MIT

ndg@mit.edu

Causal Models Revisited

Noah D. Goodman
Computational Cognitive Science Group,
MIT

IPAM GSS, UCLA
July 2007

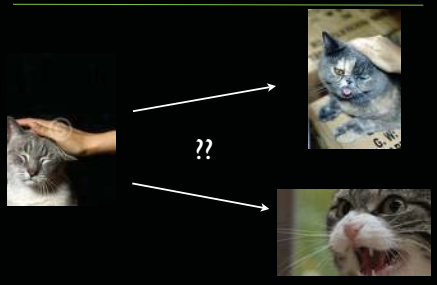
Causal Models

- Causal models represent causal structure between variables
  (e.g. Waldmann and Holyoak, 1992; Cheng, 1997; Pearl, 2000; Gopnik, et al, 2004; Griffiths and Tenenbaum, 2005),
- which are learned from contingency data (and interventions, etc.).

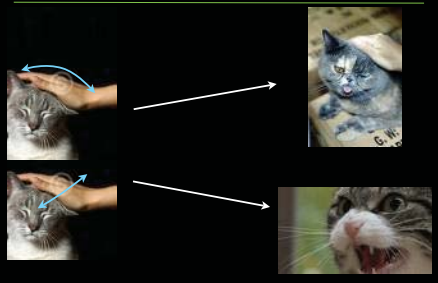
Petting:	yes	yes	no	yes	no	yes
Purring:	yes	no	no	yes	no	yes

Causal Musings

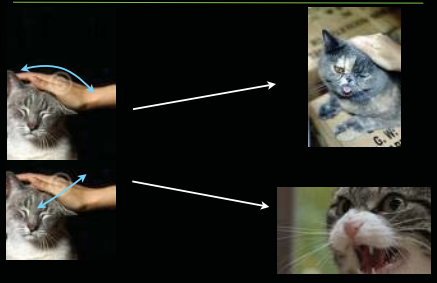
Causal Musings



Causal Musings



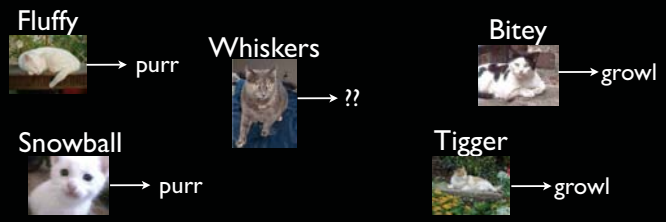
Causal Musings



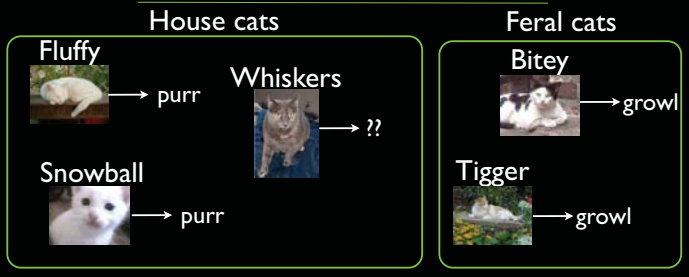
- A child learns that petting the cat leads to purring, while pounding leads to growling. What are the origins of the event concepts (variables) over which causal links are defined?

Causal Musings

Causal Musings

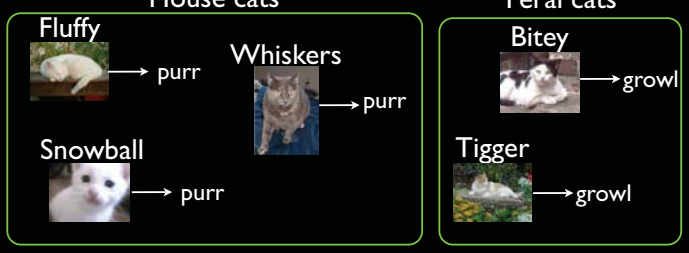


Causal Musings



Causal Musings

Causal Musings

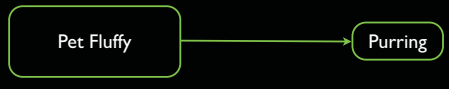


- A child learns that petting house cats leads to purring, while petting feral cats leads to biting. What is the form of this abstract causal knowledge? How can it be learned?

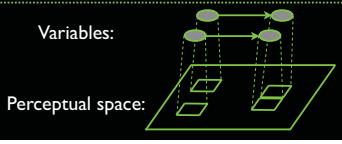
Causal Musings

Abstract Causal Knowledge:

Causal Model:



Perceptual Grounding:



MIT

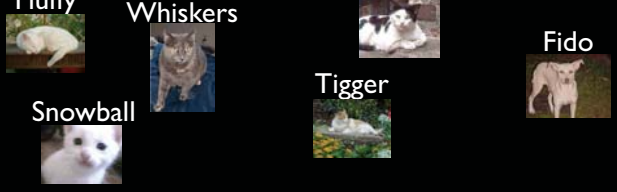
ndg@mit.edu

Part I: Causal Schemata

Noah D. Goodman
Computational Cognitive Science Group,
MIT

with Charles Kemp and Josh Tenenbaum

Schemata



Schemata

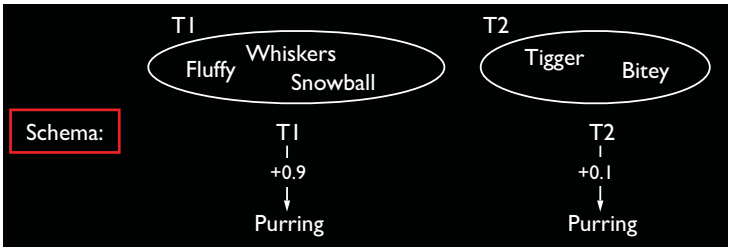
Schemata

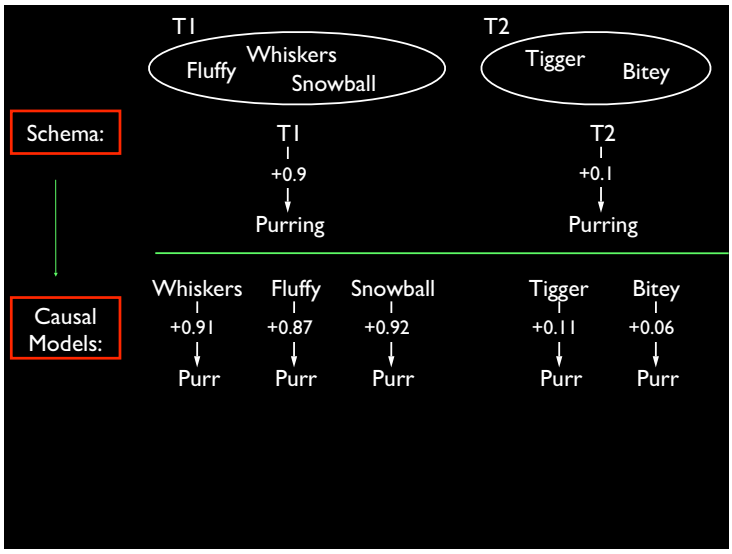
- Organize objects into causal types,
- Specify the causal powers of each type,
- Specify characteristic features of each type.

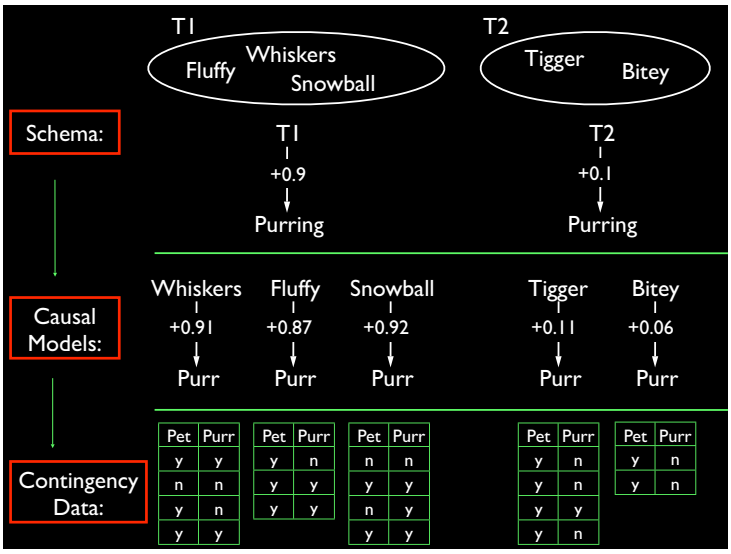
Schemata

- Organize objects into causal types,
- Specify the causal powers of each type,
- Specify characteristic features of each type.

Related work, see: Kelley (1973); Griffiths (2005); Mansingkha, et al (2006)

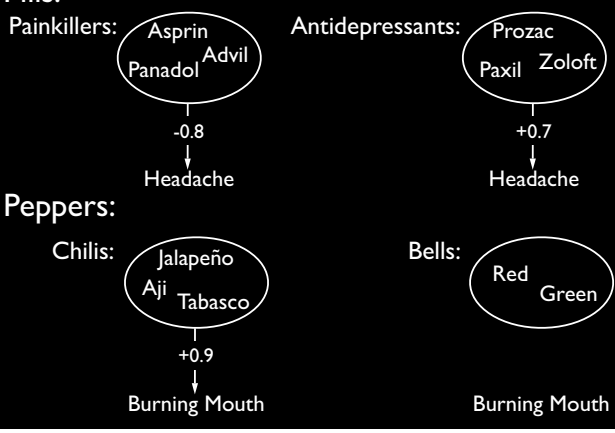




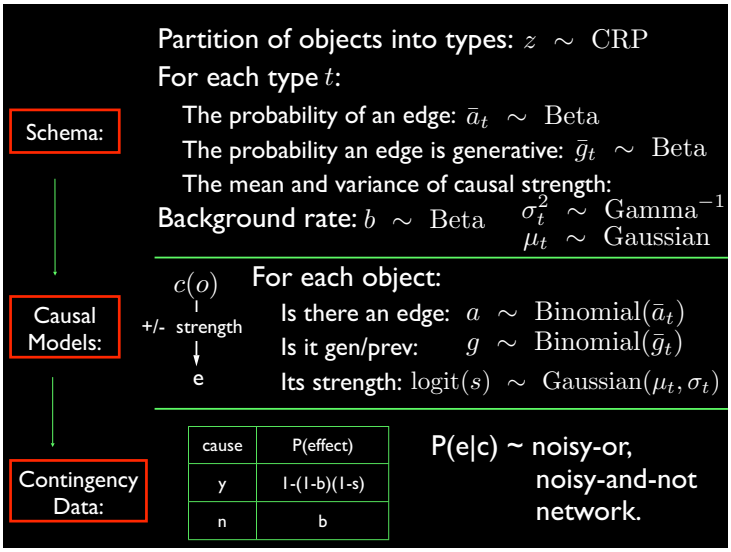


Other Examples

- Pills:



- Peppers:



Learning Schemata

- We have a probabilistic generative model giving:
  - the probability of contingency data given a schema,
  - the prior probability of a schema.
- We can invert this model, using Bayes rule, to get the posterior probability of a schema:

$$P(\mathrm{Schema}|\mathrm{Data}) = \frac{P(\mathrm{Schema})P(\mathrm{Data}|\mathrm{Schema})}{P(\mathrm{Data})}$$

Using Schemata

Using Schemata

Painkillers: Asprin, Advil, Panadol → -0.8 → Headache

Antidepressants: Prozac, Paxil, Zoloft → +0.7 → Headache

Drug X: 2 trials, no headache (b=0.5).

Using Schemata

Painkillers: Asprin, Advil, Panadol → -0.8 → Headache

Antidepressants: Prozac, Paxil, Zoloft → +0.7 → Headache

Drug X: 2 trials, no headache (b=0.5).

Using Schemata

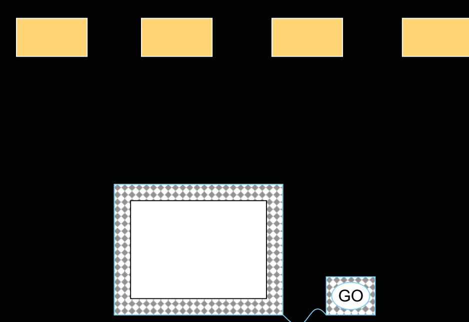
Painkillers: Asprin, Advil, Panadol → -0.8 → Headache

Antidepressants: Prozac, Paxil, Zoloft → +0.7 → Headache

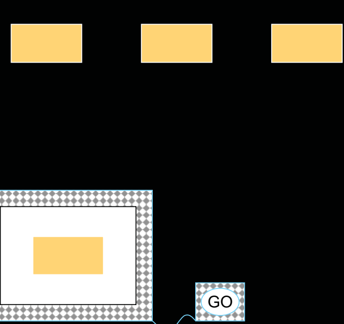
Drug X: 2 trials, no headache (b=0.5).

- A schema, once learned, constrains inferences about new objects, especially when data about the new objects is sparse.

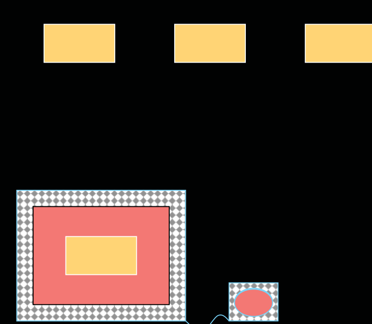
One-shot Learning: Stimuli



One-shot Learning: Stimuli



One-shot Learning: Stimuli



One-shot Learning: Design

- Learning phase:
  - $T1_{0.5}$ & $T2_{0.0}$ condition:

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	5	4	6	1	0	0	0	0
e^- :	10	5	6	4	0	10	10	10	1

  - $T1_{0.9}$ & $T2_{0.1}$ condition:

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	9	8	9	1	1	2	1	0
e^- :	10	1	2	1	0	9	8	9	1

  - $T1_0$ condition:

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6
e^+ :	0	0	0	0	0	0	0
e^- :	10	10	10	10	10	10	10

- Transfer phase: new object o_+ activates machine on one observed trial.
- Causal strength judgment.

One-shot Learning: Results

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	5	4	6	1	0	0	0	0
e^- :	10	5	6	4	0	10	10	10	1

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	9	8	9	1	1	2	1	0
e^- :	10	1	2	1	0	9	8	9	1

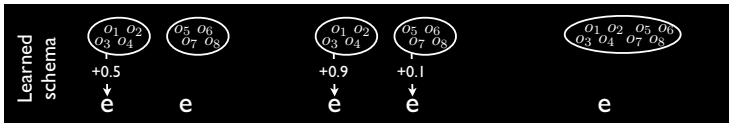
	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6
e^+ :	0	0	0	0	0	0	0
e^- :	10	10	10	10	10	10	10

One-shot Learning: Results

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	5	4	6	1	0	0	0	0
e^- :	10	5	6	4	0	10	10	10	1

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	9	8	9	1	1	2	1	0
e^- :	10	1	2	1	0	9	8	9	1

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6
e^+ :	0	0	0	0	0	0	0
e^- :	10	10	10	10	10	10	10

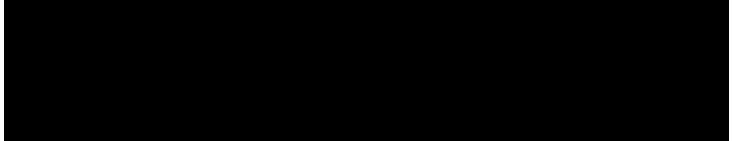


One-shot Learning: Results

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	5	4	6	1	0	0	0	0
e^- :	10	5	6	4	0	10	10	10	1

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	9	8	9	1	1	2	1	0
e^- :	10	1	2	1	0	9	8	9	1

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6
e^+ :	0	0	0	0	0	0	0
e^- :	10	10	10	10	10	10	10

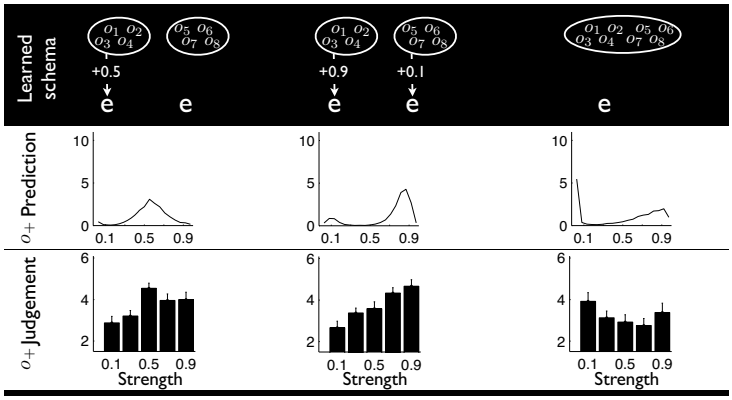


One-shot Learning: Results

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	5	4	6	1	0	0	0	0
e^- :	10	5	6	4	0	10	10	10	1

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8
e^+ :	0	9	8	9	1	1	2	1	0
e^- :	10	1	2	1	0	9	8	9	1

	$\emptyset$	o_1	o_2	o_3	o_4	o_5	o_6
e^+ :	0	0	0	0	0	0	0
e^- :	10	10	10	10	10	10	10



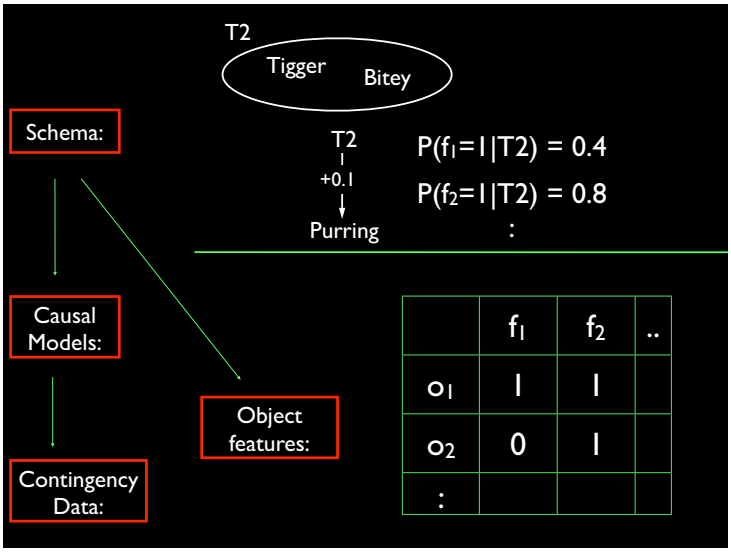
Characteristic Features

- Features of a feral cat (eyesores, scars, anger, etc) give us a hint about its causal type, before we attempt to pet it.
- How can we include knowledge about the characteristic features of members of a causal type?

Characteristic Features

- Features of a feral cat (eyesores, scars, anger, etc) give us a hint about its causal type, before we attempt to pet it.
- How can we include knowledge about the characteristic features of members of a causal type?

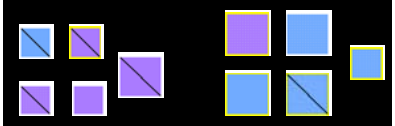
Related work, see: Lien & Cheng (2000), Waldmann & Hagmayer (2006)



Zero-shot Learning: Design

- Objects now had features which were diagnostic of their type.
  - Family-resemblance category structure.
- No trials were shown for test objects.
- Added a free-sort phase.

	$\emptyset$	o_-	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_+
e^+ :	10	0	3	2	1	2	18	18	17	19	0
e^- :	10	0	17	18	19	18	2	2	3	1	0
f_1 :		1	0	0	0	0	0	1	1	1	1
f_2 :		0	1	0	0	0	1	0	1	1	1
f_3 :		0	0	1	0	0	1	1	0	1	1
f_4 :		0	0	0	1	0	1	1	1	0	1
f_5 :		0	0	0	0	1	1	1	1	1	0



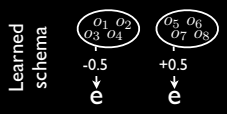
Stimuli adapted from:
Sakamoto and Love (2004)

Zero-shot Learning: Results

	$\emptyset$	o_-	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_+
e^+ :	10	0	3	2	1	2	18	18	17	19	0
e^- :	10	0	17	18	19	18	2	2	3	1	0
f_1 :		1	0	0	0	0	0	1	1	1	1
f_2 :		0	1	0	0	0	1	0	1	1	1
f_3 :		0	0	1	0	0	1	1	0	1	1
f_4 :		0	0	0	1	0	1	1	1	0	1
f_5 :		0	0	0	0	1	1	1	1	1	0

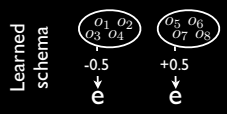
Zero-shot Learning: Results

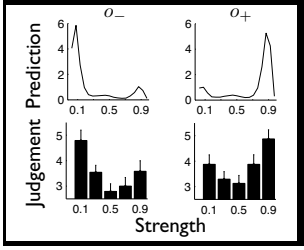
	$\emptyset$	o_-	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_+
e^+ :	10	0	3	2	1	2	18	18	17	19	0
e^- :	10	0	17	18	19	18	2	2	3	1	0
f_1 :		1	0	0	0	0	0	1	1	1	1
f_2 :		0	1	0	0	0	1	0	1	1	1
f_3 :		0	0	1	0	0	1	1	0	1	1
f_4 :		0	0	0	1	0	1	1	1	0	1
f_5 :		0	0	0	0	1	1	1	1	1	0



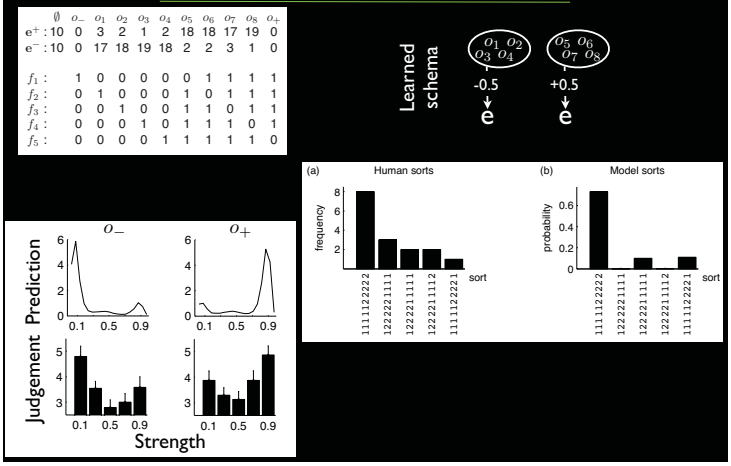
Zero-shot Learning: Results

	$\emptyset$	o_-	o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_+
e^+ :	10	0	3	2	1	2	18	18	17	19	0
e^- :	10	0	17	18	19	18	2	2	3	1	0
f_1 :		1	0	0	0	0	0	1	1	1	1
f_2 :		0	1	0	0	0	1	0	1	1	1
f_3 :		0	0	1	0	0	1	1	0	1	1
f_4 :		0	0	0	1	0	1	1	1	0	1
f_5 :		0	0	0	0	1	1	1	1	1	0

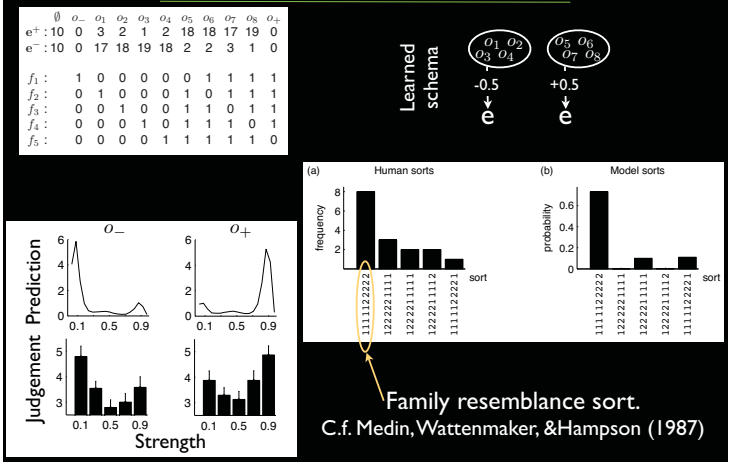




Zero-shot Learning: Results



Zero-shot Learning: Results



Family resemblance sort.
C.f. Medin, Wattenmaker, &Hampson (1987)

Discussion

- Causal schemata (object types, their causal powers and characteristic features) represent abstract causal knowledge.
  - They are rapidly learned, and used to constrain further inference.
- Preliminary evidence that children learn causal schemata quickly and robustly (Schulz, Goodman, Tenenbaum, Jenkins).

Discussion

- Some important directions:
  - More empirical work.
  - Interactions and functional form.
  - Forces, substances, etc.
  - Richer (logical?) representation.
- Where do the variables come from?

MIT

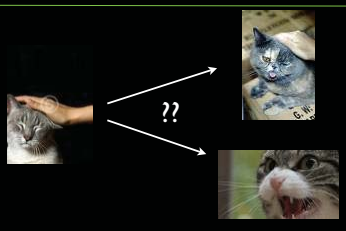
ndg@mit.edu

Part II: Grounded Causal Models

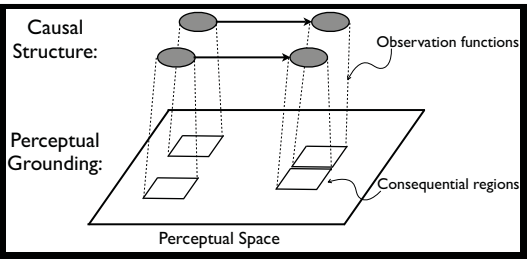
Noah D. Goodman
Computational Cognitive Science Group, MIT

with Vikash Mansinghka and Josh Tenenbaum

Causal Variables



- What are the origins of the variables (like petting) over which causal links are defined? How are they formed from, and related to, perceptual experience?



- A *Grounded Causal Model* consists of:
  - a set of (abstract) variables,
  - an observation function for each variable mapping percepts to states of the variable,
  - a causal Bayesian* network structure relating the variables.

*Or other relational structure.

Learning GrCMs

Some options:

- Variables are innate.
- Bottom-up: 'clusters-then-causes'.
- Learn variables and structure together.

Acquisition model:

Learning GrCMs

Some options:

- Variables are innate.
- Bottom-up: 'clusters-then-causes'.
- Learn variables and structure together.

Acquisition model:

Causal contingency learning.

Learning GrCMs

Some options:

- Variables are innate.
- Bottom-up: 'clusters-then-causes'.
- Learn variables and structure together.

Acquisition model:

Causal contingency learning.

Cluster percepts into variables, then causal contingency learning.

Learning GrCMs

Some options:

- Variables are innate.
- Bottom-up: 'clusters-then-causes'.
- Learn variables and structure together.

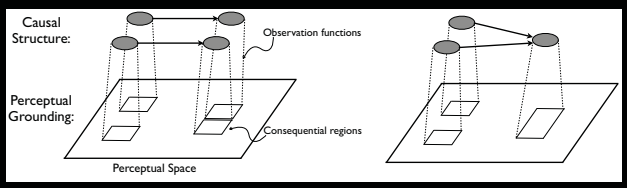
Acquisition model:

Causal contingency learning.

Cluster percepts into variables, then causal contingency learning.

???

Learning GrCMs



- The petting/pounding example suggests that causal information is crucial for variable formation....
- But causal structure between variables can't be known before the variables....
- This is a chicken-and-egg problem!

Learning GrCMs

- We want the joint posterior probability of number of variables, their observation functions, and causal structure.
- Assume (for simplicity) uniform prior probabilities.
- The posterior probability of a GrCM is proportional to the likelihood of a sequence of percepts given that GrCM:

N : Number of Vars. I : Interventions
C : Causal Str. $\mathbf{w}$: Percepts
f : Obs. Fns.

$$P(N, C, f|\mathbf{w}; I) \propto P(\mathbf{w}|N, C, f; I)$$

Learning GrCMs

- To build a likelihood, assume:
  - The observation function of each (binary) variable is given by a *consequential region* in perceptual space.
  - Percepts occur uniformly in the region of perceptual space cut out by active variables.
  - The causal structure is given by a directed graph: a variable is *active* at time *t* if any of its parents was active at time *t-1* (and there's a small chance that any variable flips state).

Learning GrCMs

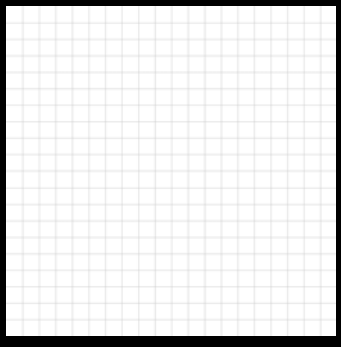
- Formal details:
  - Each state depends only on previous, add power-law decay:
    $$P(f, C|\mathbf{w}, I) \propto \prod_{t=0}^{T} P(w_t|s_{t-1}, C, f, I)^{(T-t)^{-\gamma}}$$
  - State is determined by percept (through observation function), probability of percept is inversely proportional to the size of the consequential region of this observed state:
    $$P(w_t|s^{\text{ob}}_{t-1}, C, f, I) = \frac{1}{|R_{s^{\text{ob}}_t}|} \prod_{i=1}^{N} P(s^{\text{ob}}_{i,t}|s^{\text{ob}}_{t-1}, C, I)$$
  - Nearly-deterministic-or causal structure for $P(s^{\text{ob}}_{i,t}|s^{\text{ob}}_{t-1}, C, I)$.

Experiment

- Goals:
  - See if people can learn GrCMs from the results of their own interventions, in a simple setting,
  - See if people succeed in conditions where a purely bottom-up learner should fail,
  - Test the Bayesian model.

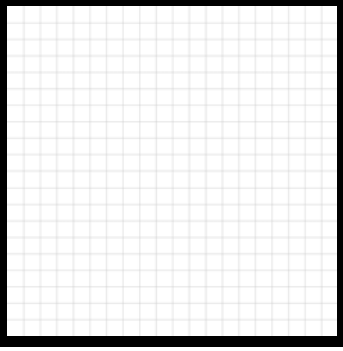
Design

- 'Alien panels': perceptual space of dots in a rectangle, variables are 'invisible buttons'.



Design

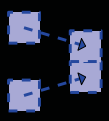
- 'Alien panels': perceptual space of dots in a rectangle, variables are 'invisible buttons'.



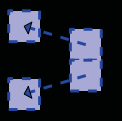
Design

- Three conditions (within subjects): structures a, b, c.
- Each participant was stopped every ~30 clicks (five times in each condition) and asked to "describe what's going on" using a simple drawing tool.

Structure a:

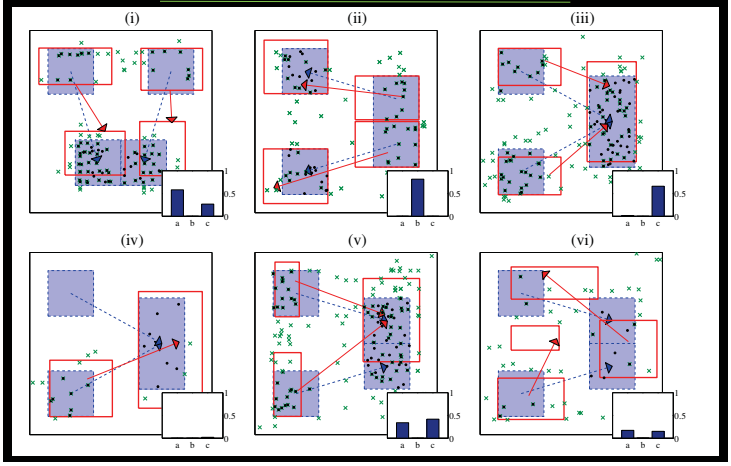


Structure b:



Structure c:

Results: Typical Responses

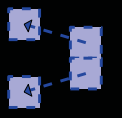


Data analysis

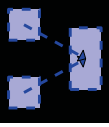
- To enable group analyses, participants' responses were coded as structure a, b, c, or 'other'.
- Model posterior probability for each response, conditioned on the specific evidence available for that response, was calculated over a set of 'reasonable' structures, including a, b, and c.

Structure a:

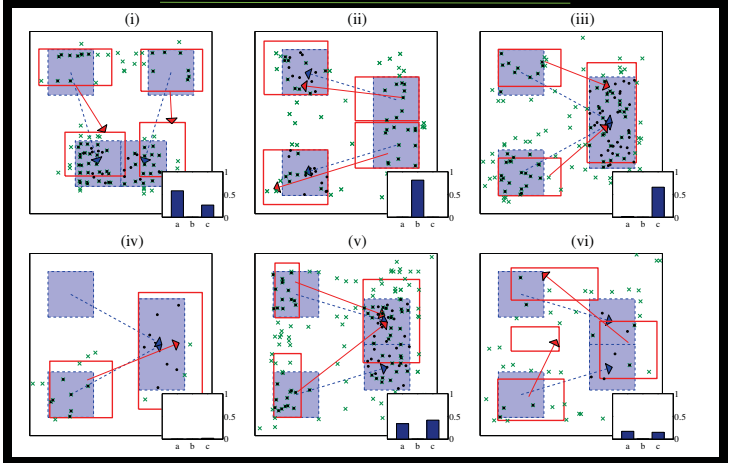
Structure b:



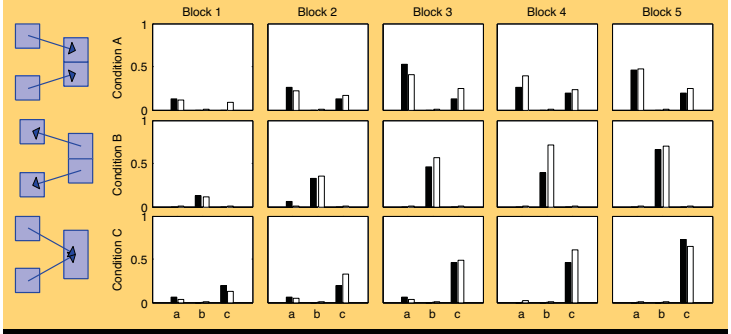
Structure c:



Results: Typical Responses

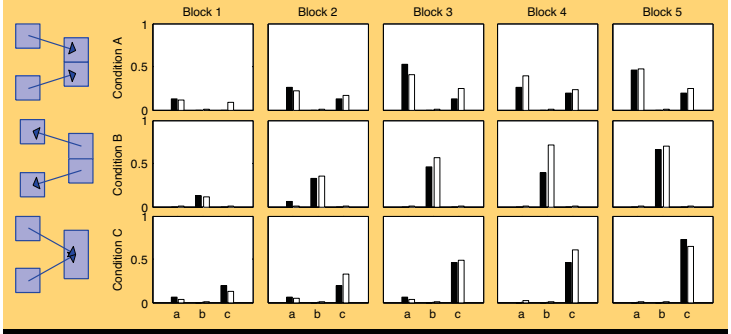


Results: Mean Responses



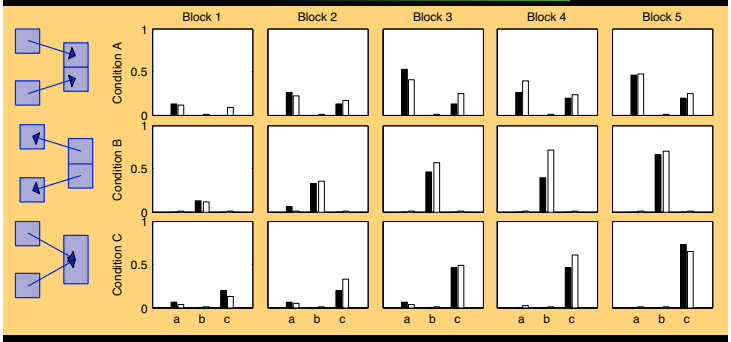
Black bars: human proportion of responses for each structure.
White bars: model posterior.

Results: Mean Responses



- Participants distinguished the three conditions,
- Participants learned the correct GCM overall,
- There were errors -- predicted by model?

Results: Mean Responses



- Yes, model predicts group means qualitatively,
- and quantitatively: $r = 0.95$
  (two free parameters)

Results: Predicting Errors

- The model posterior probability of the correct structure was significantly higher when participants made correct responses than incorrect (Mann-Whitney U, $p<0.0001$).
- This indicates that many human errors were 'rational errors': reasonable responses to the available evidence.

Discussion

- Where do observable variables come from?
- They are learned.
  - The two factors of the 'meaning' of a variable, observational grounding and causal relations, are learned together and are mutually constraining.
- Still many open questions about this idea....

Discussion

- Some directions:
  - More detailed experiments.
  - Scaling up (computationally and empirically).
  - Grounding interventions in action.
  - Prior knowledge about observation functions: causal affordances.
  - Integrating with abstract knowledge and object concepts.

Conclusion

- Causal structure is a primary tool used by the mind to tame the river of experience.
- Causal knowledge grounds in perception and exists at multiple levels of abstraction, with a rich ontology:
  - Observable variables, causal relations, causal-types of objects, etc.
- This knowledge can be learned from experience. In learning, each component/ level constrains the others.