Philos Stud (2009) 144:281–296
DOI 10.1007/s11098-008-9210-8

The myth of the categorical counterfactual

David Barnett

Published online: 12 February 2008

Abstract I aim to show that standard theories of counterfactuals are mistaken, not in detail, but in principle, and I aim to say what form a tenable theory must take. Standard theories entail a *categorical* interpretation of counterfactuals, on which to state *that, if it were that A, it would be that C* is to state something, not relative to any supposition or hypothesis, but categorically. On the rival *suppositional* interpretation, to state *that, if it were that A, it would be that C* is to state *that it would be that C* relative to the supposition *that it were that A*. The two interpretations make incompatible predictions concerning the correct evaluation of counterfactuals. I argue that the suppositional interpretation makes the correct prediction.

Keywords Counterfactuals · Conditionals · Subjunctive conditionals

A counterfactual is a statement of the form, 'If it were that A, it would be that C'. I aim to show that standard theories of counterfactuals are mistaken, not in detail, but in principle, and I aim to say what form a tenable theory must take.

Remarkably, on standard theories, counterfactuals are interpreted as *categorical* statements. On this interpretation, to state *that, if it were that A, it would be that C* is to state something, not relative to any supposition or hypothesis, but categorically. Differences in detail among the standard theories are differences over *which thing* is stated categorically. Nelson Goodman (1947) says that it is an entailment from the antecedent, together with laws of nature and particular facts about the actual world, to the consequent;[1] Robert Stalnaker (1968) says that it is a predication of a single

[1] More recent views inspired by Goodman's account include Pollock 1981, Barker 1999, and Hiddleston 2005.

D. Barnett (✉)
Department of Philosophy, University of Colorado-Boulder, 232 UCB, Boulder, CO 80309-0232, USA
e-mail: david.b.barnett@colorado.edu

possible world; and David Lewis (1973) says that it is an existential generalization over a set of possible worlds.

By contrast, W.V.O. Quine (1950), John Mackie (1973), Michael Dummett (1978), and Dorothy Edgington (1995) maintain that counterfactuals are as their surface form suggests: *suppositional* statements.[2] We all have an intuitive grasp of what it is to suppose something. If I ask you to suppose that Bin Laden had been a pacifist, you know just what to do. On the suppositional view of counterfactuals, to state *that, if it were that A, it would be that C* is to state something *relative to a supposition*. What is stated is *that it would be that C*; what is supposed is *that it were that A*. Counterfactual statements are acts of supposing-cum-stating.

I have heard it suggested that the categorical and suppositional interpretations are compatible. On this proposal, a counterfactual statement can correctly be interpreted as a categorical statement of something, and it can also correctly be interpreted as a statement of something relative to a supposition. To my knowledge, nobody has defended this proposal in print.[3] In any case, it is false: the suppositional and categorical interpretations make different, incompatible, predictions concerning the correct evaluation of counterfactuals.[4] In this paper I draw out one such difference.

The difference concerns the question of whether a certain principle governs the rational evaluation of counterfactual questions–questions such as what things would be like, if Bin Laden were a pacifist. I introduce the principle and show why it initially appears to hold for all questions (Sect. 1). Then I show how the principle fails in contexts of counterfactual suppositions (Sect. 2). It follows that the suppositional view predicts that the principle fails for counterfactual questions (Sect. 3). By contrast, the categorical view predicts that the principle must hold for counterfactual questions (Sect. 4). The principle fails for counterfactual questions (Sect. 5). Thus, the idea of a categorical counterfactual—one that states what it does outside the scope of any supposition—is a myth.

1 Field's Principle

Some questions seem to lack answers. For instance, if Harry is borderline bald, then one is initially inclined to say that there is no answer to the question of whether Harry is bald. For present purposes, the controversy over the status of this inclination will not matter: we shall simply *suppose* that the inclination is sound. On this supposition, how should we distribute our confidence among the candidate answers to the question? How confident should we be that Harry is bald, and how confident should we be that Harry is not bald? Intuitively, we should have

[2] Other views consistent with the suggestion include Ryle 1950, Adams 1965, 1966, 1975, Dudman 1994, Woods 1997, and Bennett 2003.

[3] To be sure, certain proponents of the categorical view endorse a suppositional method for *evaluating* counterfactuals. But this does not mean that they endorse the suppositional *interpretation* of counterfactuals. See e.g., Lewis 1973, Stalnaker 1981.

[4] Likewise, the suppositional and categorical views of *indicative* conditionals make a host of incompatible predictions; for a sample, see Edgington 1995, Barnett 2006.

absolutely no confidence in either proposition. And there is a simple argument to back up this intuition:

(1a) However confident we are *that Harry is bald*, we should be equally confident *that the answer to the question of whether Harry is bald is that Harry is bald*; and however confident we are *that Harry is not bald*, we should be equally confident *that the answer to the question of whether Harry is bald is that Harry is not bald*.

(2a) On the supposition that there is no answer to the question of whether Harry is bald, we should have zero confidence that the answer to this question is that Harry is bald, and zero confidence that the answer to this question is that Harry is not bald.

(3a) Hence, on the given supposition we should have zero confidence that Harry is bald and zero confidence that Harry is not bald.

To accommodate this conclusion, Hartry Field (2000) develops a non-standard calculus of rational degrees of confidence. The idea behind the calculus is that for a rational thinker to treat a question as potentially without an answer is for her degrees of confidence in the (individually complete, mutually exclusive, and jointly exhaustive) candidate answers to add to less than one—what remains is her degree of confidence that the question has no answer.[5] Here, then, is the guiding principle behind Field's calculus:

Field's Principle One's degrees of confidence in the candidate answers to a given question, together with one's degree of confidence that the question has no answer, rationally add to 1.

Considered in the abstract, Field's Principle seems right. For it seems that however confident one is in a candidate answer to a question, one should be equally confident, of that candidate, that it *is* the answer to the question. Thus, to calculate one's degree of confidence that the question has no answer, it seems that one should simply subtract one's degrees of confidence in the candidate answers from 1.

2 Field's Principle may fail in contexts of counterfactual suppositions

My aim here is not to attack Field's Principle, but to begin to draw out a difference in predictions between the suppositional and categorical views of counterfactuals. Presumably Field's Principle holds in a suitably circumscribed set of contexts. In conversation, Field has accepted that his principle does not hold in contexts of counterfactual suppositions, for the following reason.

In the context of a counterfactual supposition, an assignment of a value to a subjunctive question or statement qualifies as correct, not on the condition that what is asked or stated *have* the value, but rather on the condition that what is asked or

[5] Whereas I speak of *questions without answers*, Field speaks of *indeterminate sentences*. Field says that "for an agent to *treat A as potentially indeterminate* is for him to have degrees of belief in it and its negation that add to less than 1" (2000, p. 8).

stated *stand in a certain relation* to what is supposed. For illustration, let us suppose (Urn):

(Urn) that yesterday you had drawn randomly from some urn or other containing ninety-nine red balls and one black ball.

Sometimes when someone asks you to suppose that things had been different in some way, the person wants you also to suppose that everything else had been as similar as possible to how things actually were. This is not one of those times. I simply want you to suppose (Urn); nothing more, nothing less. To be sure, given (Urn), it might be highly likely that everything else would have been as similar as possible to how things actually were. But in any case this is not what I am asking you to suppose.

In the context of supposing (Urn), one might evaluate the subjunctive statement *that you would have drawn red* as 99% probable. This assignment of 99% probability qualifies as correct, not on the condition that what is stated have an absolute probability of 99%, but rather on the condition that what is stated stand in a certain relation to (Urn), namely, that it be made 99% probable by (Urn). Likewise for degrees of confidence: in the context of supposing (Urn), to have 99% confidence in the statement that you would have drawn red is not to have 99% *absolute* confidence in what is stated, but rather to have 99% *conditional* confidence in what is stated, relative to (Urn). A similar point applies to ascriptions of the value *answerless* to subjunctive questions. In the context of our supposition, one might assign the value *answerless* to (Q1):

(Q1) What color would you have drawn?

This assignment qualifies as correct, not on the condition that what is asked by (Q1) be absolutely without an answer, but rather on the condition that what is asked by (Q1) stand in a certain relation to (Urn), namely, that it have no answer determined by (Urn).

From here it is easy to see how Field's Principle may fail in the context of a counterfactual supposition. Imagine that we are in such a context: our supposition is *p*, and we are evaluating the subjunctive question of whether *q*. Imagine that we believe *p* and *q* to be related as follows:

(4a) *p* neither entails *q* nor entails *not-q*.
(4b) *p* makes *q* 99% probable
(4c) *p* makes not-*q* 1% probable.

There is nothing problematic about (4a)–(4c). Yet, to commit to this trio, in the context of supposing *p*, we can take up a combination of stances toward *q* that jointly violate Field's Principle:

(5a) maintain that there is no answer to the question whether *q*
(5b) maintain 99% confidence in *q*
(5c) maintain 1% confidence in *not-q*.

In the context of supposing *p*, the stance comprising (5a)–(5c) commits one to the unproblematic (4a)–(4c), and it has no further problematic commitments. Thus, it need not be irrational. Yet, by Field's Principle, it is irrational. And so we see how it is that Field's Principle may fail in contexts of counterfactual suppositions.

In the preceding section I gave an argument to motivate Field's Principle. I argued that, on the supposition that there is no answer to the question of whether borderline Harry is bald, we should have zero confidence that Harry is bald and zero confidence that Harry is not bald. To see where this argument goes wrong when applied in the context of a counterfactual supposition, let us apply it to (Q1), in the context of supposing (Urn):

(1b) However confident we are *that you would have drawn red*, we should be equally confident *that the answer to (Q1) is that you would have drawn red*; and however confident we are *that you would have drawn black*, we should be equally confident *that the answer to (Q1) is that you would have drawn black*.

(2b) On the *indicative* supposition that there *is* no answer to (Q1), we should have zero confidence that the answer to the question *is* that you would have drawn red, and zero confidence that the answer to the question *is* that you would have drawn black.

(3b) Hence, on the given indicative supposition, we should have zero confidence that you would have drawn red and zero confidence that you would have drawn black.

To see where the argument goes wrong, we need to focus our attention on a subtle distinction between two questions:

(Subjunctive) What *would have been* the answer to the indicative question of what color you drew, supposing (Urn)?

(Indicative) What *is* the answer to the subjunctive question (Q1)?

(Subjunctive) is a subjunctive question about an indicative question that would have existed, supposing (Urn). (Indicative) is an indicative question about a subjunctive question, (Q1). Now, in the context of supposing (Urn), we are to a certain degree confident *that you would have drawn red*. Given this confidence, how do the laws of rationality constrain our attitudes toward (Subjunctive) and (Indicative)?

First consider (Subjunctive). Certainly: *you would have drawn red* iff *the answer to the indicative question of what color you drew would have been that you drew red*. So, however confident we are (i) *that you would have drawn red*, we should be equally confident (ii) *that the answer to the indicative question of what color you drew would have been that you drew red*. In the context of supposing (Urn), *n%* confidence in (i) involves the same commitment as *n%* confidence in (ii), namely, that (Urn) makes it *n%* probable that you would have drawn red. So we see how our confidence that you would have drawn red rationally constrains our attitude toward (Subjunctive).

Now consider how it rationally constrains our attitude toward (Indicative). It is false that, however confident we are (i) *that you would have drawn red*, we should be equally confident (iii) *that the answer to the subjunctive question (Q1) is that you would have drawn red*. Why? Because confidence in (i) involves a very different commitment from confidence in (iii). Confidence in (i), in the context of supposing (Urn), commits one to the claim that (Urn) makes it *probable* that you would have drawn red. By contrast, confidence in (iii) commits one to the claim that (Urn) *entails* an answer to (Q1), namely, that you would have drawn red. And (Urn) might make it probable that you would have drawn red without entailing that you would have drawn red. So our confidence that you would have drawn red does not rationally translate into confidence that the answer to (Q1) is that you would have drawn red. Hence, it does not rationally constrain our attitude toward (Indicative) in the same way that it constrains our attitude toward (Subjunctive).

With the distinction between (Subjunctive) and (Indicative) carefully in mind, we are in a position to see where the preceding argument goes wrong: premise (1b) is false. In the context of supposing (Urn), confidence that you *would have* drawn red need not rationally translate into confidence that the answer to (Q1) *is* that you would have drawn red.

Generally speaking, in the context of a counterfactual supposition, our confidence that things *would have been* a certain way need not rationally translate into confidence that things *are in fact* a certain way with respect to some subjunctive question, but only into confidence that things *would have been* a certain way with respect to an indicative question that *would have* existed. Field's Principle owes its appeal to the claim that however confident one is in a candidate answer to a question, one should be equally confident, of that candidate, that it *is* the answer to the question. But in the context of a counterfactual supposition, our confidence in a candidate answer to a subjunctive question rationally translates, not into confidence, of that candidate, that it *is* the answer to the question, but rather into confidence that there *would have been* a corresponding indicative question with a corresponding indicative answer.

So we see why Field's Principle may fail in contexts of counterfactual suppositions.

3 The suppositional view predicts that counterfactuals may violate Field's Principle

On the suppositional view, to state *that, if it were that A, it would be that C* is to state, relative to the supposition that it were that *A*, that it would be that *C*. To ask *whether, if it were that A, it would be that C* is to ask, relative to the supposition that it were that *A*, whether it would be that *C*. What is stated by a counterfactual statement is *that it would be that C*; what is asked by a counterfactual question is *whether it would be that C*; and what is supposed by both is *that it were that A*.

The suppositional view predicts that counterfactuals may violate Field's Principle. For on it, to evaluate a counterfactual question is to evaluate what is asked by the question, relative to what is supposed by the question. For instance, to say of a counterfactual question that it has no answer is to say, relative to what is supposed by the question, that what is asked by the question has no answer. To evaluate a counterfactual statement is to evaluate what is stated by the statement, relative to what is supposed by the statement. For instance, to ascribe *n%* probability to a counterfactual statement is to ascribe *n*% probability to what the counterfactual states, relative to what it supposes. Thus, on the suppositional view, to evaluate a counterfactual question or statement *is* to evaluate a subjunctive question or statement relative to a counterfactual supposition. And we know from Sect. 2 that such evaluation may violate Field's Principle.

4 The categorical view predicts that counterfactuals must obey Field's Principle

On the categorical view, to state *that, if it were that A, it would be that C* is to state something, not relative to any supposition, but categorically. To ask *whether, if it were that A, it would be that C* is to ask something, not relative to any supposition, but categorically.

This view predicts that counterfactuals must obey Field's Principle. For, restricted to categorical questions, Field's Principle seems obviously right. Distributing one's *absolute* confidence exhaustively among the candidate answers to a given categorical question while denying that the question has an answer would commit one to an absurdity: a question such that (i) one of its candidate answers must be right even though (ii) it has no answer. Whatever amount of confidence one has that a given categorical question lacks an answer, one has that much less *absolute* confidence to distribute among the candidate answers to the question. Thus, if counterfactual questions are categorical, Field's Principle should apply to them.

For illustration, consider Stalnaker's proposal, which I shall take as representative of the possible-worlds theory of counterfactuals. On it, what is categorically stated by a counterfactual statement is that the nearest possible world at which the antecedent is true is a world at which the consequent true.[6] Now, let us suppose that there is no answer to the counterfactual question of *whether, if it were that A, it would be that C*. By Stalnaker's analysis, this is equivalent to supposing that there is no answer to the question of *whether C is the case in the nearest possible A-world*. But on the supposition that there is no answer to *this* question, it seems clear that we should have absolutely no confidence *that C is the case in the nearest possible A-world* and absolutely no confidence *that C is not the case in the nearest possible A-world*.

Indeed, we can defend this conclusion with an instance of our argument from Sect. 1:

(1c) However confident we are *that C is the case in the nearest A-world*, we should be equally confident *that the answer to the question of whether C is the case in the nearest A-world is that C is the case in the nearest A-world*.

[6] Stalnaker (1968, pp. 33–34) says, "Consider a possible world in which A is true, and which otherwise differs minimally from the actual world. *'If A, then B' is true (false) just in case B is true (false) in that possible world.*"

(2c) On the supposition that there is no answer to the question of whether C is the case in the nearest A-world, we should have zero confidence that the answer to this question is that C is the case in the nearest A-world and zero confidence that the answer to this question is that C is not the case in the nearest A-world.
(3c) Hence, on the given supposition we should have zero confidence that C is the case in the nearest A-word, and zero confidence that C is not the case in the nearest A-world.

(1c) is an instance of the principle that however confident one is in a candidate answer to a question, one should be equally confident, of that candidate, that it *is* the answer to the question. We know that, in the context of a counterfactual supposition, our confidence in a candidate answer to a subjunctive question need not rationally translate into confidence, of that candidate, that it *is* the answer to the question. But (1c) does not occur in the context of a counterfactual supposition; nor is it about a subjunctive question. (1c) states that however much *absolute* confidence one has in the candidate answer to a certain *indicative* question, one should be equally confident, of that candidate, that it *is* the answer to the question. And this seems right. So the argument appears to be sound.

Converting (3c), by Stalnaker's analysis, back into the language of counterfactuals, we get the following: on the supposition that there is no answer to the question of *whether, if it were that A, it would be that C*, we should have absolutely no confidence *that*, *if it were that A, it would be that C*, and we should have absolutely no confidence *that, if it were that A, it would not be that C*.

Like all categorical views, Stalnaker's view predicts that Field's Principle should apply to counterfactuals.

5 Counterfactuals may violate Field's Principle

Consider (Q2):

(Q2) If yesterday you had drawn randomly from some urn or other containing ninety-nine red balls and one black ball, what color would you have drawn?

Here are three natural responses, each of which is deemed irrational by Field's Principle:

Response #1: "I'm 99% confident that, if yesterday you had drawn randomly from some urn or other containing ninety-nine red balls and one black ball, you would have drawn red; I'm 1% confident that, if you had so drawn, you would have drawn black; of course, it would be silly to bet on the question, for there is obviously no correct answer." By Field's Principle, Response #1 is irrational. But intuitively it is not.

Response #2: "What color you would have drawn depends on how you would have reached into the urn and how the balls would have been arranged. Given certain combinations of reachings and arrangements, you would have drawn red; given certain others, you would have drawn black. Because there are no facts to determine specifically which combination would have obtained, there is no answer to (Q2). Still, it is far more likely, if you had drawn from such an urn, that you would have drawn *red*. Indeed, it is ninety-nine times more likely. This is why I am 99% confident that, if you had so drawn, you would have drawn red. Likewise, I am 1% confident that, if you had so drawn, you would have drawn black." By Field's Principle, Response #2 is irrational. But intuitively it is not.

Response #3: "Personally, I don't know whether (Q2) has an answer. But I am willing to *suppose* that it has no answer. In this context, how confident should I be in each of the candidate answers to (Q2)? Certainly, if you had drawn, one of two possibilities would have obtained: you would have drawn red or you would have drawn black. This is not to say that there is an answer to the question *which* of these two possibilities would have obtained, but only that *one of them would have*. To be sure, if you had so drawn, there *would have been* an answer to the corresponding *indicative* question of which possibility obtained, just as there would have been an answer to the indicative question of what color you drew. But this is consistent with the idea that there *is* no answer to the subjunctive question of which of the two possibilities would have obtained, if you had so drawn. Now, there are facts about how likely each of these possibilities is to have obtained. Because no combination of a reaching and an arrangement is more likely than any other to have obtained, and because the proportion of combinations that would have resulted in your drawing red is ninety-nine times greater than the proportion that would have resulted in your drawing black, it is ninety-nine times more likely that you would have drawn red, if you had so drawn, than that you would have drawn black, if you had so drawn. Thus, on the supposition that there is no answer to (Q2), I am 99% confident that you would have drawn red, if you had so drawn, and I am 1% confident that you would have drawn black, if you had so drawn." By Field's Principle, Response #3 is irrational. But intuitively it is not.

I conclude that (Q2) is a counterexample to Fields' Principle.[7]

By now it should be clear where the motivating argument for Field's Principle goes wrong when applied to counterfactuals. Here is the argument, applied to (Q2):

(1d) However confident we are *that, if you had so drawn, you would have drawn red*, we should be equally confident *that the answer to the question of what color you would have drawn, if you had so drawn, is that, if you had so drawn, you would have drawn red*; and however confident we are *that, if you*

[7] Not all counterfactuals violate Field's Principle. Suppose again that, because Harry is a borderline case of baldness, there is no answer to the question of whether Harry is bald. Now consider whether, if Harry had had the very same hair situation, but had been a fraction of an inch taller, he would have been bald. One is inclined to say that there is no answer to the question. Suppose that this is right. Then how confident should we be that Harry would have been bald, and how confident should we be that he would not have been bald? Here I think one should have no confidence that he would have been bald and no confidence that he would not have been bald. For, because a minor difference in *height* is irrelevant to whether a person qualifies as *bald*, however confident one is that Harry would have been bald, one should be equally confident that Harry *is* bald; and however confident one is that Harry would not have been bald, one should be equally confident that Harry is *not* bald. And we have already concluded that, on the present supposition, one should have zero confidence that Harry is bald and zero confidence that he is not bald.

had so drawn, you would have drawn black, we should be equally confident *that the answer to the question of what color you would have drawn, if you had so drawn, is that, if you had so drawn, you would have drawn black.*

(2d) On the indicative supposition that there is no answer to the question of what color you would have drawn, if you had so drawn, we should have zero confidence that the answer to this question is that, if you had so drawn, you would have drawn red, and zero confidence that the answer to the question is that, if you had so drawn, you would have drawn black.

(3d) Hence, on the given indicative supposition we should have zero confidence that, if you had so drawn, you would have drawn red and zero confidence that, if you had so drawn, you would have drawn black.

The argument goes wrong in (1d). (1d) is an instance of the principle that however confident one is in a candidate answer to a question, one should be equally confident, of that candidate, that it *is* the answer to the question. But in the case of a counterfactual question, our confidence in a candidate answer rationally translates, not into confidence, of that candidate, that it *is* the answer to the counterfactual question, but rather into confidence that there *would have been* a corresponding indicative question with a corresponding indicative answer.

(1d) might look attractive because it is similar to the following sound principle:

(1e) However confident we are *that, if you had so drawn, you would have drawn red*, we should be equally confident *that, if you had so drawn, there would have been a question as to what color you drew whose answer would have been that you drew red*; and however confident we are *that, if you had so drawn, you would have drawn black*, we should be equally confident *that, if you had so drawn, there would have been a question as to what color you drew whose answer would have been that you drew black.*

Of course, if we replace the false premise, (1d), with the truth, (1e), the argument becomes invalid.

I conclude that counterfactuals may violate Field's Principle.

6 Conclusion

The suppositional view predicts that counterfactuals may violate Field's Principle (Sect. 3). The categorical view predicts that counterfactuals must obey Field's Principle (Sect. 4). Counterfactuals may violate Field's Principle (Sect. 5). Hence, we should reject the categorical view in favor of the suppositional view.[8]

[8] In my 2006 I defend a particular version of the suppositional view as it applies to indicative conditionals; in an unpublished manuscript ("Zif Would Have Been If: A Suppositional View of Counterfactuals"), I defend that version as it applies to counterfactual conditionals. In the present paper, my aim is simply to show that *some* version of the suppositional view must be correct for counterfactuals (and therefore that all standard views of counterfactuals are false).

7 Questions

Here are some questions that one might have about my argument.

Question 1. Call the fact that (Q2) violates Field's Principle "The Violation." Might Stalnaker accommodate The Violation by claiming (i) that, because many worlds are tied for the nearest antecedent-world, there is no answer to the question of what color you would have drawn, if you had so drawn, and (ii) that, because you draw red in 99% of these worlds and black in 1%, one ought to be 99% confident that you would have drawn red, if you had so drawn, and 1% confident that you would have drawn black, if you had so drawn?

Answer. No. There are at least two problems with this suggestion.

First, on the supposition that there is no answer to the question which color you draw in the nearest antecedent-world, one cannot rationally be 99% confident that you draw red, and 1% confident that you draw black, in the nearest antecedent-word. For such confidence rationally requires 99% confidence that the answer to the question is that you draw red in the nearest antecedent-world, 1% confidence that the answer to the question is that you draw black in the nearest antecedent-world, and thus 100% confidence that there is an answer to the question. So, given Stalnaker's analysis, 99% confidence that you would have drawn red, if you had so drawn, and 1% confidence that you would have drawn black, if you had so drawn, is rationally inconsistent with the supposition that there is no answer to the question what color you would have drawn, if you had so drawn.

Second, even setting aside our supposition that there is no answer to the question and granting that you draw red in 99%, and black in 1%, of many worlds tied for the nearest antecedent-world, it would not, on Stalnaker's analysis, be rational to be 99% confident that you would have drawn red and 1% confident that you would have drawn black. For it would not be rational to be 99% confident that you draw red, and 1% confident that you draw black, in the nearest antecedent-world. Given that *n*% of the Fs tied for the most G are H, one should not be n% confident that the most G is H. Suppose for illustration that 100 players are tied for the tallest player in the league, and that exactly 99 of them are Canadians. One should not be 99% confident that the tallest player in the league is Canadian; rather one should be 0% confident, for the supposition entails that there is no tallest player in the league. So, given Stalnaker's analysis, being 99% confident that you would have drawn red, and 1% confident that you would have drawn black, if you had so drawn, is rationally inconsistent with the supposition that you draw red in 99%, and black in 1%, of many worlds tied for the nearest antecedent-world.

Question 2. Might one explain away the appeal of The Violation by claiming that we are confusing false claims about how likely it *is* that things would have been a certain way, if our antecedent had obtained, with true claims about how likely it *would have been* that things would turn out a certain way, had our antecedent obtained? More specifically, might one explain away the appeal of The Violation by claiming that, on the supposition that there is no answer to the question what color you would have drawn, if you had so drawn, we mistakenly accept (a) (which we are rationally required to reject) because we confuse it with (b) (which we are rationally required to accept):

(a) that it *is* 99% likely that you would have drawn red, and that it *is* 1% likely that you would have drawn black, if you had so drawn
(b) that it *would have been* 99% likely that you would draw red, and that it *would have been* 1% likely that you would draw black, if you had so drawn?

Answer. No. There are at least two problems with this suggestion.

First, although (b) is appealing, accepting it does not reduce the appeal of (a). Had you drawn from such an urn, one of two possibilities would have been realized: you would have drawn red or you would have drawn black. At least prior to your having drawn, one of these two possibilities *would have been* ninety-nine times more likely *to be* realized than the other. Still, it seems obvious that one of the two possibilities *is* ninety-nine times more likely *to have been* realized than the other. And so while (b) is plausible, it is not a plausible *surrogate* for (a).

Second, by making a small adjustment to our urn example, we can avoid the issue of distinguishing between (a) and (b) altogether. Notice that The Violation is formulated in terms of degrees of confidence, not degrees of likelihood. To be sure, the proponent of the idea that we are confusing (a) and (b) could easily reformulate (a) and (b) in terms of degrees of confidence. But, unlike degrees of likelihood, degrees of confidence are states of thinkers. And so without thinkers there would be no degrees of confidence. Now consider (Q3):

(Q3) If there had never been any thinkers, but there had been a device that once in history selected randomly from a pool of ninety-nine red berries and one black berry, what color berry would the device have selected?

Suppose that there is no answer to this question. Then how should we distribute our confidence among the candidate answers? I am 100% confident that one of two possibilities would have been realized: a red berry would have been drawn or a black berry would have been drawn. This is not to say that I am confident there is an answer to the question *which* of the two possibilities would have been realized, but only that one of them would have. Because the first possibility is ninety-nine times more likely to have been realized than the second, I am 99% confident that a red berry would have been selected, and 1% confident that a black berry would have been selected, if a device had so drawn. Because my stance is not irrational, it constitutes a violation of Field's Principle. And yet it is obvious in this case that I am not confusing my degree of confidence in the candidate answers to (Q3) with degrees of confidence that I *would have had*, if the device had so drawn; for I am certain that, if the device had so drawn, I would not have existed.

Question 3. A referee for this journal suggested that counterfactual statements are ambiguous: on one reading, they are suppositional; on another, they are categorical. On this suggestion, if-clauses are ambiguous in *function*: they can either express their contents as suppositions or contribute them to larger contents that are categorically stated by the statements in which the if-clauses are embedded. Are counterfactuals ambiguous in this way?

Answer. No. Consider (6):

(6) If yesterday you had drawn randomly from an urn containing ninety-nine red balls and one black ball, you would have drawn red.

On its face, (6) does not seem ambiguous. So, to establish that (6) is ambiguous, an argument is needed.

The referee suggests an argument. The first premise is that the question of how probable (6) is can be interpreted either as the question of how probable the consequent of (6) is, given the antecedent, or as the question of how probable it is that the whole statement, (6), is true. In other words, the question of how probable (6) is can be interpreted either as (7) or as (8):

(7) How probable is it that you would have drawn red, given that yesterday you had drawn randomly from an urn containing ninety-nine red balls and one black ball?

(8) How probable is it that the following statement is true: that, if yesterday you had drawn randomly from an urn containing ninety-nine red balls and one black ball, you would have drawn red?

The second premise is that the answer to (7) is *99%*. The third premise is that the answer to (8) is *0%*. The conclusion is that (6) is ambiguous between a suppositional and a categorical reading: on the suppositional reading, its probability is 99%; on the categorical reading, its probability is 0%.

The conclusion is not justified by the premises. For the premises are easily accommodated by the rival view that (6) is univocally suppositional. Here, again, are the premises:

(P1) the question 'How probable is (6)?' is ambiguous between (7) and (8)
(P2) the answer to (7) is *99%*
(P3) the answer to (8) is *0%*

(P1) is simply an instance of a general ambiguity that arises whenever we assign a name, '*n*', to an unambiguous statement, *S*, and then ask how probable *n* is. Are we asking (a) how probable it is *that S* or (b) how probable it is *that the statement that S is true*? For categorical statements, (a) and (b) are equivalent. But not for suppositional statements. For suppositional statements, (a) is a question about the probability of what is stated by the statement, given what is supposed by it; (b) is a question about the probability that it is correct *to ascribe truth* to what is stated by the statement, given what is stated by it. We know that, in the context of a counterfactual supposition, *p*, it need not be irrational to both (i) ascribe a high probability to a content *q* and (ii) maintain that there is no answer to the question of whether *q*, and thus that *q* is neither true nor false. The reason that (i) and (ii) are not jointly irrational, recall, is that (i) commits one to one relation between *p* and *q*, namely, that *p* makes *q* highly probable; (ii) commits one to a compatible relation, namely, that *p* does not entail *q*; and (i) and (ii) have no further jointly problematic commitments. We see, then, why (a) and (b) are not equivalent for suppositional statements. On the suppositional view of counterfactuals, (7) and (8) are simply instances of (a) and (b). Moreover, on the suppositional view, our answer to (7) should be *99%* and our answer to (8) should be *0%*. For, given that you had drawn from such an urn, it is 99% probable that you would have drawn red; it is 0% probable that there is a fact of the matter whether you would drawn red; and so it is

0% probable that *it is true* that you would have drawn red. (To be sure, it is 99% probable that the statement that you drew red *would have been* true; but that is a very different claim.) So, the view that (6) is univocally suppositional easily accommodates premises (P1)–(P3). Hence, we should not endorse the rival view—that (6) is ambiguous between a categorical and a suppositional reading—on the basis of (P1)–(P3).

I know of no other plausible argument for a categorical/suppositional ambiguity in (6). Because (6) does not appear to be ambiguous, I conclude that it is not. Generalizing, I conclude that there is no categorical/suppositional ambiguity in counterfactual statements.

Question 4. Consider (Q4):

(Q4) Will ticket #79 win the 100-ticket raffle tomorrow?

There is a view on which future contingent questions, including (Q4), lack answers. Add to this view the claim that one should be 1% confident that #79 will win and 99% confident that #79 will not win, and we get a view on which (Q4) is counterexample to Field's Principle. Add the further claim that (Q4) is a categorical question, and we get a view on which Field's Principle need not apply to categorical questions. Can one appeal to this view about future contingents to show that The Violation is compatible with a categorical interpretation of counterfactuals?

Answer. No. One reason to doubt the considered view of future contingents is that it has a counterintuitive consequence, namely, that (Q4) is a counterexample to Field's Principle. Intuitively, however confident one is that #79 will win, one should be equally confident that the answer to (Q4) is that #79 will win; and however confident one is that #79 will not win, one should be equally confident that the answer to (Q4) is that #79 will not win. Hence, given that one should be 1% confident that ticket #79 will win, and 99% confident that it will not win, one should be 100% confident that there *is* an answer to (Q4). Intuitively, then, (Q4) is not a counterexample to Field's Principle. I am not claiming that this is a decisive objection to the view that future contingent questions lack answers, but only that one cannot show that The Violation is compatible with the categorical interpretation of counterfactual questions by citing future contingents questions, for intuitively such questions obey Field's Principle.

Question 5. On the suppositional view, to say that there is no answer to a counterfactual question is to commit to the view that what is supposed by the antecedent does not entail an answer to what is asked by the consequent. But in practice we assert candidate answers to counterfactual questions even when the antecedents of the questions do not entail the candidate answers. In other words, we assert counterfactuals whose antecedents do not entail their consequents. Is this evidence against the suppositional view?

Answer. No. For illustration, consider (Q5):

(Q5) If Ray had not dropped the vase, would it still have broken?

It is easy to imagine a context in which it would be conversationally appropriate to assert (A5):

(A5) If Ray had not dropped the vase, it would not have broken.

Yet on the suppositional view an objectively correct response is to say that there is no answer to (Q5); for the supposition that Ray had not dropped the vase entails neither that the vase would have broken nor that it would not have broken. How, then, could it be conversationally appropriate to assert (A5)?

Sometimes it is appropriate to assert something because one has a high degree of confidence in that thing. On the suppositional view, to assert (A5) is to assert *that the vase would not have broken*, relative to the supposition *that Ray had not dropped it.* And a degree of confidence in (A5) is a degree of confidence *that the vase would not have broken*, relative to the supposition *that Ray had not dropped it*. It is easy to imagine a context in which it would be rational to have a high degree of confidence that the vase would not have broken, conditional on the supposition that Ray had not dropped it. In such a context, it may be conversationally appropriate (thought neither objectively correct nor objectively incorrect) to assert (A5). Hence, the suppositional view explains how it could be conversationally appropriate to assert (A5).

More generally, the suppositional view explains how it could be conversationally appropriate to assert a counterfactual whose antecedent does not entail its consequent.[9]

Question 6. How, on the suppositional view, do we characterize valid reasoning with conditionals?

Answer. Not in terms of the truth of what is stated by the premises guaranteeing the truth of what is stated by the conclusion. For when we reason with conditionals, we evaluate what they state, not categorically, but relative to the suppositions expressed by their antecedents. To accommodate this fact, Adams (1965, 1975) introduces the following notion of validity: an argument is *probabilistically valid* iff there is no probability function in which the uncertainty of the conclusion exceeds the sum of the uncertainties of the premises. The notion of probabilistic validity provides the basis for a formal logic of conditionals and other suppositional statements.[10] A noteworthy feature of the logic is that, for arguments not involving suppositional statements, an argument is probabilistically valid iff it is truth-preserving; and so classical truth-functional logic is preserved. Although Adams's logic is formulated in terms of indicatives, many of the results, as well as the spirit of the logic, can be applied to subjunctives as well.[11]

---

[9] On a related note, the suppositional view explains what is plausible about its rival categorical theories. These theories purport to give the truth conditions of what is categorically stated by counterfactuals. They do a bad job of this, since nothing is categorically stated by counterfactuals. Still, they may do a relatively good job of tracking the conditions under which what is stated by a counterfactual has a high degree of probability, conditional on what is supposed.

[10] For discussion of Adams's strategy, as well as related topics, such as how the suppositional view handles compounding, see Edgington 1995, Bennett 2003, Barnett 2006.

[11] Earlier versions of this paper were presented at the University of Texas at Austin and the University of Colorado at Boulder. For helpful comments and discussion, I am grateful to members of those audiences, and to Yuval Avnur, George Bealer, David Christensen, Hartry Field, Michael Huemer, William Lycan, Adam Morton, Adam Pautz, Derk Pereboom, and a referee for *Philosophical Studies*.

References

Adams, E. W. (1965). A logic of conditionals. *Inquiry*, *8*, 166–197.

Adams, E. W. (1966). Probability and the logic of conditionals. In J. Hintikka & Suppes P (Eds.), *Aspects of inductive logic* (pp. 256–316). Amsterdam: North Holland.

Adams, E. W. (1975). *The logic of conditionals*. Dordrecht: Reidel.

Barnett, D. (2006). Zif is If. *Mind*, *115*, 519–566.

Barker, S. (1999). Counterfactuals, probabilistic counterfactuals, and causation. *Mind*, 108, 427–469.

Bennett, J. (2003). *A philosophical guide to conditionals*. New York: Oxford University Press.

Dudman, V. H. (1994). On conditionals. *Journal of Philosophy*, *91*, 113–128.

Dummett, M. (1978). *Truth and other enigmas*. London: Duckworth.

Edgington, D. (1995). On conditionals. *Mind*, *104*, 235–330.

Field, H. (2000). Indeterminacy, degree of belief, and excluded middle. *Noûs* 34, 1–30.

Goodman, N. (1947). The problem of counterfactual conditionals. *Journal of Philosophy*, *44*, 113–128.

Hiddleston, E. (2005). A causal theory of counterfactuals. *Noûs*, *39*, 632–657.

Lewis, D. (1973). *Counterfactuals*. Cambridge, Mass: Harvard University Press.

Mackie, J. (1973). *Truth, probability and paradox*. Oxford: Clarendon Press.

Pollock, J. (1981). A refined theory of counterfactuals. *Journal of Philosophical Logic*, *10*, 239–266.

Quine, W. V. (1950). *Methods of logic*. New York: Holt, Rinehart, and Winston.

Ryle, G. (1950). 'If', 'so' and 'because'. In M. Black (Ed.), *Philosophical analysis*. Englewood Cliffs: Prentice-Hall.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Oxford: Blackwell.

Stalnaker, R. (1981). A defense of conditional excluded middle. In W. Harper, G. Pearce, & R. Stalnaker (Eds.), *Ifs: Conditionals, belief, decision, chance, and time* (pp. 87–104). Dordrecht: D. Reidel.

Woods, M. (1997). *Conditionals*. Oxford: Clarendon Press.