Computational Cognitive Science

Lecture 11: Causality

Benjamin Peters

School of Informatics

University of Edinburgh

October 20, 2025

Reading

Optional:

- “Causality” (Pearl, 2009)

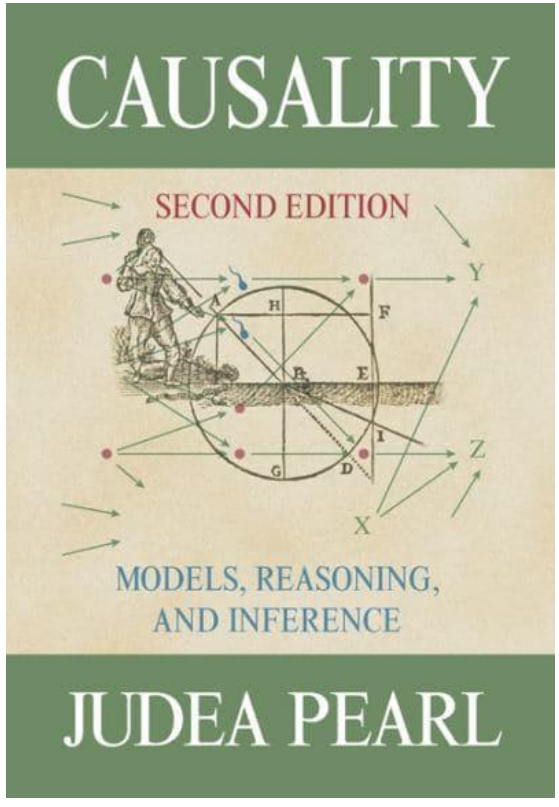
Causality and causal reasoning

"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)

Photo credit: Anne Burgess, who points out Hume probably didn't wear a toga in 18th century Edinburgh.

Causality and causal reasoning

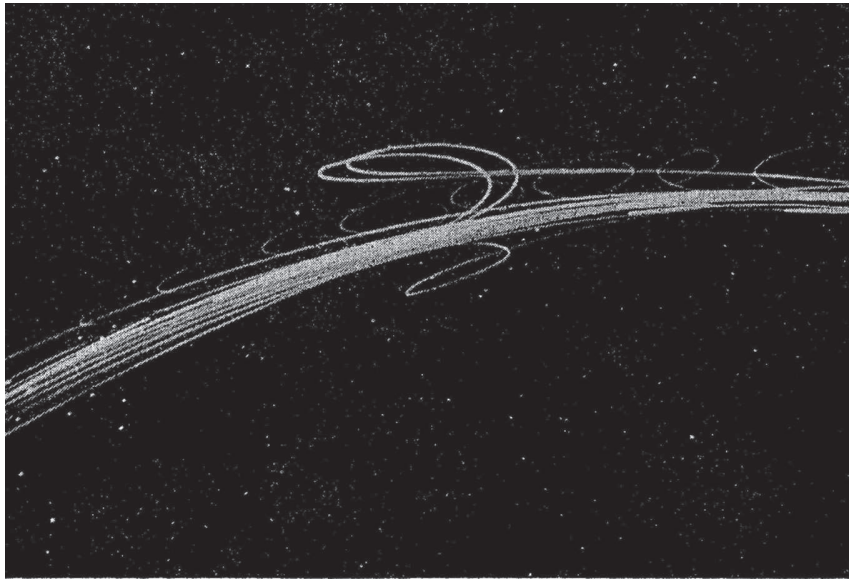
"I now take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality" (Pearl, 2009)



Causality and causal reasoning

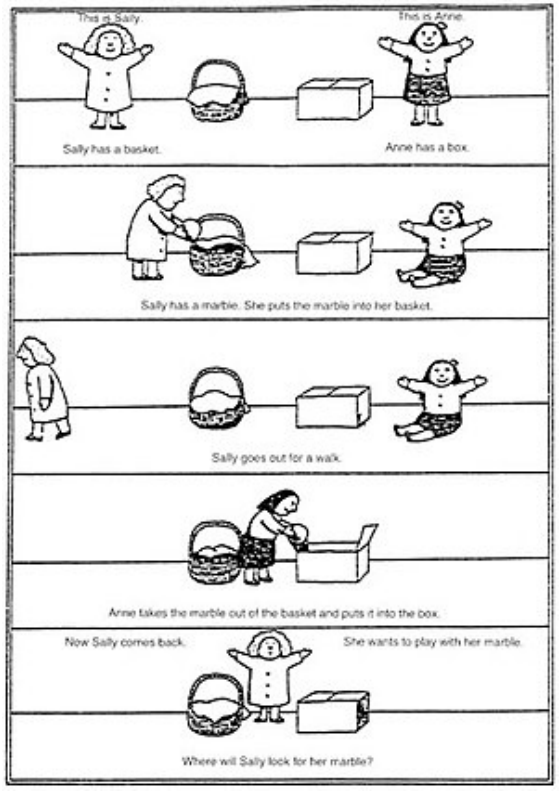
Hard to escape in cognitive science (or anywhere)

Scientific reasoning



- 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, e.g., "Developing a Theory of Mind" by Wellman, 2011; link)

Planning

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

- Which path should I take?

Categorization

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

(see, e.g., Rehder, 2010; link)

Causal attribution

- Who is responsible for the fire?
- Why did the patient die?
- "but for the doctor's actions, the patient would have survived"
- Determining the "real" causes of event with many contributing factors

(see, e.g., Lagnado and Gerstenberg, 2017; link)

Physical reasoning

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

Causality vs association

Why is it important to think about *causality*?

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

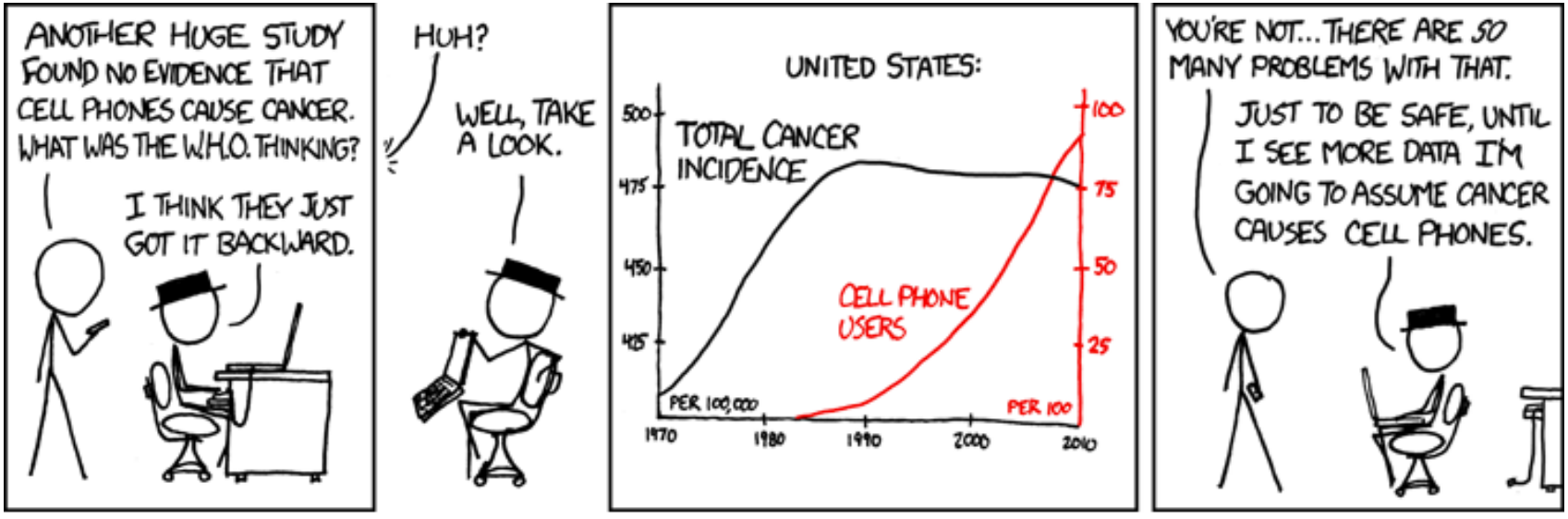
(setting aside spurious/coincidental associations for now)

Causality vs association

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



https://xkcd.com/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.

Association and causation

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 (link).

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

Rescorla-Wagner

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

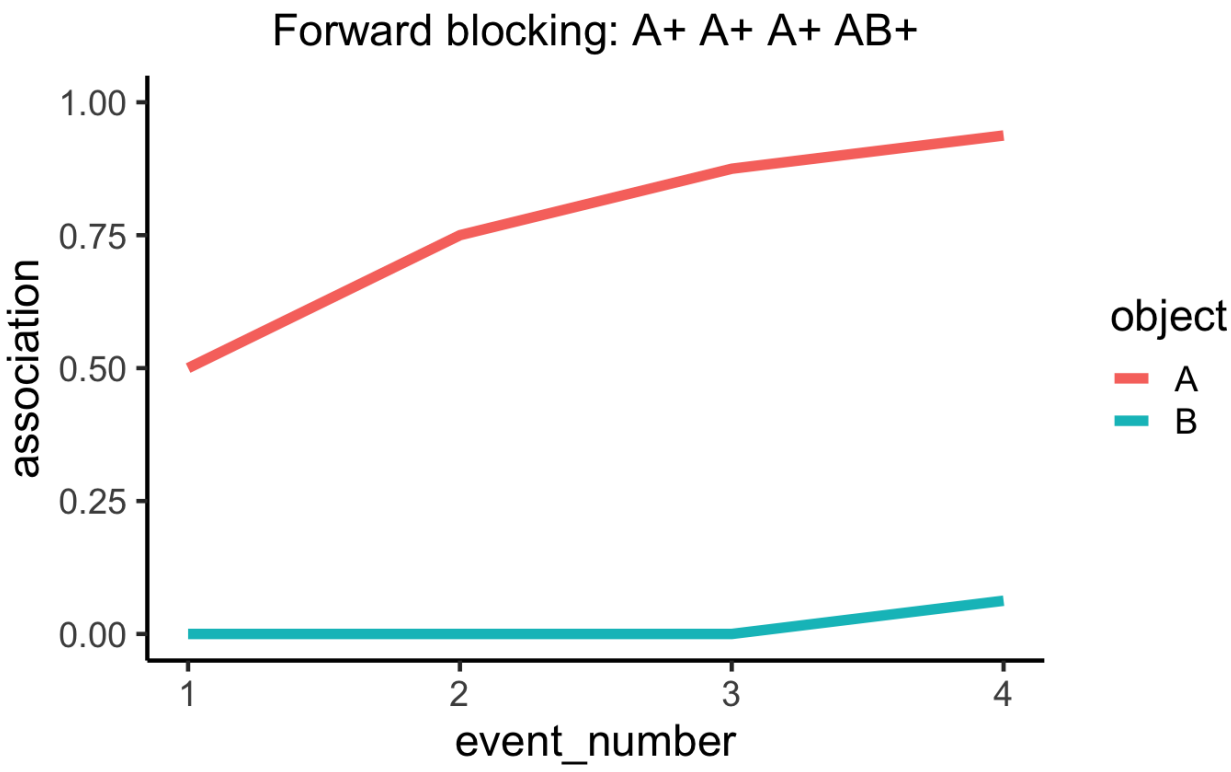
- ΔV_i: Change in the association value between stimulus i and the effect
- λ: 0/1 if effect is absent/present (in the binary case)
- α_i: The learning rate associated with cause
- β: The learning rate associated with the effect
- C: The set of causes that are present

Rescorla-Wagner

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"

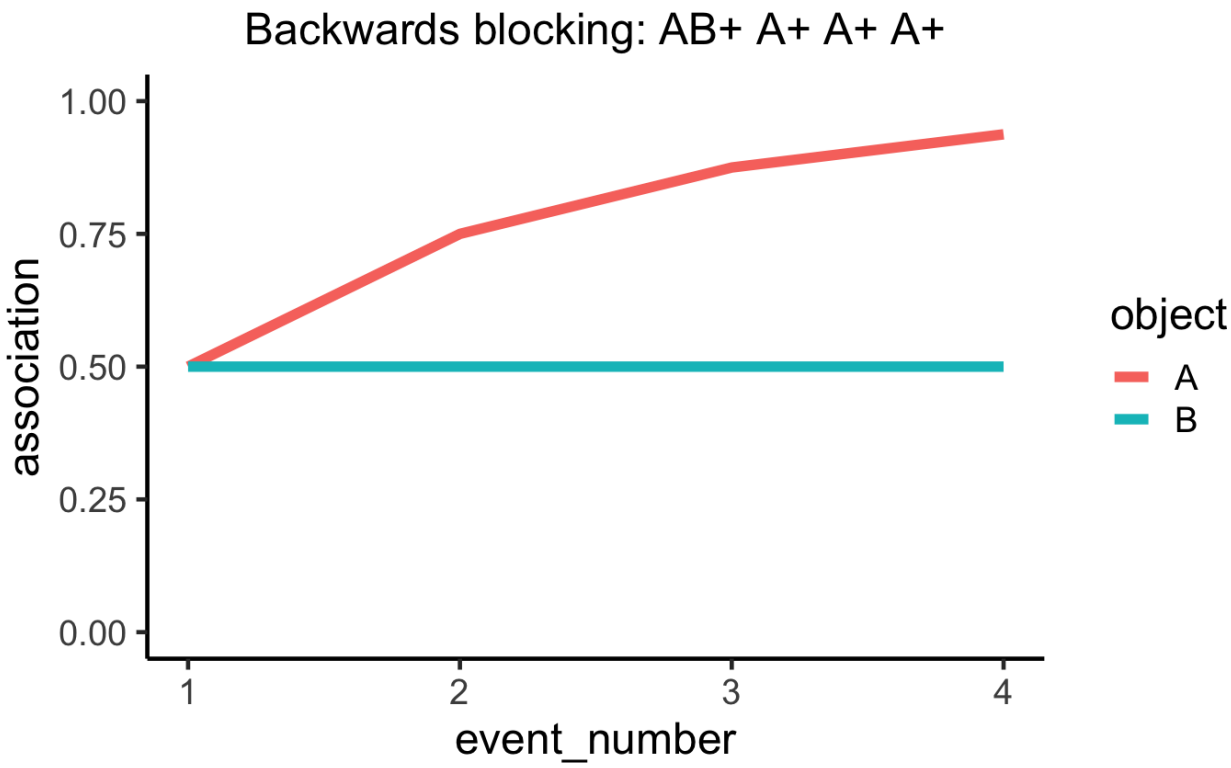
Rescorla-Wagner: Forward blocking



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

Other models

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

Coming up

- The assignment: Released Weds
- Next week (W7): no tutorials. Instead a Q&A session on MS teams (will be recorded).