Outline for lectures

- Introduction
- Cognition as probabilistic inference
- Learning concepts from examples (continued)
- Learning and using intuitive theories (more structured systems of knowledge)
Learning from just one or a few examples, and mostly unlabeled examples ("semi-supervised learning").
Simple model of concept learning

“This is a blicket.”

“Can you show me the other blickets?”
Learning to learn: what object features count for word learning?

• 24-month-olds show the shape bias with simple novel objects. 20-month-olds do not. (Landau, Smith, Jones 1988)

• Smith et al (2002) trained 17-month-olds on labels for 4 artificial categories:

• After 8 weeks of training (20 min/week), 19-month-olds show the shape bias.
Transfer to real-world vocabulary

The intuition: Learn that shape varies across categories but is relatively constant within nameable categories.

The puzzle: The shape bias is a powerful inductive constraint, yet can be learned from very little data.
Learning about feature variability
(Kemp, Perfors & Tenenbaum, Dev. Science 2007)
Learning about feature variability

(Kemp, Perfors & Tenenbaum, *Dev. Science* 2007)
A hierarchical Bayesian model

Level 3:
Prior expectations on bags in general

Level 2:
Bags in general

Level 1:
Bag proportions

Data

Simultaneously infer $p(\theta^i, \alpha, \beta \mid y^i, \lambda)$

$\alpha \sim \text{Exponential}(\lambda)$
$\beta \sim \text{Dirichlet}(1)$
$\theta^i \sim \text{Dirichlet}(\alpha \beta)$
$y^i \sim \text{Multinomial}(\theta^i)$
A hierarchical Bayesian model

Level 3:
Prior expectations on bags in general

Level 2:
Bags in general

Level 1:
Bag proportions

Data

\[ \lambda \]
\[ \alpha = 0.1 \] (within-bag variability)
\[ \beta \] (overall population distribution)

\[ \alpha \sim \text{Exponential}(\lambda) \]
\[ \beta \sim \text{Dirichlet}(1) \]
\[ \theta^i \sim \text{Dirichlet}(\alpha \beta) \]
\[ y^i \sim \text{Multinomial}(\theta^i) \]
A hierarchical Bayesian model

Level 3:
Prior expectations on bags in general

Level 2:
Bags in general

Level 1:
Bag proportions

Data

\[ \alpha \sim \text{Exponential}(\lambda) \]
\[ \beta \sim \text{Dirichlet}(1) \]
\[ \theta^i \sim \text{Dirichlet}(\alpha \beta) \]
\[ y^i \mid n^i \sim \text{Multinomial}(\theta^i) \]

\[ \alpha = 5 \] (within-bag variability)

\[ \beta \] (overall population distribution)
Learning the shape bias

(Kemp, Perfors & Tenenbaum, *Dev. Science* 2007)

Assuming independent Dirichlet-multinomial models for each dimension...

...we learn that:

- Shape varies across categories but not within categories.
- Texture, color, size vary across and within categories.
Second-order generalization test
(Kemp, Perfors & Tenenbaum, Dev. Science 2007)

This is a dax.
Show me the dax…

<table>
<thead>
<tr>
<th>Training</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
<td>1 1 2 2 3 3 4 4</td>
</tr>
<tr>
<td>Shape</td>
<td>1 1 2 2 3 3 4 4</td>
</tr>
<tr>
<td>Texture</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>Color</td>
<td>1 2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>Size</td>
<td>1 2 1 2 1 2 1 2</td>
</tr>
</tbody>
</table>

“blessing of abstraction”

Probability (normalized) that choice object belongs to the same category as the test exemplar.

1st order gen
2nd order gen
A more realistic model

Prior expectations on categories in general

Categories in general

Individual categories

Data

\[ z \sim \text{CRP}(\gamma) \]
\[ \alpha \sim \text{Exponential}(\lambda) \]
\[ \beta \sim \text{Dirichlet}(1) \]
\[ \theta^k \sim \text{Dirichlet}(\alpha \beta) \]
\[ y^i \sim \text{Multinomial}(\theta^{z_i}) \]

(Perfors & Tenenbaum, *Proc Cog Sci 2009*)
A more realistic model

Prior expectations on categories in general

Categories in general

Individual categories

Data

(Pfors & Tenenbaum, *Proc Cog Sci 2009*)
A more realistic model

Prior expectations on categories in general

Categories in general

Individual categories

Data

(Perfors & Tenenbaum, *Proc Cog Sci 2009*)

\[
\begin{aligned}
\lambda &
\downarrow \\
\alpha, \beta &
\downarrow \\
\theta^1 &
\downarrow \\
y^1 & z \\
42571507 - 1 & 23648160 - 2 \\
11577707 - 1 & 73046446 - 2 \\
41670016 - 1 & 78640370 - 2 \\
41502465 - 1 & 73616235 - 2 \\
\end{aligned}
\]

\[
\begin{aligned}
\lambda &
\downarrow \\
\alpha, \beta &
\downarrow \\
\theta^2 &
\downarrow \\
y^2 & z \\
56643025 - 3 & 30746502 - 4 \\
73014667 - 3 & 31242541 - 4 \\
78640370 - 2 & 56315442 - 3 \\
73616235 - 2 & 16616311 - 3 \\
\end{aligned}
\]

\[
\begin{aligned}
\lambda &
\downarrow \\
\alpha, \beta &
\downarrow \\
\theta^3 &
\downarrow \\
y^3 & z \\
56315442 - 3 & 30746502 - 4 \\
56643025 - 3 & 31242541 - 4 \\
73616235 - 2 & 16616311 - 3 \\
\end{aligned}
\]

\[
\begin{aligned}
\lambda &
\downarrow \\
\alpha, \beta &
\downarrow \\
\theta^4 &
\downarrow \\
y^4 & z \\
30746502 - 4 & 88998899 - 5 \\
31242541 - 4 & 88998899 - 5 \\
30252135 - 4 & 88998899 - 5 \\
60240453 - 4 & 88998899 - 5 \\
\end{aligned}
\]

\[
\begin{aligned}
\lambda &
\downarrow \\
\alpha, \beta &
\downarrow \\
\theta^5 &
\downarrow \\
y^5 & z \\
\end{aligned}
\]

2\textsuperscript{nd} order generalization:

88994271 - 5? or 42718899 - 5?
A more realistic model

Prior expectations on categories in general

Categories in general

Individual categories

Learning the base distribution of a DP mixture

\[
\lambda \\
\alpha, \beta
\]

\[
\theta^1 \quad \theta^2 \quad \theta^3 \quad \theta^4 \quad \theta^5
\]

Data

\[
\begin{align*}
\text{y} \quad \text{z} & \quad \text{y} \quad \text{z} & \quad \text{y} \quad \text{z} & \quad \text{y} \quad \text{z} & \quad \text{y} \quad \text{z} \\
42571507 - 1 & \quad 23648160 - 2 & \quad 56643025 - 3 & \quad 30746502 - 4 & \quad 88998899 - 5 \\
11577707 - 1 & \quad 73046446 - 2 & \quad 50614667 - 3 & \quad 31242541 - 4 & \\
41670016 - 1 & \quad 78640370 - 2 & \quad 56315442 - 3 & \quad 30252135 - 4 & \\
41502465 - 1 & \quad 73616235 - 2 & \quad 16616311 - 3 & \quad 60240453 - 4 & \\
\end{align*}
\]

2nd order generalization:

88994271 - 5? or 42718899 - 5?

(Perfors & Tenenbaum, Proc Cog Sci 2009)
Model vs. People

Towards more natural concepts
Towards more natural concepts

CRP mixture:
How many different ways to structure a domain?
(Shafto, Kemp, Mansingka, Tenenbaum, 2006; submitted)

“CrossCat”: nonparametric clustering over features, with a different clustering of objects for each feature-cluster.

Leopard  Python
Sheep  Alligator
Seal  Octopus
Dolphin  Penguin
Monkey  Finch
Bat  Eagle
Alligator  Owl
Iguana  Seagull
Frog  Owl
Pythons  Grasshopper
Finch  Bat
Ostrich  Eagle
Ostrich  Dragonfly
Penguin  Sheep

Leopard  Python  Finches
Sheep  Alligator  Owl
Seal  Octopus  Seagull
Dolphin  Penguin  Owl
Monkey  Finch  Finches
Bat  Eagle  Dragonfly
Alligator  Sheep  Sheep
Iguana  Sea Pigeon  Owls
Frog  Grasshopper  Bat
Pythons  Ostrich  Ostrich

Lives in lakes is an amphibian is a rodent is tall is slimy has horns has hooves has a tail has webbed feet eats nuts is smooth lives in trees is large lives in cold climates
Is dangerous is a carnivore is a predator lives in water flies is long eats leaves eats animals lives in hot climates
How many different ways to structure a domain?
(Shafto, Kemp, Mansingka, Tenenbaum, 2006; submitted)

“CrossCat”: nonparametric clustering over features, with a different clustering of objects for each feature-cluster.

Evidence for CrossCat-like learning in humans:
- Sorting natural categories
- Sorting artificial categories
- Predicting values for novel features and novel objects
Learning relational concepts

CONCEPTS
Professors
Graduate students
Undergraduates

RELATIONSHIPS
Professors give advice to Grad students and Undergrads.
Grad students give advice to Undergrads.
Undergrads give advice to no one.
Learning relational concepts

CONCEPTS
Magnets
Magnetic objects
Non-magnetic objects

RELATIONSHIPS
Magnets interact with each other.
Magnets and Magnetic objects interact.
Magnetic objects do not interact with each other.
Non-magnetic objects do not interact with anything.
Learning relational concepts

person 1 gives advice to person 9
Learning relational concepts

gives advice to

people

Prof  Grads  Ug
1  4  8  2  6  9  5  7  3

people

1  2  3  4  5  6  7  8  9
Infinite Relational Model (IRM)  
(Kemp, Griffiths, Tenenbaum, Yamada, & Ueda, 2006)

\[ z \mid \gamma \sim \text{CRP}(\gamma) \]

\[ \eta_{ab} \mid \alpha, \beta \sim \text{Beta}(\alpha, \beta) \]

\[ R_{ij} \mid z, \eta \sim \text{Bernoulli}(\eta_{z_iz_j}) \]

\[ p(z, \eta \mid R) \propto P(R \mid z, \eta) p(\eta \mid z) P(z) \]
Learning algorithm

- Continuous parameters (weights/probabilities) integrated out analytically.
- Gibbs sampling + split-merge moves:
The causal blocks world
(Tenenbaum and Niyogi, 2003)
The causal blocks world
(Tenenbaum and Niyogi, 2003)
Stage (# of objects / 3)

Likelihood of Lighting Up

Pre
Post Same Group
Post Different Group

Model predictions

Learning curves

Training

Test

# of objects observed

Probability of lighting up

Model predictions

Pre
Post Same Group
Post Different Group
Constructing semantic networks

(Collins & Quillian, 1969)
Upper level medical ontology

Biomedical predicate data used to construct ontology in UMLS (McCrae et al.):

- 134 concepts: enzyme, hormone, organ, disease, cell function ...
- 49 predicates: affects(hormone, organ), complicates(enzyme, cell function), treats(drug, disease), diagnoses(procedure, disease) ...
Learning semantic networks with IRM

Biomedical predicate data from UMLS (McCrae et al.):

- 134 concepts: enzyme, hormone, organ, disease, cell function ...
- 49 predicates: affects(hormone, organ), complicates(enzyme, cell function), treats(drug, disease), diagnoses(procedure, disease) ...
Learning semantic networks with IRM

e.g., Diseases affect Organisms
Chemicals interact with Chemicals
Chemicals cause Diseases
Learning semantic networks with IRM

Relations between clusters:

- Diseases affect Organisms
- Chemicals interact with Chemicals
- Chemicals cause Diseases
Extracting semantic networks from text via relational clustering (Kok & Domingos 2008)

Tested several algorithms for relational clustering on TextRunner data:

~ 2 million triples of the form $R(x, y)$: e.g., upheld(Court, ruling), named_after(Jupiter, Roman_god).

~ 10,214 $R$ symbols, 8942 $x$ symbols, 7995 $y$ symbols (each appears >25 times).
Annotated hierarchies model
(Roy, Kemp, Mansinghka & Tenenbaum, 2007)
Annotated hierarchies model

170 neurons in Macaque Inferior Temporal (IT) cortex
78 grey-scale images

Hung, Kreiman, Poggio, and DiCarlo (Science 2005)
The Mondrian Process
(Roy & Teh, 2008; in prep)

We can also construct an exchangeable variant of the Annotated Hierarchies model (a hierarchical block model) by moving from the unit square to a product of random trees drawn from Kingman’s coalescent prior (Kingman, 1982a). Let $\mu_d$ be Lebesgue measure.

\[
\begin{align*}
T_d & \sim KC(\lambda), \forall d \in \{1, \ldots, D\} \quad \text{for each dimension, sample a tree} \\
M \mid T & \sim MP(2\alpha, T_1, \ldots, T_D) \quad \text{partition the cross product of trees} \\
\phi_S \mid M & \sim Beta(\alpha_0, \alpha_1), \forall S \in M. \quad \text{each block } S \text{ gets a probability } \phi_S
\end{align*}
\]

Let $S_{ij}$ be the subset $S \in M$ where leaves $(i, j)$ fall in $S$. Then

\[
R_{ij} \mid \phi, M \sim Bernoulli(\phi_{S_{ij}}), i, j \in \{1, \ldots, n\}. \quad R_{ij} \text{ is true w.p. } \phi_{S_{ij}}
\]

**Algorithm 1** Conditional Mondrian $m \sim MP(\lambda, \Theta_1, \ldots, \Theta_D \mid \rho)$

1. let $\lambda' \leftarrow \lambda - E$ where $E \sim Exp(\sum_{d=1}^{D} \mu_d(\Theta_d \setminus \Phi_d))$.
2. if $\rho$ has no cuts then $\lambda'' \leftarrow 0$ else $\langle d', x', \lambda'', \rho_>, \rho_< \rangle \leftarrow \rho$.
3. if $\lambda' < \lambda''$ then take root form of $\rho$
4. if $\rho$ has no cuts then
5. return $m \leftarrow \Theta_1 \times \ldots \times \Theta_D$.
6. else $(d', x')$ is the first cut in $m$
7. return $m \leftarrow \langle d', x', \lambda', MP(\lambda'', \Theta_1, \ldots, \Theta^{<d'}_d, \ldots, \Theta_D \mid \rho_<), \Theta^{<d'}_d, \ldots, \Theta_D \mid \rho_>) \rangle$.
8. else $\lambda'' < \lambda'$ and there is a cut in $m$ above $\rho$
9. draw a cut $(d, x)$ outside $\rho$, i.e., $p(d) \propto \mu_d(\Theta_d \setminus \Phi_d), x|d \sim \frac{\mu_d(\Theta_d \setminus \Phi_d)}{\mu_d(\Theta_d \setminus \Phi_d)}$
   without loss of generality suppose $\Phi_d \subset \Theta^{<d}_d$
10. return $m \leftarrow \langle d, x, \lambda', MP(\lambda', \Theta_1, \ldots, \Theta^{<d}_d, \ldots, \Theta_D \mid \rho), \\
    \Theta^{<d}_d, \ldots, \Theta_D \rangle$.

\[
\mu_d(\Theta_d \setminus \Phi_d)
\]
The Mondrian Process

\[ \text{MP}(\lambda, \Theta_1, \ldots, \Theta_{d^x}^x, \ldots, \Theta_D) \]
Learning from just one or a few examples, and mostly unlabeled examples (“semi-supervised learning”).
Learning words for objects

“tufa”

(Collins & Quillian, 1969)

(IT population responses
Hung et al., 2005; c.f. Kiani et al. 2007)
Learning words for objects

Bayesian inference over tree-structured hypothesis space:

(Xu & Tenenbaum, *Psych. Review* 2007; Schmidt & Tenenbaum, in prep)
Learning words for objects

Bayesian inference over tree-structured hypothesis space:

(Xu & Tenenbaum, *Psych. Review* 2007; Schmidt & Tenenbaum, in prep)
Hierarchical Bayesian framework

\[ F: \text{form} \]
\[ \downarrow \]
\[ S: \text{structure} \]
\[ \downarrow \]
\[ D: \text{data} \]
Learning to learn: what is the right form of structure for the domain?

\[ F: \text{form} \]
\[ S: \text{structure} \]
\[ D: \text{data} \]
The value of structural form knowledge: inductive constraints (bias)

Mystery city …
- average annual temperature: 66 F / 19 C
- voted 60% for George W Bush in 2004
- popular foods are fried and BBQ
Property induction

Given that \( \{X_1, \ldots, X_n\} \) have property P, how likely is it that Y does?
Given that \( \{X_1, \ldots, X_n\} \) have property P, how likely is it that Y does?

Property induction

(Kemp & Tenenbaum, *Psych. Review* 2009)

\[
\text{Property } \sim N\left(0, \left(\Delta + \frac{1}{\sigma^2} I\right)^{-1}\right)
\]

(Zhu, Lafferty & Ghahramani, 2003)

\[
N\left(0, \exp\left(-\frac{1}{\sigma} \|x_i - x_j\|\right)\right)
\]

(c.f. Lawrence, 2004)
Given that \( \{X_1, \ldots, X_n\} \) have property P, how likely is it that Y does?

Property \( \sim N\left(0, (\Delta + \frac{1}{\sigma^2} I)^{-1}\right) \)
(Zhu, Lafferty & Ghahramani, 2003)
Learning structural forms

People can discover structural forms…

– Scientists
  
  Linnaeus
  
  Kingdom Animalia
  Phylum Chordata
  Class Mammalia
  Order Primates
  Family Hominidae
  Genus Homo
  Species Homo sapiens

– Children

  e.g., hierarchical structure of category labels, cyclical structure of seasons or days of the week, clique structure of social networks.

… but standard learning algorithms assume fixed forms.

– Principal components analysis: low-dimensional spatial structure
– Hierarchical clustering: tree structure
– $k$-means clustering, mixture models: flat partition.
Hypothesis space of structural forms
(Kemp & Tenenbaum, PNAS 2008)
A hierarchical Bayesian approach
(Kemp & Tenenbaum, PNAS 2008)

\[ P(F) \]

**F:** form

\[ P(S | F) \]

Simplicity

**S:** structure

\[ P(D | S) \]

Smoothness (Fit to data)

**D:** data

\[
P(S, F | D) \propto P(D | S) P(S | F) P(F)
\]
Development of structural forms as more data are observed

“blessing of abstraction”
Causal learning and reasoning

Causal model

Event data

(Griffiths & Tenenbaum; Kemp, Goodman, Tenenbaum)
Causal learning and reasoning

Causal schema

- Behaviors:
  - high-fat diet
  - working in factory
  - ...

- Diseases:
  - heart disease
  - lung cancer
  - ...

- Symptoms:
  - coughing
  - chest pain
  - ...

Cut down hypothesis space from size 521,939,651,343,829, 405,020,504,063 to 131,072

(Griffiths & Tenenbaum; Kemp, Goodman, Tenenbaum)
Causal schema

Causal model

Event data

Infinite relational model

Belief net with Dirichlet-multinomial parameterization

(Mansinghka, Kemp, Tenenbaum, Griffiths, UAI 2006)
Ground-truth causal network

Causal model
\downarrow
Event data

Causal model
\downarrow
Causal schema
\downarrow
Causal model
\downarrow
Event data

"blessing of abstraction"

(Mansinghka, Kemp, Tenenbaum, Griffiths, UAI 2006)
Conclusions

How does the mind get so much from so little, in learning about objects, classes, causes, scenes, sentences, thoughts, social systems?

A toolkit for studying the nature, use and acquisition of abstract knowledge:

- Bayesian inference in probabilistic generative models.
- Probabilistic models defined over a range of structured representations: spaces, graphs, grammars, predicate logic, schemas, programs.
- Hierarchical models, with inference at multiple levels of abstraction.
- Nonparametric models, adapting their complexity to the data and balancing constraint with flexibility.

An alternative to classic “either-or” dichotomies: Nature versus Nurture, Logic (Structure, Rules, Symbols) versus Probability (Statistics).

- How can domain-general mechanisms of learning and representation build domain-specific abstract knowledge?
- How can structured symbolic knowledge be acquired by statistical learning?

A different way to think about the development of a cognitive system.

- Powerful abstractions can be learned surprisingly quickly, together with or prior to learning the more concrete knowledge they constrain.
- Structured representations need not be rigid, static, hand-wired, brittle. Embedded in a probabilistic framework, they can grow dynamically and robustly in response to the sparse, noisy data of experience.
Open directions and challenges

• More precise relation to psychology
  – How does human cognitive processing perform approximate probabilistic inference (i.e., approximately implement rational methods of approximate inference)?

• Relation to the brain
  – How to implement structured probabilistic models in neural architectures?

• Probabilistic models for richly structured knowledge
  – How to formalize an intuitive theory of physics or psychology?

• Effective learning of structured probabilistic models
  – How to balance expressiveness/learnability tradeoff?
Goal-directed action
(production and comprehension)

(Wolpert et al., 2003)
Goal inference as inverse probabilistic planning

(Baker, Tenenbaum & Saxe, Cognition, in press)

Gergely, Csibra et al.:
Inferring social goals


Hamlin, Kuhlmeier, Wynn & Bloom:

---

Model prediction
Subject ratings
Model prediction
Subject ratings
Model prediction
Subject ratings
Model prediction
Subject ratings
Model prediction
Subject ratings
The really big questions

Where does it all end?

Across different domains and tasks, and different levels of abstraction, our probabilistic models are starting to look increasingly complex and to differ in almost arbitrarily many ways. This seems unsatisfying…

The brain appears to have a uniform circuitry. Other cognitive modeling paradigms adopt a single unifying representational primitive (production rules, predicate logic, synaptic strengths, tensors). Is there a single universal Bayesian primitive?

How does it all begin?

Can all these different kinds of representations be learned? What is the ultimate hypothesis space of innate primitives – or is it simply “turtles all the way up”? Could a universal hypothesis space for all probabilistic models possibly be searched effectively?

(C.f. Kolmogorov complexity theory, Chater & Vitanyi)
Church: a universal probabilistic language
(Goodman et al., UAI 2008)
Church: a universal probabilistic language
(Goodman et al., UAI 2008)

(define cause (mem (lambda (a b) (flip 0.5)))
(define spontaneous (mem (lambda (a t) (flip 0.01)))
(define do (mem (lambda (a t) (uniform-draw (pair '() (values a)))))
(define (parents a) (filter (lambda (y) (cause y a)) variables))

;; a noisy-or version:
(define strength (mem (lambda (a b)
    (if (cause a b)
      (beta 1 1); or some other prior on strengths
      0.0))))

(define (occurs a t)
  (or (spontaneous a t)
    (do a t)
      (fold (lambda (x y) (noisy-or (occurs x t) (strength x a) y 1.0))
        false
        (parents a)))))

Causal networks
Church: a universal probabilistic language
(Goodman et al., UAI 2008)

```
(define cause (mem (lambda (a b) (flip 0.5)))
(define spontaneous (mem (lambda (a t) (flip 0.01)))
(define do (mem (lambda (a t) (uniform-draw (pair '() (values a)))))
(define (parents a)
  ;; a noisy-or version
  (define class (mem (lambda (obj) (drawclass))))
  (define drawclass (DPmem 1.0 gensym))
  (define (irm-mean
               (mem (lambda (obj-class1 obj-class2)
                     (normal 0.0 10.0)))))
  (define (irm-value
               (mem (lambda (obj1 obj2)
                     (normal (irm-mean (class obj1) (class obj2))
                             1.0))))
  (parents a)
```

Causal networks
Relational schema
Church: a universal probabilistic language
(Goodman et al., UAI 2008)

(define cause (mem (lambda (a b) (flip 0.5)))))
(define spontaneous (mem (lambda (a t) (flip 0.01)))))
(define do (mem (lambda (a t) (uniform-draw (pair '() (values a)))))))
(define (parents a)
  (define drawclass (DPmem 1.0 gensym))
  ;; a noisy-or version
  (define class (mem (lambda (obj) (drawclass))))
  ;; physical world
  (define objects (repeat (poisson 1.0) gensym))
  (define depth (mem (lambda (object time) (depth object (- time 1))))))
  (define location (mem (lambda (object time)
    (+ (drift) (location object (- time 1))))))
  (define (drift) (uniform-draw (list 0 1 -1)))
  (define extent (mem (lambda (object) (uniform-draw (list 1 2 3))))))
  (define (object-seen location time)
    (argmin depth
      (map (lambda (o) (intersects o location time)) objects)))
  (define (view location time) (object-properties (object-seen location time))))

Causal networks
Relational schema
Physical objects
Church: a universal probabilistic language
(Goodman et al., UAI 2008)

(define cause (mem (lambda (a b) (flip 0.5)) ))
(define spontaneous (mem (lambda (a t) (flip 0.01)) ))
(define do (mem (lambda (a t) (uniform-draw (pair '() (values a)))))
(define (parents a)
  ;;a noisy-or version
  (define drawclass (DPmem 1.0 gensym))
  (define class (mem (lambda (obj) (drawclass))))

(define objects (repeat (poisson 1.0) gensym))
(define depth (mem (lambda (object time) (depth object (- time 1)))))
(define location (mem (lambda (object time) (+ (drift location) depth))
  (define (drift) (uniform-draw (list (- 1.0 depth) depth))
  (define extent (mem (lambda (object)))
  (define (object-seen location time)
    (argmin depth
      (map (lambda (o)
        (define view location time) (obj
      (define (choose-action state)
        (lex-query
          '((action (action-prior)))
          'action
          '(flip (normalize-reward
            (sample-reward action state))))))
  (define (sample-reward action state)
    (let ((next-state (state-transition state action)))
      (+ (reward next-state)
        (if (terminal? next-state)
          0
          (sample-reward (choose-action next-state) next-state)))))
Open directions and challenges

• More precise relation to psychology
  – How does human cognitive processing perform approximate probabilistic inference (i.e., approximately implement rational methods of approximate inference)?

• Relation to the brain
  – How to implement structured probabilistic models in neural architectures?

• Probabilistic models for richly structured knowledge
  – How to formalize an intuitive theory of physics or psychology?

• Effective learning of structured probabilistic models
  – How to balance expressiveness/learnability tradeoff?