Notre Dame Journal of Formal Logic
Volume XVII, Number 3, July 1976
NDJFAM

AN ANALYSIS OF THE COUNTERFACTUAL CONDITIONAL

HOWARD C. WASSERMAN

1 *Introduction* The "problem" of the counterfactual conditional consists of providing a logical analysis of statements of the form "If ϕ were the case, then ψ would be the case." The importance of making such an analysis is that (i) a great many physical "laws" are naturally stated as counterfactual conditionals, and (ii) everyday speech is saturated with counterfactuals. The difficulty in making an adequate analysis is primarily due to the informal, unanalyzed notion of the counterfactual conditional being quite fuzzy—owing, most likely, to the facts that (i) the antecedent of a counterfactual conditional is (in most cases) presumed false, making ordinary testing of the conditional impossible, and (ii) the counterfactual conditional is surely not a truth-functional connective.

Historically, the study of counterfactual conditionals has proceeded more along philosophical lines than mathematical lines. Two of the earlier papers of this type are those of Chisolm [1] in 1946, and Goodman [2] in 1946. In 1951, Hiż [4] provided an analysis more along the lines of formal logic. He presented the view of a counterfactual conditional as a metalinguistic statement given in the context of some formal system. More explicitly, for the statement form "If p had been true, then s would be true," Hiż provides the reading: "A system S based on $p, q_1, \ldots, q_n$ as axioms and $R_1, \ldots, R_m$ as rules of inference is consistent and contains the statement s. There is such a statement r which added to S gives a new system $S_1 = S + r$ which is inconsistent. But by removing p from S_1 we obtain $S_2 = S - p + r$ which again is consistent."

Von Wright's analysis [7] in 1957, less formal than that of Hiż's, divides the "problem" into two cases: The *deductive* and the *non-deductive*. The deductive case is the case in which the statement ϕ implies the statement ψ, in which case he interprets the assertion of "If ϕ were the case, then ψ would be the case" as the denying of ϕ and the asserting that ϕ implies ψ. The non-deductive case, in which ϕ does *not* imply ψ is less satisfactorily interpreted. In this case, von Wright interprets the assertion as denying ϕ and asserting the existence of (some sort of) entailment "connexion" between ϕ and ψ.

Received November 21, 1972

Nagel [6], in 1961, provides a very clear, albeit non-mathematical, analysis which appears to be compatible with Hiż's analysis. Nagel writes that ". . . a counterfactual can be interpreted as an implicit *metalinguistic* statement . . . asserting that the indicative form of its consequent clause follows logically from the indicative form of its antecedent clause, when the latter is conjoined with some law and the requisite initial conditions for the law." As a particular example, he considers a physicist accounting for the failure of some previous experiment asserting "If the length of pendulum $\mathfrak{a}$ had been shortened to one-fourth its actual length, its period would have been half its actual period." He re-states this, then, as follows: "The statement 'The period of the pendulum $\mathfrak{a}$ was half its present period' *follows logically* from the supposition 'The length of $\mathfrak{a}$ was one-fourth its present length,' when this supposition is conjoined with the law that the period of a simple pendulum is proportional to the square root of its length, together with a number of further assumptions about initial conditions for the law . . . ."

Another, semi-formal, analysis of counterfactuals was provided by Goodman [3] in 1965. He provides the rule that a counterfactual with antecedent A and consequent C ". . . is true if and only if there is some set S of true sentences (not a consequence of $\sim A$)[1] such that S is compatible with C and with $\sim C$ and such that $A \cdot S$ is self-compatible and leads by law to C; while there is no set S' (not a consequence of $\sim A$)[1] compatible with C and with $\sim C$, and such that $A \cdot S'$ is self-compatible and leads by law to $\sim C$."

As far as the present paper is concerned, it is left to the reader to make comparisons of the analysis given here with those analyses described above. It turns out that there are a number of similarities. However, the approach taken here is intended to be strictly along the lines of formal logic. The "philosophical" motivation for the formal semantics provided here for a statement of the form "If ϕ were the case, then ψ would be the case" is that such a statement is about some "world", "state-of-affairs", or, more formally, some structure S, and that the statement "means" that ψ holds in every structure which differs from S "just enough" to make ϕ true. This "philosophical" notion is formalized in the contexts of sentential logic and first-order predicate logic. It is shown that "counterfactual sentential logic" is decidable, but that the set of logically true formulas of "counterfactual first-order predicate logic" is not even recursively enumerable.

2 *Counterfactual Sentential Logic* Let $\mathcal{L}$ be the sentential language based on the connectives $\supset$ and $\sim$. Let $\mathcal{L}'$ be the superlanguage obtained from $\mathcal{L}$ by adding the binary connective $\mathrel{\supset\!\!\cdot}$ (which shall be called the *counterfactual conditional*) with no nesting of $\mathrel{\supset\!\!\cdot}$. More explicitly, $\mathcal{L}'$ is defined to be the least set of expressions closed under the following formation rules:

1. We incorporate here Goodman's footnote into his quote.

(i) Every sentential variable belongs to $\mathcal{L}'$.
(ii) If $\phi, \psi \in \mathcal{L}'$, then $(\phi \supset \psi) \in \mathcal{L}'$.
(iii) If $\phi \in \mathcal{L}'$, then $\sim\phi \in \mathcal{L}'$.
(iv) If $\phi, \psi \in \mathcal{L}$, then $(\phi \mathrel{\supset\!\!\!\rightarrow} \psi) \in \mathcal{L}'$.

For $\phi, \psi \in \mathcal{L}$, the suggested reading of $(\phi \mathrel{\supset\!\!\!\rightarrow} \psi)$ is "if ϕ were the case, then ψ would be the case."

An *assignment* (of truth-values) is a mapping of the set of all sentential variables into the set $\{0, 1\}$. An assignment, $\mathfrak{A}$, extends to a valuation on $\mathcal{L}$ (also denoted "$\mathfrak{A}$") in the usual fashion. The notion of model of a set of formulas of $\mathcal{L}$, and the derivative notions of tautology, satisfiability, and implication shall be the standard ones. Let $\phi \in \mathcal{L}$, and let $\mathfrak{A}$ be an assignment. Let $T_{\mathfrak{A}} = \{\phi\colon \phi \in \mathcal{L} \text{ and } \mathfrak{A}(\phi) = 1\}$. A *truth-set for* ϕ *in* $\mathfrak{A}$ is herewith defined as a subset of $T_{\mathfrak{A}}$, maximal with respect to joint satisfiability with ϕ. An assignment $\mathfrak{A}'$ shall be called a *ϕ-variety of* $\mathfrak{A}$ in case $\mathfrak{A}'$ is a model of $S \cup \{\phi\}$, for some truth-set, S, for ϕ in $\mathfrak{A}$. For $\psi \in \mathcal{L}$, the truth-value of $(\phi \mathrel{\supset\!\!\!\rightarrow} \psi)$ in $\mathfrak{A}$ is defined as follows:

$$\mathfrak{A}(\phi \mathrel{\supset\!\!\!\rightarrow} \psi) = 1 \text{ if and only if } \mathfrak{A}'(\psi) = 1 \text{ for every } \phi\text{-variety, } \mathfrak{A}', \text{ of } \mathfrak{A}.$$

Note that if $\mathfrak{A}(\phi) = 1$, then $T_{\mathfrak{A}}$ is the *unique* truth-set for ϕ in $\mathfrak{A}$, and hence $\mathfrak{A}(\phi \mathrel{\supset\!\!\!\rightarrow} \psi) = 1$ if and only if $\mathfrak{A}(\psi) = 1$.[2]

Lemma 1 *Let* $\phi \in \mathcal{L}$, *let* $\mathfrak{A}$ *and* $\mathfrak{A}'$ *be assignments, and suppose that* $\mathfrak{A}'(\phi) = 1$. *Then* $\mathfrak{A}'$ *is a ϕ-variety of* $\mathfrak{A}$ *if and only if* $T_{\mathfrak{A}} \cap T_{\mathfrak{A}'}$ *is a truth-set for* ϕ *in* $\mathfrak{A}$ (*and, incidentally, the only truth-set for* ϕ *in* $\mathfrak{A}$ *of which* $\mathfrak{A}'$ *is a model*).

Proof: If $T_{\mathfrak{A}} \cap T_{\mathfrak{A}'}$ is a truth-set for ϕ in $\mathfrak{A}$, then, since $\mathfrak{A}'$ is a model of $(T_{\mathfrak{A}} \cap T_{\mathfrak{A}'}) \cup \{\phi\}$, $\mathfrak{A}'$ is a ϕ-variety of $\mathfrak{A}$. Conversely, suppose that $\mathfrak{A}'$ is a ϕ-variety of $\mathfrak{A}$. Then $\mathfrak{A}'$ is a model of some truth-set, S, for ϕ in $\mathfrak{A}$. Hence $S \subseteq T_{\mathfrak{A}} \cap T_{\mathfrak{A}'}$. But $T_{\mathfrak{A}} \cap T_{\mathfrak{A}'} \subseteq T_{\mathfrak{A}}$; $(T_{\mathfrak{A}} \cap T_{\mathfrak{A}'}) \cup \{\phi\}$ is satisfiable ($\mathfrak{A}'$ is a model). Hence, the maximality of S gives that $T_{\mathfrak{A}} \cap T_{\mathfrak{A}'} = S$.

Lemma 2 *Let* $\phi \in \mathcal{L}$, *let* $\mathfrak{A}$ *be an assignment, and let* $\mathfrak{A}'$ *and* $\mathfrak{A}''$ *be ϕ-varieties of* $\mathfrak{A}$. *Then* $\mathfrak{A}' = \mathfrak{A}''$ *if and only if* $\mathfrak{A}'$ *and* $\mathfrak{A}''$ *are models of the same truth-set for* ϕ *in* $\mathfrak{A}$.

Proof: Clearly, if $\mathfrak{A}' = \mathfrak{A}''$, then $\mathfrak{A}'$ and $\mathfrak{A}''$ are models of the same truth-set for ϕ in $\mathfrak{A}$. Suppose, conversely, that $\mathfrak{A}'$ and $\mathfrak{A}''$ are models of the same truth-set, S, for ϕ in $\mathfrak{A}$. Let $S' = T_{\mathfrak{A}} \cap T_{\mathfrak{A}'}$ and let $S'' = T_{\mathfrak{A}} \cap T_{\mathfrak{A}''}$. Since $\mathfrak{A}'$ and $\mathfrak{A}''$ are ϕ-varieties of $\mathfrak{A}$, it follows from Lemma 1 that $S' = S$ and

2. An alternative approach would be always to take $\mathfrak{A}(\phi \mathrel{\supset\!\!\!\rightarrow} \psi) = 0$ when $\mathfrak{A}(\phi) = 1$. This approach is not taken here for two reasons:

(i) This approach would trivialize the notion of $\mathrel{\supset\!\!\!\rightarrow}$ in the sense that one would have $\models(\phi \mathrel{\supset\!\!\!\rightarrow} \psi)$ if and only if $\models\sim\phi$ (for all $\psi \in \mathcal{L}$).
(ii) In English, when ϕ is a true sentence, one is more inclined to consider the sentence $(\phi \mathrel{\supset\!\!\!\rightarrow} \psi)$ as "odd" rather than false (which suggests, incidentally, that an attempt to adjoin $\mathrel{\supset\!\!\!\rightarrow}$ to a three-valued logic such as Keenan's [5] is worthy of some effort).

$S'' = S$, hence $S' = S''$. Thus, for every $\psi \in T_{\mathfrak{A}}$,

$$\mathfrak{A}'(\psi) = 1 \Leftrightarrow \psi \in S' \Leftrightarrow \psi \in S'' \Leftrightarrow \mathfrak{A}''(\psi) = 1.$$

Likewise, for $\psi \in \mathcal{L} - T_{\mathfrak{A}}$, $\sim\psi \in T_{\mathfrak{A}}$; hence,

$$\mathfrak{A}'(\psi) = 1 \Leftrightarrow \mathfrak{A}'(\sim\psi) = 0 \Leftrightarrow \sim\psi \notin S' \Leftrightarrow \sim\psi \notin S'' \Leftrightarrow \mathfrak{A}''(\sim\psi) = 0 \Leftrightarrow \mathfrak{A}''(\psi) = 1.$$

Thus, $\mathfrak{A}' = \mathfrak{A}''$.

Lemma 3 *Let $\phi, \psi \in \mathcal{L}$, let $\mathfrak{A}$ be an assignment, let S be a truth-set for ϕ in $\mathfrak{A}$, and let $\mathfrak{A}'$ be a model of $S \cup \{\phi\}$. Then $\psi \in T_{\mathfrak{A}'}$ if and only if $S \cup \{\phi, \psi\}$ is satisfiable.*

Proof: If $\psi \in T_{\mathfrak{A}'}$, then $\mathfrak{A}'$ is a model of $S \cup \{\phi, \psi\}$. Assume, conversely, that $S \cup \{\phi, \psi\}$ is satisfiable. Let $\mathfrak{A}''$ be a model of $S \cup \{\phi, \psi\}$. Then $\mathfrak{A}'$ and $\mathfrak{A}''$ are models of the same truth-set S, and hence, by Lemma 2, $\mathfrak{A}' = \mathfrak{A}''$. Thus, $\psi \in T_{\mathfrak{A}'}$.

Theorem 1 *Let $\phi, \psi \in \mathcal{L}$, and let $\mathfrak{A}$ be an assignment. Then $\mathfrak{A}(\phi \supset \psi) = 1$ if and only if every formula in $T_{\mathfrak{A}}$ which is jointly satisfiable with ϕ is jointly satisfiable with $\{\phi, \psi\}$* (i.e., *for each $\eta \in T_{\mathfrak{A}}$, if $\{\eta, \phi\}$ is satisfiable, then $\{\eta, \phi, \psi\}$ is satisfiable*).

Proof: $\mathfrak{A}(\phi \supset \psi) = 1$

$\Leftrightarrow$ $\mathfrak{A}'(\psi) = 1$ for every ϕ-variety, $\mathfrak{A}'$, of $\mathfrak{A}$

$\Leftrightarrow$ $\mathfrak{A}'(\psi) = 1$ for every truth-set, S, for ϕ in $\mathfrak{A}$, and every model, $\mathfrak{A}'$, of $S \cup \{\phi\}$

$\Leftrightarrow$ $S \cup \{\phi, \psi\}$ is satisfiable, for every truth-set, S, for ϕ in $\mathfrak{A}$ (by Lemma 3)

$\Leftrightarrow$ $F \cup \{\phi, \psi\}$, is satisfiable, for all finite subsets, F, of truth-sets for ϕ in $\mathfrak{A}$ (by compactness of $\mathcal{L}$)

$\Leftrightarrow$ $F \cup \{\phi, \psi\}$ is satisfiable for every finite subset, F, of $T_{\mathfrak{A}}$ jointly satisfiable with ϕ (since every such F is extendible, by a Lindenbaum procedure, to a truth-set for ϕ in $\mathfrak{A}$)

$\Leftrightarrow$ $\{\eta, \phi, \psi\}$ is satisfiable, for every $\eta \in T_{\mathfrak{A}}$ jointly satisfiable with ϕ (by conjoining the members of F).

Theorem 2 *There is an effective procedure, which, when provided with any formulas $\phi, \psi \in \mathcal{L}$ and the restriction of any assignment $\mathfrak{A}$ to the variables occurring in ϕ and ψ, will decide the truth or falsity of $(\phi \supset \psi)$ in $\mathfrak{A}$.*

Proof: It is readily verifiable that $\mathfrak{A}(\phi \supset \psi) = 0$ if and only if either (i) $\mathfrak{A}(\phi) = 1$ and $\mathfrak{A}(\psi) = 0$, or (ii) ϕ is satisfiable but $\{\phi, \psi\}$ unsatisfiable, or (iii) $\{\phi, \psi\}$ is satisfiable, and there is a disjunctive normal form in the variables occurring in ϕ, ψ which is true in $\mathfrak{A}$, jointly satisfiable with ϕ, but not jointly satisfiable with $\{\phi, \psi\}$. Thus, an algorithm for deciding the truth-value of $(\phi \supset \psi)$ in $\mathfrak{A}$ is the following:

1. Test whether $\mathfrak{A}(\phi) = 1$ and $\mathfrak{A}(\psi) = 0$; if yes, return "0", and halt; else
2. Test the satisfiability of ϕ by a truth-table; if ϕ is unsatisfiable, return "1", and halt; else

3. Test the joint satisfiability of ϕ, ψ by a truth-table; if unsatisfiable, return "0", and halt; else
4. List all disjunctive normal forms in the variables occurring in ϕ, ψ which are true in $\mathfrak{A}$; test these in turn for joint satisfiability with ϕ and unsatisfiability with $\{\phi, \psi\}$; if this last series of tests fails for all the forms in the list, return "1" and halt; else, return "0" and halt.

Corollary *The decision problem for $\mathcal{L}'$ is solvable.*

3 *First-Order Counterfactual Logic* Given a first-order language $\mathcal{L}_K$ (where K is the set of non-logical constants of the language), let $\mathcal{L}'_K$ be the superlanguage of $\mathcal{L}_K$ obtained by adjoining the connective $\supset$ without nesting (i.e., applying $\supset$ only to pairs of formulas not containing $\supset$). Let the usual notion of interpretation be assumed (i.e., that of a mapping which associates with each variable and with each individual constant of K, a member of a domain of discourse U, with each n-ary function symbol, an n-ary operation in U, and with each n-ary relation symbol, a subset of U^n). Let the derivative notions of model, logical truth, and satisfiability also be the usual ones. Then the formal definitions leading up to and including the definition of the truth-value of a formula $(\phi \supset \psi)$ in an assignment all carry over to $\mathcal{L}'_K$ (with "interpretation" replacing "assignment"). Then the decision problem for $\mathcal{L}'_K$ is unsolvable. In fact, there is a stronger result.

Theorem *The set of logically true formulas of $\mathcal{L}'_K$ is not recursively enumerable.*

Proof: It is easily seen that for every formula $\phi \in \mathcal{L}_K$, $\models \phi$ if and only if $\models (\sim\phi \supset \phi)$, and, moreover, $\not\models \phi$ if and only if $\models \sim(\sim\phi \supset \phi)$. Now, since exactly one of $\models \phi$ and $\not\models \phi$ is the case, exactly one of $\models(\sim\phi \supset \phi)$ and $\models\sim(\sim\phi \supset \phi)$ is the case. Thus, were the set of logically true formulas of $\mathcal{L}'_K$ recursively enumerable, then by effectively enumerating this set until one "reached" $(\sim\phi \supset \phi)$ or $\sim(\sim\phi \supset \phi)$, one would decide the logical truth of an arbitrary formula $\phi \in \mathcal{L}_K$, contradicting the recursive undecidability of $\mathcal{L}_K$.

REFERENCES

[1] Chisolm, R. M., "The contrary-to-fact conditional," *Mind*, vol. LV (1946), pp. 289-307.

[2] Goodman, N., "The problem of counterfactual conditionals," *The Journal of Philosophy*, vol. 44 (1947), pp. 113-128.

[3] Goodman, N., *Fact, Fiction, and Forecast*, 2nd Edition, Bobs-Merrill, Indianapolis (1965).

[4] Hiż, H., "Comments and criticism on the inferential sense of contrary-to-fact conditionals," *The Journal of Philosophy*, vol. 48 (1951), pp. 586-587.

[5] Keenan, E., "A logical base for a transformation grammar of English" (Doctoral Dissertation), *Transformations and Discourse Analysis Papers*, No. 82, University of Pennsylvania, Philadelphia (1970).

[6] Nagel, E., *The Structure of Science; Problems in the Logic of Scientific Explanation*, Harcourt, Brace & World, Inc., New York (1961).

[7] Von Wright, G. H., *Logical Studies*, The Humanities Press, Inc., London (1957).

Queens College of C.U.N.Y.
Flushing, New York