Notre Dame Journal of Formal Logic
Volume XVI, Number 4, October 1975
NDJFAM

COUNTERFACTUALS

DONALD NUTE

A complete analysis of our ordinary use of counterfactual conditionals should provide us with a means of determining (at least in principle) the truth value of any ordinary counterfactual claim. Such an analysis is a much more ambitious project than I propose to undertake here. A more modest goal would be to provide a means of determining the *validity* of any ordinary counterfactual claim. This is still a very ambitious project, so I will concentrate on an account of the validity of counterfactuals which does not consider any problems of quantification. A number of authors have made recent attempts at developing an adequate conditional sentence logic. I will examine these attempts and pinpoint certain controversial assumptions upon which they are based. Then I will offer two new calculi which are based upon the denial of these assumptions. Finally, I will produce proof sketches of the semantical completeness and decidability of these two new systems using a method of proof for decidability unlike that of any other author writing on counterfactuals. I should warn the reader in advance that it is not my purpose to show the inadequacy of any of the systems I criticize; I am rather concerned with showing the diversity of uses we made of counterfactuals. The new logics I develop are not intended to replace those offered by others, but to augment their efforts. In short, I hope to show that we use counterfactuals on different occasions in different and even incompatible ways. Some of these usages—I would even claim some of the most common usages—have not been investigated until now.

Where "$>$" is the counterfactual connective, there are three schemata crucial to my discussion:

(1) $(A > B) \vee (A > -B)$;
(2) $(A \& B) \supset (A > B)$;
(3) $\Box(A \supset B) \supset. (B > C) \supset. \Diamond(A \& C) \supset. (A > C)$.

In their article "A Semantical Analysis of Conditional Logic,"[1] Robert Stalnaker and Richmond Thomason construct two deductive systems, C1 and C2, both of which have (1) as a theorem schema. David Lewis, in

Received July 4, 1973

his article "Completeness and Decidability of Three Logics of Counterfactual Conditionals,"[2] discards (1) as counterintuitive and constructs a system somewhat weaker than Stalnaker's and Thomason's C2 which lacks (1) as a theorem schema and which, he suggests, properly characterizes ordinary usage. However, Lewis' favorite system contains (2) as an axiom.

Before arguing that (2) may be counterintuitive also, I would like to present an informal argument intended to show that (2) follows from (1). Suppose P and Q are true propositions. Then we surely would not say that $P > -Q$. So if we accept (1) we must conclude that $P > Q$. Since (1) does entail (2) at the intuitive level, we need only discredit (2) to discredit (1).[3]

Smith and Jones chance to meet Bill Russell on the street without recognizing him. Referring to Russell, Smith says to Jones, "That man would walk exactly that way if he were a basketball player." Although Smith's statement is a counterfactual with true antecedent and consequent, it seems clear that it might yet be false. Jones might reasonably reply, "That's not true. He needn't walk that way just because he happened to be a basketball player. A lot of tall men walk that way and a lot of tall basketball players walk differently." Examples like this and others that might come to mind should convince the reader that (2) is not intuitively valid. In fact, our whole attitude toward such counterfactuals is quite different. Rather than base the truth of a counterfactual upon the truth of its component statements, we most often use the truth of the counterfactual as evidence of the truth of its antecedent. For example, if Jones had not contradicted him, Smith might have gone on to say, "I'll wager he *is* a basketball player, from that walk." Of course the truth of a counterfactual and its consequent do not logically guarantee the truth of its antecedent, but they frequently are used as evidence for the truth of the antecedent.[4] This practice is quite different from the practice represented by (2).

Let us now examine (3). One consequence of (3) is

(4) $(B > C) \supset . \Diamond(A \ \& \ C) \supset . (A \ \& \ B > C)$.

This is very similar to the much-discussed schema

(5) $(B > C) \supset (A \ \& \ B > C)$.

It seems to me there are divergent intuitions and usages connected with (4). Although he was directly concerned with (5), Stalnaker implicitly objected to (4) as well when he argued that ". . . we cannot always strengthen the antecedent of a true conditional and have it remain true. Consider 'If this match were struck, it would light,' and 'If this match had been soaked in water overnight *and* it were struck, it would light.'"[5] We might interpret Stalnaker's example in either of two ways. First, we might agree with Stalnaker that his example is a counter-example and reject (4). Or we might hold that his example shows "If this match were struck, it would light" to be false.

What can we conclude from these considerations? First, I would not want to deny that counterfactual conditionals are ever used in a way which assumes the validity of (1) and/or (2), or that (1) and (2) are categorically counterintuitive. I would instead claim that I have shown examples of ordinary usage which assumes the invalidity of both (1) and (2). Similarly, I have tried to show that a particular speaker on a particular occasion might assume either the validity or invalidity of (3). The conclusion that we must draw is that there are *several different* counterfactual connectives involved in our ordinary usage. The logic of these different connectives varies in ways which can be represented by the different conditional calculi we use to represent them. Stalnaker, Thomason, and Lewis have already constructed calculi which display the formal properties of a portion of our ordinary use of counterfactuals, but there are some usages which they have overlooked. It is these that I will now try to characterize.

Beginning with C2 as axiomatized by Stalnaker and Thomason, we can produce a logic which avoids (1) and (2). At the same time, we can construct our new logic so that it embraces (3). First we weaken their Axiom 5 to eliminate (1) and (2); then if we choose, we add (3) to the resulting axioms. The new systems, which I will call "C3" and "C4", are determined by the following rules and axioms.

Rules for C3 and C4:

R1. *From A and $A \supset B$ to infer B.*
R2. *From A to infer $\Box A$.*

Axioms for C3:

A1. *Any tautologous wff is an axiom.*
A2. $\Box(A \supset B) \supset (\Box A \supset \Box B)$.
A3. $\Box(A \supset B) \supset (A > B)$.
A4. $\Box A \supset. (A > B) \supset -(A > -B)$.
A5*. $A > (B \supset C) \supset. (A > B) \supset (A > C)$.
A6. $(A > B) \supset (A \supset B)$.
A7. $\Box(A \equiv B) \supset. (A > C) \supset (B > C)$.
A8. $A > (B \;\&\; C) .\supset. (A \;\&\; B) > C$.

Axioms for C4: A1-A8, and

A9. $\Box(A \supset B) \supset. (B > C) \supset. \Diamond(A \;\&\; C) \supset (A > C)$.

Provability and derivability are defined in the usual ways using these rules and axioms. Inconsistent$_{3,(4)}$, consistent$_{3,(4)}$, and maximal$_{3,(4)}$ sets of wffs are also defined in the standard manner.

For our semantics we need the notion of a model structure (Ms)—a structure $\langle K, R\rangle$ such that K is a nonempty class and R is a binary reflexive relation on K. $\emptyset$ will serve the purpose of the isolated element or "absurd world" (to use Stalnaker's and Thomason's phrase) in the Cms since our selection functions will pick out classes instead of single elements. A valuation on an Ms M is defined in the usual way.

D1. A *class-selection function* (*cs-function*) on an Ms $\langle K, R\rangle$ is any two-place function f taking wffs and members of K into subsets of K such that for all $\alpha, \beta \in K$, if $\beta \in f(A, \alpha)$ for some wff A, then $\alpha R \beta$.

D2. A *quasi-interpretation** I on an Ms M is an ordered pair $\langle v, f\rangle$ where v is a valuation on M and f is a cs-function on M.

D3. Where I is a quasi-interpretation* on an Ms $\langle K, R\rangle$, we define the truth value $I\alpha(A)$ for wff A and $\alpha \in K$ in the usual way for the cases in which there are wffs B and C such that A is a sentential variable, $A = -B$, $A = B \vee C$, or $A = \Box B$. Additionally, for wffs B and C such that $A = B > C$, $I\alpha(A) = \mathsf{T}$ if $I\beta(C) = \mathsf{T}$ for all $\beta \in f(B, \alpha)$ and $I\alpha(A) = \mathsf{F}$ otherwise.

D4. An *interpretation** I on an Ms $\langle K, R\rangle$ is a quasi-interpretation* on $\langle K, R\rangle$ which for all wffs A, B, and C, and all $\alpha \in K$ meets the following four conditions:

1) $I\beta(A) = \mathsf{T}$ for all $\beta \in f(A, \alpha)$;
2) If $I\alpha(A) = \mathsf{T}$, then $\alpha \in f(A, \alpha)$;
3) If $f(A, \alpha) = \emptyset$, then there is no $\beta \in K$ such that $I\beta(A) = \mathsf{T}$ and $\alpha R \beta$;
4) If $I\beta(A) = \mathsf{T}$ for all $\beta \in f(B, \alpha)$ and $I\gamma(B) = \mathsf{T}$ for all $\gamma \in f(A, \alpha)$, then $f(A, \alpha) = f(B, \alpha)$.

D5. A *regular interpretation** I on an Ms $\langle K, R\rangle$ is quasi-interpretation* on $\langle K, R\rangle$ which for all wffs A and B, and all $\alpha \in K$ meets the following condition:

1) If $\{\beta \in K\colon \alpha R \beta \text{ and } I\beta(A \,\&\, -B) = \mathsf{T}\} = \emptyset$, $I\alpha(B > C) = \mathsf{T}$ and there is a $\beta \in K$ such that $\alpha R \beta$ and $I\beta(A \,\&\, C) = \mathsf{T}$, then $I\alpha(A > C) = \mathsf{T}$.

D6. 1) A set Γ of wffs is *simultaneously satisfiable*$_{3(4)}$ if there exists a(n) (regular) interpretation* I on an Ms $\langle K, R\rangle$ and a member α of K such that for all $A \in \Gamma$, $I\alpha(A) = \mathsf{T}$.
2) A set Γ of wffs *implies*$_{3(4)}$ a wff A if $\Gamma \cup \{-A\}$ is not simultaneously satisfiable$_{3(4)}$.
3) A wff A is *valid*$_{3(4)}$ if $\emptyset$ implies$_{3(4)}$ A.

Adapting the corresponding proof by Stalnaker and Thomason,[6] we can easily show that

T7. $\vdash_{\overline{C3(C4)}} A$ *iff* A *is valid*$_{3(4)}$.

Clearly, (3) is a theorem schema of C4. Now we must make sure that (1) and (2) are theorem schemata of neither C3 nor C4. Consider the Ms

$$M = \langle \{0, 1\}, \{(1, 1), (1, 0), (0, 1), (0, 0)\}\rangle.$$

Let v be any valuation on M such that for sentential variables P and Q, $v_1(P) = \mathsf{T}$, $v_1(Q) = \mathsf{T}$, $v_0(P) = \mathsf{T}$, and $v_0(Q) = \mathsf{F}$. Then we can define a function f from wffs and members of $\{0, 1\}$ into subsets of $\{0, 1\}$ recursively such that $I = \langle v, f\rangle$ is an interpretation* on M and $f(P, 1) = \{0, 1\}$. Then we can easily show that $I_1[(P > Q) \vee (P > -Q)] = \mathsf{F}$ and $I_1[P \,\&\, Q \supset (P > Q)] = \mathsf{F}$, completing our argument.

Thomason[7] has proven decidability for C2 and Lewis[8] has proven decidability for all the systems he considers using a method different from Thomason's. The following is an algebraic proof of the decidability of C3 and C4 which differs in method from those of either Thomason or Lewis. This proof method is adapted from the work of E. J. Lemmon.[9]

D8. A structure $\mathfrak{M} = \langle M, \cup, \cap, *, \rhd \rangle$ is a *normal* algebra* iff M is a set of elements closed under the operations $\cup$, $\cap$, $*$, and $\rhd$ such that for all $x, y, z \in M$:

1) $\langle M, \cup, \cap, * \rangle$ is a Boolean algebra;
2) $(*x \cup *y) \rhd (x \cap y) = (*x \rhd x) \cap (*y \rhd y)$;
3) $(*x \rhd x) \leqslant x$ (where $x \leqslant y$ iff $x \cap y = x$ iff $x \cup y = y$);
4) $(0 \rhd 1) = 1$.

D9. 1) $\mathrm{N}x = *x \rhd x$; 2) $x \rightarrow y = *x \cup y$; 3) $x \leftrightarrow y = (x \rightarrow y) \cup (y \rightarrow x)$.

D10. An algebra is *standard* iff in addition to being normal*, it satisfies the following postulates:

5) $\mathrm{N}(x \rightarrow y) \leqslant x \rhd y$;
6) $(x \rhd y) \cap (x \rhd *y) \leqslant (x \rhd *x)$;
7) $(x \rhd (y \rightarrow z)) \cap (x \rhd y) \leqslant x \rhd z$;
8) $x \rhd y \leqslant x \rightarrow y$;
9) $x \rhd (y \cap z) \leqslant (x \cap y) \rhd z$;
10) $\mathrm{N}(x \leftrightarrow y) \cap (y \rhd z) \leqslant x \rhd z$;

D11. An algebra is *regular* iff in addition to being standard, it satisfies the following postulate:

11) $\mathrm{N}(x \rightarrow y) \cap (y \rhd z) \cap *\mathrm{N}(*(x \cap z)) \leqslant x \rhd z$.

D12. Let $\mathfrak{M} = \langle M, \cup, \cap, *, \rhd \rangle$ be a standard (regular) algebra. An *interpretation* of C3(C4) in $\mathfrak{M}$ is a one-place function I which sends wffs of C3(C4) into M and which satisfies the following conditions:

1) $\mathrm{I}(-A) = *\mathrm{I}(A)$;
2) $\mathrm{I}(A \vee B) = \mathrm{I}(A) \cup \mathrm{I}(B)$;
3) $\mathrm{I}(A > B) = \mathrm{I}(A) \rhd \mathrm{I}(B)$.

D13. A standard (regular) algebra $\mathfrak{M}$ *satisfies*$_{3(4)}$ a wff A iff for every interpretation I for C3(C4) in $\mathfrak{M}$, $\mathrm{I}(A) = 1$.

T14. *If* $\vdash_{\overline{C3(C4)}} A$, *then for every standard (regular) algebra* $\mathfrak{M}$, $\mathfrak{M}$ *satisfies*$_{3(4)}$ A.

The proof of T14 is straightforward and will be omitted.

T15. *If A is not a theorem of* C3(C4), *then there is a standard (regular) algebra* $\mathfrak{M} = \langle M, \cup, \cap, *, \rhd \rangle$ *such that* $\mathfrak{M}$ *does not satisfy*$_{3(4)}$ A *and* $\mathrm{Card}(M) \leqslant 2^{2^{\mathrm{Card}(\mathrm{SF}(A))}}$ (*where* $\mathrm{SF}(A)$ *is the set of subformulas of* A).

Proof sketch: I shall only provide a sketch for the case of C3 since the case of C4 is proven in exactly the same way. Assume A is not a theorem of C3 and let $\mathrm{M}^* = \langle K, R \rangle$ be an Ms and $\mathrm{I}^* = \langle \mathrm{v}, f \rangle$ be an interpretation* on M^* such that for some $\alpha \in K$, $\mathrm{I}^*_{\alpha}(A) = \mathbf{F}$. Such an M^* and I^* exist by T7.

For each wff B let $\mathrm{Im}(B) = \{\alpha \in K: \mathrm{I}^{*}_{\alpha}(B) = \mathrm{T}\}$. Let $\mathfrak{M} = \langle M, \cup, \cap, *, \triangleright\rangle$ such that:

1) $M = \bigcap\{\Gamma: \{\mathrm{Im}(B): B \in \mathrm{SF}(A)\} \subseteq \Gamma$ and if $\mathrm{Im}(B), \mathrm{Im}(C) \in \Gamma$, then $\mathrm{Im}(-B)$, $\mathrm{Im}(B \vee C)$, $\mathrm{Im}(B > C) \in \Gamma\}$;
2) $\cup, \cap, *$ are the set-theoretic operations of union, intersection, and complementation;

3) $\mathrm{Im}(B) \triangleright \mathrm{Im}(C) = \bigcup\{\mathrm{Im}(D) \in M: \mathrm{Im}(D) \subseteq \mathrm{Im}(B > C)\}$ for all $\mathrm{Im}(B), \mathrm{Im}(C) \in M$.

I will omit the uncomplicated proof that $\mathfrak{M}$ is a standard algebra with at most $2^{2^{\mathrm{Card}(\mathrm{SF}(A))}}$ elements. All that remains is to show that $\mathfrak{M}$ does not satisfy$_3$ A. Define a function I from the set of wffs into M recursively as follows:

1) $\mathrm{I}(P) = \bigcup\{\mathrm{Im}(D) \in M: \mathrm{Im}(D) \subseteq \mathrm{Im}(P)\}$ for all sentential variables P;
2) $\mathrm{I}(-B) = {*}\mathrm{I}(B)$;
3) $\mathrm{I}(B \vee C) = \mathrm{I}(B) \cup \mathrm{I}(C)$;
4) $\mathrm{I}(B > C) = \mathrm{I}(B) \triangleright \mathrm{I}(C)$.

By (2), (3), and (4), I is an interpretation for C3 in $\mathfrak{M}$. Furthermore, since $\mathrm{I}(P) = \mathrm{Im}(P)$ for all sentential variables $P \in \mathrm{SF}(A)$, it is clear that $\mathrm{I}(B) = \mathrm{Im}(B)$ for all wffs $B \in \mathrm{SF}(A)$. Hence, $\mathrm{I}(A) = \mathrm{Im}(A) \neq K$ (by hypothesis) and $\mathfrak{M}$ does not satisfy A. Combining T14 and T15, we get the following completeness and decidability result:

T16. $\vdash_{\overline{C3(C4)}} A$ *iff* A *is satisfied*$_{3(4)}$ *by every standard (regular) algebra with cardinality no greater than* $2^{2^{\mathrm{Card}(\mathrm{SF}(A))}}$.

We now have two additional conditional sentence logics. Unlike Stalnaker or Lewis, I am not prepared to claim that any of these is an adequate representation of the logic of counterfactuals as we ordinarily use them Instead, I would claim that we use counterfactuals in different (and incompatible) ways on different occasions with the result that several of the logics developed characterize one of the usages commonly occurring. I suspect that C3 and C4 represent the most common usages (C3 being the favorite), but this is an empirical question not to be decided by logicians.

NOTES

1. R. Stalnaker and R. Thomason, "A semantical analysis of conditional logic," *Theoria*, vol. 36 (1970), pp. 23-42.

2. D. Lewis, "Completeness and decidability of three logics of counterfactual conditionals," *Theoria*, vol. 37 (1972), pp. 75-85.

3. (2) is also a *formal* consequence of (1) in the system C2.

4. This practice was pointed out to me by Dr. Romane Clark.

5. R. Stalnaker, "A theory of conditionals," in N. Rescher (ed.), *Studies in Logical Theory*, *American Philosophical Quarterly* supplementary monograph series, Blackwell, Oxford (1968), p. 106. Adding schema (5) to either of the systems C3 or C4 developed later in this paper will produce the result $\vdash_{C3, C4} (A > B) \equiv \Box(A \supset B)$ as follows. We already have one direction by A3. Suppose $\langle K, R\rangle$ is an Ms, $\mathsf{I} = \langle v, f\rangle$ is a regular interpretation* on $\langle K, R\rangle$, and $\alpha \epsilon K$ such that $\mathsf{I}\alpha(B > C) = \mathbf{T}$. Then by (5) and R1., $\mathsf{I}\alpha(B \,\&\, - C > C) = \mathbf{T}$, i.e., $\mathsf{I}\beta(C) = \mathbf{T}$ for all $\beta \epsilon f(B \,\&\, - C, \alpha)$. Hence, $\mathsf{I}\beta(B \,\&\, - C \,\&\, C) = \mathbf{T}$ for all $\beta \varepsilon f(B \,\&\, - C, \alpha)$ and $f(B \,\&\, - C, \alpha) = \phi$ by the definition of a regular interpretation*. But then there does not exist a $\beta \varepsilon K$ such that $\alpha R \beta$ and $\mathsf{I}\beta(B \,\&\, - C) = \mathbf{T}$ by the definition of a regular interpretation*. Therefore, $\mathsf{I}\alpha\Box(B \supset C) = \mathbf{T}$. Using our completeness result, we get the proposed result. The same result can be obtained by adding (5) to any conditional sentential logic with which I am familiar, so (5) must be rejected if the counterfactual connective is to represent anything weaker than strict implication.

6. The reference is to the work cited in note 1. Their D5.4 must be changed since our function will pick out classes. The new function will be defined as follows: for all A and all $\Delta' \epsilon K$, $f(A, \Delta') = \{\Delta : A > B \epsilon \Delta\}$. Since (1) is not a theorem schema for C3, $f(A, \Delta')$ will not in general be a singleton as it is in the semantics for C2.

7. *Cf.* R. Thomason, "The finite model property in conditional sentence logic," mimeographed and distributed privately. So far as I know, this paper has not been published.

8. *Op. cit.*

9. *Cf.*, E. J. Lemmon, "Algebraic semantics for modal logics I," *The Journal of Symbolic Logic*, vol. 31 (1966), pp. 46-65.

Indiana University
Bloomington, Indiana