

Computational Cognitive Science

Lecture 11: Causality 1

Frank Keller

26 October 2026

School of Informatics
University of Edinburgh
keller@inf.ed.ac.uk

Slide credits: Chris Lucas, Benjamin Peters

Optional reading: [Pearl \(2009\)](#).

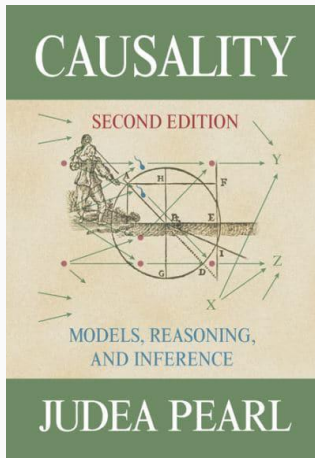
Causality and causal reasoning

“if there be any relation among objects which it imports to us to know perfectly, it is that of cause and effect. On this are founded all our reasonings concerning matter[s] of fact or existence” (Hume 1748)

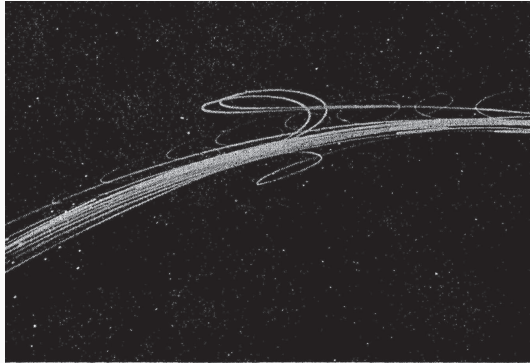


Causality and causal reasoning

"I now take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality" ([Pearl 2009](#))

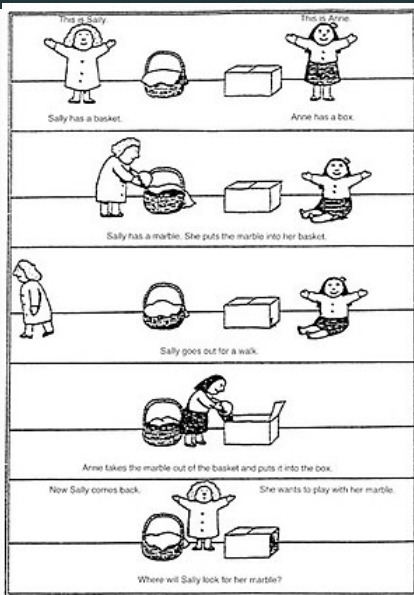


Hard to escape in cognitive science (or anywhere)



- Why do some points in the night sky move that way?
- How will my intervention shape people's judgments in my experiment?

Theory of mind and social reasoning



- "Why did they make that face?"
- "Why didn't Sally look for the marbles in the box?"
- "Why didn't he jump off the diving board?"

See [Wellman \(e.g., 2011\)](#).

Planning

- “How can I avoid getting sick?”
- “How can I pass my courses?”



- Which path should I take?

- “What makes a cat a cat?”
- Causal relationships influence category judgments
 - `has_cat_DNA` $\succ$ `is_furry`, `does_meow`

See, e.g., [Rehder \(2010\)](#).

Arthur's Seat blaze likely caused by human activity - fire service



- Who is responsible for the fire?
- Why did the patient die?
- Determining the “real” causes of event with many contributing factors

See, e.g., [Lagnado and Gerstenberg \(2017\)](#).

- “Will removing that block make the tower fall?”
- “Where should I aim my dart?”

Why is it important to think about *causality*?

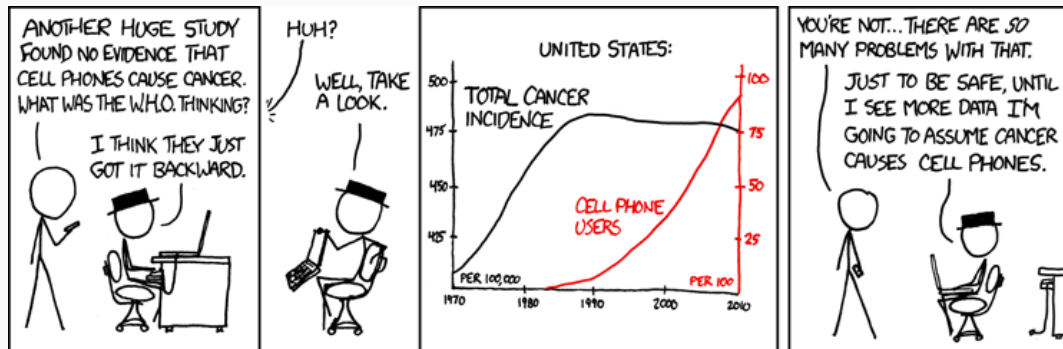
What mistakes arise if we get associations right but causality wrong?

(setting aside spurious/coincidental associations for now)

If we fail to distinguish association from causality:

- Antibiotics cause infections
- Smoking doesn't cause health problems; a propensity for risky behavior causes both
- The landing dance summons planes

Causality vs association



From [xkcd web comics 925](#)

Historical perspectives

“we may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object had not been, the second never had existed.” (Hume, 1748)

This quote seems to offer two different theories:

1. Causality is just association and temporal order.
2. Causality depends on **counterfactuals** – if the cause had changed, its effects would have changed as well.

Do people infer causal relationships from association, potentially constrained by order?

Associative learning

Idea: When C is associated with effect E that **we didn't already anticipate**, we learn to predict E from C.

Enter the Rescorla-Wagner model (RW) model ([Rescorla and Wagner 1972](#)).

RW has a long history outside causal learning – not our focus here.

We will dispense with behaviorist nomenclature (e.g., “conditioned stimulus”).

Some people still take extensions of RW seriously.s

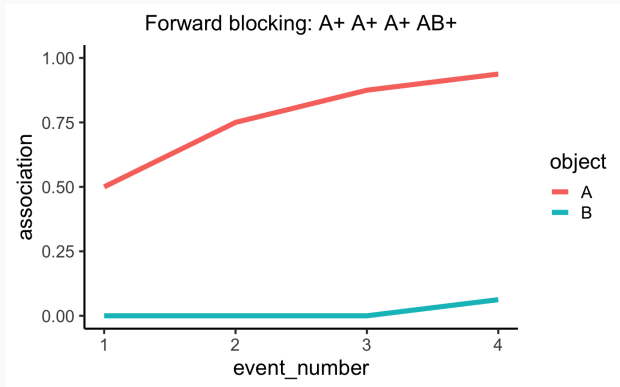
$$\Delta V_i = \alpha_i \beta (\lambda - \sum_{j \in C} V_j)$$

- ΔV_i : Change in the association value between stimulus i and the effect
- λ : 0/1 if effect is absent/present (in the binary case)
- α_i : The learning rate associated with cause
- β : The learning rate associated with the effect
- C : The set of causes that are present

Some features of RW:

- Associations can be negative
- β can vary between present and absent effects
- Simplifying assumption: $\alpha_i = \alpha$
- Can conflate:
 - “I suspect this is a reliable cause”
 - “I *know* this is a weak cause”

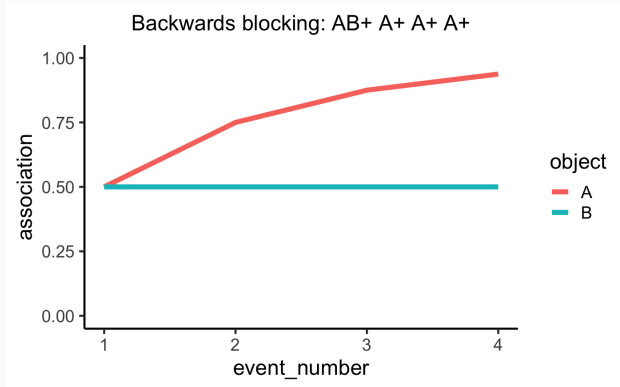
Rescorla-Wagner: Forward blocking



If *A* alone can explain the effect, it “blocks” *B*.

This is consistent with human behavior.

Rescorla-Wagner: No backwards blocking



If we reverse the order of events, learning A is a sufficient cause does not cause RW to update association for B.

People **do** revise their beliefs about B in light of later A events.

Is RW a good model?

Mixed success, empirically; can't explain some phenomena, e.g., backwards blocking

Is RW a good model?

RW also has some theoretical shortcomings:

- Has trouble w/more complex causal relationships, e.g.,
 - Enabling conditions
 - Magnitudes
- Doesn't accommodate prior knowledge
- Relies on temporal information to avoid spurious inferences
- Human experience isn't divided into trials – time is continuous
- Conflates confidence in a relationship and strength of the relationship

But like the Copernican model of planetary motion, it provides a useful stepping stone to more complex and accurate models

Next we will focus on probabilistic models that take a *counterfactual* view of causality

References

- Hume, David. 1748. *An Enquiry Concerning Human Understanding*. A. Millar.
- Lagnado, David A. and Tobias Gerstenberg. 2017. Causation in legal and moral reasoning. In Michael R. Waldmann, editor, *The Oxford Handbook of Causal Reasoning*, Oxford University Press, Oxford, pages 565–602.
- Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2nd edition.
- Rehder, Bob. 2010. Chapter 2: Causal-based categorization: A review. In *The Psychology of Learning and Motivation*, Academic Press, volume 52 of *Psychology of Learning and Motivation*, pages 39–116.
- Rescorla, Robert A. and Allan R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black and William F. Prokasy, editors, *Classical Conditioning II: Current Research and Theory*, Appleton-Century-Crofts, New York, pages 64–99.
- Wellman, Henry M. 2011. Developing a theory of mind. In Usha Goswami, editor, *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, Wiley-Blackwell, Malden, MA, pages 258–284. 2nd edition.