PAVEL TICHÝ

# SUBJUNCTIVE CONDITIONALS: TWO PARAMETERS VS. THREE

(Received 11 January, 1983)

The last two decades have witnessed an unparalleled proliferation of rival semantic accounts of subjunctive conditionals, i.e., of statements expressed by sentences of the form

(*) If it were the case that $A$, then it would be the case that $B$.

These theories come in two substantially different sorts.

Some of them take it for granted that a sentence of the form (*) is either unambiguous or its ambiguity can be blamed on the subordinate clauses $A$ and $B$. The assumption is, in other words, that someone who is familiar with the meaning of the subjunctively-conditional connective 'If it were the case that ..., then it would be the case that ...', will fully understand (*) as soon as he understands the subordinate clauses $A$ and $B$; and that once the question of what $A$ and $B$ mean is settled, the truth-value of (*) depends on nothing but extralinguistic facts. I shall call these theories *two-parameter* ones.

Other theories reject this assumption as untenable. Their authors have noted that in countless cases a sentence of the form (*) is true if understood in one perfectly plausible way and false if understood in another, equally plausible, way, even though the subclauses $A$ and $B$ are themselves understood on both occasions in the same way. To illustrate, suppose that you are catching and actually will catch a 7 p.m. flight, that now it is 5 p.m. and that it takes an hour to drive to the airport. Which of the following two subjunctive conditionals is true: 'If it were 7 p.m. I would be missing my plane' or 'If it were 7 p.m. I would be boarding my plane'? It depends. Read in one way the former conditional is true and the latter is false. But read in another, no less plausible, way, the latter is true and the former is false. And yet the two readings do not result from reading any of the component clauses in two different ways. Thus each of the conditionals must be capable of expressing at least two different statements even if the meaning of the clauses is kept fixed. The meaning of (*) thus cannot be a function of the meaning of $A$ and

$B$ alone; a third parameter must be in the play, one which is not explicitly spelled out, but merely tacitly understood. In order to know what statement (*) expresses on a given occasion one must somehow divine from the context of utterance what value the implicit third parameter is meant to take. Theories which approach subjunctive conditionals in this way will be spoken of as *three-parameter* ones.

The present article consists of five sections. In Section 1 the main two-parameter theories are surveyed and each of them is shown materially inadequate. In Section 2 several three-parameter theories are considered and it is argued that the oldest one of them (due to Mill, Ramsey, and Chisholm), is by far superior to the others in naturalness, simplicity, and intuitive appeal. The theory is then defended against an alleged refutation in Section 3, slightly amended in Section 4, and illustrated in Section 5.[1]

## 1. TWO-PARAMETER THEORIES

The two-parameter research programme was initiated by Nelson Goodman (1947). Goodman and his followers take the view that the notion of natural law plays a central rôle in the truth-condition of subjunctive conditionals. In order to state concisely their various theories about the exact nature of that rôle, it will be convenient to use the following technical terms.

Let $A$ and $B$ be propositions and $C$ a class of propositions. We shall say that $C$ *yields* $B$ *via* $A$ if $A$, (the members of) $C$, and some true laws of nature jointly imply $B$. We shall say that $C$ *yields* $B$ (simpliciter) if $C$ yields $B$ via a tautology.

### *Goodman's Problem*

A natural two-parameter truth-condition for subjunctive conditionals may seem to be the following:

(G) (*) is true iff there exists a class $C$ of true propositions such that $C$ yields $B$ via $A$.

As an illustration of how (G) is meant to work, consider

EXAMPLE 1. Let $X$ be an object and for any time $t$ let $Wm^t, Dr^t, Ox^t, Sr^t$, and $Li^t$ be the respective propositions that $X$ is a well-made match at $t$, that

$X$ is dry at $t$, that $X$ is surrounded with oxygen at $t$, that $X$ is struck at $t$, and that $X$ lights at $t$. Assume that there is a law of nature whereby any well-made, dry, and oxygen-surrounded match lights one second after being struck. Then supposing that $Wm^T$, $Dr^T$, and $Ox^T$ are in fact true and $Sr^T$ and $Li^{T+1}$ false, the conditional

(1.1) If $X$ were struck at $T$, then $X$ would light at $T$+1

is true on (G), as intuitively desired: the class $\{Wm^T, Dr^T, Ox^T\}$ clearly yields the consequent $Li^{T+1}$ via the antecedent $Sr^T$.

Goodman (1947) considered (G) and pointed out that despite its apparent naturalness, it is indefensible as it stands. For if *any* true proposition qualifies for membership of the class $C$, too many conditionals come out true. If $A$ is false, $\{\sim A\}$ yields $B$ via $A$ for any $B$ whatsoever; (G) thus makes all subjunctive conditionals with false antecedents true. Clearly, the class $C$ should be required to be compatible with the antecedent $A$ and the laws of nature. Let us then call a class *A-compatible* if it does not yield $\sim A$; a proposition will be called *A-compatible* if its singleton is. We must require at least that the class $C$ in (G) be $A$-compatible. A little reflection reveals, however, that this is not enough. If $A$ is false then for any $A$-compatible $B$, $A \supset B$ is true, $A$-compatible, and yields $B$ via $A$. Still too many conditionals come out true.

Considerations of this sort have prompted a number of authors to distinguish between logical compounds like $A \supset B$ on the one hand, and, on the other hand, propositions which cannot be analyzed into simpler ones. On their view only propositions of the latter sort and their negations – call them *rock-bottom propositions* or RBPs for short – should be allowed entry into $C$. Let us then consider the following amendment to (G):

(G+) (∗) is true iff there exists an $A$-compatible class $C$ of true RBPs such that $C$ yields $B$ via $A$.

Some of Goodman's examples show that (G+) is still inadequate.

EXAMPLE 1 (continued). Assume that $Wm^T$, $Ox^T$, and $Li^{T+1}$ are RBPs. Then $\{Wm^T, Ox^T, \sim Li^{T+1}\}$ is an $Sr^T$-compatible class of true RBPs which yields $\sim Dr^T$ via $Sr^T$. Thus the conditional

(1.2) If $X$ were struck at $T$, then $X$ would not be dry at $T$

comes out true on (G+), contrary to what would normally be intuitively expected.

The problem is that a class of true RBPs may be $A$-compatible and yet contain propositions (such as $\sim Li^{T+1}$ in Example 1) which *would not* be true if $A$ were. Such classes should also be disqualified from playing the rôle of $C$ in the truth-condition: $C$ should be required to consist of propositions which are not just true, but would still be true if $A$ were. But this requirement cannot be incorporated in (G) without making the truth-condition circular and thereby worthless. I shall speak of this impasse as *Goodman's Problem*.

Goodman himself set the problem aside as intractable. Other authors, however, have deemed Goodman's despair premature and attempted to rule out the unwanted truths in a non-circular way. In the rest of this section I shall give a brief account of the various proposals resulting from these attempts. I shall simplify slightly by ignoring the provisions some authors make for so-called counter-legals, i.e., conditionals whose antecedents are incompatible with the laws of nature. This will have the advantage of simplifying the discussion without in any way distorting what the theories have to say about the usual, 'legal', conditionals. And we shall see that none of the theories gives an adequate account of these conditionals.

The discussion will be facilitated by a few more terminological conventions. We shall assume that each RBP reports an event or state as happening or failing to happen at a definite moment of time. Thus the usual ordering of moments of time can be carried over to the RBPs themselves: we shall say that an RBP $P$ is *earlier* (or *later*) than RBP $Q$ if $P$ reports an event or state which precedes (or follows after) that reported by $Q$. Furthermore, a class $D$ will be called an *$A$-maximal subset of* a class $E$ of propositions if $D$ is an $A$-compatible subset of $E$ and no further member of $E$ can be added to $D$ without making the result $A$-incompatible.

For future reference let us formulate and prove the following simple note.

NOTE. Let one of $P$ and $R$ be the negation of the other. If $P$ is an $A$-compatible member of $E$, then not all $A$-maximal subsets of $E$ yield $R$ via $A$.

*Proof.* Since $P$ is $A$-compatible, it is clearly a member of an $A$-maximal subset of $E$, say $M$. As $M$ yields $\sim R$, if $M$ yielded $R$ via $A$, $M$ would be $A$-incompatible, in contradiction to the definition of an $A$-maximal subset of $E$. Thus $M$ does not yield $R$ via $A$, Q.E.D.

*Pollock's Theory*

The key concept of John Pollock's theory (1981) is that of a complete causal history of an event. He defines an *historical antecedent* of a true RBP $P$ as any class of earlier true RBPs which yields $P$. Pollock allows for what he calls *ungrounded* events, i.e., true RBPs which have no historical antecedent at all. A class $C$ of true RBPs is called a *complete causal history* of $Q$ if $Q$ is an element of $C$ and $C$ contains an historical antecedent of every grounded element of $C$. A complete causal history of $Q$ thus comprises direct or indirect causes (if any) of $Q$, causes of these causes, and so on, as far into the past as one can go. Pollock now calls a true RBP $P$ *undercut by* $A$ if every complete causal history of $P$ is $A$-incompatible. For any $A$, let $|A|$ be the class of true RBPs which are not undercut by $A$. Pollock's truth-condition for subjunctive conditionals is then as follows:

(P) (∗) is true if every $A$-maximal subset of $|A|$ yields $B$ via $A$.

To see how (P) copes with Goodman's problem, consider again

EXAMPLE 1 (continued). First let us note that $Dr^T$ is not likely to be undercut by $Sr^T$: there are bound to be complete causal histories of $X$'s dryness which leave it open whether $X$ is struck or not. Thus, $Dr^T$ belongs to $|Sr^T|$. Moreover, $Dr^T$ is clearly $Sr^T$-compatible. But then, by the Note, some $Sr^T$-maximal subsets of $|Sr^T|$ do not yield $\sim Dr^T$ via $Sr^T$ and (1.2) is false on (P) as desired.

To see that (1.1), on the other hand, is true on (P) let us first note that $Wm^T$ and $Ox^T$ are not undercut by $Sr^T$ for the same reason that $Dr^T$ is not. Now let us ask: is there an $Sr^T$-compatible subset of $|Sr^T|$ which can be rendered $Sr^T$-incompatible by adding $Wm^T$, $Dr^T$, or $Ox^T$ to it? Pollock takes the view that there is not. One might think that $\{Wm^T, Ox^T, \sim Li^{T+1}\}$, for example, is such a set: it is $Sr^T$-compatible and becomes $Sr^T$-incompatible when $Dr^T$ is added to it. But Pollock assumes that $\sim Sr^T$ played an essential role in the course of events which in fact brought $\sim Li^{T+1}$ about. If this is so, every complete causal history of $\sim Li^{T+1}$ is $Sr^T$-incompatible and consequently $\sim Li^{T+1}$ is undercut by $Sr^T$. Now if $Wm^T$, $Dr^T$, and $Ox^T$ can be added to every $Sr^T$-compatible subset of $|Sr^T|$, as Pollock thinks they can, then they will be elements of every $Sr^T$-maximal subset of $|Sr^T|$ and consequently every such set will yield $Li^{T+1}$ via $Sr^T$. Thus (1.1) comes out true on (P) as desired.

To see that (P) is nevertheless inadequate, let us consider

EXAMPLE 2 (*The Bowser Test*). Imagine that healthy dogs are causally necessitated to yelp one second after their tails have been stepped on, but that they are capable of yelping gratuitously, i.e., without any antecedent cause whatsoever. For any $t$, let $Dg^t$, $Hl^t$, $Yl^t$, and $Sp^t$ be the respective propositions that Bowser is a dog at $t$, that Bowser is healthy at $t$, that Bowser yelps at $t$, and that I step on Bowser's tail at $t$. Assume that as a matter of fact, $Dg^T$ and $Hl^T$ are true while $Sp^T$ and $Yl^{T+1}$ are false. Clearly no theory which unconditionally predicts that the statement

(2.1) If I stepped on Bowser's tail at $T$, then Bowser would yelp at $T+1$

is false, can be correct. Yet this is exactly the prediction made by Pollock's theory. It is enough to note that, since yelps can occur spontaneously, $\sim Yl^{T+1}$ is an ungrounded event: no class of earlier true RBPs yields it. Thus $\{\sim Yl^{T+1}\}$ is a complete causal history of $\sim Yl^{T+1}$; and as it is clearly $Sp^T$-compatible, $\sim Yl^{T+1}$ is not undercut by $Sp^T$. Thus, by the Note, not all $Sp^T$-maximal subsets of $|Sp^T|$ yield $Yl^{T+1}$ via $Sp^T$, and consequently, (2.1) is false on (P), contrary to what is intuitively expected.

*Blue's Theory*

The proposal of N. A. Blue (1981), can also be recast in terms of a notion of undercutting. Let us call an RBP $P$ *Blue-undercut by* $A$ if some $A$-compatible set of true RBPs, none of them later than $P$, becomes $A$-incompatible when $P$ is added to it. Let $\|A\|$ be the class of true RBPs which are not Blue-undercut by $A$.

Blue seems to take it for granted that if $A$ is compatible with the laws of nature then $\|A\|$ is $A$-compatible. But in the absence of some finitary assumptions which Blue never explicitly states, this need not be so.[2] Let us assume, therefore, that some additional general postulates have been laid down which ensure that for any 'legal' $A$, $\|A\|$ is $A$-compatible. Blue's truth-condition for subjunctive conditionals is then as follows:

(B) (∗) is true iff $\|A\|$ yields $B$ via $A$.

To see how (B) copes with Goodman's problem, consider again

EXAMPLE 1 (continued). $\{Wm^T, Dr^T, Ox^T\}$ is an $Sr^T$-compatible set of true RBP's, all of them earlier than $\sim Li^{T+1}$, and becomes $Sr^T$-incompatible when $\sim Li^{T+1}$ is added to it. Hence $\sim Li^{T+1}$ is Blue-undercut by $Sr^T$ and consequently there is no reason to think that $\| Sr^T \|$ yields $\sim Dr^T$ via $St^T$. If it does not, (1.2) is false on (B) as desired.

On the other hand, since $Wm^T$, $Dr^T$, and $Ox^T$ are unlikely to have played any role in the causal history of $\sim Sr^T$, no $Sr^T$-compatible set of true RBPs which take place before or at $T$ becomes $Sr^T$-incompatible when $Wm^T$, $Dr^T$ or $Ox^T$ is added to it. The three propositions thus belong to $\| Sr^T \|$ and consequently, $\| Sr^T \|$ yields $Li^{T+1}$ via $Sr^T$. Hence (1.1) is true on (B) as desired.

There is a sense in which Blue's notion of uncercutting is a direct reverse of Pollock's. If the antecedent $A$ is a false RBP, then Pollock disqualifies a true RPB $P$ if the negation of $A$, together with some other true RBPs, yields $P$; whereas Blue disqualifies $P$ if $A$, together with some true RBPs, yields the negation of $P$. As intuition demands that $P$ be disqualified whether it is of one of these two kinds or the other, it is to be expected that in some cases where Pollock's definition fails Blue's will succeed, and in some cases where Blue's definition fails Pollock's will succeed.

EXAMPLE 2 (continued). $\sim Yl^T$, as we have seen, is not Pollock-undercut. But it is Blue-undercut: $\{Dg^T, Hl^T\}$ is an $Sp^T$-compatible set of true RBPs which becomes $Sp^T$-incompatible when $\sim Yl^{T+1}$ is added to it. But $Dg^T$ and $Hl^T$, provided they played no role in the causal history of $\sim Sp^T$, both belong to $\| Sp^T \|$. Consequently, $\| Sp^T \|$ yields $Yl^{T+1}$ via $Sp^T$ and (2.1) is true on (B) as desired. Thus, while (P) fails the Bowser test (as we have seen), (B) passes it.

EXAMPLE 3 (*The Inverse Bowser Test*). Using the same notation as in Example 2, imagine that this time $Dg^T$, $Hl^T$, $Sp^T$, and $Yl^{T+1}$ are all true. Clearly no theory which unconditionally predicts that

(3.1) If I had not stepped on Bowser's tail, he would have yelped

is true, can be correct. Yet this is exactly what (B) predicts. As Bowser is free to yelp spontaneously, no $\sim Sp^T$-consistent set of true RBPs becomes $\sim Sp^T$-inconsistent when $Yl^{T+1}$ is added to it. Hence $Yl^{T+1}$ is not Blue-undercut by $\sim Sp^T$, $\| \sim Sp^T \|$ yields $Yl^{T+1}$, and (5.1) is true on (B). On the other hand, as $Sp^T$ played a causal role in bringing $Yl^{T+1}$ about, $Yl^{T+1}$ is Pollock-undercut by $\sim Sp^T$, and there may well be $\sim Sp^T$-maximal subsets of $|\sim Sp^T|$ which do not yield $Yl^{T+1}$; if there are, (3.1) is false on (P) as desired. Thus, while (B) fails the inverse Bowser test, (P) passes it.

*Tichý's Theory*

In Tichý (1978) an attempt is made to solve Goodman's problem by taking advantage of the asymmetry of the cause-effect relation. Laws of nature are construed as general statements about the cause-effect relation between (occurring and non-occurring) events. A (true or false) RBP $P$ is said to be a *causal consequence* of a class $C$ of RBPs if $P$ belongs to the closure of $C$ with respect to the cause-effect relation, in other words, if there exists an uninterrupted chain of cause-effect links starting with members of $C$ and culminating in $P$. The relation of causal consequence is clearly stronger than the yield-relation which constitutes the cornerstone of the theories of Goodman, Pollock, and Blue: $C$ may yield $P$ without having the causal power to bring $P$ about. As an illustration, consider again

EXAMPLE 1 (continued). The class $\{Wm^T, Ox^T, \sim Li^{T+1}\}$ yields $\sim Dr^T$ via $Sr^T$, since an addition of $Dr^T$ to it would make the class incompatible with $Sr^T$ and the laws of nature. Yet $\sim Dr^T$ is not a causal consequence of $\{Wm^T, Ox^T, Sr^T, \sim Li^{T+1}\}$: the state of affairs consisting in $X$'s being a well-made, oxygen-surrounded, and struck match which is not going to light in one's second's time, is clearly powerless to make $X$ currently wet.

A set $S$ of true RBPs is said to be *connectively closed* if $S$ contains the actual causes and effects of any subclass of $S$.

For the sake of simplicity, let us restrict ourselves to conditionals whose antecedents are conjunctions of RBPs; where $A$ is such an antecedent, let $A^*$ be the class of RBPs it implies. A set $S$ of true RBPs is said to be *$A^*$-admissible* if it is $A^*$-compatible and connectively closed. Furthermore, $S$ is said to be *strongly $A^*$-admissible* if it is $A^*$-admissible and, for any $A^*$-admissible set $S'$, $S$ is compatible with the class of causal consequences of $S' \cup A^*$. Tichý's definition can now be stated thus:

(T) (∗) is true iff there is a strongly $A^*$-admissible set $S$ of true RBPs such that $B$ follows from the class of causal consequences of $S \cup A^*$.

To see how (T) copes with Goodman's problem, consider again

EXAMPLE 1 (continued). Let $S$ be the smallest connectively closed set of true RBPs containing $Wm^T$, $Dr^T$, and $Ox^T$. There is no reason to expect $S$ to fail of strong $\{Sr^T\}$-admissibility. $Li^{T+1}$ is in the class of causal consequences of $S \cup \{Sr^T\}$. This means, firstly, that (1.1) is true on (T), and secondly, that no $Sr^T$-admissible class of true RBPs contains $\sim Li^{T+1}$ and there is thus no reason to expect (1.2) to be true on (T).

(T) implements the intuitive idea behind Pollock's proposal as well as the one behind Blue's. As a result, it passes both the Bowser tests.

EXAMPLE 2 (continued). Let $S$ be the smallest connectively closed set of true RPBs containing $Dg^T$ and $Hl^T$. There is no reason to expect $S$ not to be strongly $Sp^T$-admissible. Moreover, $Yl^{T+1}$ is in the class of causal consequences of $S \cup \{Sp^T\}$. Thus (2.1) is true on (T) as desired.

EXAMPLE 3 (continued). Consider any connectively closed set $S$ of true RPBs. First assume that $Yl^{T+1}$ is among the causal consequences of $S$. Then, since $Sp^T$ played an essential role in bringing $Yl^{T+1}$ about, $S$ contains $Sp^T$ and is, therefore, $\sim Sp^T$-inadmissible. Now assume that $Yl^{T+1}$ is not a causal consequence of $S$. Then clearly $Yl^{T+1}$ is not a causal consequence of $S \cup \{\sim Sp^T\}$ either. Consequently, (3.1) is not true on (T).

Alas, (T) is also wrong, as witness the following example:

EXAMPLE 4 (*The Conference Test*). Imagine that John, who lives in Tucson, Ariz., was invited to two equally attractive philosophy conferences to be held concurrently in Boston, Mass., and Los Angeles, Calif. John decided to go to Boston (by car) and arrived in the city at $T$. Under the circumstances, it is clearly true to say that

(4.1) If John had not been in Boston at $T$ he might have been in Los Angeles instead.

And as it is physically impossible to get from Massachusetts to Los Angeles in one minute, no theory which unconditionally predicts that

(4.2) If John had not been in Boston at $T$, he would have been in Massachusetts one minute before $T$

is true, can be correct. Yet this is exactly what is predicted by (T). Let $B^T$

and $M^{T-60}$ be the propositions which report John's actual spatial locations at $T$ and a minute before $T$ respectively, so that the location reported by $B^T$ is inside Boston and the one reported by $M^{T-60}$ inside Massachusetts. Let $S$ be the smallest connectively closed set of true RBPs containing $M^{T-60}$. It would be absurd to think that $B^T$ is a causal consequence of $S$; for this would mean that between $T-60$ and $T$ it was not up to John to stop the car and refrain from crossing the Boston city boundary. There is thus little reason to doubt that $S$ is strongly $\sim B^T$-admissible. Consequently, (4.2) is true and (4.1) false on (T), contrary to what is intuitively expected.

Pollock's and Blue's theories fare no better.

EXAMPLE 4 (continued). $M^{T-60}$ is obviously a member of $|\sim B^T|$ and it is hard to see how it could be possibly missing from any $\sim B^T$-maximal subset of $|\sim B^T|$. Thus (4.2) is true on (P). As for (B), note that the set of all true RBPs not later than $T-60$ is $\sim B^T$-compatible. Thus $M^{T-60}$ is a member of $\|\sim B^T\|$ and (4.2) is true on (B).

The three theories just considered share the idea that one has to somehow maximize the class of auxiliary truths which can be legitimately added to the counterfactual protasis and that the conditional (*) is then true if $A$, reinforced with these auxiliary truths, has the (logical or causal) power to make $B$ true. The idea is suggested by, and seems to work for, conditionals of a certain sort, like those in Examples 1 and 2. These are conditionals which are not weakened by attaching to the antecedent the tag '... and history deviated from its actual course as little as possible'. But Example 4 shows that not all conditionals are of this sort. Hence any attempt to work the tag into a definition of the subjunctively-conditional connective itself is bound to lead to inadequacy. Any naturally looking method of maximizing the class of auxiliary truths seems to saddle us with too many of them.

Let us then set two-parameter theories aside and address ourselves to what the literature has to offer in the way of three-parameter theories of subjunctive conditionals.

## 2. THREE-PARAMETER THEORIES

When adjudicating between rival three-parameter theories, one cannot normally use concrete examples as the ultimate court of appeal. For given a subjunctive conditional, it will usually be possible to find value-assignments for the third parameter on which the theory delivers whatever prediction concerning that conditional may seem intuitively appropriate. What can be done, however, is to see to what extent such assignments agree with, or do violence to, our pretheoretic intuitions concerning that parameter. The reduction of one concept to another is hardly illuminating if it only works at the cost of radically revising our basic intuitions concerning the latter concept. One would not be impressed by a theory which insisted that gold is really solidified water, and added that we must, of course, abandon the endemic prejudice that the stuff in the Thames river is water. We shall see that most of the recently proposed three-parameter theories of subjunctive conditionals suffer from this gold-water syndrome.

*Lewis' Theory*

For David Lewis (1973 and 1979) the third, implicit, parameter co-determining the force of (∗) is relative similarity among possible worlds. Lewis assumes that given two possible worlds, we often can and do judge one of them to be more similar, overall, to the actual world than the other. These judgments, however, fluctuate from context to context, so that before one can fully appreciate what (∗) says on a given occasion one has to know which particular relation of relative similarity is intended. Where $R$ is such a relation we shall speak, briefly, about *R-similarity* between wordls. Moreover, a world in which proposition $P$ is true will be called a $P$-world. Lewis's truth-condition for subjunctive conditionals can now be stated thus:

(L) (∗) is true relative to $R$ iff either $A$ is true in no world at all or there is an $A \& B$-world which is more $R$-similar to the actual world than is any $A \& {\sim}B$-world.

It is readily seen that if $A$ and $B$ are logically independent (L) makes (∗) true relative to some $R$s and false relative to others. So if our pretheoretic intuitions did not impose any constraints on what can sensibly count as a relation of overall similarity between worlds, (L) would be quite worthless. Let us then see whether Lewis's theory heeds some of the more obvious constraints of this sort.

One such contraint is surely the following: if two worlds differ merely in that one isolated event happens in one of them and not in the other, then the world which agrees in this respect with the actual world is more similar to the actual world than is the other one. Lewis, however, has to repudiate this constraint. To see why, consider again

EXAMPLE 3 (continued). $\sim Sp^T$-worlds divide into $\sim Sp^T$ & $Yl^{T+1}$-worlds and $\sim Sp^T$ & $\sim Yl^{T+1}$-worlds. As $Yl^{T+1}$ is actually true, the former worlds have a positive edge on the latter worlds according to the constraint at issue. Thus if the constraint were adopted, (3.1) would be true relative to any acceptable similarity relation. Thus Lewis has to insist, counterintuitively, that the agreement of a $\sim Sp^T$-world with the actual world as far as Bowser's yelping is concerned, may contribute nothing whatsoever to the similarity of the two worlds.[3]

Another obvious constraint is this: the similarity of a world to the actual world is inversely proportional to the scale on which it differs from the actual world: the more comprehensive and wide-ranging the difference, the more dissimilar the world is from the actual world. For example, one might feel on safe ground in thinking that as far as overall similarity between worlds goes, the occurrence or otherwise of an Earth-destroying holocaust will be of greater moment than the integrity or otherwise of a tiny piece of wire. This intuition is also rejected by Lewis as erroneous. To see why, consider an example given by Kit Fine (1975):

EXAMPLE 5 (*The Holocaust Test*). Suppose that a nuclear holocaust will in fact never happen and that at $T$ Nixon's office was equipped with a perfectly functioning holocaust button. Clearly no theory which makes the unconditional prediction that

(5.1) If Nixon had pressed the button at $T$, then there would have been a nuclear holocaust

is false, can be correct. But Lewis' theory would yield that prediction if it endorsed the constraint under consideration. For consider a world $W$ which is very much like ours except that Nixon presses the button at $T$ and that at $T$ the wire connecting the button with the launchers is broken. The constraint requires us to count such a world as being more similar to the actual world than is any world in which Nixon presses the button and the whole planet (including Nixon, his button *and the connecting wire*) is destroyed. But this makes (5.1) unconditionally false on (L). To avoid this undesirable consequence Lewis heroically repudiates the constraint and maintains that a relation according to which the holocaust world outstrips *W* in closeness to the actual world can still qualify as a relation of overall *similarity* between worlds.

Or, to take another example, one might feel safe in thinking that a counterfactual one-minute delay on a journey which was actually undertaken constitutes a radically smaller departure from actuality than a counterfactual journey by the same traveller in the opposite direction. Lewis, however, has to repudiate this intuition as well. To see this, consider again

EXAMPLE 4 (continued). Let *W* be a world which is just like ours except that shortly before crossing the Boston city boundary John briefly stops the car to blow his nose, and as a result arrives in Boston one minute after *T*. If the constraint under consideration were accepted, then on any similarity relation worthy of being so called, *W* would be more similar to the actual world than was any world in which John goes to the West-Coast conference and is in Los Angeles at *T*. But then (L) would make (4.2) unconditionally true and (4.1) unconditionally false. Thus Lewis has to reject this constraint as well.

We thus see that Lewis's theory affords an explication of subjunctive conditionals in terms of comparative similarity among worlds only at the cost of distorting the latter concept out of all recognition. On Lewis's own admission, the relations which have to be invoked in simple cases like those considered in Examples 4 and 5, are not the relations which are 'likely to guide our explicit judgements of similarity'. The price to be paid for accepting Lewis' theory of conditionals is thus similar to the price which would have to be paid for accepting the above-mentioned aquatic theory of gold.

Another troublesome aspect of Lewis's proposal is that, although explicitly presented as a three-parameter theory of subjunctive conditionals, it does not grant the third parameter a place in the logical structure of such conditionals. On Lewis's view, the logical form of a subjunctive conditional is simply that of an application of a binary function to two propositions. He puts forward inference schemata and whole logical calculi in which the subjunctively-conditional relation is represented by a binary propositional connective.

This is rather as if someone conceded that when a person is judged superior to another the truth of the judgment depends on the respect in which they are compared – Jones, for instance, may be superior to Smith as a pianist while Smith being superior to Jones as a carpenter – and then nevertheless proceeded to develop a theory of a *two-place* relation of superiority. He might come out with all sorts of startling 'discoveries'. One thing he might tell us is that, contrary to a hitherto widely held belief, superiority obeys no asymmetry law, as witness Jones and Smith. Surely no one would be impressed with such a 'discovery', since it is clear at once that the alleged counterexample involves a shift in the suppressed third parameter that superiority judgments depend upon.

Yet Lewis has thus far been getting away with arguments which are virtual carbon copies of this one. He maintains, for example, that, contrary to a hitherto widely held belief, subjunctive conditionals obey no contraposition law. To illustrate, he tells the following story. Olga pursues Boris, who tries to avoid her. One day they are both invited to a party. So Olga is disposed to go if Boris goes, and Boris is disposed to stay at home if Olga goes. Now we are asked to consider the conditionals

(i) If Boris went to the party then Olga would go

and

(ii) If Olga did not go to the party, then Boris would not go.

'Is (i) true?' asks Lewis and answers 'Yes'. 'Is (ii) true?' 'No,' he says, 'what is true is rather the contradictory of (ii)', namely

(~ii) If Olga did not go to the party, then Boris might go.

Thus no contraposition law holds for subjunctive conditionals, Lewis concludes.

Lewis does not seem to see that once it is admitted that the truth-value of a subjunctive conditional depends on a third, contextually determined, parameter, an example of this sort has little bite unless it can be shown that it involves no surreptitious change in the third, suppressed, parameter. Can we be sure that no such change occurs in passing from (i) to (~ii)? On the contrary, it is pretty clear that a change does occur. As a situation in which both Boris and Olga go to the party is contrary to Boris's dispositions, (i) only comes out true if one ignores Boris's dispositions without ignoring Olga's. Since, on the other hand, a situation in which Olga does not go to the party and Boris does is contrary to Olga's dispositions, (~ii) is only true if one ignores Olga's dispositions. Thus tacit premises seem to change as one passes from (i) to (~ii).

For Lewis the implicit third parameter is, of course, not a set of tacit premises but comparative similarity between worlds. But the argument is readily adapted to apply to the similarity theory. In order for (i) to be true on the similarity theory, a violation of Olga's actual dispositions must count as a greater departure from the actual world than a violation of Boris's. Otherwise we would have a tie between worlds in which both Boris and Olga go to the party and worlds in which Boris goes and Olga does not, and consequently (i) would not be true on (L). But in order for (~ii) to come out true in (L) the reverse must be the case: a violation of Boris's actual dispositions must count for a greater departure from actuality than a violation of Olga's. Thus the similarity relations operative in the two cases must be different. But if so, Lewis's example is no more interesting than the Jones-Smith 'counterexample' to the asymmetry of the superiority relation. If Lewis brought his notation in line with his own theory and symbolized (*) not as $A \mathrel{\Box\!\!\rightarrow} B$ but as, say, $A \mathrel{\Box\!\!\rightarrow_R} B$, where $R$ stands for the similarity relation at issue, it would be clear that his Boris-Olga example is evidence against the inference form $A \mathrel{\Box\!\!\rightarrow_R} B / \therefore \sim B \mathrel{\Box\!\!\rightarrow_S} \sim A$. But no one would expect this form to be valid in the first place. What would be surprising is if someone produced an intuitively appealing counterexample to $A \mathrel{\Box\!\!\rightarrow_R} B / \therefore \sim B \mathrel{\Box\!\!\rightarrow_R} \sim A$. But no such counterexample has been given.

It is also worth noting, in passing, that Lewis' definition does not work at all where the antecedent constitutes a counter-temporal rather than counterfactual hypothesis. Consider

EXAMPLE 6. Suppose that Lindsay is a regular churchgoer who never misses the Sunday, 10 a.m. mass and that it is now Wednesday at 10 p.m. It is clearly true to say that

(6.1) If it was Sunday, 10 a.m. Lindsay would be in church.

Are there any worlds at all in which it currently *is* Sunday 10 a.m.? Presumably, worlds differ from one another in what *takes place* in them at various times, not in what time it currently is in those worlds. So given that currently it is not Sunday 10 a.m., it is not Sunday 10 a.m. in any world at all. This means that (L) makes (6.1) true as desired, but vacuously so: the contrary conditional 'If it was Sunday 10 a.m., Lindsay would not be in church' is also made true (no matter what similarity relation is intended).

*Kratzer's Theory*

For Angelika Kratzer (1981), the third, contextually determined, parameter is what she calls 'partition of worlds into facts'. Where $W$ is a possible world, a *partition* of $W$ is any set of propositions which jointly imply every proposition true in $W$. A *partition function* is a function which takes every world to a partition of that world. The members of the value of a partition function at a world are called the *facts* of that world relative to the function. Partition functions, according to Kratzer, are 'fixed by the utterance situation' and vary from context to context. A given subjunctive conditional can only be evaluated relative to such a function. Using the terminology introduced in Section 1 (and ignoring the provision Kratzer makes for 'counterlegals') Kratzer's truth-condition for subjunctive conditionals can be stated thus:

(K) (*) is true relative to partition function $F$ iff every $A$-maximal subset of the value taken by $F$ at the actual world yields $B$ via $A$.

In order to see how this definition is meant to work, consider again

EXAMPLE 2 (continued). If $\sim Yl^{T+1}$ counted as a fact in its own right, (2.1) would be false on (K). For $\sim Yl^{T+1}$ is clearly $Sp^T$-compatible, hence, by the Note, at least one $Sp^T$-maximal subset of the class of (actual) facts would not yield $Yl^{T+1}$ via $Sp^T$. But suppose that $\sim Yl^{T+1}$ is not itself a fact and that it is invariably lumped together with $\sim Sp^T$. Then no $Sp^T$-maximal set of facts will yield $\sim Yl^{T+1}$. Assuming furthermore that $Dg^T$ and $Hl^T$ are not always lumped together with $\sim Sp^T$, it may well be that all $Sp^T$-maximal sets yield $Dg^T$ and $Hl^T$. If they all do, each of them yields $Yl^{T+1}$ via $Sp^T$ and (2.1) is true on (K) as desired.

It is readily seen that if $A$ and $B$ are logically independent, (K) makes (*) true relative to some $F$s and false relative to others. For let $N$ be the necessarily true proposition and $Tr$ the conjunction of all actual truths. Then (*) is true relative to any $F$ whose value at the actual world is $\{Tr\}$ and false relative to any $F$ whose value at the actual world is $\{N, Tr\}$.

Just like Lewis's theory is tenable only at the cost of outlandish judgements of similarity between worlds, Kratzer's theory is tenable only at the cost of no less outlandish and completely unfamiliar judgments as to what is or is not a fact. To see this, consider again

EXAMPLE 5 (continued). In order to enable (K) to deliver the correct prediction about (5.1), one has to defy common sense by insisting that hardly any facts take place strictly after $T$. Let $Bu^T$ be the proposition that Nixon pushed the button at $T$, and consider, for instance, the shooting of Reagan in 1981. If the shooting was a fact, then, since it is $Bu^T$-compatible, there would be $Bu^T$-maximal sets of facts which yield it via $Bu^T$. But no such set yields, of course, the holocaust via $Bu^T$. Thus if (5.1) is to be true on (K), the shooting must not count as a fact.

For another illustration, consider

EXAMPLE 7. Suppose John burped at $T$. Call this proposition $Bp^T$. Can $Bp^T$ be granted the status of a fact on Kratzer's theory? Assume that ten years before $T$ John was driving hard on the heels of a truck which suddenly stopped; John only survived because he stepped on the brake pedal in time. (Such live-saving events happen undoubtedly to everybody.) Thus it is true to say that

(7.1) If John had not stepped on the brake pedal ten years before $T$, he would not have burped at $T$.

But unless he is prepared to deny (7.1), the adherent of Kratzer's theory must deny that $Bp^T$ is a fact. For let $Sb$ be the proposition that John stepped on the brake ten years before $T$. $Bp^T$ is $\sim Sb$-compatible; hence if it was a fact then, by the Note, not all $\sim Sb$-maximal sets of facts would yield $\sim Bp^T$ and (7.1) would be false on (K). So perhaps $Bp^T$ is not a fact in its own right and must be lumped with $Sb$. But the resulting lump is not a fact either. It must itself be lumped with every event which saved John's life prior to $Sb$, and the result must in turn be lumped with every event which saved John's mother's life prior to John's birth, etc. Now while (7.1) and all the other conditionals will be readily assented to by everybody, the holistic notion that John's burp is not a fact but an inseparable ingredient of a massive epic incorporating the braking episode and selected episodes in the lives of John's ancestors right down to the apes, is something that will hardly ring a bell.

Thus as far as our prior intuitions go, conditional statements are not related to facts the way Kratzer tells us they are. We can, of course, disown our pretheoretic notion of fact in favour of one which accommodates Kratzer's theory. It is not immediately clear, however, exactly what has been achieved by reducing subjunctive conditionals to a completely novel notion of fact, one which bears no relation to what normally goes by that name, and which apparently can only be grasped through a prior understanding of subjunctive conditionals themselves.

*The Mill-Ramsey-Chisholm Theory*

The theories of Lewis and Kratzer are both completely divorced from the way subjunctive conditionals are argued over in practice. If world-similarity or world-partitioning are what the truth-value of a subjunctive conditional turns upon, how is it that disputes about conditional statements are never settled by reference to such matters? Suppose that a dispute arises as to whether some nuclear missiles would have been launched had someone pushed a certain button yesterday. Are those who think that the answer is 'Yes' ever likely to support their view by arguing that a situation in which the button was pushed and the rockets went off is more similar, overall, to the way things in fact are than is any situation in which the button was also pushed but nothing happened? Or are they likely to argue that nothing that has happened since yesterday constitutes a fact in its own right? And will those who think the answer is 'No' try to refute these world-similarity or world-gerrymandering claims? I have yet to hear someone argue that way off the premisses of a philosophy department.

Among people for whom the correct truth-value of the conditional is a matter of genuine concern, such a dispute is likely to turn very soon into a dispute over some matters of fact and of ordinary logic. Those who think the conditional is true will typically invoke some facts (like the nature and state of the electrical circuits involved) and physical laws (or what they believe to be such) and then appeal to ordinary logic to show that these, together with the imaginary pushing of the button are related to the imaginary launching as the premisses of a valid argument to its conclusion. Their opponents, on their part, are likely to try and cast doubt on the alleged facts, or on the alleged laws, or on their adversaries' logic. The two parties will normally agree on which particular matters of fact are relevant to the problem at issue: yesterday's condition of the circuits, for example, will undoubtedly be deemed relevant. Today's condition of the missiles undoubtedly won't: neither party would dream of invoking this in favour of or against the conditional. No one would take seriously a clever logic-chopper who argued that, since the rockets are in fact still in their silos, then had the button been pushed yesterday, something would have been the matter with the circuits. Not because his conditional is unacceptable in some absolute sense. But because he appeals to a fact which does not belong to the class of facts which are relevant in the present context. In other contexts, where the class of relevant facts is circumscribed differently, the conditional may be quite to the point.

These observations, if correct, suggest a very natural theory of subjunctive conditionals, according to which the third, implicit parameter is a definite set of auxiliary indicative premises tacitly added by the speaker to the protasis in order to make the apodosis fall out of it as a logical consequence. On this theory, the logical relation involved in subjunctive conditionals is the familiar one of implication or entailment: subjunctive conditionals are explained simply as elliptical statements of logical consequence.

This theory is not new. It was first adumbrated by Mill (1868, p. 92), then endorsed by Ramsey (1931, p. 248), and later resurrected by Chisholm (1955, pp. 102–705). The following concise statement of the theory is due to R. M. Walters (1961, p. 37):

> ... any counterfactual contains an implicit reference to an argument and so to unstated premises. The meaning of a counterfactual is not clear until such premises are made explicit... . If *X* asserts that 'If ravens had survived in snowy regions, they would have been black', and assumes ['All ravens are black'], then his counterfactual is true. If Y asserts that 'If ravens had survived in snowy regions, they would have been white' and assumes ['All surviving in snowy regions are white'] then his counterfactual is true as well. This is a [surprising] results ... only if ... we neglect [the character of counterfactuals] as truncated arguments.

The truth-condition for subjunctive conditionals suggested by the above authors can briefly be stated thus:

(M) (∗) is true relative to class $C$ of auxiliary indicative premises iff the members of $C$ are true and $B$ is a logical consequence of $C \cup \{A\}$.

EXAMPLE 1 (continued). Conditional (1.1) is true on (M) provided $Wm^T$, $Dr^T$, $Ox^T$, and the relevant law are among the auxiliary premises, for these premises together with $Sr^T$ imply $Li^{T+1}$. (1.2), on the other hand, normally counts as false, for it is unusual for $\sim Li^{T+1}$ to be among the intended auxiliary premises.

Although never refuted, this theory was virtually forgotten as philosophical logicians became preoccupied with a search for a two-parameter account. But those who, like Lewis and Kratzer, have recently abandoned that search have no excuse for ignoring this long-standing three-parameter theory. They ought to tell us exactly what they find unacceptable about it and why they wish to supersede it with their own, much more involved and yet intuitively less satisfactory accounts.

## 3. THE TACIT-PREMISE THEORY DEFENDED

Although it has never been pointed out by the proponents of rival three-parameter theories, there *is* something unsatisfactory about the Mill-Ramsey-Chisholm account. It is vulnerable to a serious objection from a pragmatic point of view.

It is hard to deny that speakers often assert subjunctive conditionals without intending any particular set of truths which, when conjoined with the protasis, yield the apodosis as a logical consequence. Suppose Tom, who has never heard of oxygen, is holding a well-made, dry, and oxygen-surrounded match and comments that it would light if he struck it. He is unlikely to have in mind a set of true indicative premises sustaining the conditional, for (provided he in fact refrains from lighting the match) any such set is bound to include laws and initial conditions concerning oxygen. Do we want to say that Tom's statement cannot, therefore, be correct? Surely not.

Pollock (1976) has presented this observation as a conclusive explosion of the Mill-Ramsey-Chisholm theory. If the speaker himself often does not *know* exactly which truths constitute the class of auxiliary premises on a given occasion of utterance, then the membership of the class cannot be determined by his tacit choice, by what he *means*. Pollock's conclusion is that a two-parameter theory of subjunctive conditionals must be sought after all.

But if this argument from ignorance establishes that the speaker does not choose his auxiliary premises, then a similar argument will establish that the speaker does not always choose his antecedent either. Suppose, for example, that John is sitting precariously on a rickety chair and I comment: 'If John was one stone heavier than he is, the chair would collapse'. Do I know what my antecedent is? My antecedent is hardly the contradiction that John is one stone heavier than he is. If my conditional is to stand a chance of being non-vacuously true, the operative antecedent must be the proposition, as regards John's actual weight, that John is one stone heavier than that. But if I do not happen to know how heavy John is, I will not know which proposition that is: Thus the identity of the antecedent, one may argue, cannot be the result of my, the speaker's, choice. (Similar comments apply to the consequent of a conditional: 'If the chair collapsed, the centre of John's gravity would be lower than it is'.)

But is it true that whenever a choice is made, the chooser invariably knows what he has chosen? The question can be answered only after a distinction has been drawn. Choosing may be direct or indirect. When a policeman puts a pair of handcuffs on a particular man in a line-up of suspects, he has chosen his arrestee directly. But if he merely decides to arrest the murderer of Mrs. Brown, his choice is indirect and he may well not know exactly which man to handcuff. What he chooses directly in this case is not a definite individual but an *office* occupiable by an individual, a definite status that an individual may enjoy: that of Mrs. Brown's murderer. It is not up to the policeman to decide which of the suspects occupies the office; the occupancy of the office is a matter of brute fact.

An individual office is best seen as a function which takes every world-time to the individual (if any) which occupies the office in that world at that time. One can be perfectly familiar with such an office and yet, if one does not know which world is actual or which moment is present, one may have no idea which individual is its actual, current occupant. To choose an individual indirectly is to choose it as the occupant of a directly chosen individual office; the chooser may in this case have no idea which individual has been thus chosen.

What has just been said of individuals goes for objects of other logical types. In particular, a proposition may be chosen either directly, or indirectly by way of a propositional office. I may, for instance, decide to consider the proposition that John weighs 10 stones. Alternatively, I may decide to consider John's favourite proposition. These are two different things to do even if John's favourite proposition happens to be that he weighs 10 stones. What I am directly concentrating upon in the latter case is a propositional office, a function from world-times to propositions, and I am leaving it to John to decide which proposition occupies the office in the actual world at the present moment. Or, I may consider the proposition, as regards John's weight, that John is one stone heavier than that. Here again, what I have directly selected as an object of my contemplation is not a definite proposition, but a propositional office, an office which is occupied, at any world-time, by the proposition that John is one stone heavier than $q$, where $q$ is John's weight in that world at that time. I may be perfectly familiar with that office without knowing which proposition occupies it.

In the chair example above, it is this latter office that the speaker directly selects; and what he affirms is to the effect that the current actual occupant of the office, whichever proposition that is, is such that if it were true, the chair would collapse. He selects the antecedent of his conditional indirectly, leaving it partly to the facts to decide which proposition it is.

But if the antecedent can be chosen indirectly through a propositional office, so can the tacit auxiliary premise posited by the adherent of the Mill-Ramsey-Chisholm theory. The collapse of the chair follows from the antecedent in conjunction with an auxiliary premise which describes John's current behaviour, the current structural features of the chair and the laws of mechanics. But the speaker need not know what the behaviour, the structure, and the laws actually are. All he has to choose is the propositional office which is occupied, at any world-time, by the proposition which correctly reports what the behaviour, the structure, and the laws are in that world at that time. This choise will normally be tacit, but it is easy to spell it out by expanding the conditional statement into

> If John was one stone heavier than he is *while John's behaviour, the structural features of the chair, and the laws of mechanics were still what they are in fact*, then the chair would collapse.

The statement says that the occupant (whichever proposition that is) of the office named by the explicit antecedent and the occupant (whichever proposition it is) of the tacitly understood office jointly imply the consequent.

## 4. THE TACIT-PREMISE THEORY AMENDED

The above point is not easily made in the logical symbolism which is currently standard. This symbolism reflects the almost universal penchant for eschewing possible-world and time variables and thus keeping one's notation simple at the cost of perspicuity. As a result, important logical distinctions are often notationally obliterated even where they are clearly manifested in the syntax of ordinary language.[4]

One case in point is the practice of symbolizing indicative conditionals and implication statements in syntactically isomorphic way: $A \supset B$ and $A \prec B$ respectively. The practice flies in the face of the fact that $\supset$ and $\prec$ are functions of different logical types and that propositional constructions based upon them must therefore differ structurally. The function $\supset$ takes couples of truth-values to truth-values; hence it can be directly applied not to propositions but only to the truth-values the propositions take at a world-time. Thus the correct analysis of an indicative conditional of $A$ and $B$ is $\lambda w \lambda t . A_{wt} \supset B_{wt}$,[5] where $w$ and $t$ are variables ranging over worlds and times respectively. The implication function $\prec$, on the other hand, takes couples of propositions to truth-values, hence the correct analysis of the implication of $A$ and $B$ is $\lambda w \lambda t . A \prec B$, a logical construction of a completely different structure. These analyses reveal at a glance the difference in the modal status between the two kinds of statement. The truth-value of the indicative conditional may differ from one world-time to another, because the abstraction operators work on variables which are free in their scope. The implication statement, on the other hand, is bound to have a fixed truth-value throughout the logical space, because the scope of the abstraction operators in its analysis is closed. This distinction, although clearly manifested in the syntax of ordinary language ('if ... then' connects sentences whereas 'implies' connects noun-phrases) is completely obliterated in the conventional notation. The artificial ideography is in this respect less perspicuous than the ordinary-language locutions on which it is meant to shed logical light.

It might be objected that since the abstraction operators in an implication construction are vacuous, there is no point in dragging them along and the familiar notation $A \prec B$ will do as well. This objection overlooks the fact, however, that the argument places of $\prec$ need not be always occupied by closed constructions of definite propositions; they may be occupied by open propositional constructions depending on world and time variables. Consider, for example, the statement 'John's favourite proposition implies that he weighs over nine stones', which is true assuming that John favours the proposition that he weighs ten stones. The latter proposition is not mentioned in the implication statement. What is mentioned is rather the propositional office of John's favourite proposition, a function – call it $O$ – from world-times to propositions. Before $O$ can become an argument for $\prec$, it must be applied to a world-variable and time-variable; and in the absence of the initial abstraction operators these variables would dangle. Thus the logical form of the statement is $\lambda w \lambda t. O_{wt} \prec B$, where $B$ is the proposition that John weighs more than nine stones. It is a contingent statement in spite of the fact that $\prec$ relates propositions quite independently of contingent matters.

Our amendment of the Mill-Ramsey-Chisholm theory is simply this. A subjunctively-conditional sentence expresses a construction of the form $\lambda w \lambda t. \mathscr{A} \prec \mathscr{B}$, where $w$ and $t$ may be free in $\mathscr{A}$ and/or $\mathscr{B}$; parts of the construction $\mathscr{A}$ are often tacitly understood rather than explicitly spelled out in the antecedent of the conditional sentence. (No novel truth condition is called for, for the truth condition of implication statements is well known.)

If this analysis is right, Goodman's problem is solved in a very prosaic way indeed: the range of truths which can be legitimately added to the stated antecedent of a conditional is tacitly (and mostly indirectly) circumscribed by the speaker. This may sound dull, but it is difficult to see what else can be said in the face of conditionals like those in Examples 6, 9, and 11 below.

## 5. EXAMPLES

Any conditional is acceptable if its antecedent, as it stands, entails the consequent, i.e., if there is no world-time at which the if-clause is true and the then-clause is not.

EXAMPLE 8. The conditional

(8.1) If Lindsay were a bachelor, then Lindsay would be a male

is true because Lindsay is a male at any world-time at which he is a bachelor. The sentence expresses the construction

$$\lambda w \lambda t. [\lambda v \lambda s. B_{vs} X] \prec [\lambda v \lambda s. M_{vs} X],$$

where $X$, $B$, and $M$ are , respectively, Lindsay, bachelorhood, and masculinity; $w$ and $v$ are world variables and $t$ and $s$ time variables. As the initial abstraction operators are vacuous, the constructed proposition is fact-independent.

Typically, however, the antecedent of a subjunctive conditional, as it stands, is too weak to entail the consequent, i.e., there are world-times at which the if-clause is true and the then-clause is not. But the conditional is nevertheless acceptable because no such world-times can be found among those which are like the actual world at the present time in some obviously understood aspect. The intended antecedent is then best construed as containing a tacit clause to the effect that the aspect in question is as it is in fact.

The aspects come in various logical types. Often the aspect is simply the extension of a property or a relation. Sometimes it is the value of a numerical magnitude.

EXAMPLE 9. Suppose that Lindsay is a housewife who lives in Jonesville. Then it seems to make sense to assert

(9.1) If Lindsay was a film star, then a film star would live in Jonesville.

There are, of course many world-times at which Lindsay is a film star, yet no film stars live in Jonesville. But one who asserts (9.1) clearly confines himself to world-times at which people live where they now live in fact; hence charity demands to assume that the intended antecedent contains a tacit clause to this effect. In other words, the proposition the speaker has in mind is clearly one constructed by

$$\lambda w\lambda t.[\lambda v\lambda s.[F_{vs}X]\,\&.L_{vs}=L_{wt}] \prec \lambda v\lambda s.(\exists x).[F_{vs}x]\,\&.L_{vs}xJ$$

where $J$, $F$, and $L$ are, respectively, Jonesville, film-starhood, and the relation between people and places they live. Understood in this way, (9.1) is contingently true. It is tantamount, in fact, to the statement that Lindsay lives in Jonesville. And indeed, the only sensible way to argue for (9.1) is by claiming that Lindsay lives in Jonesville, and the only sensible way to dispute it is by disputing that claim.

Suppose, furthermore, that as it happens no film star lives in Jonesville. Then

(9.2) If Lindsay was a film star, Lindsay would not live in Jonesville

is also a fair comment to make, despite there being nothing inconceivable about Lindsay being a film star and still living in Jonesville. But someone who asserts (9.2) is obviously confining himself to world-times at which the tendency, on the part of film stars, to live in Jonesville is the same as it is in fact. Thus (9.2) is naturally understood as expressing the construction

$$\lambda w\lambda t.[\lambda v\lambda s.[F_{vs}X]\,\&.[\%_{vs}F\lambda w\lambda t\lambda x.L_{wt}xJ]=[\%_{wt}F\lambda w\lambda t\lambda x.L_{wt}xJ]] \prec \lambda v\lambda s.{\sim}L_{vs}XJ,$$

where % is the (world and time dependent) operation which takes any two properties to the percentage of individuals having the first property among individuals having the second. Understood in this way (9.2) is just a roundabout way of saying that no film stars live in Jonesville. It would be fatuous to invoke the fact that Lindsay lives in Jonesville as evidence against (9.2). The only sensible way to dispute (9.2) is by arguing that some film stars do live in Jonesville.

Often the tacitly fixed aspect is not just the momentary extension of some intension, but the history of the extension throughout the time scale. To illustrate, consider again

EXAMPLE 6 (continued). Someone who asserts (6.1) is clearly confining himself to worlds in which Lindsay's church-going behaviour is as it is in fact. If challenged, the speaker would certainly defend his statement by reference to Lindsay's church-going record, and those who disagreed would do so because of scepticism about that record. Hence (6.1) is naturally construed as expressing

$$\lambda w \lambda t.[\lambda v \lambda s.S_s \,\&\, .[\lambda t.C_{vt}X] = [\lambda t.C_{wt}X]] \prec \lambda v \lambda s.C_{vs}X,$$

where $C$ is the property of being in church and $S$ is the class of instants at which it is Sunday, 10 a.m. Understood in this way, (6.1) says no more and no less than that Lindsay is in church every Sunday morning at ten. (It is, of course, equally plausible to construe the conditional as tacitly fixing more comprehensive aspects, like the full history of Lindsay's whereabouts, the full history of everybody's whereabouts etc.)

Now supposing that a mass starts invariably on Sunday at 10 a.m., that it is now Wednesday 10 p.m., and that Lindsay is at home watching TV, the statement

(6.2) If it was Sunday 10 a.m., Lindsay would be missing a mass

seems also appropriate. This time it is of course not the chronology of Lindsay's churchgoing that is tacitly fixed; rather, it is Lindsay's momentary spacial location vis-à-vis the church and the chronology of mass celebration. Thus (6.2) expresses the construction

$$\lambda w \lambda t.[\lambda v \lambda s.S_s \,\&\, [[\lambda t.M_{vt}] = \lambda t.M_{wt}] \,\&\, .[C_{vs}X] = .C_{wt}X] \prec \lambda v \lambda s.M_{vs} \,\&\, {\sim} C_{vs}X,$$

where $M$ is the proposition that a mass is on. Understood in this way (6.2) conveys the information that Lindsay is currently not in church and a mass is in progress every Sunday at 10 a.m.

In the foregoing examples the intended counterfactual hypothesis was obtained by conjunctively adding some aspect-fixing clauses to the proposition explicitly named in the if-clause. This, however, is not what happens in all cases. Often the if-clause has to be construed as involving a bound variable which reappears in the tacit aspect-fixing clauses. In such a case the full intended hypothesis is not a conjunction but a quantification.[6]

EXAMPLE 10. Suppose that as a matter of fact, Lindsay has three sons. Then it seems true to say that

(10.1) If one of Lindsay's children was not male, Lindsay would still have a son.

The acceptability of this conditional clearly depends on the actual sex of Lindsay's children. Yet the sex of no particular child can be tacitly fixed, for the counterfactual assumption can be realized by altering the sex of any one of them. Clearly what is meant is that if one of the children was not male and the sex *of the others* remained as it is in fact, Lindsay would still have a son. Thus the sentence must be construed along the following lines:

$$\lambda w \lambda t.[\lambda v \lambda s.(\exists y).[H_{vs}yX] \,\&\, [\sim .M_{vs}y] \,\&\, [H_{vs} = H_{wt}] \,\&\, (\forall x).x \neq y \supset .\,[E_{vs}x] = .E_{wt}x] \prec \lambda v \lambda s(\exists x).[H_{vs}xX] \,\&\, .M_{vs}x$$

where $H$ is the relation between children and their parents and $E$ the operation which takes each person to his or her sex.

In many cases the range of worlds that the asserter of a conditional is tacitly confining himself to consists of worlds in which everything is as in the actual world except that certain antecedently undetermined events turn out differently, deflecting history from its actual course. To put this deflection idea in rigorous terms, let us use the following terminology. Where $C$ and $E$ are RBPs true in world $W$, $E$ is said to be *an etiological descendant of $C$ in $W$* if $C$, perhaps in conjunction with some other RBPs true in $W$, either necessitates $E$ in $W$ or contributes to $E$'s propensity to take place in $W$. The *etiological progeny* of a class $U$ of RBPs is the smallest class containing every member of $U$ and every etiological descendant in $W$ of any of its members.

Where $D$ is any class of RBPs, by the *W-projection* of $D$ – symbolically $W^D$ – we shall understand the class of members of $D$ or their negations which are true in $W$. Worlds $W_1$ and $W_2$ are said to be mutual *D-alternatives* if $W_1^D$ and $W_2^D$ are disjoint and for any RBP $E$, if $E$ is true in one of $W_1$, $W_2$ and false in the other, then it belongs either to the etiological progeny of $W_1^D$ in $W_1$ or to the etiological progeny of $W_2^D$ in $W_2$.

EXAMPLE 11. Consider the electric circuit shown in Figure 1. Element $e^0$ is a manual switch. For $i = 1, 2$, or 3, $e^i$ is a random switch: positive potential at its input at $t$ reappears at $t+1$ either at its left-hand output or at its right-hand output (but not both); we shall say that at $t+1$ $e^i$ *switches left* – symbolically, $L^{i,t+1}$ – or *switches right* – symbolically $R^{i,t+1}$. For $i = 4, ...,$ or 8, element $e^i$ is a bulb; positive potential at its input at $t$ has the (deterministic)

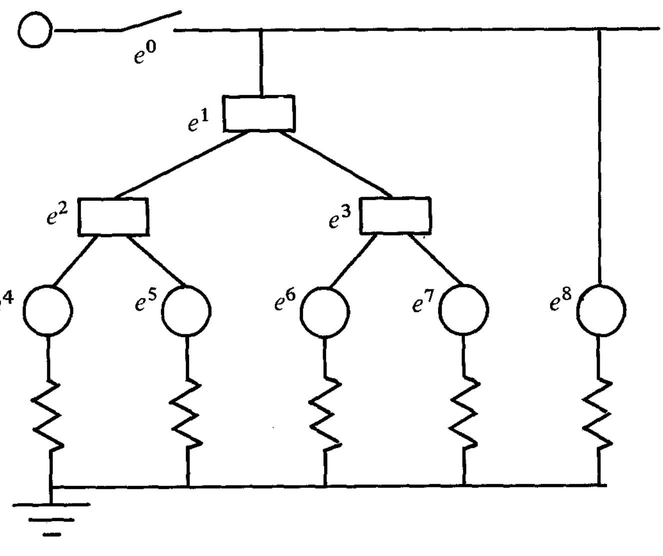


Fig. 1.

effect of $e^i$ *being on* at $t+1$ – symbolically, $O^{i,t+1}$. Suppose the actual history of the circuit is as follows. Switch $e^0$ was closed at time 0. As a result, $e^8$ came on at time 1 and the following sequence of events was set off: $L^{1,1}$, $L^{2,2}$, $O^{4,3}$. Clearly $L^{2,2}$ is an etiological descendant of $L^{1,1}$, and $O^{4,3}$ is an etiological descendant of $L^{2,2}$. Let $D$ be $\{L^{1,1}, R^{1,1}\}$ and $W_1$ the actual world. Then $W_1^D$ is $\{L^{1,1}, \sim R^{1,1}\}$ and its etiological progeny in $W_1$ contains each of $L^{1,1}, L^{2,2}$, and $O^{4,3}$. Now consider a world $W_2$ which is like $W_1$ until (and including) time 0, whereupon $e^1$ switches in the opposite direction and the following course of events ensues: $R^{1,1}$, $L^{3,2}$, $O^{6,3}$. $W_2^D$ is $\{\sim L^{1,1}, R^{1,1}\}$ and its etiological progeny in $W_2$ contains each of $R^{1,1}$, $L^{3,2}$, and $O^{6,3}$. As any RPB which does not have the same truth-value in $W_1$ and $W_2$ belongs either to the etiological progeny of $W_1^D$ in $W_1$ or to the etiological progeny of $W_2^D$ in $W_2$, $W_1$ and $W_2$ are mutual $D$-alternatives. In other words, $W_2$ is one of the worlds which might have been actualized had history been deflected from its actual course by $e^1$ switching right rather than left at time 1.

Now let $U$ be an arbitrary class of RBPs. World $W_2$ is said to be a *revision of* world $W_1$ *with respect to* $U$ if there is a subclass $D$ of $U$ such that $W_2$ is a $D$-alternative of $W_1$. The intuitive idea sketched above can now be restated rigorously thus: when asserting a conditional, the speaker often has a definite deflection set $U$ in mind and his intention is to confine himself to worlds which are revisions of the actual world w.r.t. $U$.

EXAMPLE 11 (continued). The conditional

(11.1) If $e^1$ had not switched left at time 1, $e^8$ would still have been on at time 1

will normally be considered as true. Yet an unsympathetic hearer can argue that if $e^1$ had not switched left at 1, it might have been because $e^0$ had not been closed at 0, in which case $e^8$ would not have been on at 1. A well-disposed hearer will gather that the speaker does not wish to consider revisions of the actual world which go as far back as time 0; he is confining himself to worlds which differ from the actual world solely as a result of $e^1$ switching, counterfactually, right at time 1. In other words, the charitable hearer will assume that the speaker has in mind the set $\{L^{1,1}, R^{1,1}\}$ and that he confines himself to revisions of the actual world with respect to that set. The hearer will thus construe (11.1) as expressing the construction

$$\lambda w \lambda t . [\lambda v \lambda s . R_{vs}^{1,1} \,\&\, . Q_{vs} [\lambda p . p = R^{1,1} \vee . p = L^{1,1}] w] < O^{8,1},$$

where $Q$ is the relation which holds in any world $W$ between classes of RBPs and worlds which are revisions of $W$ with respect to those classes.

For another application of the revision relation, consider again

EXAMPLE 5 (continued). The holocaust statement (5.1) is also naturally construed as a revision conditional, i.e., as expressing the construction

$$\lambda w \lambda t.[\lambda v \lambda s.Bu^{T}_{vs} \,\&\, .Q_{vs}[\lambda p.p = Bu^{T}]w] \prec Ho,$$

where *Ho* is the proposition that a holocaust has occurred since *T*, and *Q* is as in Example 11.

(5.1) affords a particularly good opportunity to compare the naturalness and intuitive appeal of the three three-parameter theories we have considered. Lewis' theory (as we have seen) directs us to conjecture that the speaker tacitly uses a similarity relation on which a world featuring a total devastation of the Earth bears a closer resemblance to the actual, no-holocaust world than does any no-holocaust world in which the button is pushed. In my experience at any rate, speakers who are happy to affirm (5.1) are invariably reluctant to confirm this conjecture when asked. Kratzer's theory (as we have seen) directs us to impute to the speaker, *inter alia*, a tacit assumption that it is not a fact that Hinkley shot Reagan in 1981. In my experience at least, speakers who willingly assent to (5.1) invariably refuse to own up to any assumption of this sort. On the Mill-Ramsey-Chisholm theory, all one needs to assume is that the speaker is tacitly restricting himself to situations which deviate from the actual situation only as far as the button-pushing and its etiological progeny is concerned. (Note that since the etiological progeny of a small-scale event may be huge, this allows for situations which are very unlike the actual one.) In other words, all we need to assume, on the tacit-premise theory, is that (5.1) is to be read as short for

> If Nixon had pushed the button at *T and things had otherwise been as they are in fact except for the etiological consequences of the pushing*, then a holocaust would have occurred.

It is my experience that speakers who assert (5.1) readily agree that this is exactly what they mean.

The last two examples might suggest that when a conditional is based on the revision idea, the relevant deflection set is always uniquely determined by the explicit antecedent. To see that this is not the case, consider again

EXAMPLE 11 (continued). Is the conditional

(11.2) If $e^4$ had not been on at time 3, then $e^5$ would have been on at 3 true? It can be plausibly argued that it is: if $e^4$ had not been on at 3, then $e^2$ would not have switched left at 2, in which case it would have switched right, making $e^5$ come on at 3. But an equally plausible argument seems to show that (11.2) is false: if $e^4$ had not been on at 3, it might have been because $e^1$ had switched right at 1, in which case neither $e^4$ nor $e^5$ would have been on at 3. Yet there is no paradox here, for the two arguments are based on two different construals of (11.2). The former argument takes the intended deflection set to be $\{L^{2,2}, R^{2,2}\}$ while the latter takes it to contain $\{L^{1,1}, R^{1,1}\}$.

When the antecedent, as in the last example, denies an event which has an actual etiological history, the history has to be revised. This raises the question, how far back may or must the revision go? Going no farther than is absolutely necessary to make room for the antecedent often leads to unacceptable results, as witness Example 4. But going all the way (down to the Big Bang) is obviously absurd. The only reasonable conclusion seems to be that in each particular case a limit to the back-tracking is tacitly set by the speaker in the form of a deflection set.

This is why no two-parameter theory of subjunctive conditionals would work even if all such conditionals were of the deflection kind, like those considered in Examples 4 and 11. But not all conditionals are of this kind anyway. Imagine, for example, that John tries and fails to lift a heavy box. Then the laws of gravity may justify us in saying that had the Earth's mass been only one tenth of what it is, John would have managed. Here we are unlikely to mean that John lifts the box in every world whose history is a revision of actual history, drastic enough to accommodate a ten times smaller Earth. Revisions on such a monumental scale would hardly preserve some of the facts that the truth of the conditional turns upon: John being around, exerting exactly the same force on the box, the box having its actual mass etc. Some theorists – e.g., Jackson (1977) and Lewis (1979) – have suggested that in cases like these the relevant worlds are worlds which are like the actual one until shortly before the time in question when a sudden *miracle* makes the antecedent true in defiance of the laws of nature. This, however, cannot be right, for nobody is prepared to maintain that if the mass of the Earth had been a tenth of what it is then it would have lost nine tenths of its mass in the last couple of minutes and the laws of nature would have been violated. There is a simpler and less fanciful explanation of why the box-lifting conditional is true.

The explanation is that this conditional has nothing to do at all with revising the history of the actual world, minimally or otherwise. What is meant is roughly this. The mass of the Earth, the mass of the box, the upward pull exerted by John on the box and the acceleration of the box relative to the Earth, form a system of magnitudes of which the last is, due to the actual laws of nature and the absence of other forces, functionally dependent on the other three. What the conditional statement says is to the effect that one tenth of the actual value of the first magnitude and the actual values of the second and third constitute an argument-triple at which the function takes a positive value.

## NOTES

1 The author is indebted to Angelika Kratzer, John Pollock, David Lewis, Alan Musgrave, Graham Oddie, and Victor Flynn, who have been kind enough to comment on an early draft of this paper. Kratzer's and Pollock's comments were especially helpful in finalizing the last two sections.

2 Imagine that event $E^T$ (taking place at $T$) was in fact caused, in conformity with a law of nature, by an infinite class $\{C^{T-1}, C^{T-0.5}, C^{T-0.25}, \ldots\}$ of antecedent and causally independent events, each of which was essential in bring $E^T$ about. Since for each member of the class there is a later one, none need be Blue-undercut by $\sim E^T$ and $\sim E^T$ may well be $\sim E^T$-incompatible.

3 An analogous counterexample to Lewis' theory was given in Tichý (1976).

4 In the theory of probability, the suppression of possible-world variables is known to have engendered a *prima facie* paradox. See Howson and Oddie (1979).

5 For a detailed exposition of the notation used in this and the following sections, see Tichý (1980a and 1980b).

6 I owe this point to Angelika Kratzer (private communication).

## BIBLIOGRAPHY

Blue, N. A.: 1981, 'A metalinguistic interpretation of counterfactual conditionals', Journal of Philosophical Logic 10, pp. 179–200.

Chisholm, Roderick M.: 1955, 'Law statements and counterfactual inference', Analysis 15, pp. 97–105.

Fine, Kit: 1975, Critical notice of Lewis (1973), Mind 84, pp. 451–458.

Goodman, Nelson: 1947, 'The problem of counterfactual conditionals', Journal of Philosophy 44, pp. 113–128.

Jackson, Frank: 1977, 'A causal theory of counterfactuals', Australasian Journal of Philosophy 55, pp. 3–21.

Howson, Colin and Oddie, Graham: 1979, 'Miller's so-called paradox of information', British Journal for the Philosophy of Science 30, pp. 253–278.

Kratzer, Angelika: 1981, 'Partition and revision: the semantics of counterfactuals', Journal of Philosophical Logic 10, pp. 201–216.

Lewis, David: 1973, Counterfactuals (Harvard University Press, Cambridge, Massachusetts).

Lewis, David: 1979, 'Counterfactual dependence and time's arrow', Noûs 13, pp. 455–475.
Mill, John Stuart: 1868, System of Logic (Longmans, London).
Pollock, John L.: 1976, Subjunctive Reasoning (D. Reidel, Dordrecht).
Pollock, John L.: 1981, 'A refined theory of counterfactuals', Journal of Philosophical Logic 10, pp. 239–266.
Ramsey, F. P.: 1931, 'General propositions and causality', in The Foundations of Mathematics (Humanities, New York).
Tichý, Pavel: 1976, 'A counterexample to the Stalnaker-Lewis analysis of counterfactuals', Philosophical Studies 29, pp. 271–273.
Tichý, Pavel: 1978, 'A new theory of subjunctive conditionals', Synthese 37, pp. 433–457.
Tichý, Pavel: 1980a, 'The logic of temporal discourse', Linguistics and Philosophy 3, pp. 343–369.
Tichý, Pavel: 1980b, 'The semantics of episodic verbs', Theoretical Linguistics 7, pp. 263–296.
Walters, R. S.: 1961, 'The problem of counterfactuals', Australasian Journal of Philosophy 39, pp. 30–46.

*Department of Philosophy,*
*University of Otago,*
*Box 56, Dunedin,*
*New Zealand*