

ACTUAL CAUSATION: ANALYSIS, MODELING & JUDGMENT

Christopher Hitchcock
California Institute of Technology

Co-authors

Josh Knobe
Yale,
Philosophy



Joe Halpern
Cornell,
Computer Science



The McDonnell Collaborative

Philosophy

John Campbell (UCB)
David Danks (CMU)
Frederick Eberhardt
(Wash. St. L.)
Clark Glymour (CMU)
Jim Woodward (Pitt)

Psychology

Alison Gopnik (UCB)
Tom Griffiths (UCB)
Tamar Kushnir (Cornell)
Tania Lombrozo (UCB)
Andy Meltzoff (Uwash)
Josh Tenenbaum (MIT)
Henry Wellman (UMich)

Interdisciplinary Research

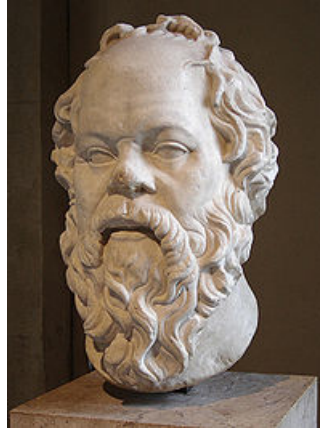
- ~~□ Division of labor~~
- Communication
- Brainstorming
- Consilience

Philosophical Analysis: The Platonic Model

What is...

- ☐ Justice?
- ☐ Knowledge?
- ☐ Virtue?
- ☐ Piety?
- ☐ Friendship?
- ☐ Love?
- ☐ Courage?

Plato



Republic

Socrates

...if a sane man lends weapons to his friend and then asks for them back when he is out of his mind, the friend shouldn't return them, and wouldn't be acting justly if he did

Republic

Polemarchus

...it is just to give to each what is owed to him

Socrates

...what he has lent to you is something that is owed to him isn't it?

Polemarchus

Yes

What is Causation?

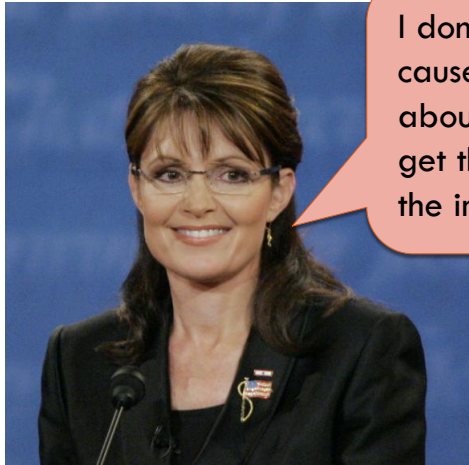
If cause were non-existent everything would have been produced by everything and at random. Horses, for instance, might be born, perchance, of flies, and elephants of ants...

Sextus Empiricus

Causation [is]...the cement of the universe

David Hume, John Mackie

On the Causes of Climate Change



I don't want to argue about the causes, what I want to argue about is how are we going to get there to positively affect the impacts we have.

On the Causes of Climate Change



I dream of a day when my children will live in a world without the shackles of cause and effect, when things will happen for no understandable reason, and we can respond randomly. What a magical time that will be!...Life will be full of surprises!

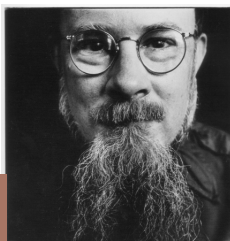
Counterfactual Definition

C causes E if and only if...

- If C had not occurred, E would not have occurred
(Anglo-American common law, David Hume, David Lewis)



David Hume



David Lewis

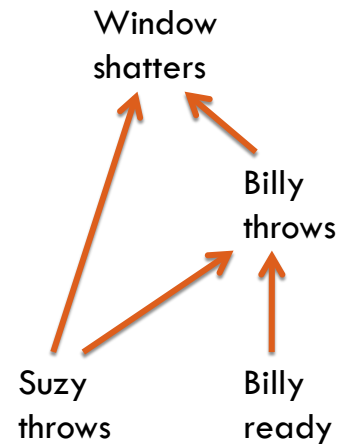
Counterexample

- Suzy throws a rock at a window, which shatters. Billy stands by, ready, willing, and able to throw his rock through the window if Suzy does not.
- Suzy's throw caused the window to break
- But if Suzy hadn't thrown, the window still would have broken (because Billy would have thrown his rock through the window)

Causal Models

- Modifiable Structural Equation Models
- Graphical Causal Models
 - Wright (1921)
 - Frisch (1938)
 - Haavelmo (1943)
 - Strotz and Wold (1960)
 - Spirtes, Glymour and Scheines (1993)
 - Pearl (2000)

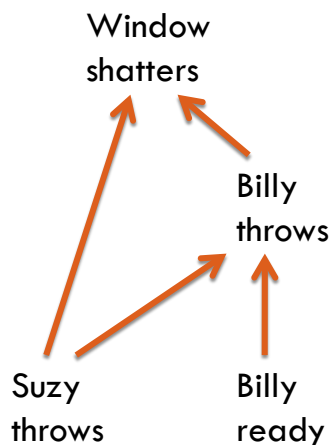
Causal Models



Window shatters
= Suzy throws or
Billy throws

Billy throws
= Billy ready &
not(Suzy throws)

Counterfactuals and Interventions



Window shatters
= Suzy throws or
Billy throws

~~Billy throws
= Billy ready &
not(Suzy throws)~~

Billy throws = yes

Terminology

- What the graph and equations represent is the *causal structure* of this scenario
- The judgment that Suzy's throw caused the window to shatter is a judgment of *actual causation*
- The challenge: to provide an account of *actual causation* in terms of *causal structure*

Metaphysics

- The causal structure is an objective feature of the world
- Its structure is more complex than 'C causes E'
- Actual causation is one specific causal notion
- Judgments of actual causation involve the retrospective assignment of causal responsibility

The Bastard Child

- They have sought a single causal relation that is fully objective, like causal structure
- Has the logical form 'C causes E', like actual causation
- Is revealed by our retrospective judgments of the form 'C caused E', like actual causation
- This runs together two different concepts

A Mistake

- Philosophers have been misled by the use of the simple verb 'cause' in actual causation claims
- Suzy's throw *caused* the window to shatter
- This led them to assume that *actual causation* is *causation* tout court

A Puzzle

- Knowledge of the *causal structure* lets us:
 - ▣ Predict what will happen for any initial conditions
 - ▣ Predict the effects of interventions on the system
 - ▣ Evaluate any counterfactuals
- What more could we want?
- Why, in addition to all that, do we also have the concept of *actual cause*?
- What *work* does this do?

A Puzzle

- This question emerges, in part, from having adequate representational tools
- It is motivated, in part, by psychological considerations
 - ▣ Cognitive limitations, need for efficiency
- It is partly a descriptive, psychological question: what role does this concept play in our cognitive economy?
- But it is also a normative question: what is this concept *good for*?

Knobe and Fraser

Subjects are then asked to rate the extent to which they agree or disagree with *one of the the* following statements:

- Professor Smith caused the problem
- The administrative assistant caused the problem

Subjects give a much higher rating to Professor Smith (2.2 vs -1.2 on a -3 to +3 scale)

Knobe and Fraser

The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take pens, but faculty members are supposed to buy their own...

On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later, that day, the receptionist needs to take an important message...but she has a problem. There are no pens left on her desk.

'Egalitarianism'

[T]his invariable sequence subsists...between a consequent and the sum of several antecedents; the concurrence of all of them being requisite to produce... the consequent. In such cases it is very common to single out one only of the antecedents under the denomination of Cause...The real cause, is the whole of these antecedents; and we have, philosophically speaking, no right to give the name of cause to one of them, exclusively of the others.

John Stuart Mill (1843)

‘Egalitarianism’

We sometimes single out one among all the causes of some event and call it ‘the’ cause, as if there were no others. Or we single out a few as the ‘causes’, calling the rest mere ‘causal factors’ or ‘causal conditions’... I have nothing to say about these principles of invidious discrimination.

David Lewis (1973)

One Hypothesis

- In the pen vignette (and others like it), subjects are confusing judgments of causation and moral responsibility
- In an online study of over 300 professional philosophers, Knobe and I found the same effect
- Moreover, the effect occurs even when there is no question of moral responsibility

‘Egalitarianism’

- Egalitarianism seems motivated by the idea that causation is objective
- This is appropriate for causal structure, but not necessarily for actual causation
- Hitchcock and Knobe: causal selection is important to understanding the function of actual causation

The Battery

A machine is set up in such a way that it will short circuit if both the black wire and the red wire touch the battery at the same time. The machine will not short circuit if just one of these wires touches the battery. The black wire is designated as the one that is supposed to touch the battery, while the red wire is supposed to remain in some other part of the machine.

One day, the black wire and the red wire both end up touching the battery at the same time. There is a short circuit.

Our Hypothesis

- What is driving judgments of actual causation in these examples is *normality* and *abnormality*
- Abnormal events are more likely to be picked out as causes

Normality

Normal

- Frequent, expected, (stereo-)typical
- Conforms with a norm
- Social, moral, legal, policy
- Proper functioning – organism, machine, social institution

Abnormal

- Rare, surprising, atypical
- Violates a norm

Normality

- As philosophers, our instinct is to distinguish these different senses of 'normal'
- In ordinary thought we jumble them together
- Often use one as a *heuristic* for another (in the sense of Kahneman and Tversky)

Cognitive Model

- Rather than constructing a separate causal model for each situation we encounter...
- We have templates for how things *normally* go
- When we encounter a novel situation, we only need to note the features that are *abnormal*
- Actual causes are the abnormal features that make a difference against the normal background
- It makes sense to target interventions that work in *normal* circumstances

Work in Progress

- Halpern and I are working on a project to model this process
- How to derive causal models using a template plus specifications of abnormal events
- Makes contact with a branch of logic called 'default logic'

Portable causal dependence

- Alison Gopnik, "Explanation as Orgasm"
- Jim Woodward, "Sensitive and Insensitive Causation"
- David Danks and David Rose, cognitive process model of actual causation
- Tania Lombrozo, "Explanation for Export"

Tania Lombrozo

Berkeley
Psychology



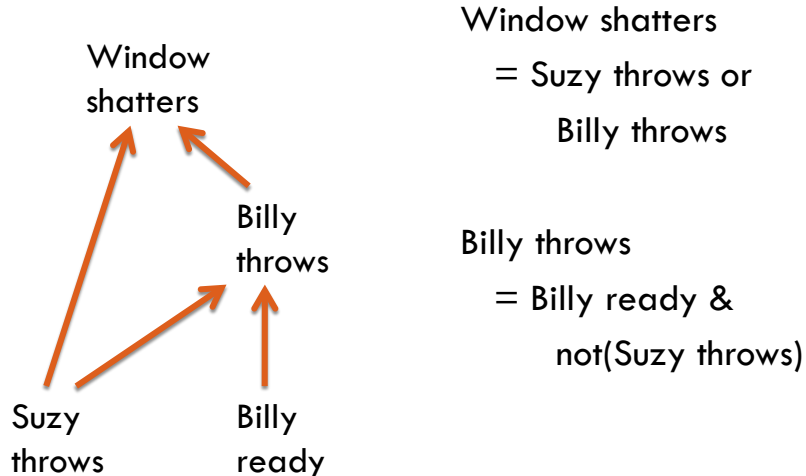
Lombrozo, "Explanation for Export"

- When do we accept functional explanations?
- How satisfying do we find functional explanations?
- When does the existence of a causal process increase our willingness to judge that something is a cause?
- When are we more inclined to judge that intentional actions are causes?

Lombrozo, “Explanation for Export”

- Interaction effects
- If an action is done unintentionally, we are more likely to judge it to be a cause if there is a causal process
- If there is no causal process (e.g. double prevention) we are more likely to judge an intentional action to be a cause

Actual Causation



Lombrozo, “Explanation for Export”

- She explained this in terms of *portability*
- Would the dependence relation hold in other circumstances?
- E.g. intentions are flexible: they can bring about their outcomes in a variety of ways
- Accidentally causing an outcome is often sensitive to the precise circumstances
- Simple causal processes tend to be insensitive to the precise circumstances, etc.

Actual Causation

- The problem: intervening on whether Suzy throws makes no difference for whether the window shatters
- Idea developed by Pearl 2000, Hitchcock 2001, Halpern and Pearl 2005 (related to an idea in Lewis 1986)
- Actual causation is counterfactual dependence in a ‘permissible’ modification of the original model

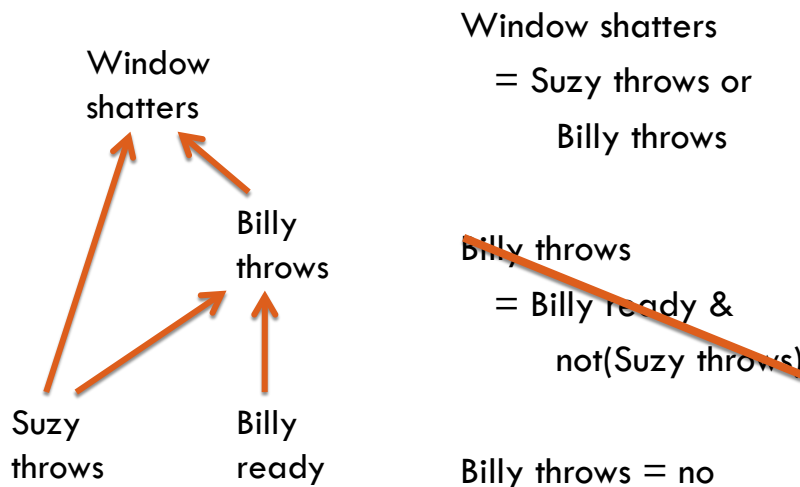
Actual Causation

- This account ends up being too permissive
- Hall 2007, Hitchcock 2007, Halpern 2008, Halpern and Hitchcock 2010: allow considerations of normality constrain the allowed modifications of the model
- You can modify a model to make it more normal, but not less normal

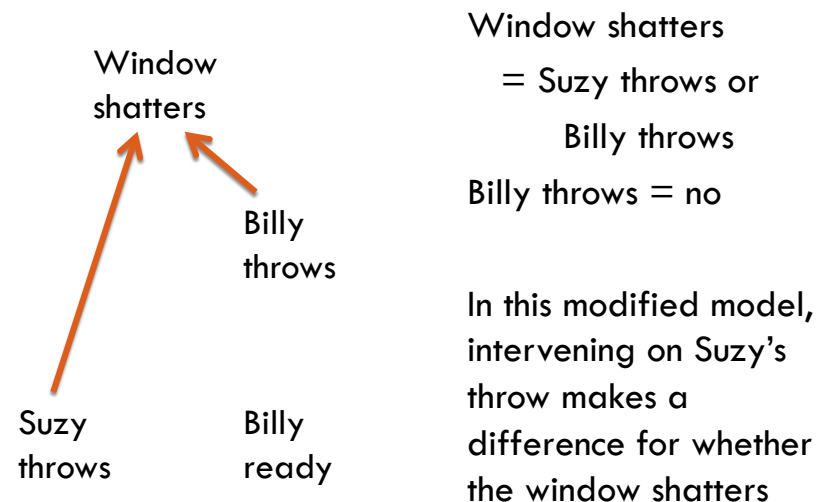
Counterfactual Availability

- Kahneman and Tversky on counterfactual thinking
- We are more likely to entertain counterfactual situations in which an abnormal feature of a situation is made normal than vice versa
- E.g., if Jones left work early, took his normal route home, was hit by another car...
- ...we are more likely to think 'if only he had left at his normal time', rather than 'if only he had taken some other route home...'

Counterfactuals and Interventions



Counterfactuals and Interventions

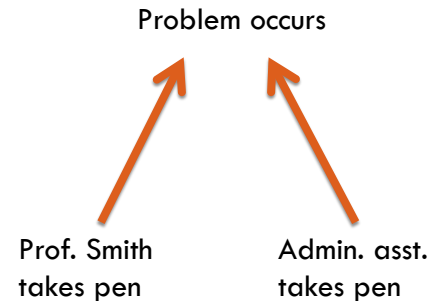


Graded Causation

- Halpern and Hitchcock (MS)
- Actual causation is not all or nothing
- Causes are 'graded' according to the normality of the model where they make a counterfactual difference

The Pen Vignette

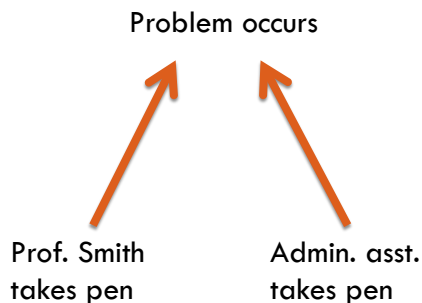
Problem = Prof & Admin



The problem counterfactually depends on both Prof and Admin

So the standard theory says both are causes, unlike the judgments of experimental subjects

The Pen Vignette



Problem = Prof & Admin

Normal: Prof = 0

Normal: Admin = 0 or 1

Prof makes a difference in the *normal* case when Admin = 1

Admin only makes a difference only in the *abnormal* case when Prof = 1

So Prof = 1 receives a higher 'grade'

