Informative Counterfactuals

Adam Bjorndahl & Todd Snider
Cornell University

Logic, Grammar & Meaning
June 7, 2014

- We use counterfactuals all the time:

(1) If Alice had gone to the party, Bob would have stayed home.

(2) If the movie had been any good, I wouldn't have fallen asleep.

(3) If there hadn't been traffic, we would have been on time.

- We can use them to talk about things we know to be false or things we're uncertain about
- (1) typically implies that Alice didn't go to the party and Bob did
- It also communicates some relation between the two events

- There are different ways for the events in (1) to be related

(1) If Alice had gone to the party, Bob would have stayed home.

- Does Bob try to avoid Alice?
  - Maybe he's shy
  - Maybe he doesn't like her
- Do other circumstances prevent them from attending parties together?
  - Maybe they're a couple on a tight budget
  - Maybe Bob is actually Alice in disguise
- Does Alice try to avoid Bob?
  - Unlike the other scenarios, this one does not seem to jive with (1)...

Our Proposal

- Counterfactuals denote sets of relationships between events
- We use the mechanics of *structural equation models* to represent these relationships
- This provides a rich set of tools we use to define a typology of *explanatory strategies*
- Our analysis cleanly distinguishes two different kinds of 'backtracking'
- It also provides a principled account of certain kinds of mutually incompatible counterfactuals

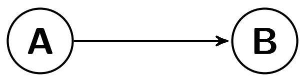
Outline

Outline

- To capture relationships between events, we use *structured possible worlds* (Starr 2014)
- Worlds are event variables, their values, and **dependencies** between them
  - Just like truth values, we can use the (non)existence of dependencies to discriminate among worlds
- We model these dependencies using Structural Equation Models (SEMs) as formalized in Pearl 2000

Structural Equation Models (SEMs)

- Allows for the modeling not only of variables but also dependencies
- Models consist of:
  - Nodes Circles Variables/Events
  - Edges Arrows Dependencies
    - Labeled with equations

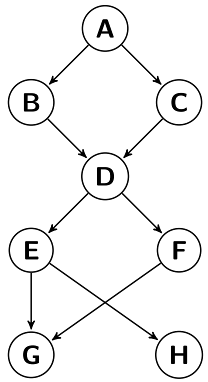


$B = \neg A$

- For convenience and simplicity, our examples are
  - Two-valued
  - Deterministic
- This framework and analysis also handles multi-valued and/or probabilistic systems

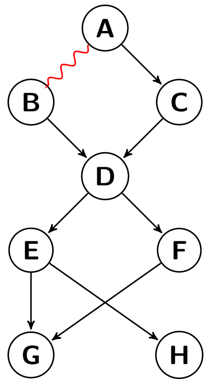
- Unlike Pearl, we take the SEM not as a *given* but as a *goal*
- Rather than structures within which to evaluate the truth of a counterfactual, we interpret SEMs as *candidate explanations*
- Counterfactuals denote sets of such explanations

Graph as *given*

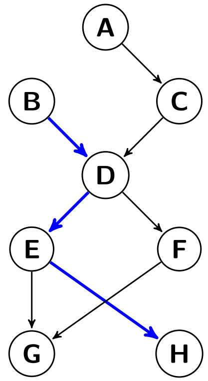


Start with a model

Hear
B > H

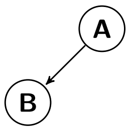


Intervene on the model



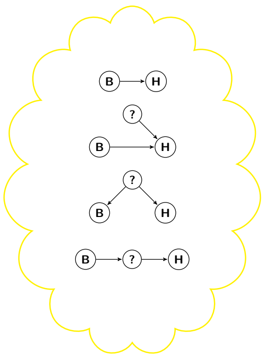
Evaluate truth in updated model

Graph as *goal*

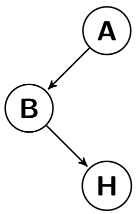


Start with minimal assumptions

Hear
B > H



Consider possible explanations



Build new model

Outline

- Counterfactuals assert some degree of covariance between the antecedent and consequent

(4) If I had pushed this button, the rocket would have launched.

- They implicate a direct (causal) dependence of consequent on antecedent ($C = A$)
- This implicature can be canceled (5) or strengthened (6):

(5) If I had pushed this button, the rocket would have launched, but pushing this button doesn't directly cause the rocket to launch.

(6) If I had pushed this button, the rocket would have launched, and (in fact) pushing this button directly causes the rocket to launch.

- Sometimes the direct dependency is problematic

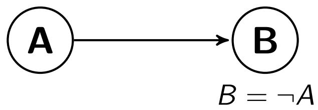
Rejecting explanations

- There are many reasons to reject an explanation (including the implicated direct dependency)
  - It might contradict prior knowledge
  - It might violate a law of good explanations
    - e.g. by positing an effect that is temporally prior to its cause
  - It might not satisfy the contextual parameter for specificity

- Any of these reasons might make us reject the simple direct dependency of the consequent on the antecedent
  - In other words, reject the $C = A$ edge
- But the counterfactual stipulates some covariance
- Trying to maintain the cooperativity of the speaker's contribution, we search for an explanation to make the counterfactual true
- Three possible ways to deal with this problematic dependence:
  - Additional cause
  - Common cause
  - Intermediate cause
- Call these **explanatory strategies**

(1) If Alice had gone to the party, Bob would have stayed home.

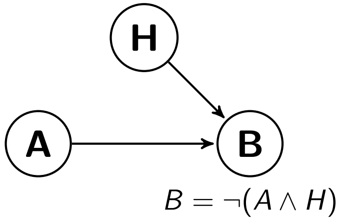
- The implicated simple dependency of (1):



- But Alice's attendance doesn't directly cause Bob to be elsewhere
- There are other *explanations*

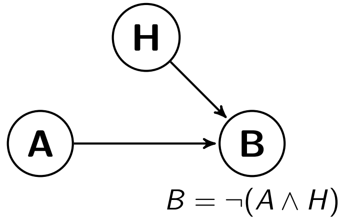
Additional cause

- The hearer might suppose that the consequent is dependent not solely on the antecedent but also on some additional cause
- For example, a common interpretation of (1) might lead one to believe that Bob hates Alice
- We can consider Bob's hatred of Alice as an additional node in our model



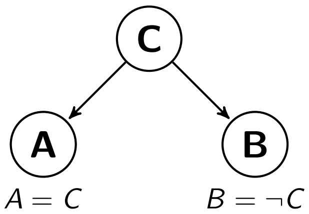
Additional cause

- The dependence of B on A is still present, but it's been modified
  - The $B = \neg A$ edge is no longer part of the model
- The antecedent and consequent covary only in the right H-conditions



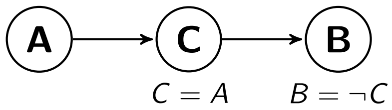
Common cause

- The hearer might suppose that the consequent isn't dependent upon the antecedent at all
- Instead, both antecedent and consequent depend on some common cause
- They still covary, but have no interdependence
- For example, imagine that Alice & Bob flip a coin to determine who attends



Intermediate cause

- The hearer might suppose that the consequent depends on the antecedent only by means of some intermediate cause
- The antecedent and consequent still covary, but without positing a direct causal dependency
- For example, imagine that Alice brings her cat wherever she goes, and Bob is deathly allergic to cats



A fourth explanatory strategy?

- Reversing the simple causal relationship also allows the antecedent and consequent to covary

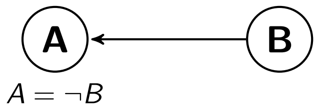
(1) If Alice had gone to the party, Bob would have stayed home.

- This classical *backtracker* has the consequent as the cause
- This model is rejected as an interpretation of (1)
- It's available with a double-auxiliary construction, as in (7)

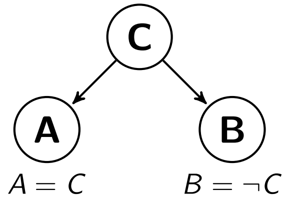
(7) If Alice had gone to the party, Bob would have had to have stayed home.

A note on *backtracking*

- Two different things referred to as *backtracking*



- Reversing causal direction
- Classic philosophy literature
- Double-aux environment



- 'Upstream' reasoning
- Recent psychology literature
- Available in (1)

How do we update with what we've learned?

- Once an acceptable explanation is found, we have to integrate it with our extant body of knowledge
- With structured possible worlds, our knowledge includes not just facts about variables but also dependencies
- We can model our knowledge as one persistent SEM
- When consolidating, we integrate dependencies, not variable values
- Counterfactuals *can* inform us about actual values via presupposition, accommodation
- We don't want to update with Alice's counterfactual attendance

- There are at least two mechanisms involved in consolidation

1. Addition
   - Extending the graph
   - Possibly add new nodes
   - Add new dependencies among nodes
2. Expansion
   - Looking deeper into the internal mechanism of a single node
   - Explode one node into multiple nodes
   - Retains incoming/outgoing dependencies of the original node

- After consolidation, deduce values of new nodes, if necessary

- Consider a world where Alice and Bob are married, and live with their young son Doug

(1) If Alice had gone to the party, Bob would have stayed home.

(8) If Alice had gone to the party, Doug would have been home alone.

- (1) and (8) are each felicitous individually
- A felicitous utterance of one precludes a felicitous utterance of the other
- Any account of how we update our knowledge with counterfactuals should explain this

- This consolidation process gives us insight into interactions between counterfactuals

(1) If Alice had gone to the party, Bob would have stayed home.

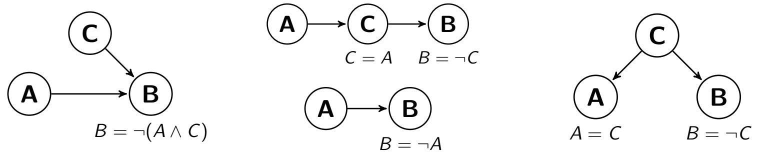
(8) If Alice had gone to the party, Doug would have been home alone.

- Updating with (1) adds a covariance between A and $\neg B$ to our knowledge base
  - Alice and Bob have opposite party-attendance values
- Updating with (8) requires that A and B have the same value
- The model we build after hearing one of (1)/(8) precludes the other

Outline

Conclusion

- We can use structured possible worlds to model dependencies, not just facts
- We propose using them to model informative counterfactuals
- Doing so gets us a natural way to typologize explanatory strategies



Conclusion

- Our analysis also neatly captures the distinction between different senses of *backtracking*
  - Classical philosophical backtrackers reverse the generally implicated direction of dependence
  - Recent psychological uses of the term refer to explanations including at least one instance of COMMON CAUSE
- Also provides insight into the mechanism that explains mutually infelicitous counterfactuals

Thank You!

Questions?

Eric Hiddleston. A causal theory of counterfactuals. *Noûs*, 39(4):632–657, 2005.

Boris Kment. Counterfactuals and explanation. *Mind*, 115(458):261–310, 2006.

Angelika Kratzer. An investigation of the lumps of thought. *Linguistics and Philosophy*, 12(5):607–653, 1989.

David K. Lewis. Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2(4):418–446, 1973.

David K. Lewis. Counterfactual dependence and time's arrow. *Noûs*, 13:455–476, 1979a.

David K. Lewis. Conditionals: Ordering semantics & Kratzer's semantics. Ms. Princeton University., 1979b.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge Univ Press, 2000.

William B. Starr. Structured possible worlds. Ms. Cornell University, 2014.

Pavel Tichý. A counterexample to the Stalnaker-Lewis analysis of counterfactuals. *Philosophical Studies*, 29:271–273, 1976.