Outline

- Theory-based Bayesian causal induction
- On intuitive theories: their structure, function, and origins

Domain theory generates hypothesis space of causal models

- Ontology of types and predicates.
  - What is there?
- Constraints on causal relations between predicates.
  - What can/must/is likely to cause what?
- Functional forms of causal relations.
  - How does an effect depend functionally on its causes?

Domain theory generates hypothesis space of causal models

- Ontology of types and predicates.
  - What is there? Nodes.
- Constraints on causal relations between predicates.
  - What can/must/is likely to cause what? Edges.
- Functional forms of causal relations.
  - How does an effect depend functionally on its causes? Parameterizations and parameters.

Theories as probabilistic logic

- **Ontology**
  - **Types:** Block, Detector, Trial
  - **Predicates:**

    Contact(Block, Detector, Trial)

    Active(Detector, Trial)
- **Constraints on causal relations**
  - For any Block *b* and Detector *d*, with probability *q* : Cause(Contact(*b*,*d*,*t*), Active(*d*,*t*))
- **Functional form of causal relations**
  - Causes of Active(*d*,*t*) are independent mechanisms, with causal strengths w_i. A background cause has strength w_0. Assume a near-deterministic mechanism: $w_i \sim 1$, $w_0 \sim 0$.

Bayesian inference

- Evaluating causal network hypotheses in light of data:

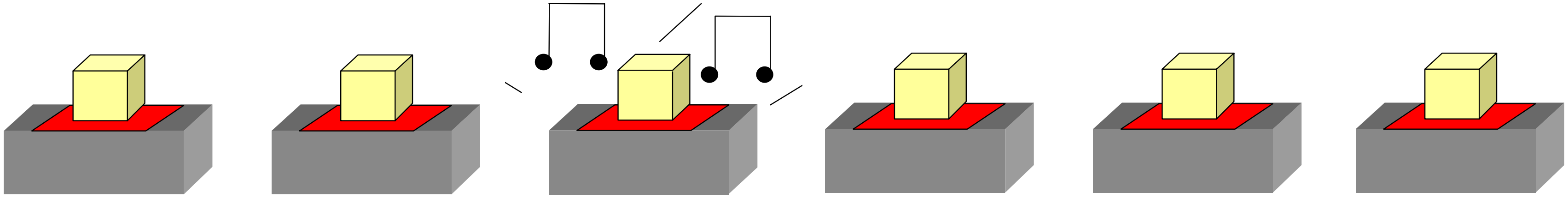
$$P(h_i \mid d) = \frac{P(d \mid h_i)P(h_i)}{\sum_{h_j \in H} P(d \mid h_j)P(h_j)}$$

- Inferring a particular causal relation:

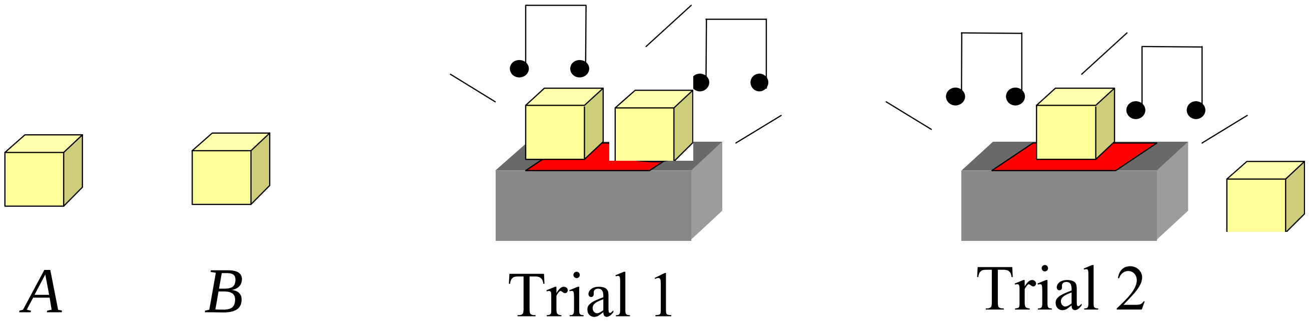
$$P(A \rightarrow E \mid d) = \sum_{h_j \in H} P(A \rightarrow E \mid h_j)P(h_j \mid d)$$

Manipulating the prior

I. Pre-training phase: Blickets are rare . . . .



II. Backwards blocking phase:



After each trial, adults judge the probability that each object is a blicket.

Rare condition

Common condition

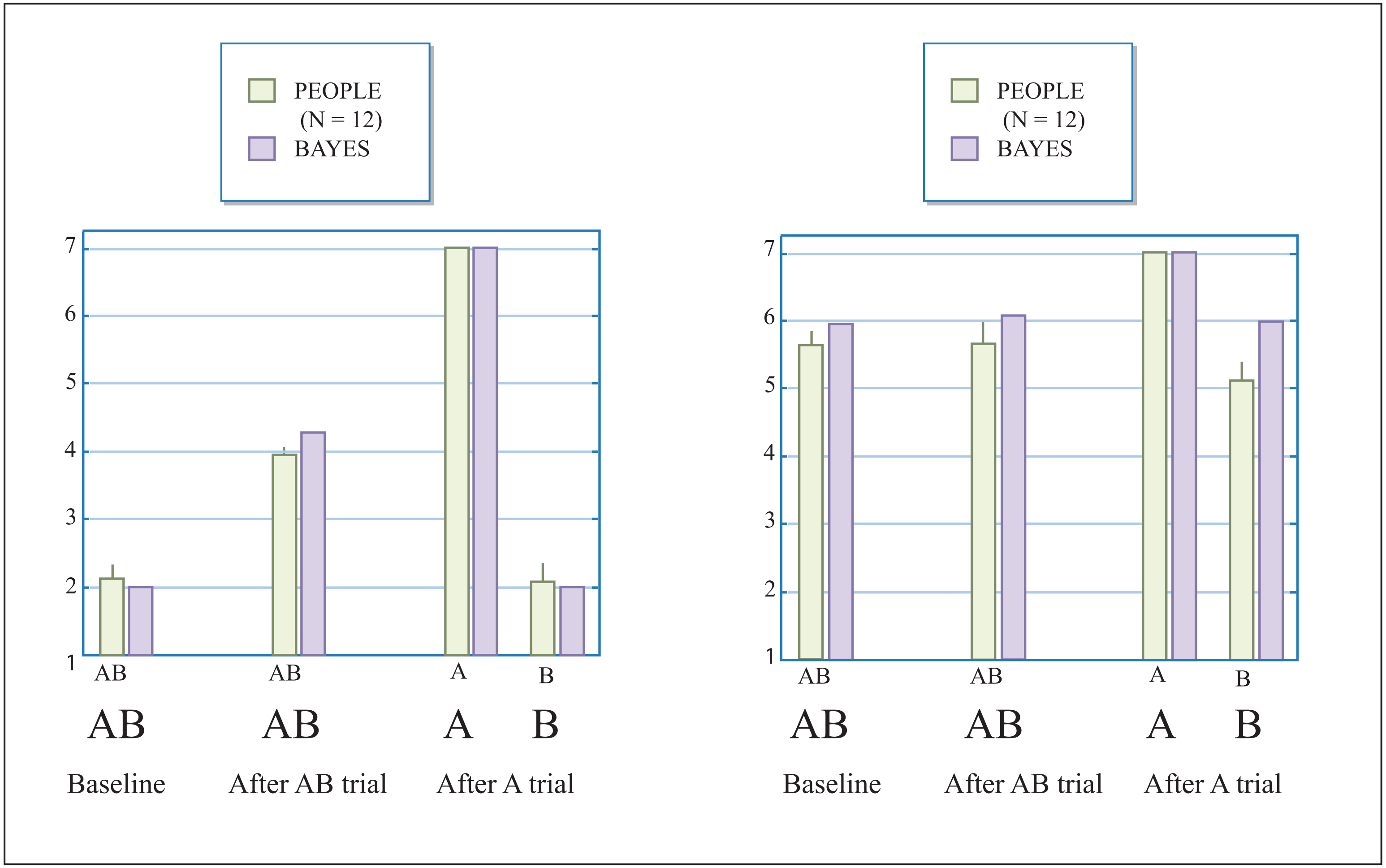


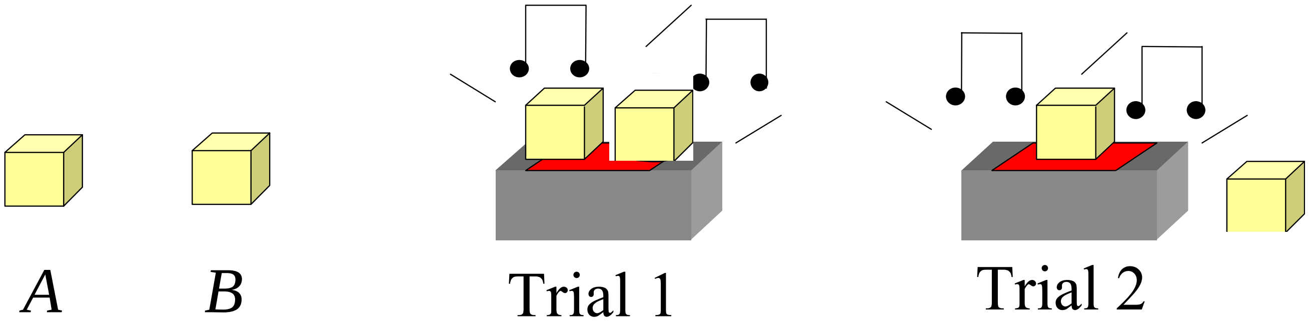
Figure by MIT OCW.

Manipulating the priors of 4-year-olds

(Sobel, Tenenbaum & Gopnik, 2004)

I. Pre-training phase: Blickets are rare.

II. Backwards blocking phase:



Rare condition:

A: 100% say “a blicket”

B: 25% say “a blicket”

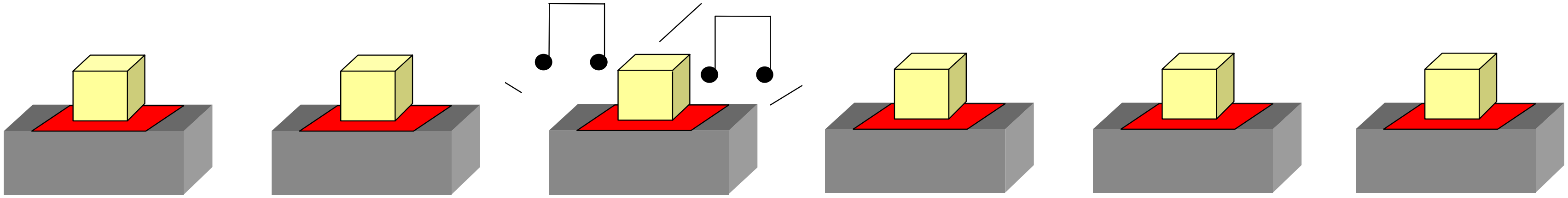
Common condition:

A: 100% say “a blicket”

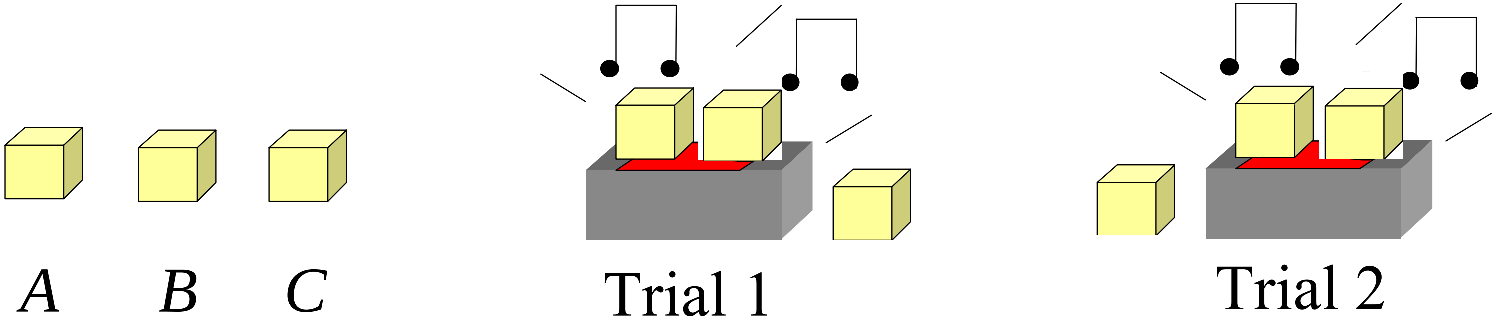
B: 81% say “a blicket”

Inferences from ambiguous data

I. Pre-training phase: Blickets are rare . . . .



II. Two trials: A B → detector, B C → detector

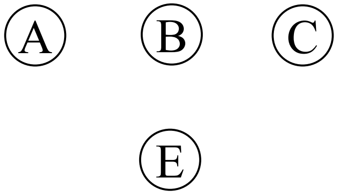


After each trial, adults judge the probability that each object is a blicket.

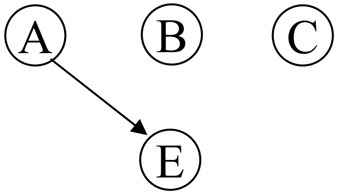
Same domain theory generates hypothesis space for 3 objects:

- Hypotheses:

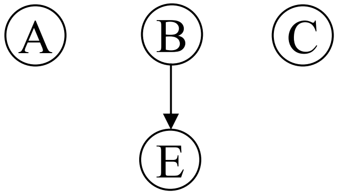
$h_{000} =$



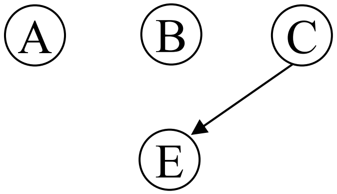
$h_{100} =$



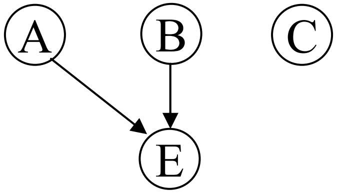
$h_{010} =$



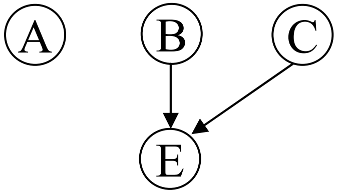
$h_{001} =$



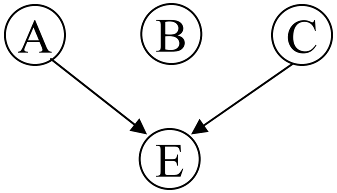
$h_{110} =$



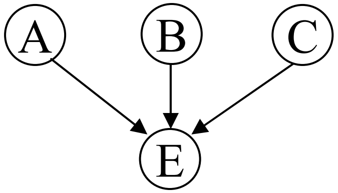
$h_{011} =$



$h_{101} =$



$h_{111} =$



- Likelihoods: $P(E{=}1|\, A, B, C; h) = 1$ if $A = 1$ and $A \longrightarrow E$ exists,
  or $B = 1$ and $B \longrightarrow E$ exists,
  or $C = 1$ and $C \longrightarrow E$ exists,
  else 0.

- “Rare” condition: First observe 12 objects on detector, of which 2 set it off.

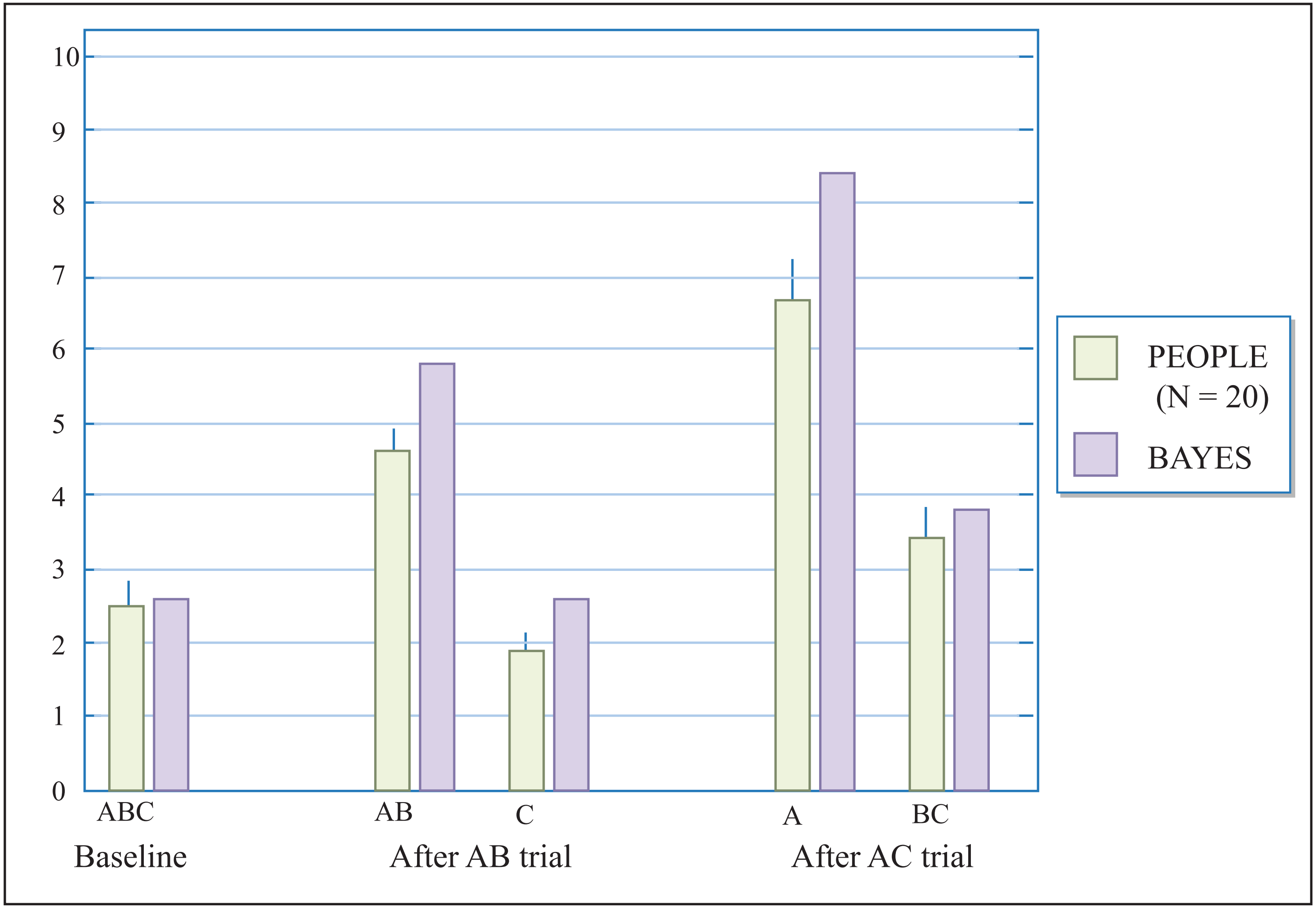
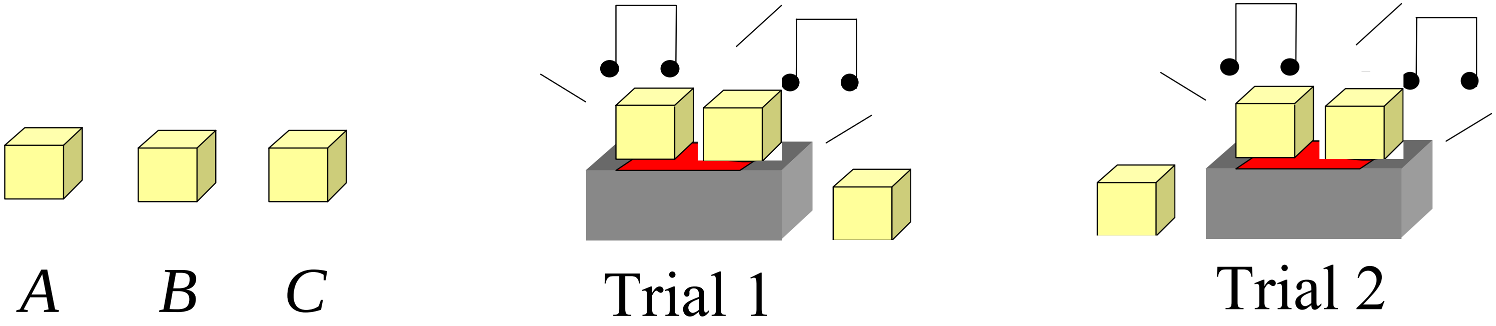


Figure by MIT OCW.

Ambiguous data with 4-year-olds

I. Pre-training phase: Blickets are rare.

II. Two trials: A B → detector, B C → detector



Final judgments:

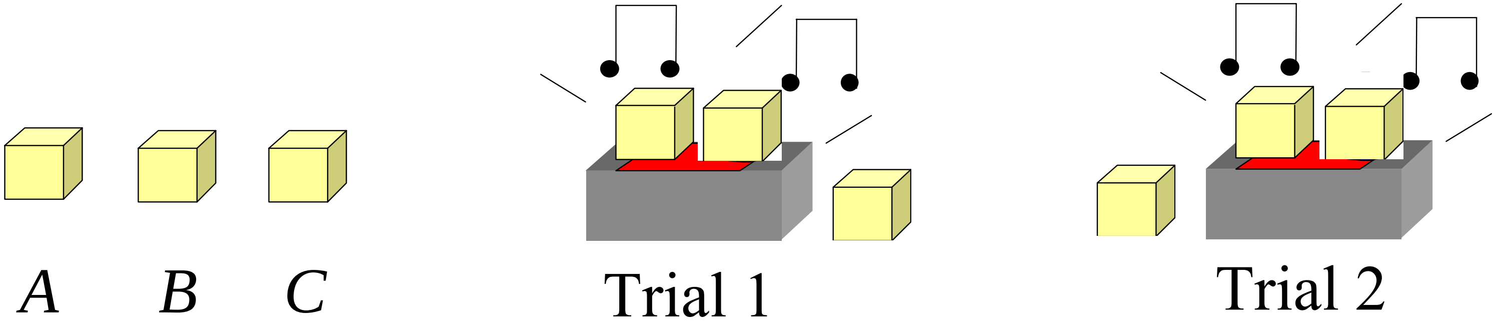
A: 87% say “a blicket”

B or *C*: 56% say “a blicket”

Ambiguous data with 4-year-olds

I. Pre-training phase: Blickets are rare.

II. Two trials: A B → detector, B C → detector



Final judgments:

A: 87% say “a blicket”

B or *C*: 56% say “a blicket”

Backwards blocking (rare)

A: 100% say “a blicket”

B: 25% say “a blicket”

The role of causal mechanism knowledge

- Is mechanism knowledge necessary?
  - Constraint-based learning using χ^2 tests of conditional independence.

- How important is the deterministic functional form of causal relations?
  - Bayes with “probabilistic independent generative causes” theory (i.e., noisy-OR parameterization with unknown strength parameters; c.f., Cheng’s causal power).

Bayes with correct theory:

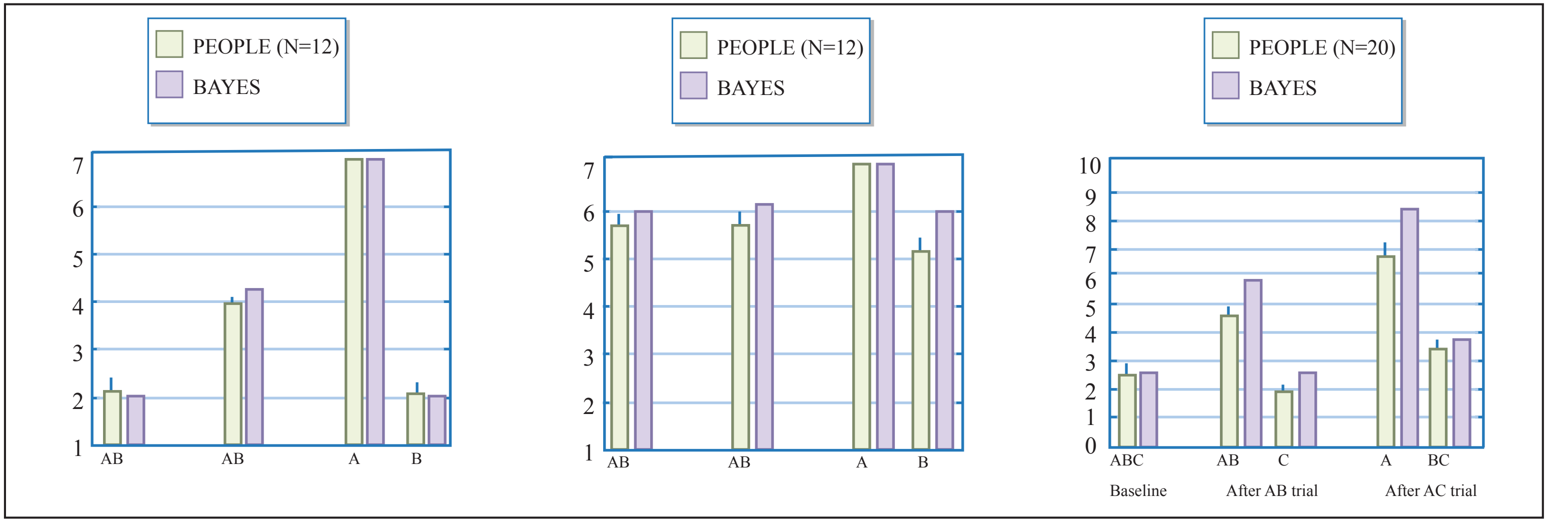


Figure by MIT OCW.

Independence test with fictional sample sizes:

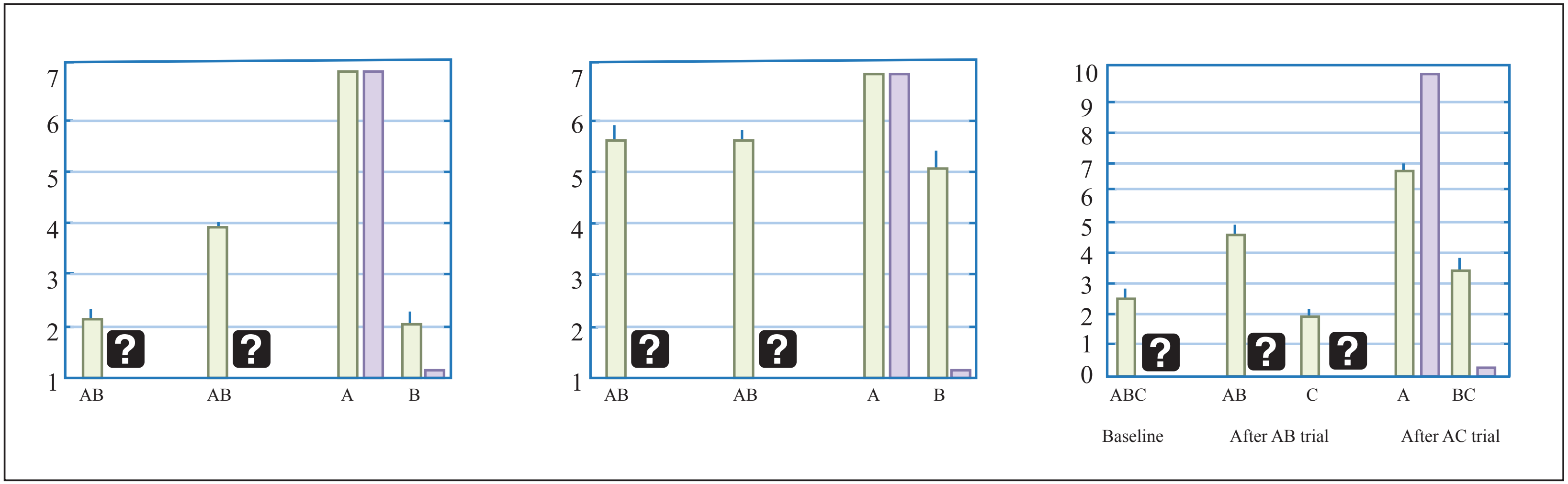


Figure by MIT OCW.

Bayes with correct theory:

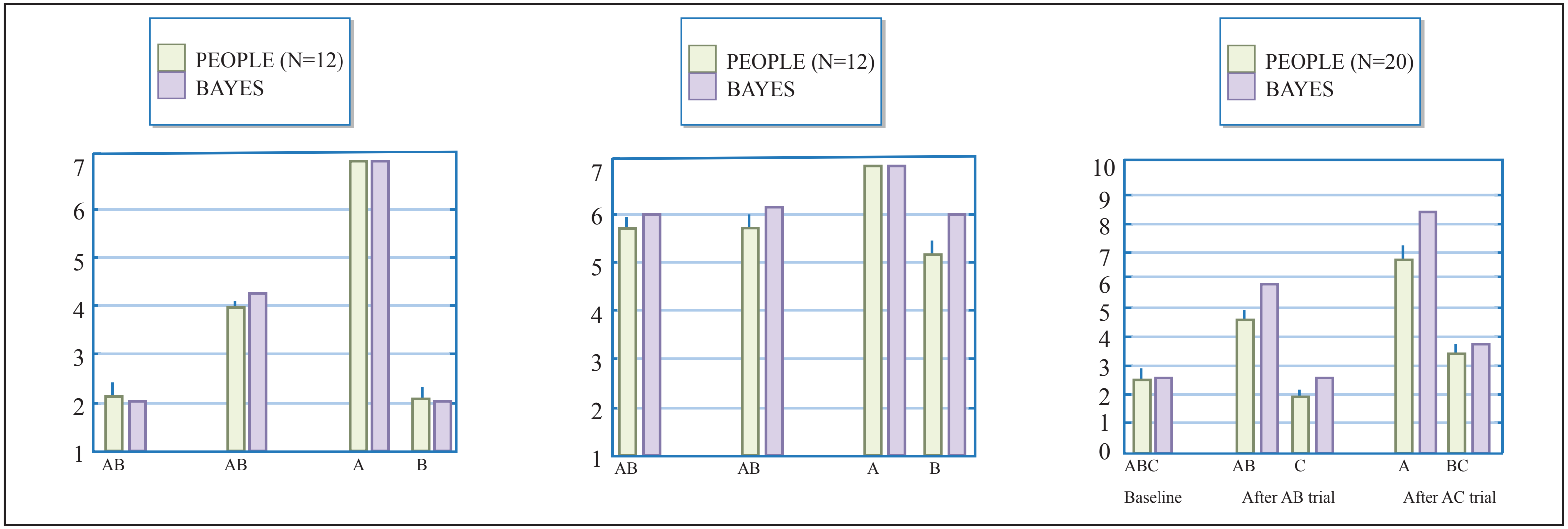


Figure by MIT OCW.

Bayes with "noisy sufficient causes" theory:

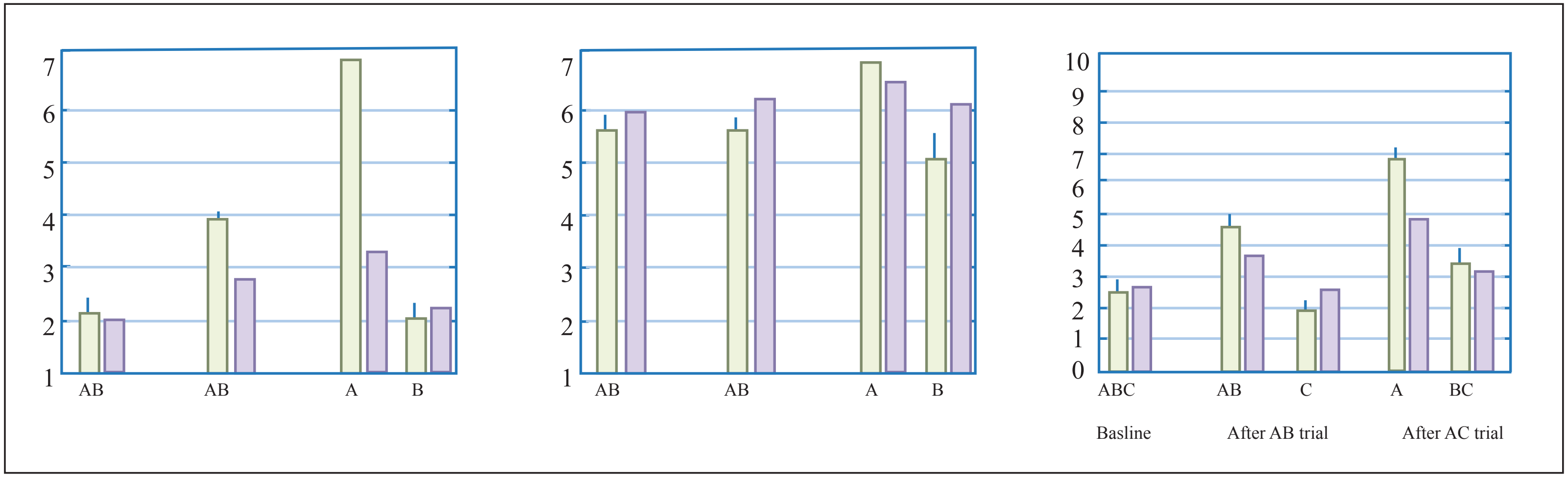


Figure by MIT OCW.

Blicket studies: summary

- Theory-based Bayesian approach explains one-shot causal inferences in physical systems.
- Captures a spectrum of inference:
  - Unambiguous data: adults and children make all-or-none inferences
  - Ambiguous data: adults and children make more graded inferences
- Extends to more complex cases with hidden variables, dynamic systems, ….

Learning a probabilistic causal relation

Given a random sample of mice:

	Injected with X	Not injected with X
Expressed Y	6	3
Did not express Y	2	5

- "To what extent does chemical X cause gene Y to be expressed?"
- Or, "What is the probability that X causes Y?"

Theory

- **Ontology**
  - **Types:** Chemical, Gene, Mouse
  - **Predicates:**

    Injected(Mouse, Chemical)

    Expressed(Mouse, Gene)
- **Constraints on causal relations**
  - For any Chemical *c* and Gene *g*, with prior probability *q* :

    Cause(Injected(*m*,*c*), Expressed(*m*,*g*))
- **Functional form of causal relations**
  - Causes of Expressed(*m*,*g*) are independent probabilistic mechanisms, with causal strengths w_i. An independent background cause is always present with strength w_0.

Judging the probability that C → E

(Buehner & Cheng, 1997; 2003)

Image removed due to copyright considerations.

Parameter estimation models

Image removed due to
copyright considerations.

How important is the theory?

Image removed due to copyright considerations.

Learning causal structure without causal mechanism knowledge

- Constraint-based: χ^2 test of independence.

Image removed due to copyright considerations.

- "Mechanism-free" Bayes: no constraints on the functional form of $P(E|B,C)$

Image removed due to copyright considerations.

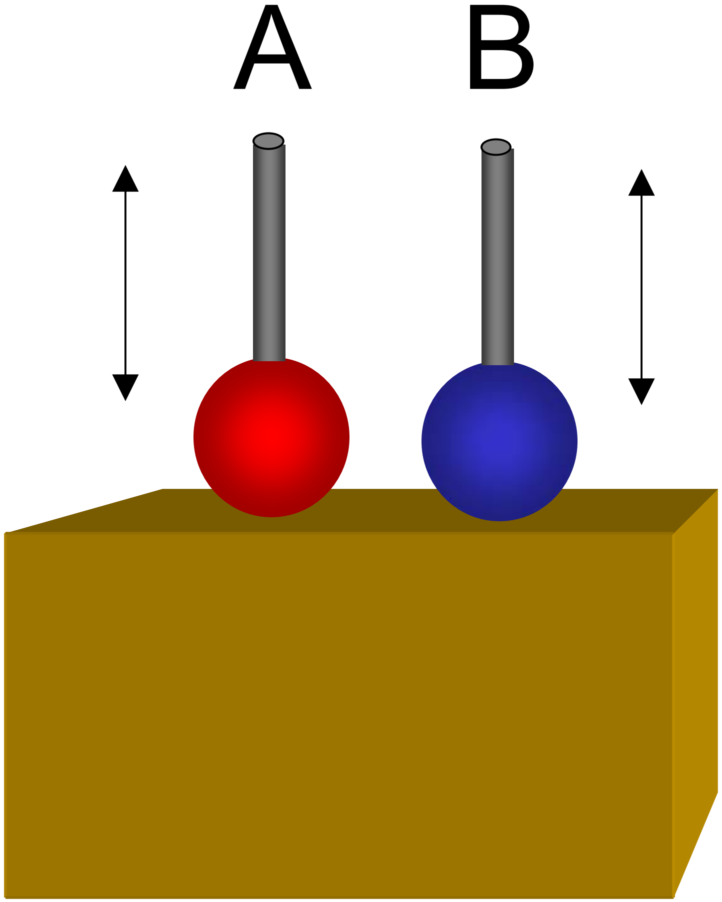
Data for inhibitory causes

Image removed due to copyright considerations.

Probabilistic causal relations: summary

- Much weaker theory: causes may have any degree of strength.
- Thus, inferences about causal structure still graded after many observations.

The stick-ball machine



T. Kushnir, A. Gopnik, L Schulz, and D. Danks. "Inferring Hidden Causes." *Proceedings of the Twenty-Fifth Annual Meeting of the Cognitive Science Society*. Boston, MA: Cognitive Science Society, 2003, pp. 699-703.

Inferring hidden causal structure

Common unobserved cause

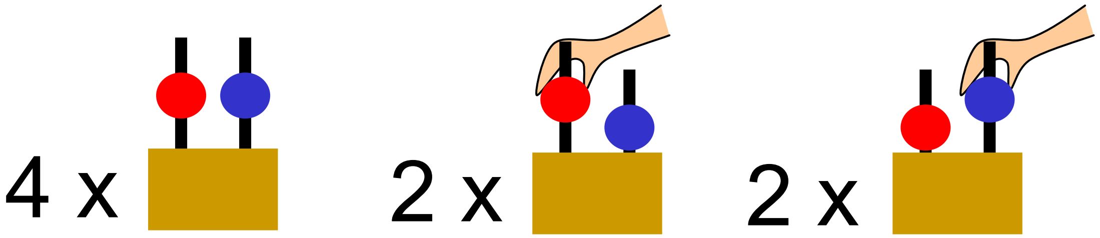
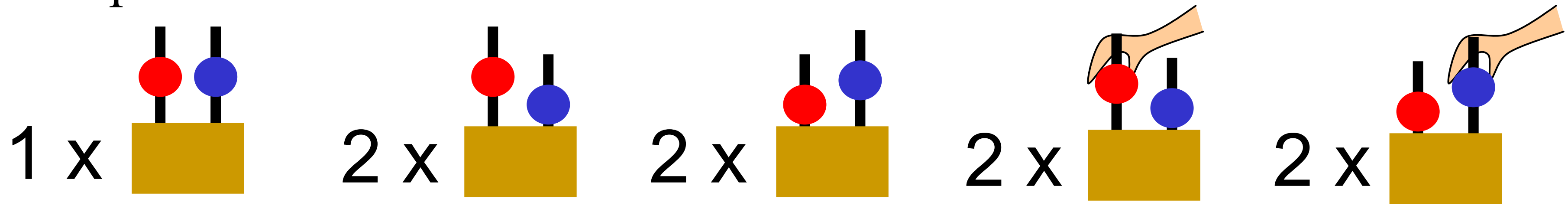
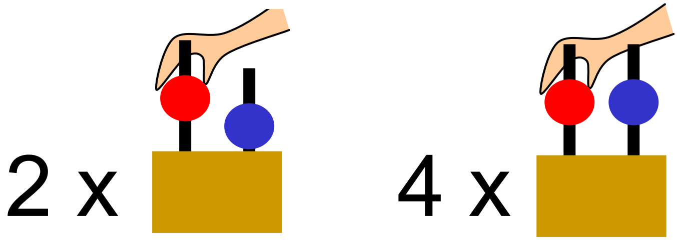


Image removed due to copyright considerations. Please see:
T. Kushnir, A. Gopnik, L Schulz, and D. Danks. "Inferring Hidden Causes." *Proceedings of the Twenty-Fifth Annual Meeting of the Cognitive Science Society*. Boston, MA: Cognitive Science Society, 2003, pp. 699-703.

Independent unobserved causes



One observed cause



Theory

- **Ontology**
  - **Types:** Ball, Source, Trial
  - **Predicates:**

    Moves(Ball, Trial)

    Moves(Source, Trial)
- **Constraints on causal relations**
  - Possible for any Ball or Source *x* and any Ball *b*: Cause(Moves(*x*,*t*), Moves(*b*,*t*))
- **Functional form of causal relations**
  - Causes of Moves(*b*,*t*) are independent mechanisms, with causal strengths β.
  - If Moves(*x*,*t*) has no causes, it occurs with probability α.

Image removed due to copyright considerations.

c.f., infinite mixture model

Image removed due to copyright considerations.

Modeling bi-directional influences

Image removed due to copyright considerations. Please see:

T. Kushnir, A. Gopnik, L Schulz, and D. Danks. “Inferring Hidden Causes.” In *Proceedings of the Twenty-Fifth Annual Meeting of the Cognitive Science Society*. Boston, MA: Cognitive Science Society, 2003, pp. 699-703.

Common unobserved cause

Image removed due to copyright considerations.

Independent unobserved causes

Image removed due to copyright considerations.

One observed cause

Image removed due to copyright considerations.

Image removed due to copyright considerations. Please see: T. Kushnir, A. Gopnik, L Schulz, and D. Danks. “Inferring HiddenCauses.” In *Proceedings of the Twenty-Fifth Annual Meeting of the Cognitive Science Society*. Boston: Cognitive Science Society, 2003, pp. 699-703.

$\alpha = 0.3$

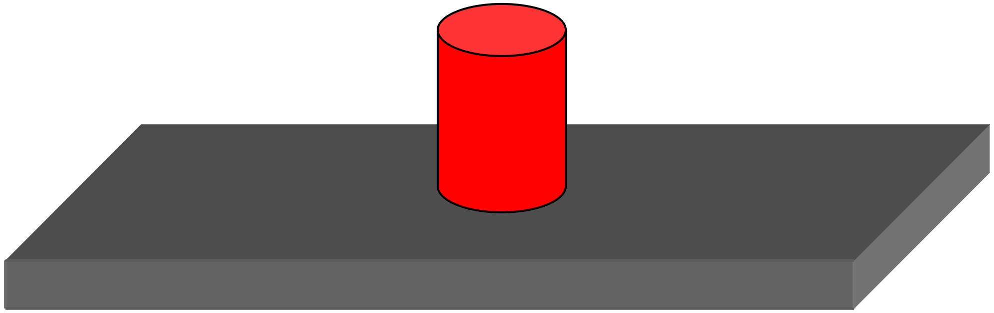
$\omega = 0.8$

$r = 0.94$

Stick-ball studies: summary

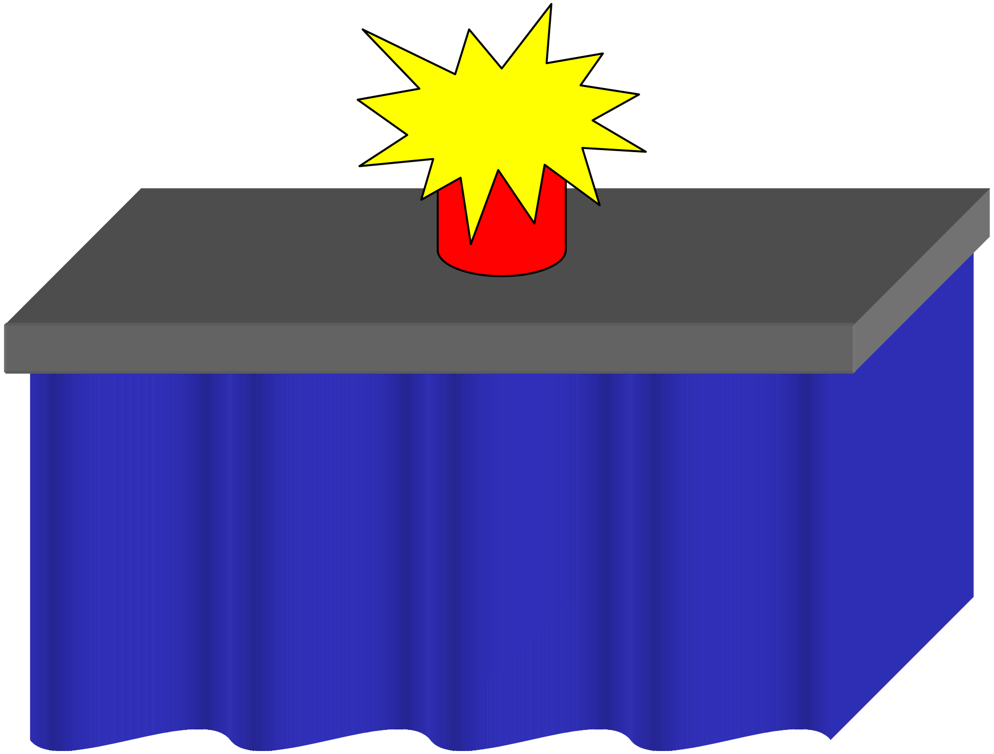
- More flexible theory: causal mechanisms are not deterministic, hidden causes may exist.
- Thus, inferences require more data, and are not quite all-or-none.

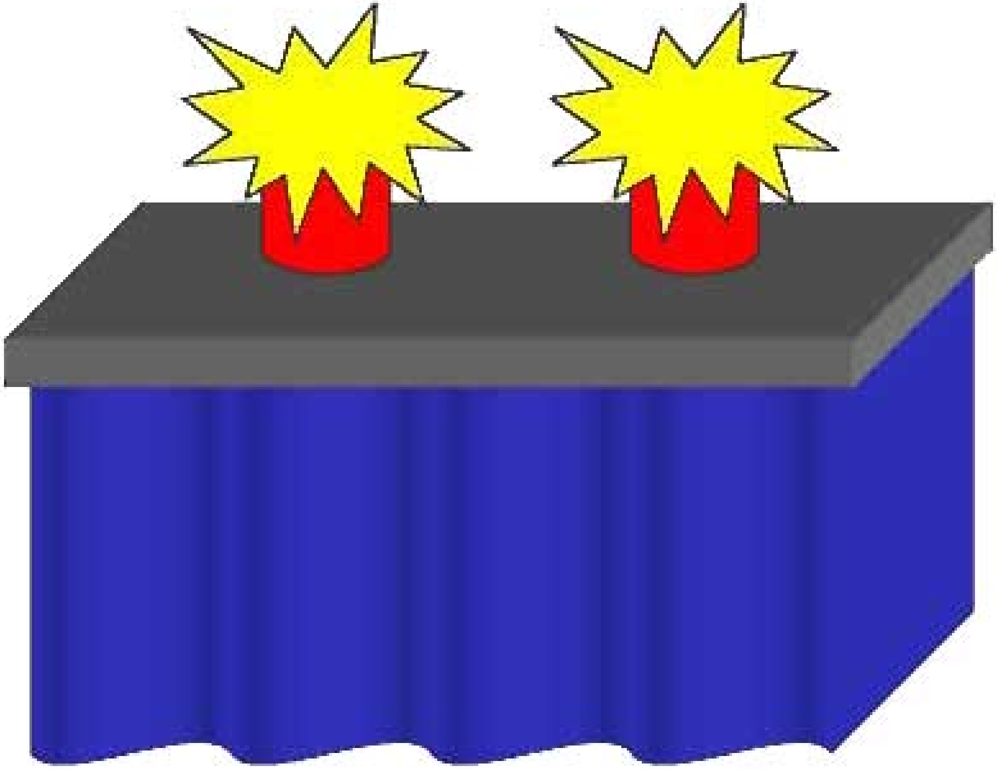
Nitro X

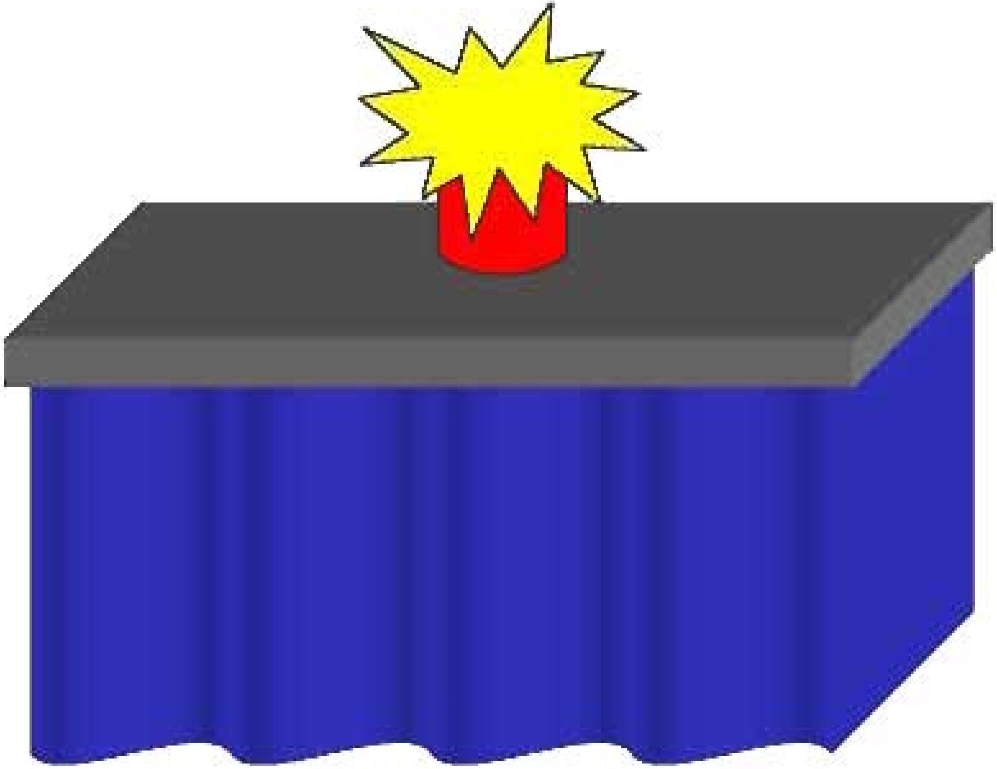


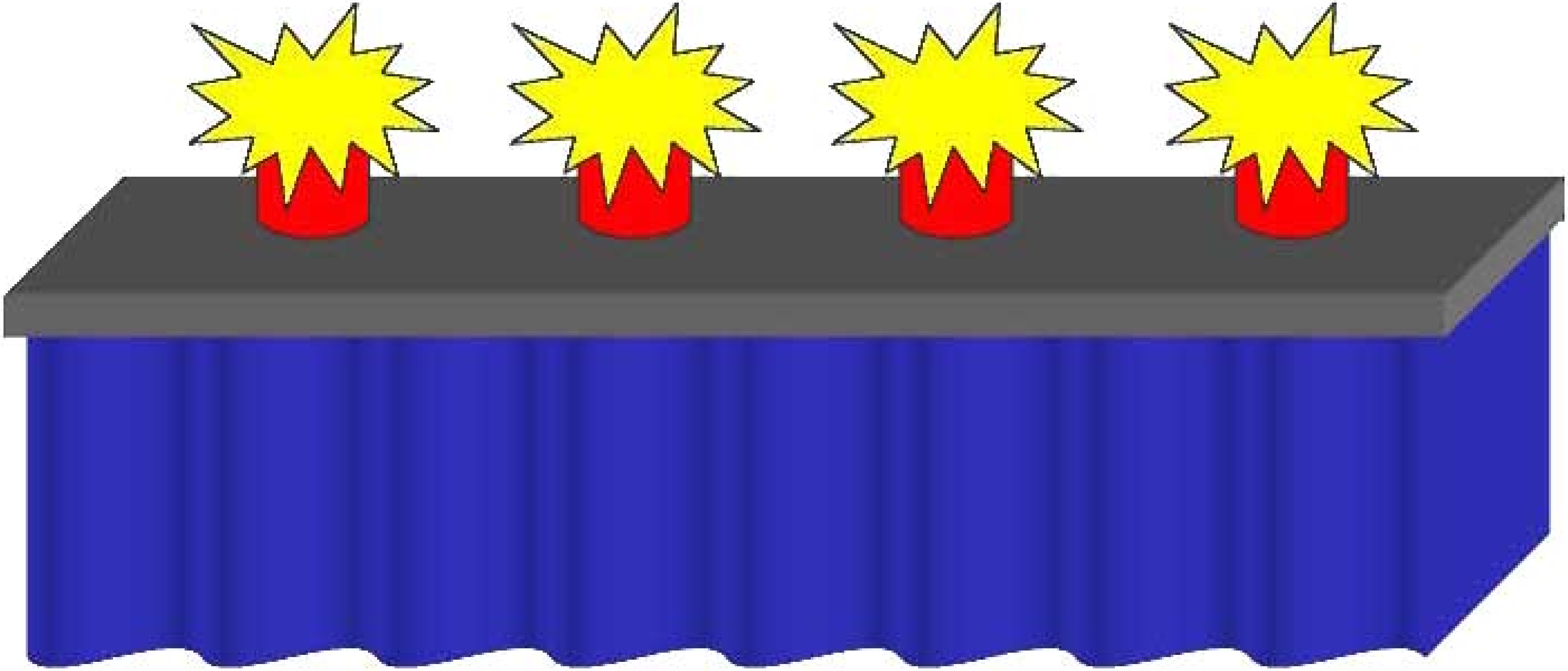
- More extreme test of ability to infer hidden causes
  - – single datapoint
  - – no mention of hidden cause in instructions
- More sophisticated physical theory
- Importance of *statistical* inference

Nitro X



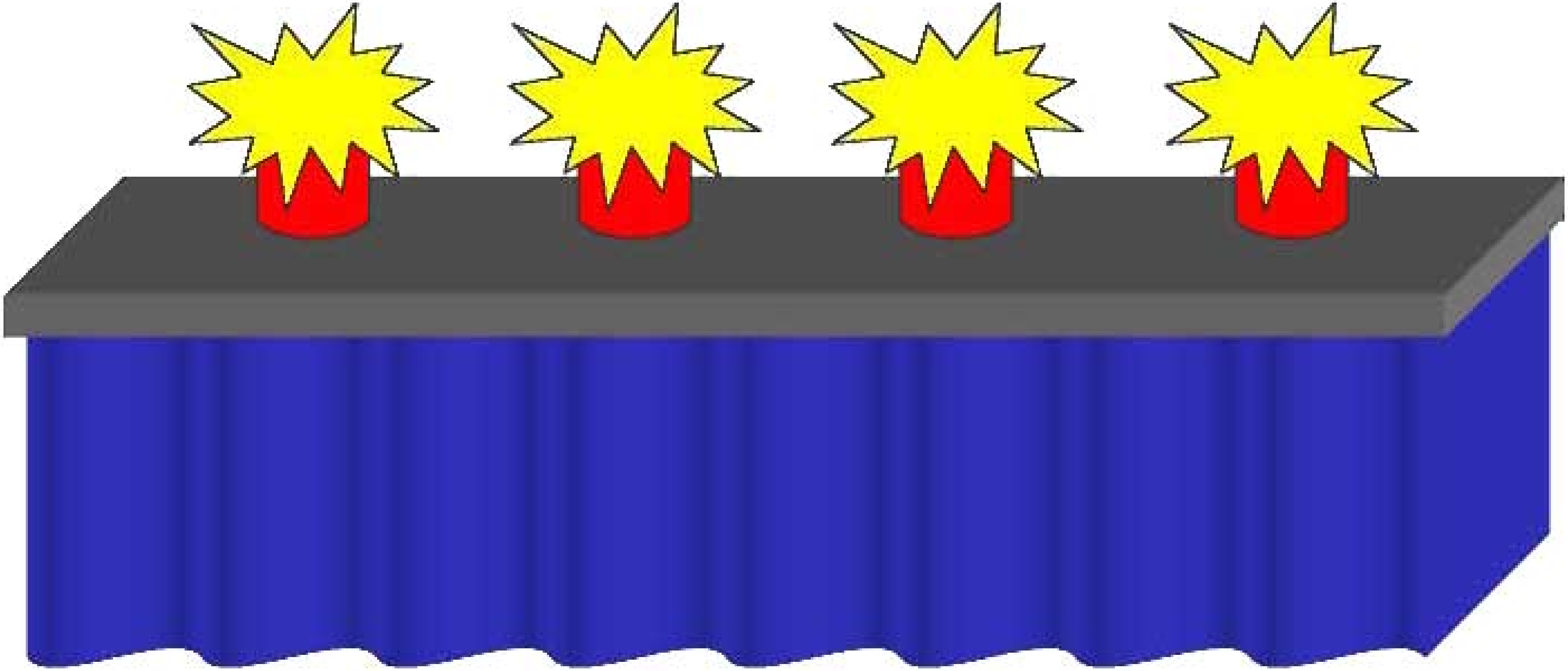






Test trials

- Show explosions involving multiple cans
  - allows inferences about causal structure
- For each trial, choose one of:
  - chain reaction
  - spontaneous explosions
  - other
- No explicit reference to a hidden cause



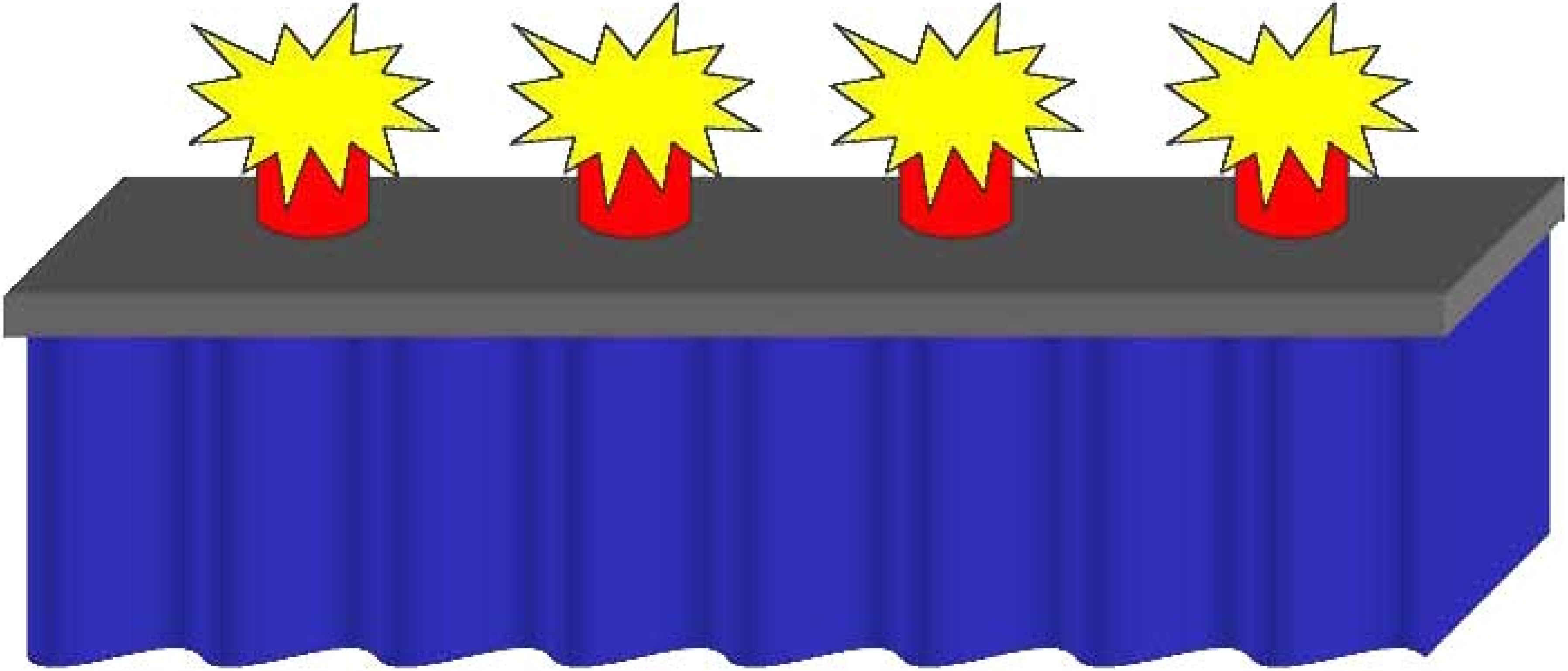


Image removed due to copyright considerations.

c.f., infinite mixture model

Image removed due to copyright considerations.

People are sensitive to statistical evidence for a hidden cause

Nitro-X studies: summary

- Perceptual inferences about hidden causal structure can be explained in the same theory-based Bayesian framework.
  - Perceptual settings are information-rich: a single observation is often sufficient to provide strong evidence for a hidden cause.

- Some questions:
  - What do we gain by describing causal reasoning, causal learning, and causal perception in the same framework?
  - Are there any kinds of causal inferences that can’t be explained in this framework?

Theories + Bayes in causal induction

- Theory is a logical structure that generates a space of probabilistic causal models -- the hypothesis space for Bayesian inference.
- Parallel with theory-based Bayesian models for learning about concepts and properties (c.f., Ta' grammars for domain structures).
    - *Something missing: learning the theory?*
- General theme: merging of two major traditions in cog sci and AI, previously seen as opposing:
    - structured symbolic representations
    - probabilistic inference and statistical learning

Outline

- Theory-based Bayesian causal induction
- On intuitive theories: their structure, function, and origins

Two lines of questioning

- How do we infer networks of causal relations?
  - Given only observational data (correlation)?
  - Given just a few observations?
  - In the presence of hidden causes?
  - With dynamic systems?

- How to characterize intuitive causal theories? (e.g., theories of physics, biology, mind)
  - What is the content of causal theories?
  - How are they used in causal inference?
  - How are the theories themselves learned?

Some background on theories

- A primary motivation comes from cognitive development. *(But related to learning in AI...)*
- The big question: what develops from childhood to adulthood?
  - One extreme: basically everything
    - Totally new ways of thinking, e.g. logical thinking.
  - The other extreme: basically nothing
    - Just new facts (specific beliefs), e.g., trees can die.
  - Intermediate view: something important
    - New theories, new ways of organizing beliefs.

Outline

- What are causal theories?
  - Bayes nets
  - Probabilistic graph grammars
  - Probabilistic logics (TBB, RPMs, BLOG)
- How are theories used in causal inference?
  - Generate constrained, domain-dependent hypothesis spaces for domain-general Bayesian inference
- How can theories be learned?
  - Bayesian inference in a space of possible theories.
  - Example: Learning concepts based on causal relations.

Causal networks in cognition

- Learning networks of causal relations
  - e.g., Danks & McKenzie, Griffiths & Tenenbaum, Schulz & Gopnik, Sloman & Lagnado, Sobel, Steyvers & Tenenbaum, Waldmann
- Judging the strength of cause-effect relations
  - e.g., Buehner & Cheng, Cheng & Novick, Shanks, White
- Categories as causal networks
  - e.g., Ahn, Rehder, Waldmann

Causal networks = theories?

Example: network of diseases, effects, and causes

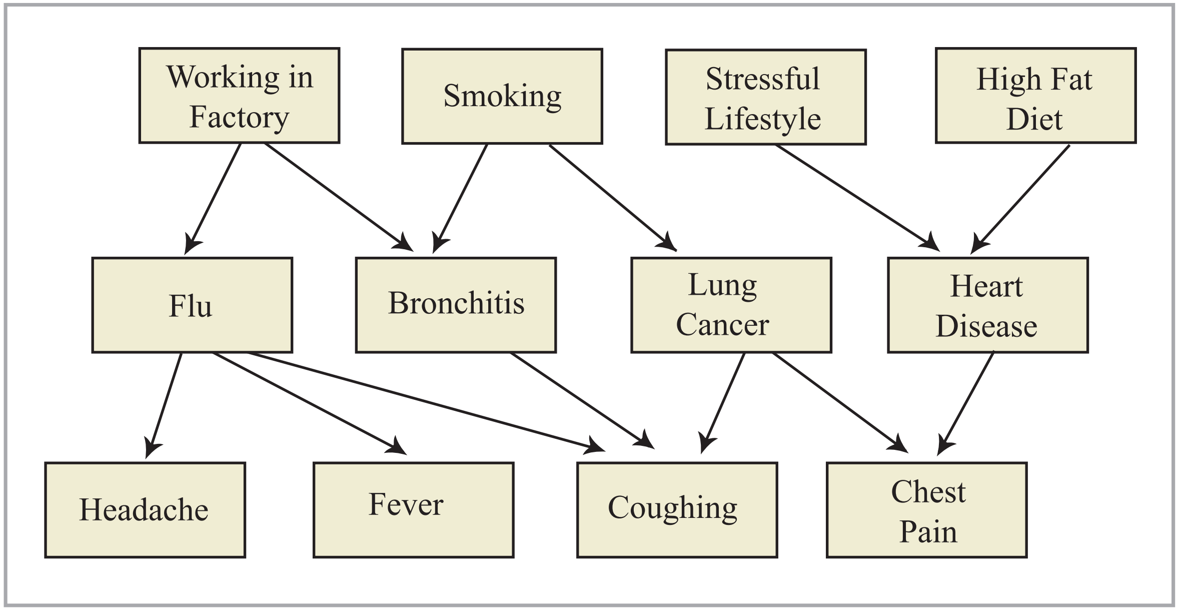


Figure by MIT OCW.

Theory-like properties:

- Generates hypotheses for causal inference (diagnosis).
- Provides causal explanation of observed correlations.
- Supports intervention and action planning.

Causal networks = theories?

Example: network of diseases, effects, and causes

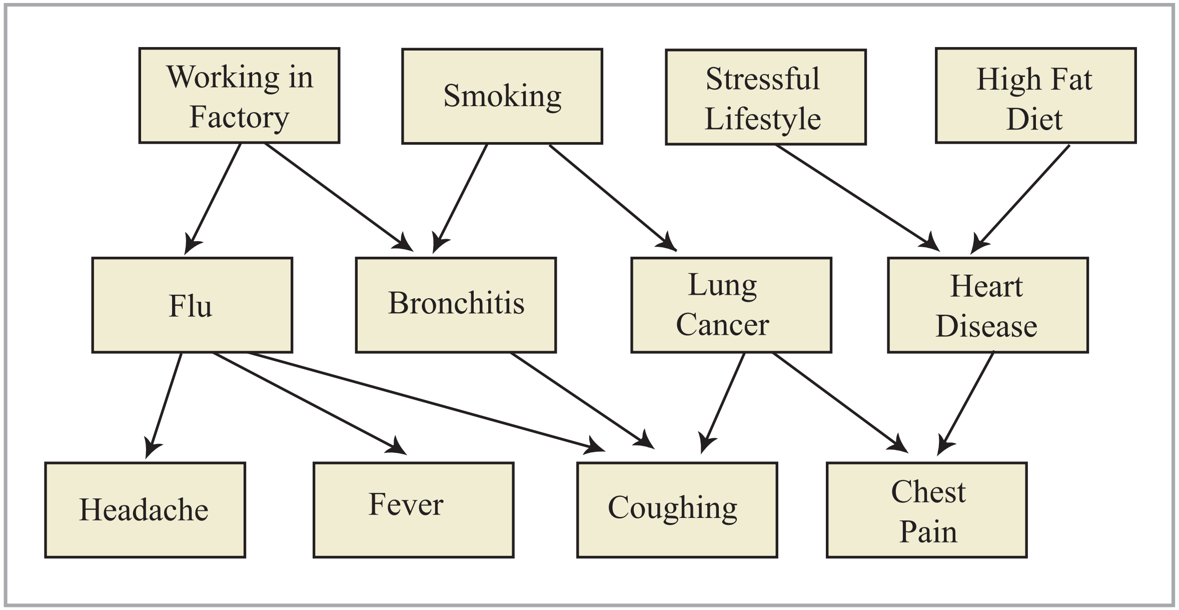


Figure by MIT OCW.

Missing a key level of abstraction:

Domain-specific laws that constrain the causal network structures considered when learning and reasoning in a given domain.

Different causal networks, *Same* domain theory

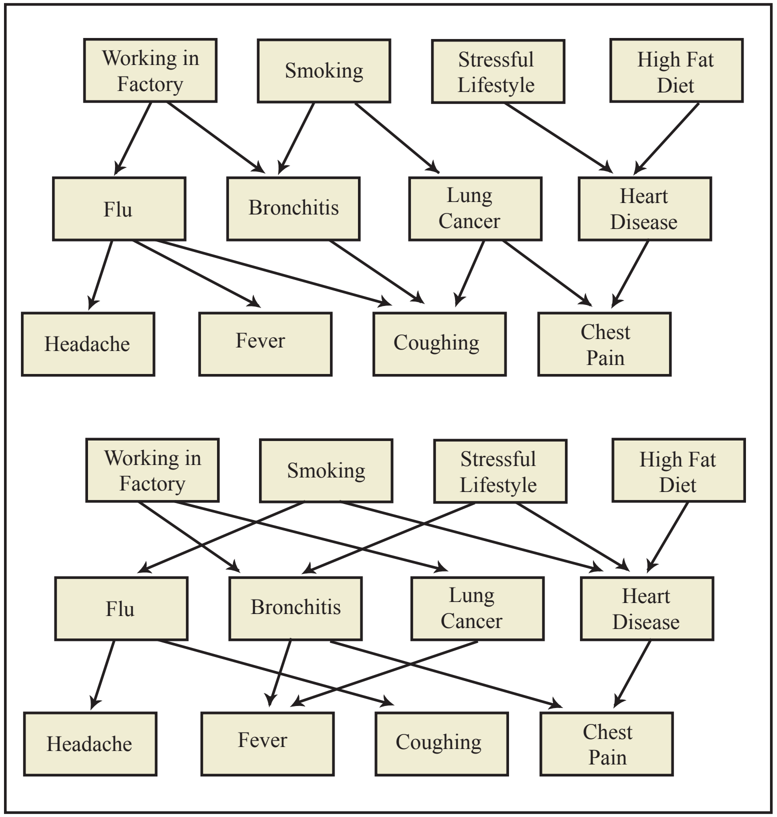


Figure by MIT OCW.

Different domain theories

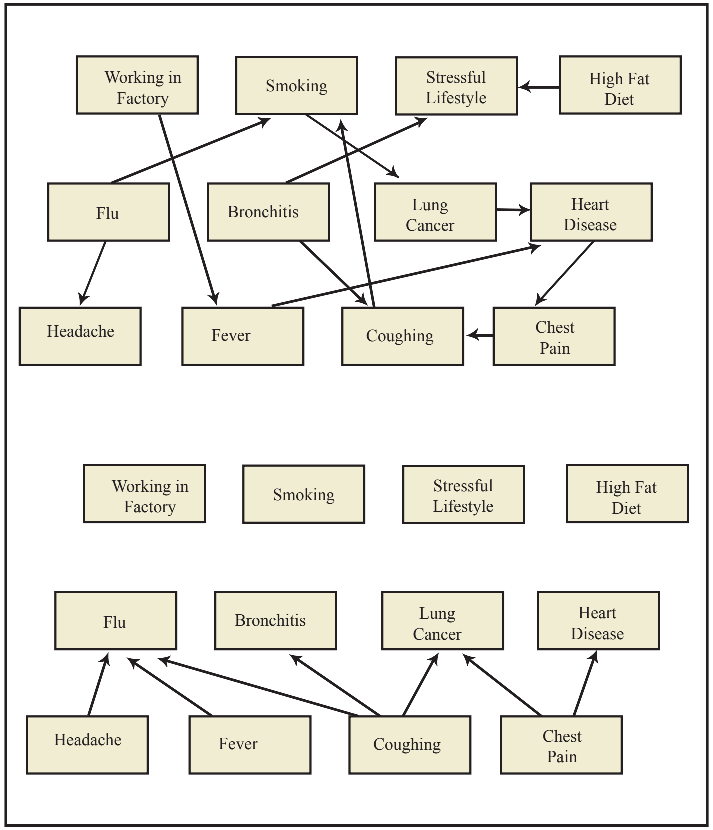
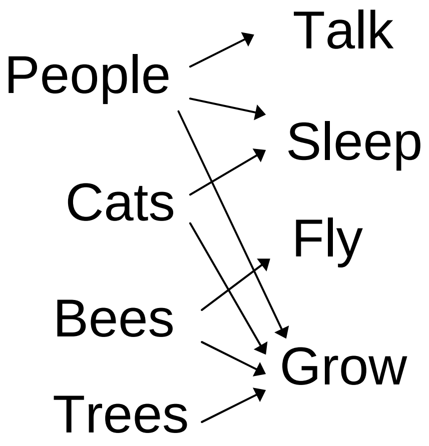


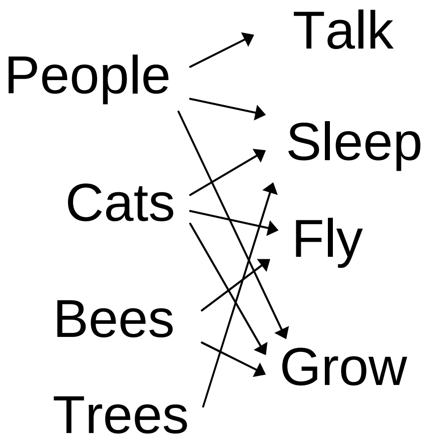
Figure by MIT OCW.

The grammar analogy

Different semantic networks,
Same linguistic grammar

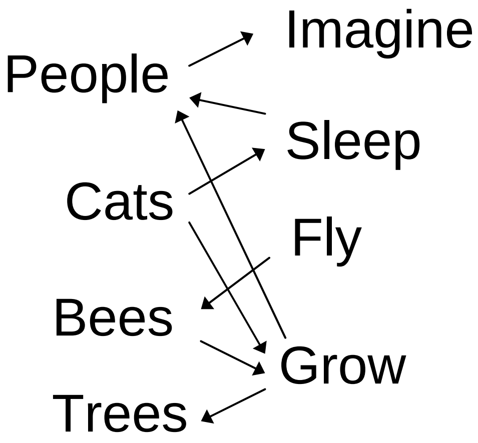


Talk
People
Sleep
Cats
Fly
Bees
Grow
Trees

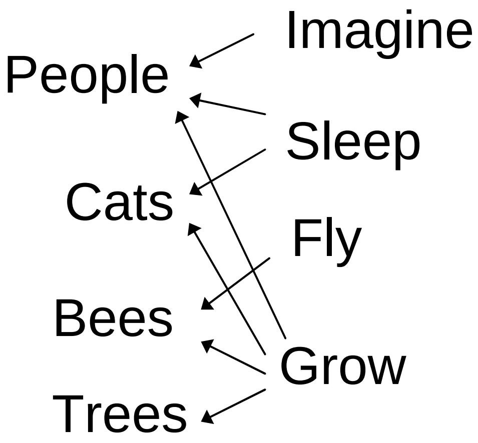


Talk
People
Sleep
Cats
Fly
Bees
Grow
Trees

Different grammars



Imagine
People
Sleep
Cats
Fly
Bees
Grow
Trees



Imagine
People
Sleep
Cats
Fly
Bees
Grow
Trees

The grammar analogy

"The grammar of a language can be viewed as a theory of the structure of this language. Any scientific theory is based on a certain finite set of observations and, by establishing general laws stated in terms of certain hypothetical constructs, it attempts to account for these observations, to show how they are interrelated, and to predict an indefinite number of new phenomena…. Similarly, a grammar is based on a finite number of observed sentences… and it 'projects' this set to an infinite set of grammatical sentences by establishing general 'laws'… [framed in terms of] phonemes, words, phrases, and so on…."

Chomsky (1956), "Three models for the description of language"

Theories in Cognitive Development

"A theory consists of three interrelated components: a set of phenomena that are in its domain, the causal laws and other explanatory mechanisms in terms of which the phenomena are accounted for, and the concepts in terms of which the phenomena and explanatory apparatus are expressed."

Carey (1985), "Constraints on semantic development"

Theories as graph grammars

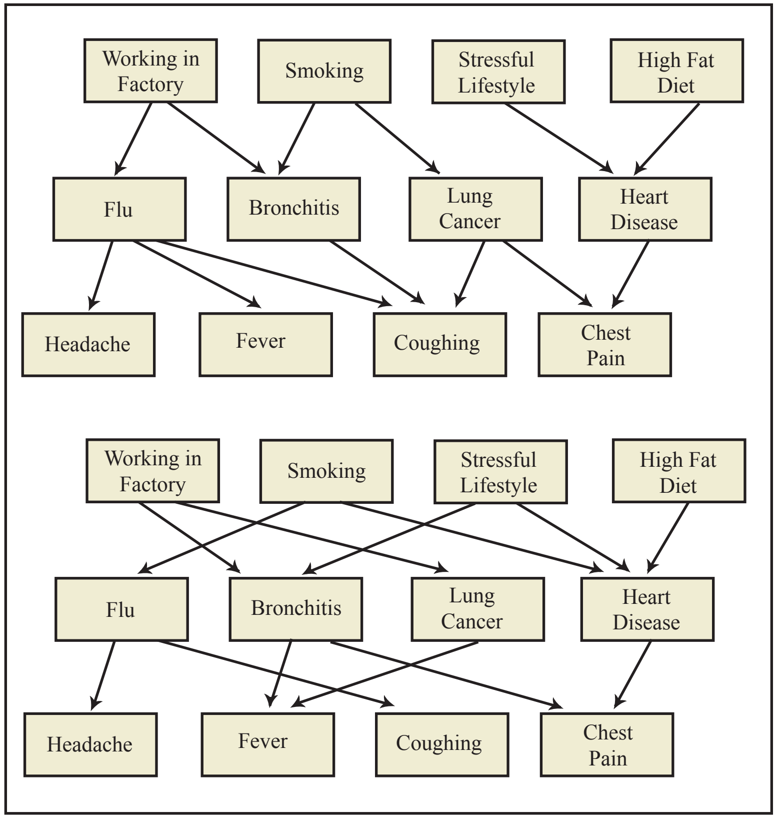


Figure by MIT OCW.

Classes = {B, D, S}

Laws = $\{B \dashrightarrow D, D \dashrightarrow S\}$

($\dashrightarrow$: possible causal link)

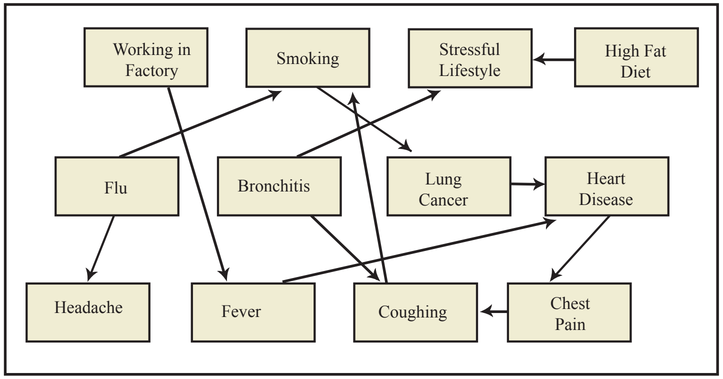


Figure by MIT OCW.

Classes = {C}

Laws = $\{C \dashrightarrow C\}$

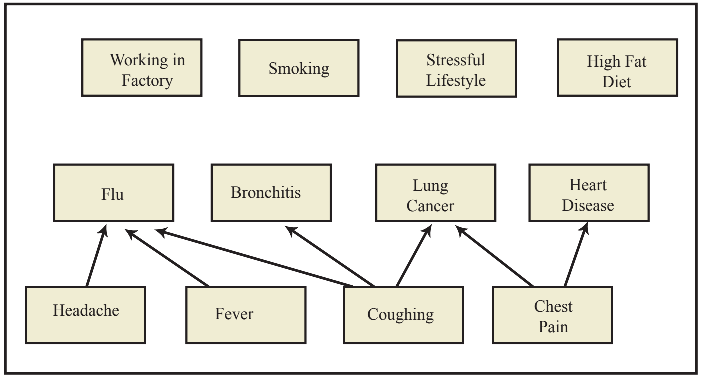


Figure by MIT OCW.

Classes = {B, D, S}

Laws = $\{S \dashrightarrow D\}$

Theories as graph grammars

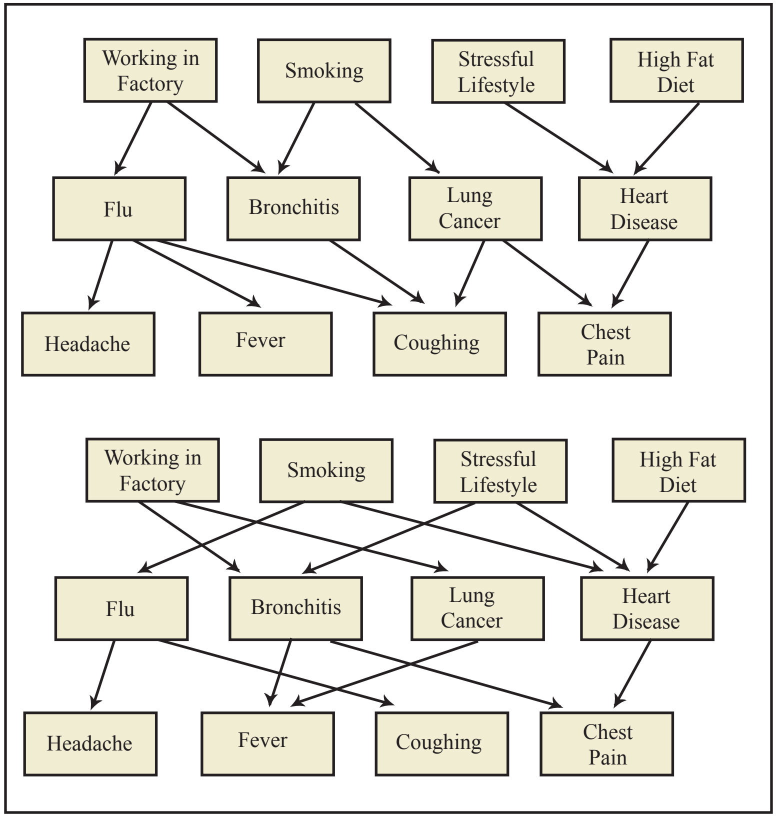


Figure by MIT OCW.

Classes = {B, D, S}

Laws = {B ····► D, D ····► S}

(····► : possible causal link)

Stronger regularities:

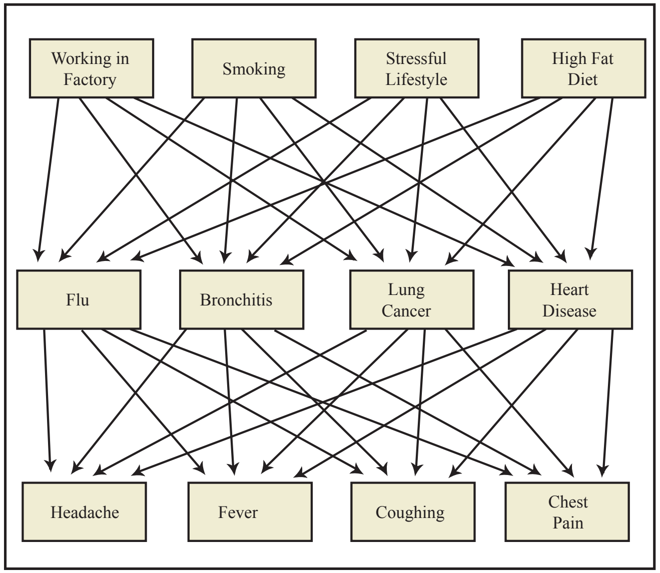


Figure by MIT OCW.

Classes = {B, D, S}

Laws = {B → D, D → S}

(→ : necessary causal link)

Also could have probabilistic link rules.

The grammar analogy

Natural Language Grammar

Abstract classes: N, V, ….

Production rules: S → [N V], ….

(“Nouns precede verbs”)

N → {people, cats, bees, trees, ….}

V → {talks, sleep, fly, grow, ….}

Linguistic Grammar

↓

Syntactic structures

↓

Observed sentences
in the language

Causal Theory

Abstract classes: D, S, ….

Causal laws: [D ⇢ S], ….

(“Diseases cause symptoms”)

D → {flu, bronchitis, lung cancer,….}

S → {fever, cough, chest pain, ….}

Causal Theory

↓

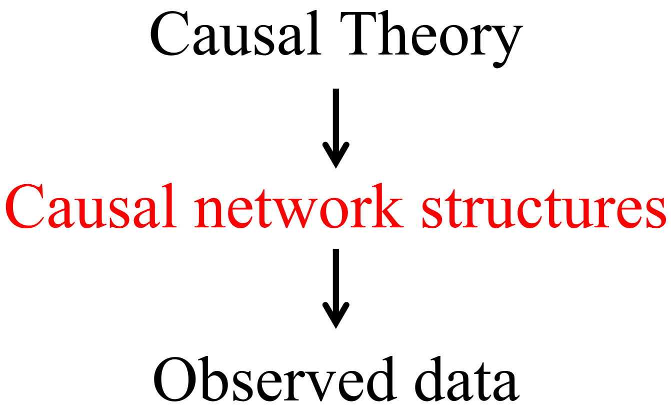
Causal network structures

↓

Observed data
in the domain

Specific theories versus “framework theories”

Wellman (1990; Gelman & Wellman, 1992)



Causal Theory

↓

Causal network structures

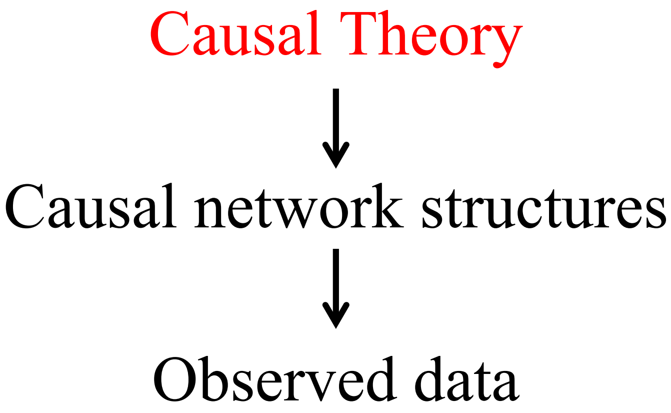
↓

Observed data

“*Specific theories* are detailed scientific formulations about a delimited set of phenomena.”

Specific theories versus “framework theories”

Wellman (1990; Gelman & Wellman, 1992)



Causal Theory

↓

Causal network structures

↓

Observed data

“*Framework theories* outline the ontology and the basic causal devices for their specific theories, thereby defining a coherent form of reasoning about a particular set of phenomena.”

“*Specific theories* are detailed scientific formulations about a delimited set of phenomena.”