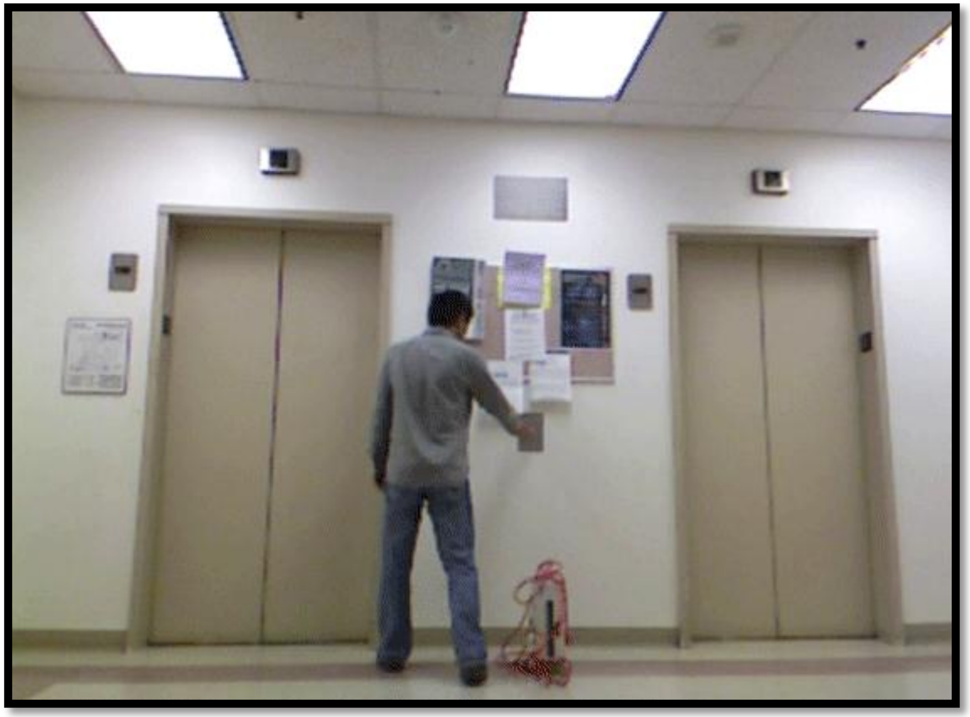
Learning Perceptual Causality from Video

Amy Fire and Song-Chun Zhu

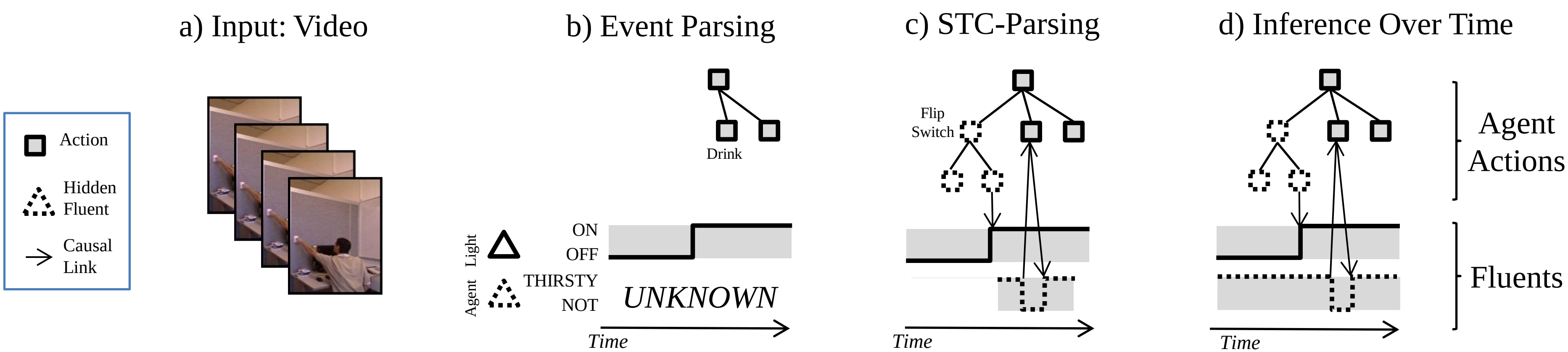
Center for Vision, Cognition, Learning, and Art

UCLA

Ideally:
Learn Causality from Raw Video

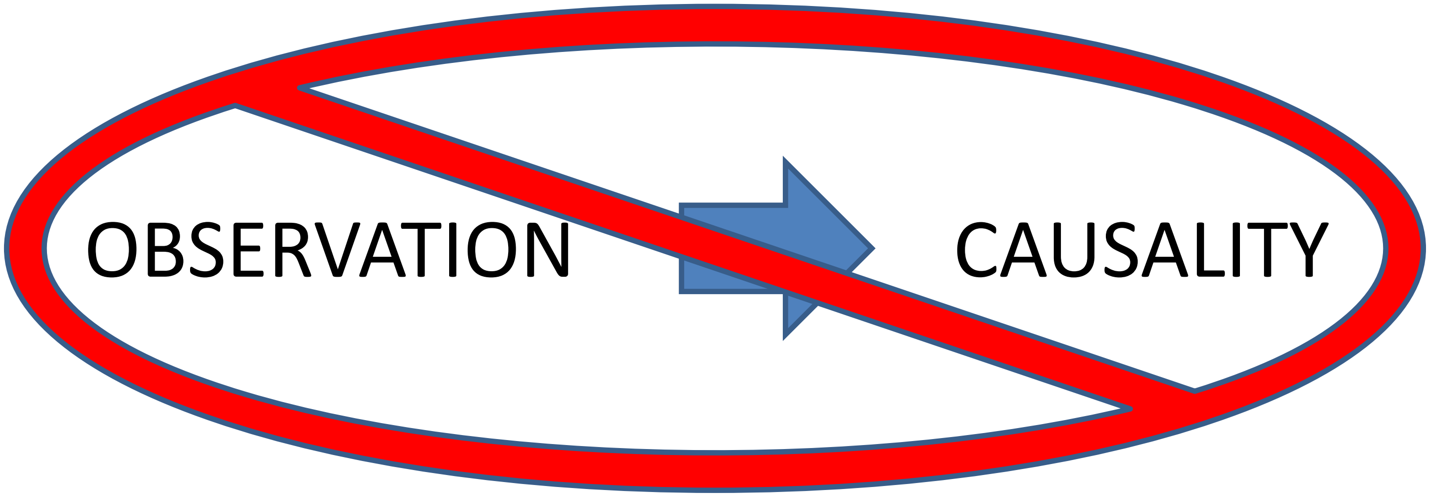


Inference Using Learned Causal Structure



- Answer why events occurred
- Joint STC: Infer misdetections and hidden objects/actions
- Infer triggers, goals, and intents

But…



(generally)

SO…WHERE ARE WE NOW?

Vision Research and Causal Knowledge

- Use pre-specified causal relationships for action detection
  - E.g., PADS (Albanese, et al. 2010)
  - Model Newtonian mechanics (Mann, Jepson, and Siskind 1997)
- Use causal measures to aid action detection
  - E.g., Prabhakar, et al. 2010
- Use infant perceptions of motion to learn causality
  - Using cognitive science (Brand 1997)
- Needed: Learn causality from video, integrating ST learning strategies at pixel level

Causality and Video Data: Often Disjoint

- Learning Bayesian networks
  - Constraint satisfaction (Pearl 2009)
  - Bayesian formulations (Heckerman 1995)
  - Intractable on vision sensors

- Commonsense reasoning (Mueller 2006) – first order logic.
  - Do not allow for ambiguity/probabilistic solutions

- MLNs (Richardson and Domingos 2006)
  - Intractable
  - Used for action detection (Tran and Davis 2008)
    - KB formulas not learned

MOVING FORWARD: OUR PROPOSED SOLUTION

Cognitive Science as a Gateway: Perceptual Causality

- Causal Induction from Observation in Infancy
  - Agentive actions are causes (Saxe, Tenenbaum, and Carey 2005)

  - Co-occurrence of events and effects (Griffiths and Tenenbaum 2005)

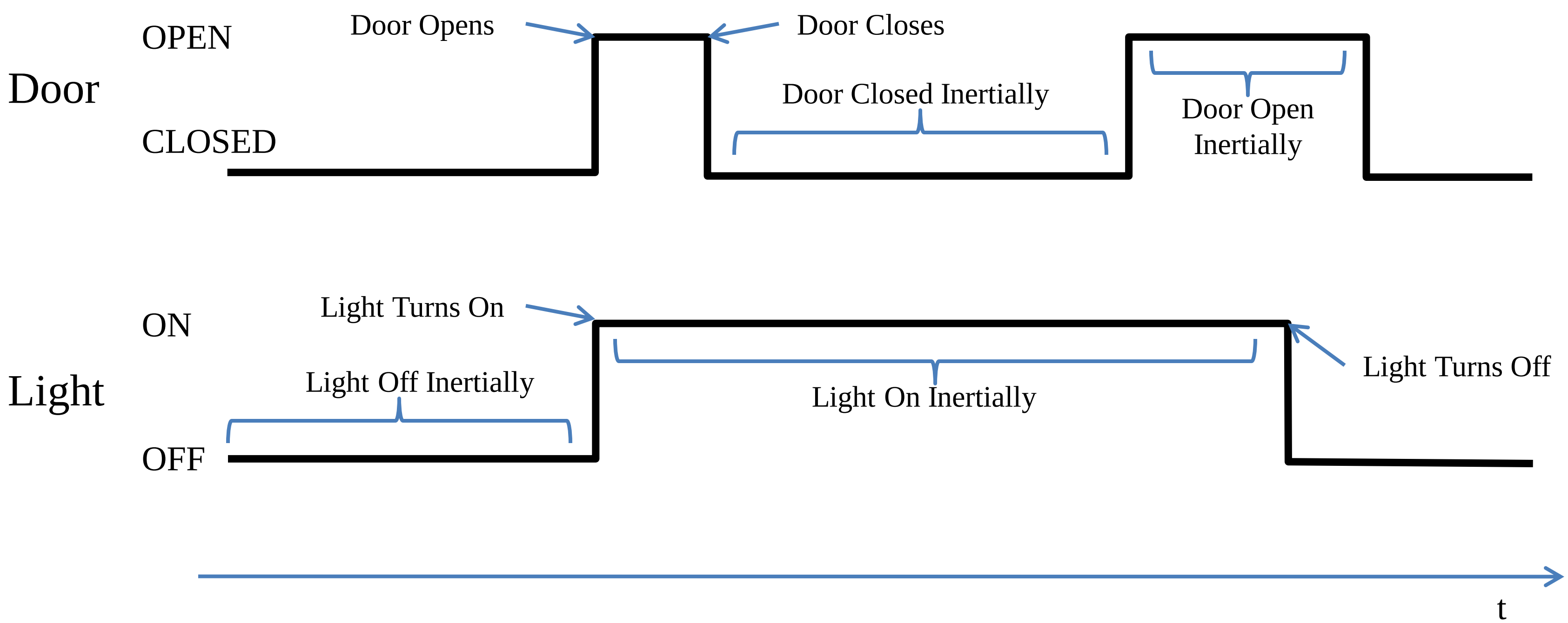
cr :

	Action	¬Action
Effect	c_0	c_1
¬Effect	c_2	c_3

  - Temporal lag between the two is short (Carey 2009)
  - Cause precedes effect (Carey 2009)
- Note: NOT the same as necessary and sufficient causes

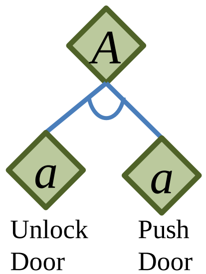
MODIFIED GOAL:
LEARN AND INFER *PERCEPTUAL CAUSALITY*

What are the effects? Fluent changes.

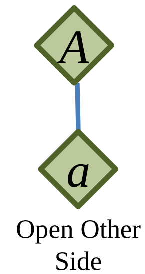


- Fluents are time-varying statuses of objects
  - Mueller – Commonsense Reasoning 2006

What are the causes? Actions.



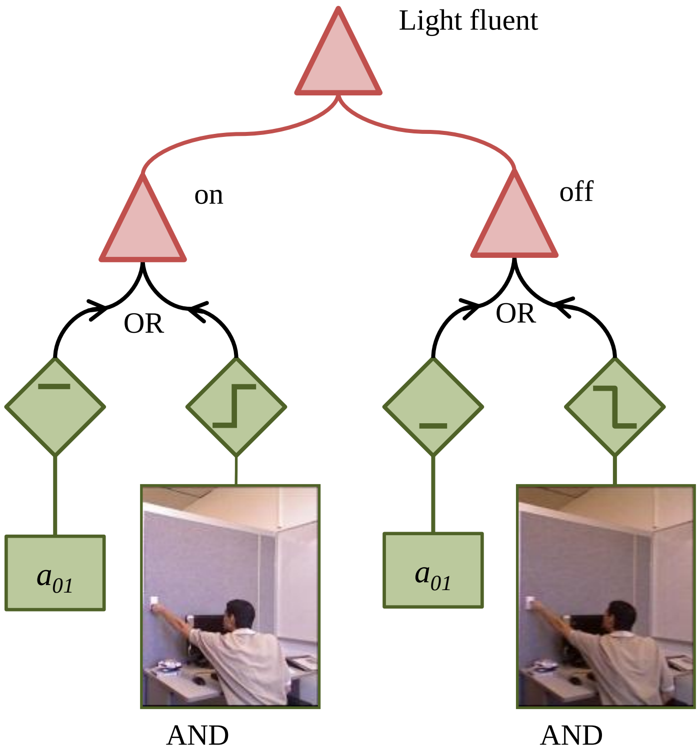
And-Nodes Compose Single Causes (multiple sub-actions)



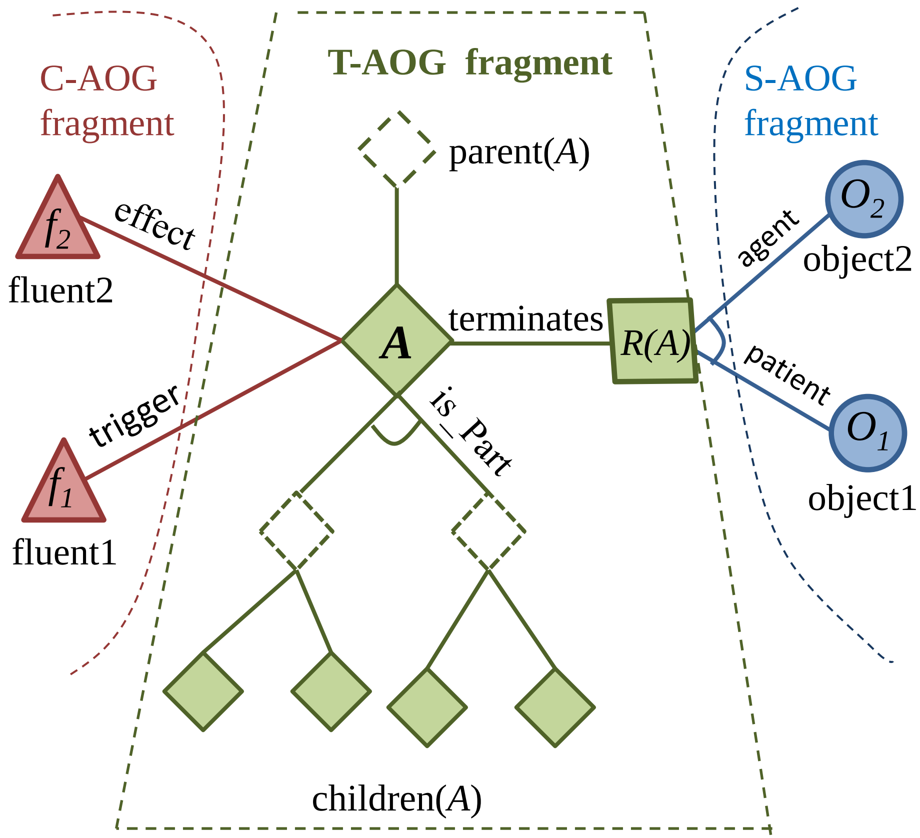
Or-Nodes Give Alternative Causes

- Probabilistic Graphical Representation for Causality
  - And-Or Graph

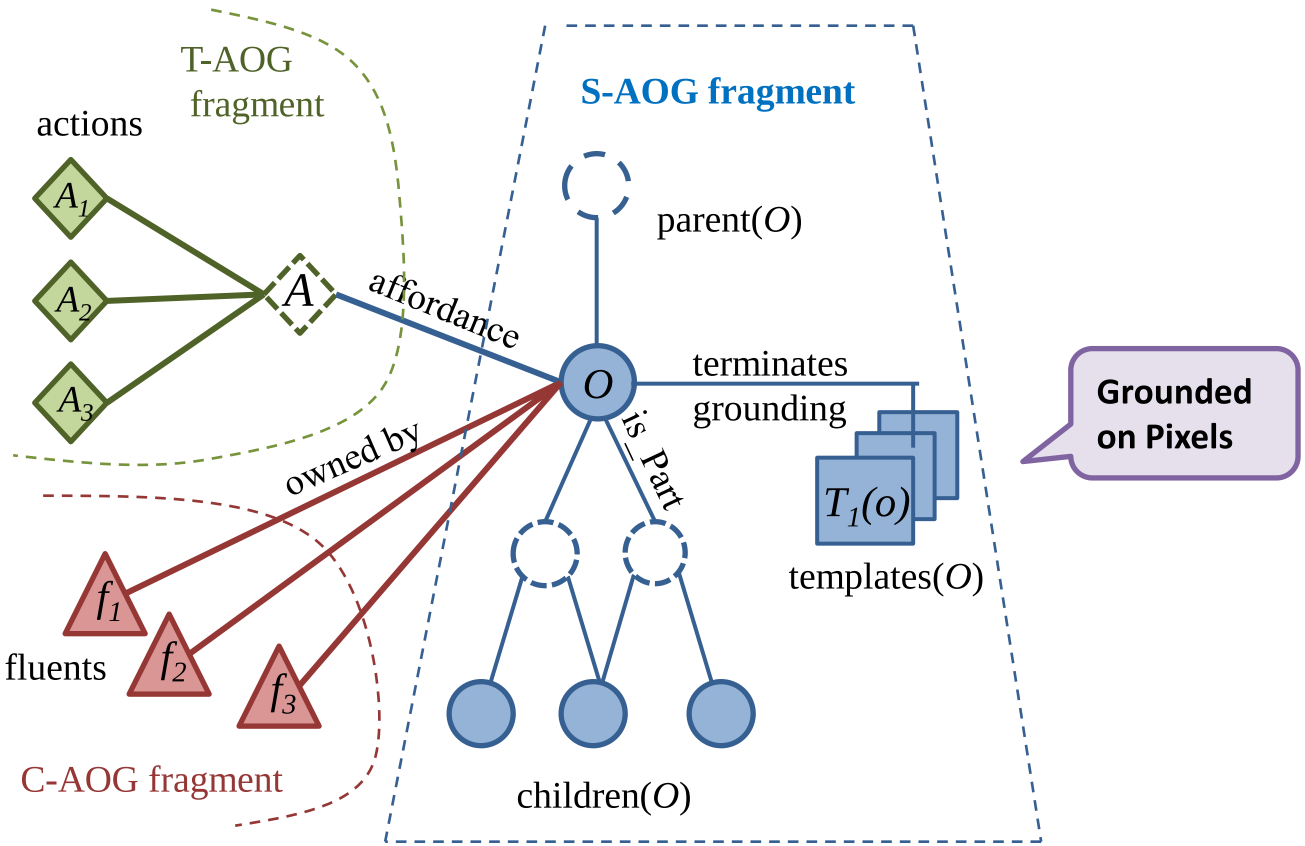
Causal AOG



Connecting Temporal to Causal and Spatial



Grounding on Pixels: Connecting S/T/C-AOG



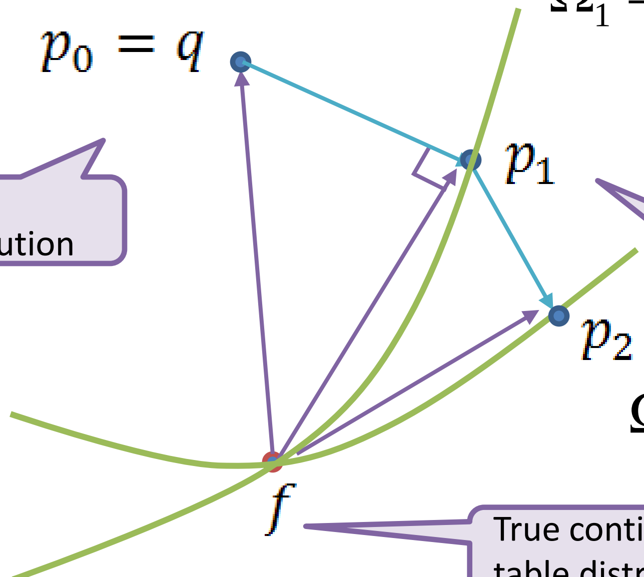
Preliminary Theory

LEARNING PERCEPTUAL CAUSALITY

Principled Approach: Information projection

$$\Omega_1 = \left\{ p : E_p\left(\mathbf{cr}_1(I)\right) = E_f\left(\mathbf{cr}_1(I)\right) \right\}$$

Match the statistics of the contingency table



Initial distribution

Select cause/effect relationship that makes p_1 closest (KL) to p to preserve learning history while maximizing Info. Gain

$$\Omega_2 = \left\{ p : E_p\left(\mathbf{cr}_2(I)\right) = E_f\left(\mathbf{cr}_2(I)\right) \right\}$$

Repeat

True contingency table distribution

$$KL(f \| p) = KL(f \| p_+) + KL(p_+ \| p)$$

$$\max \left[KL(f \| p) - KL(f \| p_+) \right] = \max KL(p_+ \| p)$$

DellaPietra, DellaPietra, Lafferty, 97
Zhu, Wu, Mumford, 97

Learning Pursuit: Add Causal Relations

- Model Pursuit

$$p_0 \to p_1 \to \ldots \to \boxed{p \to p_+} \to \ldots \to p_k \approx f$$

(On ST-AOG)

$$p_+(pg) = \frac{1}{z_+} p(pg) \exp\left(-\langle \lambda_+, \mathbf{cr}_+ \rangle\right)$$

cr	Effect	¬ Effect
Action	c_o	c_2
¬ Action	c_1	c_3

- Proposition 1: Find parameters
  - Model formed by min $KL\ (p_+ \,||\, p)$, matching statistics

$$E_{p_+}(\mathbf{cr}_+) = E_f(\mathbf{cr}_+) \qquad \lambda_{+,i} = \log\left(\frac{h_i}{h_0} \times \frac{f_0}{f_i}\right)$$

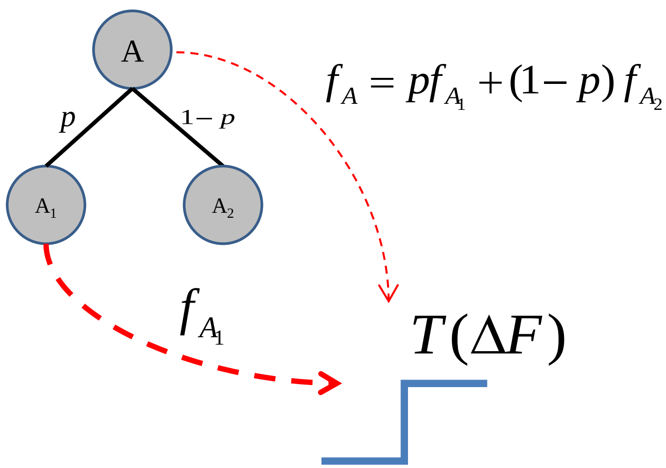
h_i is c_i under p
f_i is c_i from f

- Proposition 2: Pursue **cr**.

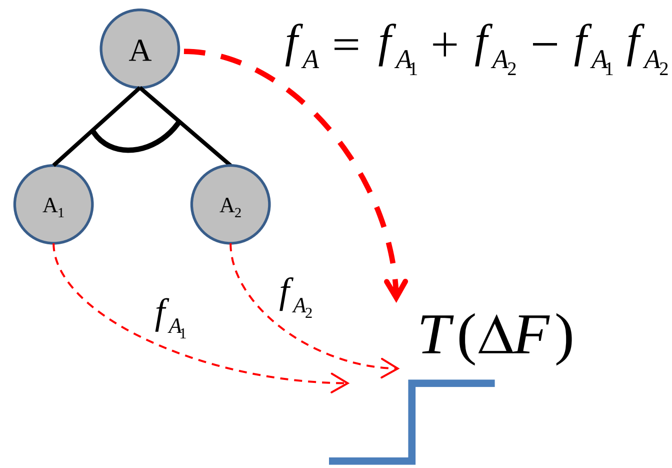
$$\mathbf{cr}_+ = \underset{\mathbf{cr}}{\operatorname{argmax}}\, KL\left(p_+ \,||\, p\right) = \underset{\mathbf{cr}}{\operatorname{argmax}}\, KL\left(\mathbf{f} \,||\, \mathbf{h}\right)$$

Selection from ST-AOG

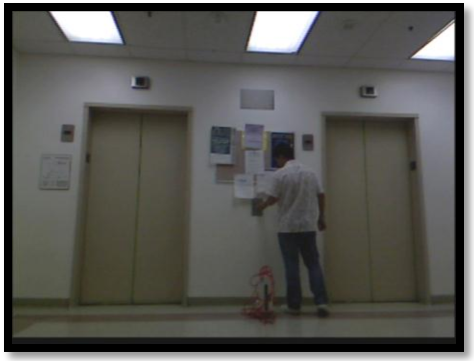
Or-node

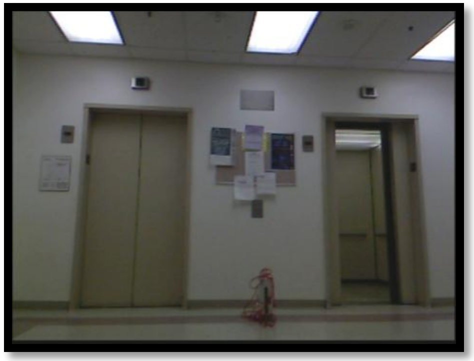


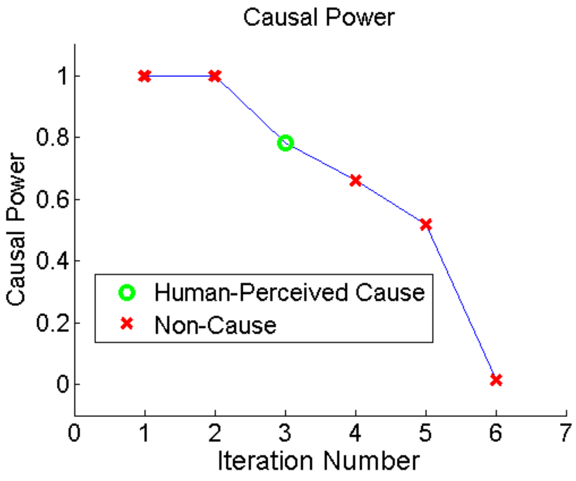
And-node

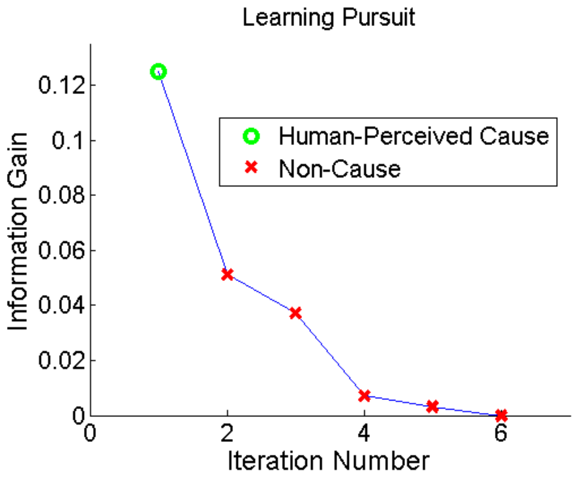


Performance vs. TE

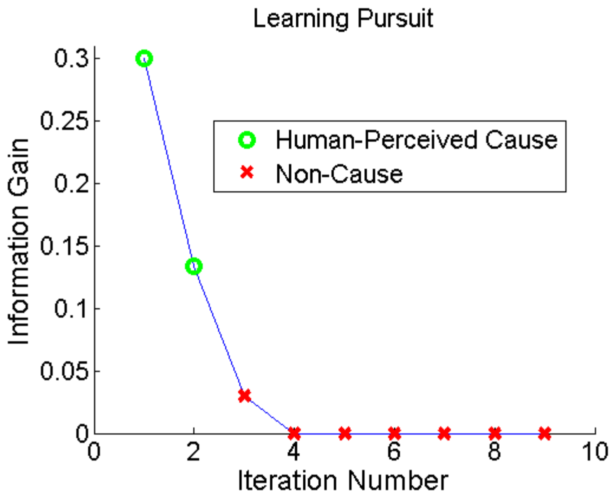


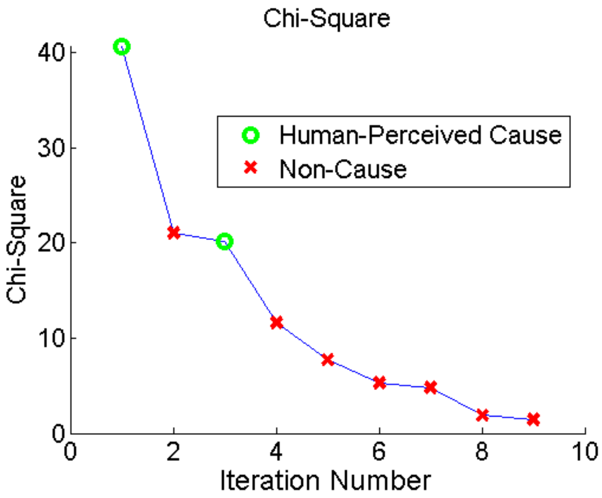




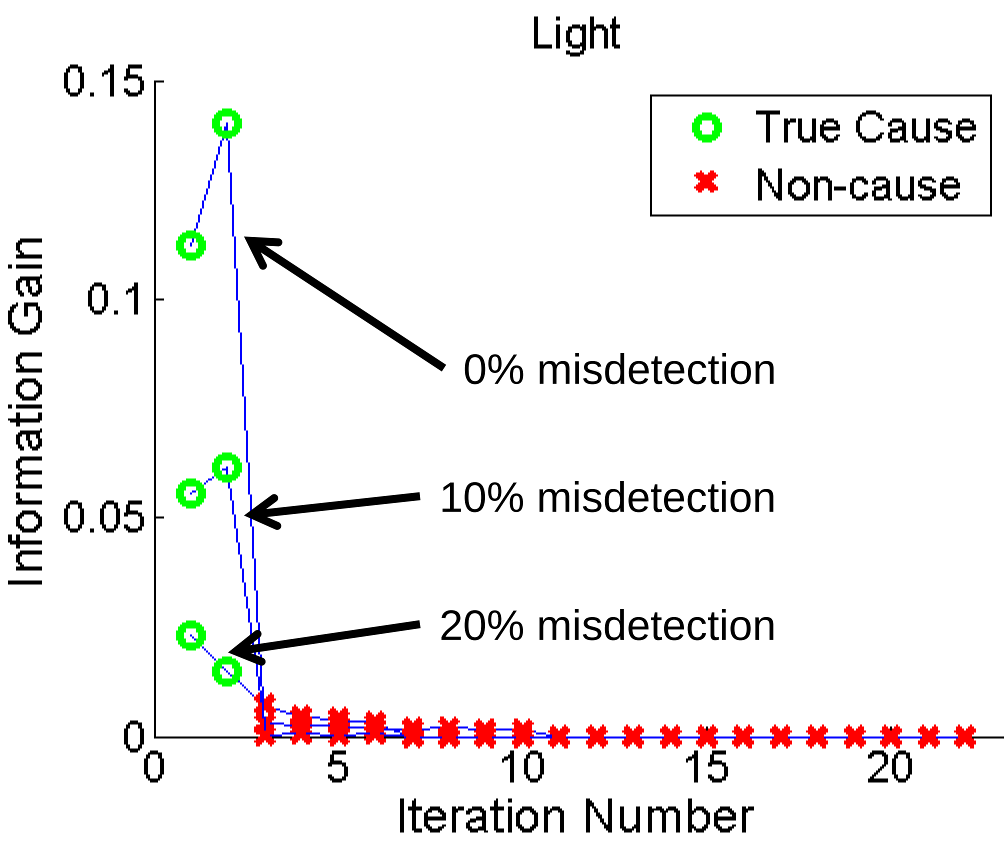


Performance vs. Hellinger χ^2





Increasing Misdetections
(Simulation)



STC-Parsing Demo

Looking Forward:

- Finish learning the C-AOG
- Increase reasoning capacity of the C-AOG
- Integrate experiment design