Counterfactual Reasoning, Qualitative: Philosophical Aspects
Stathis Psillos, University of Athens, Athens, Greece
© 2015 Elsevier Ltd. All rights reserved.
This article is a revision of the previous edition article by A. Hájek, volume 4, pp. 2872–2874, © 2001, Elsevier Ltd.

Abstract

This article reviews the two major approaches to counterfactual conditionals: the metalinguistic or ‘support’ approach and the possible worlds approach. It identifies the major problems they face and explores the fact that the core idea behind counterfactual reasoning is to assert that there are not good inductive reasons to affirm simultaneously a realization and the possibility of an exception to it. It also examines the role of counterfactuals in causal inference, causation, and laws of nature.

Subjunctive conditionals or counterfactual conditionals are probably as old as language itself: since they give speakers the means to talk about what would or might happen if something happened or if events had been different. In ordinary language, they have the form:

If x were (not) the case, then y would (not) be the case,
or
If x had (not) been the case, then y would (not) have been the case.

Subjunctive conditionals leave open the possibility of the realization of whatever is expressed in the antecedent, for example, if John were to come to the party, Mary would not go.

If we assume that classical semantics apply to indicative conditionals (the indicative conditional is true if either the antecedent is false or the consequent is true), trying to apply classical semantics to counterfactuals leads to their trivialization: Given the actual falsity of the antecedent of a counterfactual, both the counterfactual with the actual consequent and the counterfactual with the negation of the actual consequent end up being true.

Example: Given that the vase was not struck with a hammer, both of the following two conditionals (treated as material conditionals) are true:

If this vase had been struck with a hammer, it would have broken, and
If this vase had been struck with the hammer, it would not have broken.

The failure of the three principles and this unwanted consequence is a reflection of the view that classical semantics apply to counterfactuals. But then, what is the right semantics for counterfactuals? What are the truth-conditions of a counterfactual conditional? Or, at least, what are their assemblibility conditions?

Counterfactuals fail a number of principles that indicative conditionals satisfy. Most important, they are nonmonotonic, that is, they fail the principle of strengthening of the antecedent:

X→Y does not entail X&Z→Y.

Example: If John had been poisoned, he would have died. This does not entail: If John had been poisoned and taken an antidote he would have died.

X→Y and Y→Z does not entail X→Z.

Example: If John had gone to the market, he would have taken the bus; if John had taken the bus, then he would have gone to his office. These two do not entail: If John had gone to the market, then he would have gone to his office.

X→Y does not entail not→Y→not→X.

Example: If John had lived in a Euro-zone country, he would have used Euros. This is not equivalent to: If John had used Euros, he would have lived in a Euro-zone country.

The failure of the three principles and this unwanted consequence is a reflection of the view that classical semantics apply to counterfactuals. But then, what is the right semantics for counterfactuals? What are the truth-conditions of a counterfactual conditional? Or, at least, what are their assemblibility conditions? This problem came under sharp focus in the 1940s, when philosophers started to realize that the concept of counterfactual conditionals is instrumental for the explication and understanding of a number of other philosophical concepts. As Nelson Goodman put it in one of the first papers to deal with this issue, “(…) if we lack the means of interpreting the counterfactual conditionals, we can hardly claim to have any adequate philosophy of science” (1947, 113).

Note that in assessing a counterfactual assertion X→Y, we should replace, as it were, the actual nonoccurrence of X with the supposition that X has occurred. But given that the laws of nature and the actual course of events led to non-X, in supposing the actual occurrence of X, we need to make counterfactual suppositions concerning either the laws or the actual course of events, such that X actually occurred. In particular, we have to assume that either some laws were broken (so that X did happen after all) or that some actual particular matters of fact did not occur. Hence, in specifying the semantics of counterfactuals, we have to take into account considerations concerning the laws of nature and other particular matters of fact leading the conditions specified in the antecedent of the counterfactuals.

There are two major views concerning the semantics of counterfactuals, the first being introduced by Goodman himself, whereas the second was developed by Robert Stalnaker.
2 Counterfactual Reasoning, Qualitative: Philosophical Aspects

and David Lewis (but introduced by William Todd in 1964). Let us examine them in turn.

### The Metalinguistic or ‘Support’ View

On the first major view, known as ‘support view’ or ‘metalinguistic view’, a counterfactual conditional X → Y is an elliptic or telescoped argument (or a linguistic construction about an argument) such that the antecedent X (taken in its indicative) can be true together with suitable auxiliary premises entails the consequent (taken in its indicative form). Hence, X → Y should not be taken to be a statement at all; its assertoric content is captured by the following argument type:

\[ X \& S \& L \rightarrow Y \]

where L are statements capturing laws of nature and S are singular statements capturing background or collateral conditions that should be ‘cotenable’ with the antecedent X and express necessary conditions for the consequent to follow. X&S&L (materially) imply Y,

where these conditions (collectively designated by S) are not sufficient for the lighting of the match; various laws are required (collectively designated by L). Hence, in asserting the counterfactual ‘if this match had been struck, it would have lit’ (Y), for a struck match to light, it is necessary that the match is well made; that it is dry; that there is oxygen; and so on. But the main problem with this view concerns the characterization of the relevance relation when it comes to the background-collateral conditions. It cannot be too permissive. If we allowed all true statements to be relevant to the argument, the falsity of the antecedent X (which is actually false) would be relevant, too; however, the consequent Y then would be trivially true. It cannot be too restrictive either. The consequent of the counterfactual is false as well. Hence, not-Y is the case. It is not hard to see that given (X&S&L → Y) and not-X and L, it follows (by obvious steps) that X → not-S. (Assuming, for simplicity that S is ‘the match is dry’, the conclusion would be as follows: if the match is struck, it will not be dry!) The point then is that only those background or collateral conditions that are ‘cotenable’ with the antecedent should be admitted. But which are they? Those conditions S, which are such that if X had been true, S would have been true, too. This is a counterfactual assertion and Goodman thought that this kind of circularity entails the metalinguistic view.

Goodman thought that this kind of circularity is captured in the subjunctive mood: ‘If this were an A-situation, it would be accompanied by a B-situation’. The counterfactual content of a law then is seen as a ‘contextual implication’ of a law-statement.

This idea of ‘contextual implication’ is captured by the supposition view of counterfactuals, which is akin to (although interestingly different from) the metalinguistic view, as this was developed by John Stadlin (1973). According to this view, we assert something like X → Y is to assert Y within the scope of the supposition that X. In other words, we suppose X and then we envisage various possibilities and consequences. This idea brings to light the contextuality of counterfactual conditionals, which is not resolvable without some degree of arbitrariness: X did not happen; suppose that X did happen, what else do we have to assume or suppose? What features of the background (including laws and particular matters of fact) should we retain or change? There is no uniquely determined answer to this question, although contextual matters (including a fuller specification of the antecedent of the conditional) might (and as a rule do) help us.

Example: If I had let go of this stone, I would have fallen to the ground true or false (or assertible or not assertible)? It depends on the context. There are certain conversational contexts in which it would be false to assert it, for example, if this were a precious stone and the owner was very careful with it, so if the stone were to be let go, she would have caught it in midair. Another example: Consider the following pair of counterfactuals: ‘If Julius Caesar had been in charge of United Nations Forces during the Korean War, then he would have used nuclear weapons’ and ‘If Julius Caesar had been in charge of United Nations Forces during the Korean War, then he would have used catapults’. Only contextual assumptions can tell us which one, if any, and in what context, is true (or assertible).

The supposition view takes it that counterfactuals are not truths about possible worlds but are ways to express an attitude.
toward a possible state of affairs made within the scope of a supposition. The cotenability problem is solved by a legiti-
mate weakening of the cotenability condition. The cotenable
premises are taken to be those that are thought to be coten-
able in a certain conversational context. But how, within this view,
can it be explained that laws support counterfactuals while
ascribing an interest not? The difference is not in the context of a
statement expressing a causal law as opposed to the content of
an accidentally true generalization. Rather, the difference is
in the circumstances under which it is legitimate (or accept-
able) to combine the supposition that X is the case (i.e., the
antecedent of the counterfactual X \( \rightarrow \) Y) with the law L as
premises of the relevant argument whose conclusion is Y.
Suppose that the sole ground for believing the law L (e.g., All Fs
are G) is an enumeration of actual instances (Fai\&Gai) of L.
Then adding the supposition X, that is, that a further a is F,
removes the ground for accepting L. We can no longer draw the
conclusion that this further a is G. Hence, we cannot assert the
counterfactual X \( \rightarrow \) Y. More generally, if the reasons for
accepting L survive placing L within the scope of the supposi-
tion that there are further instances of the law's subject term, then we can
say that the law supports the relevant counterfactual condi-
tional. According to Mackie the required reasons are ordinary
inductive reasons, that is, good inductive evidence for the law.
Good inductive evidence, in other words, is evidence for the
noted, the evidence plays a double role. It first establishes
inductively a generalization. But then, "it continues to operate separately in making it reasonable to assert the
counterfactual conditionals which look like an extension of
the law into merely possible worlds" (Mackie, 1974: 203).

So the interesting related thought comes from Julius Weinberg
(1951) who claimed the following: A counterfactual X \( \rightarrow \) Y
is not best seen as the indicative statement the statement of a
generalization) X \( \rightarrow \) Y plus some further antecedent
conditions (including that X did not actually happen), but
rather as asserting something about the evidence there is for
X \( \rightarrow \) Y, that is, that there is evidence for, and no evidence
generally against, the generalization (for all X \( \rightarrow \) Y). Hence, the
additional strength of counterfactual is supposed to have over
the corresponding generalization is captured by the evidence
there is for the generalization.

The Possible Worlds View

Taking literally the view that counterfactuals are used in
contemplating possibilities, the second major view of the
semantics of counterfactuals appeals to possible worlds. In
first suggesting this view, Todd (1964: 107) noted that when we
allow for the possibility that the antecedent of a counterfactual
be true, we are "hypothetically substituting a different world
for the actual one." On this view, the core meaning of a
counterfactual X \( \rightarrow \) Y is (roughly): In the possible (but
not actual) world where X, Y too.

A possible world is a way the world might be or might have
been. For instance, it is possible that gold is not yellow, or that
planets describe circular orbits, or that birds do not fly, or that
deer does not need yeast to brew. But are there really possible
worlds? There are three views here. The first is that talk of
possible worlds is a mere façon de parler, although useful when
it comes to assessing counterfactuals (cf Mackie, 1974: 199).
I take it that an extension of this view is that possible worlds
are useful fictions. The second is 'extreme realism', according
to which the way the world actually is, is one among the
many ways the world could be; hence, the actual world is
one among the many possible worlds, the latter being no less
real than the actual. The chief advocate of this view was
David Lewis (1973). The third view is 'abstract realism',
according to which possible worlds are maximally consistent
sets of propositions: total ways things might be. A 'possible
world' then is fit to represent a complete reality, but only one
of them actually represents anything, that is, the actual world
(cf Bennett, 2003).

Stalnaker (1968) developed the core meaning of
counterfactuals as follows. Consider a possible world W in which X is true but other-
wise is similar to the actual world X \( \rightarrow \) Y true if Y is
true in W. The similarity relation among worlds (a selection function,

\[ \Psi \] ) is an ordering of possible worlds with respect to their resemblance to the actual world.

Calling an X-world a possible world in which X hold,
counterfactuals might be taken to be strict conditionals of the
following form:

\[ \Psi X \rightarrow Y \]

is true in a world W if Y is true in all X-worlds such that
where the blank is filled by a general condition
that X-words should satisfy. Hence, whatever goes into the
blank places a restriction on the admissible (or accessible) possible worlds. This idea would model
counterfactuals along the lines of strict conditionals of the
form:

\[ \Phi X \rightarrow Y \]

it is physically necessary that

\[ \Phi \]

or

\[ \Phi \]

where the first restriction is to all worlds with the same laws as
the actual, whereas the second 'restriction' would be to all possible worlds simpliciter.

But this analysis cannot be correct. There is no set of
two possible worlds W such that X \( \rightarrow \) Y throughout W (this is
another way to state the fact that counterfactuals are non-
monotonic). So Lewis (1973) suggested that counterfactuals
X \( \rightarrow \) Y are suitably strict conditionals: each of them is a strict
conditional, that is, every X-world of a certain sort is a
Y-world, but the relevant set of worlds varies with different
conditionals.

Like Stalnaker, Lewis took it that worlds are ordered in terms
of similarity, or closeness to the actual world. According to this
prominent notion of 'comparative overall similarity': 'we may say
that one world is closer to actuality than another if the first
resembles our actual world more than the second does, taking
account of all the respects of similarity and difference and
balancing them off against one another' (1986: 163).

But unlike Stalnaker, Lewis took it that in assessing the
counterfactual X \( \rightarrow \) Y, it does not make good sense to talk
about the closest-to-actual possible X-world. It is not just that
there might be more than one closest-to-the actual possible
worlds. It is mainly that there might not be even one rightly
Counterfactual Reasoning, Qualitative: Philosophical Aspects

deemed the closest (even in a limiting sense). Hence, according to Lewis’s view:
\[ X \rightarrow Y \quad \text{true at a world } W \text{ if some (accessible) } X\text{-world in which } Y \text{ holds is closer to } W \text{ than any } X\text{-worlds that } Y \text{ does not hold.} \]

For instance, take the counterfactual that if this pen had been left unsupported, it did not fall to the floor. Take all X-worlds. The counterfactual \( X \rightarrow Y \) is true (in \( @ \)) if the X-worlds in which Y is true (i.e., the pen is left unsupported and falls to the floor) are closer to \( @ \) than any of the X-worlds in which Y is false (i.e., the pen is left unsupported but does not fall to the ground, e.g., it stays still in midair). As Lewis (1986, 164) put it, “[A] counterfactual \( \neg X \rightarrow Y \) is true if it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent.”

The key idea behind the possible-world semantics is that in specifying the truth-conditions of a counterfactual conditional, we should imagine a set of affairs in which X obtains and which is such that all else is pretty much as they actually were. But as noted already, this is not quite possible. In the possible world in which X did happen, many other things (including the laws) were different from the actual world \( @ \) in which X did not occur. Can we find comfort in the notion of comparative similarity? Now, although “comparative overall similarity” is not defined strictly, a lot can be said of it. Notably, it imposes a weak ordering on the set of possible worlds that are accessible from \( @ \), that is, the relation of comparative similarity is connected and transitive. (It also imposes a centering assumption: \( @ \) is closer to itself than any other world is to it.) More important, however, similarity is clearly not one dimensional, but rather it is the result of many component similarities. Lewis (1986, 47–48) ranked possible worlds according to the following dimensions of similarity (put in order of importance):

- Avoid big, widespread violations of the laws of nature of the actual world (very important).
- Maximize the spatiotemporal perfect match of particular matters of fact (not at all important).
- Avoid small, localized violations of the laws of nature of the actual world.
- Secede approximate similarity of particular matters of fact (not at all important).

So, a world \( W_1 \) that has the same laws of nature as the actual world \( @ \) is closer to \( @ \) than a world \( W_2 \) that has different laws. But insofar as there is exact similarity of particular facts in large spatiotemporal regions between \( @ \) and a world \( W_0 \), Lewis allows that \( W_0 \) is close to \( @ \) even if some of the laws that hold in \( @ \) are violated in \( W_0 \).

All this implies that there is quite a lot of vagueness in the notion of overall comparative similarity, which accounts for the fact that counterfactuals themselves are vague, at least in the sense that it is a contextual matter as to what to keep fixed and what to change when we assert a counterfactual conditional. A more serious worry relates to the issue of the motivation behind the foregoing ranking of dimensions of similarity among worlds. It has been observed by many that Lewis’s initial theory yielded the wrong truth-values for a type of counterfactual conditional that can be schematized as follows:

\[ X \rightarrow Y \quad \text{true at a world } W \text{ if some (accessible) } X\text{-world in which } Y \text{ holds is closer to } W \text{ than any } X\text{-worlds that } Y \text{ does not hold.} \]

For instance:

\[ (C) \text{ If the president had pressed the button, a nuclear war would have ensued.} \]

Intuitively, \( (C) \) is true. But on Lewis’s initial account, it would be false. For a possible world \( W_1 \) in which the president did press the button and a nuclear war did erupt, \( Y \) is more distant from (because more dissimilar to) actuality \( @ \) in a different way, a really big miracle would have to occur. For all the many and tiny traces of the button pushing would have to be wiped out. Hence, appearances to the contrary, \( W_1 \) would be more distant from (because more dissimilar to) actuality \( @ \) than \( W_2 \). The big violation of laws of nature in \( W_2 \) is out-weighed by the maximization of the perfect spatiotemporal match of particular matters of fact between \( W_1 \) and \( @ \). So, with the help of the refined criteria of similarity among possible worlds, the president counterfactual becomes true. Still, one may follow Horwich (1987, 171–172) in wondering how psychologically plausible Lewis’s theory becomes: The similarity criteria are so tailored that the right counterfactuals become true, but they have little to do with our pretheoretical understanding of judgments of similarity.

As noted in relation to the ‘support’ view, an adequate theory of counterfactuals has to jump two hurdles. The first relates to counterenability. Lewis solved this problem by taking it that some conditions \( S \) are cotenable with \( X \) (the antecedent of the counterfactual \( X \rightarrow Y \)) if some \( X \)-world is closer to the actual world than any \( S \)-world. The second hurdle relates to the distinction between laws and accidents. Here, the possible world approach is on safe ground, although the ground can support any decent theory of counterfactuals. David Lewis (1973) revamped a long tradition that goes back to John Stuart Mill, via Frank Ramsey, according to which the regularities that constitute the laws of nature are those that are expressed by the axioms and theorems of an ideal deductive system of our knowledge of the world, and in particular, of a deductive system that strikes the best balance between simplicity and strength. Simplicity is required because it disallows extraneous elements from the system of laws. Strength is required because the deductive system should be as informative as possible about the laws that hold in the world. Whatever regularity is not part of this best
A key objection to the possible world approach to counterfactuals is that counterfactual conditionals are not purely objective; an immediately subjective element enters into the judgment of similarity (and arguably, to the distinction between laws and accidents). Not only are the truth-conditions of counterfactuals “a highly volatile matter” as Lewis (1973: 92) himself noted, but also model choices are true or false. The experimental units are chosen and separated into two groups (the experimental group and the control group) by randomization. To simplify matters, let the treatment set be defined and determined. In particular, it is assumed that even if subject u actually is given treatment t and Y(c, u) is the value that would be observed if the unit u were exposed to treatment t and Y(c, u) is the value that would be observed on the same unit if it were exposed to c. A key assumption of Rubin’s model is that both values Y(c, u) and Y(c, u) are well defined and determined. In particular, it is assumed that even if subject u actually is given treatment t and has response Y[c, u], there is still a fact of the matter about what the subject’s u response would have been, had she been given treatment c. The task is to figure out the individual causal effect, that is the difference
\[ \tau(u) = Y(t, u) - Y(c, u), \]
which measures the effect of treatment t on u, relative to treatment c.

In each particular experiment, either Y[t, u] or Y[c, u] (but not both) ceases to be counterfactual. Yet, given that one of Y[t, u] and Y[c, u] becomes testable, the other has to be untestable. Holland has called a situation such as this “the fundamental problem of causal inference”. Does it follow that figuring out \( \tau(u) \) is impossible?

Suppose that we give treatment t to u and we observe Y[t, u]. The question then is how could we possibly figure out the counterfactual value of Y[c, u]? According to the present model, when certain assumptions are in place, there are ways to assess counterfactuals such as the above. Here is how:

Given that unit u got treatment t, we may try treatment c to a different unit u’, which is very much like u, except that it was given treatment c instead. That is, instead of testing the counterfactual conditional Y[c, u’], we test the indicative conditional Y[c, u’] - the response of subject u’ if it is given treatment c - and claim that this tells indirectly what the value of Y[c, u] is. For this move to be plausible at all, we need an assumption of unit homogeneity: that u and u’ are so similar...
that the actual response of $u'$ to treatment $c$ is the same as the response that unit $u$ would have to treatment $c$. Under this assumption, we take it that $Y(u, a) = Y(u, c, a)$ and $Y(c, u) = Y(c, u')$.

Then, the individual causal effect can be calculated, since eqn [1] becomes thus:

$$\tau(u) = Y(u, a) - Y(u, c, a) = Y(u, a) - Y(c, u').$$ \[2\]

Although eqn [1] involves essentially a counterfactual conditional $(Y(u, a))$, eqn [2] does not. Eqn [2] is indeed testable, but the counterfactuals are gone. Instead, eqn [2] has two indicative conditionals, one for unit $u$ who received treatment $t$ and another for unit $u'$ who received treatment $c$, in sense, the unit homogeneity assumption renders the counterfactual conditional $Y(u, a)$ not so much a claim about the specific unit $u$ but rather a claim about any of the homogeneous units. It is because of this fact that the counterfactual is supposed to become testable.

We might proceed in another way to calculate $\tau(u)$. Instead of giving treatment $t$ to unit $u$ and treatment $c$ to (uniform) unit $u'$, we give treatment $c$ to unit $u$ at time $t_1$ and treatment $t$ to the very same unit $u$ at a later time $t_2$. This move requires another assumption, that is, temporal stability or the constancy of response over time. It also requires an assumption of ‘causal transience’ to avoid situations like this. The subject’s taking a placebo at time $t_1$ changes some properties of her enough to affect her response to taking an aspirin at a later time $t_2$. Under these assumptions, we take it that $Y(t_1, u) = Y(t_2, u)$ and $Y(t_2, u) = Y(t_2, u)$. If this is so, then the individual causal effect can be calculated, since eqn [1] becomes:

$$\tau(u) = Y(t_1, u) - Y(t_2, u) - Y(t_2, u) = Y(t_1, u) - Y(t_2, u).$$ \[3\]

The remarks made about eqn [2] can be applied about eqn [3], too. Eqn [1] has no counterfactuals and it seems that the content of eqn [1] - which does involve the counterfactual $Y(c, u)$ - values to the joined content of two indicative conditionals $Y(t_1, u) = Y(t_2, u)$ and $Y(t_2, u) = Y(t_2, u)$ together with two further assumptions of causal transience and temporal stability.

The key point then is that the alleged testability of counterfactual conditionals is predicated on the plausibility and success of certain general assumptions noted previously. These assumptions might fail if, however, there are reasons to believe they do not, that is, if there is evidence for the general assumptions, then causal inference seems quite safe. I would suggest, that these assumptions are characteristics of stable causal homomorphic structures. Consider unit homogeneity. For it to hold, it must be the case that two units $u$ and $u'$ are alike in all causally relevant respects other than treatment status. If this is so, we can substitute $u$ for $u'$ and vice versa. This simply means that there is a causal law connecting the treatment and its characteristic effect, which holds for all homogeneous units and hence it is independent of the actual unit chosen (or could have been chosen) to test it. In effect, this holds for temporal stability too, since the latter is the temporal version of unit homogeneity.

**Interventionist Counterfactuals**

James Woodward recently has introduced the claim that only counterfactuals that are related to interventions can be of help when it comes to assessing their test or assertibility conditions. An intervention gives rise to an ‘active counterfactual’, that is, to a counterfactual whose antecedent is made true by (hypothetical) interventions.

Woodward (2003: 3) characterized the appropriate counterfactuals in terms of experiments: "They are understood as claims about what would happen if a certain sort of experiment were to be performed".

Take Ohm’s law (that the voltage of a current is equal to the product of its intensity and the resistance of the wire) and consider the following two subjunctives:

1. If the resistance were set to $R = \text{at time } t$, then the intensity $I$ would be $I = e\times R$ at $t$.
2. If the resistance were set to $R = \text{at time } t$, then the voltage $E$ would be $E = e\times R$ at $t$.

According to Woodward, we can perform the experiments at a future time $t^*$ to see whether (1) or (2) are true. If, however, we are interested in finding out what would have happened, had we performed the experiment in a past time $t$ (although we never did), Woodward invited us to rely on the ‘very good evidence’ we have that the behavior of the circuit is stable over time. Given this evidence, we can assume that the actual performance of the experiment at a future time $t^*$ is as good for the assessment of (1) and (2) as a hypothetical performance of the experiment at the past time $t$. An obvious advantage of this approach is that the truth-conditions of (the right sort of) counterfactual conditionals are not specified by means of an abstract metaphysical theory, as in the possible worlds approach. But there is a residual tension in this view. Woodward (rightly) insisted that counterfactual conditionals have determinate meaning and truth-conditions independent of the actual and hypothetical interventions. So there is a distinction between truth-conditions and test-conditions for a counterfactual. But then there must be a way for counterfactuals to get their truth-conditions fixed independent of their test-conditions. It is not quite clear what this way might be. The unclarity is accentuated by the fact that if there is such a way to specify the truth-conditions of a counterfactual conditional independent of its test-conditions (which are related to hypothetical interventions), then this way will offer truth-conditions to counterfactuals that do not (or might not) have test-conditions at all.

What if we were to collapse the truth-conditions of counterfactuals to their test-conditions? One can see the prima facie attraction of this move. Because evidence-conditions are specified in terms of actual and hypothetical experiments, the right sort of counterfactuals (the active counterfactuals) and only those end up being meaningful and truth-valuable. But there is an important drawback. Recall the subjunctive assertion in (1). On the option presently considered, what makes (1) true is that its evidence-conditions obtain. Under this option, counterfactual conditionals lose, so to speak, their counterfactuality. (1) becomes a shorthand for a future prediction or the evidence that supports the relevant law. If it is a future time, (1) gives way to an indicative conditional (a prediction). If it is a past time, then, given that there is good evidence for Ohm’s law, all that (1) asserts under the present option is that there has been good evidence for the law.
In any case, Woodward is keen to keep evidence- and truth-objects (e.g., fragility or elasticity and the like) has been into a full-blown theory. The sufficient condition of a counterfactual: Actual and hypothetical experiments) are connected with the truth-conditions of a counterfactual. Actually, ascription of dispositions to objects is analyzed in terms of (1) a specific counterfactual conditional and (2) an unspecified categorical claim about the nature of the object. Although the owner was caused in some other way.

Counterfactual Reasoning, Qualitative: Philosophical Aspects

Acknowledgment

This research has been cofinanced by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program ‘Education and Lifelong Learning’ of the National Strategic Reference Framework (NSRF) - Research Funding Program: THALIS – UOA - AIPRePoSMA.

See also: 63006; 63007; 63016.

References


This position, intuitively compelling though it may be, faces a number of important difficulties that has led to various ad hoc additions and modifications. (For a discussion of them see my 2002 paper.) The key point is that this approach explains how the effect depends on the cause without entailing anything as to how the cause is connected to the cause.

Those who think that there is a special kind of connection between cause and effect take counterfactual dependence to be a symptom of the presence of a power in the cause to bring about the effect. Causation amounts to a power’s producing its manifestation. (This idea goes back to Leibniz who took that causes are ‘producers’.) Actually, ascription of dispositions to objects (e.g., fragility or elasticity and the like) has been analyzed in terms of counterfactual (and subjunctive) conditionals. So, to ascribe a disposition to an object x is to say that if x were to be given stimulus S, the characteristic result would be R. Example: x is fragile if x were to be struck, it would break.

Rom Harré and Edward Madden (1975) offered a general analysis of powers along the foregoing lines: x has the power to be F x were subject to stimuli or conditions of an appropriate kind, then x would do F, “in virtue of its intrinsic nature”. So power-ascriptions to objects is analyzed in terms of (1) a specific counterfactual conditional and (2) an unspecified categorical claim about the nature of the object.

Despite its initial promise, this view faces important counterexamples, most of which point to the claim that the meaning of dispositional ascriptions cannot be captured by counterfactual conditionals. For it is possible either that the antecedent of the counterfactual is realized and the characteristic response does not obtain (e.g., because something else blocks the manifestation of the disposition) or that a disposition exists even if there are no manifestations of it (and hence no relevant counterfactuals to be entailed by it). Hence, the power-based attempts to understand the causal connection should either disqualify causation from counterfactuals or take “being disposed to” as a special (not further reducible) relation (see Mumford and Anjum, 2011).

But

p0385

intrinsic nature

to F if x were subject to stimuli or conditions of an appropriate kind, then x would do F, “in virtue of its intrinsic nature”. So power-ascriptions to objects is analyzed in terms of (1) a specific counterfactual conditional and (2) an unspecified categorical claim about the nature of the object.

Despite the various difficulties we have presented, counterfactual conditionals are a stable part of the philosophical arsenal. Their being modal in character has invited the thought that they can capture a special connection between the antecedent and the consequent most typically a causal connection. According to a popular counterfactual analysis of causation, to say (roughly) that event e causes event e is to say that e is counterfactually dependent on c, that is, that if c had not happened, e would not have happened either. This idea goes back to David Hume, but has been developed by Lewis (1974) into a full-blown theory. The sufficiency part of the definition is straightforward: If two events e and e are actual, and e is counterfactually dependent on c, then e is the cause of e. Example: Let c be the actual short circuit and e be the actual fire. If it is the case that if c had not occurred, then e would not have occurred, then the short circuit is the cause of the fire. But causation is transitive, whereas (as we have seen) counterfactual dependence is not. Example: e is the cause of e, e is the cause of e.

Although the owner’s insurance company (e) is counterfactually dependent on the short circuit (c), e is not counterfactually dependent on e. The owner would have got the insurance compensation (e) even if the short circuit (c) had not occurred, assuming that the fire was caused in some other way. To make counterfactual dependence a necessary condition for causation, Lewis introduced a way to enforce the transitivity of counterfactual dependence. The sequence of events must form a causal chain. A sequence of events e', e', e' is a chain of causal (counterfactually) dependence if e causally (counterfactually) depends on e, e causally (counterfactually) depends on e, and so on.

This position, intuitively compelling though it may be, faces a number of important difficulties that has led to various ad hoc additions and modifications. (For a discussion of them see my 2002 paper.) The key point is that this approach explains how the effect depends on the cause without entailing anything as to how the cause is connected to the cause.

Those who think that there is a special kind of connection between cause and effect take counterfactual dependence to be a symptom of the presence of a power in the cause to bring about the effect. Causation amounts to a power’s producing its manifestation. (This idea goes back to Leibniz who took that causes are ‘producers’.) Actually, ascription of dispositions to objects (e.g., fragility or elasticity and the like) has been analyzed in terms of counterfactual (and subjunctive) conditionals. So, to ascribe a disposition to an object x is to say that if x were to be given stimulus S, the characteristic result would be R. Example: x is fragile if x were to be struck, it would break.

Rom Harré and Edward Madden (1975) offered a general analysis of powers along the foregoing lines: x has the power to be F x were subject to stimuli or conditions of an appropriate kind, then x would do F, “in virtue of its intrinsic nature”. So power-ascriptions to objects is analyzed in terms of (1) a specific counterfactual conditional and (2) an unspecified categorical claim about the nature of the object.

Despite its initial promise, this view faces important counterexamples, most of which point to the claim that the meaning of dispositional ascriptions cannot be captured by counterfactual conditionals. For it is possible either that the antecedent of the counterfactual is realized and the characteristic response does not obtain (e.g., because something else blocks the manifestation of the disposition) or that a disposition exists even if there are no manifestations of it (and hence no relevant counterfactuals to be entailed by it). Hence, the power-based attempts to understand the causal connection should either disqualify causation from counterfactuals or take “being disposed to” as a special (not further reducible) relation (see Mumford and Anjum, 2011).

This research has been cofinanced by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program ‘Education and Lifelong Learning’ of the National Strategic Reference Framework (NSRF) – Research Funding Program: THALIS – UOA – AIPRePoSMA.

See also: 63006; 63007; 63016.

References


