Sid Su
Causation in Counterfactuals - 1790 Words
Background
A conditional is any "if... then..." statement. They are common in almost every form of writing. For example, the conditional (1) is adapted from Shakespeare, and (2) is a Roman adage.

(1) If he hears you, then you will anger him.
(2) If you want peace, then prepare for war.

Conditionals are expressed symbolically as A→B (if A, then B). A is known as the antecedent, and B is known as the consequent.

Example (1) is an indicative conditional, and (2) is a command conditional. However, this paper will focus on counterfactual conditionals.

Counterfactuals are a type of conditional that includes the word "would" in the consequent. For example, (3) is a counterfactual conditional:

(3) If I were 10 feet tall, then I would not fit through the doorway.

Several theories seek to evaluate conditionals. All of them seek to answer two questions: When is a conditional true? And what "good" inferences can a conditional give us?

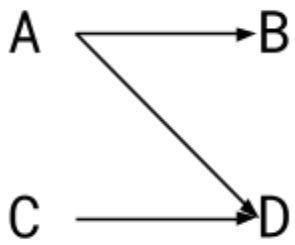
Introduction
I will begin by introducing the Causal Theory of counterfactuals. Then, present advantages of the causal theory over Possible Worlds semantics. Finally, I will discuss the main issue with causal models, defining what a "good model" is.

The Causal Theory of Counterfactuals
A *causal model* is a set of *random variables* and *structural equations* which take the form of mathematical relationships. Convention writes these equations with similar syntax to most computer programming languages. For instance, here's a simple causal model.

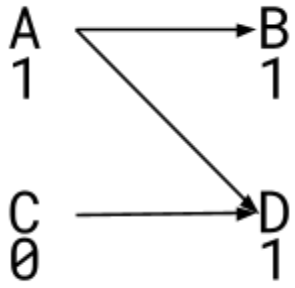
```
--RANDOM VARIABLES
A: Abstract variable
B: Abstract variable
C: Abstract variable
D: Abstract variable
--STRUCTURAL EQUATIONS
A=1          --A is true
B=A          --B is set equals to A
C=0          --C is false
D=MAX(A,C)   --D is the maximum value of A or C
```

A, B and D are all true and equal to 1, while C is false and equal to 0.

You can also represent *causal models* using *directed graphs*, so for our simple model



Or if we resolve the values,



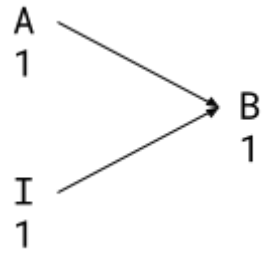
Causal models can be used to model counterfactuals. Consider (4)

(4) If I were born in 1989, then I would have been born the same year as Taylor Swift (A→B)

A reasonable set of *structural equations* would be

```
--RANDOM VARIABLES
A: Speaker being born in 1989
B: Speaker being born in the same year as Taylor Swift
I: Taylor Swift being born in 1989
--STRUCTURAL EQUATIONS
A=1                          --Set the antecedent as true
I=BORN("Taylor Swift", 1989)  --Compares TS birth year with 1989
B=MIN(A,I)                   --B is the minimum value of A and I
```

Using a *directed graph*



A is set equal to 1 because it's the antecedent. This is known as an *intervention*, because the speaker could be born in a different year, but is imagining a scenario where he was born in 1989.

Also notice that this model had to add a new random variable, 'I', to account for evaluating whether or not Taylor Swift was born in 1989. Taylor Swift was born in 1989, so the `BORN()` function returns true. If Taylor Swift had in fact been born in 1990, or any other year, then the `BORN()` function returns false, causing the `MIN()` function to return false as well. 'I' is an *exogenous variable*, because it is affected by factors outside the model, in this case, the birth year of Taylor Swift.

'A' and 'B' are both *endogenous variables*, because they are determined by factors in the model.

Advantages

The Causal Model Theory has the advantage of handling so-called Morganbesser-type cases better than Possible Worlds Semantics, a competing theory. Morganbesser-type cases are those with *indeterministic* events after the antecedent. Consider a situation where a man plays slots at a casino. When he calls it for the night and goes to leave, the person who goes on the same machine after him gets triple 7's, and wins a jackpot. The man disappointingly thinks

(5) If I had played another spin, I would have won a jackpot.

The following *causal model* emerges

```
--RANDOM VARIABLES
A: Playing another spin
B: Winning a jackpot
I: The next spin is a Jackpot
--STRUCTURAL EQUATIONS
A=1         --Intervention to set the antecedent to true
I=1         --The slot machine has a Jackpot in the next spin
B=MIN(A,B) --The man win a jackpot
```

The Causal Model is able to handle this counterfactual because it is able to hold fixed 'I', the next spin is a Jackpot. In a competing theory, Possible Words Semantics, the most "similar" antecedent containing possible worlds are selected, and "similarity" seeks to keep exact history the same for as long as possible, with approximate historical similarity having little to no value. Should the man, contrary to the actual world, decide to take another spin on the slot machine, then the next spin being a jackpot becomes indeterminate, because after the split from the actual world, the random event of a slot machine spin occurs again. Similarity has been broken, and the counterfactual becomes predicted as false, because there are possible worlds where the man doesn't win if he spins again.

Possible Worlds semantics have tried to respond to this by saying that sometimes approximate similarity matters, but this breaks other aspects of Possible Worlds Semantics. For instance, if approximate similarity matters, the counterfactual (6) becomes false.

(6) If I was on vacation, I would be in Paris

A model that allows for approximate similarity would result in a situation where the speaker does not go to Paris, but instead some variation of getting blocked from going on vacation such as his tickets getting cancelled, to get back to an approximately similar actual world where he stays at home, but if the speaker of (6) has the adequate means and desire to go, then it's unreasonable to declare (6) false.

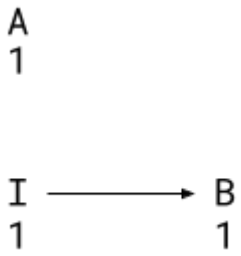
It's important to remember that Causal Models model *causation*. Consider (7), adapted from Shakesphere.

(7) If there were a rose by any other name, then it would smell just as sweet.

A causal model would look something like this

```
--RANDOM VARIABLES
A: Another name for rose in English
B: The differently named rose smelling just as sweet
I: Pseudo-rose has the same properties as a regular rose
--STRUCTURAL EQUATIONS
A=1      --Set the antecedent as true
I=1      --The name different, so everything else is same
B=I      --The rose smells just as sweet
```

As a *directed graph*



As the name of the pseudo-rose doesn't affect the smell (barring any psychological bouba/kiki connection to name), this *causal model* can still be evaluated as true, but it is independent of the antecedent

Causal Models also give an easy framework to evaluate more complex counterfactuals. For instance, consider (8)

(8) If Alexander the Great had gone West instead of East, then Rome would not have become an empire.

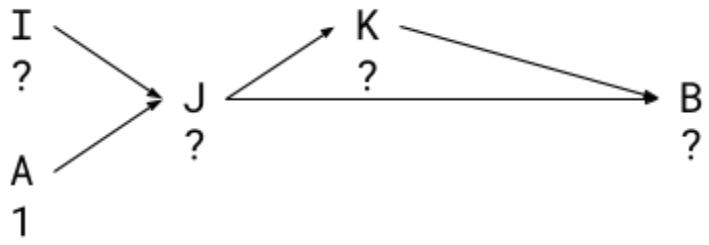
Under Possible Worlds semantics, one would have to pinpoint a break from the actual world, and "play out" what would happen from there. In the case of (8), some neurons in Alexander the Great's brain fire differently, causing him to cross the Adriatic Sea instead of the Hellespont in 334. However, using possible worlds doesn't help to evaluate whether or not the consequent holds for the antecedent. With a causal model, the factors as to whether or not Rome would become an empire can be considered with many paths to Empire.

Should the maniple system truly be able to best a Macedonian phalanx, then Rome could become an empire by defeating Macedon. Should the Romans get conquered, but much like in history Alexander's empire breaks up after his death, giving the Romans an opportunity to bully out their neighbours and becoming an empire. It's much easier to evaluate the consequent with the Causal Model Theory of Conditionals.

```
--RANDOM VARIABLES
A: Alexander the Great goes West instead of East
B: Rome becomes an empire
I: The Roman maniple can stand up to the Macedonian phalanx
J: Rome gets conquered by Alexander the Great
K: Rome reestablishes itself when Alexander the Great dies
--STRUCTURAL EQUATIONS
A=1           --Antecedent has intervention
I=?           --A debatable issue
```

```
J=MIN(A,I)    --Do the Romans win versus Alexander?
K=REEST(J)    --If they lose the battle, can they reestablish?
B=MAX(K,J)    --True for empire, false for no empire
```

As a *directed graph*



The Causal Model Theory creates an easy way to evaluate B. However, this is also the greatest criticism against this theory, because what is a "good model"?

The Issue of a "Good Model"
The Theory of Causal Models does run into a problem of what to include in a "good model".

Hiddleston presents "good model" as one that:
1. If the model includes an event, then the case needs to as well.
2. The Laws of the model are accurate enough to the case.
3. The model is complete enough to represent the causal relations in the case.

The Hiddleston notion of a "Good Model" is based on the idea that a given case will give a complete understanding of a situation, but this is not the case for some counterfactuals. Consider the (8) Alexander the Great case. The model I provided could lead to some good predictions of true or false, but it doesn't account for all situations. What if after the death of Alexander the Great, Rome is able to reestablish itself, but it instead takes the form of a Greek-style city-state instead of an empire? What if Alexander the great crosses the Adriatic, but upon entering Italy, he is not used to the climate and dies of disease the day before the battle, so there is no battle. What if Alexander the Great becomes enchanted with Italian country life, and leaves his empire to Rome, thus Rome inherits the Macedonian Empire? Should we assign new Random Variables 'L', 'M' and 'N'? This case seems infinitely more definable, so it's impossible to say what is complete enough.

However, I believe this shortcoming to not be because of a shortcoming of the Theory of Causal models, but because by definition counterfactuals quantify over events that did not happen. Trying to define with specificity what does or does not matter is only an educated guess given the information that we do have. Ideally, there we would limit ourselves to events that have direct evidence. So to answer the question of whether or not 'L', 'M' and 'N' should be added as variables in relation to (8), 'L' should, because

Rome at the time before the hypothetical Macedonian Invasion was a city-state, 'M' shouldn't because there's no reason to believe that the Italian climate is too hostile and 'N' shouldn't as well, because there's no reason to believe that Alexander the Great had Italian sympathies. Causal models are a useful and fascinating tool when trying to evaluate a counterfactual conditional.