Computational cognitive science: Generative models, probabilistic programs, and common sense

Josh Tenenbaum

MIT

Brains, Minds and Machines Summer School 2015

Two notions of intelligence: *Classifying/recognizing/predicting data* vs. *Explaining/understanding/modeling the world*

- What’s the difference between classification and explanation?
- What makes a good explanation?

Two notions of intelligence:
Classifying/recognizing/predicting data vs. *Explaining/understanding/modeling the world*

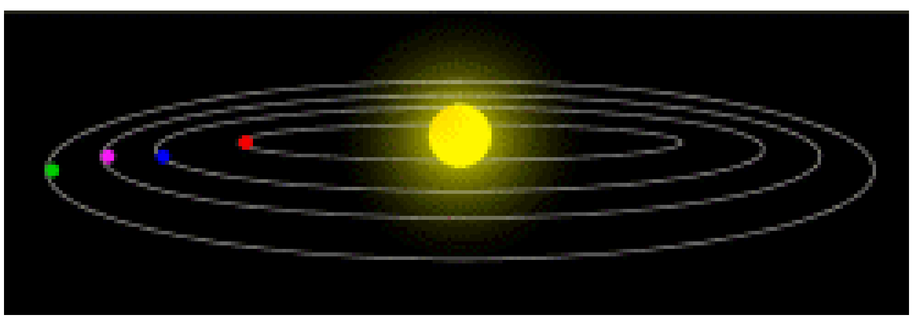
Both notions have roles to play, but here I'll emphasize *explanation*, because it is at the heart of human intelligence, and much of current AI, machine learning, computational neuroscience is so focused on *classification*.

(Why? Building machines that explain and understand is harder than building machines that merely recognize and classify. Classification is easier to map to neural networks and neural circuits.)

But not only are both probably essential, they can interact in powerful, probably essential ways! We'll talk about how deep neural networks can help model-based methods work more quickly, efficiently – or how model-based methods can help model-free methods become richer and more flexible.

Two notions of intelligence:
Classifying/recognizing/predicting data vs. *Explaining/understanding/modeling the world*

- What's the difference between classification and explanation?
- What makes a good explanation?
  - Compact / unifying / nonarbitrary / "hard to vary"
  - Generative: Output is the world, not how we should perform a task.
  - Causal / actionable for an endless range of tasks, via planning
  - Compositional / flexible / extensible

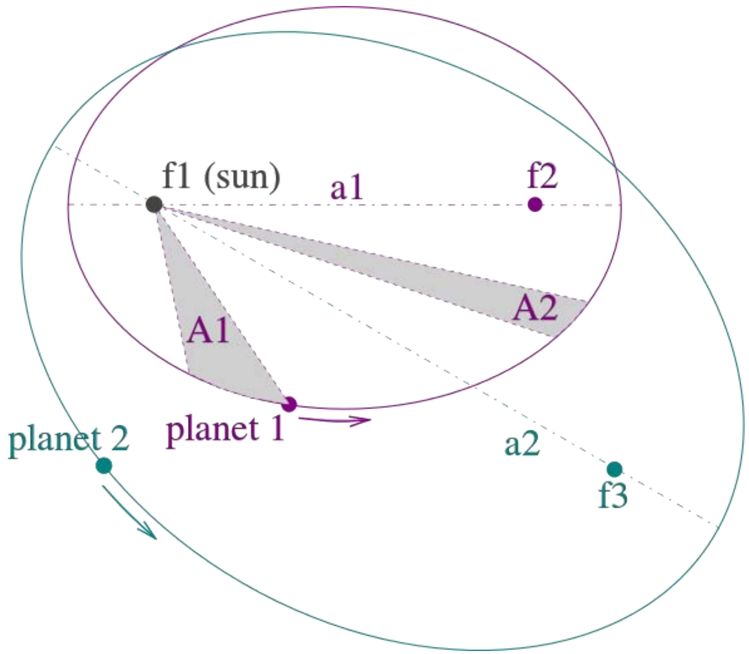


Phenomena: the motion of objects in the solar system.

Contrast Kepler's laws and Newton's laws….

- Kepler's laws:

1. The orbit of a planet is an ellipse with the Sun at one of the two foci.
2. A line segment joining a planet and the Sun sweeps out equal areas during equal intervals of time.[1]
3. The square of the orbital period of a planet is proportional to the cube of the semi-major axis of its orbit.



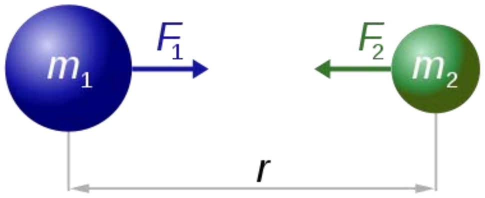
- Newton's laws:

First law: When viewed in an inertial reference frame, an object either remains at rest or continues to move at a constant velocity, unless acted upon by an external force.[2][3]

Second law: The vector sum of the external forces $\mathbf{F}$ on an object is equal to the mass m of that object multiplied by the acceleration vector $\mathbf{a}$ of the object: $\mathbf{F} = m\mathbf{a}$.

Third law: When one body exerts a force on a second body, the second body simultaneously exerts a force equal in magnitude and opposite in direction on the first body.

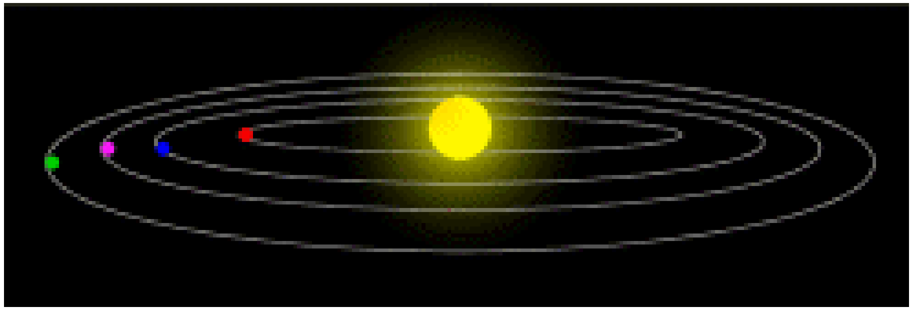
Law of gravitational force:



$$F_1 = F_2 = G\frac{m_1 \times m_2}{r^2}$$

Two notions of intelligence:
Classifying/recognizing/predicting data vs. *Explaining/understanding/modeling the world*

- What's the difference between classification and explanation?
- What makes a good explanation?
    - Compact / unifying / nonarbitrary / "hard to vary"
    - Generative: Output is the world, not how we should perform a task.
    - Causal / actionable for an endless range of tasks, via planning
    - Compositional / flexible / extensible



Newton but not Kepler explains…

- Not just the orbits of planets, but other solar-system objects.
- Not just the motion of planets, but also the apple I drop right here on Earth.
- Why some things orbit other things, but not others.
- How you could get a man to the moon, and back again.
- How you could build a rocket or solar sail or sling shot to escape Earth's gravity,

The brain as a generative modeling engine

The Nature of Explanation (1943):

One of the most fundamental properties of thought is its power of predicting events…. It enables us, for instance, to design bridges with a sufficient factor of safety instead of building them haphazard and waiting to see whether they collapse… If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it. Most of the greatest advances of modern technology have been instruments which extended the scope of our sense-organs, our brains or our limbs. Such are telescopes and microscopes, wireless, calculating machines, typewriters, motor cars, ships and aeroplanes. Is it not possible, therefore, that our brains themselves utilize comparable mechanisms to achieve the same ends and that these mechanisms can parallel phenomena in the external world as a calculating machine can parallel the development of strains in a bridge?

Portrait photo removed due to copyright restrictions.

Kenneth Craik

(1914-1945)

The big question

How does the mind get so much out of so little?

Our minds build rich models of the world and make strong generalizations from input data that is sparse, noisy, and ambiguous – in many ways far too limited to support the inferences we make.

How do we do it?

The big question

How does the mind get so much out of so little, so quickly, so flexibly, on such little energy?

Visual scene perception

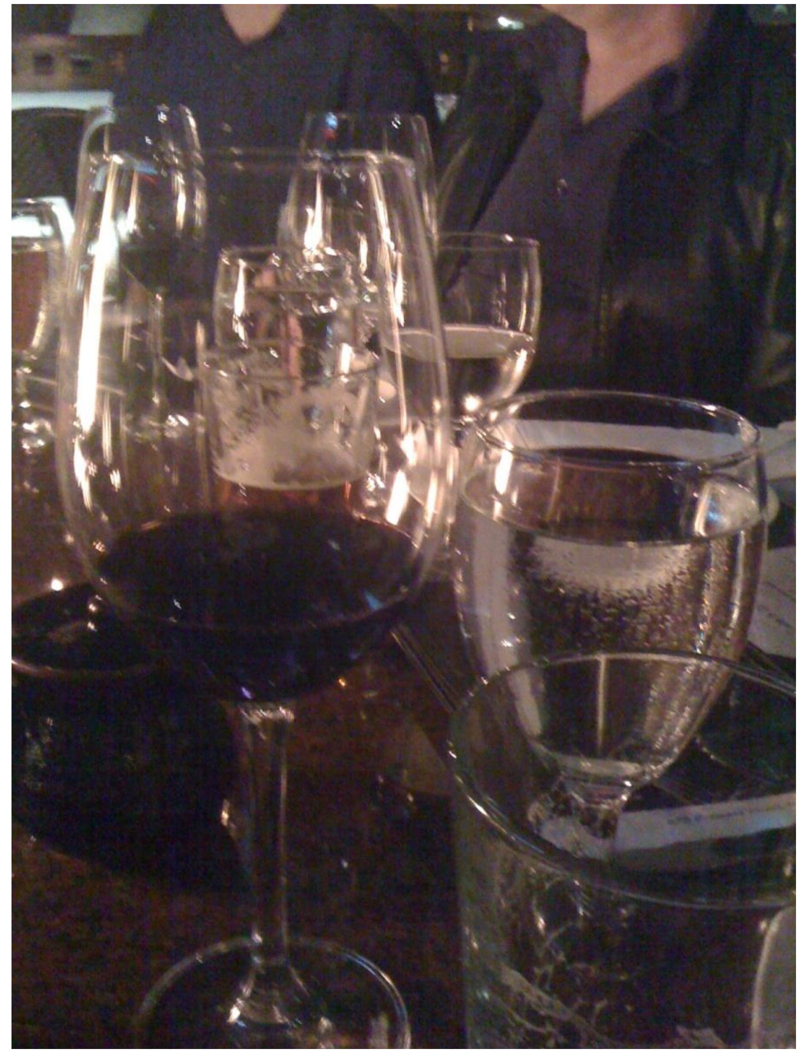
But…
look around you!

Where are the people?

Where are the people?

… books?

… glasses?



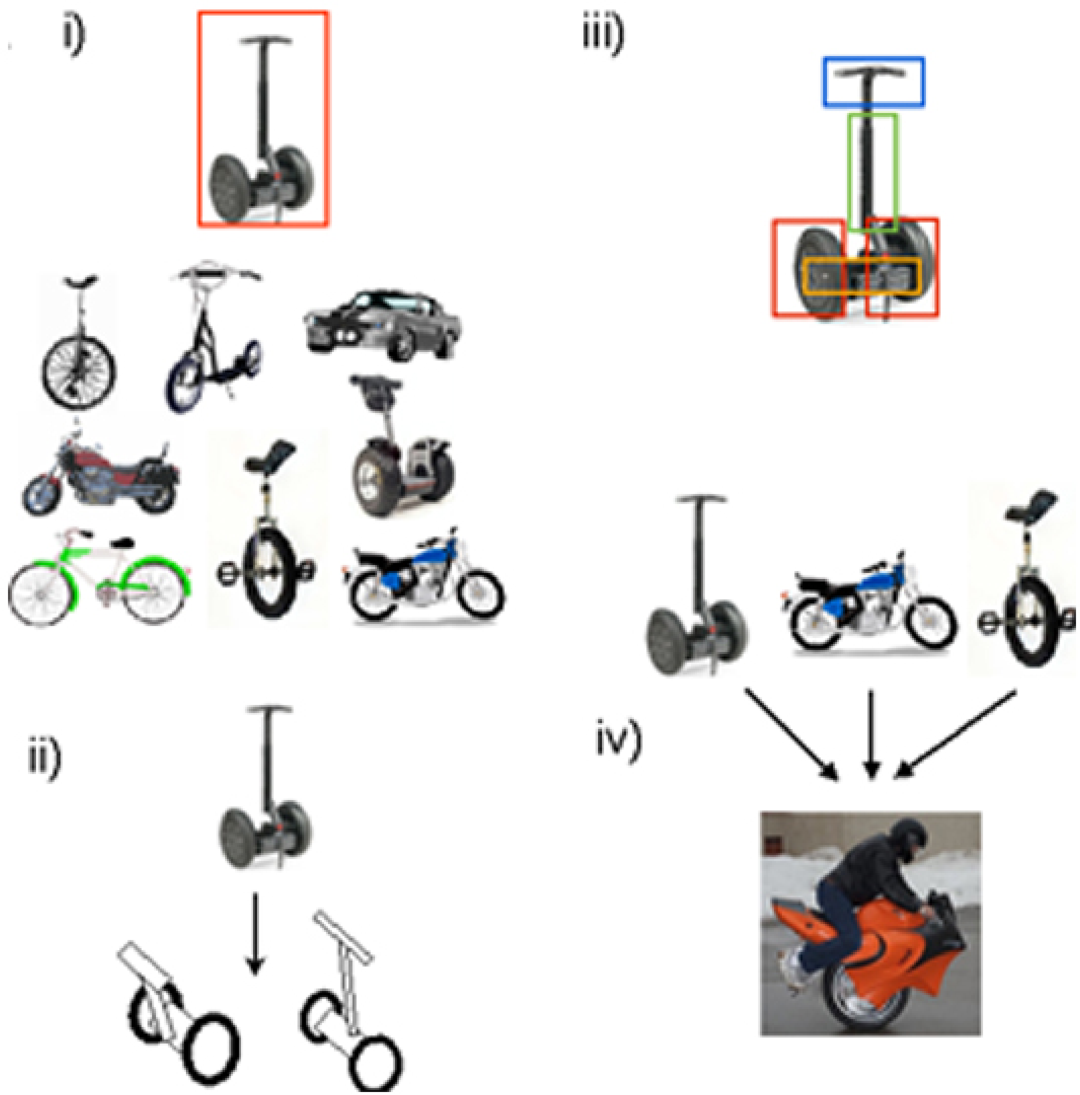
Learning and generalizing object concepts

What’s this?

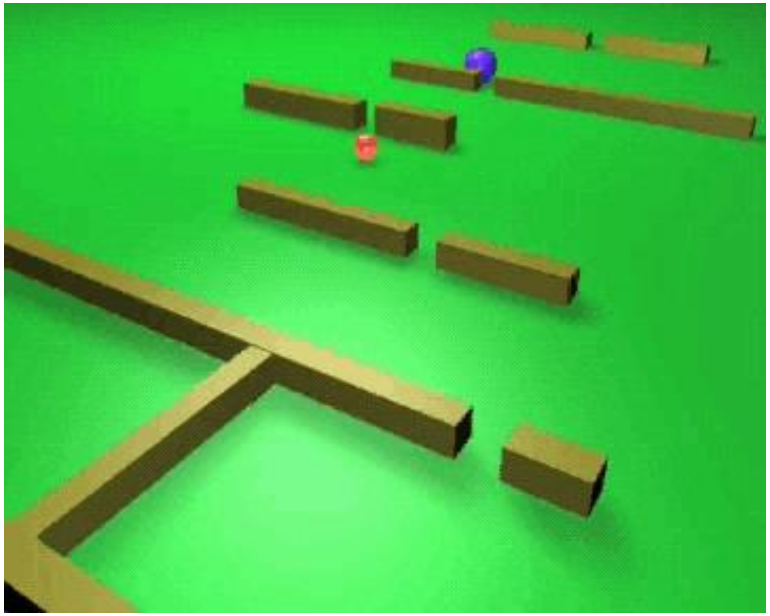
Photo of rock climbing equipment removed due to copyright restrictions. See video.

Photo of rock climbing equipment removed due to copyright restrictions. See video.

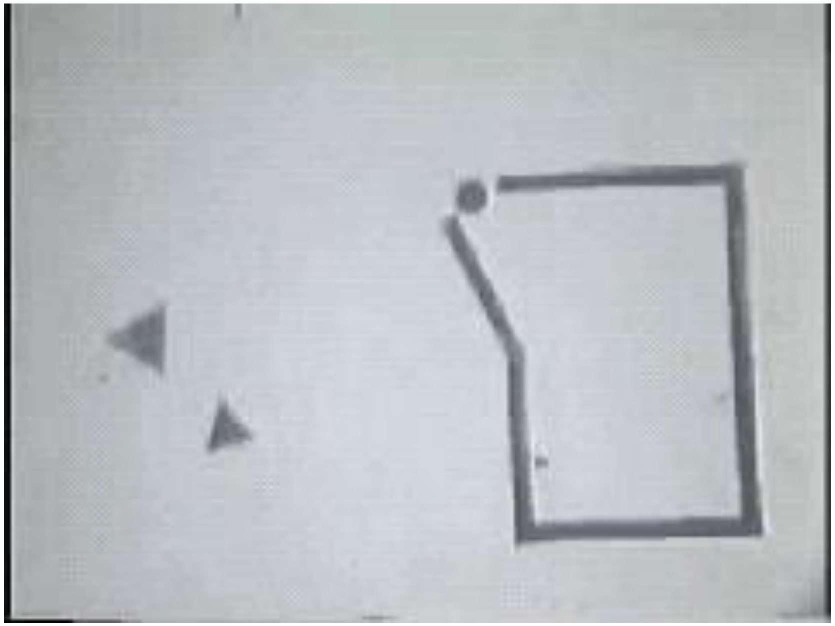
Concept learning is not simply classification



Understanding events with common-sense theories



(Southgate and Csibra)

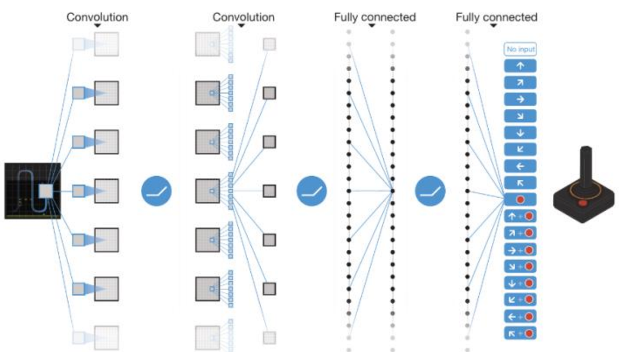


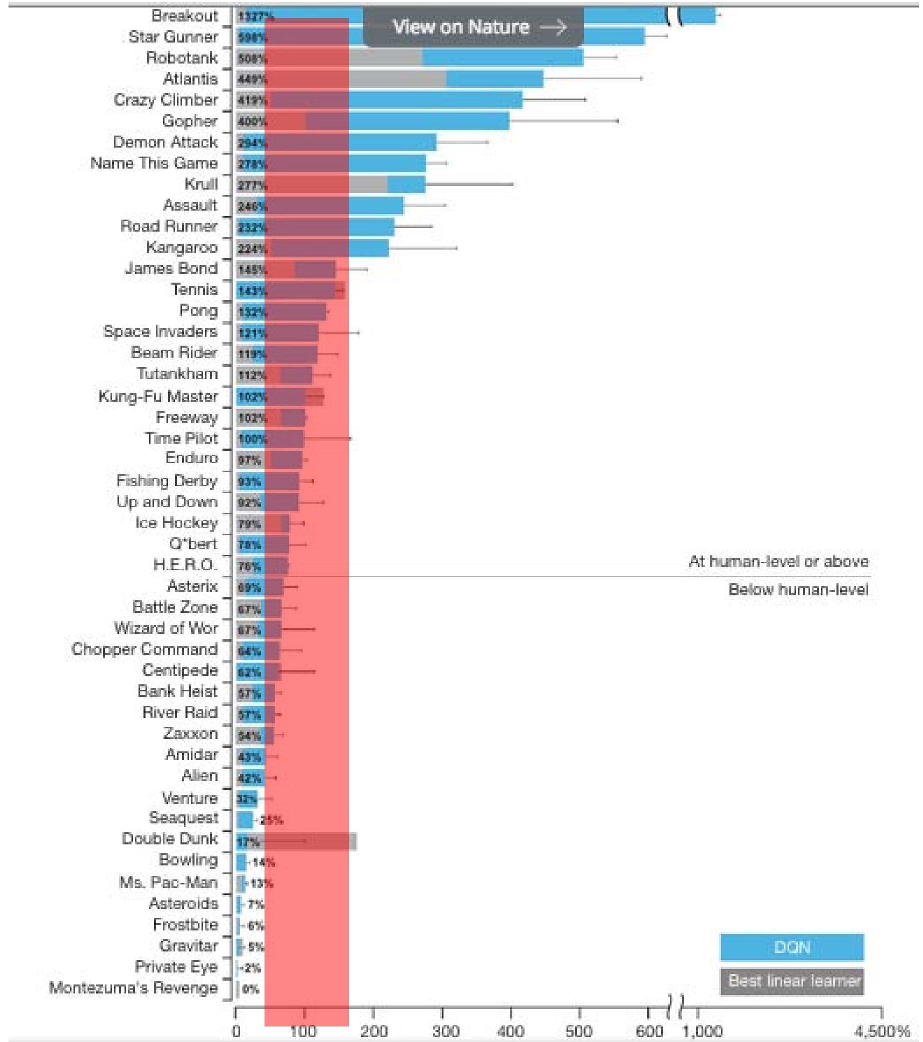
(Heider and Simmel)

Source: Heider, F., & Simmel, M. (1944) "An experimental study in apparent behavior." The American Journal of Psychology, 57, 243-259.

Intuitive physics: objects, forces and masses
Intuitive psychology: beliefs and desires
Intuitive sociology: us and them
Intuitive morality: good and bad

Learning to play video games the way people do?



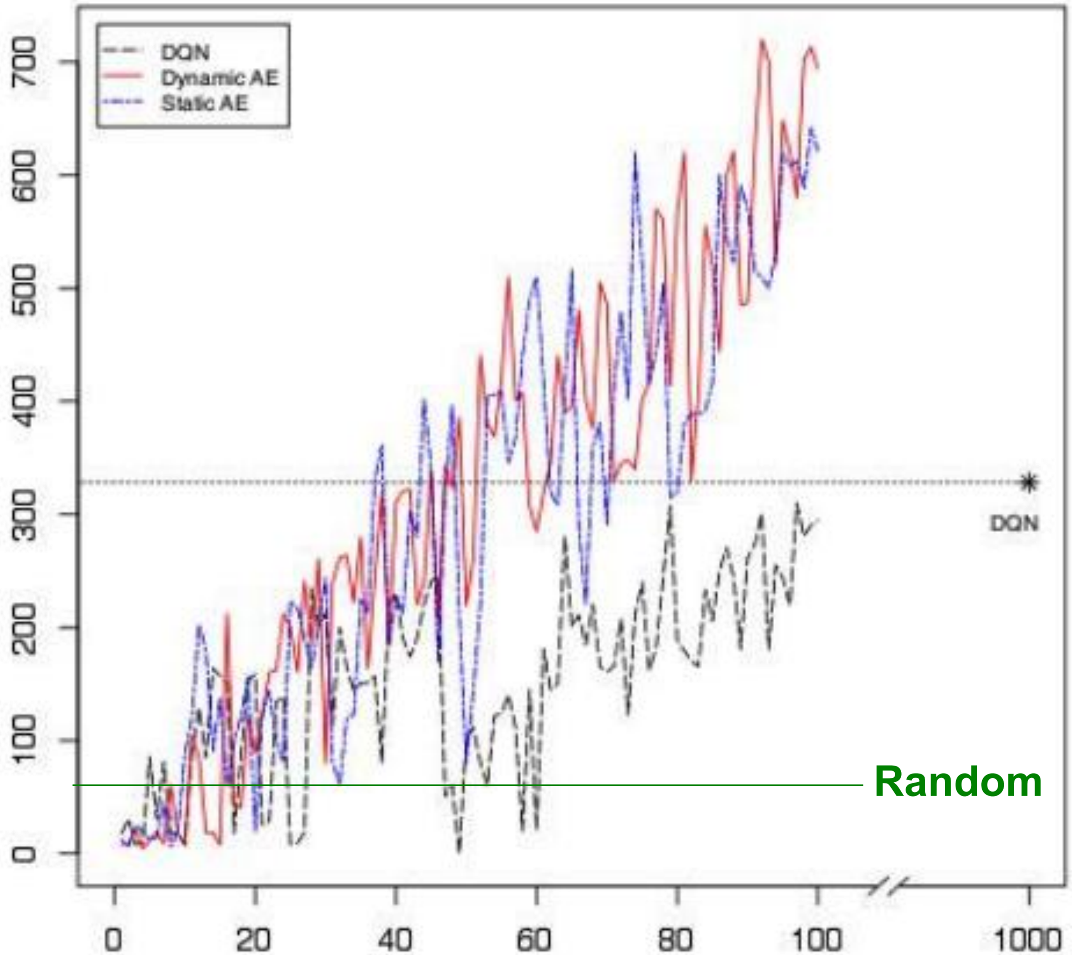


Deep Mind, 2015

Learning to play video games the way people *really* do

Courtesy of sean dreilinger on Flickr. License CC BY-NC-SA.

Learning to play video games the way people *really* do



(Stadie, Levine, Abbeel 2015)

Source: Stadie, Bradly C., Sergey Levine, and Pieter Abbeel. "Incentivizing exploration in reinforcement learning with deep predictive models." arXiv preprint arXiv:1507.00814 (2015).

The big question

How does the mind get so much out of so little?

- Recovering the entire world around you, from a glance, in a flash.
- Learning a generalizable concept from just one example.
- Discovering causal relations from just a single observed event.
- Seeing forces, and seeing inside other minds, from just the motion of a few two-dimensional shapes.
- Learning to play games, solve problems, and act in a whole new world – all in under one minute.
- Understanding the words you’re reading now.

The goal: A computational framework for understanding how people make these inferences, and how they can be successful, expressed in engineering terms.

The problems of induction

Abstract knowledge.

(Constraints / Inductive bias / Priors)

1. How does abstract knowledge guide learning and inference from sparse data?
2. What form does abstract knowledge take, across different domains and tasks?
3. How is abstract knowledge itself constructed, from some combination of innate specifications and experience?

…

The "Generative models" approach

1. How does abstract knowledge guide learning and inference from sparse data?

   Bayesian inference in probabilistic generative models.

$$P(h \mid d) = \frac{P(d \mid h)P(h)}{\sum_{h_i \in H} P(d \mid h_i)P(h_i)}$$

2. What form does that knowledge take, across different domains and tasks?

   Probabilities defined richly structured symbolic representations: spaces, graphs, grammars, logical predicates, schemas...

   Probabilistic Programs

3. How is that knowledge itself constructed?

   Hierarchical models, with inference at multiple levels. Learning models as probabilistic inference; "learning to learn", transfer learning, learning representations and learning inductive biases not fundamentally different.

The approach (cont'd)

4. How can learning and inference proceed efficiently and accurately, even with very complex hypothesis spaces?

   Sampling-based algorithms for approximate inference, e.g., MCMC, sequential Monte Carlo ("particle filtering"), importance sampling. Cost-sensitive sampling ("One and done"). Fast initialization with bottom-up recognition models ("Neural networks").

5. How can probabilistic inferences be used to drive action?

   Utility-based frameworks for decision and planning under uncertainty and risk, such as Bayesian decision theory or Markov decision processes (MDPs).

6. How could these computations be implemented in neural hardware, or massively parallel computing machines?

   Probabilistic interpretations of cortical circuitry and neural population codes; stochastic digital circuits.

1990s-present: Cognition as probabilistic inference

Visual perception [Yuille, Weiss, Simoncelli, Adelson, Richards, Freeman, Feldman, Kersten, Knill, Maloney, Olshausen, Jacobs, Pouget, ...]

Language processing and acquisition [Brent, de Marken, Niyogi, Klein, Manning, Jurafsky, Chater, Keller, Levy, Hale, Johnson, Griffiths, Perfors, Tenenbaum, Frank, Piantadosi, O'Donnell, Goodman…]

Motor learning and motor control [Ghahramani, Jordan, Wolpert, Koerding, Kawato, Doya, Todorov, Shadmehr, Maloney, …]

Reinforcement learning [Dayan, Daw,, Niv, Frank, Gershman, Gureckis, …]

Memory [Anderson, Schooler, Shiffrin, Steyvers, Griffiths, McClelland, Gershman …]

Attention [Mozer, Huber, Torralba, Oliva, Geisler, Yu, Itti, Baldi, Vul, …]

Categorization and concept learning [Anderson, Nosfosky, Rehder, Navarro, Griffiths, Feldman, Tenenbaum, Rosseel, Goodman, Kemp, Mansinghka, …]

Reasoning [Chater, Oaksford, Sloman, McKenzie, Heit, Tenenbaum, Kemp, Goodman…]

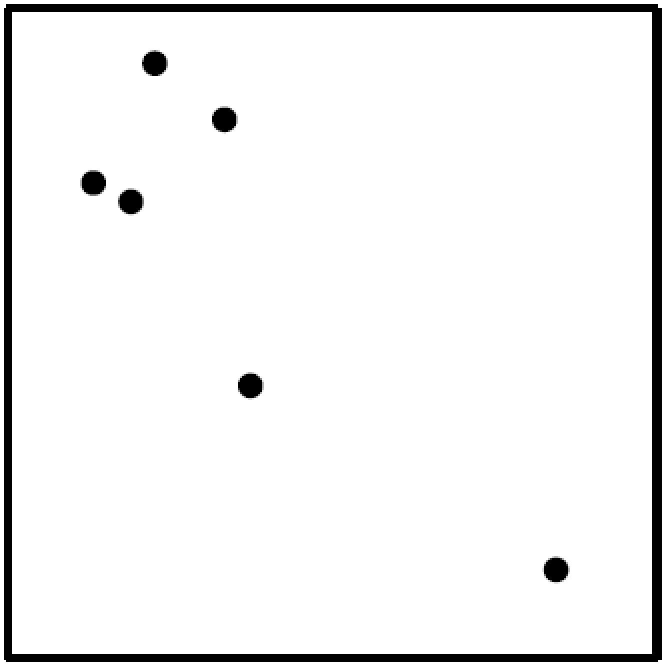
Causal inference and learning [Waldmann, Sloman, Steyvers, Griffiths, Tenenbaum, Yuille, Lu, Holyoak, Lagnado, …]

Basic cognitive capacities as intuitive probabilistic inference

- Similarity (Tenenbaum & Griffiths, *BBS* 2001; Kemp & Tenenbaum, *Cog Sci* 2005)
- Representativeness and evidential support (Tenenbaum & Griffiths, *Cog Sci* 2001)
- Causal judgment (Steyvers et al., 2003; Griffiths & Tenenbaum, *Cog. Psych.* 2005)
- Coincidences and causal discovery (Griffiths & Tenenbaum, *Cog Sci* 2001; *Cognition* 2007; *Psych. Review*, in press)
- Diagnostic inference (Krynski & Tenenbaum, *JEP: General* 2007)
- Predicting the future (Griffiths & Tenenbaum, *Psych. Science* 2006)

Causes and coincidences:
Mere randomness or a hidden cause?

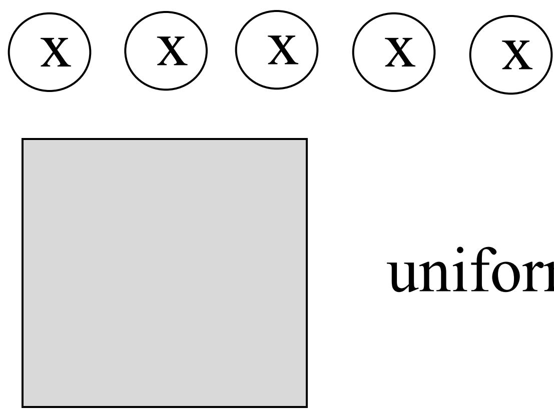
(Griffiths & Tenenbaum, Cognition 2007; Psych. Review, 2009)



Source: Griffiths, T. L., and J. B. Tenenbaum. "Theory-Based Causal Induction." *Psychological Review* 116, no. 4 (2009): 661-716.

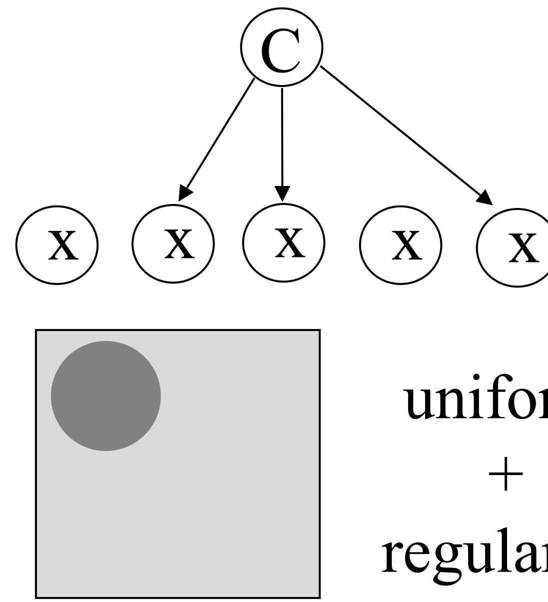
Bayesian measure of evidence: $\log \frac{P(d \mid latent)}{P(d \mid random)}$

Random:



uniform

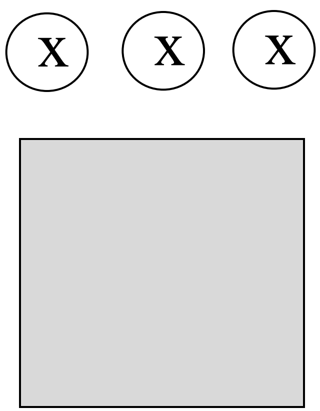
Latent common cause:



uniform
+
regularity

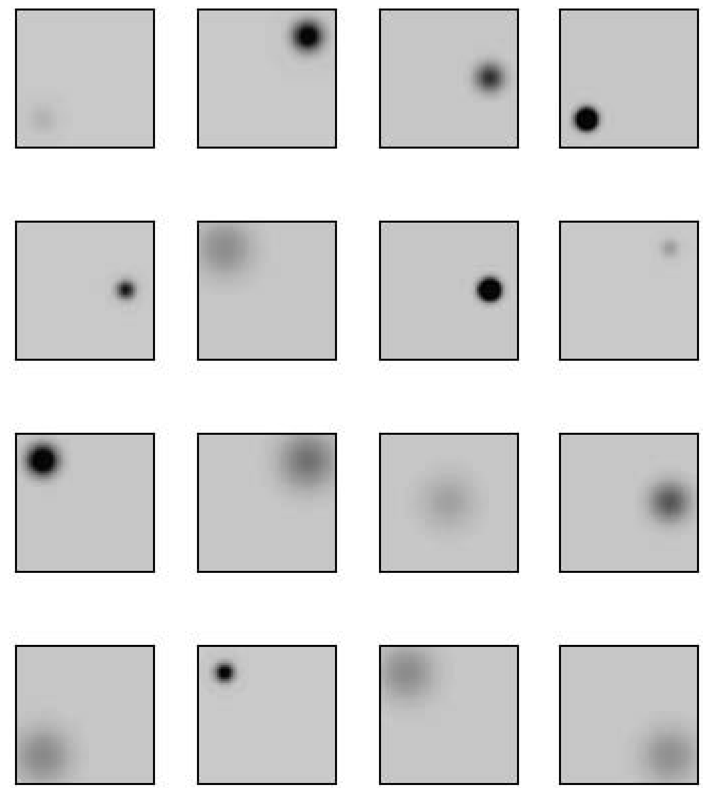
Bayesian measure of evidence: $\log \dfrac{P(d \mid latent)}{P(d \mid random)}$

Random:



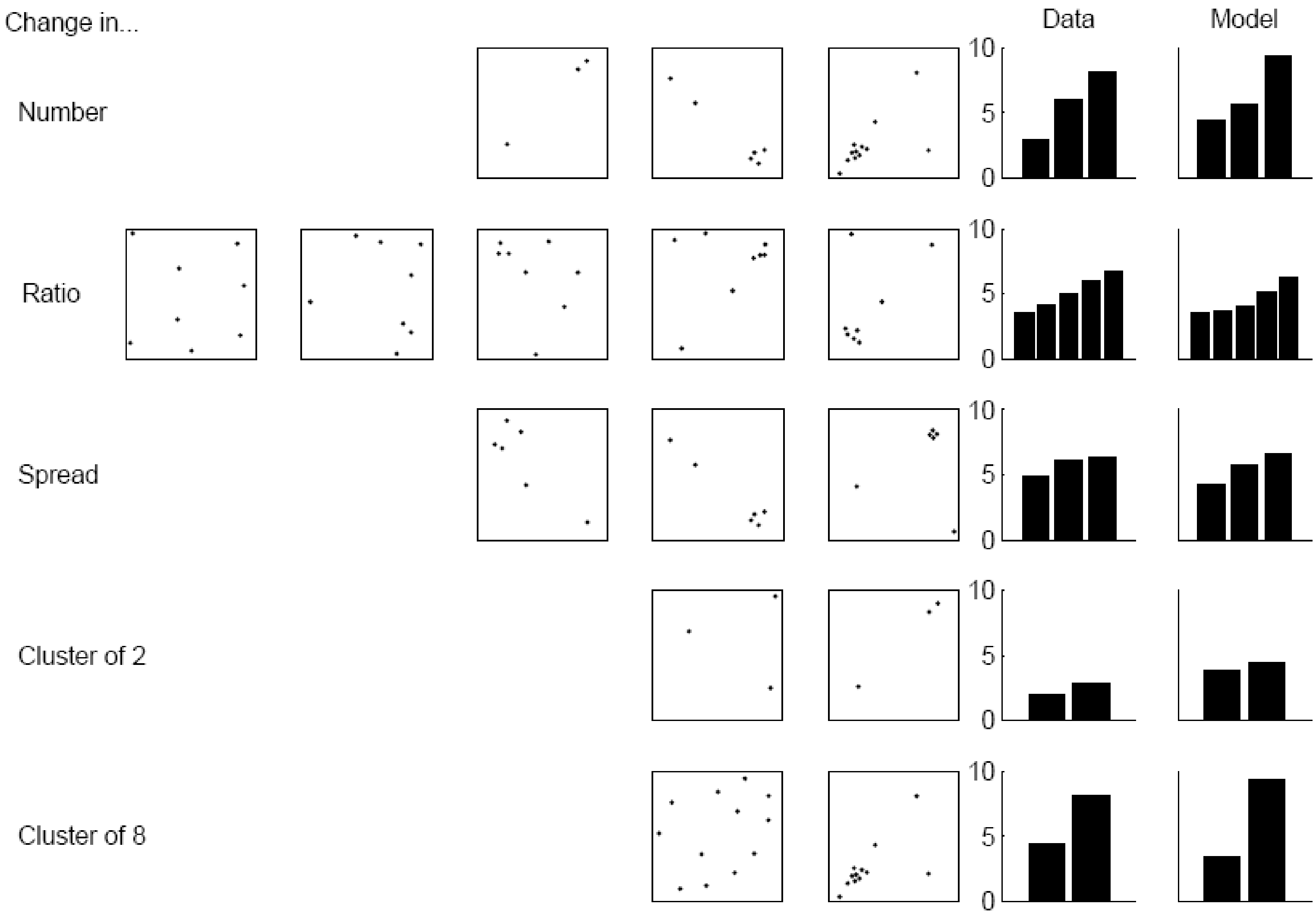
uniform

Latent common cause:

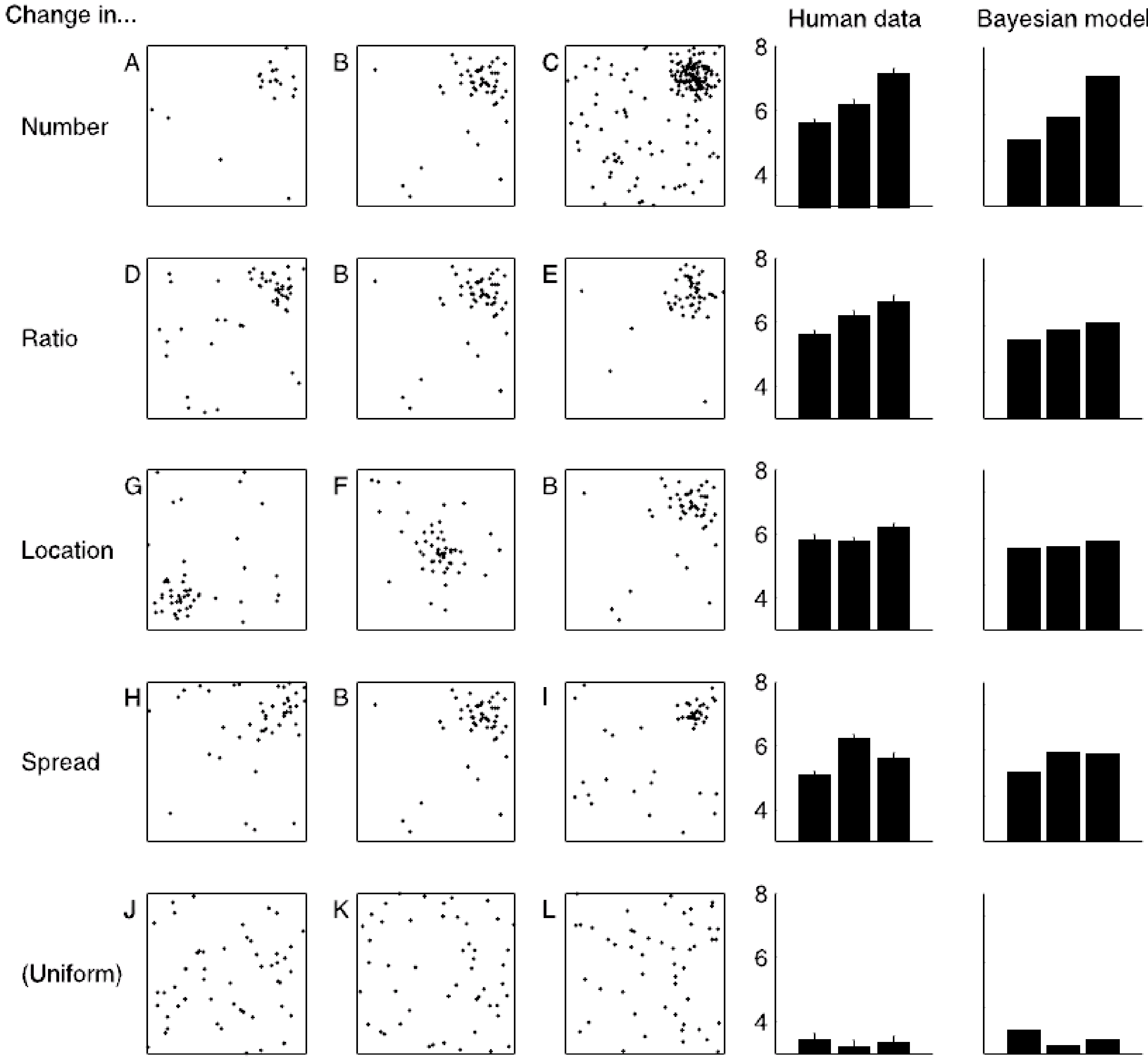


Cancer clusters?

Judging the probability of a hidden environmental cause



Source: Griffiths, T. L., and J. B. Tenenbaum. "Theory-Based Causal Induction." *Psychological Review* 116, no. 4 (2009): 661-716.

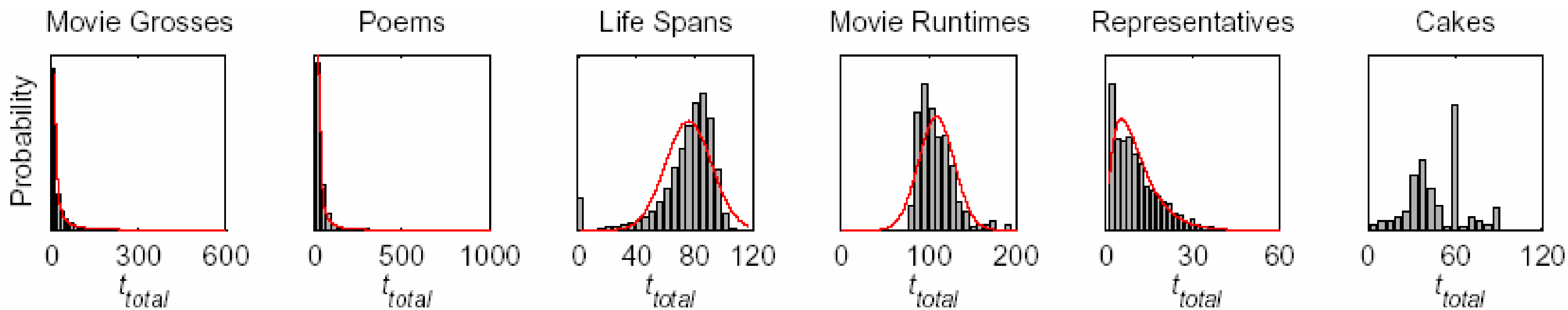

Change in...
Human data
Bayesian model
Number
Ratio
Location
Spread
(Uniform)
A
B
C
D
E
F
G
H
I
J
K
L
8
6
4

Everyday prediction problems

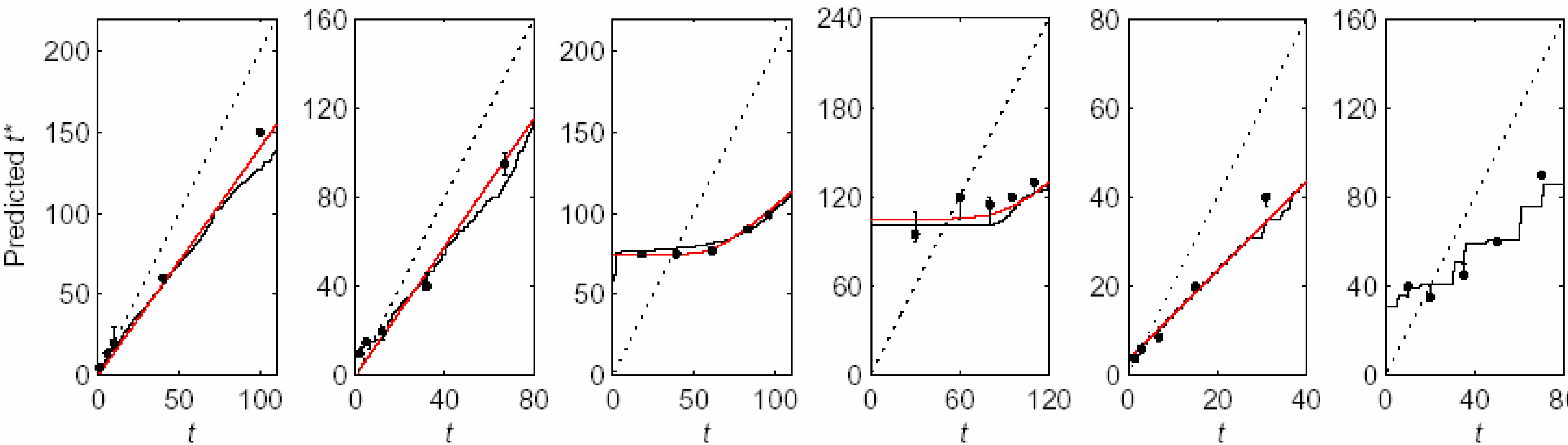
(Griffiths & Tenenbaum, Psych Science 2006)

- You read about a movie that has made $60 million to date. How much money will it make in total?
- You see that something has been baking in the oven for 34 minutes. How long until it's ready?
- You meet someone who is 78 years old. How long will they live?
- Your friend quotes to you from line 17 of his favorite poem. How long is the poem?
- You meet a US congressman who has served for 11 years. How long will he serve in total?
- You encounter a phenomenon or event with an unknown extent or duration, t_{total}, at a random time or value of $t < t_{total}$. What is the total extent or duration t_{total}?

Priors $P(t_{total})$ based on empirically measured durations or magnitudes for many real-world events in each class:



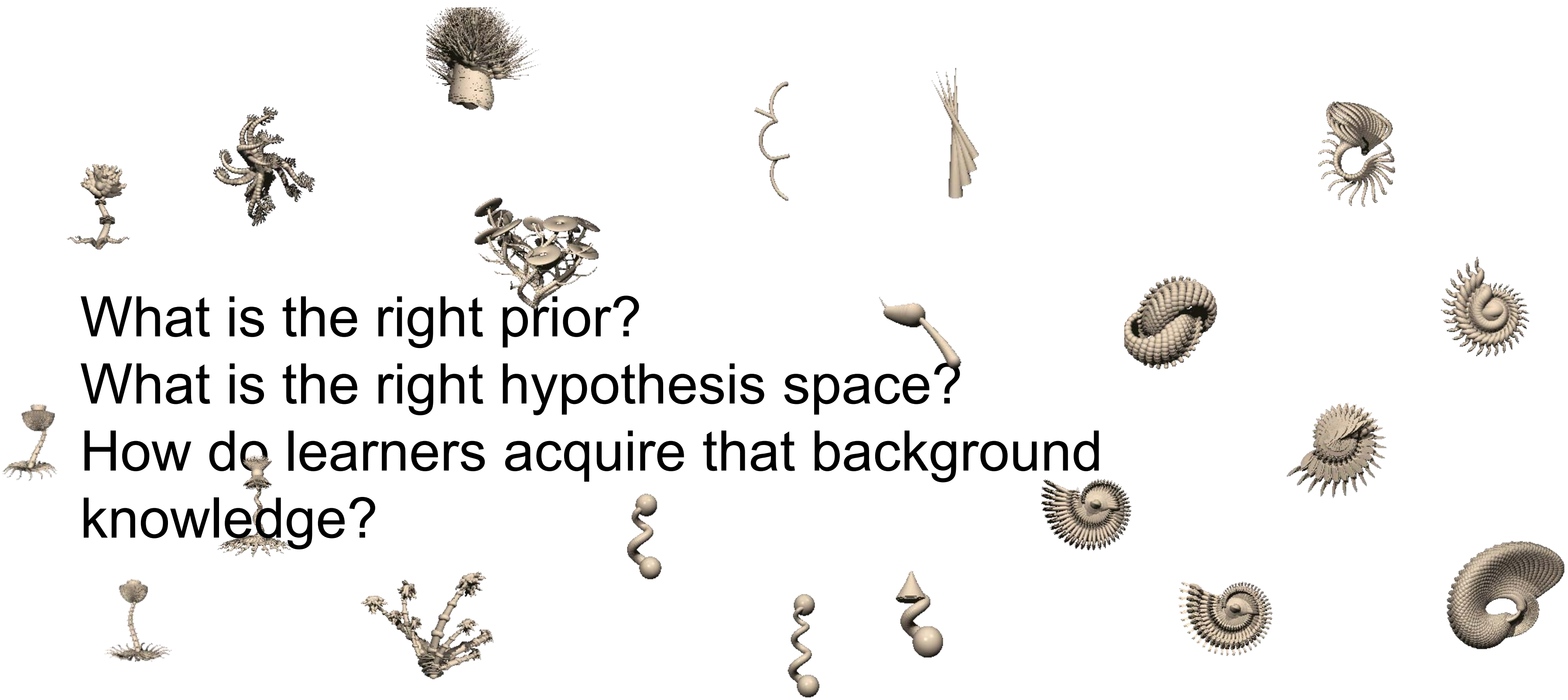
Median human judgments of the total duration or magnitude t_{total} of events in each class, given one random observation at a duration or magnitude t, versus Bayesian predictions (median of $P(t_{total}|t)$).



See Griffiths and Tenenbaum, "Optimal Predictions in Everyday Cognition." *Psychological Science* 17, no. 9 (2006)

Learning words for objects

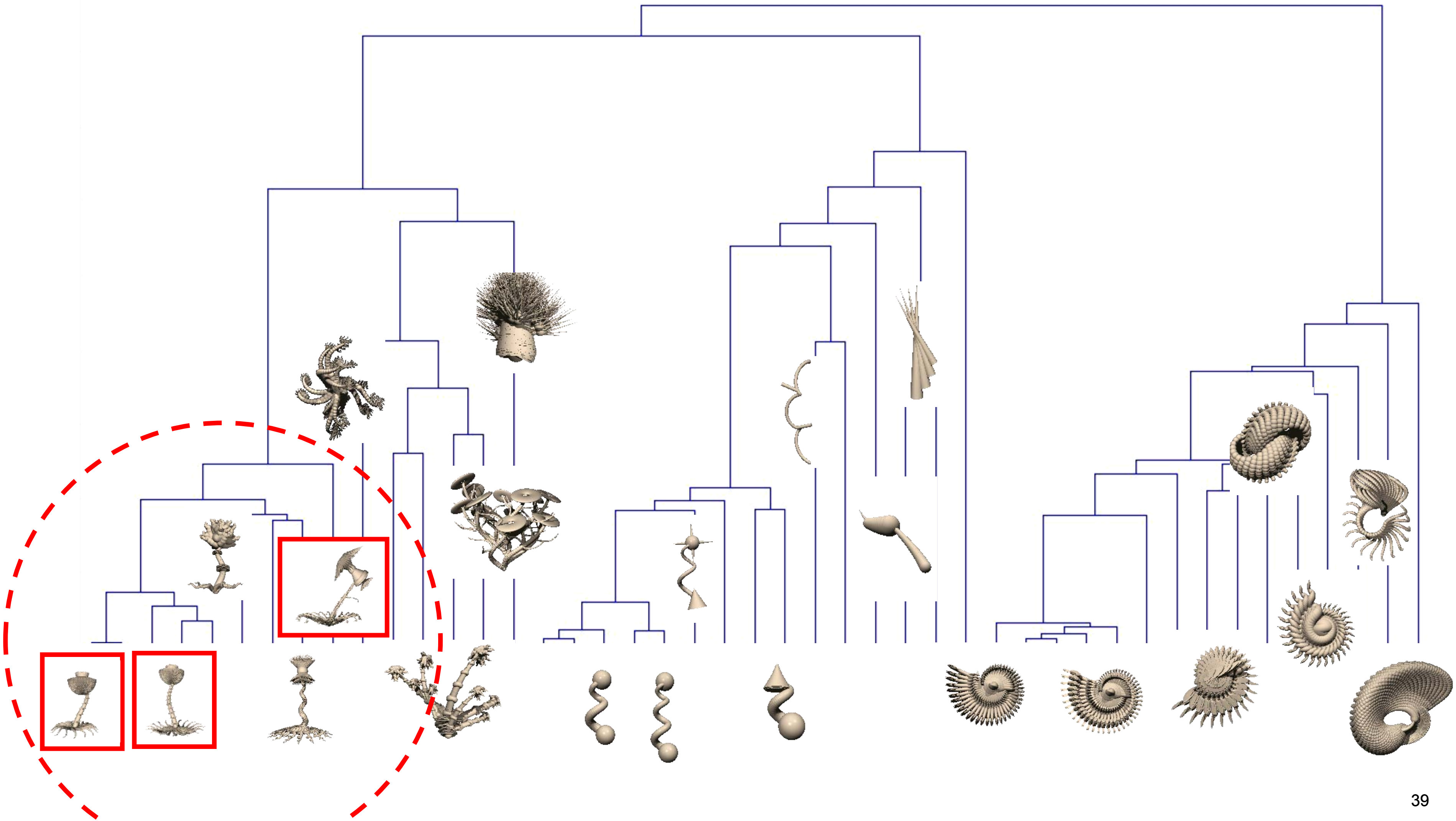
Word learning as Bayesian inference
(Xu & Tenenbaum, *Psych Review*, 2007)



What is the right prior?
What is the right hypothesis space?
How do learners acquire that background knowledge?

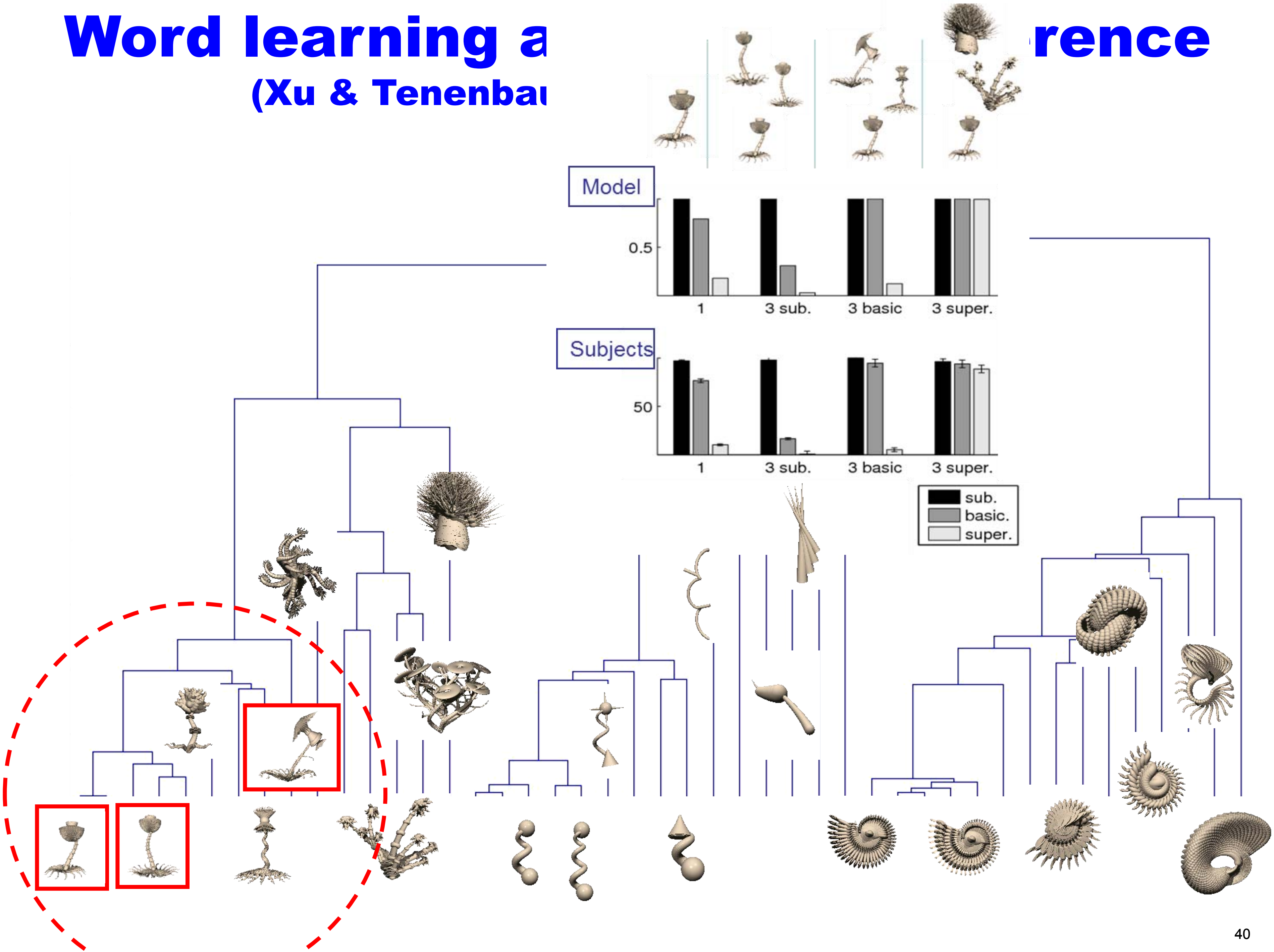
Word learning as Bayesian inference

(Xu & Tenenbaum, *Psych Review*, 2007)



Word learning a… …rence

(Xu & Tenenba…



Property induction

Gorillas have T9 hormones.
Seals have T9 hormones.

Horses have T9 hormones.

"Similarity"
"Typicality"
"Diversity"

Gorillas have T9 hormones.
Seals have T9 hormones.

Anteaters have T9 hormones.

Gorillas have T9 hormones.
Chimps have T9 hormones.
Monkeys have T9 hormones.
Baboons have T9 hormones.

Horses have T9 hormones.

Experiments on property induction
(Osherson, Smith, Wilkie, Lopez, Shafir, 1990)

- 20 subjects rated the strength of 45 arguments:

X_1 have property P. (e.g., Cows have T4 hormones.)
X_2 have property P.
X_3 have property P.

All mammals have property P. [General argument]

- 20 subjects rated the strength of 36 arguments:

X_1 have property P.
X_2 have property P.

Horses have property P. [Specific argument]

Feature rating data
(Osherson and Wilkie)

- People were given 48 animals, 85 features, and asked to rate whether each animal had each feature.

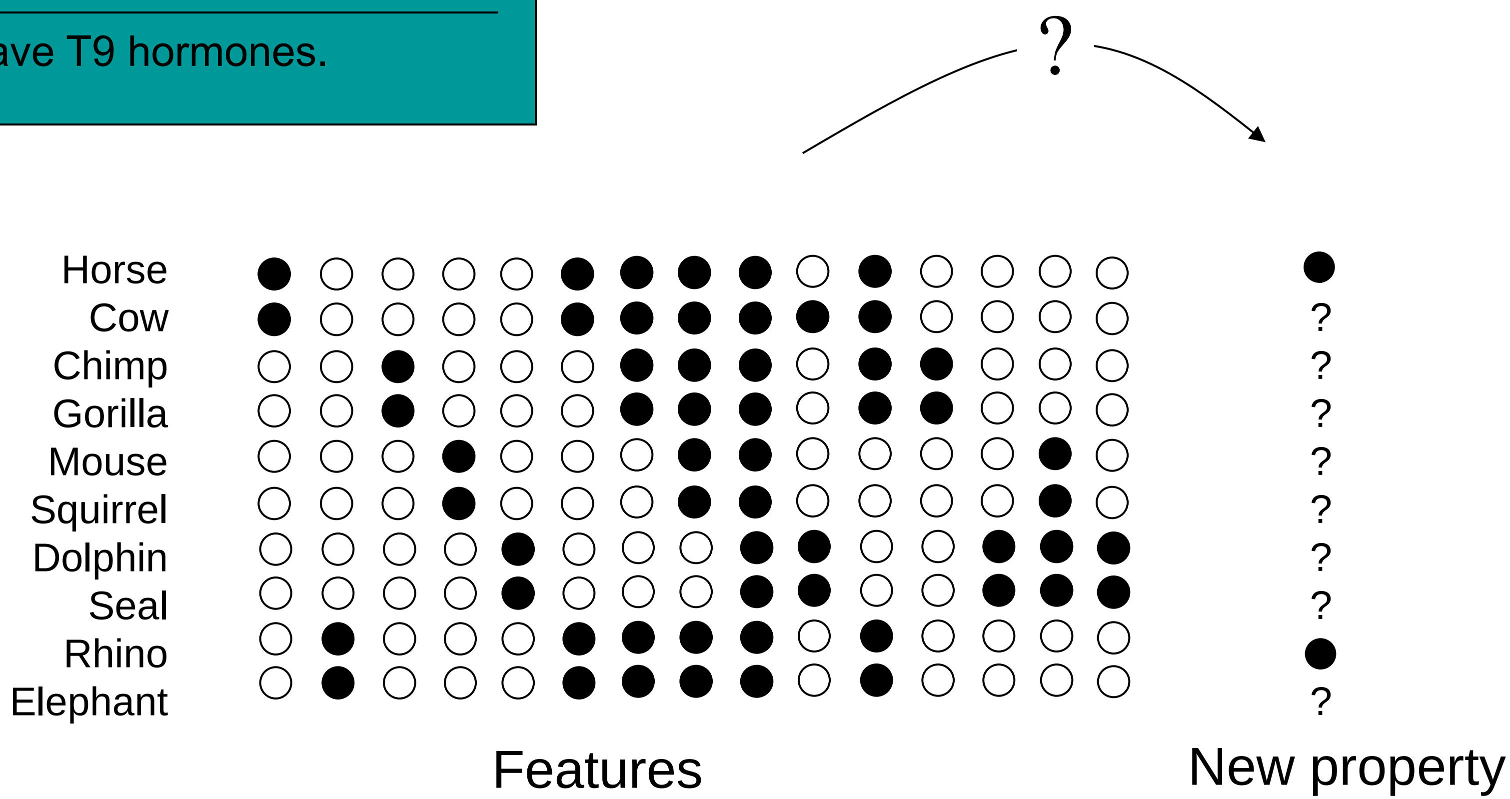
  E.g., elephant:

'gray' 'hairless' 'toughskin'
'big' 'bulbous' 'longleg'
'tail' 'chewteeth' 'tusks'
'smelly' 'walks' 'slow'
'strong' 'muscle' 'quadrapedal'
'inactive' 'vegetation' 'grazer'
'oldworld' 'bush' 'jungle'
'ground' 'timid' 'smart'
'group', …

The computational problem

Horses have T9 hormones.
Rhinos have T9 hormones.

Cows have T9 hormones.



Cf. semi-supervised learning,
sparse matrix completion

Hierarchical Bayesian Framework

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

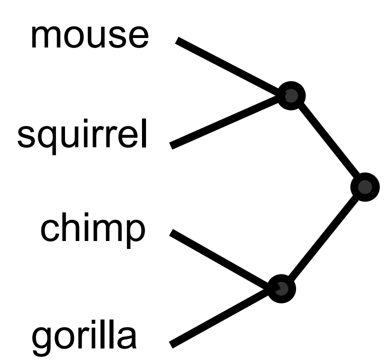
P(form)

F: form

Tree with species at leaf nodes

P(structure | form)

S: structure



P(data | structure)

D: data

	F1	F2	F3	F4		Has T9 hormones
mouse	●	○	○	●		?
squirrel	●	○	○	○	…	?
chimp	○	●	●	●		●
gorilla	○	●	●	●		?

A graph-based prior
(c.f., diffusion model of genetic variation)

Let d_{ij} = length of the edge between objects i and j

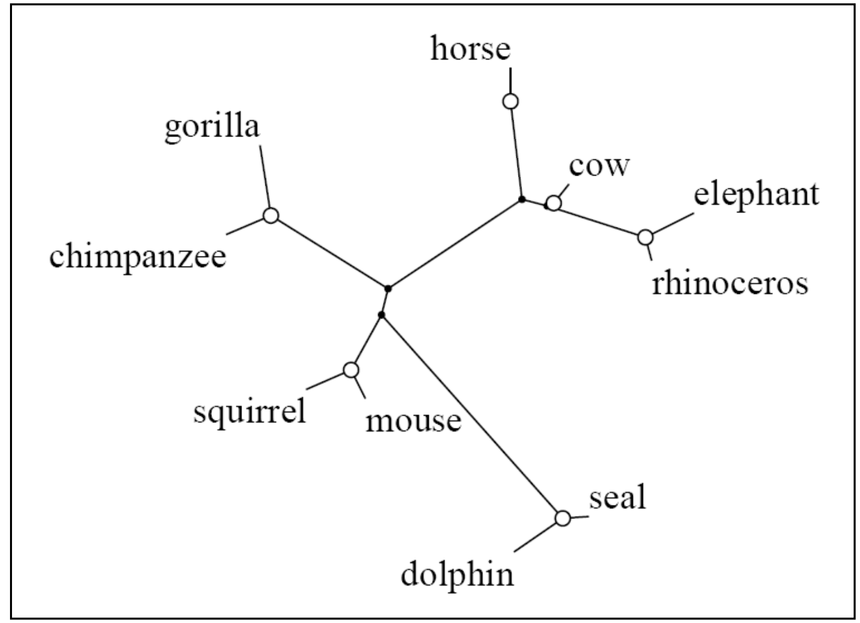
($= \infty$ if i and j are not connected in S),

f_i = value of the feature for object i.

$$p(f \mid S) \propto \exp\left(-\frac{1}{4} \sum_{ij} \frac{(f_i - f_j)^2}{d_{ij}} - \frac{1}{2\sigma} f^T f \right)$$

A Gaussian prior $\sim N(0, \Sigma)$, with $\Sigma = \tilde{\Delta}^{-1}(S)$.
(Zhu, Lafferty & Ghahramani, 2003)

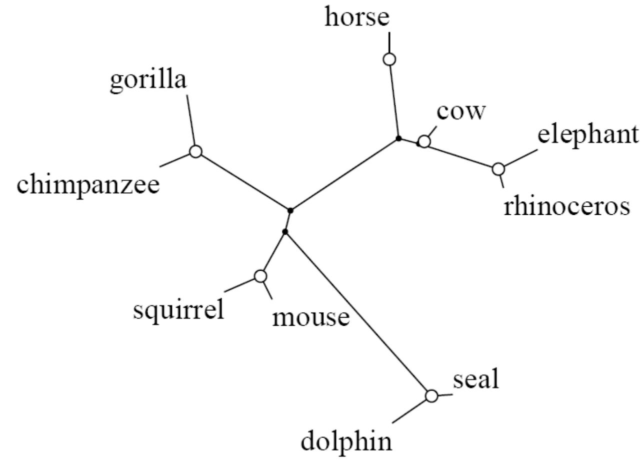
Structure S



Species 1	●	○	○	○	○	●	●	●	●	○	●	○	○	○	○	●
Species 2	●	○	○	○	○	●	●	●	●	●	●	○	○	○	○	?
Species 3	○	○	●	○	○	○	●	●	●	○	●	●	○	○	○	?
Species 4	○	○	●	○	○	○	●	●	●	○	●	●	○	○	○	?
Species 5	○	○	○	●	○	○	○	●	●	○	○	○	○	●	○	?
Species 6	○	○	○	●	○	○	○	●	●	○	○	○	○	●	○	?
Species 7	○	○	○	○	●	○	○	○	●	●	○	○	●	●	●	?
Species 8	○	○	○	○	●	○	○	○	●	●	○	○	●	●	●	?
Species 9	○	●	○	○	○	●	●	●	●	○	●	○	○	○	○	●
Species 10	○	●	○	○	○	●	●	●	●	○	●	○	○	○	○	?

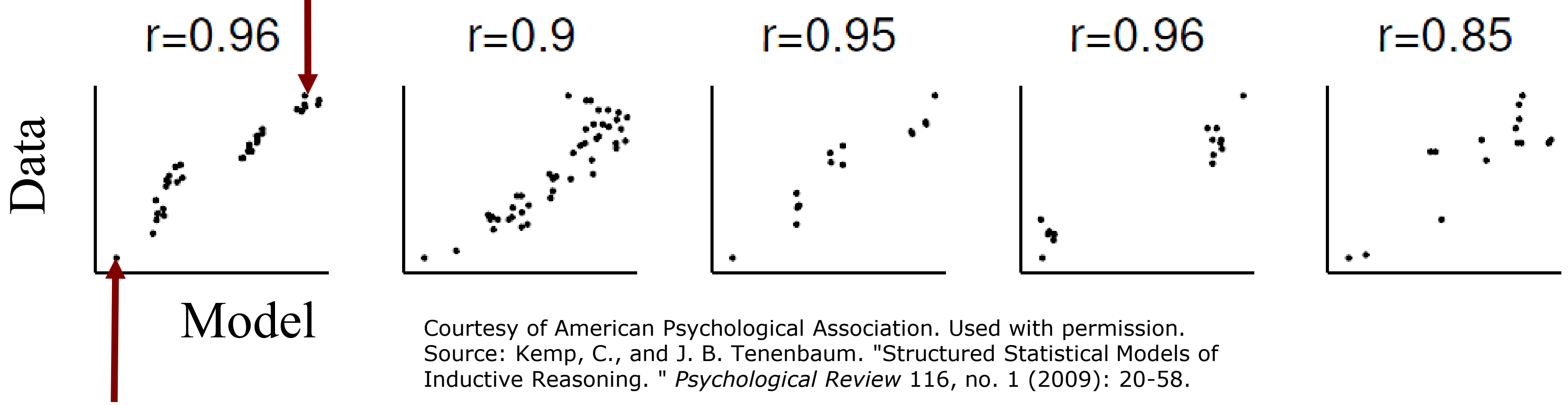
Features f New property

Results



Cows have property P.
Elephants have property P.

Horses have property P.



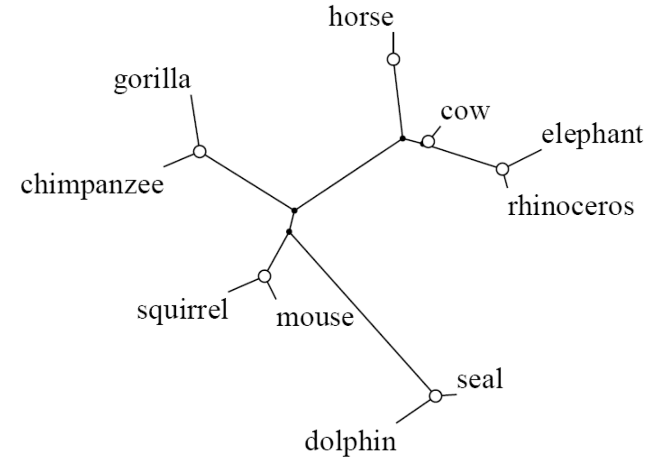
Source: Kemp, C., and J. B. Tenenbaum. "Structured Statistical Models of Inductive Reasoning. " *Psychological Review* 116, no. 1 (2009): 20-58.

Dolphins have property P.
Seals have property P.

Horses have property P.

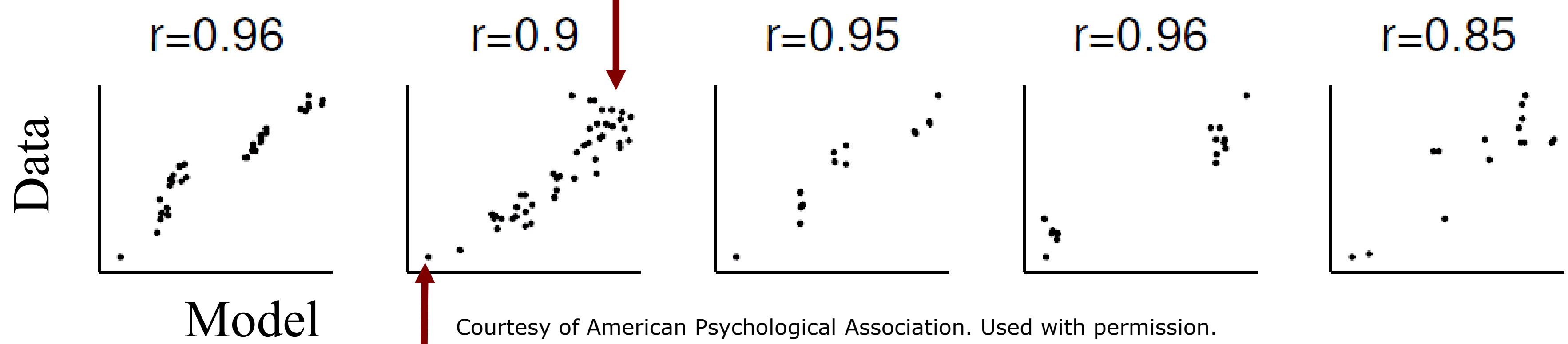
(Osherson et al, Smith et al)

Results



Gorillas have property P.
Mice have property P.
Seals have property P.

All mammals have property P.



Courtesy of American Psychological Association. Used with permission.
Source: Kemp, C., and J. B. Tenenbaum. "Structured Statistical Models of Inductive Reasoning." *Psychological Review* 116, no. 1 (2009): 20-58.

Cows have property P.
Elephants have property P.
Horses have property P.

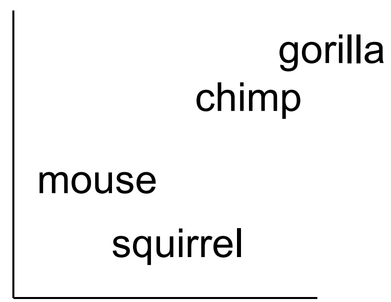
All mammals have property P.

Hierarchical Bayesian Framework

F: form

Low-dimensional space of species

S: structure

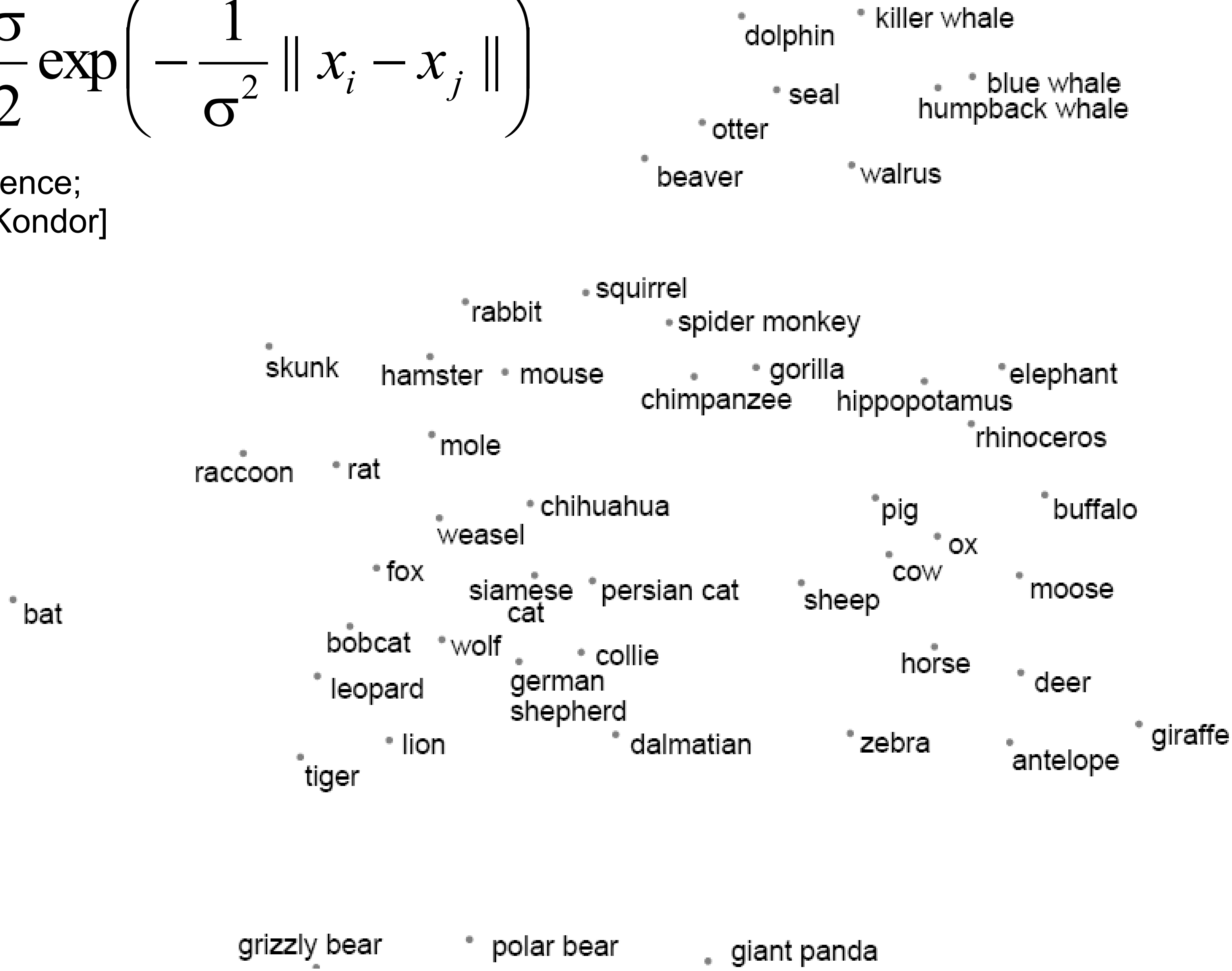


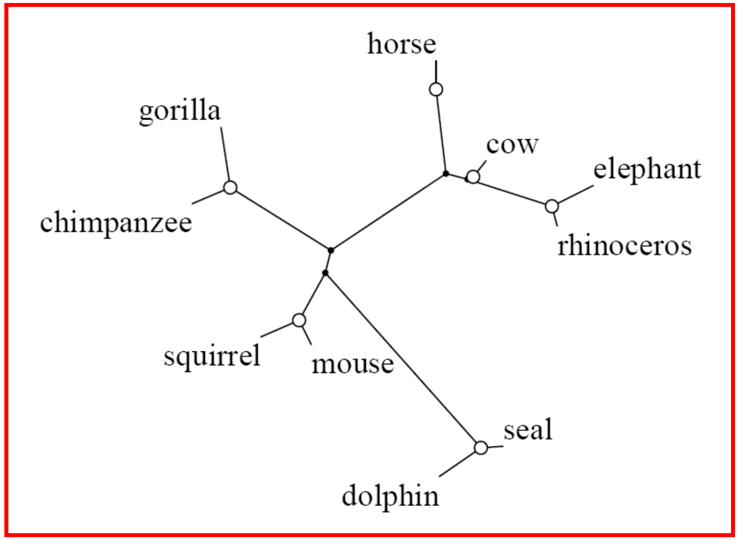
D: data

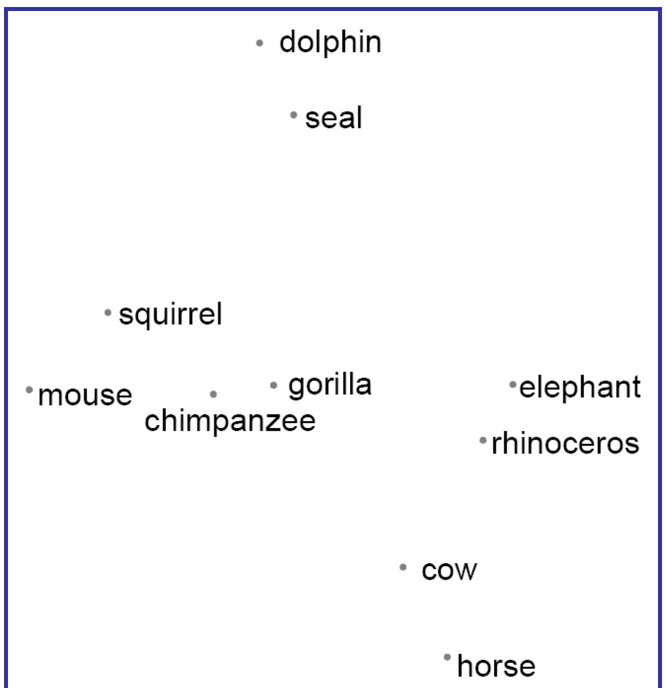
	F1	F2	F3	F4		Has T9 hormones
mouse	●	○	○	●		?
squirrel	●	○	○	○	…	?
chimp	○	●	●	●		●
gorilla	○	●	●	●		?

$$\Sigma_{ij} = \frac{\sigma}{2}\exp\left(-\frac{1}{\sigma^2}\| x_i - x_j \|\right)$$

[c.f., Lawrence;
Smola & Kondor]

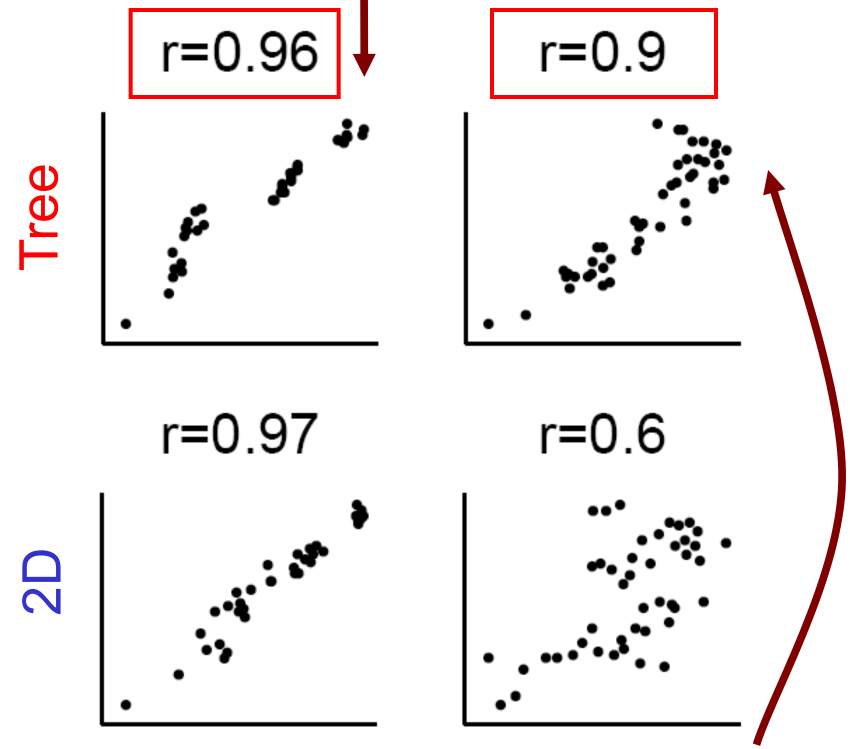






Cows have property P.
Elephants have property P.

Horses have property P.



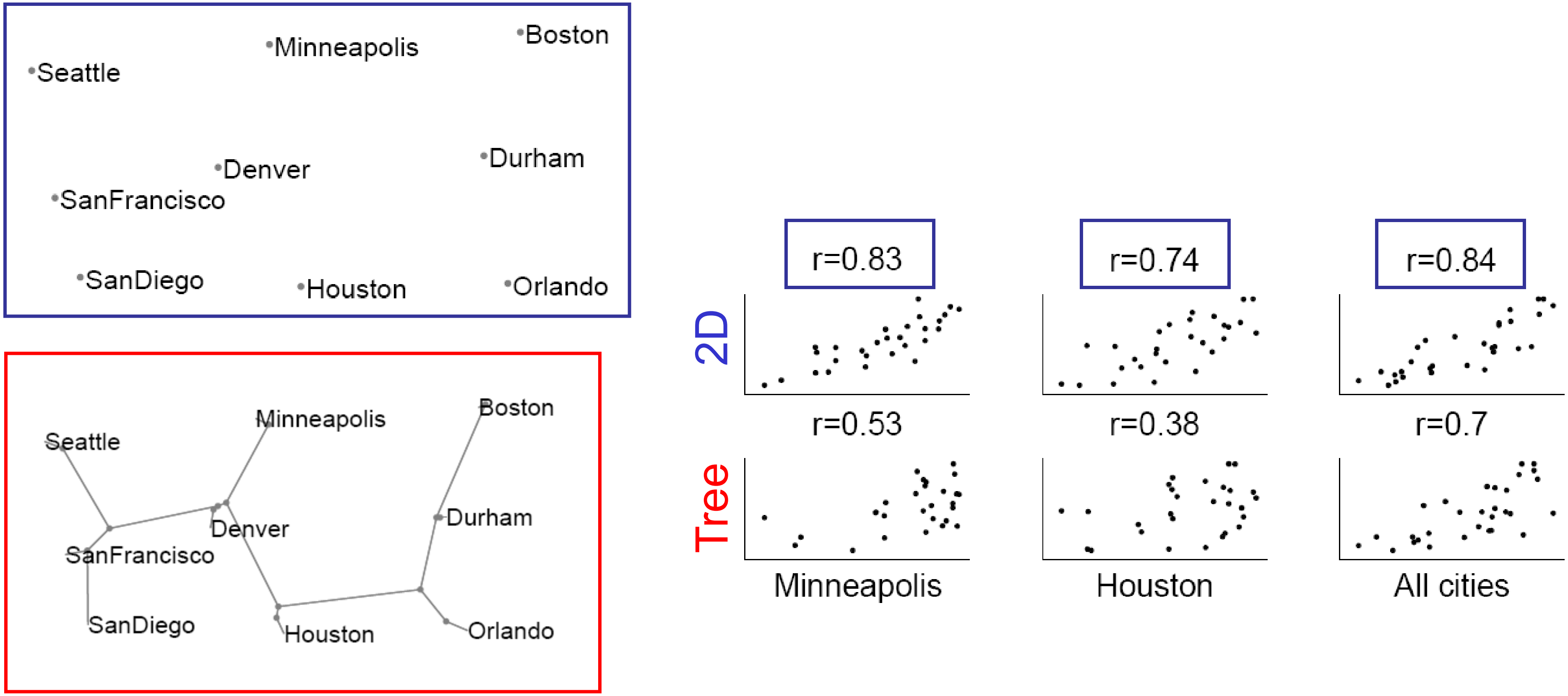
Courtesy of American Psychological Association. Used with permission. Source: Kemp, C., and J. B. Tenenbaum. "Structured Statistical Models of Inductive Reasoning. " *Psychological Review* 116, no. 1 (2009): 20-58.

Gorillas have property P.
Mice have property P.
Seals have property P.

All mammals have property P.

Reasoning about spatially varying properties

Geographic inference task: e.g., "Given that a certain kind of native American artifact has been found in sites near city X, how likely is the same artifact to be found near city Y?"



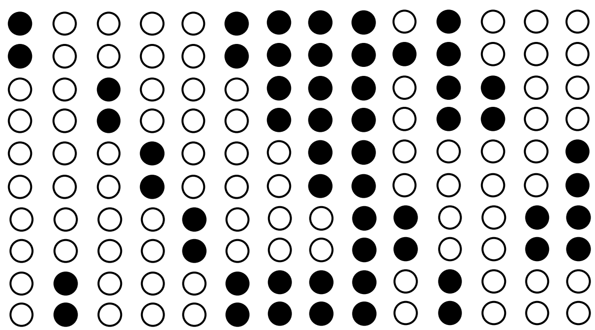
Source: Kemp, C., and J. B. Tenenbaum. "Structured Statistical Models of Inductive Reasoning. " *Psychological Review* 116, no. 1 (2009): 20-58.

Do people learn explicit structures of different forms?

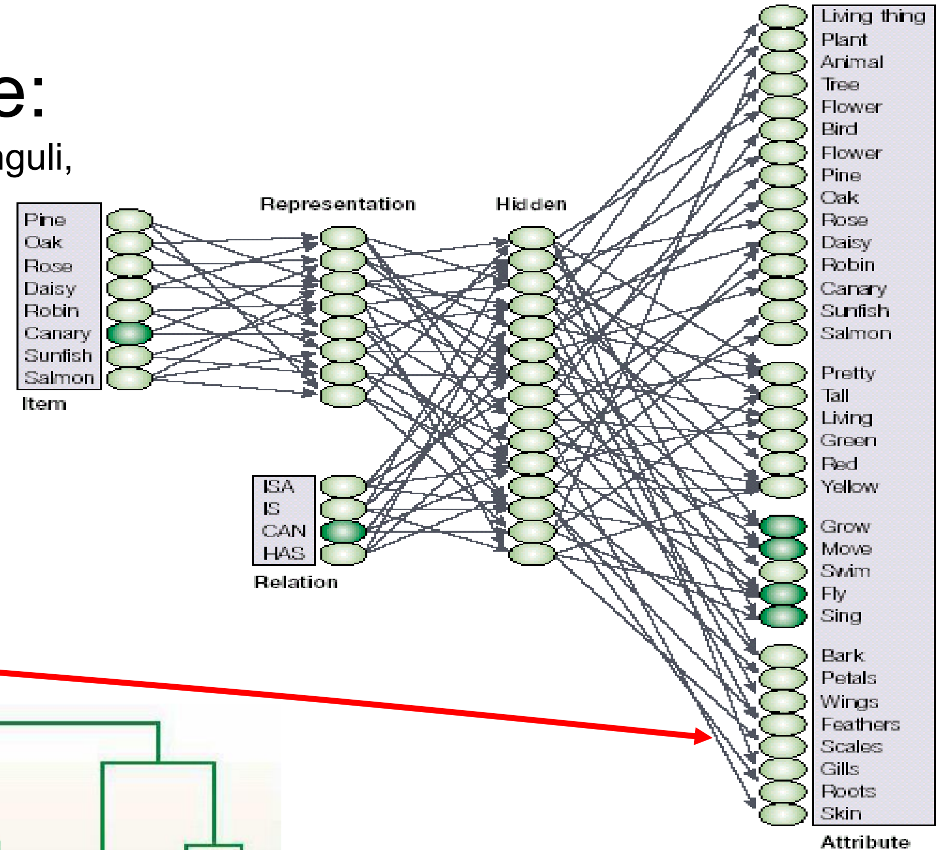
A neural-network alternative:

(Rogers and McClelland, 2004; Saxe, McClelland, Ganguli, 2013)

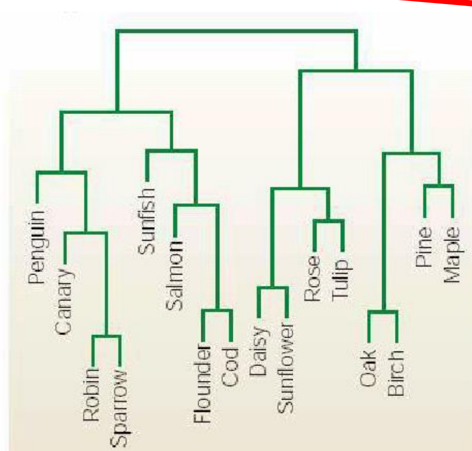
Species



Features



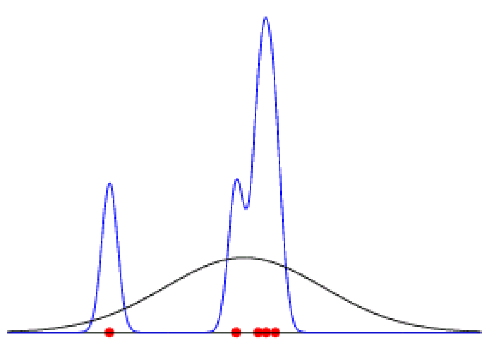
Emergent structure: clustering on hidden unit activation vectors



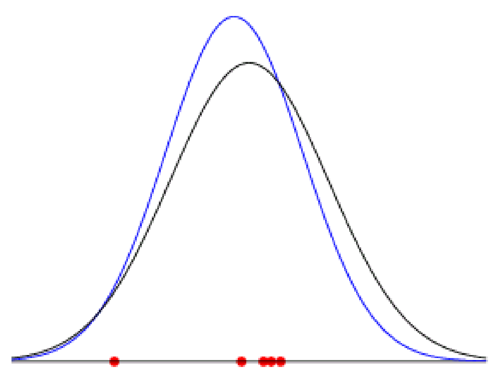
The need for inductive bias

- Learning from sparse data requires constraints or a prior on the hypothesis space.

- An analogy: Learning a smooth probability density by local interpolation (kernel density estimation).



$N = 5$



$N = 5$

Assuming an appropriately structured form for density (e.g., Gaussian) leads to better generalization from sparse data.

Beyond similarity-based induction

- Reasoning based on dimensional thresholds: (Smith et al., 1993)

Poodles can bite through wire.

German shepherds can bite through wire.

Dobermans can bite through wire.

German shepherds can bite through wire.

- Reasoning based on causal relations: (Medin et al., 2004; Coley & Shafto, 2003)

Salmon carry E. Spirus bacteria.

Grizzly bears carry E. Spirus bacteria.

Grizzly bears carry E. Spirus bacteria.

Salmon carry E. Spirus bacteria.

Different priors from different kinds of causes

Chimps have T9 hormones.

Gorillas have T9 hormones.

Taxonomic similarity

Poodles can bite through wire.

Dobermans can bite through wire.

Jaw strength

Salmon carry E. Spirus bacteria.

Grizzly bears carry E. Spirus bacteria.

Food web relations

Property type

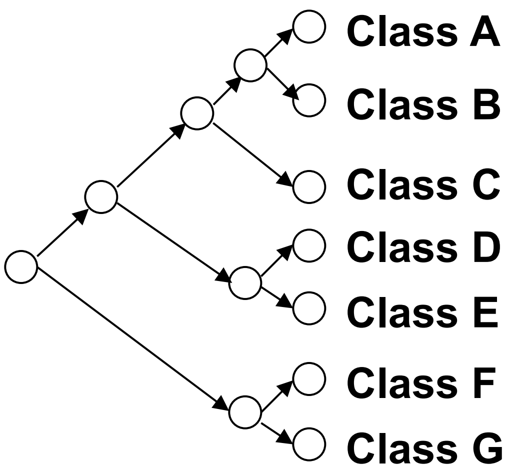
“has T9 hormones” “can bite through wire” “carry E. Spirus bacteria”

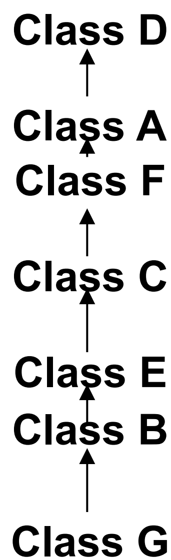
Causal Structure

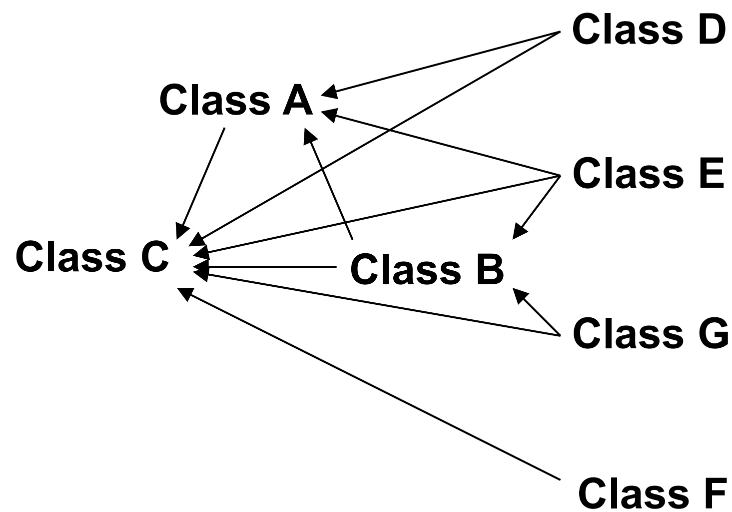
taxonomic tree
+ diffusion process

directed chain
+ drift process

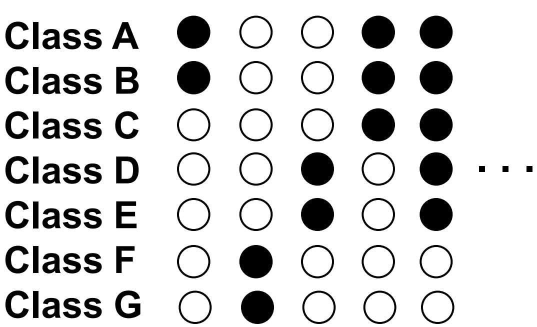
directed network
+ noisy transmission

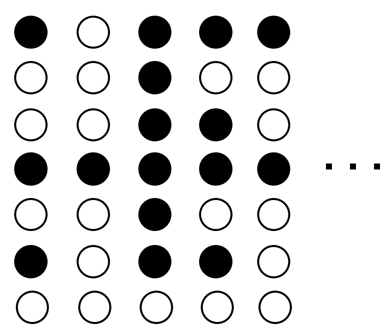


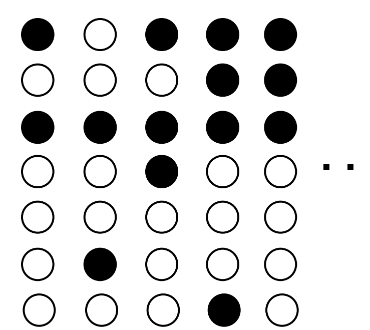




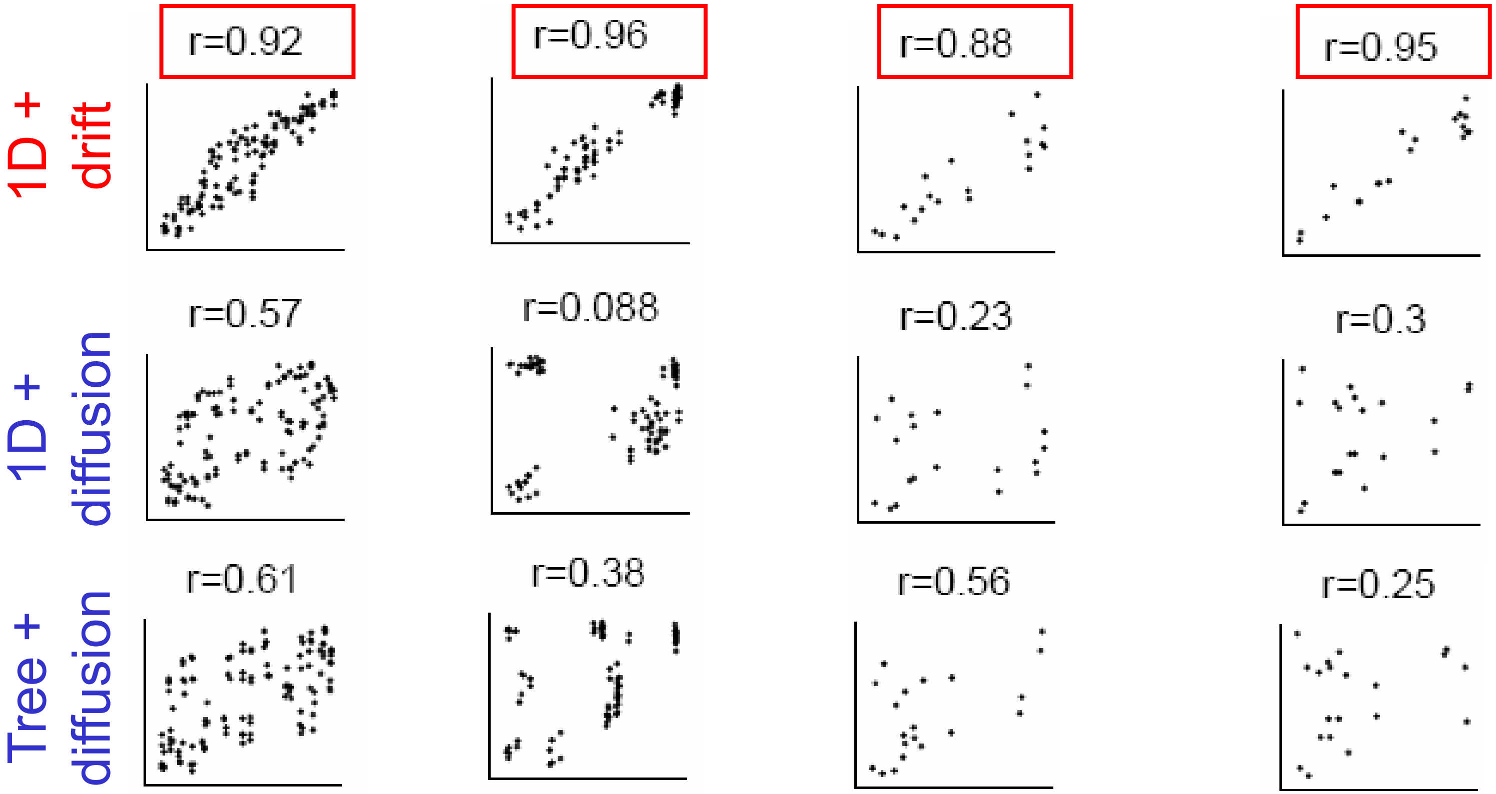
Properties

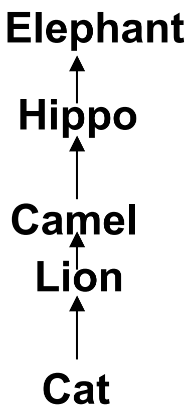






Reasoning with linear-threshold properties





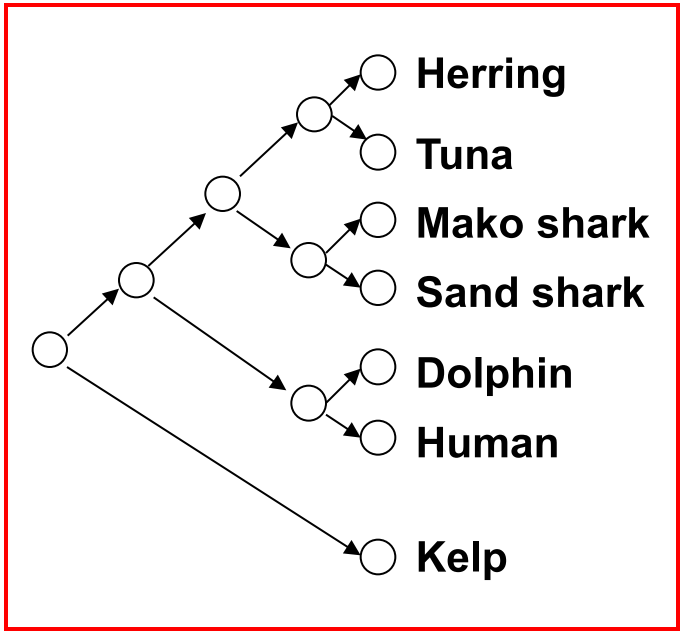
Source: Kemp, C., and J. B. Tenenbaum. "Structured Statistical Models of Inductive Reasoning. " *Psychological Review* 116, no. 1 (2009): 20-58.

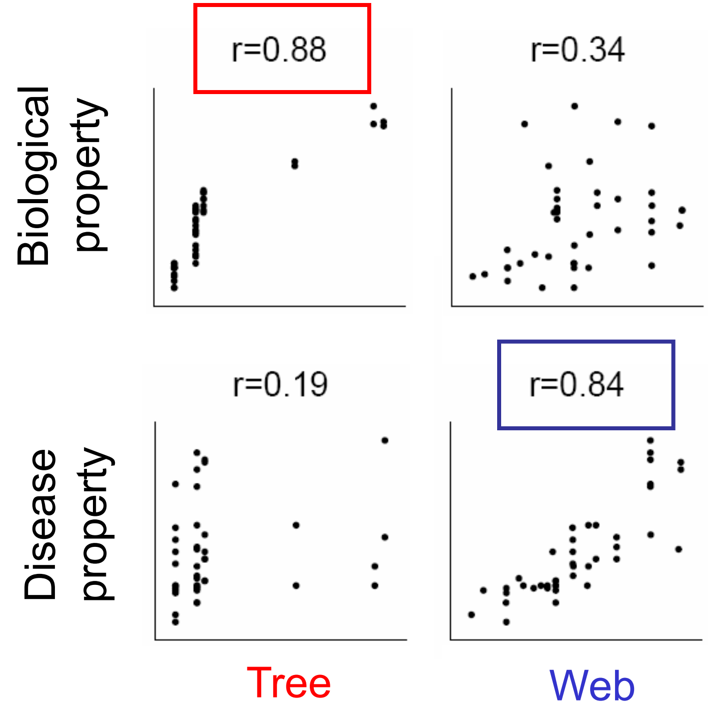
e.g., “has skin that is more resistant to penetration than most synthetic fibers”

Blok et al. 4 colleges	Blok et al. 5 colleges	Smith et al. Adapts to dark	Smith et al. Thick skin

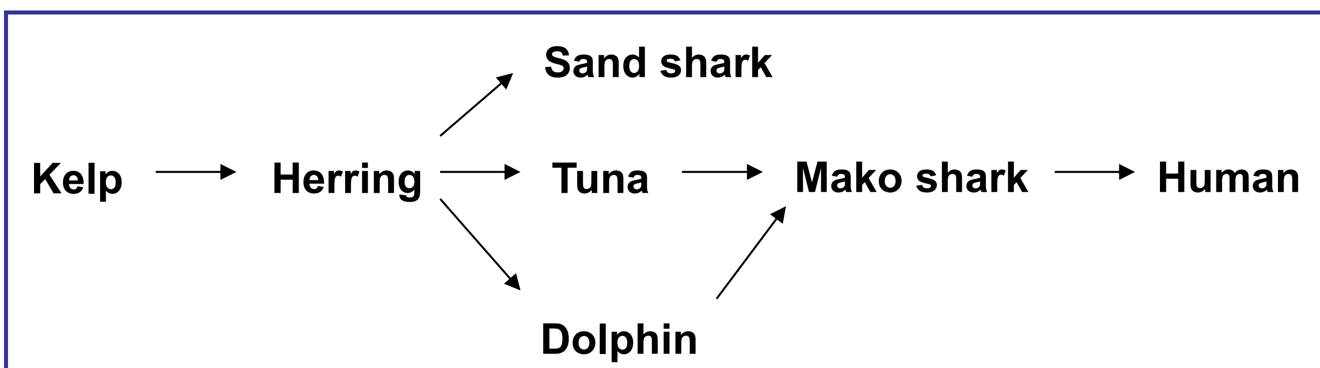
Reasoning with two property types

"Given that X has property P, how likely is it that Y does?"

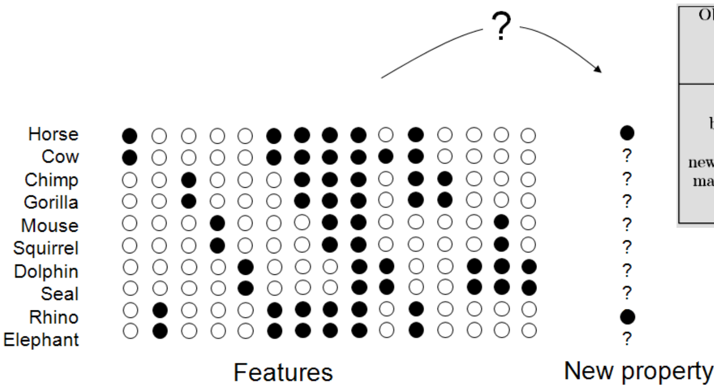




(Shafto, Kemp, Bonawitz, Coley & Tenenbaum)

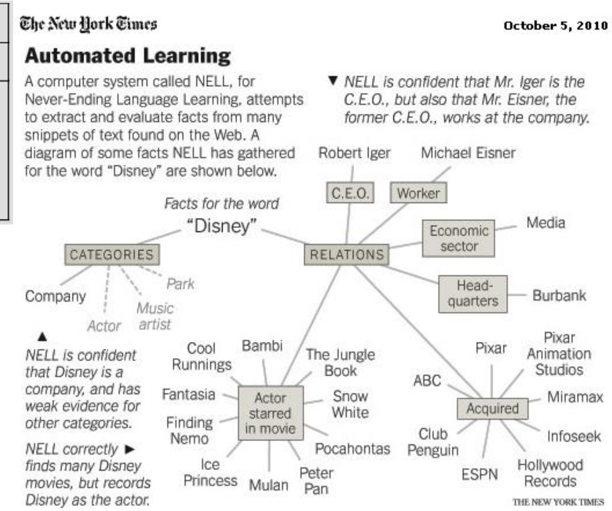


"Common-sense reasoning" as sparse matrix completion: at the heart of classical associationism, probabilistic or connectionist models of semantic cognition, contemporary machine-learning approaches to building general AI systems.

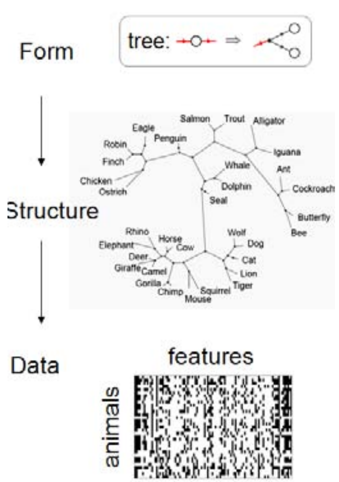


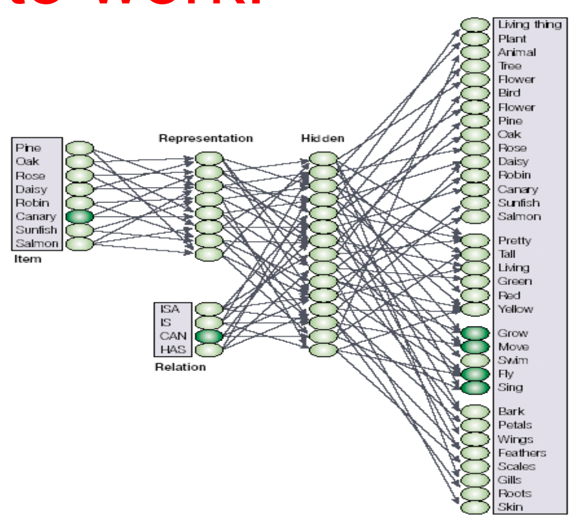
Objects	…	contains knowledge *contain knowledge*	has pages *have page*	is cold *be cold*	is for reading *be read*	…
		Properties *(with simplified form)*				
⋮		⋮	⋮	⋮	⋮	
book	…	x	x		x	…
ice	…		-	x		…
newspaper	…		x		x	…
magazine	…	x	x		x	…
⋮		⋮	⋮	⋮	⋮	

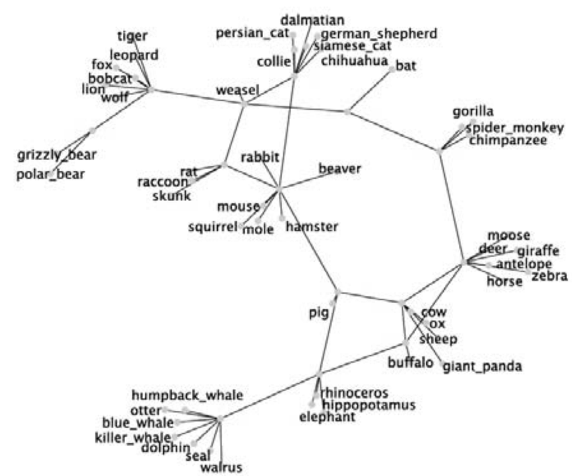
a ball may be made of metal
a ball shades you from the sun
a ball weighs less than ten tons
a balloon can be any color
a balloon can be black
a balloon can break
a ballpoint pen can be made in Singapore
a ballpoint pen is a man-made thing
a ballpoint pen is used for drawing
a ballpoint pen is used to write things
abalones are a kind of seafood
abalones are edible by humans
abalones are food
abalones are found in the ocean
abalones are good to eat



But it's not going to work.







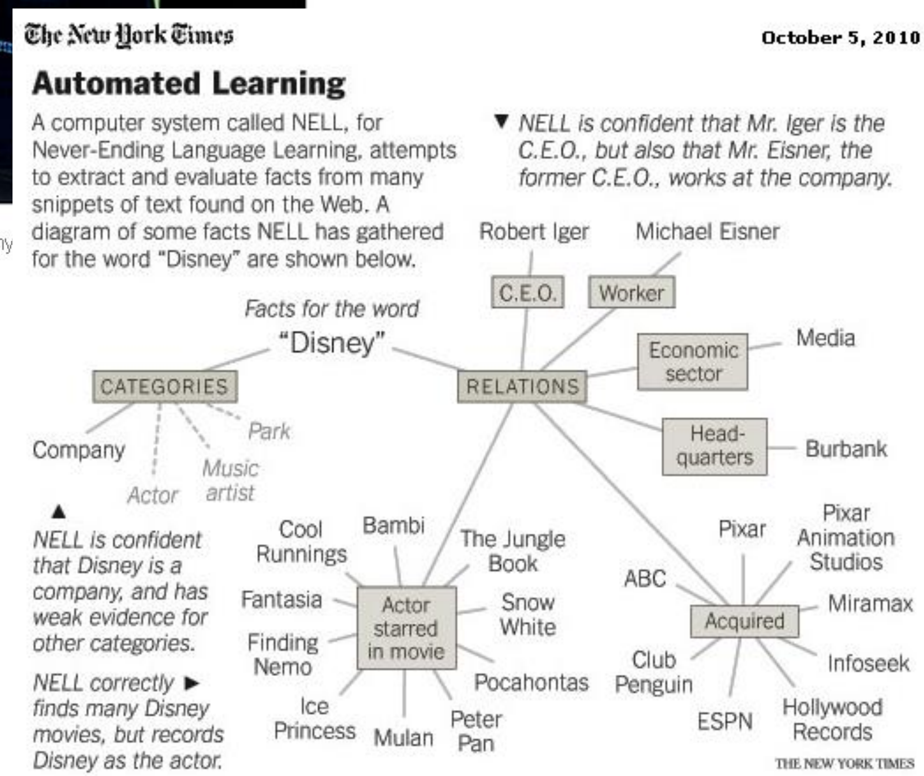
SMARTER THAN YOU THINK

Aiming to Learn as We Do, a Machine Teaches Itself

NELL'S TEAM Tom M. Mitchell, center, and, from left, William Cohen, Jayant Krishnamurthy and Bryan Kisiel.

By STEVE LOHR
Published: October 4, 2010

Never-Ending Language Learning (NELL)



The New York Times — October 5, 2010

Automated Learning

A computer system called NELL, for Never-Ending Language Learning, attempts to extract and evaluate facts from many snippets of text found on the Web. A diagram of some facts NELL has gathered for the word "Disney" are shown below.

▼ *NELL is confident that Mr. Iger is the C.E.O., but also that Mr. Eisner, the former C.E.O., works at the company.*

▲ *NELL is confident that Disney is a company, and has weak evidence for other categories.*

NELL correctly finds many Disney movies, but records Disney as the actor. ▶

THE NEW YORK TIMES

Lohr, Steve. "Aiming to Learn as We Do, a Machine Teaches Itself." The New York Times, October 4, 2010.

Never-Ending Language Learning (NELL)

Text excerpt removed due to copyright restrictions.
See Lohr, Steve. "Aiming to Learn as We Do, a Machine Teaches Itself."
The New York Times, October 4, 2010.

Engineering common sense: what, and how?

The roots of common sense

Images of children playing removed due to copyright restrictions. Please see the video or http://providencechildrensmuseum.blogspot.com/2012/05/loosen-up.html.

Several other similar photos removed due to copyright restrictions. See video.

Engineering common sense: what, and how?

What: The “common sense core”

Human thought is structured around a basic understanding of physical objects, intentional agents, and their interactions – intuitive **physics** (forces, masses…) and **psychology** (desires, beliefs, plans…) [Spelke, Baillargeon, Gergeley, Csibra, Carey, Kanwisher, Saxe, Dehaene, Tomasello…]

Develops early in infancy

Shared to some extent with other species

Enriched and extended massively in humans

The targets of understanding visual scenes, language, and action planning.

How can these internal models be realized computationally?

How can they be studied rigorously in behavior?

How are they instantiated in neural circuits?

How are they built, through evolution, development and learning?

The development of object knowledge in infancy

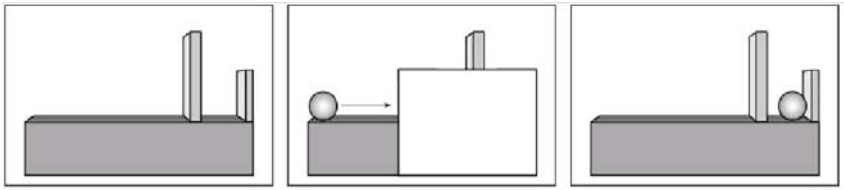
http://www.bbc.com/news/technology-19637175

https://www.youtube.com/watch?v=0jaxzURLylc

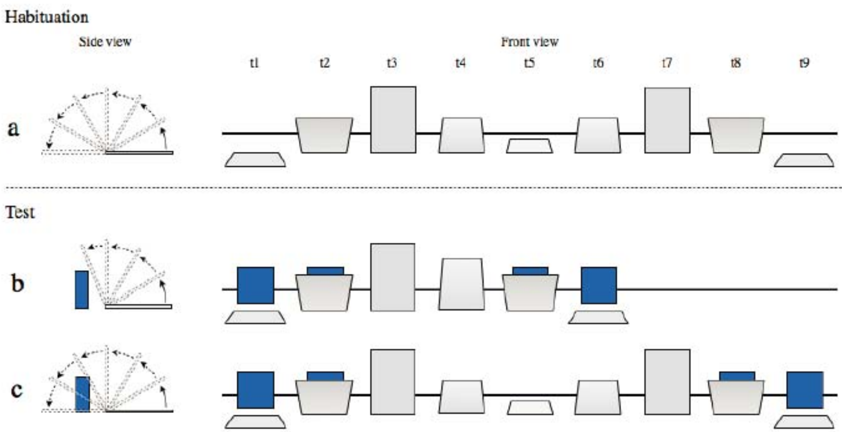
The development of object knowledge in infancy

(Baillargeon, Spelke et al)

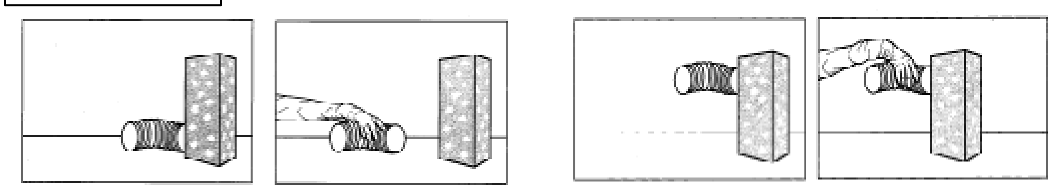
2-3 months



4-5 months



8 months



Violation detected at each stage

3 months

Initial Concept:
Contact/No contact

5 months

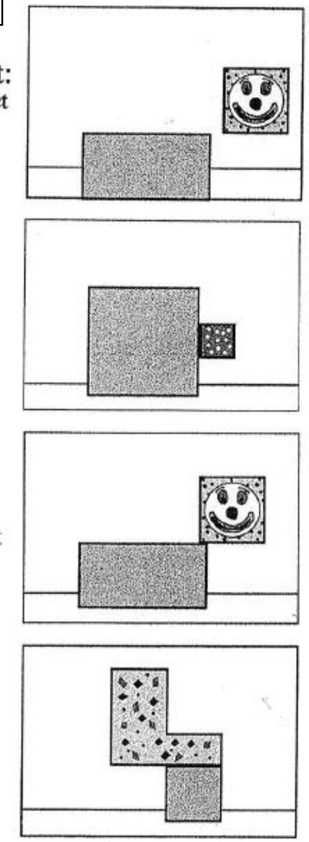
Variable:
Type of contact

6.5 months

Variable:
Amount of contact

12 months

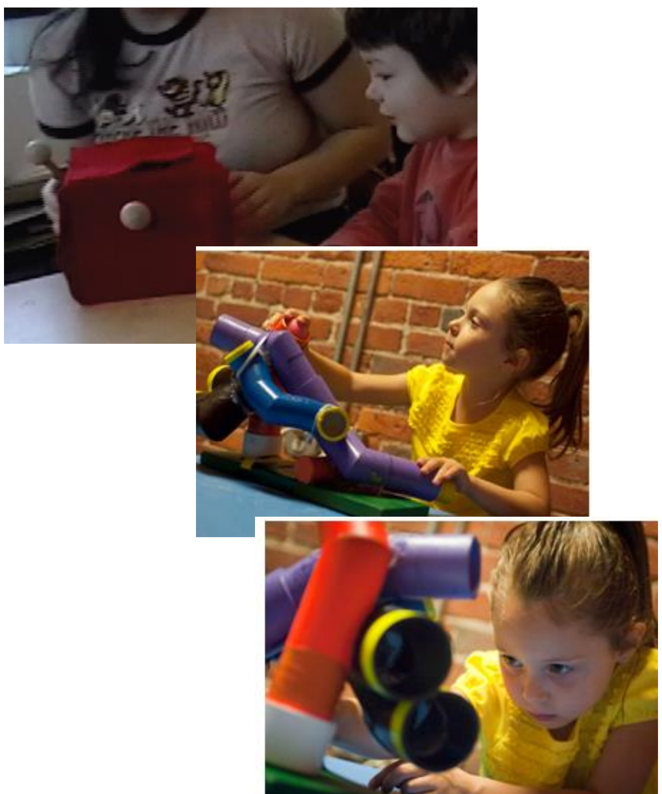
Variable:
Shape of the box

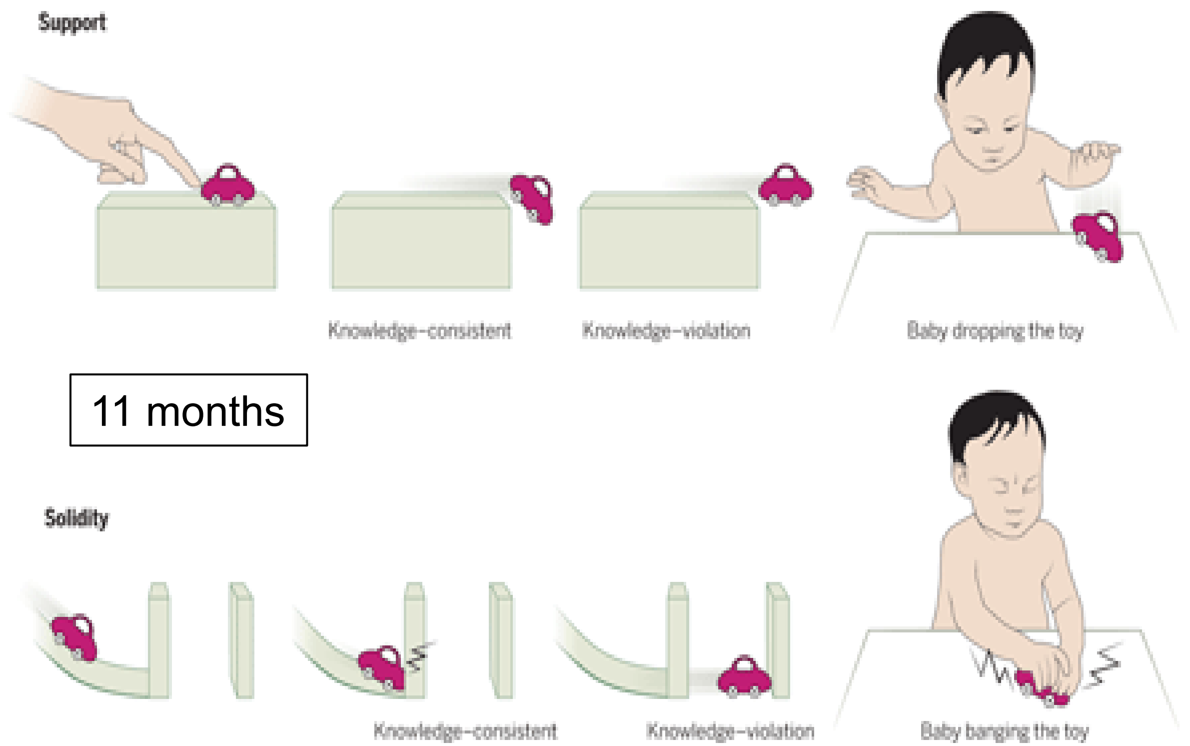


Source: Baillargeon, Renée. "Infants' understanding of the physical world." Journal of the Neurological Sciences 143, no. 1-2 (1996): 199-199.

How knowledge grows

Learning and abstraction as theory-building (or, the "child as scientist", not "data analyst"). Knowledge grows through hypothesis- and explanation-driven interpretations of sparse data, causal learning, learning theories, learning compositional abstractions, learning to learn, exploratory learning, social learning. [Carey, Karmiloff-Smith, Gopnik, Schulz, Feigenson, …]





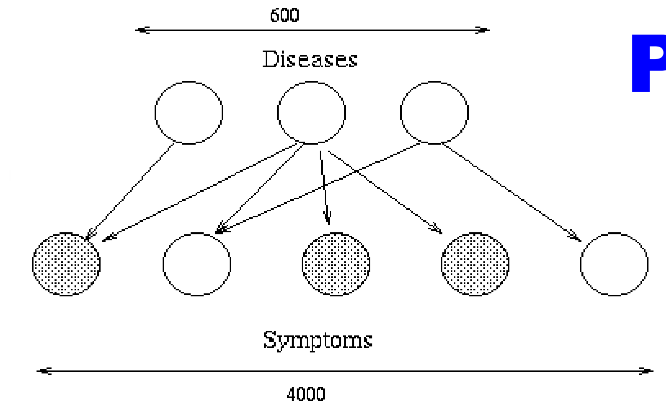
Engineering common sense: what, and how?

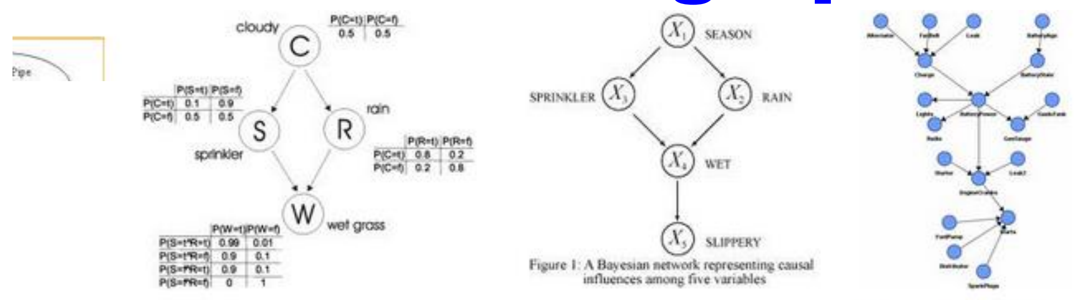
What: The "common sense core"

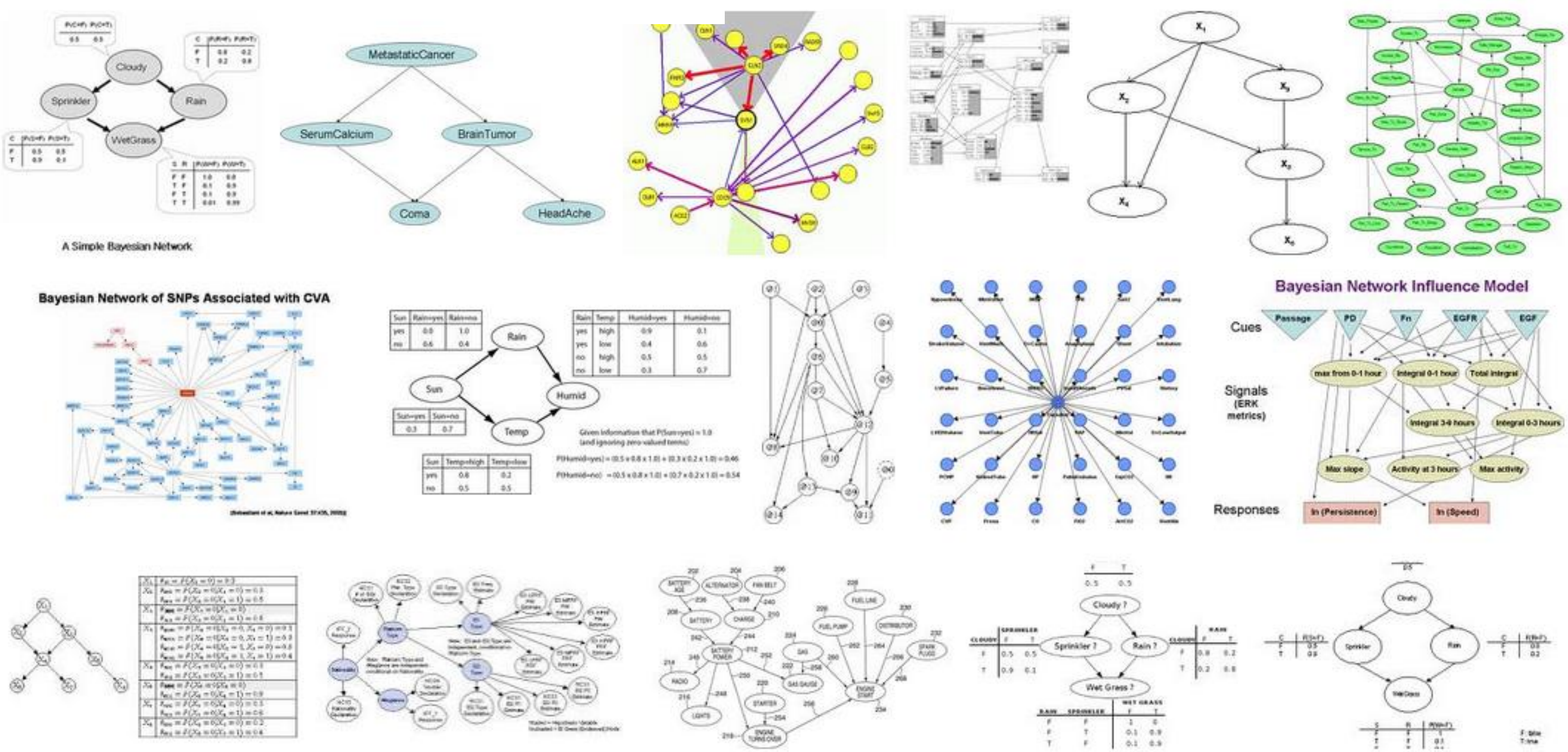
How: A modeling engine built on probabilistic programs

(Goodman, Mansinghka, Roy, Freer, …) **[See: probmods.org]**

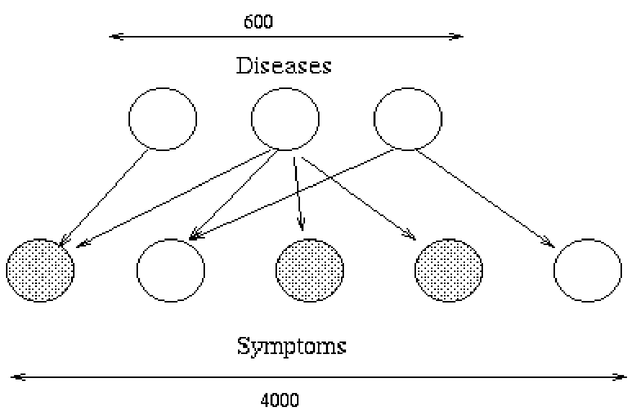
Bayesian networks: Probabilities on graphs

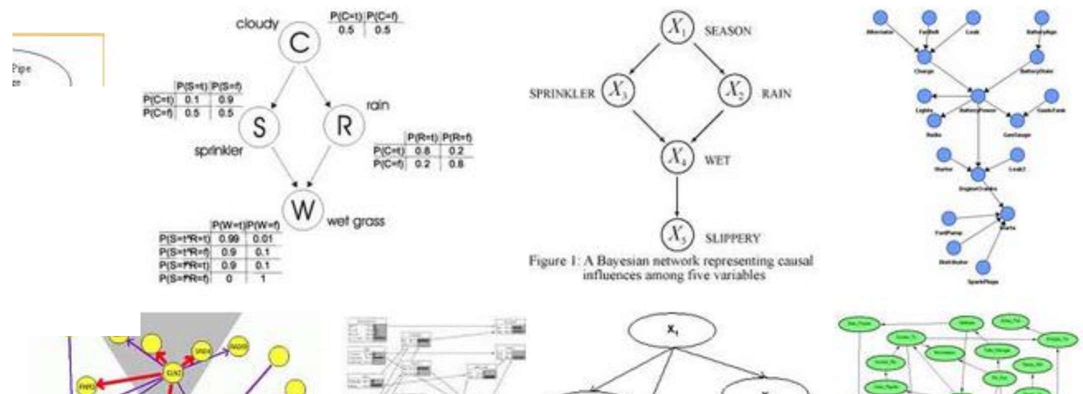


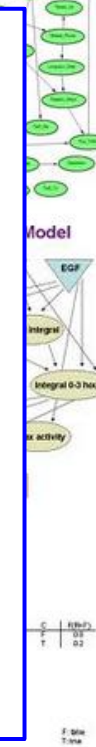




Bayesian networks: Probabilities on graphs







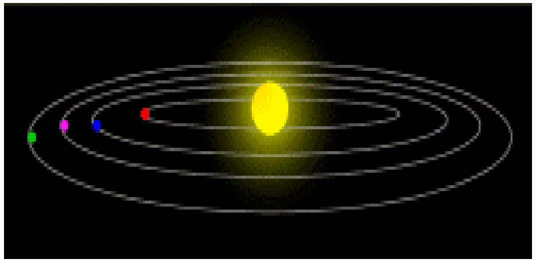
Bayes nets (and probabilistic graphical models more generally) brought a potent combination to AI and many fields:

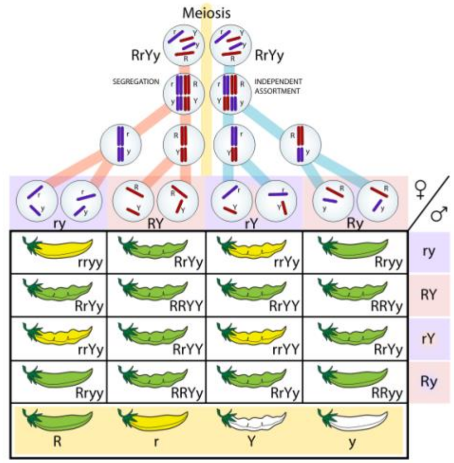
1. General-purpose languages for representing the structure of the world.

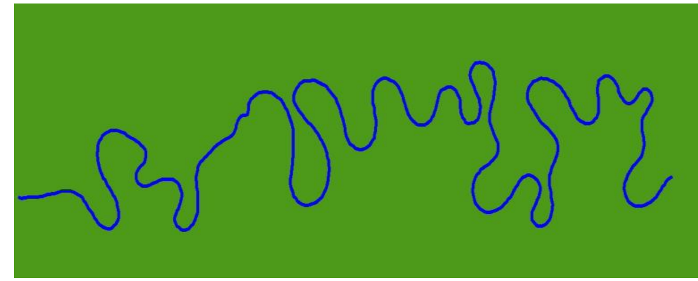
2. General-purpose algorithms for inference and decision under uncertainty.

But they're not enough.

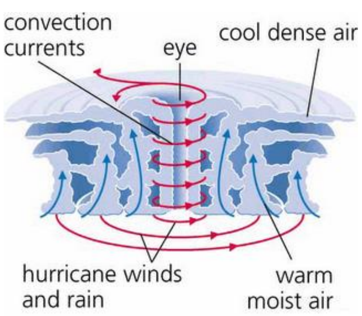
Modeling the world with programs

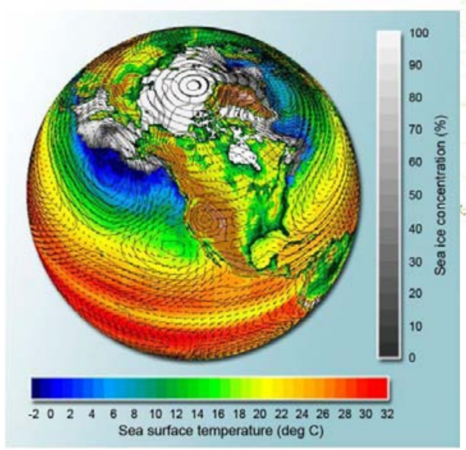


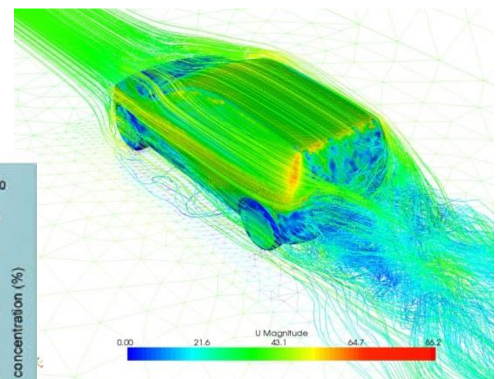


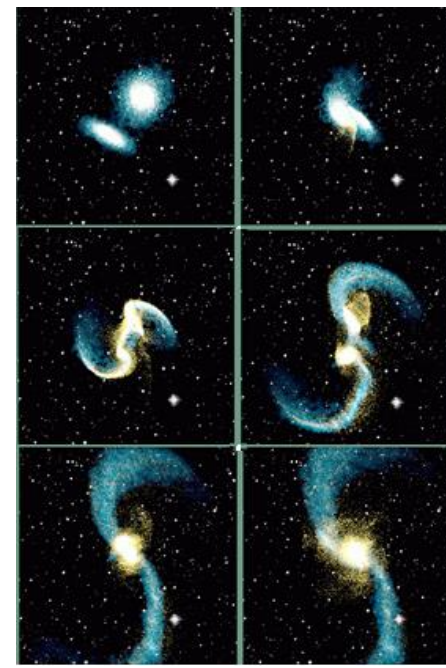


$$F = \frac{GMm}{r^2}$$

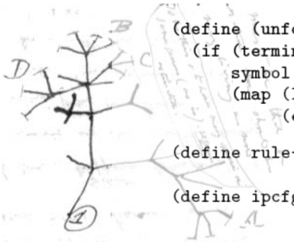




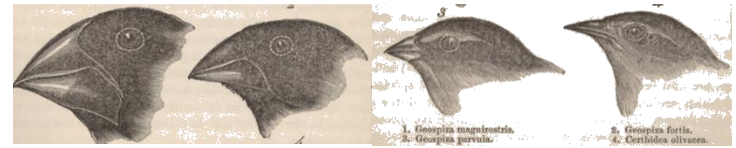




Model building as program learning

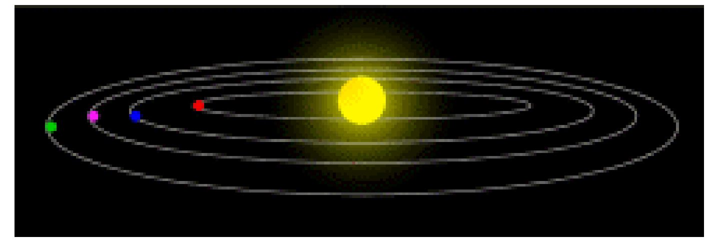


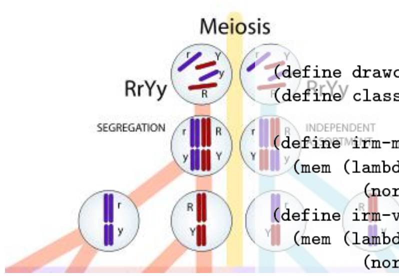
```
(define (unfold expander symbol)
  (if (terminal? symbol)
      symbol
      (map (lambda (x) (unfold expander x))
           (expander symbol) )))

(define rule-type (mem (lambda symbol)
                       (if (flip) 'terminal 'binary-production))
(define ipcfg-expander (DPmem 1.0 (lambda (symbol)
                                     (if (eq? (rule-type symbol) 'terminal)
                                         (multinomial terms term-probs)
```



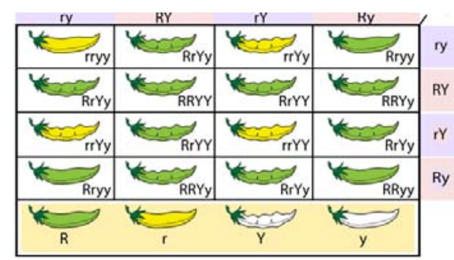
$$F = \frac{GMm}{r^2}$$

```
(define (unfold expander symbol)
  (if (terminal? symbol)
      symbol
      (map (lambda (x) (unfold expander x))
           (expander symbol) )))

(define get-symbol (DPmem 1.0 gensym))
(define get-observation-model (mem (lambda (symbol) (make-die))))
(define ihmm-transition (DPmem 1.0 (lambda (state)
                                      (if (flip) 'stop (get-symbol))
```






```
(define drawclass (DPmem 1.0 gensym))
(define class (mem (lambda (obj) (drawclass))))

(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 ))))
```



$$p(Model \mid Data) = \frac{p(Data \mid Model)\, p(Model)}{p(Data)}$$

Probabilistic programs: Church, probmods.org

(Goodman, Mansinghka, Roy, Bonawitz & Tenenbaum 2008; Goodman & Tenenbaum, 2014)

Probabilistic Models of Cognition

All chapters | Login

Example: Reasoning about the Tug of War

Returning to the earlier example of a series of tug-of-war matches, we can use query to ask a variety of different questions. For instance, how likely is it that Bob is strong, given that he's been in a series of winning teams? (Note that we have written the `winner` function slightly differently here, to return the labels `'team1` or `'team2` rather than the list of team members. This makes for more compact conditioning statements.)

```
(define samples
  (mh-query 1000 10

    (define strength (mem (lambda (person) (gaussian 0 1))))

    (define lazy (lambda (person) (flip (/ 1 3))))

    (define (total-pulling team)
      (sum
         (map
          (lambda (person) (if (lazy person) (/ (strength person) 2) (strength person)))
          team)))

    (define (winner team1 team2)
      (if (> (total-pulling team1) (total-pulling team2)) 'team1 'team2))

    (strength 'bob)

    (and (eq? 'team1 (winner '(bob mary) '(tom sue)))
         (eq? 'team1 (winner '(bob sue) '(tom jim))))))

(display (list "Expected strength: " (mean samples)))
(density samples "Bob strength" true)
```

From https://probmods.org

Probabilistic programs: Church, probmods.org

(Goodman, Mansinghka, Roy, Bonawitz & Tenenbaum 2008; Goodman & Tenenbaum, 2014)

Probabilistic Models of Cognition

All chapters Login

Example: Reasoning about the Tug of War

Returning to the earlier example of a series of tug-of-war matches, we can use query to ask a variety of different questions. For i
that we have w
than the list of t

```
(define sample
 (mh-query 10

  (define st

  (define la

  (define (t
   (sum
    (map
     (lam
     team

  (define (w
   (if (> (

  (strength

  (and (eq?
       (eq?

(display (list
(density sampl
```

Probabilistic Models of Cognition

All chapters Login

Example: Causal Inference in Medical Diagnosis

This classic Bayesian inference task is a special case of conditioning. Kahneman and Tversky, and Gigerenzer and colleagues, have studied how people make simple judgments like the following:

> The probability of breast cancer is 1% for a woman at 40 who participates in a routine screening. If a woman has breast cancer, the probability is 80% that she will have a positive mammography. If a woman does not have breast cancer, the probability is 9.6% that she will also have a positive mammography. A woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer?

What is your intuition? Many people without training in statistical inference judge the probability to be rather high, typically between 0.7 and 0.9. The correct answer is much lower, less than 0.1, as we can see by evaluating this Church query:

```
(define samples
 (mh-query 100 100
   (define breast-cancer (flip 0.01))

   (define positive-mammogram (if breast-cancer (flip 0.8) (flip 0.096)))

   breast-cancer

   positive-mammogram
 )
)
(hist samples "breast cancer")
```

Tversky & Kahneman (1974) named this kind of judgment error *base rate neglect*, because in order to make the correct judgment, one must realize that the key contrast is between the *base rate* of the disease, 0.01 in this case, and the *false alarm rate* or probability of a positive mammogram given no breast cancer, 0.096. The false alarm rate

Probabilistic programs: Church, probmods.org

(Goodman, Mansinghka, Roy, Bonawitz & Tenenbaum 2008; Goodman & Tenenbaum, 2014)

Probabilistic Models of Cognition

All chapters Login

Example: Inverse intuitive physics

We previously saw how a generative model of physics—a noisy, intuitive version of Newtonian mechanics—could be used to make judgements about the final state of physical worlds from initial conditions. We showed how this forward simulation could be used to model judgements about stability. We can also use a physics model to reason backward: from final to initial states.

Imagine that we drop a block from a random position at the top of a world with two fixed obstacles:

```
;set up some bins on a floor:
(define (bins xmin xmax width)
  (if (< xmax (+ xmin width))
      ;the floor:
      '( (("rect" #t (400 10)) (175 500)) )
      ;add a bin, keep going:
      (pair (list '("rect" #t (1 10)) (list xmin 490))
            (bins (+ xmin width) xmax width))))

;make a world with two fixed circles and bins:
(define world (pair '(("circle" #t (60)) (60 200))
                    (pair '(("circle" #t (30)) (300 300))
                          (bins -1000 1000 25))))

;make a random block at the top:
(define (random-block) (list (list "circle" #f '(10))
                             (list (uniform 0 worldWidth) 0)))

;add a random block to world, then animate:
(animatePhysics 1000 (pair (random-block) world))
```

Probabilistic Models of

Example: Reasoning about t

Returning to the earlier example of a seri
questions. For i
that we have w
than the list of t

```
(define sample
  (mh-query 10

    (define st

    (define la

    (define (t
      (sum
        (map
          (lam
          team

    (define (w
      (if (> (

    (strength

    (and (eq?
         (eq?

(display (list
(density sampl
```

Probabilistic

Example: C

This classic Bayesian i
colleagues, have studie

> The probability
> woman has bre
> woman does n
> mammography.
> What is the prol

What is your intuitio
high, typically betwee
this Church query:

```
(define samples
 (mh-query 100 100
  (define breast-cancer (flip 0.01))

  (define positive-mammogram (if breast-cancer (flip 0.8) (flip 0.096)))

  breast-cancer

  positive-mammogram
 )
)
(hist samples "breast cancer")
```

Tversky & Kahneman (1974) named this kind of judgment error *base rate neglect*, because in order to make the correct judgment, one must realize that the key contrast is between the *base rate* of the disease, 0.01 in this case, and the *false alarm rate* or probability of a positive mammogram given no breast cancer, 0.096. The false alarm rate

Probabilistic programs: Church, probmods.org

(Goodman, Mansinghka, Roy, Bonawitz & Tenenbaum 2008; Goodman & Tenenbaum, 2014)

Probabilistic Models of Cognition

All chapters Login

Probabilistic Models of

Example: Reasoning about t

Returning to the earlier example of a seri
questions. For i
that we have w
than the list of t

```
(define sample
  (mh-query 10
    (define st
    (define la
    (define (t
      (sum
        (map
         (lam
         team
    (define (w
      (if (> (
    (strength
    (and (eq?
         (eq?
(display (list
(density sampl
```

Probabilistic

Example: C

This classic Bayesian i
colleagues, have studie

The probability
woman has bre
woman does n
mammography.
What is the prol

What is your intuitio
high, typically betwee
this Church query:

```
(define samples
 (mh-query 100 100
   (define breast-cancer (flip 0.01))

   (define positive-mammogram (if breast-cancer (

   breast-cancer

   positive-mammogram
 )
)
(hist samples "breast cancer")
```

Tversky & Kahneman (1974) named this kind of ju
correct judgment, one must realize that the key con
and the *false alarm rate* or probability of a positive ma

Example: Inverse intuitive physics

We previously saw h
be used to make judg
forward simulation c
backward: from final

Imagine that we drop

```
;set up some bins o
(define (bins xmin
  (if (< xmax (+ xm
      ;the floor:
      '( (("rect" #
      ;add a bin, k
      (pair (list '
            (bins (

;make a world with
(define world (pair

;make a random bloc
(define (random-blo

;add a random block
(animatePhysics 100
```

Probabilistic Models of Cognition

Social Cognition

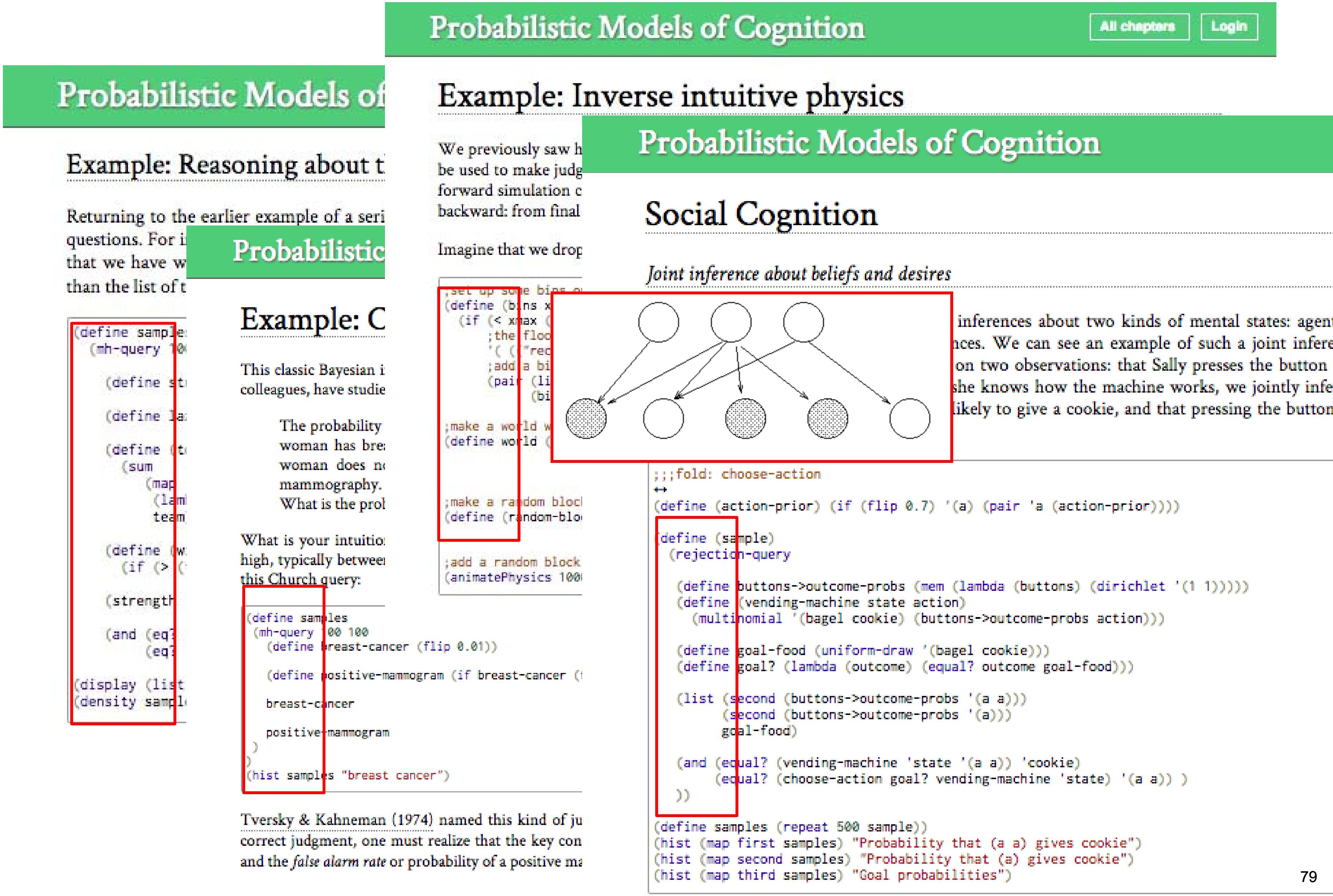
Joint inference about beliefs and desires

In social cognition, we often make joint inferences about two kinds of mental states: agents
world and their desires, goals or preferences. We can see an example of such a joint inferen
machine scenario. Suppose we condition on two observations: that Sally presses the button tw
results in a cookie. Then, assuming that she knows how the machine works, we jointly infer
cookie, that pressing the button twice is likely to give a cookie, and that pressing the button o
give a cookie.

```
;;;fold: choose-action
↔
(define (action-prior) (if (flip 0.7) '(a) (pair 'a (action-prior))))

(define (sample)
  (rejection-query

   (define buttons->outcome-probs (mem (lambda (buttons) (dirichlet '(1 1)))))
   (define (vending-machine state action)
     (multinomial '(bagel cookie) (buttons->outcome-probs action)))

   (define goal-food (uniform-draw '(bagel cookie)))
   (define goal? (lambda (outcome) (equal? outcome goal-food)))

   (list (second (buttons->outcome-probs '(a a)))
         (second (buttons->outcome-probs '(a)))
         goal-food)

   (and (equal? (vending-machine 'state '(a a)) 'cookie)
        (equal? (choose-action goal? vending-machine 'state) '(a a)) )
   ))

(define samples (repeat 500 sample))
(hist (map first samples) "Probability that (a a) gives cookie")
(hist (map second samples) "Probability that (a) gives cookie")
(hist (map third samples) "Goal probabilities")
```

Probabilistic programs: Church, probmods.org

(Goodman, Mansinghka, Roy, Bonawitz & Tenenbaum 2008; Goodman & Tenenbaum, 2014)



Engineering common sense: what, and how?

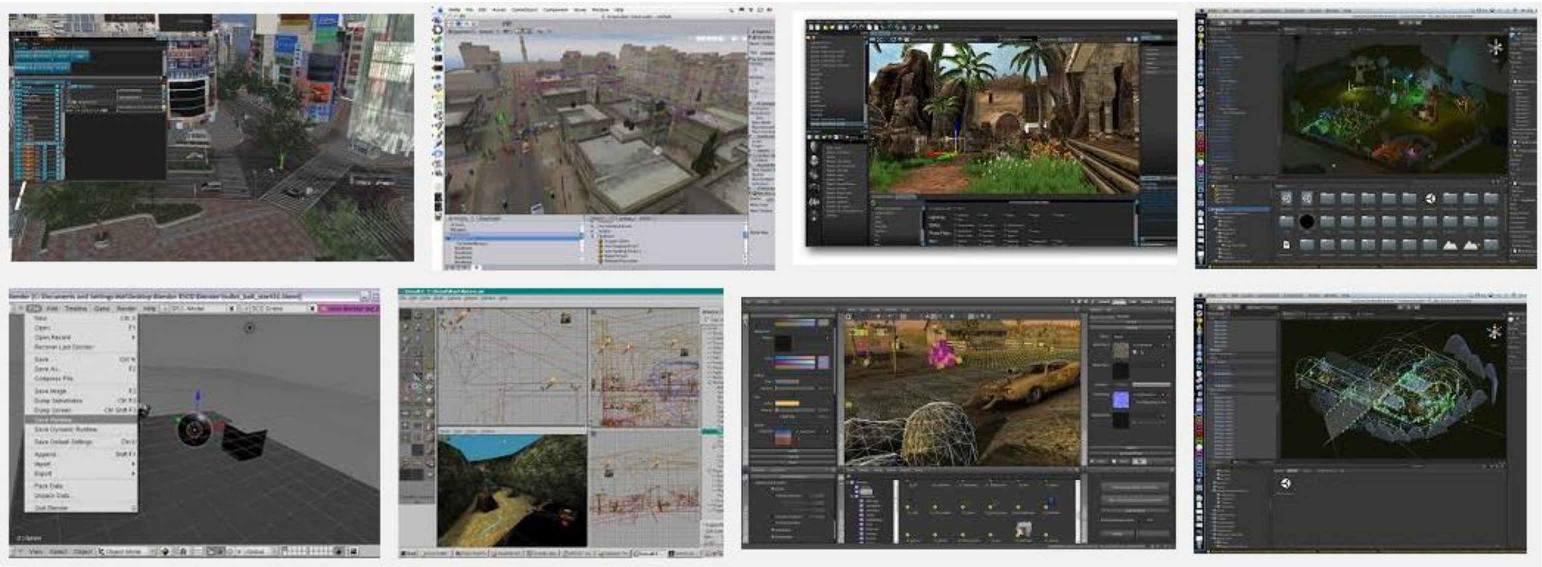
What: The "common sense core"

How: A modeling engine built on probabilistic programs

(Goodman, Mansinghka, Roy, Freer, …) **[See: probmods.org]**

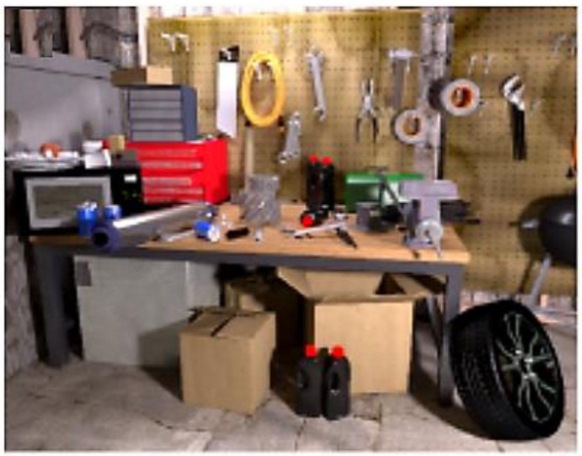
Representations: the "game engine in your head"

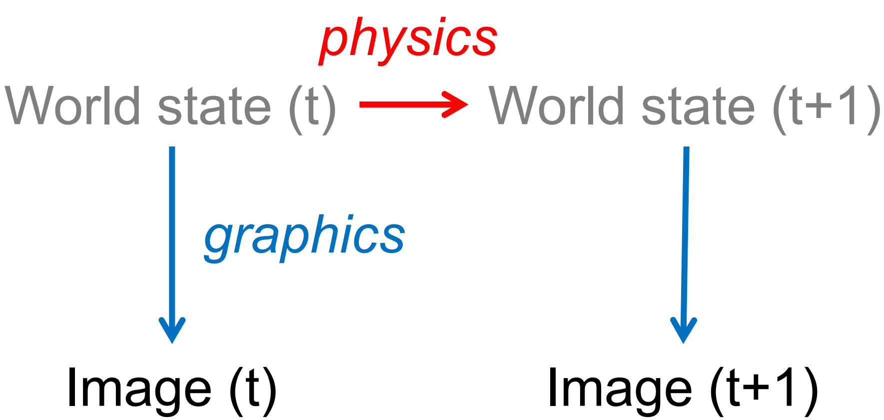
(graphics engine, physics engine, planning engine)

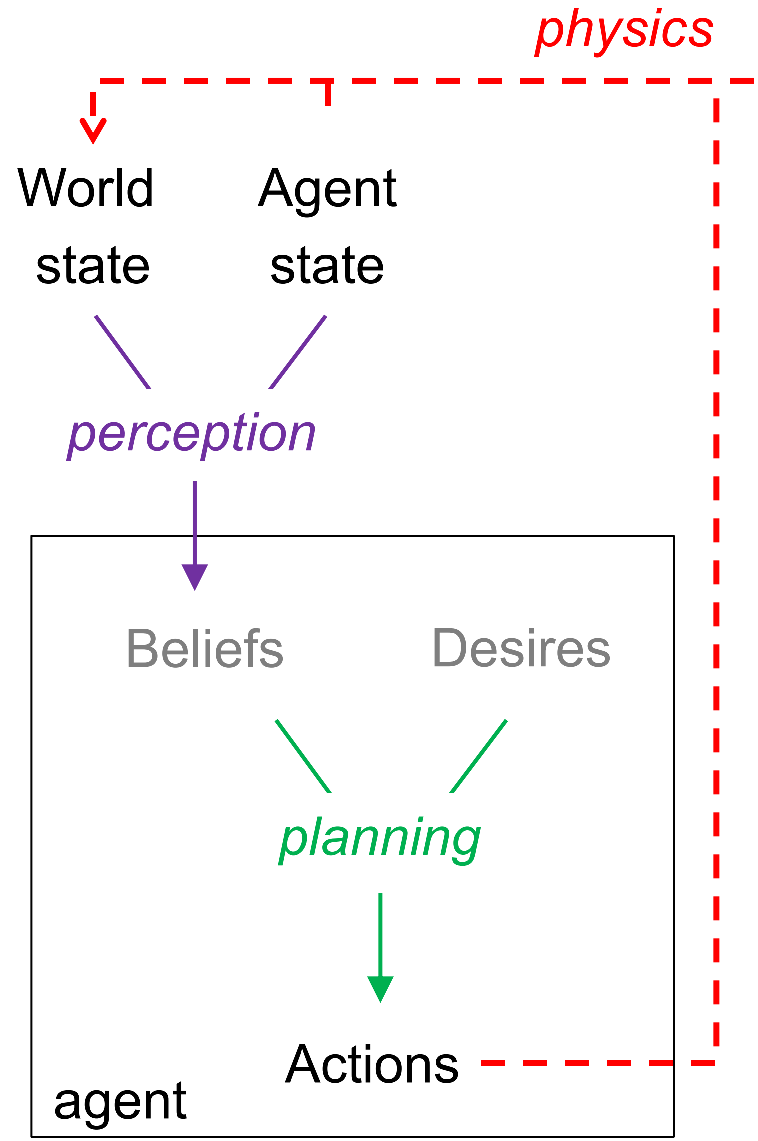


Probabilistic programs

Photo of young students in crosswalk, with crossing guard, removed due to copyright restrictions.







Engineering common sense: what, and how?

What: The "common sense core"

How: A modeling engine built on probabilistic programs

(Goodman, Mansinghka, Roy, Freer, …) **[See: probmods.org]**

Representations: the "game engine in your head"

(graphics engine, physics engine, planning engine)

Algorithms: "inference programs"

Really fast: Bottom-up guesses based on cached experience. (Perception)

Fast: Forward simulation (Prediction, imagination, top-down percepts)

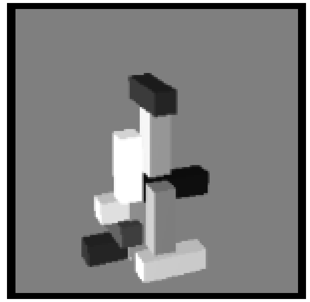
Slower: Sampling by reverse simulation. (Thinking, reasoning)

Slow (& really slow, & really really slow): Stochastic search (Learning, development, evolution)

The intuitive physics engine

(Battaglia, Hamrick, Tenenbaum, PNAS 2013)

Vision as inverse graphics

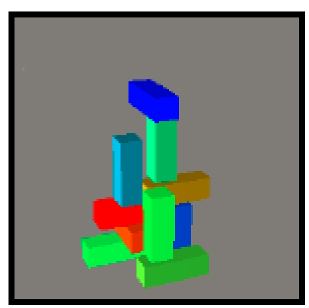


World state (t)

Prob. approx. rendering

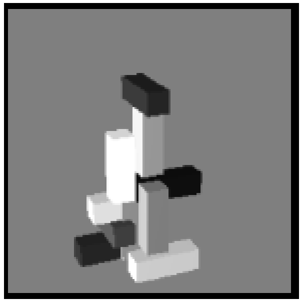
↓

Image (t)



Source: Battaglia, P. W., et al. "Simulation as an engine of physical scene understanding." PNAS 110 no. 45 (2013): 18327-18332

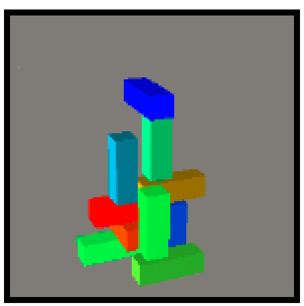
Vision as inverse graphics

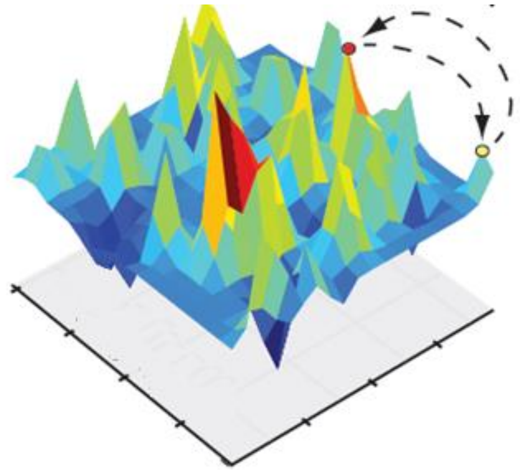


Scene

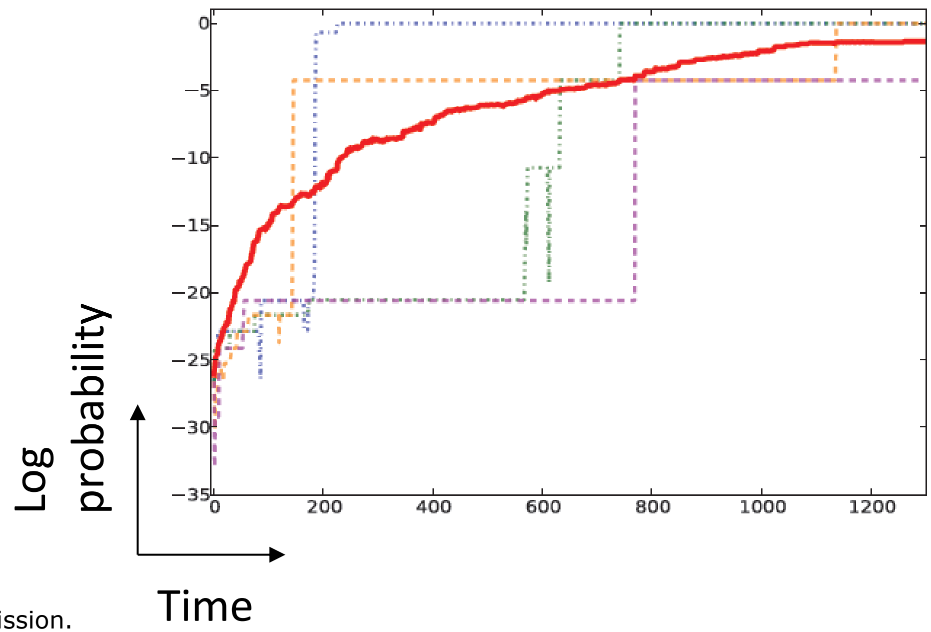
Prob. approx. rendering

Image



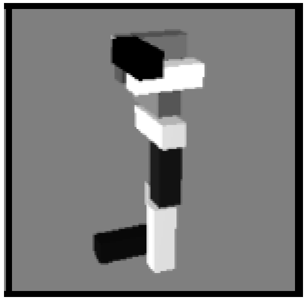


Markov Chain
Monte Carlo (MCMC):
Metropolis-Hastings



Source: Battaglia, P. W., et al. "Simulation as an engine of physical scene understanding." PNAS 110 no. 45 (2013): 18327-18332

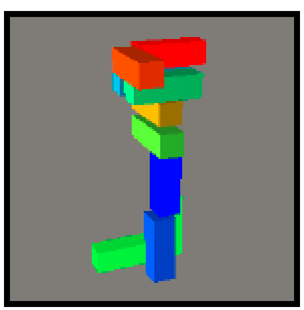
Vision as inverse graphics

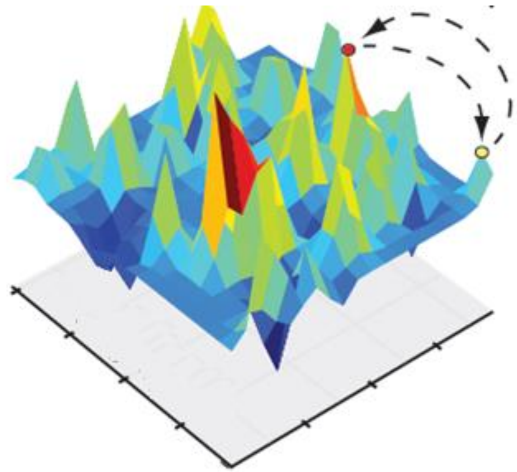


Scene

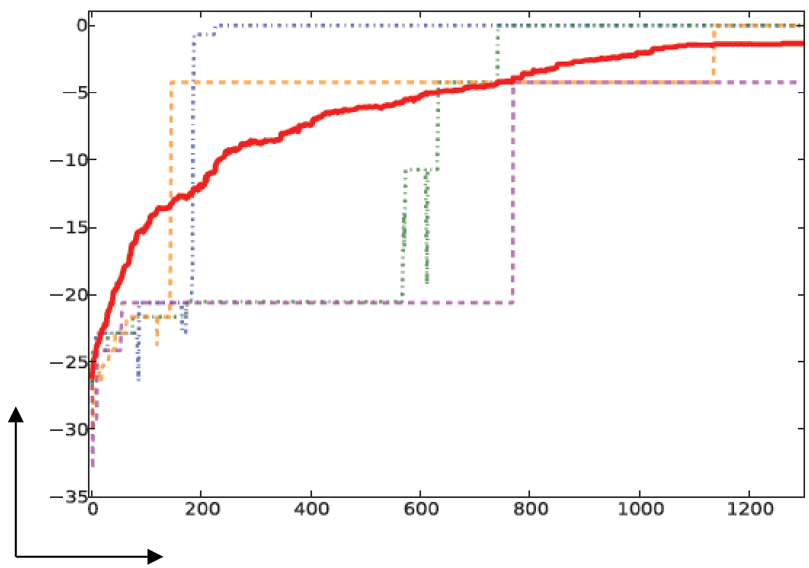
Prob. approx. rendering

Image





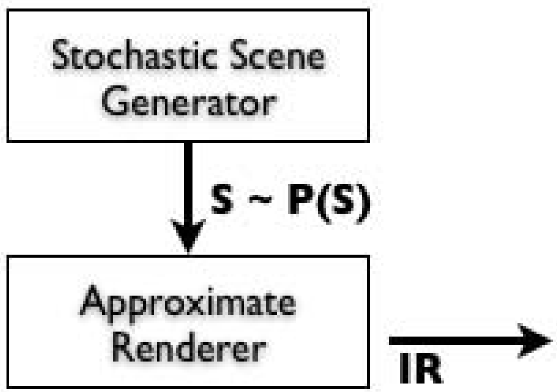
Markov Chain
Monte Carlo (MCMC):
Metropolis-Hastings



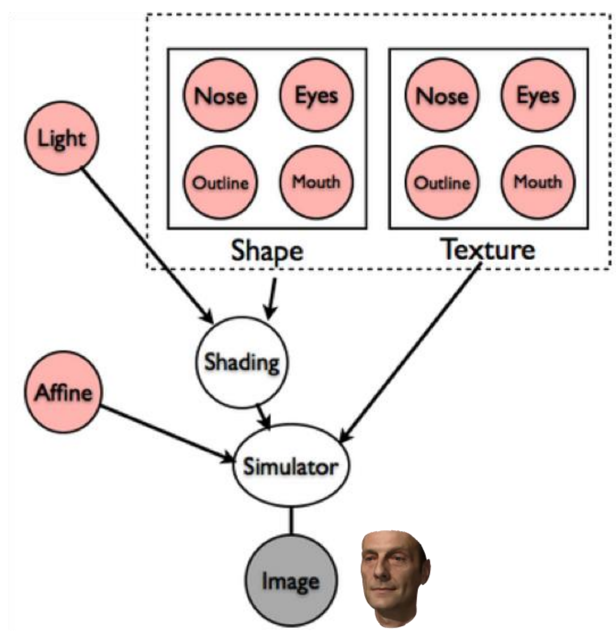
Log probability

Time

Architecture (Kulkarni et al., CVPR 2015)



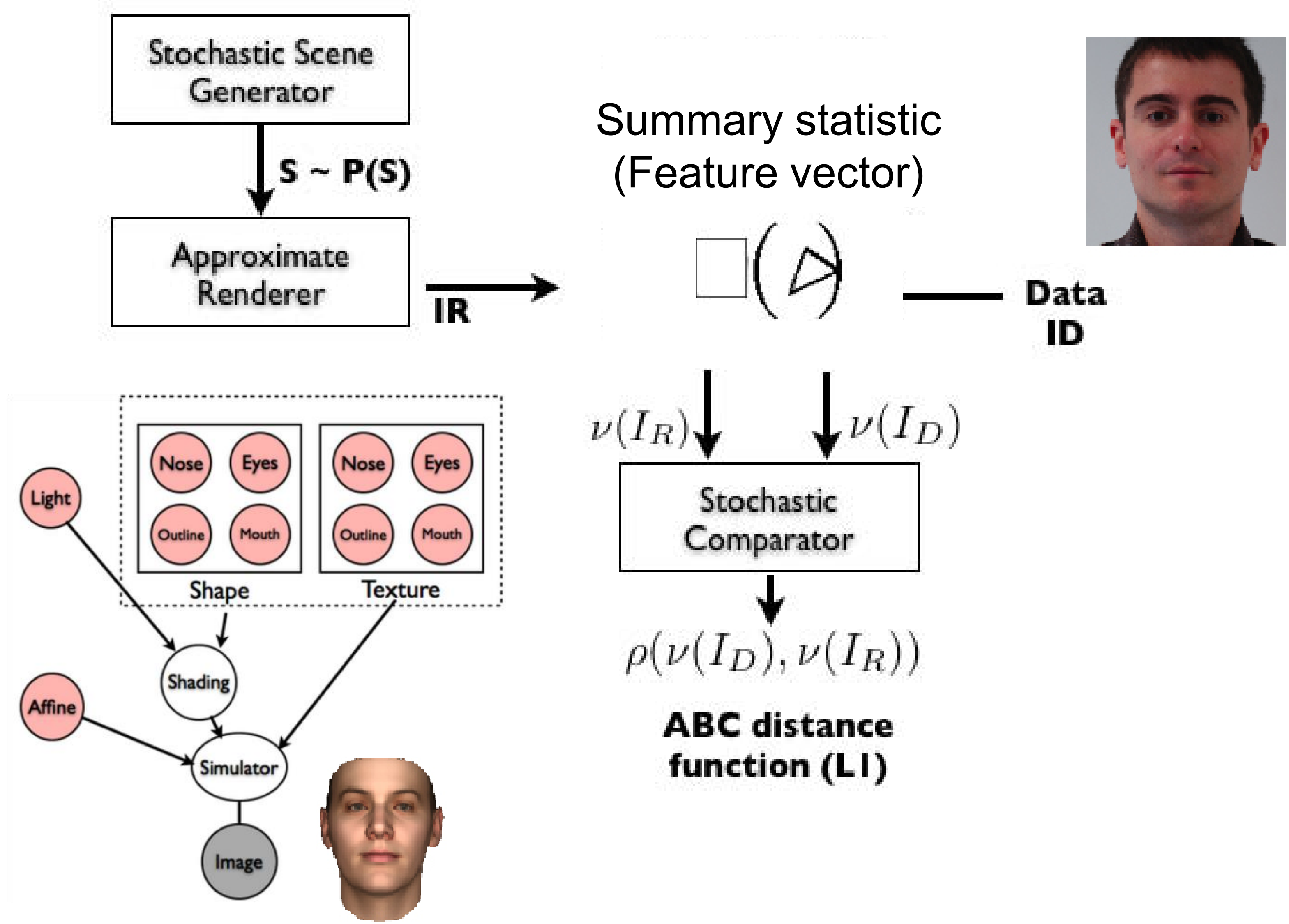
Random samples …



Courtesy of Ilker Yildirim. Used with permission.

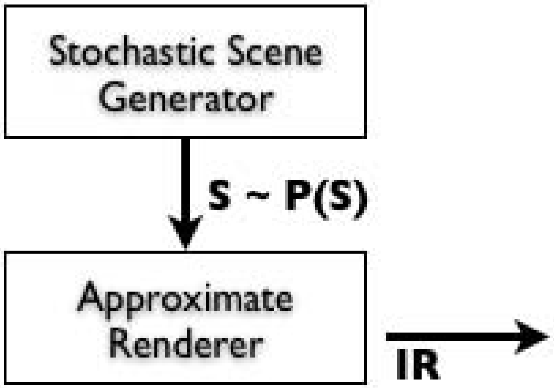
From Yildirim, Kulkarni, Friewald, and Tenenbaum (2015)

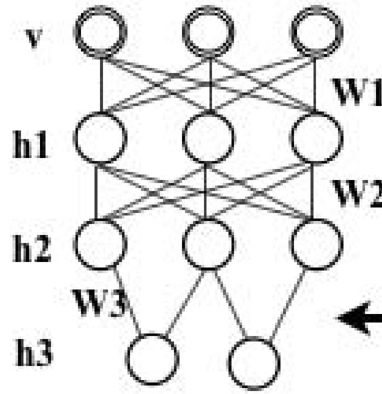
Architecture (Kulkarni et al., CVPR 2015)

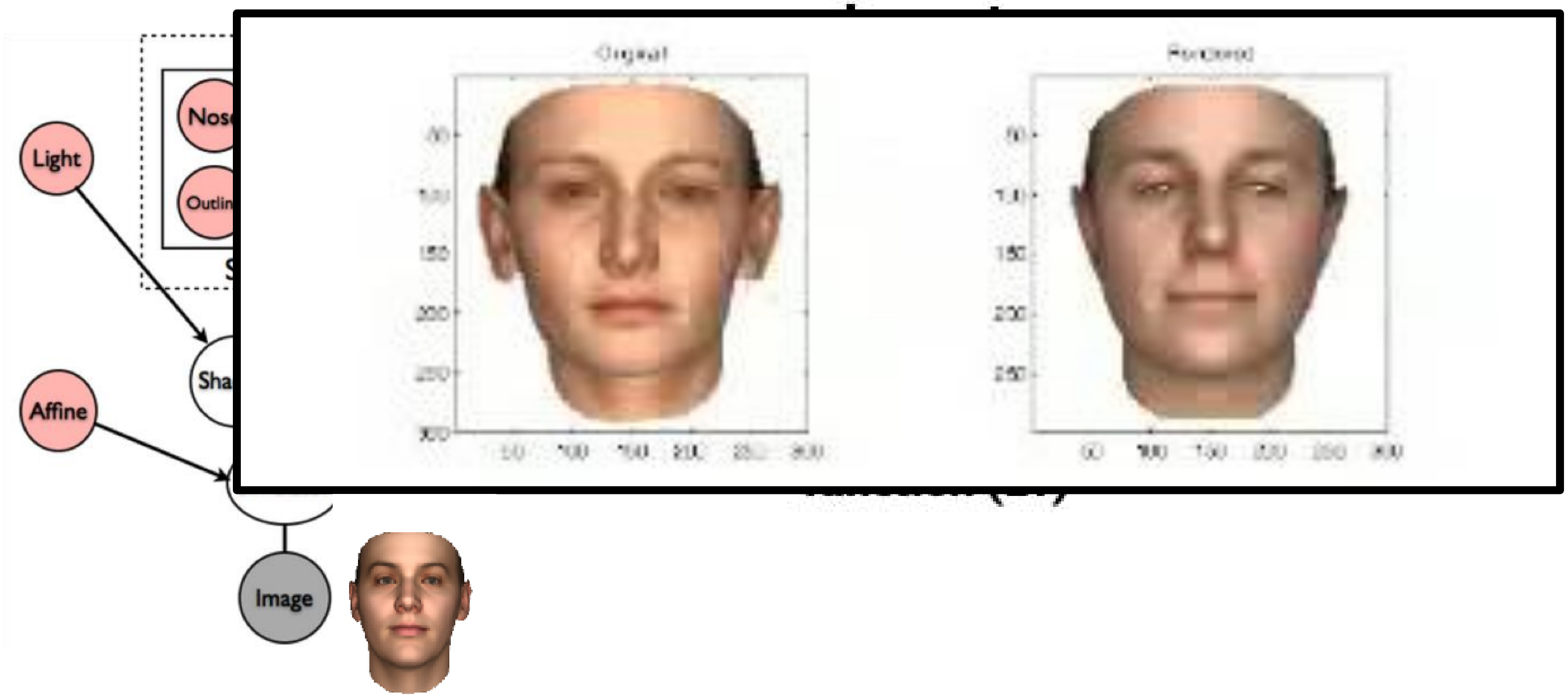


Architecture (Kulkarni et al., CVPR 2015)

Convolutional neural network

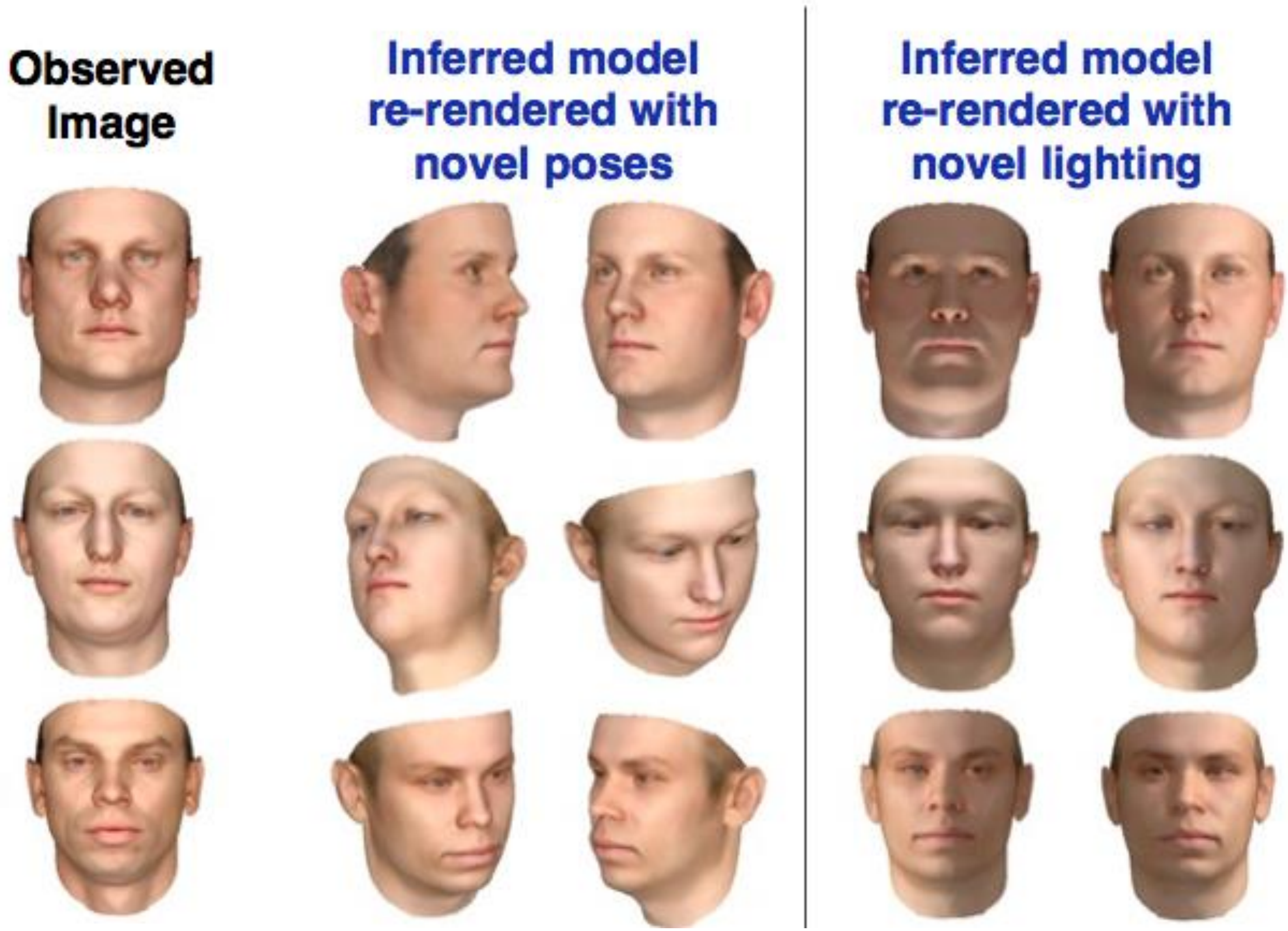




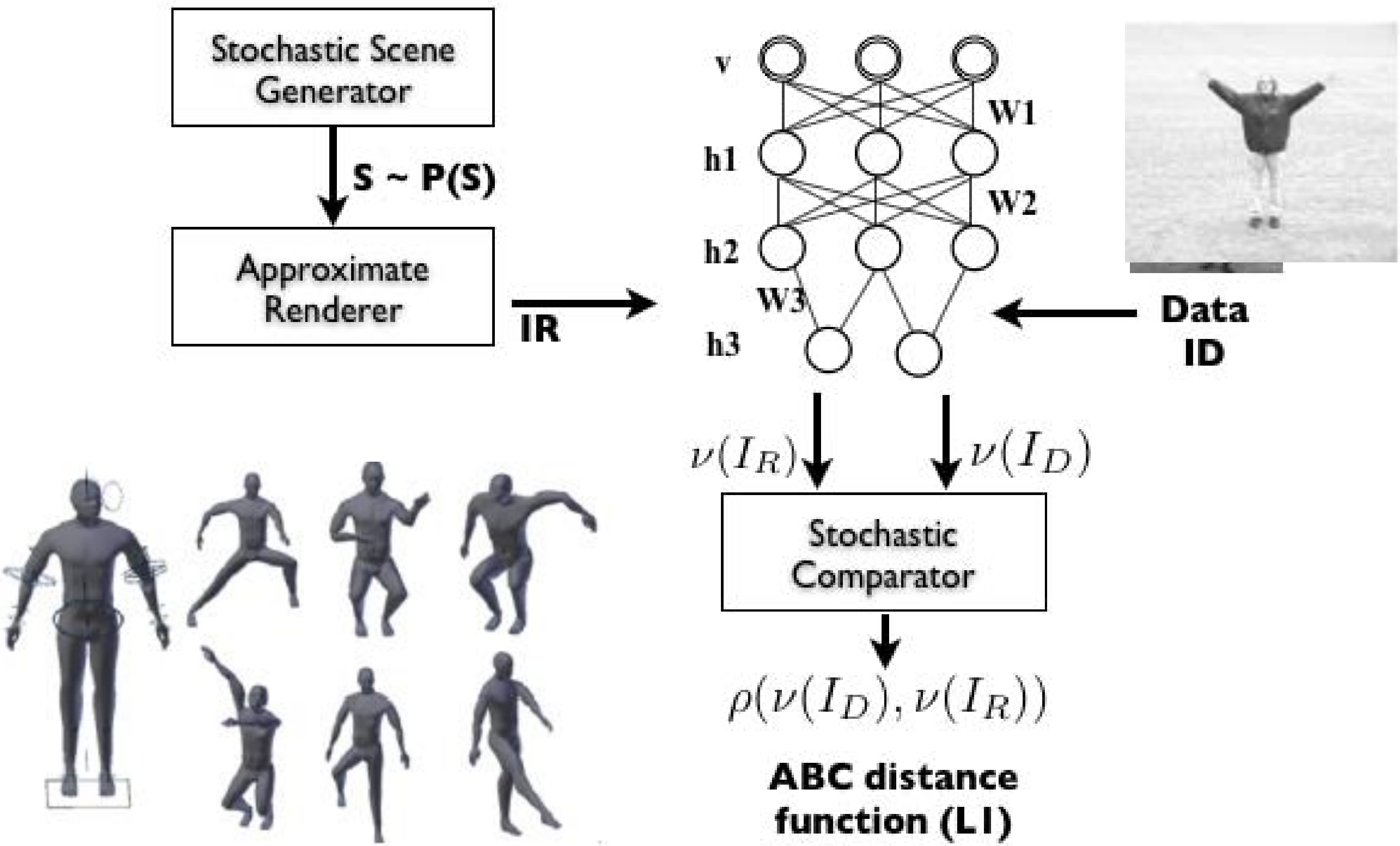


Generalizing across viewing conditions

(Kulkarni et al., CVPR 2015)

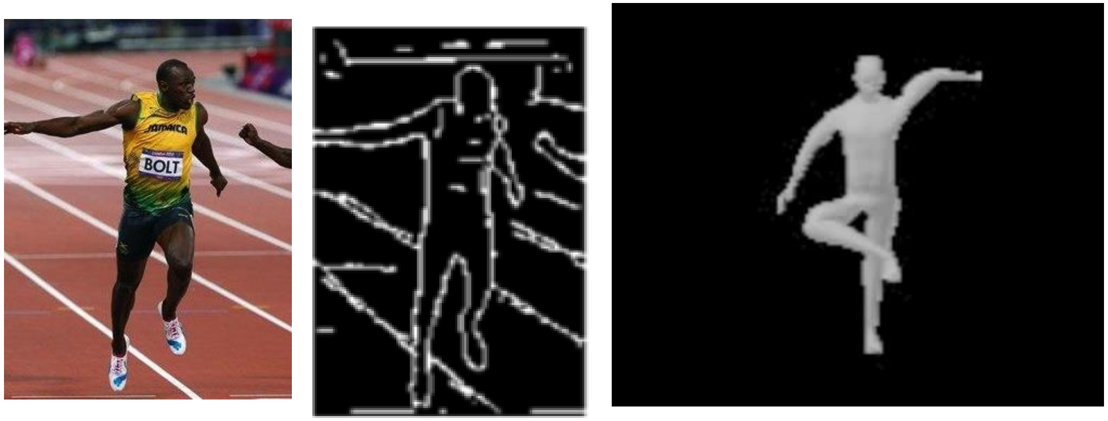


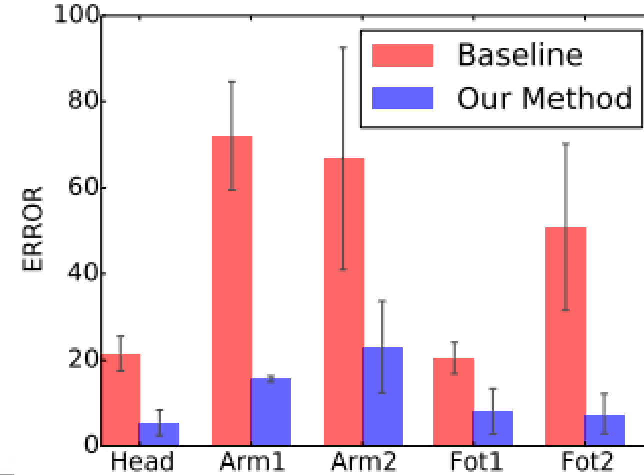
Human body pose estimation

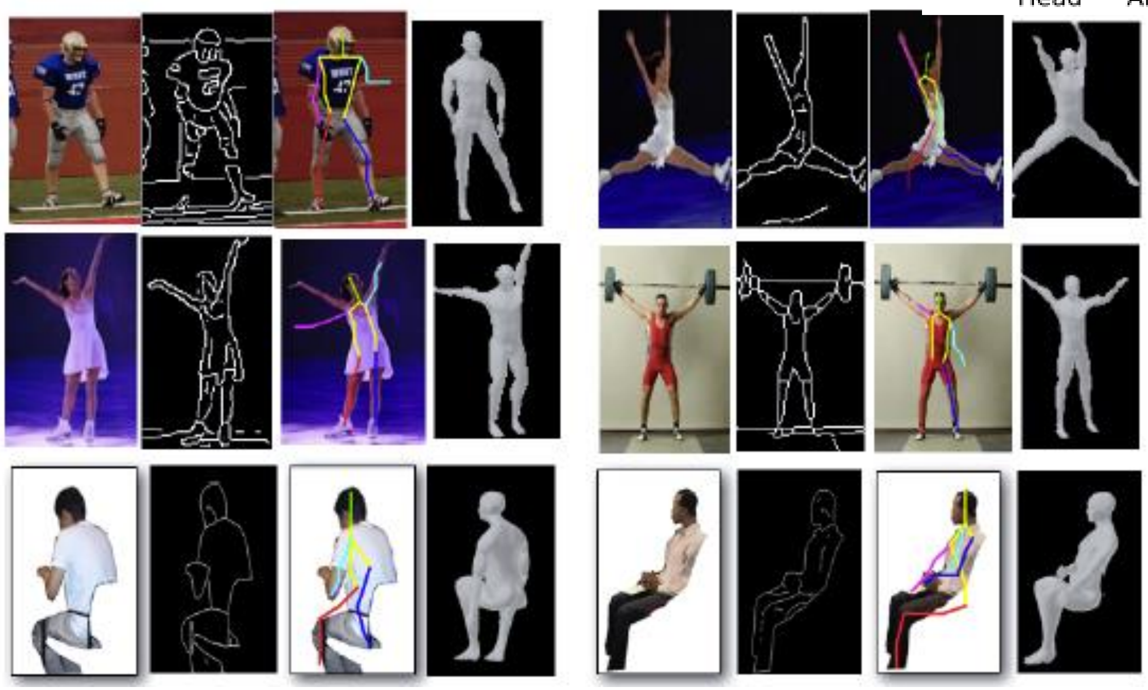


(Kulkarni et al., CVPR 2015)

Human body pose estimation

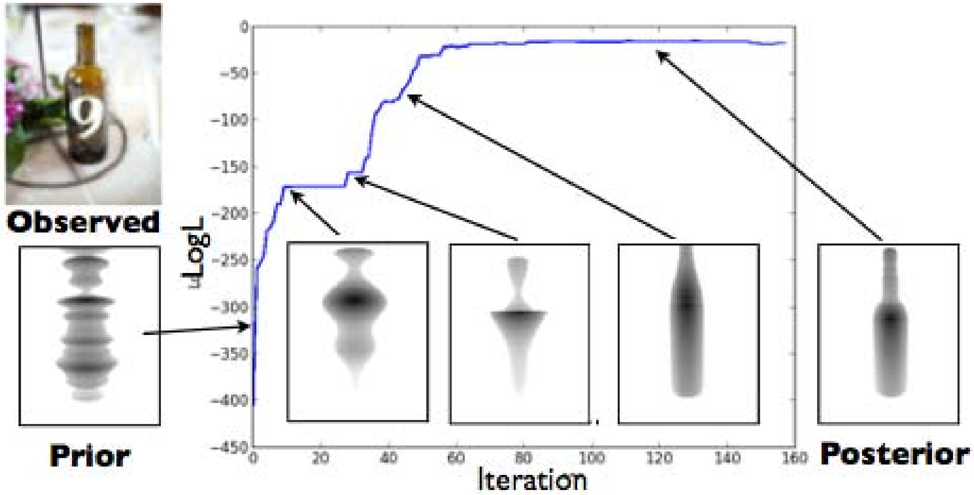






(Kulkarni et al., CVPR 2015)

Inferring generic 3D shape

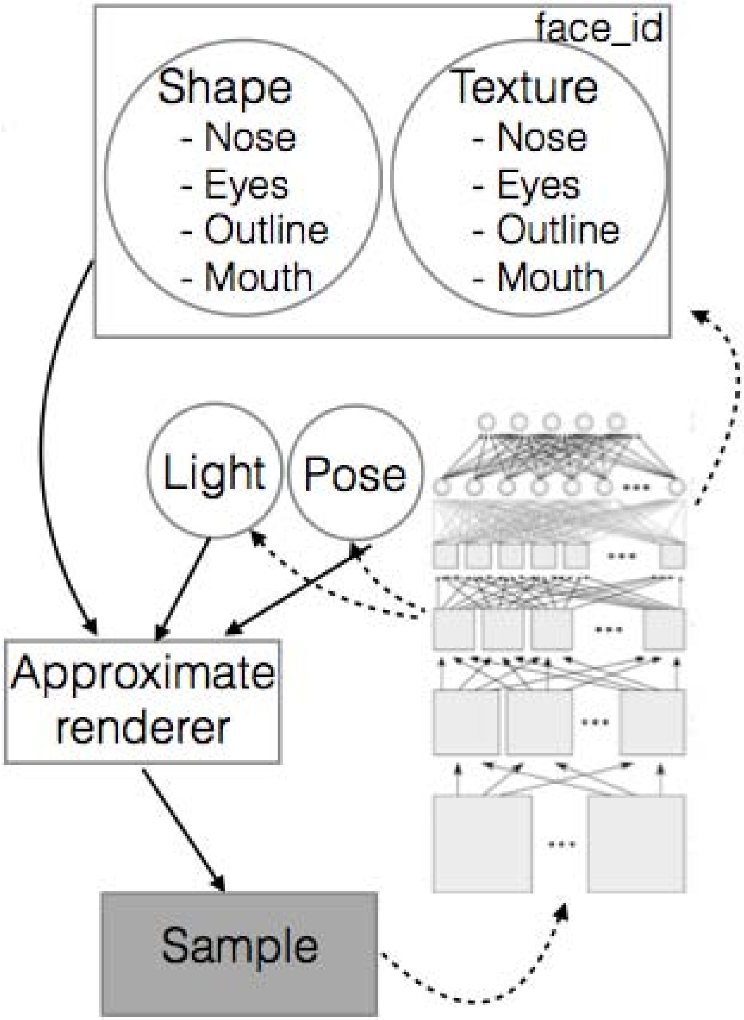


Courtesy of Tejas Kulkarni. Used with permission.

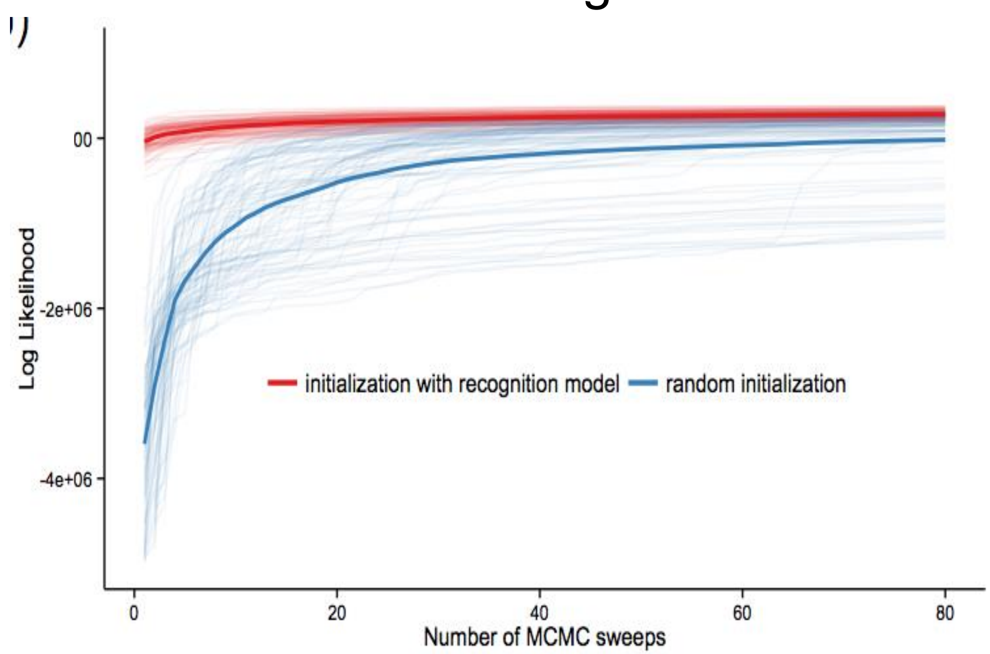
(Kulkarni et al., CVPR 2015)

Faster (and more brain-like) inference

(Yildirim, Kulkarni, Freiwald, Tenenbaum, Cog Sci '15, in prep)



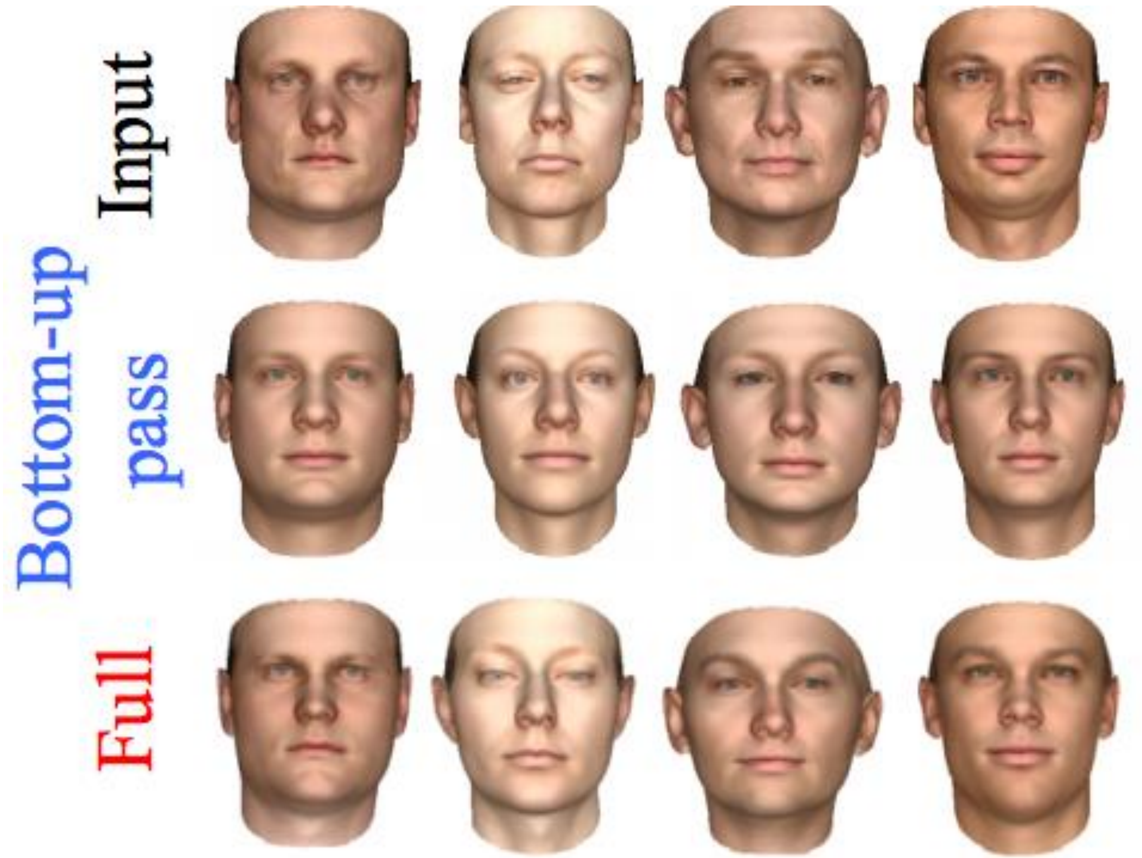
Learning to do inference a la Helmholtz machine (Hinton et al., 1995):
- Initialize inference with recognition model (a deep neural network).
- Trained in a self-supervised way from fantasies of the generative model.



Courtesy of Ilker Yildirim. Used with permission.

Psychophysics and neural data

(Yildirim, Kulkarni, Freiwald, Tenenbaum, Cog Sci ‘15, in prep)



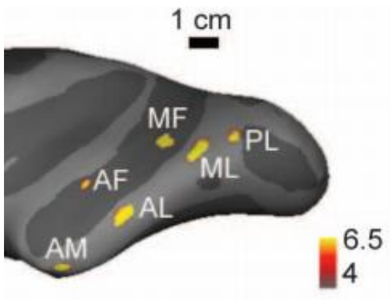
Courtesy of Ilker Yildirim. Used with permission.

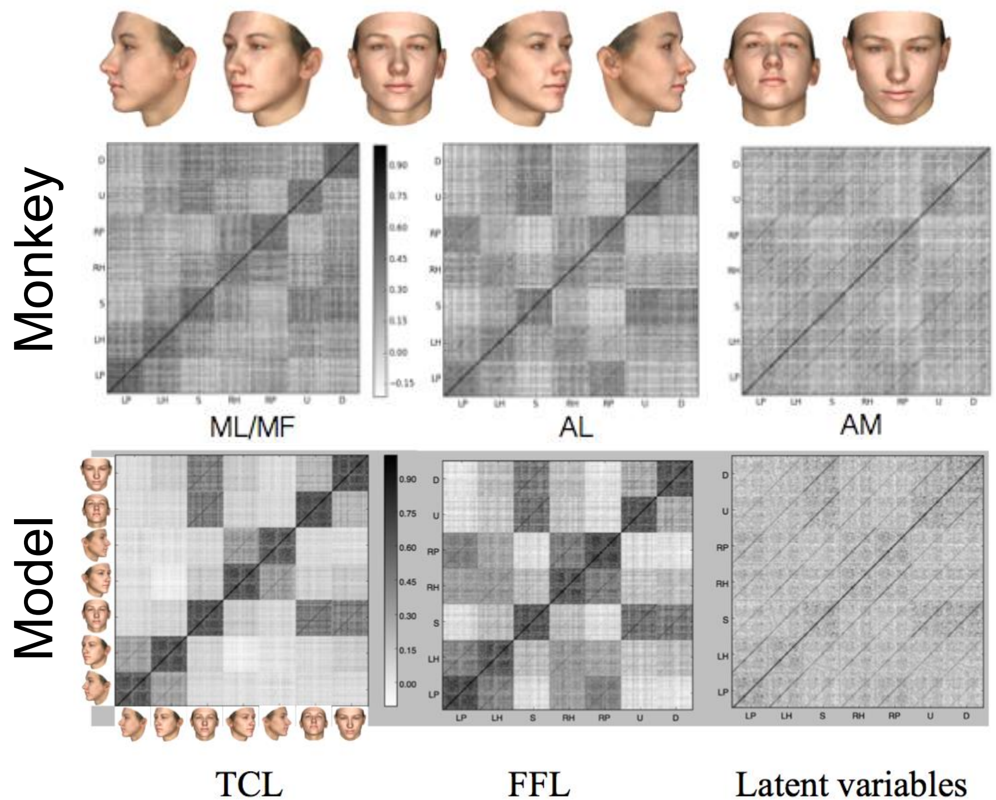
Psychophysics and neural data

(Yildirim, Kulkarni, Freiwald, Tenenbaum, Cog Sci '15, in prep)

Courtesy of Ilker Yildirim. Used with permission.

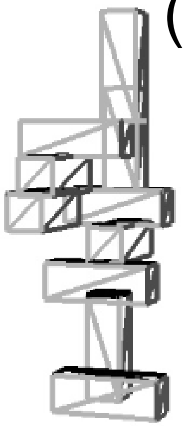
(Freiwald and Tsao, 2010)



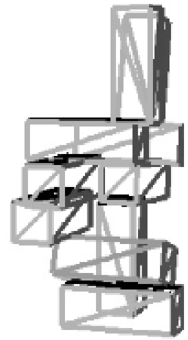


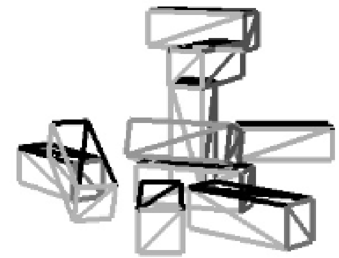
The intuitive physics engine

(Battaglia, Hamrick, Tenenbaum, PNAS 2013)



Prob.
approx.
Newton

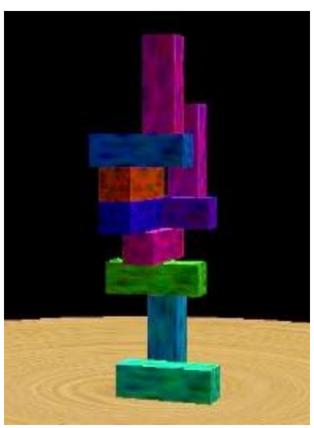




… World state (t-1) → World state (t) → World state (t+1) …

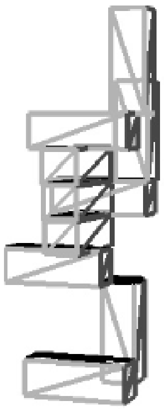
Prob. approx. rendering

Image (t-1) Image (t) Image (t+1)

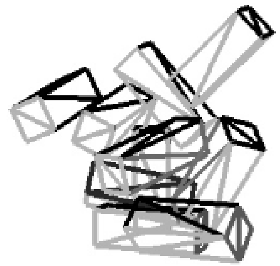


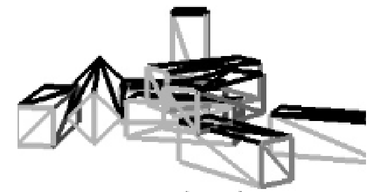
The intuitive physics engine

(Battaglia, Hamrick, Tenenbaum, PNAS 2013)



Prob.
approx.
Newton

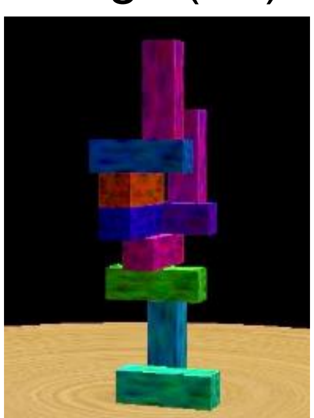




… World state (t-1) → World state (t) → World state (t+1) …

Prob. approx. rendering

Image (t-1)   Image (t)   Image (t+1)

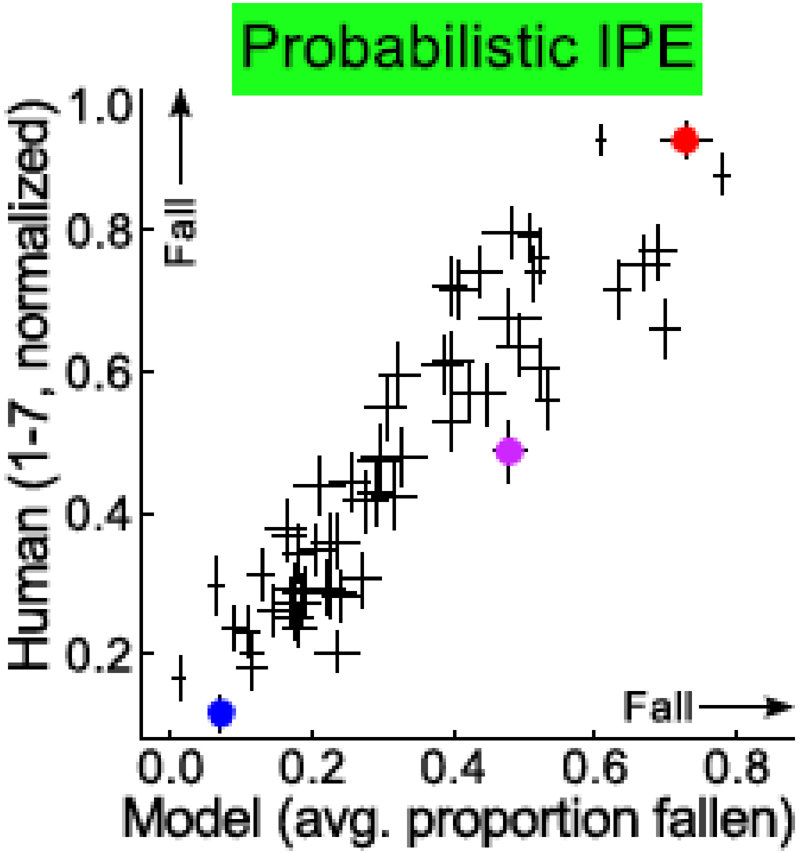


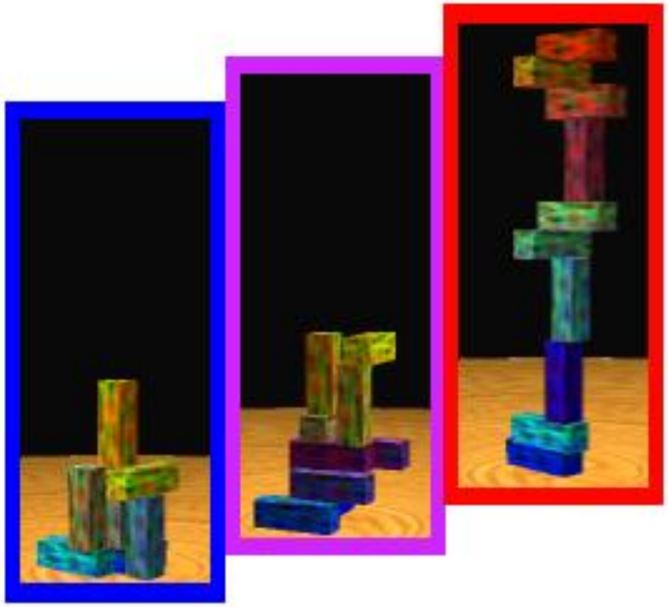
σ = state uncertainty

ϕ = latent force magnitude

The intuitive physics engine

(Battaglia, Hamrick, Tenenbaum, PNAS 2013)

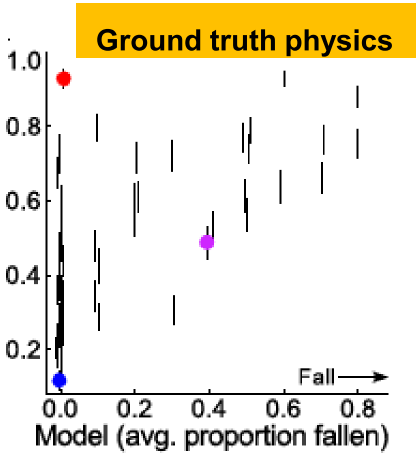




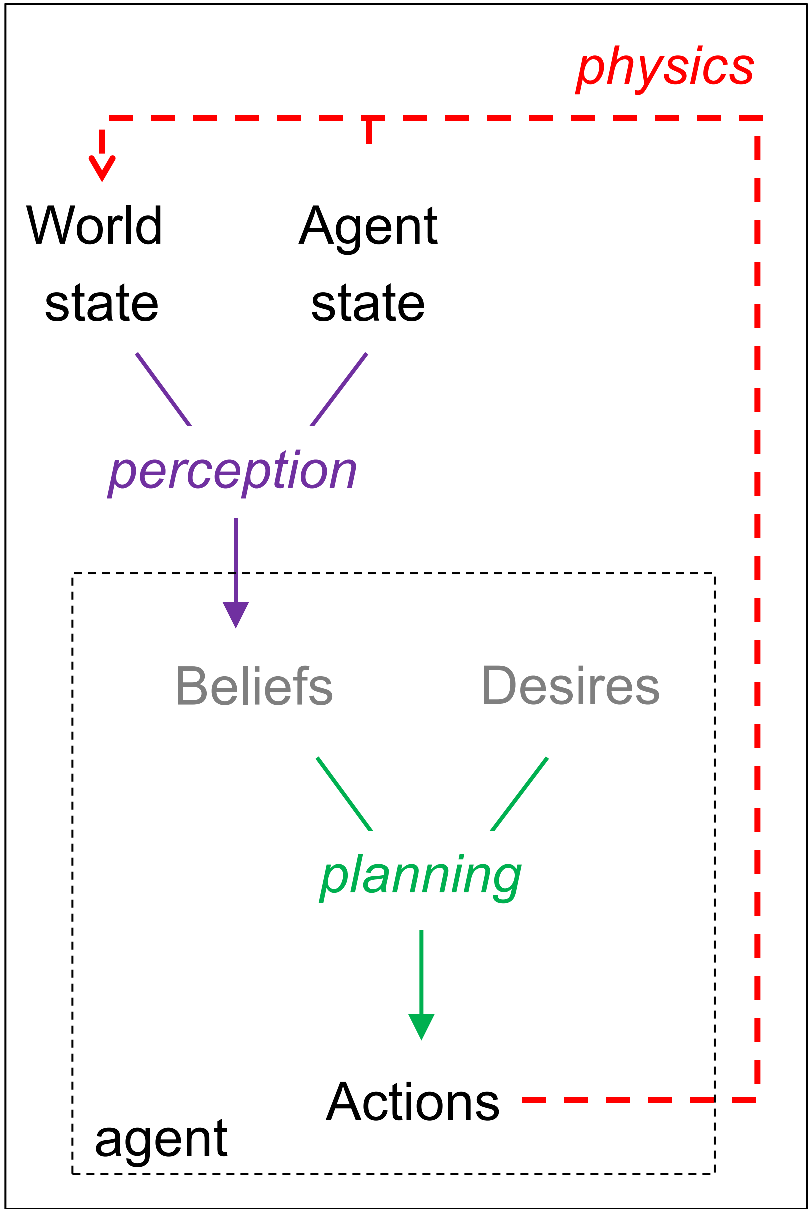
Courtesy of National Academy of Sciences, U. S. A. Used with permission. Source: Battaglia, P. W., et al. "Simulation as an engine of physical scene understanding." PNAS 110 no. 45 (2013): 18327-18332 Copyright © 2013 National Academy of Sciences, U.S.A.

$\sigma = 0.2$

$\phi = 0.2$

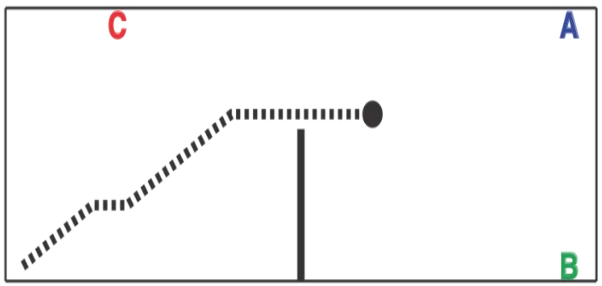


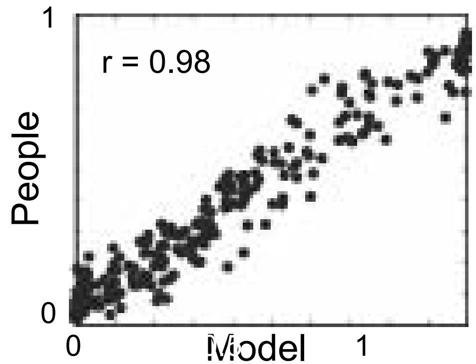
The intuitive psychology engine

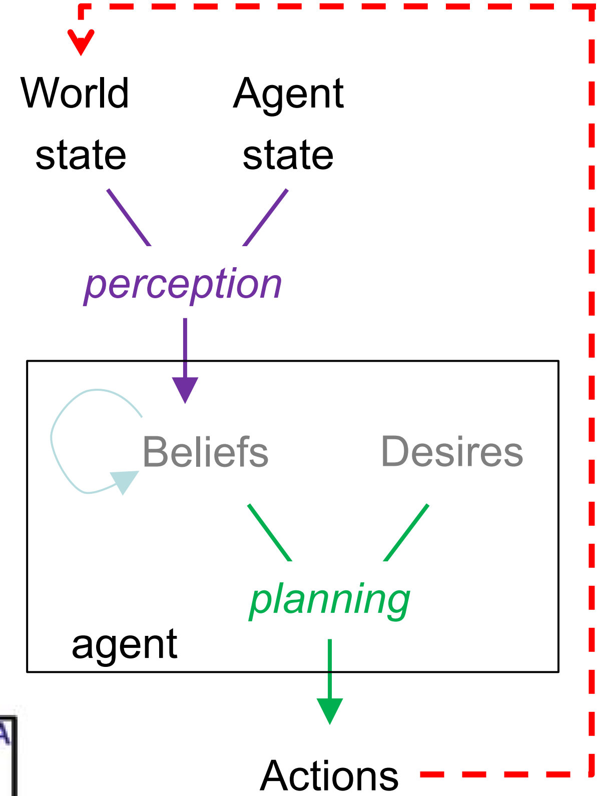


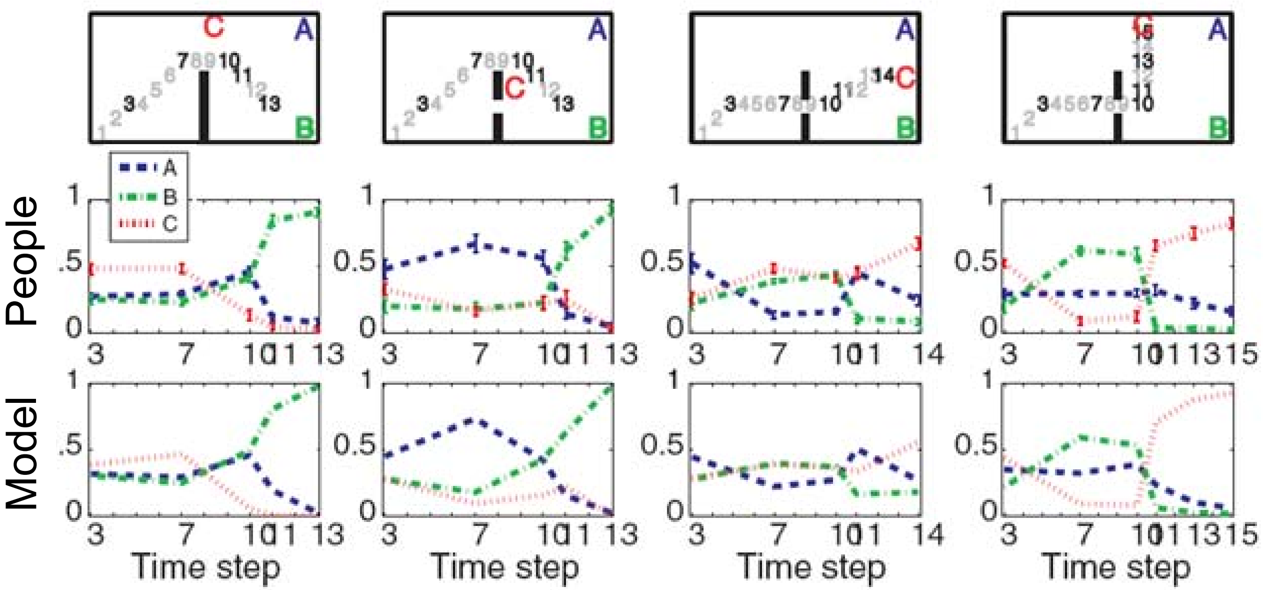
Photos (1) from TV show *The Office (2)* young students in crosswalk, with crossing guard, removed due to copyright restrictions.

Goal inference as inverse planning
(Baker, Saxe, Tenenbaum, 2009)









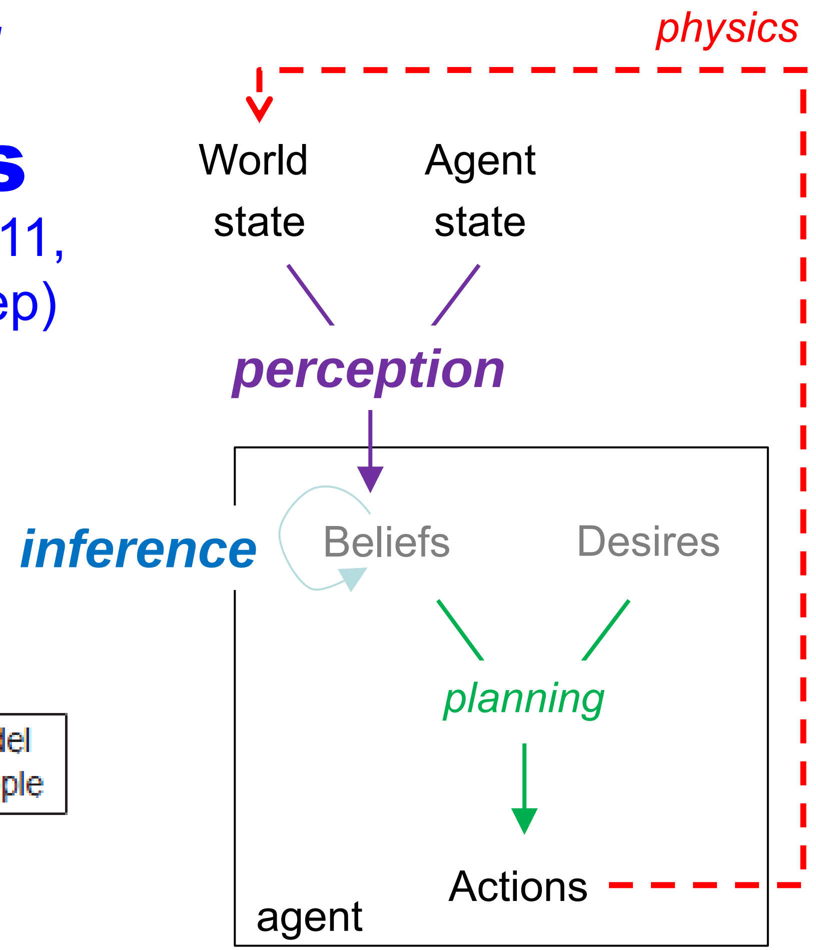
$$U(a,s) = R(s) - C(a)$$

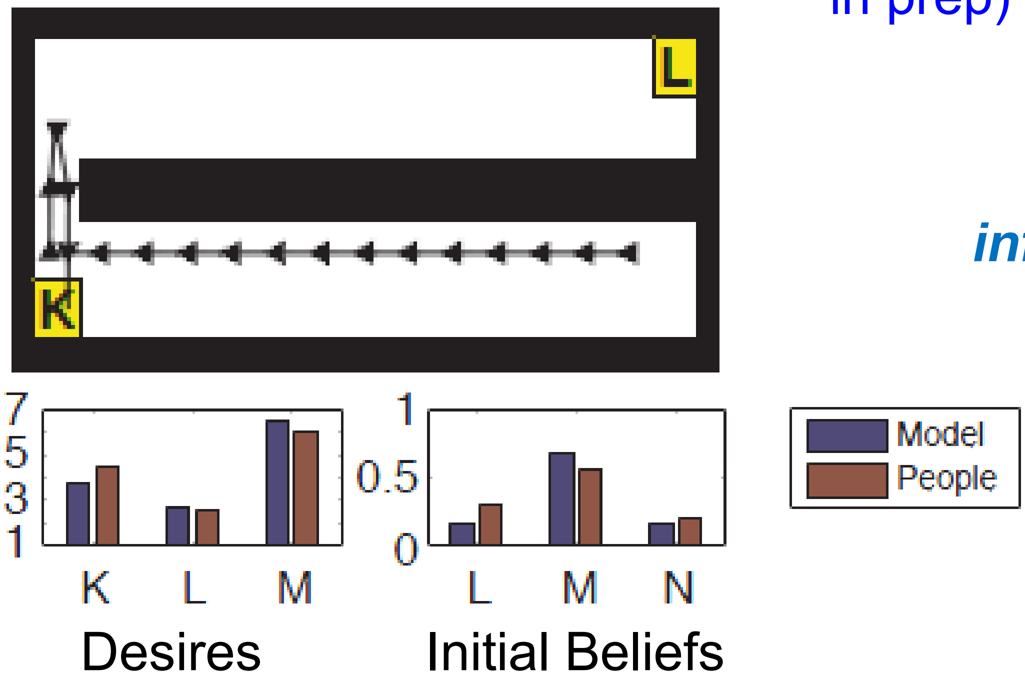
$R(s)$: large reward for achieving goal

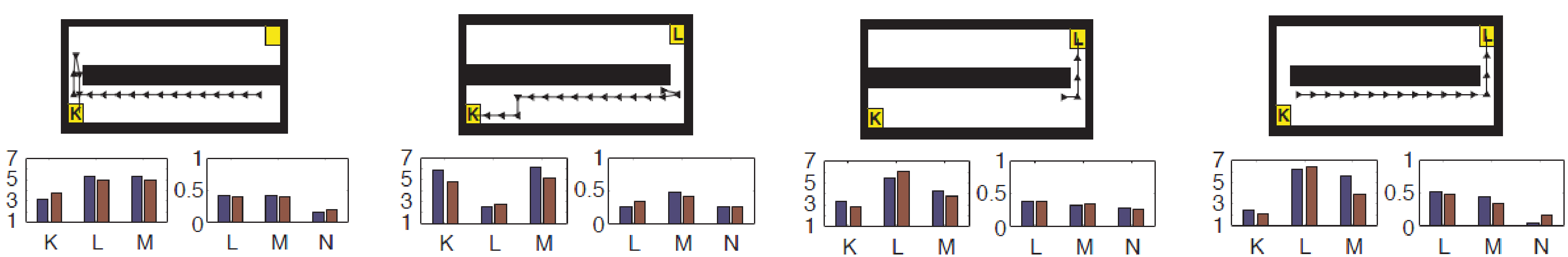
$C(a)$: small cost per step

Joint inference of beliefs and desires

(Baker, Saxe, Tenenbaum, Cog Sci 2011, in prep)

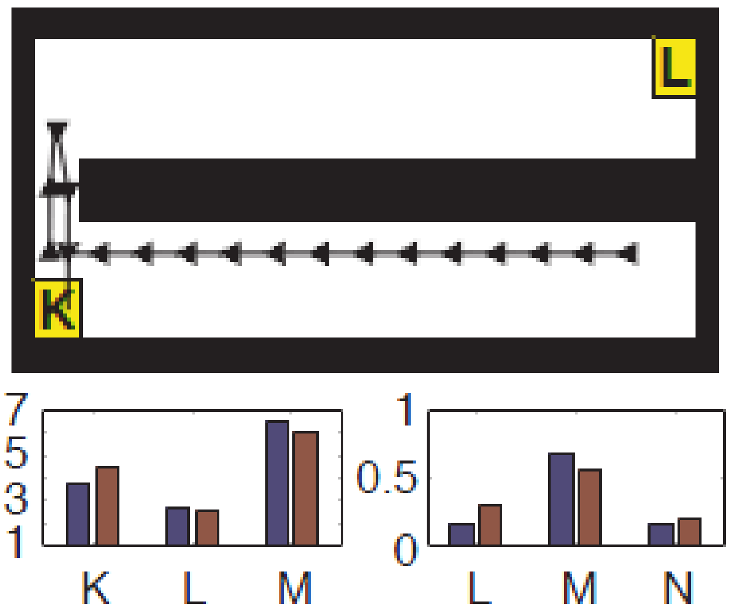


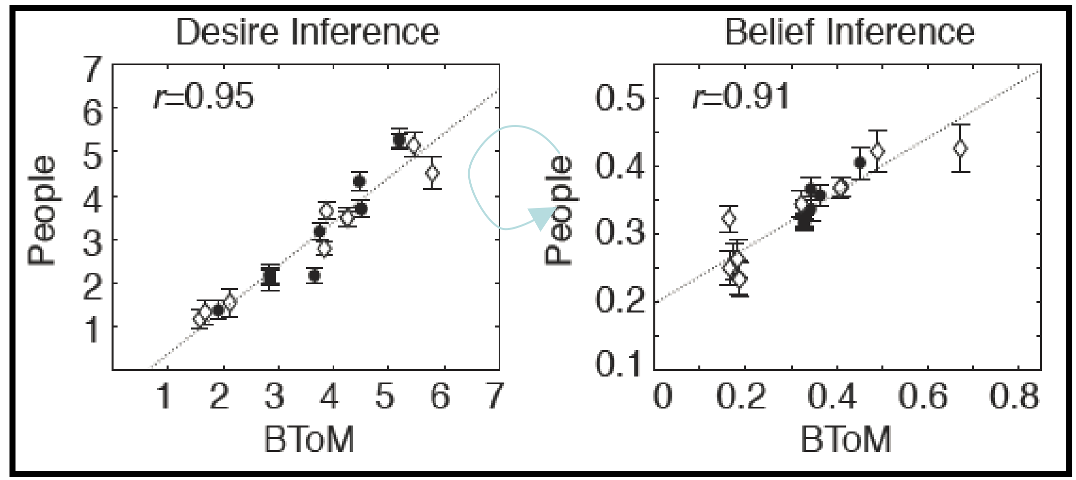


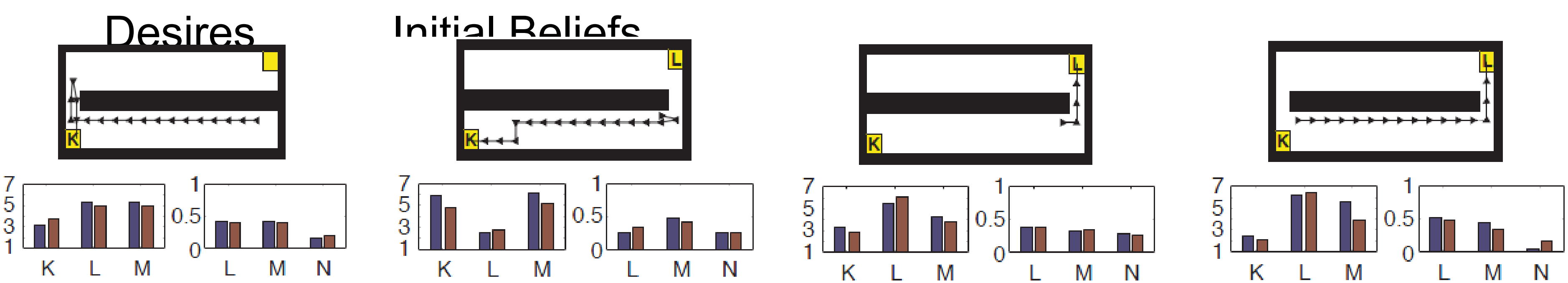


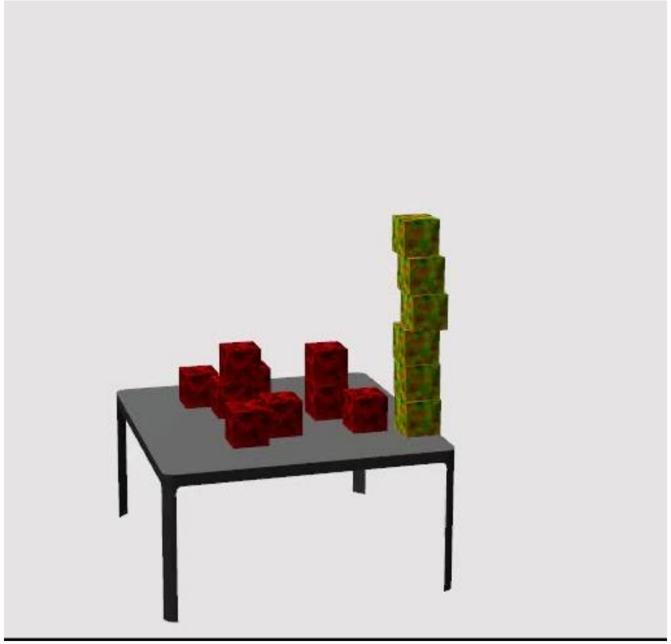
Joint inference of beliefs and desires

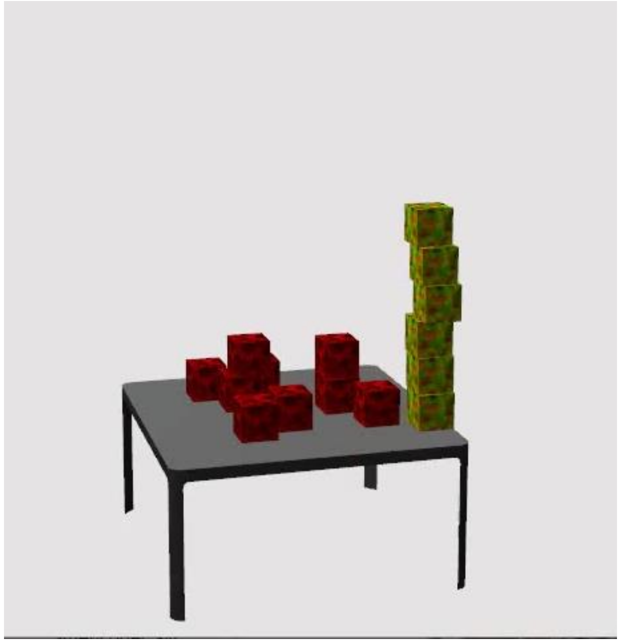
(Baker, Saxe, Tenenbaum, Cog Sci 2011, in prep)



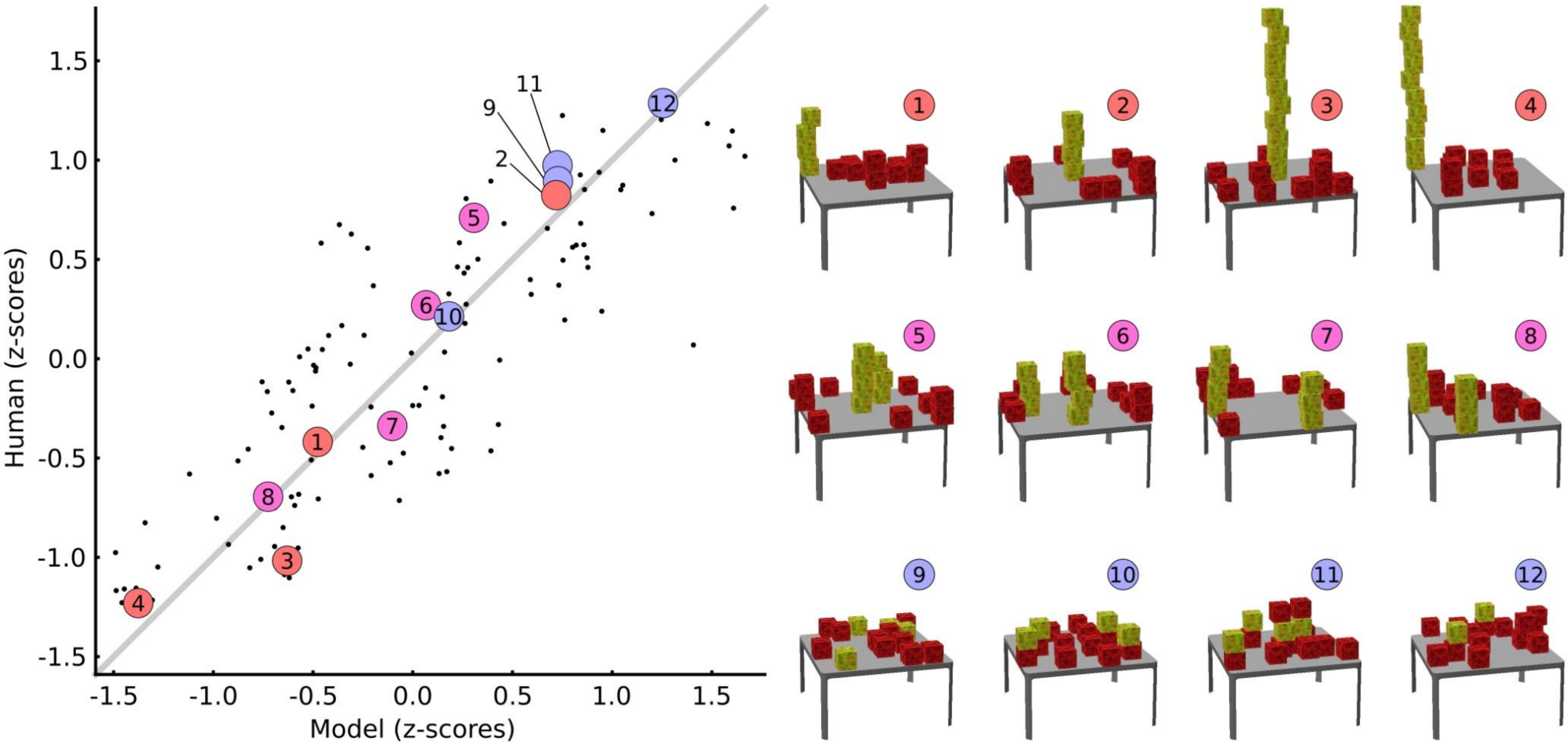








If you bump the table...



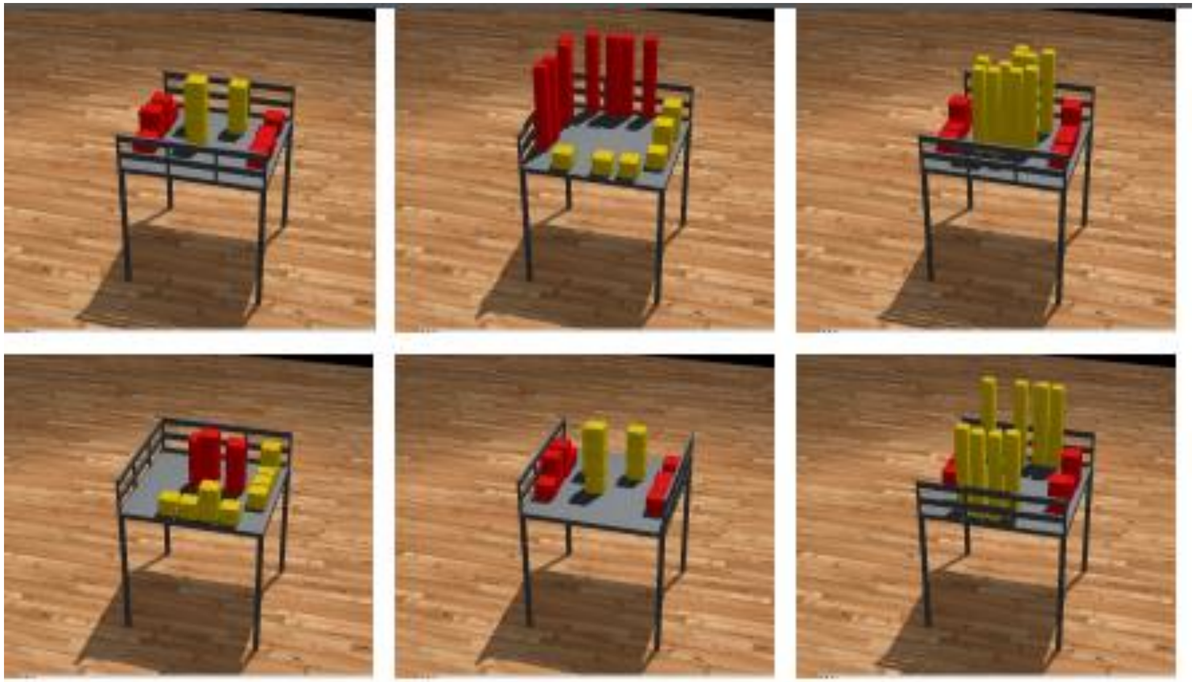
100% yellow

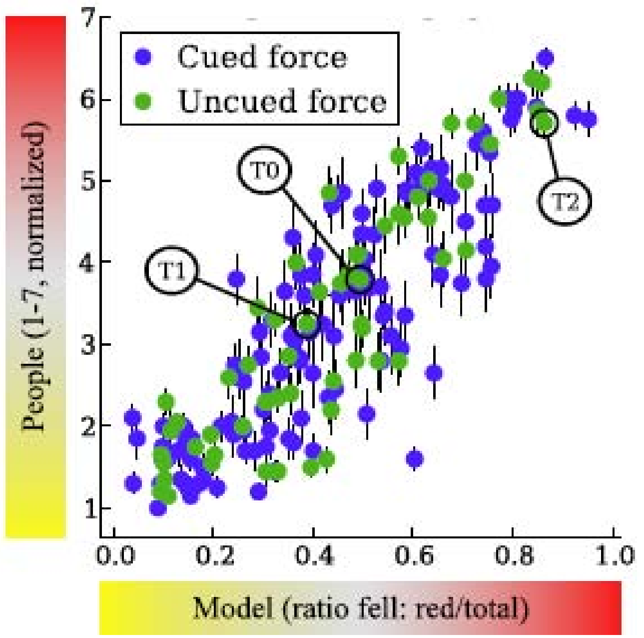
100% red

Model simulates table “bumps” integrating over a range of force magnitudes and directions. (R = 0.84)

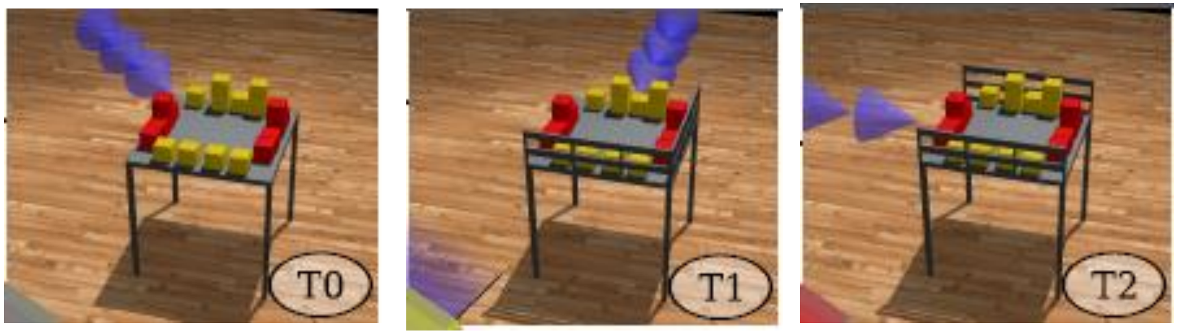
Varying objects, constraints, forces

Uncued forces

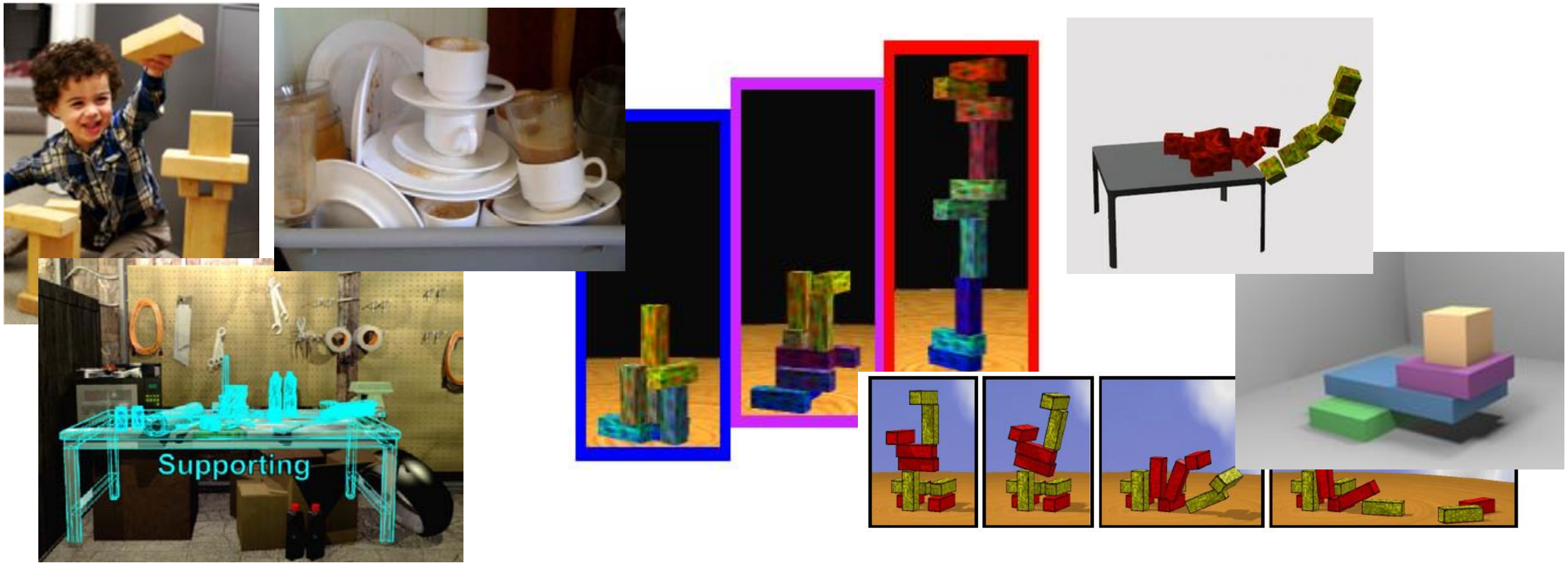


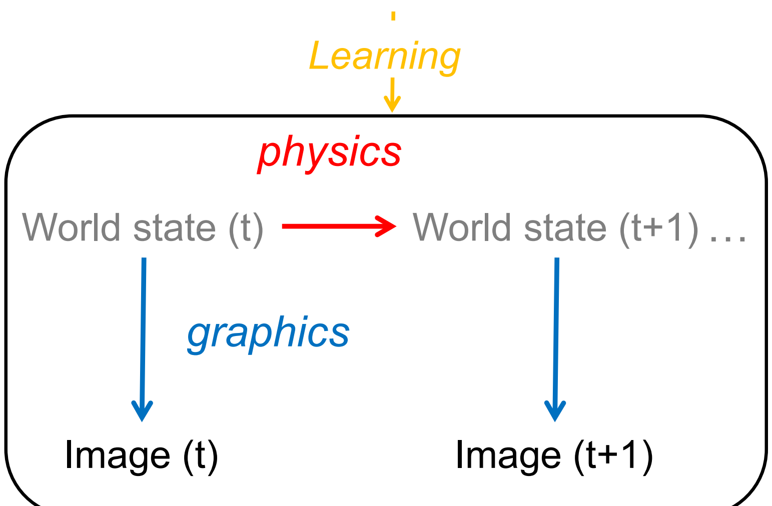


Cued forces



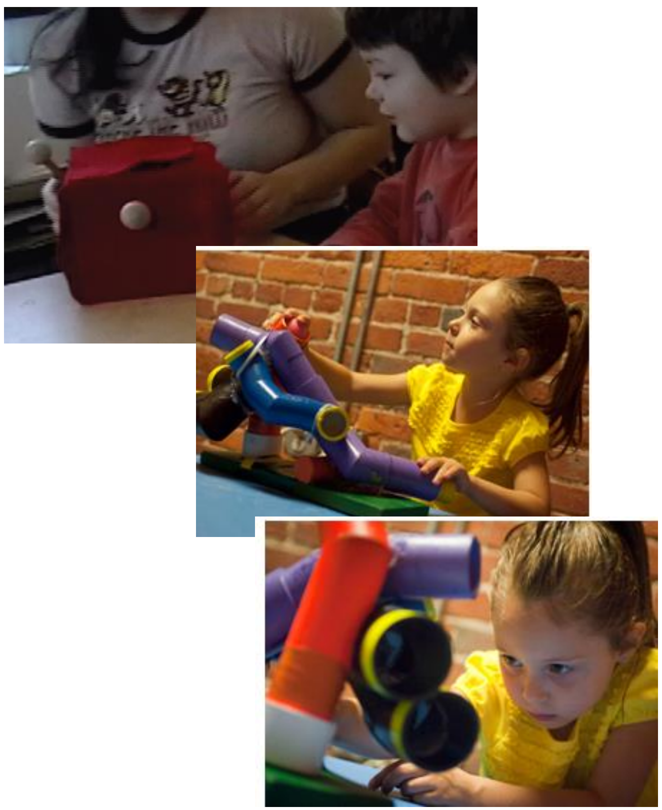
Probabilistic programs for model building (“program-learning” programs)

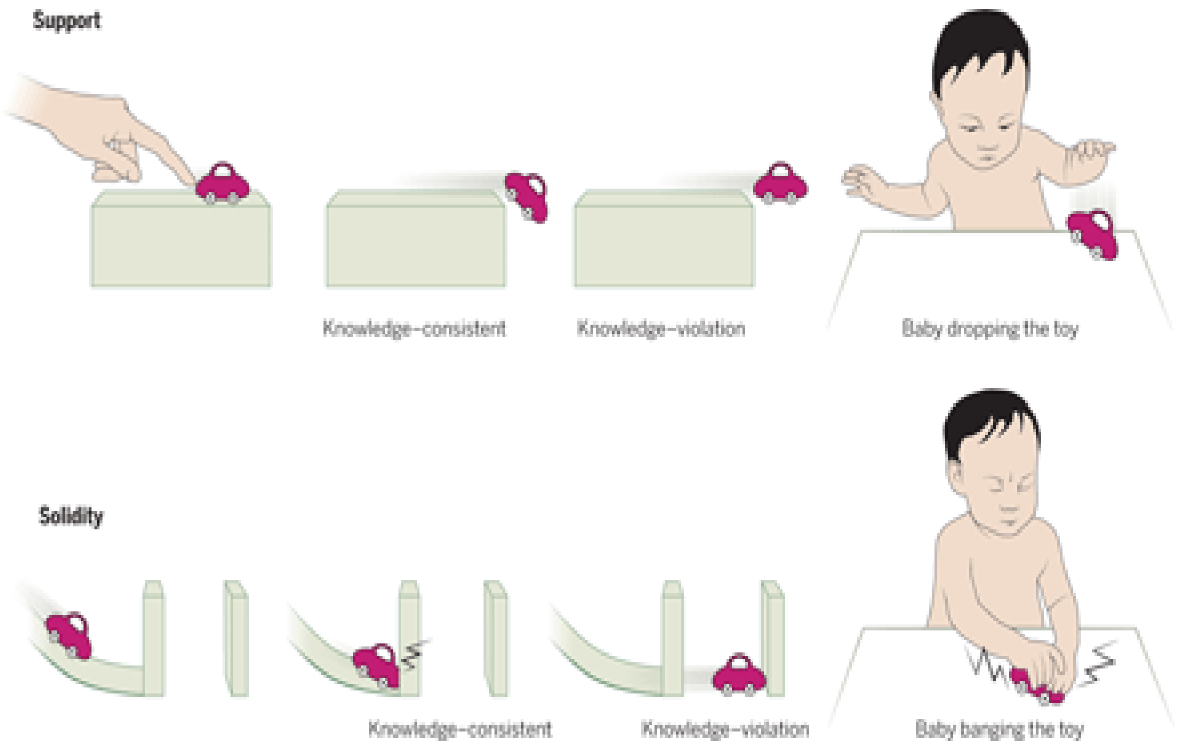




The child as scientist

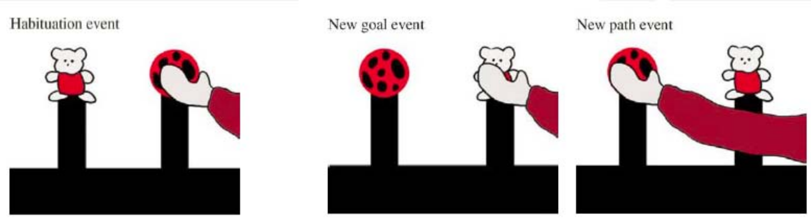
Learning as "theory building", not "data analysis".
Knowledge grows through hypothesis- and explanation-driven interpretations of sparse data, causal learning, learning theories, learning compositional abstractions, learning to learn, exploratory learning, social learning.
[Carey, Karmiloff-Smith, Gopnik, Schulz, Feigenson…]





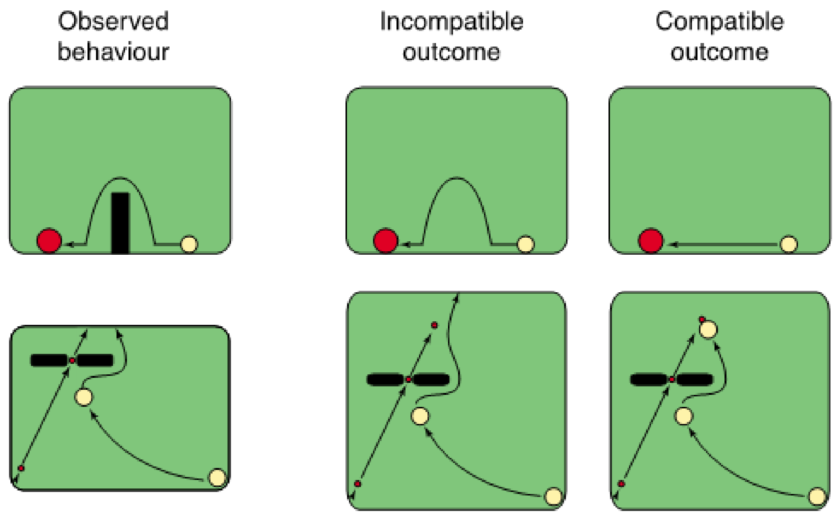
Explaining the dynamics of development? (w/ T. Ullman, Spelke, others)

9 months

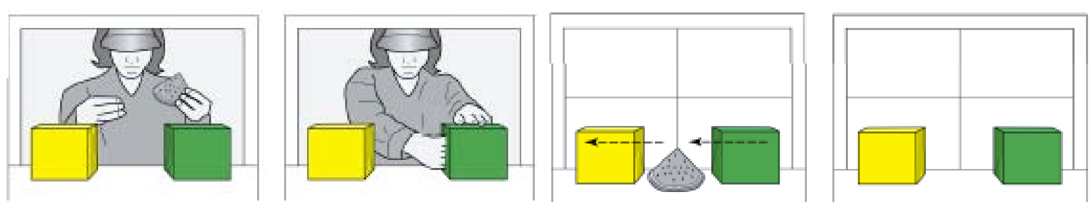


Courtesy of Elsevier, Inc., http://www.sciencedirect.com. Used with permission. Source: Sommerville, Jessica A., Amanda L. Woodward, and Amy Needham. "Action experience alters 3-month-old infants' perception of others' actions." Cognition 96, no. 1 (2005): B1-B11.

12 months

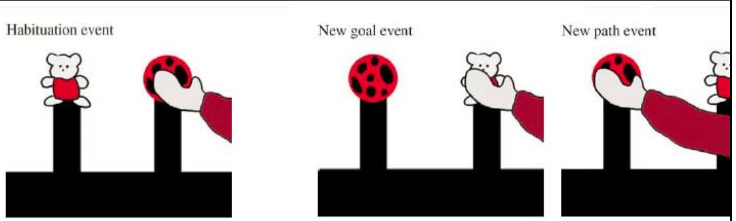


15 months



Explaining the dynamics of development? (w/ T. Ullman, Spelke, others)

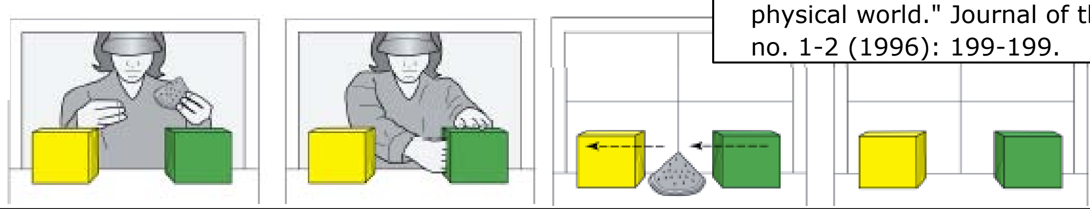
9 months



Capture different knowledge stages with a sequence of probabilistic programs?

Explain the trajectory of stages as rational statistical inference in the space of programs?

15 months



3 months

Initial Concept:
Contact/No contact

5 months

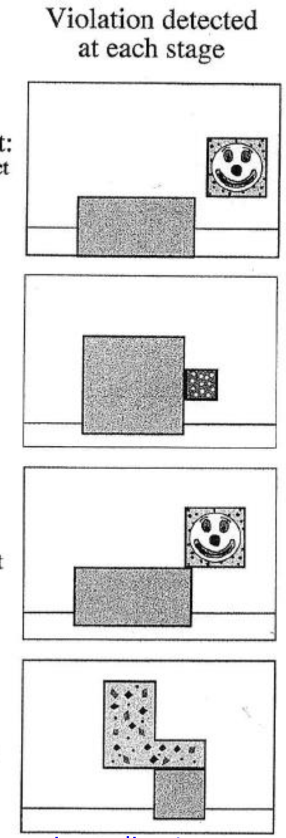
Variable:
Type of contact

6.5 months

Variable:
Amount of contact

12.5 months

Variable:
Shape of the box



Conclusion

What makes us so smart?

1. **How we start:** Common-sense core theories of intuitive physics and intuitive psychology.
2. **How we grow:** Learning as theory construction, revision and refinement.

The tools of probabilistic programs and program induction are beginning to let us reverse-engineer these capacities, with languages that are:

- Probabilistic.
- Generative.
- Causally structured
- Compositionally structured: flexible, fine-grained dependencies, hierarchical, recursive, unbounded

We have to view the brain not simply as a pattern-recognition device, but as a *modeling engine*, an *explanation engine* – and we have to understand how these views work together.

Much promise but huge engineering and scientific challenges remain… full of opportunities for bidirectional interactions between cognitive science, neuroscience, developmental psychology, AI and machine learning.