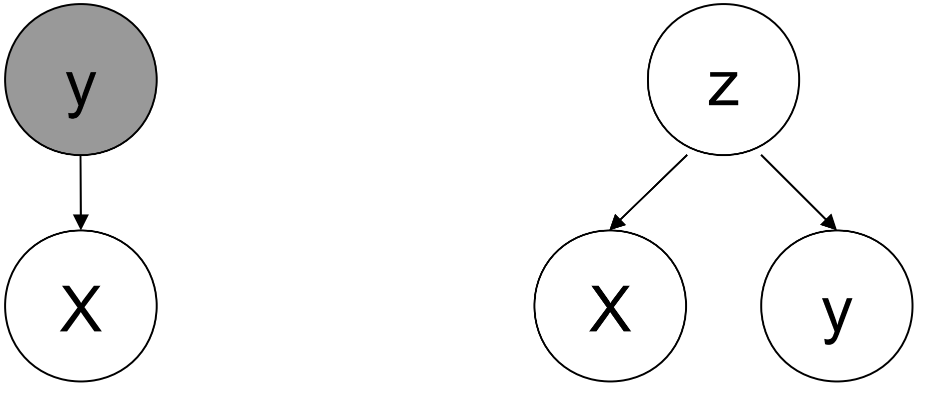
- Psychological Goals
  - Linguistic
  - Feature overlap
  - Similar function
- Computational Goal
  - Minimize prediction error

Goals of Categorization

- Prediction:

$$P(y_N = j | x_N, \mathbf{x}_{N-1}, \mathbf{y}_{N-1}) = \frac{P(x_N | y_N = j, \mathbf{x}_{N-1}, \mathbf{y}_{N-1}) P(y_N = j | \mathbf{y}_{N-1})}{\sum_{y=1}^{J} P(x_N | y_N = y, \mathbf{x}_{N-1}, \mathbf{y}_{N-1}) P(y_N = y | \mathbf{y}_{N-1})}$$

- Category learning



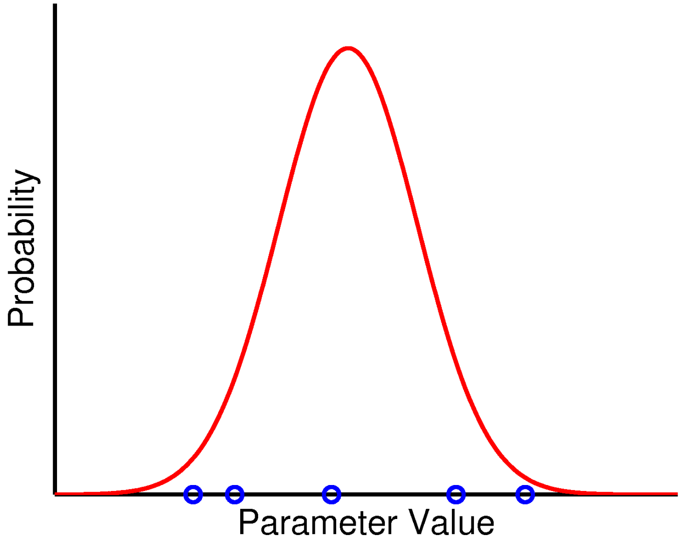
Environmental Structure

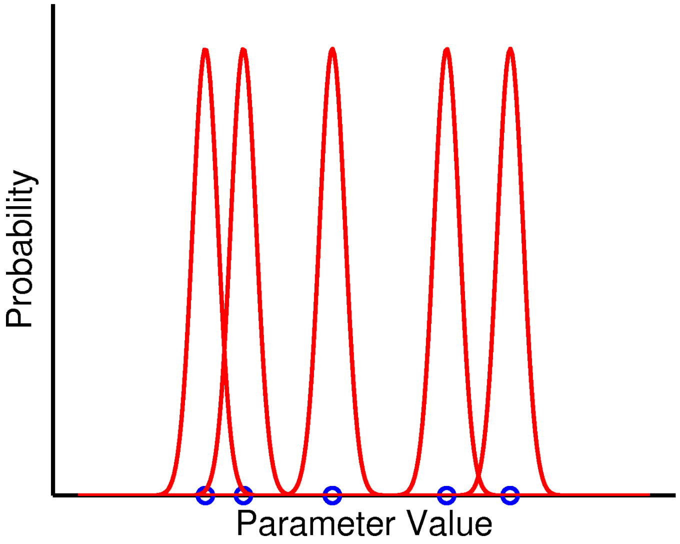
- Can view the environment as a partition of objects into clusters
  - Species

- Given a cluster, features are independent

Representation in the RMC

- Flexible representation can interpolate between prototype and exemplar models (density estimation)





Prior on the Partitions

- Each item has a particular coupling probability c, which does not depend on the number of previously seen objects

$$P(z_i = k|\mathbf{z}_{i-1}) =$$

$$\begin{cases} \frac{cM_k}{(1-c)+c(i-1)} & \text{if } M_k > 0 \text{ (i.e., } k \text{ is old)} \\ \frac{(1-c)}{(1-c)+c(i-1)} & \text{if } M_k = 0 \text{ (i.e., } k \text{ is new)} \end{cases}$$

Equivalent to the Chinese Restaurant Process

- Neal (1998) showed that $\alpha = (1-c)/c$ Anderson's prior is a CRP

$$P(z_i = k | \mathbf{z}_{i-1}) =$$

$$\begin{cases} \frac{M_k}{i-1+\alpha} & \text{if } M_k > 0 \text{ (i.e., } k \text{ is old)} \\ \frac{\alpha}{i-1+\alpha} & \text{if } M_k = 0 \text{ (i.e., } k \text{ is new)} \end{cases}$$

Computational Constraints

- Commitment to a specific hypothesis

$$P(\mathbf{x}_N, \mathbf{y}_N) = \sum_{\mathbf{z}_N} P(\mathbf{x}_N, \mathbf{y}_N | \mathbf{z}_N) P(\mathbf{z}_N)$$

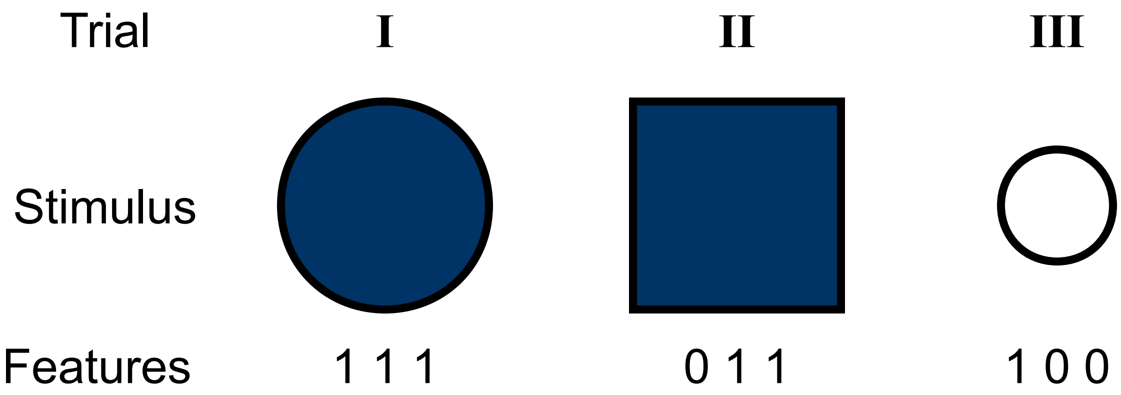
$$P(\mathbf{x}_N, \mathbf{y}_N) \approx P(\mathbf{x}_N, \mathbf{y}_N | \mathbf{z}_N)$$

- Incremental updates of the distribution

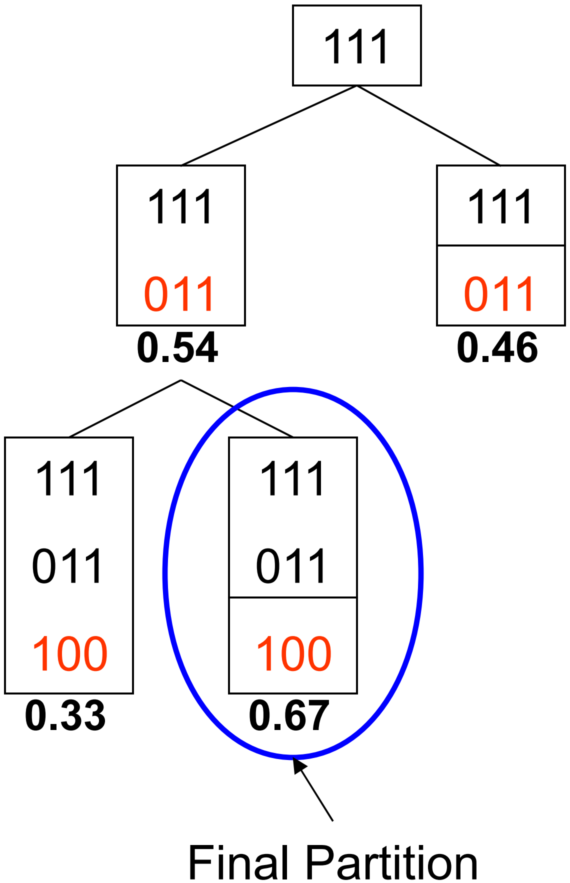
$$P(z_i = k | \mathbf{z}_{i-1}, x_i, \mathbf{x}_{i-1}, y_i, \mathbf{y}_{i-1}) \propto$$
$$P(x_i | z_i = k, \mathbf{z}_{i-1}, \mathbf{x}_{i-1}) P(y_i | z_i = k, \mathbf{z}_{i-1}, \mathbf{y}_{i-1}) P(z_i = k | \mathbf{z}_{i-1})$$

A Toy Example of Clustering

- Three stimuli are presented sequentially
- The stimuli each have values on three binary features
- The goal of the person in the task is to infer how the stimuli should be grouped

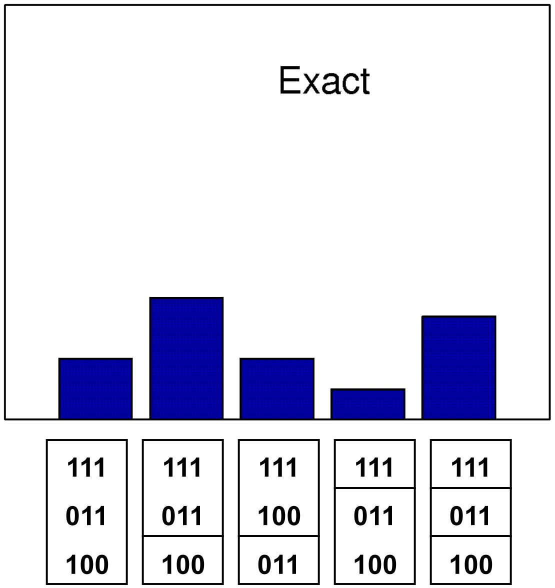


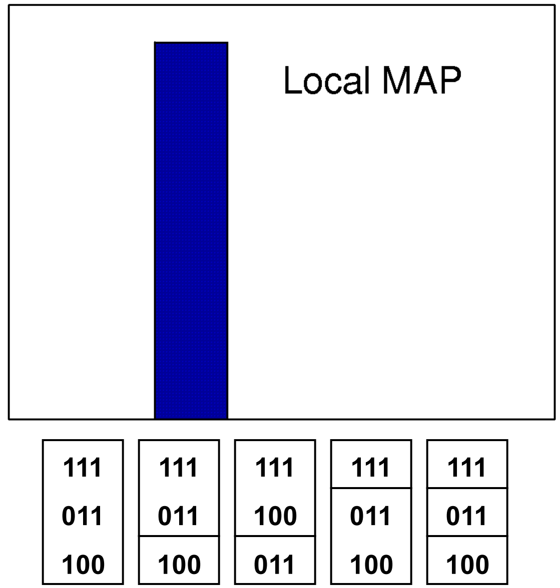
RMC Updating



- The RMC update algorithm uses assigns a stimulus to the cluster with the highest posterior probability
- We will call this algorithm the local MAP

Approximating the Posterior





- For a particular order, the Local MAP will produce a single partition

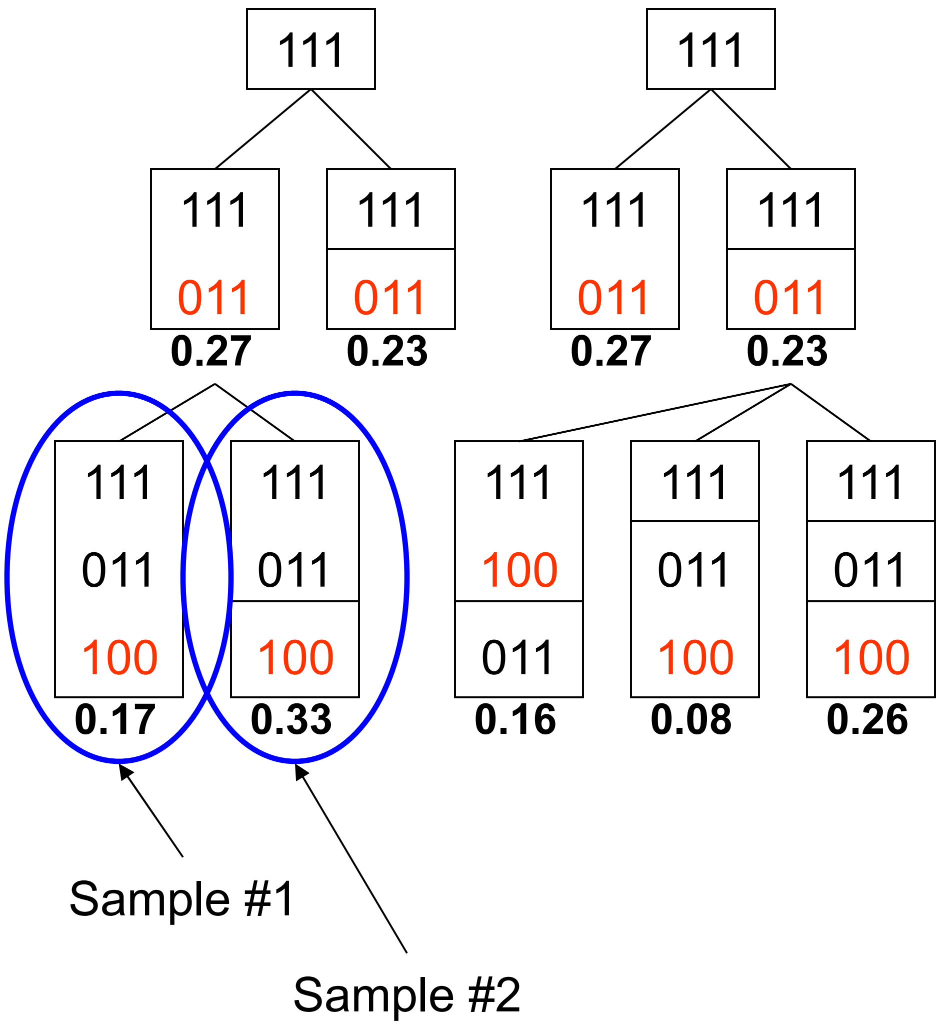
Other Approximation Methods

- There are many methods for approximating the posterior of a mixture model
  - Gibbs sampling
  - Particle filtering
- Identifying the RMC as a DPMM allows us to use these methods

Particle Filters

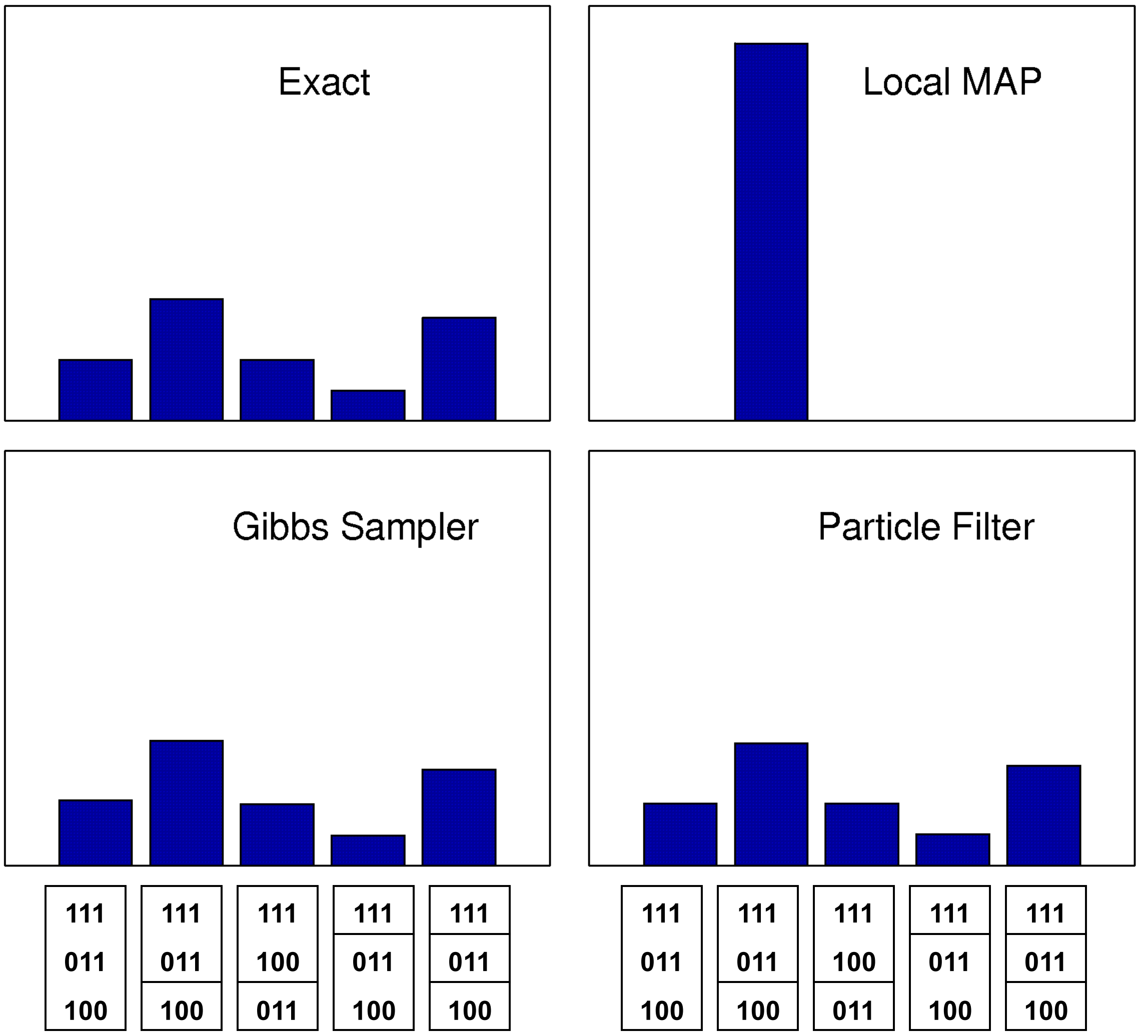
- A sequential Monte Carlo technique for using samples to provide an approximation to the posterior distribution
- Set of particles is an approximation to the posterior distribution on each trial
- Many particle filter schemes are possible

Particle Filter for the DPMM



- Assignments are made probabilistically
- A fixed number of particles are carried over from each trial

Approximating the Posterior



- For a single order, the Local MAP will produce a single partition
- The Gibbs Sampler and Particle Filter will approximate the exact DPMM distribution

Medin and Schaffer (1978) Experiment 1

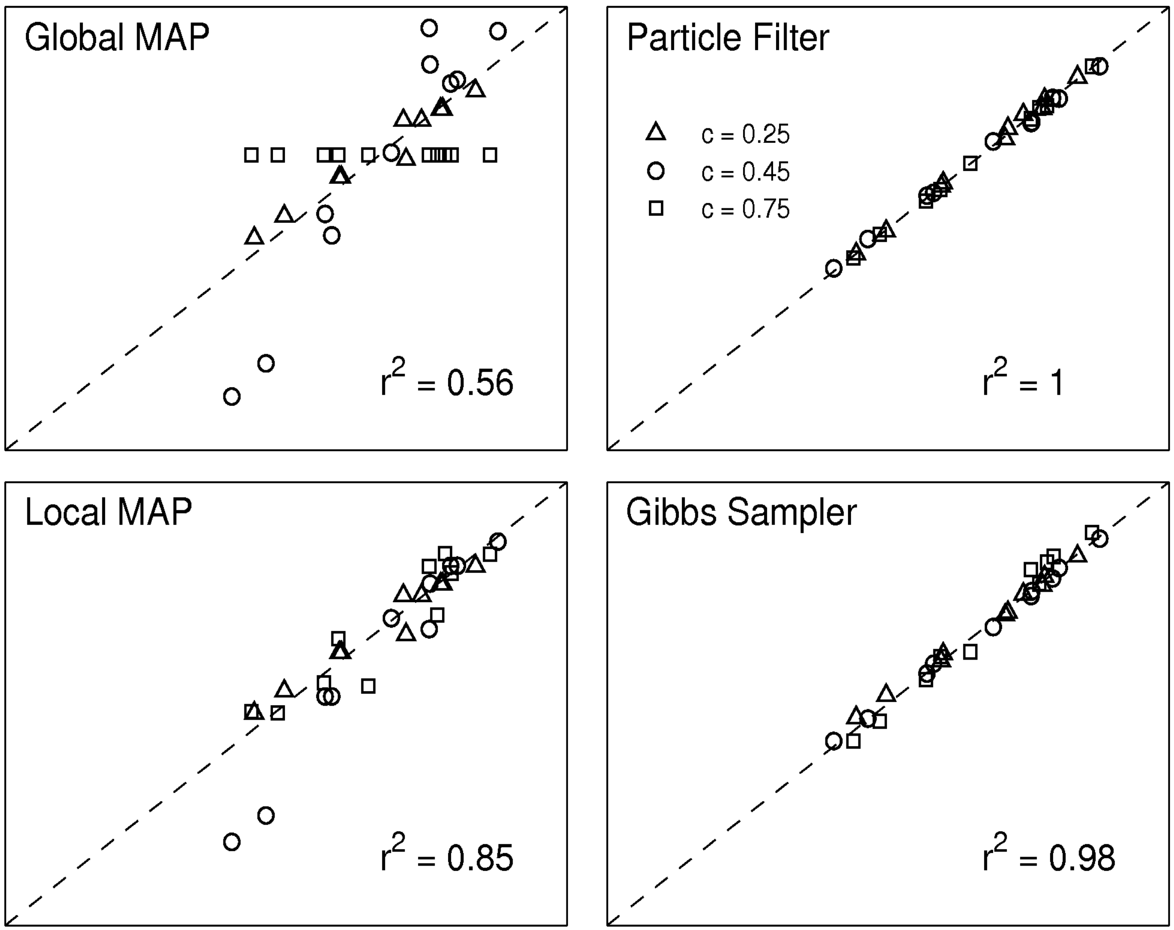
Training Stimuli

"A" Stimuli						"B" Stimuli					
Stimulus Number	Dimension Values				Rating	Stimulus Number	Dimension Values				Rating
	F	S	C	P			F	S	C	P	
6	1	1	1	1	4.8	10	0	0	0	0	5.2
7	1	0	1	0	4.6	15	1	0	1	1	4.5
9	0	1	0	1	4.8	16	0	1	0	0	4.9

New Transfer Stimuli

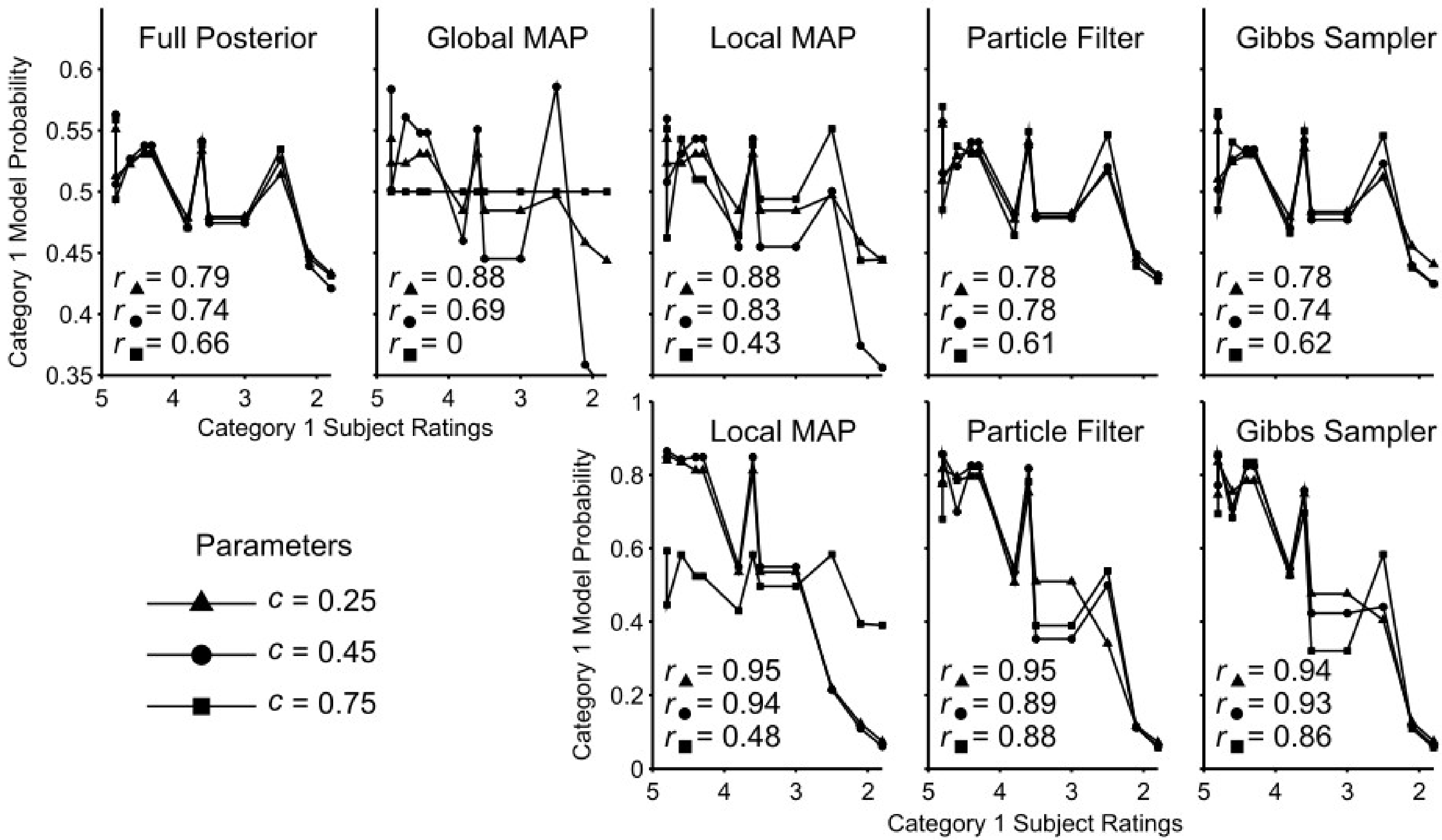
"A"-Predicted						"B"-Predicted					
Stimulus Number	Dimension Values				Rating	Stimulus Number	Dimension Values				Rating
	F	S	C	P			F	S	C	P	
5	0	1	1	1	4.3	3	1	0	0	0	3.5
13	1	1	0	1	4.4	8	0	0	1	0	4.0
4	1	1	1	0	3.6	14	0	0	0	1	3.2

Approximating the Posterior



- The Global MAP is the single partition with the highest probability
- The Local MAP was run on every possible order of the six stimuli and the results were averaged

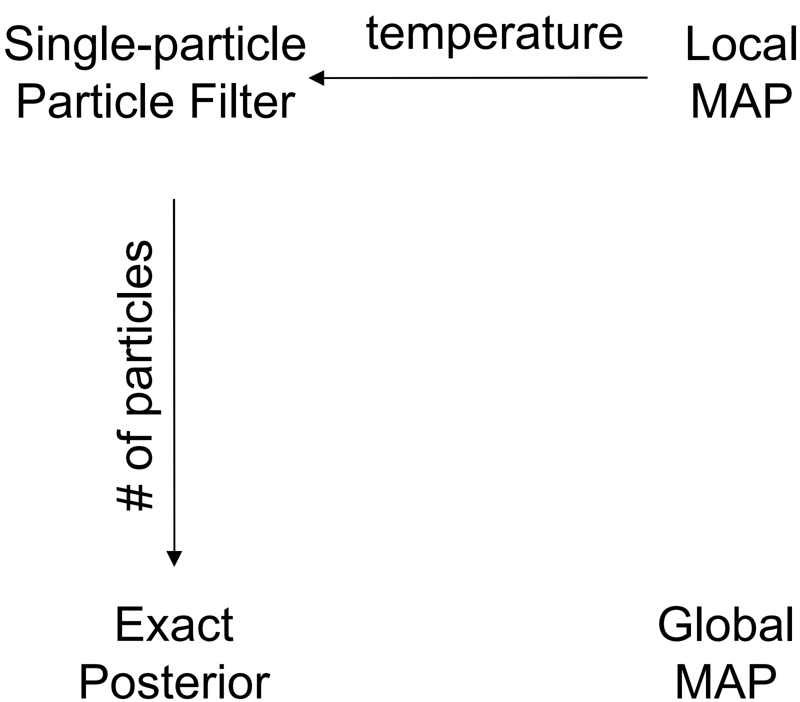
Fitting the Medin & Schaffer Results



Psychological Plausibility

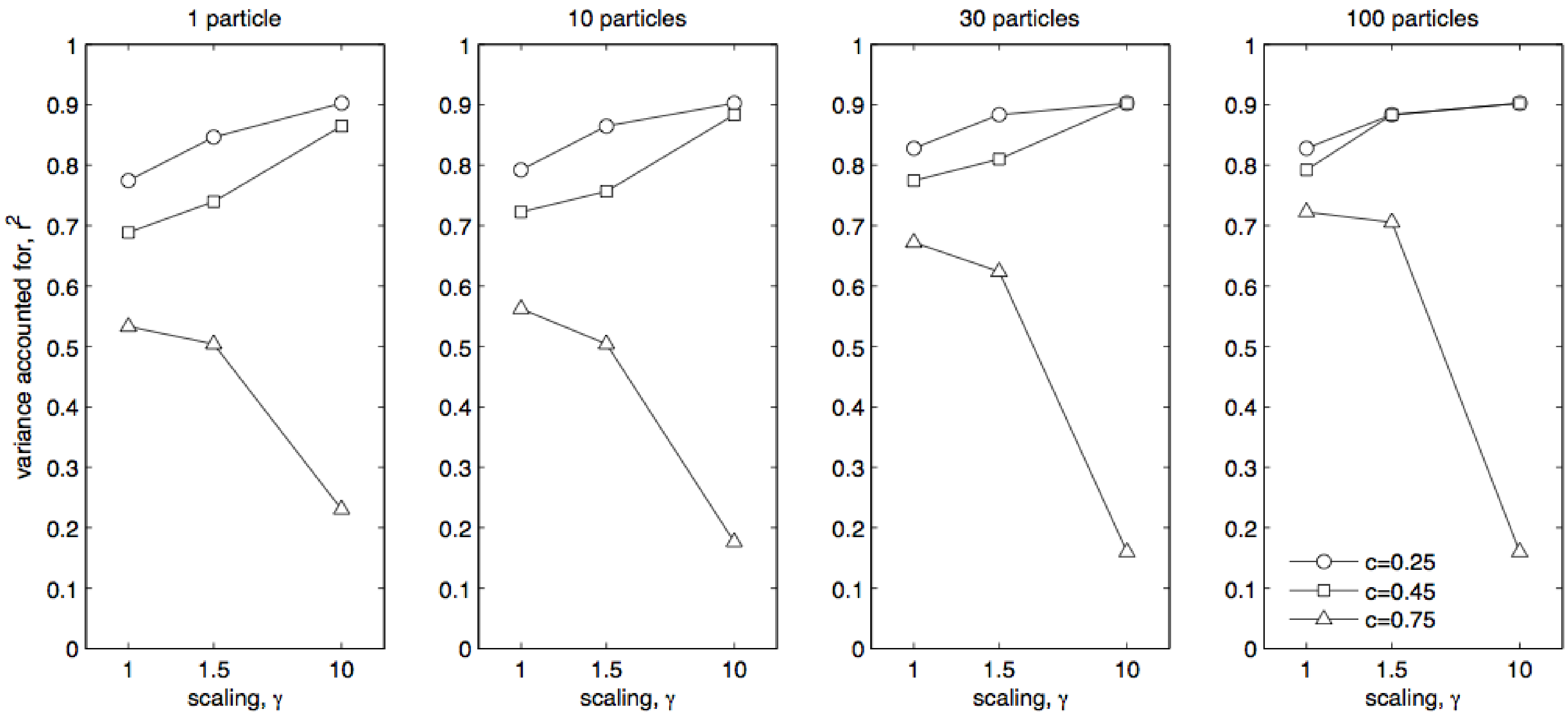
	Sequential Updating	Fixed Assignment
Local MAP	YES	YES
Gibbs Sampler	NO	NO
Particle Filter (many particles)	YES	NO
Particle Filter (1 particle)	YES	YES

Relationships Between the Local MAP and Particle Filter



- Lowering the temperature of the posterior distribution (by raising each probability to a constant power and normalizing) produces Local MAP behavior from a particle filter

Results for a Range of Particles and Temperatures



Order Effects in Human Data

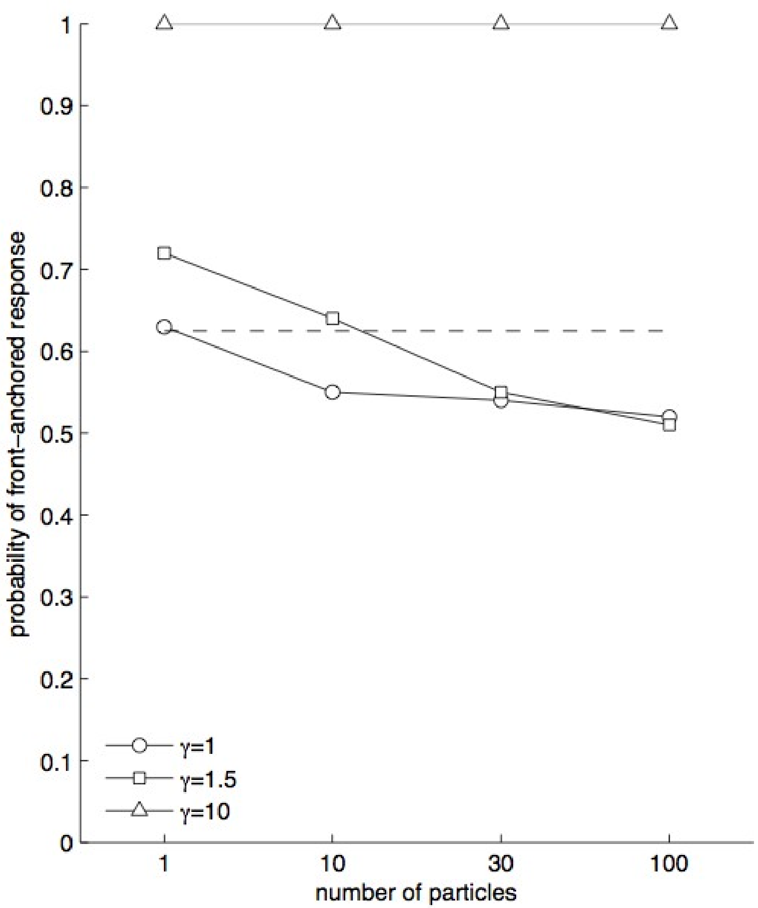
- The DPMM does not produce any order effects
- Human data shows order effects (Medin & Bettger, 1994)
- Anderson and Matessa tested the order effects in a categorization experiment with unsupervised learning

Anderson and Matessa Order Experiment

Front-Anchored Order	End-Anchored Order
scadsporm	snadstirb
scadstirm	snekstirb
sneksporb	scadsporm
snekstirb	sceksporm
sneksporm	sneksporm
snekstirm	snadsporm
scadsporb	scedstirb
scadstirb	scadstirb

- Subjects were shown all sixteen stimuli that had four binary features
- Front-anchored ordered stimuli emphasized the first two features in the first eight trials; end-anchored ordered emphasized the last two

Anderson & Matessa Results



Subjects divided stimuli into two equal groups

Order bias in the Local MAP is very strong

A single-particle particle filter is weakly biased

A Rational Process Model

- Connections between machine learning and human learning can be used to identify other psychologically plausible algorithms
- The particle filter is a useful approximation that can be tested as general-purpose psychological heuristic
- Using a psychologically plausible approximation can change a rational model into a rational process model