DAVID LEWIS

COUNTERFACTUALS AND COMPARATIVE POSSIBILITY*

In the last dozen years or so, our understanding of modality has been much improved by means of possible-world semantics: the project of analyzing modal language by systematically specifying the conditions under which a modal sentence is true at a possible world. I hope to do the same for counterfactual conditionals. I write $A \Box\!\!\rightarrow C$ for the counterfactual conditional with antecedent A and consequent C. It may be read as 'If it were the case that A, then it would be the case that C' or some more idiomatic paraphrase thereof.

1. Analyses

I shall lead up by steps to an analysis I believe to be satisfactory.

ANALYSIS 0. *$A \Box\!\!\rightarrow C$ is true at world i iff C holds at every A-world such that* —. 'A-world', of course, means 'world where A holds'.

The blank is to be filled in with some sort of condition restricting the A-worlds to be considered. The condition may depend on i but not on A. For instance, we might consider only those A-worlds that agree with i in certain specified respects. On this analysis, the counterfactual is some fixed strict conditional.

No matter what condition we put into the blank, Analysis 0 cannot be correct. For it says that if $A \Box\!\!\rightarrow \bar{B}$ is true at i, $\bar{B}$ holds at every A-world such that —. In other words, there are no AB-worlds such that —. Then $AB \Box\!\!\rightarrow \bar{C}$ and $AB \Box\!\!\rightarrow C$ are alike vacuously true, and $-(AB \Box\!\!\rightarrow C)$ and $-(AB \Box\!\!\rightarrow \bar{C})$ are alike false, for any C whatever. On the contrary: it can perfectly well happen that $A \Box\!\!\rightarrow \bar{B}$ is true, yet $AB \Box\!\!\rightarrow \bar{C}$ is non-vacuous, and $AB \Box\!\!\rightarrow C$ is false. In fact, we can have an arbitrarily long sequence like this of non-vacuously true counterfactuals and true denials of their opposites:

$$A \Box\!\!\rightarrow \bar{B} \quad \text{and} \quad -(A \Box\!\!\rightarrow B),$$
$$AB \Box\!\!\rightarrow \bar{C} \quad \text{and} \quad -(AB \Box\!\!\rightarrow C),$$

Journal of Philosophical Logic 2 (1973) 418–446. *All Rights Reserved*

$ABC \Box\!\!\rightarrow \bar{D}$ and $-(ABC \Box\!\!\rightarrow D)$,
etc.

Example: if Albert had come to the party, he would not have brought Betty; for, as he knows, if he had come and had brought Betty, Carl would not have stayed; for, as Carl knows, if Albert had come and had brought Betty and Carl had stayed, Daisy would not have danced with him; ... Each step of the sequence is a counterexample to Analysis 0. The counterfactual is not any strict conditional whatever.

Analysis 0 also says that $A \Box\!\!\rightarrow C$ implies $AB \Box\!\!\rightarrow C$. If C holds at every A-world such that —, then C holds at such of those worlds as are B-worlds. On the contrary: we can have an arbitrarily long sequence like this of non-vacuously true counterfactuals and true denials of their opposites:

$A \Box\!\!\rightarrow Z$ and $-(A \Box\!\!\rightarrow Z)$,
$AB \Box\!\!\rightarrow Z$ and $-(AB \Box\!\!\rightarrow Z)$,
$ABC \Box\!\!\rightarrow Z$ and $-(ABC \Box\!\!\rightarrow Z)$,
etc.

Example: if I had shirked my duty, no harm would have ensued; but if I had and you had too, harm would have ensued; but if I had and you had too and a third person had done far more than his duty, no harm would have ensued... For this reason also the counterfactual is not any strict conditional whatever.

More precisely, it is not any one, fixed strict conditional. But this much of Analysis 0 is correct: (1) to assess the truth of a counterfactual we must consider whether the consequent holds at certain antecedent-worlds; (2) we should not consider all antecedent-worlds, but only some of them. We may ignore antecedent-worlds that are gratuitously remote from actuality.

Rather than any fixed strict conditional, we need a *variably strict conditional*. Given a far-fetched antecedent, we look perforce at antecedent-worlds remote from actuality. There are no others to look at. But given a less far-fetched antecedent, we can afford to be more fastidious and ignore the very same worlds. In considering the supposition 'if I had just let go of my pen...' I will go wrong if I consider bizarre worlds where the law of gravity is otherwise than it actually is; whereas in considering the supposition 'if the planets traveled in spirals...' I will go just as wrong if I ignore such worlds.

It is this variable strictness that accounts for our counter-example sequences. It may happen that we can find an A-world that meets some stringent restriction; before we can find any AB-world we must relax the restriction; before we can find any ABC-world we must relax it still more; and so on. If so a counterexample sequence of the first kind definitely will appear, and one of the second kind will appear also if there is a suitable Z.

We dream of considering a world where the antecedent holds but everything else is just as it actually is, the truth of the antecedent being the one difference between that world and ours. No hope. Differences never come singly, but in infinite multitudes. Take, if you can, a world that differs from ours *only* in that Caesar did not cross the Rubicon. Are his predicament and ambitions there just as they actually are? The regularities of his character? The psychological laws exemplified by his decision? The orders of the day in his camp? The preparation of the boats? The sound of splashing oars? Hold *everything* else fixed after making one change, and you will not have a possible world at all.

If we cannot have an antecedent-world that is otherwise just like our world, what can we have? This, perhaps: an antecedent-world that does not differ gratuitously from ours; one that differs only as much as it must to permit the antecedent to hold; one that is closer to our world in similarity, all things considered, than any other antecedent world. Here is a first analysis of the counterfactual as a variably strict conditional.

ANALYSIS 1. *$A \mathrel{\Box\!\!\rightarrow} C$ is true at i iff C holds at the closest (accessible) A-world to i, if there is one.* This is Robert Stalnaker's proposal in 'A Theory of Conditionals', *Studies in Logical Theory* (*A.P.Q.* supplementary monograph series, 1968), and elsewhere.

It may be objected that Analysis 1 is founded on comparative similarity – 'closeness' – of worlds, and that comparative similarity is hopelessly imprecise unless some definite respect of comparison has been specified. Imprecise it may be; but that is all to the good. Counterfactuals are imprecise too. Two imprecise concepts may be rigidly fastened to one another, swaying together rather than separately, and we can hope to be precise about their connection. Imprecise though comparative similarity may be, we *do* judge the comparative similarity of complicated things like cities or people or philosophies – and we do it often without benefit of any definite respect of comparison stated in advance. We balance off various similarities and dissimilarities according to the importances we attach to various respects of comparison and according to the degrees of similarity in the various respects. Conversational context, of course, greatly affects our weighting of respects of comparison, and even in a fixed context we have plenty of latitude. Still, not anything goes. We have concordant mutual expectations, mutual expectations of expectations, etc., about the relative importances we will attach to respects of comparison. Often these are definite and accurate and firm enough to resolve the imprecision of comparative similarity to the point where we can converse without misunderstanding. Such imprecision we can live with. Still, I grant that a counterfactual based on comparative similarity has no place in the language of the exact sciences.

I imposed a restriction to *A*-worlds 'accessible' from *i*. In this I follow Stalnaker, who in turn is following the common practice in modal logic. We might think that there are some worlds so very remote from *i* that they should always be ignored (at *i*) even if some of them happen to be *A*-worlds and there are no closer *A*-worlds. If so, we have the wherewithal to ignore them by deeming them *inaccessible* from *i*. I can think of no very convincing cases, but I prefer to remain neutral on the point. If we have no need for accessibility restrictions, we can easily drop them by stipulating that all worlds are mutually interaccessible.

Unfortunately, Analysis 1 depends on a thoroughly implausible assumption: that there will never be more than one closest *A*-world. So fine are the gradations of comparative similarity that despite the infinite number and variety of worlds every tie is broken.

Example: *A* is 'Bizet and Verdi are compatriots', *F* is 'Bizet and Verdi are French', *I* is 'Bizet and Verdi are Italian'. Grant for the sake of argument that we have the closest *F*-world and the closest *I*-world; that these are distinct (dual citizenships would be a gratuitous difference from actuality); and that these are the two finalists in the competition for closest *A*-world. It might be that something favors one over the other – for all I know, Verdi narrowly escaped settling in France and Bizet did not narrowly escape settling in Italy. But we can count on no such luck. The case may be perfectly balanced between respects of comparison that favor the *F*-world and respects that favor the *I*-world. It is out of the question, on

Analysis 1, to leave the tie unbroken. That means there is no such thing as *the* closest A-world. Then anything you like holds at the closest A-world if there is one, because there isn't one. If Bizet and Verdi had been compatriots they would have been Ukranian.

ANALYSIS 2. *$A \Box\!\!\rightarrow C$ is true at i iff C holds at every closest (accessible) A-world to i, if there are any.* This is the obvious revision of Stalnaker's analysis to permit a tie in comparative similarity between several equally close closest A-worlds.

Under Analysis 2 unbreakable ties are no problem. The case of Bizet and Verdi comes out as follows. $A \Box\!\!\rightarrow F$, $A \Box\!\!\rightarrow \bar{F}$, $A \Box\!\!\rightarrow I$, and $A \Box\!\!\rightarrow \bar{I}$ are all false. $A \Box\!\!\rightarrow (F \vee I)$ and $A \Box\!\!\rightarrow (\bar{F} \vee \bar{I})$ are both true. $A \Box\!\!\rightarrow FI$ and $A \Box\!\!\rightarrow \bar{F}\bar{I}$ are both false. These conclusions seem reasonable enough.

This reasonable settlement, however, does not sound so good in words. $A \Box\!\!\rightarrow F$ and $A \Box\!\!\rightarrow \bar{F}$ are both false, so we want to assert their negations. But negate their English readings in any straightforward and natural way, and we do not get $-(A \Box\!\!\rightarrow F)$ and $-(A \Box\!\!\rightarrow \bar{F})$ as desired. Rather the negation moves in and attaches only to the consequent, and we get sentences that seem to mean $A \Box\!\!\rightarrow \bar{F}$ and $A \Box\!\!\rightarrow \bar{\bar{F}}$ – a pair of falsehoods, together implying the further falsehood that Bizet and Verdi could not have been compatriots; and exactly the opposite of what we meant to say.

Why is it so hard to negate a whole counterfactual, as opposed to negating the consequent? The defender of Analysis 1 is ready with an explanation. Except when A is impossible, he says, there is a unique closest A-world. Either C is false there, making $-(A \Box\!\!\rightarrow C)$ and $A \Box\!\!\rightarrow \bar{C}$ alike true, or C is true there, making them alike false. Either way, the two agree. We have no need of a way to say $-(A \Box\!\!\rightarrow C)$ because we might as well say $A \Box\!\!\rightarrow \bar{C}$ instead (except when A is impossible, in which case we have no need of a way to say $-(A \Box\!\!\rightarrow C)$ because it is false).

There is some appeal to the view that $-(A \Box\!\!\rightarrow C)$ and $A \Box\!\!\rightarrow \bar{C}$ are equivalent (except when A is impossible) and we might be tempted thereby to return to Analysis 1. We might do better to return only part way, using Bas van Fraassen's method of supervaluations to construct a compromise between Analyses 1 and 2.

ANALYSIS $1\frac{1}{2}$. *$A \Box\!\!\rightarrow C$ is true at i iff C holds at a certain arbitrarily chosen one of the closest (accessible) A-worlds to i, if there are any. A sen-*

tence is super-true iff it is true no matter how the arbitrary choices are made, super-false iff false no matter how the arbitrary choices are made. Otherwise it has no super-truth value. Unless a particular arbitrary choice is under discussion, we abbreviate 'super-true' as 'true', and so on. Something of this kind is mentioned at the end of Richmond Thomason, 'A Fitch-Style Formulation of Conditional Logic', *Logique et Analyse* 1970.

Analysis $1\frac{1}{2}$ agrees with Analysis 1 about the equivalence (except when A is impossible) of $-(A \Box\!\!\rightarrow C)$ and $A \Box\!\!\rightarrow \bar{C}$. If there are accessible A-worlds, the two agree in truth (i.e. super-truth) value, and further their biconditional is (super-)true. On the other hand, Analysis $1\frac{1}{2}$ tolerates ties in comparative similarity as happily as Analysis 2. Indeed a counterfactual is (super-)true under Analysis $1\frac{1}{2}$ iff it is true under Analysis 2. On the other hand, a counterfactual false under Analysis 2 may either be false or have no (super-)truth under Analysis $1\frac{1}{2}$. The case of Bizet and -Verdi comes out as follows: $A \Box\!\!\rightarrow F$, $A \Box\!\!\rightarrow \bar{F}$, $A \Box\!\!\rightarrow I$, $A \Box\!\!\rightarrow \bar{I}$, and their negations have no truth value. $A \Box\!\!\rightarrow (F \vee I)$ and $A \Box\!\!\rightarrow (\bar{F} \vee \bar{I})$ are (super-)true. $A \Box\!\!\rightarrow FI$ and $A \Box\!\!\rightarrow \bar{F}\bar{I}$ are (super-)false.

This seems good enough. For all I have said yet, Analysis $1\frac{1}{2}$ solves the problem of ties as well as Analysis 2, provided we're not too averse to (super-) truth value gaps. But now look again at the question how to deny a counterfactual. We have a way after all: to deny a 'would' counterfactual, use a 'might' counterfactual with the same antecedent and negated consequent. In reverse likewise: to deny a 'might' counterfactual, use a 'would' counterfactual with the same antecedent and negated consequent. Writing $A \Diamond\!\!\rightarrow C$ for 'If it were the case that A, then it might be the case that C' or some more idiomatic paraphrase, we have these valid-sounding equivalences:

(1) $-(A \Box\!\!\rightarrow C)$ is equivalent to $A \Diamond\!\!\rightarrow \bar{C}$,

(2) $-(A \Diamond\!\!\rightarrow C)$ is equivalent to $A \Box\!\!\rightarrow \bar{C}$.

The two equivalences yield an explicit definition of 'might' from 'would' counterfactuals:

$$A \Diamond\!\!\rightarrow C =^{\mathrm{df}} -(A \Box\!\!\rightarrow \bar{C});$$

or, if we prefer, the dual definition of 'would' from 'might'. According to this definition and Analysis 2, $A \Diamond\!\!\rightarrow C$ is true at i iff C holds at some closest (accessible) A-world to i. In the case of Bizet and Verdi, $A \Diamond\!\!\rightarrow F$,

$A \Diamond\!\!\rightarrow F$, $A \Diamond\!\!\rightarrow I$, $A \Diamond\!\!\rightarrow \bar{I}$ are all true; so are $A \Diamond\!\!\rightarrow (F \vee I)$ and $A \Diamond\!\!\rightarrow (F \vee \bar{I})$; but $A \Diamond\!\!\rightarrow FI$ and $A \Diamond\!\!\rightarrow F\bar{I}$ are false.

According to the definition and Analysis 1 or $1\frac{1}{2}$, on the other hand, $A \Diamond\!\!\rightarrow C$ and $A \Box\!\!\rightarrow C$ are equivalent except when A is impossible. That should put the defender of those analyses in an uncomfortable spot. He cannot very well claim that 'would' and 'might' counterfactuals do not differ except when the antecedent is impossible. He must therefore reject my definition of the 'might' counterfactual; and with it, the equivalences (1) and (2), uncontroversial though they sound. He then owes us some other account of the 'might' counterfactual, which I do not think he can easily find. Finally, once we see that we do have a way to negate a whole counterfactual, we no longer appreciate his explanation of why we don't need one. I conclude that he would be better off moving at least to Analysis 2.

Unfortunately, Analysis 2 is not yet satisfactory. Like Analysis 1, it depends on an implausible assumption. Given that some A-world is accessible from i, we no longer assume that there must be *exactly* one closest A-world to i; but we still assume that there must be *at least* one. I call this the *Limit Assumption*. It is the assumption that as we proceed to closer and closer A-worlds we eventually hit a limit and can go no farther. But why couldn't it happen that there are closer and closer A-worlds without end – for each one, another even closer to i? Example: A is 'I am over 7 feet tall'. If there are closest A-worlds to ours, pick one of them: how tall am I there? I must be $7+\varepsilon$ feet tall, for some positive ε, else it would not be an A-world. But there are A-worlds where I am only $7+\varepsilon/2$ feet tall. Since that is closer to my actual height, why isn't one of these worlds closer to ours than the purportedly closest A-world where I am $7+\varepsilon$ feet tall? And why isn't a suitable world where I am only $7+\varepsilon/4$ feet even closer to ours, and so ad infinitum? (In special cases, but not in general, there may be a good reason why not. Perhaps $7+\varepsilon$ could have been produced by a difference in one gene, whereas any height below that but still above 7 would have taken differences in many genes.) If there are A-worlds closer and closer to i without end, then any consequent you like holds at every closest A-world to i, because there aren't any. If I were over 7 feet tall I would bump my head on the sky.

ANALYSIS 3. *$A \Box\!\!\rightarrow C$ is true at i iff some (accessible) AC-world is closer* *to i than any $A\bar{C}$-world, if there are any (accessible) A-worlds.* This is my final analysis.

Analysis 3 looks different from Analysis 1 or 2, but it is similar in principle. Whenever there are closest (accessible) A-worlds to a given world, Analyses 2 and 3 agree on the truth value there of $A \mathbin{\Box\!\!\rightarrow} C$. They agree also, of course, when there are no (accessible) A-worlds. When there are closer and closer A-worlds without end, $A \mathbin{\Box\!\!\rightarrow} C$ is true iff, as we proceed to closer and closer A-worlds, we eventually leave all the $A\bar{C}$-worlds behind and find only AC-worlds.

Using the definition of $A \mathbin{\Diamond\!\!\rightarrow} C$ as $-(A \mathbin{\Box\!\!\rightarrow} \bar{C})$, we have this derived truth condition for the 'might' counterfactual: $A \mathbin{\Diamond\!\!\rightarrow} C$ is true at i iff for every (accessible) $A\bar{C}$-world there is some AC-world at least as close to i, and there are (accessible) A-worlds.

We have discarded two assumptions about comparative similarity in going from Analysis 1 to Analysis 3: first Stalnaker's assumption of uniqueness, then the Limit Assumption. What assumptions remain?

First, the *Ordering Assumption*: that for each world i, comparative similarity to i yields a *weak ordering* of the worlds accessible from i. That is, writing $j \leqslant_i k$ to mean that k is not closer to i than j, each $\leqslant_i$ is *connected* and *transitive*. Whenever j and k are accessible from i either $j \leqslant_i k$ or $k \leqslant_i j$; whenever $h \leqslant_i j$ and $j \leqslant_i k$, then $h \leqslant_i k$. It is convenient, if somewhat artificial, to extend the comparative similarity orderings to encompass also the inaccessible worlds, if any: we stipulate that each $\leqslant_i$ is to be a weak ordering of *all* the worlds, and that j is closer to i than k whenever j is accessible from i and k is not. (Equivalently: whenever $j \leqslant_i k$, then if k is accessible from i so is j.)

Second, the *Centering Assumption*: that each world i is accessible from itself, and closer to itself than any other world is to it.

2. Reformulations

Analysis 3 can be given several superficially different, but equivalent, reformulations.

2.1. *Comparative Possibility*

Introduce a connective $\prec$. $A \prec B$ is read as 'It is less remote from actuality that A than that B' or 'It is more possible that A than that B' and is true at a world i iff some (accessible) A-world is closer to i than is any B-world. First a pair of modalities and then the counterfactual can be defined from this new connective of comparative possibility, as follows. (Let $\bot$ be a sentential constant false at every world, or an arbitrarily chosen contradiction; later, let $\top =^{\mathrm{df}} - \bot$.)

$$\Diamond A =^{\mathrm{df}} A \prec \bot; \quad \Box A =^{\mathrm{df}} -\Diamond - A;$$

$$A \mathbin{\Box\!\!\rightarrow} C =^{\mathrm{df}} \Diamond A \supset (AC \prec A\bar{C}).$$

The modalities so defined are interpreted by means of accessibility in the usual way. $\Diamond A$ is true at i iff some A-world is accessible from i, and $\Box A$ is true at i iff A holds throughout all the worlds accessible from i. If accessibility restrictions are discarded, so that all worlds are mutually interaccessible, they became the ordinary 'logical' modalities. (We might rather have defined the two modalities and comparative possibility from the counterfactual.

$$\Box A =^{\mathrm{df}} \bar{A} \mathbin{\Box\!\!\rightarrow} \bot; \quad \Diamond A =^{\mathrm{df}} -\Box - A;$$

$$A \prec B =^{\mathrm{df}} \Diamond A \,\&((A \vee B) \mathbin{\Box\!\!\rightarrow} A\bar{B}).$$

Either order of definitions is correct according to the given truth conditions.)

Not only is comparative possibility technically convenient as a primitive; it is of philosophical interest for its own sake. It sometimes seems true to say: It is possible that A but not that B, it is possible that B but not that C, C but not D, etc. Example: A is 'I speak English', B is 'I speak German' (a language I know), C is 'I speak Finnish', D is 'A dog speaks Finnish', E is 'A stone speaks Finnish', F is 'A number speaks Finnish'. Perhaps if I say all these things, as I would like to, I am equivocating – shifting to weaker and weaker noncomparative senses of 'possible' from clause to clause. It is by no means clear that there are enough distinct senses to go around. As an alternative hypothesis, perhaps the clauses are compatible comparsions of possibility without equivocation: $A \prec B \prec C \prec D \prec E \prec F$. (Here and elsewhere, I compress conjunctions in the obvious way.)

2.2. *Cotenability*

Call B *cotenable* at i with the supposition that A iff some A-world accessible from i is closer to i than any $\bar{B}$-world, or if there are no A-worlds accessible from i. In other words: iff, at i, the supposition that A is either more possible than the falsity of B, or else impossible. Then $A \mathrel{\Box\!\!\rightarrow} C$ is true at i iff C follows from A together with auxiliary premises $B_1, \ldots$, each true at i and cotenable at i with the supposition that A.

There is less to this definition than meets the eye. A conjunction is cotenable with a supposition iff its conjuncts all are; so we need only consider the case of a single auxiliary premise B. That single premise may always be taken either as $\bar{A}$ (if A is impossible) or as $A \supset C$ (otherwise); so 'follows' may be glossed as 'follows by truth-functional logic'.

Common opinion has it that laws of nature are cotenable with any supposition unless they are downright inconsistent with it. What can we make of this? Whatever else laws may be, they are generalizations that we deem especially important. If so, then conformity to the prevailing laws of a world i should weigh heavily in the similarity of other worlds to i. Laws should therefore tend to be cotenable, unless inconsistent, with counterfactual suppositions. Yet I think this tendency may be overridden when conformity to laws carries too high a cost in differences of particular fact. Suppose, for instance, that i is a world governed (in all respects of the slightest interest to us) by deterministic laws. Let A pertain to matters of particular fact at time t; let A be false at i, and determined at all previous times to be false. There are some A-worlds where the laws of i are never violated; all of these differ from i in matters of particular fact at all times before t. (Nor can we count on the difference approaching zero as we go back in time.) There are other A-worlds exactly like i until very shortly before t when a small, local, temporary, imperceptible suspension of the laws permits A to come true. I find it highly plausible that one of the latter resembles i on balance more than any of the former.

2.3. *Degrees of Similarity*

Roughly, $A \mathrel{\Box\!\!\rightarrow} C$ is true at i iff either (1) there is some degree of similarity to i within which there are A-worlds and C holds at all of them, or (2) there are no A-worlds within any degree of similarity to i. To avoid the questionable assumption that similarity of worlds admits somehow of numerical measurement, it seems best to identify each 'degree of similarity to i' with a set of worlds regarded as the set of all worlds within that degree of similarity to i. Call a set S of worlds a *sphere* around i iff every S-world is accessible from i and is closer to i than is any $\bar{S}$-world. Call a sphere *A-permitting* iff it contains some A-world. Letting spheres represent degrees of similarity, we have this reformulation: $A \mathrel{\Box\!\!\rightarrow} C$ is true at i iff $A \supset C$ holds throughout some A-permitting sphere around i, if such there be.

To review our other operators: $A \mathrel{\Diamond\!\!\rightarrow} C$ is true at i iff AC holds somewhere in every A-permitting sphere around i, and there are such. $\Box A$ is true at i iff A holds throughout every sphere around i. $\Diamond A$ is true at i iff A holds somewhere in some sphere around i. $A \prec B$ is true at i iff some sphere around i permits A but not B. Finally, B is cotenable at i with the supposition that A iff B holds throughout some A-permitting sphere around i, if such there be.

Restated in terms of spheres, the Limit Assumption says that if there is any A-permitting sphere around i, then there is a smallest one – the intersection of all A-permitting spheres is then itself an A-permitting sphere. We can therefore reformulate Analysis 2 as: $A \mathrel{\Box\!\!\rightarrow} C$ is true at i iff $A \supset C$ holds throughout the smallest A-permitting sphere around i, if such there be.

These systems of spheres may remind one of neighborhood systems in topology, but that would be a mistake. The topological concept of closeness captured by means of neighborhoods is purely local and qualitative, not comparative: adjacent vs. separated, no more. Neighborhoods do not capture comparative closeness to a point because arbitrary supersets of neighborhoods of the point are themselves neighborhoods of a point. The spheres around a world, on the other hand, are nested, wherefore they capture comparative closeness: j is closer to i than k is (according to the definition of spheres and the Ordering Assumption) iff some sphere around i includes j but excludes k.

2.4. *Higher-Order Quantification*

The formulation just given as a metalinguistic truth condition can also be stated, with the help of auxiliary apparatus, as an explicit definition in the object language.

$$A \mathrel{\Box\!\!\rightarrow} C =^{\mathrm{df}} \Diamond A \supset \exists S(\Phi S \mathbin{\&} \Diamond SA \mathbin{\&} \Box(SA \supset C)).$$

Here the modalities are as before; 'S' is an object-language variable over propositions; and Φ is a higher-order predicate satisfied at a world i by a proposition iff the set of all worlds where that proposition holds is a sphere around i. I have assumed that every set of worlds is the truth-set of some – perhaps inexpressible – proposition.

We could even quantify over modalities, these being understood as certain properties of propositions. Call a modality *spherical* iff for every world i there is a sphere around i such that the modality belongs at i to all and only those propositions that hold throughout that sphere. Letting $\blacksquare$ be a variable over all spherical modalities, and letting $\blacklozenge$ abbreviate $-\blacksquare-$, we have

$$A \mathrel{\Box\!\!\rightarrow} C =^{\mathrm{df}} \Diamond A \supset \exists\blacksquare(\blacklozenge A \mathbin{\&} \blacksquare(A \supset C)).$$

This definition captures explicitly the idea that the counterfactual is a variably strict conditional.

To speak of variable strictness, we should be able to compare the strictness of different spherical modalities. Call one modality *(locally) stricter* than another at a world i iff the second but not the first belongs to some proposition at i. Call two modalities *comparable* iff it does not happen that one is stricter at one world and the other at another. Call one modality *stricter* than another iff they are comparable and the first is stricter at some world. Call one *uniformly stricter* than another iff it is stricter at every world. Comparative strictness is only a partial ordering of the spherical modalities: some pairs are incomparable. However, we can without loss restrict the range of our variable $\blacksquare$ to a suitable subset of the spherical modalities on which comparative strictness is a linear ordering. (Perhaps – iff the inclusion orderings of spheres around worlds all have the same order type – we can do better still, and use a subset linearly ordered by uniform comparative strictness.) Unfortunately, these linear sets are not uniquely determined.

Example: suppose that comparative similarity has only a few gradations. Suppose, for instance, that there are only five different (nonempty) spheres around each world. Let $\Box_1 A$ be true at i iff A holds throughout the innermost (nonempty) sphere around i: let $\Box_2 A$ be true at i iff A holds throughout the innermost-but-one; and likewise for $\Box_3$, $\Box_4$, and $\Box_5$. Then the five spherical modalities expressed by these operators are a suitable linear set. Since we have only a finite range, we can replace quantification by disjunction:

$$A \mathbin{\Box\!\!\rightarrow} C =^{\mathrm{df}} \Diamond A \supset .(\Diamond_1 A \mathbin{\&} \Box_1(A \supset C)) \\ \vee \cdots \vee (\Diamond_5 A \mathbin{\&} \Box_5(A \supset C))$$

See Louis Goble, 'Grades of Modality', *Logique et Analyse* 1970.

2.5. *Impossible Limit-Worlds*

We were driven from Analysis 2 to Analysis 3 because we had reason to doubt the Limit Assumption. It seemed that sometimes there were closer and closer *A*-worlds to *i* without limit – that is, without any closest *A*-worlds. None, at least, among the *possible* worlds. But we can find the closest *A*-worlds instead among certain *impossible* worlds, if we are willing to look there. If we count these impossible worlds among the worlds to be considered, the Limit Assumption is rescued and we can safely return to Analysis 2.

There are various ways to introduce the impossible limits we need. The following method is simplest, but others can be made to seem a little less *ad hoc*. Suppose there are closer and closer (accessible, possible) *A*-worlds to *i* without limit; and suppose Σ is any maximal set of sentences such that, for any finite conjunction *C* of sentences in Σ, $A \mathbin{\Diamond\!\!\rightarrow} C$ holds at *i* according to Analysis 3. (We can think of such a Σ as a full description of one – possible or impossible – way things might be if it were that *A*, from the standpoint of *i*.) Then we must posit an impossible limit-world where all of Σ holds. It should be accessible from *i* alone; it should be closer to *i* than all the possible *A*-worlds; but it should be no closer to *i* than any possible world that is itself clossr than all the possible *A*-worlds. (Accessibility from, and comparative similarity to, the impossible limit-worlds is undefined. Truth of sentences there is determined by the way in which these worlds were introduced as limits, not according to the ordinary truth conditions.) Obviously the Limit Assumption is satisfied once these impossible worlds have been added to the worlds under consideration. It is easy to verify that the truth values of counterfactuals at possible worlds afterwards according to Analyses 2 and 3 alike agrees with their original truth values according to Analysis 3.

The impossible worlds just posited are impossible in the least objectionable way. The sentences true there may be *incompatible*, in that not all of them hold together at any possible world; but there is no (correct) way to derive any contradiction from them. For a derivation proceeds from finitely many premises; and any finite subset of the sentences true at one of the limit-worlds *is* true together at some possible world. Example: recall the failure of the Limit Assumption among possible worlds when *A* is 'I am over 7 feet tall'. Our limit-worlds will be impossible worlds where *A* is true but all of 'I am at least 7.1 feet tall', 'I am at least 7.01 feet tall', 'I am at least 7.001 feet tall' etc. are false. (Do not confuse these with possible worlds where I am infinitesimally more than 7 feet tall. For all I know, there are such; but worlds where physical magnitudes can take 'non-standard' values differing infinitesimally from a real number presumably differ from ours in a very fundamental way, making them far more remote from actuality than some of the standard worlds where I am, say, 7.1 feet tall. If so, 'Physical magnitudes never take non-standard values' is false at any possible world where I am infinitesimally more than 7 feet tall, but true at the impossible closest *A*-worlds to ours.)

How bad is it to believe in these impossible limit-worlds? Very bad, I think; but there is no reason not to reduce them to something less objectionable, such as sets of propositions or even sentences. I do not like a parallel reduction of possible worlds, chiefly because it is incredible in the case of the possible world *we* happen to live in, and other possible worlds do not differ in kind from ours. But this objection does not carry over to the impossible worlds. We do not live in one of those, and possible and impossible worlds do differ in kind.

2.6. *Selection Functions*

Analysis 2, vindicated either by trafficking in impossible worlds or by faith in the Limit Assumption even for possible worlds, may conveniently be reformulated by introducing a function f that selects, for any antecedent A and possible world i, the set of all closest (accessible) A-worlds to i (the empty set if there are none). $A \mathrel{\Box\!\!\rightarrow} C$ is true at a possible world i iff C holds throughout the selected set $f(A, i)$. Stalnaker formulates Analysis 1 this way, except that his $f(A, i)$ is the unique member of the selected set, if such there be, instead of the set itself.

If we like, we can put the selection function into the object language; but to do this without forgetting that counterfactuals are in general contingent, we must have recourse to *double indexing*. That is, we must think of some special sentences as being true or false at a world i not absolutely, but in relation to a world j. An ordinary sentence is true or false at i, as the case may be, in relation to any j; it will be enough to deal with ordinary counterfactuals compounded out of ordinary sentences. Let $ƒA$ (where A is ordinary) be a special sentence true at j in relation to i iff j belongs to $f(A, i)$. Then $ƒA \supset C$ (where C is ordinary) is true at j in relation to i iff, if j belongs to $f(A, i)$, C holds at j. Then $\Box(ƒA \supset C)$ is true at j in relation to i iff C holds at every world in $f(A, i)$ that is accessible from j. It is therefore true at i in relation to i itself iff C holds throughout (fA, i) – that is, iff $A \Box\!\!\rightarrow C$ holds at i. Introducing an operator † such that $\dagger B$ is true at i in relation to j iff B is true at i in relation to i itself, we can define the counterfactual:

$$A \Box\!\!\rightarrow C =^{\mathrm{df}} \dagger\Box(ƒA \supset C).$$

An $ƒ$-operator without double indexing is discussed in Lennart Åqvist, 'Modal Logic with Subjunctive Conditionals and Dispositional Predicates', *Filosofiska Studier* (Uppsala) 1971; the †-operator was introduced in Frank Vlach, ' "Now" and "Then" '(in preparation).

2.7. *Ternary Accessibility*

If we like, we can reparse counterfactuals as $[A \Box\!\!\rightarrow]C$, regarding $\Box\!\!\rightarrow$ now not as a two-place operator but rather as taking one sentence A to make a one-place operator $[A \Box\!\!\rightarrow]$. If we have closest A-worlds – possible or impossible – whenever A is possible, then each $[A \Box\!\!\rightarrow]$ is a necessity operator interpretable in the normal way by means of an accessibility relation. Call j *A-accessible* from i (or *accessible from i relative to A*) iff j is a closest (accessible) A-world from i; then $[A \Box\!\!\rightarrow]C$ is true at i iff C holds at every world A-accessible from i. See Brian F. Chellas, 'Basic Conditional Logic' (in preparation).

3. Fallacies

Some familiar argument-forms, valid for certain other conditionals, are invalid for my counterfactuals.

Transitivity	Contraposition	Strengthening	Importation
$A \Box\!\!\rightarrow B$			
$B \Box\!\!\rightarrow C$	$A \Box\!\!\rightarrow C$	$A \Box\!\!\rightarrow C$	$A \Box\!\!\rightarrow (B \supset C)$
$A \Box\!\!\rightarrow C$	$\bar{C} \Box\!\!\rightarrow \bar{A}$	$AB \Box\!\!\rightarrow C$	$AB \Box\!\!\rightarrow C$

However, there are related valid argument-forms that may often serve as substitutes for these.

$$\begin{array}{l} A \mathbin{\Box\!\!\rightarrow} B \\ AB \mathbin{\Box\!\!\rightarrow} C \\ \hline A \mathbin{\Box\!\!\rightarrow} C \end{array} \qquad \begin{array}{l} \bar{C} \\ A \mathbin{\Box\!\!\rightarrow} C \\ \hline \bar{C} \mathbin{\Box\!\!\rightarrow} \bar{A} \end{array} \qquad \begin{array}{l} A \mathbin{\Diamond\!\!\rightarrow} B \\ A \mathbin{\Box\!\!\rightarrow} C \\ \hline AB \mathbin{\Box\!\!\rightarrow} C \end{array} \qquad \begin{array}{l} A \mathbin{\Diamond\!\!\rightarrow} B \\ A \mathbin{\Box\!\!\rightarrow} (B \supset C) \\ \hline AB \mathbin{\Box\!\!\rightarrow} C \end{array}$$

Further valid substitutes for transitivity are these.

$$\begin{array}{l} \\ A \mathbin{\Box\!\!\rightarrow} B \\ \Box (B \supset C) \\ \hline A \mathbin{\Box\!\!\rightarrow} C \end{array} \qquad \begin{array}{l} B \mathbin{\Box\!\!\rightarrow} A \\ A \mathbin{\Box\!\!\rightarrow} B \\ B \mathbin{\Box\!\!\rightarrow} C \\ \hline A \mathbin{\Box\!\!\rightarrow} C \end{array} \qquad \begin{array}{l} B \mathbin{\Diamond\!\!\rightarrow} A \\ A \mathbin{\Box\!\!\rightarrow} B \\ B \mathbin{\Box\!\!\rightarrow} C \\ \hline A \mathbin{\Box\!\!\rightarrow} C \end{array}$$

4. True antecedents

On my analysis, a counterfactual is so called because it is suitable for non-trivial use when the antecedent is presumed false; not because it implies the falsity of the antecedent. It is conversationally inappropriate, of course, to use the counterfactual construction unless one supposes the antecedent false; but this defect is not a matter of truth conditions. Rather, it turns out that a counterfactual with a true antecedent is true iff the consequent is true, as if it were a material conditional. In other words, these two arguments are valid.

$$(-)\,\frac{A,\quad \bar{C}}{-(A \mathbin{\Box\!\!\rightarrow} C)} \qquad (+)\,\frac{A,\ C}{A \mathbin{\Box\!\!\rightarrow} C}.$$

It is hard to study the truth conditions of counterfactuals with true antecedents. Their inappropriateness eclipses the question whether they are true. However, suppose that someone has unwittingly asserted a counterfactual $A \mathbin{\Box\!\!\rightarrow} C$ with (what you take to be) a true antecedent A. Either of these replies would, I think, sound cogent.

(−) Wrong, since in fact A and yet not C.

(+) Right, since in fact A and indeed C.

The two replies depend for their cogency – for the appropriateness of the word 'since' – on the validity of the corresponding arguments.

I confess that the case for (−) seems more compelling than the case for (+). One who wants to invalidate (+) while keeping (−) can do so if he is prepared to imagine that another world may sometimes be just as similar to a given world as that world is to itself. He thereby weakens the Centering Assumption to this: each world is self-accessible, and at least as close to itself as any other world is to it. Making that change and keeping everything else the same, $(-)$ is valid but $(+)$ is not.

5. Counterpossibles

If A is impossible, $A \mathrel{\Box\!\!\rightarrow} C$ is vacuously true regardless of the consequent C. Clearly some counterfactuals with impossible antecedents are asserted with confidence, and should therefore come out true: 'If there were a decision procedure for logic, there would be one for the halting problem'. Others are not asserted by reason of the irrelevance of antecedent to consequent: 'If there were a decision procedure for logic, there would be a sixth regular solid' or '... the war would be over by now'. But would these be confidently *denied*? I think not; so I am content to let all of them alike be true. Relevance is welcome in the theory of conversation (which I leave to others) but not in the theory of truth conditions.

If you do insist on making discriminations of truth value among counterfactuals with impossible antecedents, you might try to do this by extending the comparative similarity orderings of possible worlds to encompass also certain impossible worlds where not-too-blatantly impossible antecedents come true. (These are worse than the impossible limit-worlds already considered, where impossible but consistent infinite combinations of possibly true sentences come true.) See recent work on impossible-world semantics for doxastic logic and for relevant implication; especially Richard Routley, 'Ultra-Modal Propositional Functors' (in preparation).

6. Potentialities

'Had the Emperor not crossed the Rubicon, he would never have become Emperor' does *not* mean that the closest worlds to ours where there is a unique emperor and he did not cross the Rubicon are worlds where there is a unique emperor and he never became Emperor. Rather, it is *de re* with respect to 'the Emperor', and means that he who actually is (or was at the time under discussion) Emperor has a counterfactual property, or *potentiality*, expressed by the formula: 'if x had not crossed the Rubicon, x

would never have become Emperor'. We speak of what would have befallen the actual Emperor, not of what would have befallen whoever would have been Emperor. Such potentialities may also appear when we quantify into counterfactuals: 'Any Emperor who would never have become Emperor had he not crossed the Rubicon ends up wishing he hadn't done it' or 'Any of these matches would light if it were scratched'. We need to know what it is for something to have a potentiality – that is, to satisfy a counterfactual formula $A(x) \Box\!\!\rightarrow C(x)$.

As a first approximation, we might say that something x satisfies the formula $A(x) \Box\!\!\rightarrow C(x)$ at a world i iff some (accessible) world where x satisfies $A(x)$ and $C(x)$ is closer to i than any world where x satisfies $A(x)$ and $\bar{C}(x)$, if there are (accessible) worlds where x satisfies $A(x)$.

The trouble is that this depends on the assumption that one and the same thing can exist – can be available to satisfy formulas – at various worlds. I reject this assumption, except in the case of certain abstract entities that inhabit no particular world, and think it better to say that concrete things are confined each to its own single world. He who actually is Emperor belongs to our world alone, and is not available to cross the Rubicon or not, become Emperor or not, or do anything else at any other world. But although he himself is not present elsewhere, he may have *counterparts* elsewhere: inhabitants of other worlds who resemble him closely, and more closely than do the other inhabitants of the same world. What he cannot do in person at other worlds he may do vicariously, through his counterparts there. So, for instance, I might have been a Republican not because I myself am a Republican at some other world than this – I am not – but because I have Republican counterparts at some worlds. See my 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* 1968.

Using the method of counterparts, we may say that something x satisfies the formula $A(x) \Box\!\!\rightarrow C(x)$ at a world i iff some (accessible) world where some counterpart of x satisfies $A(x)$ and $C(x)$ is closer to i than any world where any counterpart of x satisfies $A(x)$ and $\bar{C}(x)$, if there are (accessible) worlds where a counterpart of x satisfies $A(x)$. This works also for abstract entities that inhabit no particular world but exist equally at all, if we say that for these things the counterpart relation is simply identity.

A complication: it seems that when we deal with relations expressed by counterfactual formulas with more than one free variable, we may need to mix different counterpart relations. 'It I were you I'd give up' seems to mean that some world where a character-counterpart of me is a predicament-counterpart of you and gives up is closer than any where a character-counterpart of me is a predicament-counterpart of you and does not give up. (I omit provision for vacuity and for accessibility restrictions.) The difference between Goodman's sentences

(1) If New York City were in Georgia, New York City would be in the South.

(2) If Georgia included New York City, Georgia would not be entirely in the South.

may be explained by the hypothesis that both are *de re* with respect to both 'New York City' and 'Georgia', and that a less stringent counterpart relation is used for the subject terms 'New York City' in (1) and 'Georgia' in (2) than for the object terms 'Georgia' in (1) and 'New York City' in (2). I cannot say in general how grammar and context control which counterpart relation is used where.

An independent complication: since closeness of worlds and counterpart relations among their inhabitants are alike matters of comparative similarity, the two are interdependent. At a world close to ours, the inhabitants of our world will mostly have close counterparts; at a world very different from ours, nothing can be a very close counterpart of anything at our world. We might therefore wish to fuse closeness of worlds and closeness of counterparts, allowing these to balance off. Working with comparative similarity among *pairs* of a concrete thing and the world it inhabits (and ignoring provision for vacuity and for accessibility restrictions), we could say that an inhabitant x of a world i satisfies $A(x) \mathrel{\Box\!\!\rightarrow} C(x)$ at i iff some such thing-world pair $\langle y, j\rangle$ such that y satisfies $A(x)$ and $C(x)$ at j is more similar to the pair $\langle x, i\rangle$ than is any pair $\langle z, k\rangle$ such that z satisfies $A(x)$ and $\bar{C}(x)$ at k. To combine this complication and the previous one seems laborious but routine.

7. Countercomparatives

'If my yacht were longer than it is, I would be happier than I am' might be handled by quantifying into a counterfactual formula: $\exists x, y$ (my yacht is

x feet long & I enjoy y hedons & (my yacht is more than x feet long $\Box\!\!\rightarrow$ I enjoy more than y hedons)). But sometimes, perhaps in this very example, comparison makes sense when numerical measurement does not. An alternative treatment of countercomparatives is available using double indexing. (Double indexing has already been mentioned in connection with the †-operator; but if we wanted it both for that purpose and for this, we would need triple indexing.) Let A be true at j in relation to i iff my yacht is longer at j than at i (more precisely: if my counterpart at j has a longer yacht than my counterpart at i (to be still more precise, decide what to do when there are multiple counterparts or multiple yachts)); let C be true at j in relation to i iff I am happier at j than at i (more precisely: if my counterpart...). Then $A \Box\!\!\rightarrow C$ is true at j in relation to i iff some world (accessible from j) where A and C both hold in relation to i is closer to j than any world where A and $\bar{C}$ both hold in relation to i. So far, the relativity to i just tags along. Our countercomparative is therefore true at i (in relation to any world) iff $A \Box\!\!\rightarrow C$ is true at i in relation to i itself. It is therefore $\dagger(A \Box\!\!\rightarrow C)$.

8. Counterfactual probability

'The probability that C, if it were the case that A, would be r' cannot be understood to mean any of:

(1) $\text{Prob}\,(A \Box\!\!\rightarrow C) = r$,

(2) $\text{Prob}\,(C \mid A) = r$, or

(3) $A \Box\!\!\rightarrow \text{Prob}(C) = r$.

Rather, it is true at a world i (with respect to a given probability measure) iff for any positive ε there exists an A-permitting sphere T around i such that for any A-permitting sphere S around i within T, $\text{Prob}(C \mid AS)$, unless undefined, is within ε of r.

Example. A is 'The sample contained abracadabrene', C is 'The test for abracadabrene was positive', Prob is my present subjective probability measure after watching the test come out negative and tentatively concluding that abracadabrene was absent. I consider that the probability of a positive result, had abracadabrene been present, would have been 97%. (1) I know that false negatives occur because of the inherently indeterministic character of the radioactive decay of the tracer used in the test, so I am convinced that no matter what the actual conditions were, there might have been a false negative even if abracadabrene had been present. $\mathrm{Prob}(A \Diamond\!\!\rightarrow \bar{C}) \approx 1$; $\mathrm{Prob}(A \Box\!\!\rightarrow C) \approx 0$. (2) Having seen that the test was negative, I disbelieve C much more strongly than I disbelieve A; $\mathrm{Prob}(AC)$ is much less than $\mathrm{Prob}(A)$; $\mathrm{Prob}\ (C \mid A) \approx 0$. (3) Unknown to me, the sample was from my own blood, and abracadabrene is a powerful hallucinogen that makes white things look purple. Positive tests are white, negatives are purple. So had abracadabrene been present, I would have strongly disbelieved C no matter what the outcome of the test really was. $A \Box\!\!\rightarrow \mathrm{Prob}(C) \approx 0$. (Taking (3) *de re* with respect to 'Prob' is just as bad: since actually $\mathrm{Prob}(C) \approx 0$, $A \Box\!\!\rightarrow \mathrm{Prob}(C) \approx 0$ also.) My suggested definition seems to work, however, provided that the outcome of the test at a close A-world does not influence the closeness of that world to ours.

9. Analogies

The counterfactual as I have analyzed it is parallel in its semantics to operators in other branches of intensional logic, based on other comparative relations. There is one difference: in the case of these analogous operators, it seems best to omit the provision for vacuous truth. They correspond to a doctored counterfactual $\Box\!\!\Rightarrow$ that is automatically false instead of automatically true when the antecedent is impossible: $A \Box\!\!\Rightarrow C =^{\mathrm{df}} \Diamond A \ \&\ (A \Box\!\!\rightarrow C)$.

Deontic: We have the operator $A \Box\!\!\Rightarrow_d C$, read as 'Given that A, it ought to be that C', true at a world i iff some AC-world evaluable from the standpoint of i is better, from the standpoint of i, than any $A\bar{C}$-world. Roughly (under a Limit Assumption), iff C holds at the best A-worlds. See the operator of 'conditional obligation' discussed in Bengt Hansson, 'An Analysis of Some Deontic Logics', *Noûs* 1969.

Temporal: We have $A \Box\!\!\Rightarrow_f C$, read as 'When next A, it will be that C', true at a time t iff some AC-time after t comes sooner after t than any $A\bar{C}$-time; roughly, iff C holds at the next A-time. We have also the past mirror image: $A \Box\!\!\Rightarrow_p C$, read as 'When last A, it was that C'.

Egocentric (in the sense of A. N. Prior, 'Egocentric Logic', *Noûs* 1968): We have $A \Box\!\!\Rightarrow_e C$, read as 'The A is C', true for a thing x iff some AC-thing in x's ken is more salient to x than any $A\bar{C}$-thing; roughly, iff the most salient A-thing is C.

To motivate the given truth conditions, we may note that these operators all permit sequences of truths of the two forms:

$$\begin{array}{lcl} A \mathbin{\Box\!\!\Rightarrow} \bar{B}, & & A \mathbin{\Box\!\!\Rightarrow} Z, \\ AB \mathbin{\Box\!\!\Rightarrow} \bar{C}, & \text{and} & AB \mathbin{\Box\!\!\Rightarrow} Z, \\ ABC \mathbin{\Box\!\!\Rightarrow} \bar{D}, & & ABC \mathbin{\Box\!\!\Rightarrow} Z, \\ \text{etc.}; & & \text{etc.} \end{array}$$

It is such sequences that led us to treat the counterfactual as a variably strict conditional. The analogous operators here are likewise variably strict conditionals. Each is based on a binary relation and a family of comparative relations in just the way that the (doctored) counterfactual is based on accessibility and the family of comparative similarity orderings. In each case, the Ordering Assumption holds. The Centering Assumption, however, holds only in the counterfactual case. New assumptions hold in some of the other cases.

In the deontic case, we may or may not have different comparative orderings from the standpoint of different worlds. If we evaluate worlds according to their conformity to the edicts of the god who reigns at a given world, then we will get different orderings; and no worlds will be evaluable from the standpoint of a godless world. If rather we evaluate worlds according to their total yield of hedons, then evaluability and comparative goodness of worlds will be absolute.

In the temporal case, both the binary relation and the families of comparative relations, both for 'when next' and for 'when last', are based on the single underlying linear order of time.

The sentence $(A \vee \bar{B}) \mathbin{\Box\!\!\Rightarrow}_f AB$ is true at time *t* iff some *A*-time after t precedes any $\bar{B}$-time after *t*. It thus approximates the sentence 'Until *A*, *B*', understood as being true at *t* iff some *A*-time after *t* is not preceded by any $\bar{B}$-time after *t*. Likewise $(A \vee \bar{B}) \mathbin{\Box\!\!\Rightarrow}_p AB$ approximates 'Since *A*, *B*', with 'since' understood as the past mirror image of 'until'. Hans Kamp has shown that 'since' and 'until' suffice to define all possible tense operators, provided that the order of time is a complete linear order; see his *Tense Logic and the Theory of Order* (U.C.L.A. dissertation, 1968). Do my approximations have the same power? No; consider 'Until $\top$, $\bot$', true at *t* iff there is a next moment after *t*. This sentence cannot be translated using my operators. For if the order of time is a complete linear order with discrete stretches and dense stretches, then the given sentence will vary in truth value; but if in addition there is no beginning or end of time, and if there are no atomic sentences that vary in truth value, then no sentences that vary in truth value can be built up by means of truth-functional connectives, $\Box\!\!\Rightarrow_f$, and $\Box\!\!\Rightarrow_p$.

Starting from any of our various $\Box\!\!\Rightarrow$-operators, we can introduce one-place operators I shall call the *inner modalities*:

$$\boxdot A =^{\mathrm{df}} \mathsf{T} \Box\!\!\Rightarrow A,$$
$$\Diamond A \;=^{\mathrm{df}}\; -\boxdot - A,$$

and likewise in the analogous cases. The inner modalities in the counterfactual case are of no interest (unless Centering is weakened), since $\boxdot A$ and $\Diamond A$ are both equivalent to A itself. Nor are they anything noteworthy in the egocentric case. In the deontic case, however, they turn out to be slightly improved versions of the usual so-called obligation and permission operators. $\boxdot_d A$ is true at i iff some (evaluable) A-world is better, from the standpoint of i, than any $\bar{A}$-world; that is, iff either (1) there are best (evaluable) worlds, and A holds throughout them, or (2) there are better and better (evaluable) worlds without end, and A holds throughout all sufficiently good ones. In the temporal case, $\boxdot_f A$ is true at t iff some A-time after t comes sooner than any $\bar{A}$-time; that is, iff either (1) there is a next moment, and A holds then, or (2) there is no next moment, and A holds throughout some interval beginning immediately and extending into the future. $\boxdot_f A$ may thus be read 'Immediately, A'; as may $\Diamond_f A$, but in a somewhat different sense.

If no worlds are evaluable from the standpoint of a given world – say, because no god reigns there – it turns out that $\boxdot_d A$ is false and $\Diamond_d A$ is true for any A whatever. Nothing is obligatory, everything is permitted. Similarly for $\boxdot_f A$ and $\Diamond_f A$ at the end of time, if such there be; and for $\boxdot_p A$ and $\Diamond_p A$ at its beginning. Modalities that behave in this way are called *abnormal*, and it is interesting to find these moderately natural examples of abnormality.

10. Axiomatics

The set of all sentences valid under my analysis may be axiomatised taking the counterfactual connective as primitive. One such axiom system – not the neatest – is the system **C1** of my paper 'Completeness and Decidability of Three Logics of Counterfactual Conditionals', *Theoria* 1971, essentially as follows.

Rules:

If A and $A \supset B$ are theorems, so is B.
If $(B_1 \;\&\; \cdots) \supset C$ is a theorem, so is

$$((A \Box\!\!\rightarrow B_1) \;\&\; \cdots) \supset (A \Box\!\!\rightarrow C).$$

Axioms:

All truth-functional tautologies are axioms.

$$A \Box\!\!\rightarrow A$$

$$(A \Box\!\!\rightarrow B) \;\&\; (B \Box\!\!\rightarrow A) . \supset . (A \Box\!\!\rightarrow C) \equiv (B \Box\!\!\rightarrow C)$$

$$((A \vee B) \Box\!\!\rightarrow A) \vee ((A \vee B) \Box\!\!\rightarrow B) \vee (((A \vee B) \Box\!\!\rightarrow C) \equiv (A \Box\!\!\rightarrow C) \;\&\; (B \Box\!\!\rightarrow C))$$

$$A \Box\!\!\rightarrow B . \supset . A \supset B$$

$$AB \supset . A \Box\!\!\rightarrow B$$

(Rules and axioms here and henceforth should be taken as schematic.) Recall that modalities and comparative possibility may be introduced via the following definitions: $\Box A =^{\mathrm{df}} \bar{A} \Box\!\!\rightarrow \bot$; $\Diamond A =^{\mathrm{df}} - \Box - A$; $A \prec B =^{\mathrm{df}} \Diamond A \;\&\; ((A \vee B) \Box\!\!\rightarrow A\bar{B})$.

A more intuitive axiom system, called **VC**, is obtained if we take comparative possibility instead of the counterfactual as primitive. Let $A \preccurlyeq B =^{\mathrm{df}} -(B \prec A)$.

Rules:

If A and $A \supset B$ are theorems, so is B.
If $A \supset B$ is a theorem, so is $B \preccurlyeq A$.

Basic Axioms:

All truth-functional tautologies are basic axioms.

$$A \preccurlyeq B \preccurlyeq C . \supset . A \preccurlyeq C$$

$$A \preccurlyeq B . \vee . B \preccurlyeq A$$

$$A \preccurlyeq (A \vee B) . \vee . B \preccurlyeq (A \vee B)$$

Axiom **C**:

$$A\bar{B} \supset . A \prec B$$

Recall that modalities and the counterfactual may be introduced via the following definitions: $\Diamond A =^{\mathrm{df}} A \prec \bot$; $\Box A =^{\mathrm{df}} -\Diamond - A$; $A \mathrel{\Box\!\!\rightarrow} C =^{\mathrm{df}} \Diamond A \supset (AC \prec A\bar{C})$.

VC and **C1** turn out to be definitionally equivalent. That is, their respective definitional extensions (via the indicated definitions) yield exactly the same theorems. It may now be verified that these theorems are exactly the ones we ought to have. Since the definitions are correct (under my truth conditions) it is sufficient to consider sentences in the primitive notation of **VC**.

In general, we may define a *model* as any quadruple $\langle I, R, \leqslant, [\![\ ]\!] \rangle$ such that

(1) I is a nonempty set (regarded as playing the role of the set of worlds);

(2) R is a binary relation over I (regarded as the accessibility relation);

(3) $\leqslant$ assigns to each i in I a weak ordering $\leqslant_i$ of I (regarded as the comparative similarity ordering of worlds from the standpoint of i) such that whenever $j \leqslant_i k$, if iRk then iRj;

(4) $[\![\ ]\!]$ assigns to each sentence A a subset $[\![A]\!]$ of I (regarded as the set of worlds where A is true);

(5) $[\![-A]\!]$ is $I - [\![A]\!]$, $[\![A \,\&\, B]\!]$ is $[\![A]\!] \cap [\![B]\!]$, and so on;

(6) $[\![A \prec B]\!]$ is $\{i \varepsilon I$: for some j in $[\![A]\!]$ such that iRj, there is no k in $[\![B]\!]$ such that $k \leqslant_i j\}$.

The *intended models*, for the counterfactual case, are those in which I, R, $\leqslant$, and $[\![\ ]\!]$ really are what we regarded them as being: the set of worlds, some reasonable accessibility relation, some reasonable family of comparative similarity orderings, and an appropriate assignment to sentences of truth sets. The Ordering Assumption has been written into the very definition of a model (clause 3) since it is common to the counterfactual case and the analogous cases as well. As for the Centering Assumption, we must impose it on the intended models as a further condition:

(C) R is reflexive on I: and $j \leqslant_i i$ only if $j = i$.

It seems impossible to impose other purely mathematical conditions on the intended models (with the possible exception of (U), discussed below). We therefore hope that **VC** yields as theorems exactly the sentences valid – true at all worlds – in all models that meet condition (C). This is the case.

VC is sound for models meeting (C); for the basic axioms are valid, and the rules preserve validity, in all models; and Axiom **C** is valid in any model meeting (C).

VC is complete for models meeting (C): for there is a certain such model in which only theorems of **VC** are valid. This model is called the *canonical model* for **VC**, and is as follows:

(1) I is the set of all maximal **VC**-consistent sets of sentences;

(2) iRj iff, for every sentence A in j, $\Diamond A$ is in i;

(3) $j \leqslant_i k$ iff there is no set Σ of sentences that overlaps j but not k, such that whenever $A \preccurlyeq B$ is in i and A is in Σ then B also is in Σ;

(4) i is in $[\![A]\!]$ iff A is in i.

In the same way, we can prove that the system consisting of the rules, the basic axioms, and *any* combination of the axioms listed below is sound and complete for models meeting the corresponding combination of conditions. Nomenclature: the system generated by the rules, the basic axioms, and the listed axioms — is called **V**—. (Note that the conditions are not independent. (C) implies (W), which implies (T), which implies (N). (S) implies (L). (A−) implies (U−). (W) and (S) together imply (C). (C) and (A−) together imply (S) by implying the stronger, trivializing condition that no world is accessible from any other. Accordingly, many combinations of the listed axioms are redundant.)

Axioms

N: $\Box \top$

T: $\Box A \supset A$

W: $AB \supset . \Diamond A \,\&\, A \preccurlyeq B$

C: $A\bar{B} \supset A \prec B$

L: (no further axiom, or some tautology)

S: $A \mathrel{\Box\!\!\rightarrow} C . \vee . A \mathrel{\Box\!\!\rightarrow} \bar{C}$

U: $\Box A \supset \Box\Box A$ and $\Diamond A \supset \Box\Diamond A$

A: $A \preccurlyeq B \supset \Box(A \preccurlyeq B)$ and $A \prec B \supset \Box(A \prec B)$.

Conditions

(N) (normality): For any i in I there is some j in i such that iRj.

(T) (total reflexivity): R is reflexive on I.

(W) (weak centering): R is reflexive on I; for any i and j in I, $i \leqslant_i j$.

(C) (centering): R is reflexive on I; and $j \leqslant_i i$ only if $j = i$.

(L) (Limit Assumption): Whenever iRj for some j in $[\![A]\!]$, $[\![A]\!]$ has at least one $\leqslant_i$-minimal element.

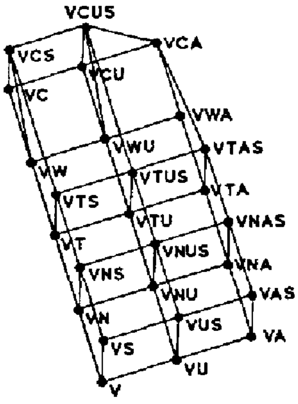
(S) (Stalnaker's Assumption): Whenever iRj for some j in $[\![A]\!]$, $[\![A]\!]$ has exactly one $\leqslant_i$-minimal element.

(U−) (local uniformity): If iRj, then jRk iff iRk.

(A−) (local absoluteness): If iRj, then jRk iff iRk and $h \leqslant_j k$ iff $h \leqslant_i k$.

The Limit Assumption (L) corresponds to no special axiom. Any one of our systems is sound and complete both for a combination of conditions without (L) and for that combination plus (L). The reason is that our canonical models always are rich enough to satisfy the Limit Assumption, but our axioms are sound without it. (Except S, for which the issue does not arise because (S) implies (L).) Moral: the Limit Assumption is irrelevant to the logical properties of the counterfactual. Had our interest been confined to logic, we might as well have stopped with Analysis 2.

Omitting redundant combinations of axioms, we have the 26 distinct systems shown in the diagram.



The general soundness and completeness result still holds if we replace the local conditions (U−) and (A−) by the stronger global conditions (U) and (A).

(U) (uniformity): For any i, j, k in I, jRk iff iRk.

(A) (absoluteness): For any h, i, j, k in I, jRk iff iRk and $h \leqslant_j k$ iff $h \leqslant_i k$.

Any model meeting (U –) or (A –) can be divided up into models meeting (U) or (A). The other listed conditions hold in the models produced by the division if they held in the original model. Therefore a sentence is valid under a combination of conditions including (U) or (A) iff it is valid under the combination that results from weakening (U) to (U –), or (A) to (A –).

In the presence of (C), (W), or (T), condition (U) is equivalent to the condition: for any i and j in I, iRj. **VCU** is thus the correct system to use if we want to drop accessibility restrictions. **VW**, or perhaps **VWU**, is the correct system for anyone who wants to invalidate the implication from A and C to $A \mathbin{\Box\!\!\rightarrow} C$ by allowing that another world might be just as close to a given world as that world is to itself. **VCS**, or **VCUS** if we drop accessibility restrictions, is the system corresponding to Analysis 1 or 1½. **VCS** is definitionally equivalent to Stalnaker's system **C2**.

The systems given by various combinations of **N**, **T**, **U**, and **A** apply, under various assumptions, to the deontic case. **VN** is definitionally equivalent to a system **CD** given by Bas van Fraassen in 'The Logic of Conditional Obligation' (forthcoming), and shown there to be sound and complete for the class of what we may call *multi-positional models* meeting (N). These differ from models in my sense in that a world may occur at more than one position in an ordering $\leqslant_i$. (Motivation: different positions may be assigned to one world *qua* realizer of different kinds of value.) Technically, we no longer have a direct ordering of the worlds themselves; rather, we have for each i in I a linear ordering of some set V_i and an assignment to each world j such that iRj of one or more members of V_i, regarded as giving the positions of j in the ordering from the standpoint of i. $A \prec B$ is true at i iff some position assigned to some A-world j (such that iRj) is better according to the given ordering than any position assigned to any B-world. My models are essentially the same as those multi-positional models in which no world does have more than one assigned position in any of the orderings. Hence **CO** is at least as strong as **VN**; but no stronger, since **VN** is already sound for all multi-positional models meeting (N).

All the systems are decidable. To decide whether a given sentence A is a theorem of a given system, it is enough to decide whether the validity of A under the corresponding combination of conditions can be refuted by a *small* countermodel – one with at most 2^n worlds, where n is the number of subsentences of A. (Take (U) and (A), rather than (U –) and (A –), as the conditions corresponding to U and A.) That can be decided by examining finitely many cases, since it is unnecessary to consider two models separately if they are isomorphic, or if they have the same I, R, $\leqslant$, and the same $[\![P]\!]$ whenever P is a sentence letter of A. If A is a theorem, then by soundness there is no countermodel and *a fortiori* no small countermodel. If A is not a theorem, then by completeness there is a countermodel $\langle I, R, \leqslant, [\![\ ]\!]\rangle$. We derive thence a small countermodel, called a *filtration* of the original countermodel, as follows. Let D_i, for each i in I, be the conjunction in some definite arbitrary order of all the subsentences of A that are true at i in the original countermodel, together with the negations of all the subsentences of A that are false at i in the original countermodel. Now let $\langle I^*, R^*, \leqslant^*, [\![\ ]\!]^*\rangle$ be as follows:

(1) I^* is a subset of I containing exactly one member of each nonempty $[\![D_i]\!]$;
(2) for any i and j in I^*, iR^*j iff i is in $[\![\Diamond D_j]\!]$;
(3) for any i, j, k in I^*, $j \leqslant^*_i k$ iff i is in $[\![D_j \preccurlyeq D_k]\!]$;
(4) for any sentence letter P, $[\![P]\!]^*$ is $[\![P]\!] \cap I^*$; for any compound sentence B, $[\![B]\!]^*$ is such that $\langle I^*, R^*, \leqslant^*, [\![\ ]\!]^*\rangle$ meets conditions (5) and (6) in the definition of a model.

Then it may easily be shown that $\langle I^*, R^*, \leqslant^*, [\![\ ]\!]^*\rangle$ is a small countermodel to the validity of A under the appropriate combination of conditions, and thereby to the theoremhood of A in the given system.

Princeton University

NOTE

* The theory presented in this paper is discussed more fully in my book *Counterfactuals* (Blackwell and Harvard University Press). My research on counterfactuals was supported by a fellowship from the American Council of Learned Societies.