Faculty Publications

1-1-2008

What if Every "If Only" Statement Were True?: The Logic of Counterfactuals

Kevin W. Saunders
Michigan State University College of Law, saunde44@law.msu.edu

Follow this and additional works at: http://digitalcommons.law.msu.edu/facpubs

Part of the Jurisprudence Commons, and the Other Law Commons

Recommended Citation

WHAT IF EVERY "IF ONLY" STATEMENT WERE TRUE?: THE LOGIC OF COUNTERFACTUALS

Kevin W. Saunders[*]

This Symposium is all about counterfactual conditionals. To ask "what if" is to posit a situation that is not actually the case and ask what else would be true under the hypothesized facts. These counterfactuals are difficult to analyze from the point of view of logic. Simple propositional logic will not do. The problem is that, from the point of view of propositional logic, every counterfactual conditional, every "what if" statement, is true.

Propositional logic is truth-functional. The truth or falsity of complex statements is a function of the truth and falsity of the simple propositional components of the complex statement and the logical operators connecting them. Introductory logic students learn to work with truth tables, such as the following table for the conjunction $p \& q$. The first two columns present all the possible combinations for the truth and falsity of the propositions p and q. The third column, the column for the conjunction, shows its truth value for each combination. The table asserts that the conjunction is true when, and only when, both conjoined propositions are true.

p	q	p&q
T	T	T
T	F	F
F	T	F
F	F	F

This would seem to be consistent with our general understanding of such a statement. Someone who asserts the truth of a conjunction is asserting the truth of each conjoined proposition.

The truth table for the disjunction $p \lor q$ is shown below. It is based on the treatment of "or" in a nonexclusive way. The proposition $p \lor q$ is true, as long as at least one of the disjuncts is true.

---

    *   Professor of Law, Michigan State University. A.B., Franklin & Marshall College; M.S., M.A., Ph.D., University of Miami; J.D., University of Michigan.

p	q	p∨q
T	T	T
T	F	T
F	T	T
F	F	F

The exclusive "or," that is, one or the other, but not both, would require another connective, or it may be expressed as *(p∨q)&~(p&q)*, where *~r* represents the negation of *r*. The truth table for the exclusive *or* built up a step at a time is as follows, and yields the result that it is true when exactly one of the two propositions is true:

p	q	p∨q	p&q	~(p&q)	(p∨q)&~(p&q)
T	T	T	T	F	F
T	F	T	F	T	T
F	T	T	F	T	T
F	F	F	F	T	F

The truth function problem with the counterfactual conditional, if it were to be treated as the material implication of propositional logic, arises from the truth table definition for that form of the conditional. In propositional logic, the truth table for *if p, then q* is as follows:

p	q	p->q
T	T	T
T	F	F
F	T	T
F	F	T

The propositional logic treatment of the conditional asserts that whenever *p* is true, *q* must be true, and no more than that. That explains the second line in the truth table. Since *p* was true and *q* was false, the proposition *p->q*, or

whenever p is true, q is true, is false. The top line is also easily determined, where both are true. What may be more difficult to see is the reasonableness of the last two lines. In those lines, *p* is false, so the proposition *p->q*, which again states that *when p is true, q is true*, cannot be false. It makes no claim as to the truth or falsity of *q*, when *p* is itself false.

Coming back to the issue of counterfactual conditionals, we can now see that any statement that begins with "If I were the King of France" is true in propositional logic, since I am not the King of France. It makes no difference what comes after the antecedent clause. Thus, every *if only* proposition, as in *If only I were the King of France* or *If only the Patent and Trademark Office did not issue business method patents*, is true, if we limit its analysis to propositional logic.

What is required here is a more complex logic, and that logic is a variety of modal logic. Modal logic considers propositions and their truth or falsity in this and in alternative worlds. There are propositions that are true in this world, but if we were in some alternative world, if the facts were different, the proposition would be false.

Similarly, there are propositions that are false in this world but true in some alternative world. In those situations, while *p* is true or false in this world, the proposition *possibly p* or *Mp* is true in this world. That is, *Mp* is true because in some alternative world, *p* is true. Note that, if *p* is true in the real world, then *Mp* is also true in this world, as long as we consider this world to be an alternative. But from the truth of *Mp*, we know nothing of the truth of *p* in this world.
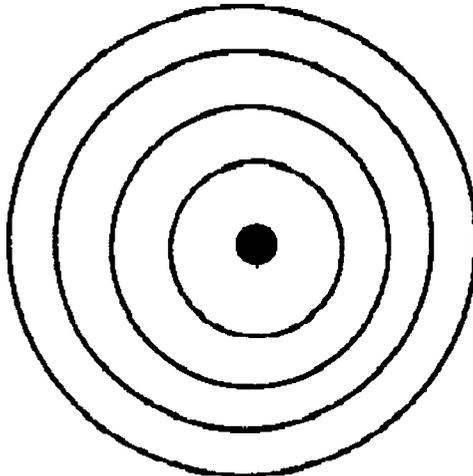
There are other propositions that must be true in every possible world. This is most easily seen with propositions such as *Either Elizabeth II is the Queen of England or Elizabeth II is not the Queen of England.* This proposition is true in this world, but it would also be true in an alternative world in which there is no English monarchy or someone else is the monarch.[1] More abstractly, any proposition of the form $p \lor \sim p$ is always true, whatever the factual truth or falsity of *p*. Thus, $p \lor \sim p$ is true in every alternative world. The proposition is said to be necessarily true. When a proposition *q* is true in every alternative world, then *q is necessary* or *Lq* is true in this world. Note that, if *Lq* is true in this world, then treating this world as an alternative to itself makes *q* also true in this (and every other) world. But, from the simple truth of *q* in this world, we cannot infer the truth of *Lq*.

While there is a form of implication based on necessity—strict implication—that, too, is an inadequate treatment. Strict implication is of the

---

1.   This does leave open the question of how to handle an alternative world in which there is no England. The proposition might simply be considered false, since there is no England for Elizabeth II to be queen of. This sort of problem is also somewhat avoidable under the nesting of worlds, based on similarity, discussed *infra* note 3 and in accompanying text.

form *L(p->q)*, or *necessarily p implies q*. It states that *p* implies *q* in every possible world. Clearly, one who asserts that if *p* were true, *q* would be true, is not making that strong a claim. The overly strong claim is that no matter what else changed, along with *p* now being true, *q* would be true. The only valid claims of that strength would be ones in which *p->q* is tautologically true and would probably not be all that interesting.

The idea of the truth of the conditional in alternative worlds is, however, a start down the road to understanding the logic of hypotheticals. The approach is to order the alternative worlds according to their similarity to the actual world. In the diagram that follows, the actual world is indicated by the dot in the center.[2] The concentric rings around that dot contain alternative worlds that differ from the actual world and do so more and more, as one progresses to the outer rings. The worlds in the innermost rings are quite similar to the actual world,[3] while the more bizarre worlds are quite removed from the center.


The person who asserts that, if *p* were true, *q* would be true, is presumably making the claim that a change in the truth of *p*, with other facts remaining the same, would mean that *q* would be true. A speaker who says "If I had struck this match it would have lighted" is making the claim that in an alternative world, where the match had in fact been struck, it lighted. One may respond by suggesting that it was raining in the alternative world or that the match in that world was covered with grease and would not be affected by friction. The original speaker's response would be to explain that she meant that if other facts had not changed, that there was no rain or grease, it would have lighted. On the other hand, the speaker is unlikely to be asserting that only the fact of striking the match has changed, and that everything else has to remain exactly the same. Striking the match will have created at least some additional micro effects, such as breeze and heat.

So, the speaker does not mean to assert that the change in the truth of *p* will result in *q* being true in every possible world. And the speaker does not mean to assert simply that, in a world in which only the truth value of *p* has changed, *q* will be true. The nested rings in the diagram below can help understand what is meant. A curve is run through the rings indicating the worlds in which a proposition is true. The letter standing for the proposition is placed on the side of the curve containing the worlds in which the proposition is true. On the other side of the curve, the proposition is false.

The following diagram shows curves for propositions *p* and *q*. In both cases, the right side, or inside, of the curve contains the worlds in which the proposition is true. Note that, since the center dot is not on the true side of the curves, both *p* and *q* are false in the actual world. Something else is also true of the relationship between *p* and *q* in the diagram. Every world that is on the true side of the *p* curve is also on the true side of the *q* curve. This means that in every world in which *p* is true, *q* is true, so *L(p->q)* holds.
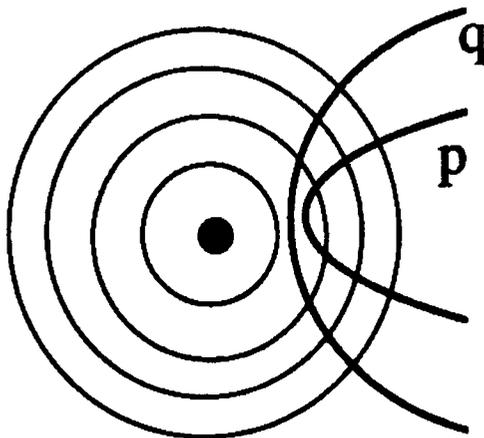
For the hypothetical conditional, we are not concerned about q being true in every world in which p is true. We are only interested in the state of affairs in those worlds most similar to the actual world. We want to know about the most similar worlds in which p is true. To examine those worlds, we look at the part of the truth curve in the innermost ring it intersects. In the above diagram, q is also true, but that was to be expected since the implication there was a strict implication that is true in every possible world.

The diagram below shows the more interesting case. There, there are worlds in which p is true and q is false, so the implication is not necessary. But, if we confine our examination to the innermost ring in which p may be true, we see that in every world in that area in which p is true (the shaded area), q is also true. This is what we mean by the counterfactual conditional.
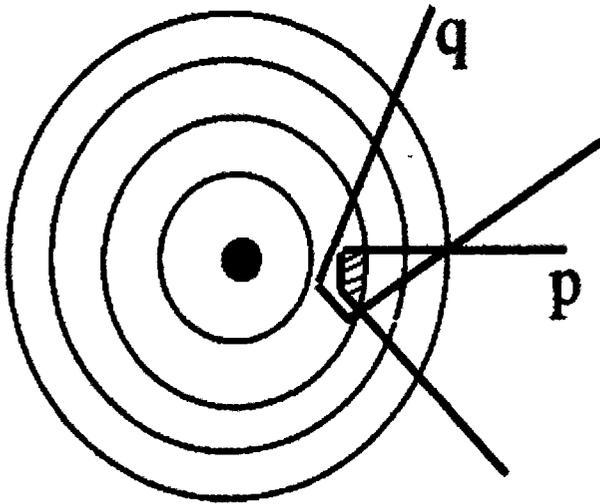

This use of diagrams may, or may not, have seemed interesting, but it does explain a bit about the logic of counterfactuals, and the diagrams can further be used to show some oddities of that logic. One such oddity has to do with the property of transitivity. In ordinary logic, if p implies q, and q implies r, then p implies r, or symbolically $((p\text{->}q)\&(q\text{->}r))\text{->}(p\text{->}r)$. David Lewis offers an example. He argues that, if we take as true *If J. Edgar Hoover had been born a Russian, then he would have been a Communist* and *If he had been born a Russian, he would have been a traitor*,[4] we cannot infer the truth of *If he had been born a Russian, he would have been a trai-*

---

4.   *Id.* at 33. Lewis credits the example to Robert C. Stalnaker, *A Theory of Conditionals, in* STUDIES IN LOGICAL THEORY 98 (Nicholas Rescher ed., 1968).

tor. As Lewis explains it: "A Communist Hoover is nowhere to be found at worlds near ours, but a Russian-born Hoover is still more remote."[5] That explanation may not be particularly enlightening, but the diagram he suggests,[6] adapted to the notation used here, does explain why transitivity does not hold. Interestingly, the diagram would seem to suggest a position that Hoover was a traitor in the real world.

In the diagram, p is the proposition *Hoover was born a Russian*, and q is the proposition *Hoover was a Communist*. From the diagram, we see that in every world in the innermost ring in which p is true, where p is true (area A), q is also true. Letting r stand for *Hoover was a traitor*, we see that in every world in the innermost ring in which q is true, where q is true (area B), r is also true. Thus, both counterfactuals are true. But, in all the worlds in the innermost ring in which p is true, there are worlds where p is true but r is false. The counterfactual *if p were true, r would be true* is then false, and transitivity fails.
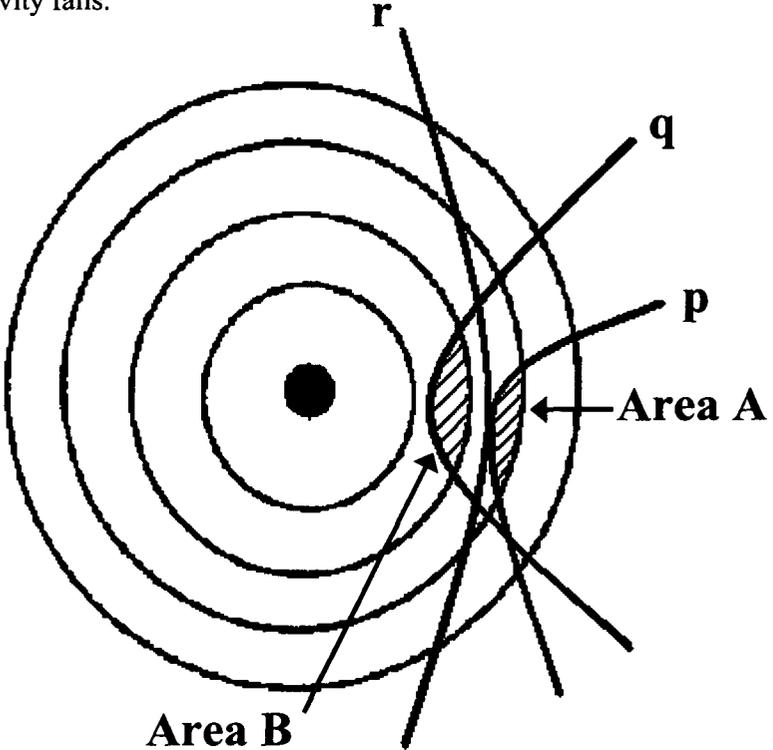

One other example of the oddity of the logic of counterfactuals is the fact that contraposition is not a valid inference. In propositional logic, from

5.   LEWIS, *supra* note 2, at 33.
6.   *See id.* at 34 fig.4(A) (notation adapted).

p->q you may infer ~*q*->~*p*. That is, if whenever *p* is true, *q* is true, then if *q* is false, *p* must have been false. This makes sense, but it is not true with counterfactuals. Lewis again gives an example. Accepting *If Boris had gone to the party, Olga would still have gone* as true, *If Olga had not gone, Boris would still not have gone* may be false.[7] As Lewis explains it, suppose Olga is interested in Boris, but Boris wants to avoid Olga.[8] So, if Boris were to go, so would Olga, but if Olga does not go, Boris may well go.

Again, a diagram is helpful.[9] Let *p* stand for *Boris goes to the party* and *q* stand for *Olga goes*. The premise is shown by the fact that in all the worlds in the innermost ring in which *p* is true, where *p* is true (area A), *q* is true. But, in the innermost ring in which *q* is false (or ~*q* is true), where *q* is false (area B), *p* may be either true or false.

Area A

---
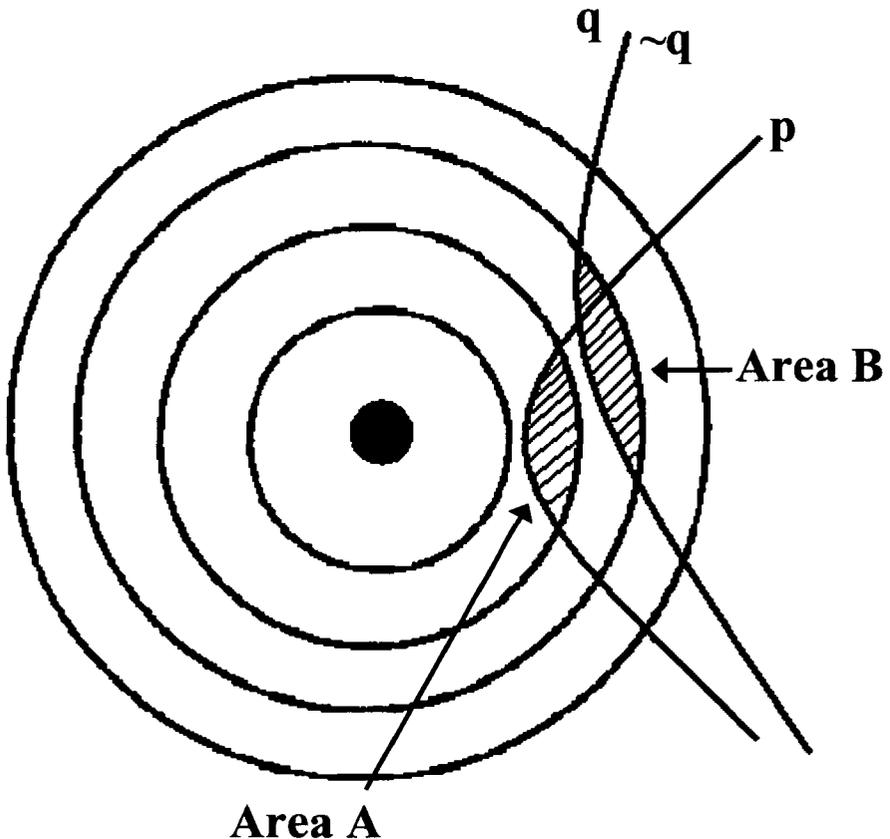
7. *Id.* at 35.
8. *See id.*
9. *See id.* at 34 fig.4(B) (notation adapted).

This discussion was intended as something far short of a complete exposition on the logic of counterfactuals.[10]   Even if not beyond the logic background of most readers, it is most likely beyond the interest level of the average reader of this Symposium issue. What was intended was a caution on the intricacies of making inferences under the assumption of a fact not currently true. Hypothetical reasoning is certainly worth doing, but it must be done carefully.

---

10.   A treatment in far greater detail may be found in LEWIS, *supra* note 2. For those who might be interested in another application to law, see Wesley Newcomb Hohfeld, *Some Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 23 YALE L.J. 16 (1913). An analysis of several of Hohfeld's jural relations would seem best treated as counterfactuals. A claim that I have a legal power over you, after all, is a claim that if I were to do something I have not yet done, I would change one of your legal relations. For an analysis of Hohfeld's relations, see Kevin W. Saunders, *A Formal Analysis of Hohfeldian Relations*, 23 AKRON L. REV. 465 (1990).