Counterfactual Conditionals – by Robert Brown and John Watling

We have an impressive number of uses for counterfactual conditional statements, and to some recent writers it has seemed that the analysis of these into synonymous indicative statements is a problem of considerable importance. Professor Chisholm, for instance, wrote that “the philosophical problems which this question involves are fundamental to metaphysics, epistemology, and the general philosophy of science”. However, both he and Professor Goodman, after detailed examination, conclude that the problem is still unsolved. Is it?

Consider the peculiarities of the project: an attempt to replace a counterfactual conditional of the form ‘if x were $\phi$, then x would be $\psi$’ by an indicative statement having the same meaning. Since the rather complicated formulae which can and have been produced for this purpose do not work, we may mistakenly accept Professor Chisholm’s alternatives. We may, that is, either tinker with a formula for eliminating counterfactual statements in favour of indicative ones, or believe that counterfactuals, and with them subjunctive conditionals, refer to some kind of connexion between relations, qualities, events and things. Such “irreducible connexions” are a vested interest with which we can dispense, and the formulae are irrelevant as well as unsuccessful. We suggest that the reason why both of these assertions are true is that the problem, so-called, is misconceived; we cannot eliminate counterfactuals, nor should we want to do so.

I

Briefly, the analysis of counterfactuals should parallel that of “fugitive propositions”. Somewhat as Mr. Nowell-Smith has shown that tensed sentences refer implicitly to a context which cannot, without paradox, be made explicit in the sentences themselves, so it can be shown that the use of counterfactuals presupposes contextual sentences that are neither implied by the counterfactuals nor part of their meaning. This is most clearly indicated if we recall the way in which we know whether a singular subjunctive conditional is being used counterfactual or not. It is perfectly correct to say ‘If this bear were a Kodiak, it would be brown’, when as a matter of fact we don’t know if it is a Kodiak. The additional information that it is not a Kodiak must be given to us before we can know that we are making a counterfactual statement. Thus singular counterfactuals cannot be of the form ‘$\neg \phi(a) \cdot \phi(a) \supset \psi(a)$’ because ‘$\neg \phi(a)$’ represents the information supplied by the context. For the same reason the

---

3 *Analysis*, vol. x, no. 5.
fact that the counterfactuals are not directly testable is hardly puzzling. It would be puzzling only if counterfactuals implied the falsity of their antecedents instead of presupposing that falsity. It has been suggested that "A man who says 'If this bird were a raven, it would be black' thinks and thinks rightly, that he is saying more than 'This bird is not a raven'." But the man is not saying the latter at all, he is presupposing it. There is no sinister problem arising from the circumstance that false antecedents can never be realized and tested, for while we cannot directly test counterfactuals we can supply evidence for them and give rational grounds for their assertion, namely, universal statements and initial condition statements. Some cases are quite simple. The singular subjunctive conditional used by Mr. Pears5 "If this bird were a raven, it would be black" holds, and its contrary 'Even if this bird were a raven, it would not be black' does not hold, when the statements made by the following sentences hold:
(a) 'Any raven is a bird' (b) 'Any bird which is a raven is black' (c) 'This is a bird'. When we are challenged on the subjunctive conditional just quoted, all we can do is to produce statements like those made by (a), (b) and (c), claiming that the first is analytic and the latter two are well confirmed. We can point out that if this bird happens to be a raven, so that (c) is fulfilled, then (b) entails that it will be black; but neither the counterfactual nor its contrary implies that the bird will be or will not be a raven. That extra piece of information must be given to us.

Most cases are more complicated. These include the counterfactuals used in history, law and social science, ones like 'If the Labour Party had called a general election in 1949 it would have lost'. The difficulties of supplying acceptable evidence here are obvious. There are fewer relevant, well tested universal and initial condition statements available for this example than there were for the previous example. From (a), (b) and (c) together we can derive the singular material conditional. If this bird is a raven, it will be black'. Since it does not have the same truth value as a counterfactual they cannot be equivalent but the additional knowledge (d) 'This bird is a raven' allows us to infer the testable prediction sentence (e) 'This bird will be black'. In defending a counterfactual we appeal to the statements corresponding to (a), (b), (c), (d) and (e) as evidence. We say, 'It so happens that ~(d) is the case, yet if ~ (d) were the case, then ~(d) in conjunction with (a), (b) and (c) would entail (e) '. If this evidence were not acceptable, either because our universal and initial condition statements were not well confirmed or else because of a logical error, then short of abandoning the counterfactual, we would have to produce other evidence and derivations. But what kind of evidence can we supply in the example of the Labour Party? By assembling a long list of trend sentences, inexact functions and general or universal

---

4 Analysis, vol. x, no. 6.
5 Analysis, vol. x, no. 3.
sentences, along with a vast number of initial conditions sentences, we might be able to show that relatively to these our counterfactual has a certain amount of confirmation, although the evidence would be precarious and liable to challenge on several grounds, including that of ignoring relevant conditions. This is the situation with regard to counterfactuals used in economics and politics, as well as with those applied to people’s characters and motives. In point of fact, arguments about motive explanations are often disagreements over the choice and independent testability of the motive statements serving as initial conditions in the explanation.

The amount of evidence we can produce varies considerably. In some instances the counterfactual is entailed by its evidence statements. A play in a game, for example, can so occur that it is both the last permissible play and the one upon which the gain or loss of the game depends. Assume that it is a penalty kick made by the Bruins at their opponents’ goal. After the play is completed and the Bruins have won, we can say counterfactually, ‘If the Bruins had missed their penalty kick, they would have lost the game’. This statement is jointly entailed by certain rules of the game and the statement describing the circumstances of the last play. At the other extreme there are counterfactuals for which no evidence can be produced because of the conditions under which they are made; yet we still want to say that such statements are either true or false. Consider the counterfactual ‘If I had thrown the dice once more I would have won’. Confirming evidence for statements like this is often completely lacking. The ‘Why?’ question is answered by ‘I had a hunch’ or ‘I just know I would’ where these remarks are not replaceable to the speaker’s satisfaction by any description of chances, of relevant laws or special conditions operating. It is logically precluded that any evidence pro or con can be given for the results of dice-throwing are due to chance, and hunches about single throws are attempts to predict or retrodict when no grounds for rational belief are obtainable. By education, of course, we might be able to eliminate the use of counterfactuals under these circumstances.6

Various misguided efforts have been made so to interpret counterfactuals that they will imply the statements serving as their evidence, and then to substitute for counterfactuals a synonymous or equivalent set of indicative statements as desired. Once the possibility and desirability of the programme are accepted, obstacles to its fulfilment multiply generously. The course of the predicament runs like this. First, due to the important uses we have for counterfactuals, we want to say that much of our knowledge is of what would occur if conditions were other than they actually are. We then go on to ask quite naturally, ‘What kind of unusual knowledge is this? Can knowledge of what does not occur or could not occur be expressed in indicative statements? What is the relation between antecedent and consequent in a counterfactual statement? Do counterfactuals give us untestable knowledge?’ These

---

6 We owe most of this paragraph to Professor A. J. Ayer.

3
questions create a puzzle which becomes rapidly inflated; until it is solved we do not see able to provide an adequate interpretation of natural laws, and without this nothing can be done in the philosophy of science.

Professors Chisholm and Goodman realized, of course, that in counterfactuals the consequents follow from the antecedents only when the antecedents are conjoined with certain other relevant statements which hold in fact. They wanted to be able to characterize these relevant statements in such a way that the description could be incorporated into an indicative formula that would enable us to distinguish the conditions under which counterfactuals held. It is not necessary to discuss the inadequacies of the attempt, for no attempt need be made. The distinction desired is obtained by an appeal to evidence, and our evidence for asserting a counterfactual cannot be implied by that counterfactual. A counterfactual refers to its evidence statements in the sense that unless they are true or to some extent confirmed we cannot defend our counterfactual and so do not employ it unless we are ignorant, careless or lying. To have counterfactual knowledge is simply to have knowledge obtained by interpolation and extrapolation on the basis of natural laws. The “irreducible connexion” presupposed by counterfactuals is that asserted by natural laws. In asking what kind of connexion holds between antecedent and consequent in a counterfactual we are asking about natural law connexions without intending to, and we are assuming also that the relation of antecedent and consequent is something like material implication or entailment, but yet perversely different.

II

Is it correct then to argue that all material implications are irrelevant to counterfactuals? This cannot be the case, of course, if natural laws are equivalent to or synonymous with universal material implications, for these laws are presupposed as evidence in our use of counterfactuals and are hardly irrelevant. Are natural law statements equivalent to (or synonymous with) universal material implications? Neither entail singular counterfactual conditionals; both entail singular subjunctive conditionals. The latter also follow from universal subjunctive conditionals and these, in turn, are equivalent to universal material implications.

It is clear that not all universal material implications are universal scientific laws. For instance when the Beast says to Beauty “You will never see me again, except every evening at 7 o’clock” his announcement can be written as a universal material implication but hardly counts as a universal law. Certain restrictions must be put upon the kind of statements occurring as antecedent and consequent before a universal material implication can be said to be a general law. But without considering what
these restrictions must be it is still possible to argue that all scientific laws are expressible as material implications. Whether this will turn out to be an interesting assertion will depend upon the interest of the arguments that may be given for or against it.

Against the argument that laws can be expressed as material implications it is said that such expressions might be true only by accident. That is, perhaps the real law would differentiate between several sets of conditions for the fulfilment of the antecedent, whilst the material implication is true because in fact, and by accident, only one set of these conditions ever turns up. Someone living in the tropics might formulate the law ‘Everyone likes ice cream’ and this be true by accident in the tropics because the temperature there is always high. And if high temperatures prevailed all over the world all of the time then the accidental Character of this law might never be realized. And the philosopher arguing against the material interpretation might say, ‘It is the connexion between hot weather and the liking for ice cream that is important and this connexion is not stated in a law that maybe true only by accident’. The most extreme type of this argument is found in examples such as ‘All four legged swans are black’ which are true as material implications because there are no four legged swans. Once more the argument is that it is only accident that makes these laws true; no connexion between the antecedent and the consequent is asserted.

Another way of stating the same argument is to maintain that accidentally true material implications do not allow the assertion of counterfactual conditionals. ‘If this man were living in a cold climate he would like ice cream’ or ‘…not like ice cream’, is, according to this argument not a legitimate deduction from ‘All men like ice cream’ because ‘All men like ice cream’ is only accidentally true. Yet another way of expressing the argument is to object to the confirmation of a law by instances where the antecedent is false.

A similar problem arises in the analysis of counterfactual conditionals themselves. Any analysis which makes a counterfactual equivalent to the logical product of a singular or universal conditional and the negation of the antecedent of this conditional allows it to be verified by merely verifying the falsity of the antecedent.

Mr. Kneale arguing against the equivalence of natural laws and material implications gives an interpretation of 

\[(x). \text{raven}(x) \supset \text{black}(x)\] 

The statement for which this is the formula he says, is equivalent to the statements made by the following three sentences : “(1) There has never been an observed raven that was not black; (2) There has never been an unobserved raven that was not black; (3) There will never be a raven that is not black.” Only the statement made by (1) is confirmed by the evidence; (2) and (3) are not, but only they are useful to us in our inference from the observed to the unobserved.
Let us consider the argument from accident. A law states the conditions which are causally necessary for the truth of the consequent. Variations in any of the factors mentioned in the antecedent will affect those mentioned in the consequent and no variations in any other factors will do this. Only then can the law be used for prediction.

A law that ignores one factor may be said to be true by accident if it happens that this factor never varies. If this factor does vary the law may be falsified and we might say that previously it had only been true by accident. A law may be said to be true by accident if we can identify a factor which was not mentioned in the law and whose constancy was not questioned when the law was applied successfully: But though the new law is better than the old, still we cannot be sure that circumstances will not turn up revealing the necessity for extending the antecedent even further. For there is no a priori limit to the number of factors that may causally interact. So that there always remains the possibility that subsequent investigation will reveal other factors upon whose variation the event depends, and therefore that previous confirmations of the law maybe said to be accidental. But how are we, given two laws both of which have never been discredited by any observations, to distinguish one as accidental and the other as not? We might decide by the amount and variety of confirmation that the law has received, or by consideration of other laws with which these two are compatible or incompatible. In Mr. Kneale’s example of ‘All ravens are black’ this is said to be accidental because if ravens had lived in polar regions they might well have become white and then the law would be untrue. That is, there is an appeal to another law, that all birds living in snowy lands are white, in order to suggest that ‘All ravens are black’ is accidental, not a law at all. ‘Any four legged swan is black’ can be sensibly affirmed or denied; how much evidence we can produce one way or the other will depend upon how much knowledge we possess about swans, especially about the relations between their legs and their colour. Similarly, in the case of natural laws about freely falling bodies and frictionless surfaces, our evidence will consist of exemplified natural laws. In all cases then, the word ‘accidental’ is used for recognizable characteristics of universal material implications; either they appeared true for a while and were then falsified or there is good reason to believe, because of other laws or insufficient confirmation, that conditions for their falsification might arise.

There are circumstances in which we can recognize a law as accidental but none where we may be certain that it is not accidental. The parallel in this respect between ‘accidental law-like’ and ‘falsified-verified’ is exact. Just as we have no use for ‘completely verified’ so we have no use for ‘certainly non-accidental’; but we do not withhold the name of ‘law’ from universal statements that are not completely verified, and neither should we withhold it from universal material implications that are misleadingly called, ‘accidental’.
The use of ‘accidental’ and ‘non-accidental’ does not depend on the truth of counterfactuals which refer to conditions that have never occurred, but on observations or on other established laws. Professor Popper has proposed that “it may be useful and, for some purposes, necessary to assume a general principle stating that every kind of event that is compatible with the accepted natural laws does in fact occur in some (finite) space region”. The argument that there must be no instances of vacuous truth if universal material implications are to count as laws runs like this: If there were possible values of the antecedent that did not occur, then it is possible that if they had occurred the event referred to in the consequent might not have occurred. As long as vacuous cases of the implication remain there is still a chance that the implication is only accidentally true. But in the first place this proposal would be insufficient to achieve the required result. A law could only be said to be non-accidental if it could be shown to hold no matter what the state of all other factors might be, including those which, though observable, remain to be discovered. This would require an infinity of worlds. In the second place we have seen that there is a perfectly good sense of ‘accidental’ which does not need such absurd conditions for its satisfaction.

Consider Boyle’s law: by it we can calculate from the volume of a gas at a certain temperature to its pressure at that temperature. Interpreting it as a material implication allows us to do this for any volume that we may find experimentally; it also allows us to state counterfactuals about volumes we have not measured, for these become vacuously true when the antecedent is false and ordinary factual statements when the antecedent is fulfilled. The statement of universal material implication does not symbolize an intrinsic connexion between the volume of a gas and its pressure. But as there could not, logically, be any observation of such a necessary connexion or causal relation it is no demerit of universal material implications that they do not symbolize one.

Having examined this asserted equivalence, consider the question of whether those material implications which are not laws are irrelevant to counterfactuals. It is true that universal material implications do not imply or entail singular counterfactuals, because the latter are not of the form

\[-\varphi(a) . \varphi(a) \supset \psi(a)\]

though they are only used when ‘~\varphi(a)’ represents what is believed to be the case or holds in fact. But it is not true to say that the universal material implication “There has never been a raven that was not black and there will never be a raven that is not black” is irrelevant to the counterfactual “If this bird were a raven, it would be black”. It is argued that “There has never been a raven that was not black” is compatible with the statement that if ravens had lived in a snowy environment their offspring would
have been white. It is, the argument continues, only an historical accident that they have not lived in such regions, just as it is merely an accident that there have never been any non-black ravens. Knowing, as in fact we do not, that there never has been such ravens would not permit us to claim "If some inhabitants of snowy regions were ravens, they would be black". 7

This argument owes what plausibility it has to the example chosen. Nevertheless it is unsound. A counterfactual cannot, strictly speaking, be incompatible with a universal material implication, but two evidence statements for a counterfactual can be contradictories. Yet the absence of incompatibility is not relevant in this case. The material implication in question is relevant in that it constitutes part of the evidence upon which the counterfactual is asserted. The material implication does not so commit us that if our evidence for making it were to be different (e.g. arctic ravens) from what it is, we should still make it. With different evidence we might well frame different material implications.

"If some inhabitants of snowy regions were ravens, they would be black" is supported by referring to the universal material implication “There never has been a raven that was not black and there will never be a raven that is not black”. What gives credence to the belief that the second statement is not evidence for the first? Only that we are reminded of another material implication about protective colouration in animals, and if this were so it might be something like 'There never has been a raven whose colour was not matched to its environment'. Black arctic ravens would provide a counter example to this, just as white arctic ravens would to 'all ravens are black'. Depending upon which of these universal material implications we believe the better confirmed, we shall accept the one or the other, and with each its respective counterfactual, “If some inhabitants of snowy regions were ravens they would be black” and “If some inhabitants of snowy regions were ravens, they would be white”. An appeal to the difference of evidence ought not to be mistaken, however, for the denial that a particular universal material implication is evidence for a particular counterfactual.

III

When the opponents of the thesis that laws are universal material implications state that counterfactuals could not be obtained from them they are wrong. They may be right in asserting that counterfactuals, analysed in the way they choose, cannot be obtained. They choose to analyse them as though the statements are not counterfactual but factual. It is not even ‘If he hadn’t crossed the Rubicon, then . . . , but he did’ rather it is ‘If he hadn’t crossed the Rubicon, then . . . , and he didn’t’. Whether a law may be legitimately applied in such cases is scarcely worth arguing, for the application of the law will certainly not lead to any observable consequences. That a universal material implication

7 Analysis, vol. x, no. 6.
does not provide such deductions is, therefore, no argument. It is not enough that Boyle’s law has been confirmed and will be confirmed in the future; it is also required that it should be confirmed in all possible worlds.

When someone is faced with several courses of action the form of argument used will be a conditional, but not a counterfactual conditional. A scientific law in the form of a universal material implication will provide such conditionals, for we are only interested in the results of the argument when the antecedent is fulfilled. The view of counterfactuals that we have been arguing against amounts to carrying this procedure over to past events, ignoring the actual decision which has taken place, and considering the conditional obtained as verified true in some other sense than its determination by the actual truth values.

Suppose we say that the law is just a rule for inferring one sentence from another. The rule is used by substituting a determined sentence for the propositional function forming the antecedent and so obtaining a determined sentence for the propositional function forming the consequent. We might then say that the rule was no more than an apparatus for making these inferences and that apart from the possible substitutions it was meaningless.

Now suppose that instead of substituting determined sentences we substitute sentences which have not been determined empirically, giving us a scheme of a large number of possible inferences. These possible inferences will yield inferences from one sentence to another when the sentences are tested and the truth values put in. This way of carrying out the substitution arrives at the same result as the first way but splits the process up into two stages. The interpretation of counterfactuals as independent of material implications amounts to stopping at the first stage of this substitution process. But at this stage the usage of the symbols still only provides rules into which substitutions may be made, and these are just as meaningless apart from the substitutions as the original material implications between propositional functions. To suppose that laws in some way describe a world in which all these inapplicable inferences can be made, and that because universal material implications do not do this they fail as laws, is obviously incorrect. It is precisely the same fallacy as to suppose that laws describe necessary relations between cause and effect. Of course there is no objection to entertaining a sentence which has not been tested and making deductions from this provided that it is realised that the application of these calculations requires sentences that have been decided by observation.

In supposing that a law is a rule of inference it is necessary to allow it a truth value, so as to distinguish those rules that are to be applied from those that are not. And the decision of this truth value requires the processes of confirmation and disconfirmation. By virtue of this truth value the rule of inference has a descriptive use. Those counterfactuals which are asserted quite independently of any
evidence, as in hunches about dice throws, are statements analogous to the results of the first stage of substitution in a law; but the law from which they might be obtained has no degree of confirmation.

On this interpretation conditionals, even when particular but undecided sentences have been substituted, provide only schemata from which inferences may be constructed. Counterfactuals, derived from conditionals whose antecedents have already been decided as false, are inference schemata that will never have any application, but which are not thereby meaningless. For we could recognise situations in which the schema would apply, although we know empirically that such situations will not arise.

The main points we have tried to make are these: that the analysis of counterfactuals must be such that it is physically impossible, or believed impossible, to decide them; that the evidence for counterfactuals is evidence for general laws and the applicability of these laws, but that a counterfactual is not a statement of a law or of initial conditions. Neither is it a conditional obtained from a law by substitution, nor the conjunction of such a conditional with the denial of its antecedent, for both of these will be decidable at once. Rather it states the validity of an inference, usually derived from a law, that can for practical reasons never be made. Since there are no facts to decide these inferences no general law need describe them.