Causation and conditionals

Christian Wüthrich

http://philosophy.ucsd.edu/faculty/wuthrich/

130 Metaphysics
Fall 2012
**Terminology**

If **ANTECEDENT**, then **CONSEQUENT**.

**Definition (Types of conditionals)**

A *counterfactual conditional* is a conditional of which the antecedent is not true, expressing (in the subjunctive tense) what would be the case, *if something were the case that is not*. An *indicative conditional* is a conditional of which the antecedent may or may not be true, expressing what is in fact the case, *if its antecedent is in fact true*. 
On the difference between indicative and counterfactual conditionals

You can accept ‘If Oswald didn’t kill Kennedy, someone else did’ as true, while rejecting ‘If Oswald hadn’t killed Kennedy, someone else would have’ as false.


- antecedent of counterfactual conditional expresses state of affairs which does not actually obtain, i.e. it is ‘contrary to fact’
- auxiliary verbs are ‘subjunctive’ in mood (though not all subjunctive conditionals are counterfactuals)
- close connection to causation
- counterfactuals are modal in character, and will require metaphysics of modality to get a full treatment
What are the truth conditions of counterfactual conditionals of the general form ‘If it had been the case that p, then it would have been the case that q’? Truth conditions are given by stating necessary and sufficient conditions for the truth of counterfactuals without using counterfactual conditionals.

‘If it had been the case that p, then it would have been the case that q’ is true if and only if q is true in the closest possible world in which p is true. (Lowe, 140)
Some comments

- analysis assumes that there is a possible world closest to the actual world in which a given proposition $p$ is true (with $p$ arbitrary)

- Of course, there could be ties, or it could be that for every possible $p$-world, there is a $p$-world closer to the actual world.

  $\Rightarrow$ analysis would have to be refined, as Lowe shows (140-142), such that ‘conditional excluded middle’ fails and a new type of counterfactuals (so-called ‘might’ counterfactuals) arises
(In)valid inferences for conditionals

On the above analysis, some inferences patterns which are arguably valid for indicative conditionals do not hold for counterfactual conditionals. In particular, the following are invalid for counterfactuals:

(Trans)

\[ p \rightarrow q, \quad q \rightarrow r, \text{ therefore } p \rightarrow r. \]

(Contr)

\[ p \rightarrow q, \text{ therefore } \neg q \rightarrow \neg p. \]

(Stren)

\[ p \rightarrow q, \text{ therefore } (p \& r) \rightarrow q. \]

i.e., counterfactuals are not transitive, not contraposable, and not subject to strengthening.
The failure of strengthening

\[(\text{Stren})\]

\[p \rightarrow q, \text{ therefore } (p \& r) \rightarrow q.\]

intuitive that this should fail for counterfactuals, as we can accept ‘If I had struck this match, it would have lit’ as true without also accepting ‘If I had struck this match and held it under water, then it would have lit’ as true
The failure of transitivity

(Trans)

\[ p \rightarrow q, \quad q \rightarrow r, \quad \text{therefore} \quad p \rightarrow r. \]

- less obvious; famous counterexample from Robert Stalnaker: “‘If J. Edgar Hoover had been born in the Soviet Union, then he would have been a communist; and if J. Edgar Hoover had been a communist, then he would have been a traitor; therefore, if J. Edgar Hoover had been born in the Soviet Union, he would have been a traitor.’” (Lowe, 143)—may also be considered an equivocation.

- exercise (on the blackboard, or at home): think about this in terms of the analysis above, involving possible worlds, etc.

- issue not academic, given the relevance of counterfactuals for causation, and given the fact that most take causation to be a transitive relation.
‘Closeness’ between possible worlds as measured by their similarity: “one world, $w_1$, is ‘closer’ to the actual world than is another world, $w_2$, just in case $w_1$ is, overall, more similar to the actual world than $w_2$ is.” (145)

But just how should we measure similarity?

It won’t do to just count propositions which are true in both worlds. (There are infinitely many worlds, not all propositions are equally important, e.g., propositions expressing laws may be more important than propositions expressing particular facts, etc)

Most likely: we will not be able to find truly objective measure of similarity, as it may also depend on subjectively given relative weight of different traits...
The trouble with deterministic worlds

That laws are always more important than particular facts when assessing similarity is far from clear, as the consideration of deterministic worlds shows:

*If the laws of this, the actual world, are deterministic in character, and thus permit no exceptions, then any world which is exactly similar to the actual in respect of its laws, but which differs from the actual world in respect of any individual matter of fact which obtains at a certain time, must also differ from the actual world in respect of individual matters of fact which obtain at earlier times—and thus must differ in respect of a great many individual matters of fact.* (147)

⇒ if two deterministic worlds with identical laws differ in just one particular matter of fact, they have to differ in uncountably many; perhaps a small difference in laws would make for greater similarity
The philosophy of causation

Acknowledgments: I thank Michael Baumgartner for making pertinent material available.

- **Vantage point:** concept of causation not primitive, i.e. it’s reducible to non-causal concepts
- ∃ more fundamental relations than causal relation
- **Central issue:** find necessary and sufficient conditions for the presence of causal process
- What are the truth conditions of propositions such as ‘x causes y’ or ‘x is causally relevant for y’?
- **wanted:** biconditional of form

\[ x \text{ causes (is causally relevant for) } y \iff P \]  (1)

with \( P \) being a complex expression not containing causal concepts

- **Relata of causal relation:** events, facts, propositions, or ‘factors’
- **General vs singular causation**: type vs token
- ex for general causation: ‘Smoking causes cancer’
- ex for singular causation given by concrete instance of general dependence: ‘John’s smoking causes John’s lung cancer’
- **Properties of general and singular causal relation?**
  - symmetrical, asymmetrical, non-symmetrical?
  - reflexive, irreflexive, non-reflexive?
  - transitive, intransitive, non-transitive?
- Lowe: asymmetrical, irreflexive, transitive
- **Principles of causation?**
- Which of the following principles are valid?
Principles of Causation

**Principle (Causality)**

*Every event has a cause.*

**Principle (‘Determinism’)**

*Given the same types of causes, the same types of effects are instantiated.*

The principles of causality and of ‘determinism’ usually taken to jointly imply

**Principle (Causal Determinism)**

*Every event is causally determined or necessitated by antecedent events and conditions (together with the laws of nature).*

Causal determinism is arguably violated in quantum mechanics.
Analysis (Sufficient conditions)

A factor x is causally relevant for a factor y if and only if x is a sufficient condition for y (x → y).

- Counterexample: Whenever the barometer drops, the weather gets bad.
- Counterexample: Whenever a match is struck and it has a blue head, it is lit.
- Objection: presence of inhibiting factors (no oxygen, etc) may preclude effect
The ‘Humean’ analysis of causation

Analysis (David Hume)

“Event c was a cause of event e if and only if (a) c preceded e and (b) c and e are, respectively, events of $T_1$ and $T_2$ such that every event of type $T_1$ is followed by an event of type $T_2$.” (158)

- Problem 1: determining how ‘types’ of events get individuated; why should this be possible in general?
- Problem 2: there are cases in which RHS is satisfied even though events are causally unrelated, e.g., if events are very rare;

$\Rightarrow$ “fails to distinguish between genuinely causal conjunctions of events and purely accidental conjunctions of events”

also supported by claim that analysis does not imply counterfactual ‘If c had not occurred, e would not have occurred’, because it only trades in what is actually the case
Necessary conditions (Hobbes)

Analysis (Necessary conditions)

A factor $x$ is causally relevant for a factor $y$ if and only if $x$ is a necessary condition for $y$ ($\neg x \rightarrow \neg y$).

- Counterexample: Match may catch fire even if not struck
- General objection: many effects can be produced by alternative factors
Analysis (INUS-conditions)

A factor $x$ is causally relevant for a factor $y$ if and only if $x$ is an Insufficient but Necessary part of a condition which is Unnecessary but Sufficient for $y$. 
Counterfactuals (Hume/Lewis)

Analysis (Counterfactuals)

A factor $x$ is causally relevant for a factor $y$ if and only if had $x$ not occurred, $y$ would not have occurred.

- Counterexample: even if match had not been struck, it still could have caught fire
- Counterexample: is ‘If Napoleon’s birth had not occurred, then his death wouldn’t have occurred’ really causal?
- also events related to one another as part and whole may satisfy the RHS without being causal (arm and hand both going up)
- General objection: truth conditions of counterfactuals highly non-trivial
Analysis (Interference/Manipulation)

A factor $x$ is causally relevant for a factor $y$ if and only if one can manipulate the world such as to make $x$ occur in order to produce $y$.

- Objection: account is circular because it assumes interference/manipulability to be more fundamental than causation; but how should interference/manipulation be understood if not in causal terms...??
- Objection: unduly anthropocentric
- Newer accounts (Woodward, and formal version of essentially this approach by Glymour/Spirtes/Scheines and Pearl) don’t claim to reduce causation to manipulation, but merely regard manipulation as a systematic mean to identify causal relations in some contexts.
Probabilistic causation (Suppes)

**Analysis (Probabilistic causation)**

*A factor $x$ is causally relevant for a factor $y$ if and only if $P(y|x) > P(y)$ (or, equivalently, $P(y|x) > P(y|\neg x)$) (and some screening off condition is satisfied).*

- **Motivation:** get an account of the many cases of imperfect knowledge about deterministic systems (or inherently chancy systems)
- **Objection:** spurious correlations (lung cancer and yellow fingers) $\Rightarrow$ screening off condition: If $P(y|x \land z) = P(y|z)$, then $z$ is said to **screen $x$ off from $y$**. Roughly, $z$ renders $x$ probabilistically irrelevant to $y$.
- **Objection:** Mr Jones and the golf ball (bad shot causes ball to go into cup for hole in one)
Objection: what causes events of probability zero?

Objection: prob raising is symmetrical (if $P(y|x) > P(y|\neg x)$, then $P(x|y) > P(x|\neg y)$)

⇒ attempt combination of probabilistic approach with counterfactual, or with interference/manipulation accounts into formal causal modeling theories (as Glymour/Spirtes/Scheines and Pearl do)
Transference (Fair, Salmon, Dowe)

Analysis (Transference)

A factor $x$ is causally relevant for a factor $y$ if and only if there is a transfer of energy or momentum from $x$ to $y$.

- Motivation: anchor causation in physical world, deliver objective analysis
- Objection: switching on the light causes the bulb to light up but there is no transfer of energy or momentum from the switch to the bulb, the flu virus causes a heightened body temperature without the virus transferring energy to the body
- Objection: cases of transmission of energy/momentum that would not ordinarily be considered causal
- Objection: sociohistorical processes not amenable to this analysis (e.g. what kind of energy transfer occurs between the causes of the French Revolution and the events in and around the Bastille on 14 July 1789?)
∃ many versions of transference analyses, e.g. Dowe’s conserved quantities theory (1992, 1995):

1. A causal interaction is an intersection of world lines which involves exchange of a conserved quantity.

2. A causal process is a world line of an object which possesses a conserved quantity.

Dowe 1995, p 323

- Objection: omissions and preventions (e.g. the neighbour’s failure to water my plants during my absence caused the plants to die)

- Objection: worries about conserved quantities (can only be defined in closed systems, which shouldn’t then be defined as a system not involved in causal interactions with the environment)

- Objection: worries about this “empirical analysis”, and about reduction
Analysis (Simple Counterfactual Analysis of causation—SCA)

“Event c was a cause of event e if and only if c occurred and e occurred and if c had not occurred, then e would not have occurred.” (Lowe, 161, 174)

- SCA fails for two events which stand in the part-whole relation as they will come out as causing one another.

⇒ amend the analysis:

Analysis (SCA+)

“Event c was a cause of event e if and only if (a) c and e are wholly distinct events, (b) c occurred and e occurred, and (c) if c had not occurred, then e would not have occurred.” (Lowe, 174)
1. Given that simultaneous causation is possible, among other things, how can we explicate what it means for two events to be ‘wholly distinct’?
   - “The obvious and correct thing to say is that two events are wholly distinct just in case there is no event which is a common part of both of them.” (175)
   - But that presupposes we can identify and individuate parts of events!

2. \( \text{SCA}^+ \) cannot always distinguish between cause and effect: the only asymmetric clause in \( \text{SCA}^+ \) is (c); but generally, if (c) holds, the following clause also holds: (c’) if \( e \) had not occurred, then \( c \) would not have occurred.
   - If that’s the case, what does this entail?
regarding (2): Threatening symmetry

⇒ event e was also a cause of event c!
⇒ threatening symmetry of causation (or at least non-symmetry)
  - Example: reliable mechanism from button to explosion of bomb, the explosion was the cause of the button being pressed!
  - We should not stipulate the cause to be earlier, as this would a priori rule out backward causation (as would the requirement that the relevant counterfactual should not be ‘backtracking’).
⇒ need an account of the direction of causation which does not merely stipulatively equate it with the direction of time
SCA+ sometimes fails to distinguish between a pair of events related by causation and a pair of events which are both effects of a common cause.

- Example: bomb example, where pressing the button not only detonates the bomb, but also activates a warning light.
  ⇒ SCA+ entails that the bomb caused the light, and vice versa.
- Proposed solution makes reference to ‘causal processes’, thereby presupposing a prior grasp of what this means.
  ⇒ Threatening circularity.
4 cases of causal overdetermination pose difficulties for SCA+:

(4a) **actual overdetermination**: “e has another actual cause, d, in addition to c, such that, even if c had not occurred, d would still have occurred and would still have caused e.” (179)

(4b) **pre-emption**: “another event, d, occurs, such that although d is not actually a cause of e, if c had not occurred, then d would still have occurred and would then have caused e.” (ibid.)

(4c) **fail-safe case**: “if c had not occurred, then another event, d—which did not actually occur—would have occurred and then would have caused e.” (ibid.)

- main difference: status of event d
- common to all: “if c had not occurred, then d would have occurred and would have caused e, so that e would still have occurred” (ibid.)

⇒ RHS of SCA+ not true, even though we said that c was e’s cause!
A possible response to causal overdetermination

Only responses to (4b) are considered, which Lowe argues are the hardest cases.

- The event that would have been caused by \( d \) had \( c \) not occurred to cause \( e \) would not have been \( e \), but rather some very similar, but numerically distinct, event \( e' \).

\[ \Rightarrow \] need to identify events across possible worlds, i.e., we need a criterion of transworld identity for events

- But what belongs to an event’s ‘individual essence’?
- Probably not its exact time of occurring.
- The totality of the event’s causes and effects?
- Lowe: this is circular, as it appeals to a prior understanding of what the event’s causes and effects are.
- Lowe: furthermore, causal relation obtain by metaphysical necessity
- CW: in a way, this trivializes causation!
A very different strategy to deal with causal overdetermination is to modify the analysis.

David Lewis: ‘complex counterfactual analysis of causation’ (CCA, below)

Definition (Counterfactual dependence)

An event e is counterfactually dependent upon another event c just in case if c had not occurred, then e would not have occurred.

Analysis (CCA)

“Event c was a cause of event e if and only if (a) c and e are wholly distinct events, (b) c occurred and e occurred, and (c) a chain of counterfactually dependent events linked c to e.” (Lowe, 183)
CCA and pre-emption

- Nota bene: CCA rests on presumption that counterfactual conditionals are non-transitive, for otherwise, the clause (c) of CCA would collapse to the clause (c) of SCA+ (exercise for the reader, cf. p. 184)

- CCA avoids most pre-emption cases: on CCA, that \( c \) caused \( e \) is perfectly consistent with its being true that \( e \) would still have occurred even if \( c \) did not (given the non-transitivity)

- CCA: \( d \) was not actually the cause of \( e \) (there may have been no chain from \( d \) to \( e \))

- CCA: there might have been a chain from \( d \) to \( e \) had \( c \) not occurred

\[ \Rightarrow \] on CCA, \( d \) would have caused \( e \), had \( c \) not occurred

- Even permits to uphold the transitivity of causation (despite the non-transitivity of counterfactuals)—exercise!

- However: cases of late pre-emption can’t be dealt with
The following two are difficulties for SCA+, CCA, as well as for any counterfactual analysis:

1. **Circularity**: “we very often need to appeal to causal considerations for the purpose of interpreting counterfactual conditionals and evaluating them as being true or false” (186) (e.g., similarity of worlds in terms of natural laws, which are causal)

2. **Objectivity**: judgments as to the similarity of worlds rely on respects in which worlds may or may not be similar and hence are partially subjective (Lewis’s helps himself to a primitive “relation of comparative over-all similarity among possible worlds”; Lewis, 559)