Understanding "why"

Causation, Counterfactuals, and Imagination

@tobigerstenberg

http://cicl.stanford.edu
Why I'm here today

@tobigerstenberg
Looking forward to it!

@eliasbareinboim
The WHY-19 will be happening from Mar/25-27 @ Stanford. The theme this year is "Beyond Curve Fitting: Causation, Counterfactuals, Imagination-based AI". We have great speakers, including J. Pearl, Y. Bengio, K. Imai, J. Ioannidis. Don't miss!! why19.causalai.net #bookofwhy

Hi Tobias, how are you doing? I saw your tweet and realized that you could be a great speaker and/or panelist to our symposium. Would you be interested in joining? If so, I'll send a formal invite.

@tobigerstenberg
Mar 7
Our lab studies the role of causality in our understanding of the world, and of each other.
The kind of causal inference seen in natural human thought can be “algorithmitized” to help produce human-level machine intelligence.

BY JUDEA PEARL

The Seven Tools of Causal Inference, with Reflections on Machine Learning


Intuitive theories as probabilistic programs

Battaglia, Hamrick & Tenenbaum (2013) Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*

Beyond structural equations

How do we **do ( )** in a probabilistic program?

How do we simulate counterfactuals?
Outline

Causal judgments

1. colliding balls

Multi-modal inference

2. dropping ball
Did A cause B to go through the gate?
What happened?

Actual situation

B went through the gate

What would have happened?

Counterfactual situation

B would have missed the gate


Quantitative predictions

What happened?

Actual situation
B went through the gate

What would have happened?

Counterfactual situation
B would have missed the gate
B would have missed the gate
B would have missed the gate

Quantitative predictions

What happened?

Actual situation

B went through the gate

What would have happened?

Counterfactual situation

B would have missed the gate

B would have gone through gate

B would have gone through gate

Quantitative predictions

Causal system

Causal judgment

Probabilistic program

Chater & Oaksford (2013) Programs as causal models: Speculations on mental programs and mental representation. Cognitive Science
What else?
Actualist theories of causation

What happened?

What would have happened?

- Salmon (1994) Causality without counterfactuals. Philosophy of Science
- Walsh & Sloman (2011) The meaning of cause and prevent: The role of causal mechanism. Mind & Language
Counterfactuals are necessary

Counterfactuals are necessary

Counterfactuals are necessary

Did A prevent B from going through the gate?

Counterfactuals are necessary

Spontaneous counterfactual simulation

Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. *Psychological Science*
Spontaneous counterfactual simulation

Did B completely miss the gate?

1/2 speed

Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. *Psychological Science*
Spontaneous counterfactual simulation

Did A prevent B from go through the gate?

Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. *Psychological Science*
Spontaneous counterfactual simulation

Did $B$ completely miss the gate?

Did $A$ prevent $B$ from going through the gate?

Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. *Psychological Science*
Counterfactual simulation model of causal judgment

- causal judgments are well-explained by the observer's beliefs about whether the candidate cause made a difference to the outcome

- counterfactual contrasts are necessary for explaining people's causal judgments

- people spontaneously engage in counterfactual simulation when making causal judgments


Did ball A cause/enable/help ball B to go through the gate? Gerstenberg & Tenenbaum (2017) Intuitive Theories. Oxford Handbook of Causal Reasoning

the language of causation

Sosa, Ullman, Gershman, Tenenbaum & Gerstenberg (submitted) Moral Dynamics.

How bad was what Blue did to Green?
Outline

1. Causal judgments
   - colliding balls

2. Multi-modal inference
   - dropping ball
Causal inference: Multi-modal integration through mental simulation


Prediction: Where will the ball land?
Prediction: Where will the ball land?
Prediction: Where will the ball land?
Prediction: Where will the ball land?
Prediction: Where will the ball land?

Ullman, Spelke, Battaglia, & Tenenbaum (2017) Mind Games: Game Engines as an Architecture for Intuitive Physics. *Trends in Cognitive Sciences*

Prediction: Where will the ball land?

\[ r = 0.99 \]
\[ \text{RMSE} = 36.5 \]
Inference: In which hole was the ball dropped?
Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

\[ \exp\left(-\frac{d(b_{\text{final},x}, b_{\text{hole},x})}{2\sigma^2}\right) \]
Inference: In which hole was the ball dropped?
Inference: In which hole was the ball dropped?
Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

\[ \exp\left(-\frac{d(b_{\text{hole}}, b_{\text{final}})}{2\sigma^2}\right) \]
Inference: In which hole was the ball dropped?

$$t = [37, 77]$$

average temporal distance between time points

$$\sum_i^N \exp\left(-\frac{d(\text{sound}_{true_i}, \text{sound}_{simulation_i})}{2\sigma^2}\right)$$

$$t_1 = [16, 60, 99]$$
$$t_2 = [16, 56, 99]$$
$$t_3 = [15, 81, 95]$$

$$t_1 = [37, 79]$$
$$t_2 = [37, 78]$$
$$t_3 = [37, 75]$$

$$t_1 = [45]$$
$$t_2 = [45]$$
$$t_3 = [45]$$

$$t = [37, 77]$$ + penalty
$$t_1 = [16, 60, 99]$$

$$t = [37, 77]$$ + penalty
$$t_1 = [37, 79]$$

$$t = [37, 77]$$ + penalty
$$t_1 = [45]$$
Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

$$
\exp\left(-\frac{d(\text{ball}_x_{\text{final}}, \text{ball}_x_{\text{hole}})}{2\sigma^2}\right)
$$

average temporal distance between time points

$$
\sum_i^N \exp\left(-\frac{d(\text{sound}_{true_i}, \text{sound}_{simulation_i})}{2\sigma^2}\right) \frac{1}{N}
$$

$$t = [37, 77]$$

$$t_1 = [37, 79]$$

$$t_2 = [37, 78]$$

$$t_3 = [37, 75]$$

$$t_1 = [45]$$

$$t_2 = [45]$$

$$t_3 = [45]$$

multiplicative integration
Inference: In which hole was the ball dropped?
Conclusion

• we build rich mental models of the world
• we simulate these models to:
  - predict the future
  - infer the past
  - evaluate counterfactuals
• together, these capabilities allow us to understand why something happened

Lake, Ullman, Tenenbaum, & Gershman (2016) Building machines that learn and think like people. Behavioral and Brain Sciences
Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. Psychological Science
