PREJUDICES, PRESUPPOSITIONS AND THE THEORY OF COUNTERFACTUALS

Frank Veltman

Abstract.[0] Two theories of truth for counterfactual conditionals are developed. Both differ from current theories in that their explanations do not stop at the level of a comparative similarity relation of worlds. They differ from one another in that the first one traces comparative similarity, and with it counterfactual truth, back to the *beliefs* of a speaker, whereas the second one takes the *knowledge* of a speaker as its starting point. It is argued that the first theory is preferable to the second one, and that both these clear-cut 'coherence' theories are preferable to any, necessarily vague, 'correspondence' theory.

Turning, then, to questions of presupposition it is shown that a counterfactual conditional can be used appropriately only if its antecedent is believed to be false. This necessary condition for the conversational correctness of a counterfactual statement is discussed in detail, and an attempt is made to justify it in view of some maxims of conversation.

Logics are discussed in an appendix.

§ 0 Introduction

The following pages contain an essay on some of the pragmatic and some of the semantic peculiarities of some kinds of conditional sentences. To be more specific: I shall settle, formally, questions of presupposition and truth in connection with two conditional operators $\boxed{c}\!\!\rightarrow$, and $\langle c\rangle\!\!\rightarrow$ to be read successively as:

If it were the case that...., then it would be the case that....
If it were the case that...., then it might be the case that....

I hope to formally settle these questions such that the result may be regarded as a set-theoretic characterization of the semantic and pragmatic behaviour of counterfactual conditionals in natural language.

In setting out my views, I shall adopt without scruples the method of *possible worlds*. Moreover, in comparing my views with competing ones, I shall

restrict myself to the theories of those who, like myself, have employed this method.[1]

What are the truth conditions for statements 'If it were the case that A, then it would be the case that B'? Several[2] answers to this question have been offered in the past few years, yet none of them is completely satisfactory. Part of the problem is still to be solved: an important part, as some of the authors involved admit; a part which cannot be solved without falling into vagueness, as the others argue. To explicate this point I shall here compare the views of Stalnaker (1968) with those of Lewis (1973).

Loosely speaking[3], we can say that both their accounts are alike in that they take a counterfactual (A $\boxed{c}\!\!\rightarrow$B) to be true in a possible world **w** iff the consequent B is true in all worlds **w**' in which the antecedent A is true and *which in other respects resemble* **w** *as closely as possible*. Yet differences appear when we ask them for an elaboration of the italicized phrase.

Stalnaker takes for granted that exactly one such antecedent-world **w**' resembling **w** *in other respects* as closely as possible can be found, and he stresses the point that '*the context of utterance, the purpose of the assertion, and the beliefs of the speaker or his community*'[4] may make a difference to the particular world **w**' which has this property. In other words, counterfactuals are *pragmatically ambiguous*. Unfortunately he does not specify *how* these contextual features make that difference to this world **w**', but contents himself with specifying some formal constraints on the possible outcome.

I believe that Stalnaker is right on this point of pragmatic ambiguity, and I hope to supply full evidence for it in due course. Moreover, much of what follows can be seen as an attempt to bridge the gap between contextual features on the one hand, and that particular antecedent-world **w**' resembling **w** in other respects as closely as possible on the other. It will appear, however, that we must drop his assumption that there is always to be found exactly one such world **w**'. Actually, we cannot even be certain of finding one *set* of worlds **w**' in which the antecedent is true and which *in other respects* resemble **w** as closely as possible. We are up against several such sets, the members of which *in other respects* all resemble **w** as closely as possible, but different members of different such sets each *in their own other respects*. 'Each in their own other respects': perhaps we ought to stop talking in terms of resemblance as soon as this notion forces us to use this kind of language.

Lewis, for his part, is rather free and easy with the 'other respects' in which the antecedent-worlds w' should resemble w as closely as possible. He simply advises us to look whether the consequent is true in all antecedent-worlds w' which are most similar, *overall*, to w, and introduces a comparitive overall similarity relation between possible worlds in his formal set up to settle this affair. He admits that this notion is vague, but '*vague in a well-understood way....it is just the sort of vagueness that we must use to give a correct analysis of something that is itself undeniably vague*'[5]. In other words, counterfactuals are, in Lewis's view, *semantically vague* rather than pragmatically ambiguous. He admits that contextual features may be somewhat helpful in narrowing the range of this vagueness, but comparative similarity, and with it counterfactual truth, do not *wholly* depend on context. There are, indeed, only vague borderlines, but that does not mean there are no borderlines at all.

It is my purpose, however, to show that we can do without such a pre-established, vague relation of comparative overall similarity. Instead I shall take Popper's maxim '*similarity presupposes the adoption of a point of view*'[6] as my starting point, and show how something like a comparative similarity relation, though it be no comparative overall similarity relation, is *induced* in a set of possible worlds by what I shall call the prejudices and assumptions a speaker holds at a certain moment. I do of course realize that in doing so counterfactual truth will turn out to be speaker-dependent, and that I shall only barter vagueness for subjectivity. Nevertheless, my aim is to convince the reader that the latter is more 'natural', and much to be preferred to the former.

§ 1 Preliminaries

Giving a set-theoretic description of a phenomenon in natural language is like drawing a cartoon: leave out everything that is not important, exaggerate the few things left, and you may end with the most striking characterization of what is going on.

The first thing to be retained in my cartoon is a non-empty set $\mathbb{S}$ of so-called *speakers*. A second set we cann ill afford to do without is a non-empty set $\mathbb{W}$ of so-called *(possible) worlds*. Within our set-theoretical model these possible worlds will serve as the situation our speakers might discuss, hear of, think of, or come across in any other conceivable way. A special non-empty subset $\mathbb{H}$ of $\mathbb{W}$ will serve as the set of those situation that together make up the 'real' *history*. And, although we shall neglect problems of time and tense as much as possible, we shall sometimes heuristically refer to members of $\mathbb{H}$ as '*moments (of time)*'. Our speakers are supposed to utter their sentences at some of these

moments, and to this purpose we introduce a non-empty subset $\mathbb{C}$ of $\mathbb{S} \times \mathbb{H}$ of so-called *(utterance) contexts*. In doing so we can take ordered couples $\langle A, \langle s, w \rangle \rangle$, where A is a sentence and $\langle s, w \rangle \in \mathbb{C}$, to represent *assertions*, i.e. utterances of a certain sentence by a certain speaker at a certain moment.

I want to stress that, from the point of view adopted here, questions like "What is a possible world?" and "Do possible worlds really exist?" are rather misleading. A possible world is a *set*, and you may take anything that is a set in virtue of, for instance, the Zermelo-Fraenkel axioms as a possible world. Thus questions about the essence and existence of possible worlds become questions about the essence and existence of sets. These are questions to be asked by people *discussing* set-theory, not by people like us who are *using* it. To put it differently: we are talking about possible worlds in connection with natural language in the same way as physicists talk about electron orbits in connection with atomic spectra. Both electron orbits and possible worlds are nothing but theoretical entities *in a mathematical model* constructed in order to explain some ill-understood phenomenon.

Now you could, of course, restate the first question, and ask: "What kind of ill-understood things are *modeled* by these theoretical entities called possible worlds?" But then you should not be surprised that I can only but repeat the phrase 'any situation one might discuss, hear of, think of, or come across in any other conceivable way'. And in answer to a rewording of the second question, I would say that I believe it is precisely one of these situations which really exists at this very moment.

Of course, we must constantly see to it that our possible worlds do not, in our theoretical frame, behave differently from their alleged sources. But thus far nothing has been mentioned about possible worlds which cannot be said about any situation anyone of us might conceive. I have nowhere stated that any of our speakers $s \in \mathbb{S}$ is capable of conceiving, let alone describing, any possible world $w \in \mathbb{W}$ completely. Moreover, some of our speakers may (partially) conceive of a particular situation, and yet choose the wrong words to describe it; they may think of something that they cannot put into words at all. And, confronted with a situation which they did not conceive of themselves, they may not only choose the wrong words for something they rightly think to be the case, but also wrongly estimate what in fact is going on.

So much for the notion of possible worlds. A more complicated notion is the notion of a proposition. It is easy enough to stipulate that we will call every subset of $\mathbb{W}$ a proposition, but to explain how these propositions are to function within our formal cartoon, and how they are related to the 'propositions' philosophers talk about is a more difficult affair.

Remember the definition of an assertion: an assertion is a sentence uttered by a speaker $s \in \mathbb{S}$ in a context $\langle s,w \rangle \in \mathbb{C}$, in short, an ordered couple $\langle A,\langle s,w \rangle\rangle$. Let us call every triple $\langle A,\langle s,w \rangle,w' \rangle$ such that $\langle A,\langle s,w \rangle\rangle$ is an assertion and $w' \in \mathbb{W}$ a *judgment*, and let us reserve the word *statement* for judgments $\langle A,\langle s,w \rangle,w' \rangle$ such that $w = w'$. Judgments are to be taken as assertions *about* a certain possible world, and statements are to be taken as judgments in which a speaker utters a sentence about his own world. We can now state our program thus: we want to define, recursively, what it means for certain judgments to be true. That is, we want to define the truth value of judgments $\langle A,\langle s,w \rangle,w' \rangle$, with A a (complex) sentence of a language we still have to specify, in terms of the truth values of judgments $\langle B,\langle s,w \rangle,w'' \rangle$, with B a (proper) component sentence of A[7].

Notice that within my formal scheme truth values are predicated of ordered triples: sentences/uttered in a certain context/about a certain situation. And notice that within natural language truth and falsity vary with these variables as well. Of course, I would agree if you remarked that *in a primary sense we are always talking about the situation in which we are talking*[8], that you can say "that's true" or "that's false" as soon as I give you a sentence and the context in which this sentence is uttered. By way of example, take the sentence 'I am twenty-six years old', and yourself as the utterer at the moment of your reading this paper. I agree that it is 'natural' to take the resulting assertion as an assertion about the moment that you are reading this, and not as an assertion about a moment four years ago. But assertions about other situations are often made *implicitly*: if you were to claim that the (complex) sentence 'I was twenty-six four years ago', uttered by you at the very moment of your reading this paper (and 'naturally' about this moment as well), is true, then you would implicitly claim that the (simpler) sentence 'I am twenty-six', uttered by you at this moment about the moment four years ago, is true. Formulating truth conditions is a matter of making explicit what a statement implicitly claims to be the case. That is where assertions about other worlds enter the scene. That is why we are interested in truth conditions for judgments in general, and not only in truth conditions for statements.

We may now indicate how the above-mentioned *propositions* will function within our formal frame. I hope that it is clear that we will have amply fulfilled the desire for truth conditions for judgments $\langle A,\langle s,w \rangle,w' \rangle$ as soon as we have defined, for each assertion $\langle A,\langle s,w \rangle\rangle$, the *set of possible worlds* w''

such that the judgment $\langle A,\langle s,w \rangle,w'' \rangle$ is true. If we make this our strategy, then sets of possible worlds are going to play a prominent role in our formal set-up, as we have to construct our truth definition in such a way that the set of possible worlds w'' such that a (complex) judgment $\langle A,\langle s,w \rangle,w'' \rangle$ is true can be defined in terms of *other* sets, each containing the possible worlds about which a (proper) component sentence of A can be truthfully asserted. Therefore, it becomes desirable to introduce a special notation '$[\![A]\!]_{\langle s,w \rangle}$' for the set of worlds w' such that the judgment $\langle A\langle s,w \rangle,w' \rangle$ is true. Read '$[\![A]\!]_{\langle s,w \rangle}$' as '*the proposition expressed* by the assertion $\langle A,\langle s,w \rangle\rangle$'; then, instead of our previous clause 'the judgment $\langle A,\langle s,w \rangle,w' \rangle$ is true', you may write '$w' \in [\![A]\!]_{\langle s,w \rangle}$', and read this as 'the proposition $[\![A]\!]_{\langle s,w \rangle}$ *holds* for w' '.

The main reason for introducing the rather abstract notion of a proposition in addition to the much more familiar notion of a judgment is this: we can now talk about what an assertion expresses without reference to the assertion that expresses it. Propositions are to be understood as non-linguistic entities that may be expressed by many speakers at many moments, using many different sentences. Sometimes a speaker may not know *how* to express something which he nevertheless wants to put into words, and sometimes there may be no way of putting a proposition into words at all. Propositions, so understood, are very abstract entities indeed, exhibiting many of the properties traditionally ascribed to them. But notice that our propositions are not supposed to capture the meaning of an assertion; they only capture truth conditions. They represent the meaning of an assertion just as far as knowing the meaning of an assertion boils down to knowing the kind of conceivable situations about which the assertion can truly be made.

§ 2 Prejudices and Assumptions

Up to now our set-theoretical cartoon consists of speakers, possible worlds, a 'real' history, and a set of contexts. It is time to introduce the key notion of this article: we are going to supply our speakers with *opinions*. We introduce a function $\mathbb{O}$ that assigns to each context $\langle s,w \rangle$ and each possible world w' a set of propositions $\mathbb{O}^{w'}_{\langle s,w \rangle}$. If $p \in \mathbb{O}^{w'}_{\langle s,w \rangle}$ then we shall call p an opinion (belief, view) of the speaker s at the moment w on the world w'. A speaker may concurrently have different opinions on different worlds. Different speakers may simultaniously differ in their opinions on the same world. And the same speaker may have different opinions on the same world at different moments. I hope that this is adequate as an explanation of the different arguments of $\mathbb{O}$.

We don't suppose that our speakers have opinions about all possible worlds

$\mathbf{w}' \in \mathbb{W}$: it may be the case that $\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$ is empty. On the other hand, we neither suppose our speakers to have opinions only about the world in which they are speaking, or only about the 'real' worlds $\mathbf{w}' \in \mathbb{H}$; they are entitled to hold views of their own on worlds described in fiction or depicted on the screen, to say nothing about the worlds they dream up themselves as well.

Let us for a moment concentrate on one context $<\mathbf{s},\mathbf{w}>$, one world $\mathbf{w}'$, and the corresponding set $\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$. Let '$|\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}|$' denote the *intersection of* $\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$ *within* $\mathbb{W}$, i.e. the set of worlds $\mathbf{w}'' \in \mathbb{W}$ such that $\mathbf{w}'' \in \mathbf{p}$ for all $\mathbf{p} \in \mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$. Notice that $|\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}| = \mathbb{W}$ if $\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>} = \emptyset$. We shall call the elements of $\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$ *mutually incompatible* iff $|\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}| = \emptyset$.

There is a difference between on the one hand propositions $\mathbf{p}$ such that $|\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}| \subset \mathbf{p}$ but $\mathbf{p} \notin \mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$, and on the other hand propositions $\mathbf{q}$ such that $\mathbf{q} \in |\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}|$. Both kinds of propositions hold for all $\mathbf{w}'' \in |\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}|$, i.e. they hold for all worlds $\mathbf{w}''$ that answer to the idea which $\mathbf{s}$ has formed of $\mathbf{w}'$. But a proposition $\mathbf{q} \in \mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$ is so to speak constitutive for this idea, and we may expect the $\mathbf{s}$ in question to present and defend it as a truth about $\mathbf{w}'$, whereas it may take his opponents quite a while to show $\mathbf{s}$ that he does hold a proposition $\mathbf{p}$ – $|\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}| \subset \mathbf{p}$ but $\mathbf{p} \notin \mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$ – at least *implicitly* in virtue of the opinions that he *explicitly* defends. We shall call a proposition $\mathbf{p}$ such that $|\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}| \subset \mathbf{p}$ an *implicit opinion* and reserve the phrase 'opinions' (rather than 'explicit opinions') for the elements $\mathbf{p} \in \mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$.

Our speakers are not supposed to adhere to all their views in the same way. We shall split up every $\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$ into two parts: *prejudices and assumptions*. Within our frame, prejudices will function as the opinions a speaker in a particular context is not willing to give up, come what may, when he is discussing matters pertaining to any of the 'real' worlds $\mathbf{w}'' \in \mathbb{H}$, i.o.w., prejudices are to function as propositions a speaker considers to be *laws of nature*.[9]

In order to settle this, we can take a function $\mathbb{P}$ from $\mathbb{C} \times \mathbb{W}$ into the powerset of the powerset of $\mathbb{W}$ such that (i) $\mathbb{P}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>} = \emptyset$ if $\mathbf{w}' \notin \mathbb{H}$; (ii) $\mathbb{P}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>} = \mathbb{P}^{\mathbf{w}''}_{<\mathbf{s},\mathbf{w}>}$ if $\mathbf{w}',\mathbf{w}'' \in \mathbb{H}$; and (iii) $\mathbb{P}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>} \subset \mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$.

The following remarks are in order: first, we take $\mathbb{P}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>} = \emptyset$ if $\mathbf{w}' \notin \mathbb{H}$ trusting that our speakers are willing to cast off their prejudices if they are confronted with "just fairy-tales" . Secondly, all prejudices are opinions that a speaker in a certain context holds on all worlds $\mathbf{w}' \in \mathbb{H}$. But not *vice versa*: not all opinions that a speaker holds in a certain context about all real worlds $\mathbf{w}' \in \mathbb{H}$ have to be prejudices. The reason for this: I want prejudices to function in our cartoon as the laws of nature a speaker is *using* in a certain context; but, aside from using them, it is a good habit to occasionally *test* your theories as well. In such a context you are trying to bring about a situation in which that law of nature does not hold. You have asked yourself "What would be the case, if this law were not to hold?", and in looking for an answer to that question you may have taken some of your other theories for granted, but not the one you are willing to test. On the other hand, it will do no harm if you nevertheless continue to *assume* that the law in question holds for all $\mathbf{w}' \in \mathbb{H}$.

The reader may not appreciate my jumping together the most distinguished scientific theories with the vilest prejudices. I admit immediately that, from a *methodological* point of view, this is unfair, if not insulting. But, as stated before, I am interested in laws in use, and not in the question of how they came into use. I want each $\mathbb{P}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$ to function as the set of propositions that constitute a speaker's theoretical *field of view* at a certain moment. And then, from a *logical* point of view, there is no distinction to be made between a thorough scholar, willing to change the spectacles that fix *his* field of view as soon as we offer him better ones, and a narrow-minded what-do-you-call'em, who will keep wearing the very same blinkers for the rest of his life.

At this point we can put our considerations together, and state that our set-theoretical cartoon for its nonlinguistic part must consist of a so-called *frame*:

Def. 1 A frame $\mathbb{F}$ is a sextuple $<\mathbb{W},\mathbb{S},\mathbb{H},\mathbb{C},\mathbb{O},\mathbb{P}>$ such that $\mathbb{W} \neq \emptyset$; $\mathbb{S} \neq \emptyset$; $\emptyset \neq \mathbb{H} \subset \mathbb{W}$; $\emptyset \neq \mathbb{C} \subset \mathbb{S} \times \mathbb{H}$; $\mathbb{O}$ is a function; $\mathrm{dom}(\mathbb{O}) = \mathbb{C} \times \mathbb{W}$; $\mathrm{ran}(\mathbb{O}) \subset \mathrm{pow}(\mathrm{pow}(\mathbb{W}))$; $\mathbb{P}$ is a function; $\mathrm{dom}(\mathbb{P}) = \mathbb{C} \times \mathbb{W}$; $\mathrm{ran}(\mathbb{P}) \subset \mathrm{pow}(\mathrm{pow}(\mathbb{W}))$; for all $\mathbf{c} \in \mathbb{C}$ and $\mathbf{w} \notin \mathbb{H}$: $\mathbb{P}^{\mathbf{w}}_{\mathbf{c}} = \emptyset$; for all $\mathbf{c} \in \mathbb{C}$ and $\mathbf{w},\mathbf{w}' \in \mathbb{H}$: $\mathbb{P}^{\mathbf{w}}_{\mathbf{c}} = \mathbb{P}^{\mathbf{w}'}_{\mathbf{c}}$; for all $\mathbf{c} \in \mathbb{C}$ and $\mathbf{w} \in \mathbb{W}$: $\mathbb{P}^{\mathbf{w}}_{\mathbf{c}} \subset \mathbb{O}^{\mathbf{w}}_{\mathbf{c}}$.

The elements of $\mathbb{W}$ are called '(possible) worlds'; the elements of $\mathbb{S}$ 'speakers'; $\mathbb{H}$ is called 'history'; the elements of $\mathbb{C}$ 'contexts'; if $\mathbf{c} = <\mathbf{s},\mathbf{w}> \in \mathbb{C}$ and $\mathbf{w}' \in \mathbb{W}$ then the elements of $\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$ are called 'opinions of $\mathbf{s}$ in $<\mathbf{s},\mathbf{w}>$ on $\mathbf{w}'$'; the elements of every $\mathbb{P}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>}$ are called 'prejudices of $\mathbf{s}$ in $<\mathbf{s},\mathbf{w}>$ about $\mathbf{w}'$'; the elements of every $(\mathbb{O}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>} \sim \mathbb{P}^{\mathbf{w}'}_{<\mathbf{s},\mathbf{w}>})$ are called 'assumptions of $\mathbf{s}$ in $<\mathbf{s},\mathbf{w}>$ about $\mathbf{w}'$'. Subsets of $\mathbb{W}$ are called 'propositions'.

Our speakers are still in need of a *language*, and for our purpose any of the languages answering to the following definition will do:

<u>Def.2</u> A language $\mathcal{L}$ is a fourtuple $\langle AT,UN,BI,SEN\rangle$ such that
- $AT \neq \emptyset$; the elements of AT are called 'atomic sentences'.
- UN contains seven elements: $\neg$, $\Box$, $\Diamond$, $\boxtimes$, $\diamondplus$, *may*, *must*; the elements of UN are called 'unary operators'.
- BI contains six so-called 'binary operators': &, v, $\supset$, $\equiv$, $\boxed{c}\!\!\rightarrow$, and $\Diamond\!\!\!\!\!\;c\rightarrow$.

AT, UN and BI are pairwise disjoint sets. Let) and (be two sets distinct from the elements of $AT \cup UN \cup BI$. Let $\mathcal{E}$ be the set consisting of the finite concatenations of elements of $AT \cup UN \cup BI \cup \{(,)\}$. (We shall freely indicate concatenation by juxtaposition.)
- SEN, the set of so-called 'sentences', is the smallest subset of $\mathcal{E}$ such that (i) $AT \subset SEN$; (ii) if $\ominus \in UN$ and $A \in SEN$, then $\ominus A \in SEN$; (iii) if $\oplus \in BI$ and $B,C \in SEN$, then $(B \oplus C) \in SEN$.

I have already indicated that we will construct our truth definition as a function mapping assertions on propositions. In definition 3 truth conditions are presented for sentences not containing any of our special conditional operators.

<u>Def.3</u> Let $\mathbb{F}$ be a frame and $\mathcal{L}$ be a language. An *interpretation* of $\mathcal{L}$ into $\mathbb{F}$ is a function $[\![\;]\!]$, where $dom([\![\;]\!]) = SEN \times \mathbb{C}$ and $ran([\![\;]\!]) \subset pow(\mathbb{W})$, such that for every $A \in SEN$ and $\langle s,w\rangle \in \mathbb{C}$:
- if $A \in AT$ then $[\![A]\!]_{\langle s,w\rangle} = [\![A]\!]_{\langle s',w'\rangle}$ for every $\langle s',w'\rangle \in \mathbb{C}$
- if $A = \neg B$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $w' \notin [\![B]\!]_{\langle s,w\rangle}$
- if $A = (B\&C)$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $w' \in [\![B]\!]_{\langle s,w\rangle}$ and $w' \in [\![C]\!]_{\langle s,w\rangle}$
- if $A = (BvC)$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $w' \in [\![B]\!]_{\langle s,w\rangle}$ or $w' \in [\![C]\!]_{\langle s,w\rangle}$
- if $A = (B \supset C)$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $w' \notin [\![B]\!]_{\langle s,w\rangle}$ or $w' \in [\![C]\!]_{\langle s,w\rangle}$
- if $A = (B \equiv C)$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $w' \in [\![B]\!]_{\langle s,w\rangle}$ iff $w' \in [\![C]\!]_{\langle s,w\rangle}$
- if $A = \Box B$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $w'' \in [\![B]\!]_{\langle s,w\rangle}$ for every $w'' \in \mathbb{W}$
- if $A = \Diamond B$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $w'' \in [\![B]\!]_{\langle s,w\rangle}$ for some $w'' \in \mathbb{W}$
- if $A = \boxtimes B$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $|\mathbb{P}^{w'}_{\langle s,w\rangle}| \subset [\![B]\!]_{\langle s,w\rangle}$
- if $A = \diamondplus B$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $|\mathbb{P}^{w'}_{\langle s,w\rangle}| \cap [\![B]\!]_{\langle s,w\rangle} \neq \emptyset$
- if $A = mustB$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $|\mathbb{O}^{w'}_{\langle s,w\rangle}| \subset [\![B]\!]_{\langle s,w\rangle}$
- if $A = mayB$ then $w' \in [\![A]\!]_{\langle s,w\rangle}$ iff $|\mathbb{O}^{w'}_{\langle s,w\rangle}| \cap [\![B]\!]_{\langle s,w\rangle} \neq \emptyset$

I shall here only expound on the merits of the fourth and the last four lines of this definition. Hopefully, the reader is familiar with the remaining ones.

Let us take the fourth line first: it says that, given an atomic sentence A, all our speakers in all contexts express the same proposition by A. I must admit

that on this point our cartoon does give too rosy a picture of what is going on in reality. Discussions carried on in a natural language often end in a merely verbal dispute: a seeming difference in opinion turns out to be nothing but a disagreement on the meaning of some term or another. I want to avoid these things happening in our cartoon, as these kinds of discussions do not interest me here. The fourth line accounts for that distortion[10].

The proposed truth conditions for *must*B - read:'it must be the case that B' - and *may*B - read: 'it is maybe the case that B' - are basically the same as the ones proposed by Groenendijk and Stokhof (1975)[11]. *may*B uttered by a speaker s in a context $\langle s,w\rangle$ about a world w' is true iff there is a world w'' for which (i) all the opinions of s in $\langle s,w\rangle$ on w' hold, and (ii) $[\![B]\!]_{\langle s,w\rangle}$ holds. *must*B uttered by a speaker s in a context $\langle s,w\rangle$ about a world w' is true iff $[\![B]\!]_{\langle s,w\rangle}$ holds for every w'' for which all the opinions of s in the context $\langle s,w\rangle$ on the world w' hold.

The truth condition for $\diamondplus$B is somewhat weaker than the one proposed for *may*B. It is therefore tempting to read $\diamondplus$B as 'it might be the case that B'. Notice that a judgment $\langle \diamondplus B,\langle s,w\rangle,w'\rangle$ is true if there exists, in the light of the theories that s in $\langle s,w\rangle$ holds on w', a possible world for which $[\![B]\!]_{\langle s,w\rangle}$ holds. If s happens to hold no theories on w', for instance if $w' \notin \mathbb{H}$, then $\langle \diamondplus A,\langle s,w\rangle,w'\rangle$ is true if and only if $\langle \Diamond A,\langle s,w\rangle,w'\rangle$ is true: in the worlds of fiction everything *logically* possible might happen. Likewise: if we restrict ourselves to worlds $w' \in \mathbb{H}$ then a judgment $\langle \boxtimes B,\langle s,w\rangle,w'\rangle$ is true iff $[\![B]\!]_{\langle s,w\rangle}$ holds in all worlds w'' that are *physically* possible according to s in $\langle s,w\rangle$. But if $w' \notin \mathbb{H}$ then $w' \in [\![\boxtimes B]\!]_{\langle s,w\rangle}$ iff $w' \in [\![\Box B]\!]_{\langle s,w\rangle}$. I propose to read $\boxtimes$B as 'it should be the case that B' or equivalently(!) as 'it's a natural law that B', but I want in passing to note that most of us start showing off our learning with a simple 'It's an accomplished fact that....'

I can only think of one serious argument against the proposed truth conditions: I can imagine an opponent who rightly notes that judgments $\langle mustA,\langle s,w\rangle,w'\rangle$ and $\langle \boxtimes A,\langle s,w\rangle,w'\rangle$ can be true, whereas *in fact* $[\![A]\!]_{\langle s,w\rangle}$ does not hold for w', and that judgments $\langle mayA,\langle s,w\rangle,w'\rangle$ and $\langle \diamondplus A,\langle s,w\rangle,w'\rangle$ can be false, whereas *in fact* $[\![A]\!]_{\langle s,w\rangle}$ does hold for w'[12]. More generally: opinions that do not *in fact* hold, mere beliefs, may make a difference to the truth value of *must*-, *may*-, $\boxtimes$- and $\diamondplus$-judgments. Our opponent is really worried by this, and he presents us with the following, puzzling case with respect to *must*-sentences (hopefully the reader can think of the analogues with respect to $\boxtimes$-, $\diamondplus$- and *may*-sentences): suppose that the statement $\langle mustA,\langle s,w\rangle,w\rangle$ is true on behalf of prejudices and assumptions that do not in fact hold for w. Suppose, furthermore, that at a moment w' later than w s discovers that $[\![A]\!]_{\langle s,w\rangle}$ did not in fact hold for w. Don't you think that s in the context $\langle s,w'\rangle$ would conclude that his statement $\langle \textit{must}A,\langle s,w\rangle,w\rangle$ was false after all?

My answer to this question runs as follows (and the reader may fill in my answer to the analogous questions): No! The only thing s has to conclude in the context $\langle s,w'\rangle$ is that the statement $\langle A,\langle s,w\rangle,w\rangle$ would have been false, and that the best thing he can now do is *change his mind* with respect to w. Then $\langle \textit{must}A,\langle s,w'\rangle,w\rangle$ will become false, but s may still take $\langle \textit{must}A,\langle s,w\rangle,w\rangle$ to have been true in that particular context $\langle s,w\rangle$. (Moreover, he still may truly say, at w', with respect to w, that *it should have been* the case that A. See below.)

My opponent, however, takes the other horn of the dilemma; he advocates the following correction of our truth conditions: Let ${}^{*}\mathbb{O}^{w'}_{\langle s,w\rangle}$ be the set of all $p \in \mathbb{O}^{w'}_{\langle s,w\rangle}$ such that (i) $w' \in p$ and (ii) if $p \in \mathbb{P}^{w'}_{\langle s,w\rangle}$ then $\mathbb{H} \subset p$. Let ${}^{*}\mathbb{P}^{w'}_{\langle s,w\rangle}$ be the set of all $p \in \mathbb{P}^{w'}_{\langle s,w\rangle}$ such that $\mathbb{H} \subset p$. Define an *interpretation* * to be a function $[\![\ ]\!]^{*}$ just like $[\![\ ]\!]$, except for the last four lines in the definition of $[\![\ ]\!]$, where every occurrence of '$\mathbb{O}^{w'}_{\langle s,w\rangle}$' has to be replaced by an occurrence of '${}^{*}\mathbb{O}^{w'}_{\langle s,w\rangle}$', and every occurrence of '$\mathbb{P}^{w'}_{\langle s,w\rangle}$' by an occurrence of '${}^{*}\mathbb{P}^{w'}_{\langle s,w\rangle}$'. In doing so it will never happen that $\langle \textit{must}A,\langle s,w\rangle,w'\rangle$ or $\langle \boxtimes A,\langle s,w\rangle,w'\rangle$ are true, whereas in fact $\langle A,\langle s,w\rangle,w'\rangle$ is false, and that $\langle \textit{may}A,\langle s,w\rangle,w'\rangle$ or $\langle \oplus A,\langle s,w\rangle,w'\rangle$ are false, whereas in fact $\langle A,\langle s,w\rangle,w'\rangle$ is true: erroneous beliefs do not interfere in matters of truth and falsity any longer.

I admit, of course, that this is another way of solving the puzzle(s). But I wonder why my opponent has resorted to such a weighty manoeuvre; perhaps he has missed the first, somewhat subtler way of dealing with it by overlooking the fact that *must*-sentences (as well as $\boxtimes$-, $\oplus$- and *may*-sentences) are pragmatically ambiguous. In any case, I am sure he has caused our cartoon to give a wrong picture of the logical behaviour of all the judgments at issue.

Firstly: According to my opponent a judgment $\langle A,\langle s,w\rangle,w'\rangle$ is true if $\langle \textit{must}A,\langle s,w\rangle,w'\rangle$ is true. But, as Groenendijk and Stokhof (1975) rightly note: 'A statement like (i) *John must be at home* is weaker than (ii) *John is at home*.... (i) is used when, given a certain amount of information, it is almost certain that the situation described by (ii) does in fact occur. For instance, if someone has the information that John turns out the light before going out at night, and one evening he passes John's house and sees that the lights are on, then he will use (i)....(ii) on the other hand is used when someone has seen for himself that John is at home'[13].

Secondly: According to my opponent statements $\langle \boxtimes A,\langle s,w\rangle,w\rangle$ are true only if $\mathbb{H} \subset [\![A]\!]_{\langle s,w\rangle}$, and the reader may consider it to be a point in his favour that statements 'it's a natural law that A' are true only if A can be truthfully asserted about all real worlds $w' \in \mathbb{H}$. But in my opinion my opponent is missing the point about the lawfulness of lawful statements. A law *in use* is only a *bet* that a situation that does not obey this law will never occur. Our theories above all frame what we *expect* and do not *expect* of reality; they set out, as stated before, our *field of view*. Therefore, I take a statement 'It's a law that A' to be true if the speaker in question - rightly or wrongly - does not *reckon with* a situation for which $[\![A]\!]_{\langle s,w\rangle}$ does not hold. I want to account for the fact that our theories serve to guide us, and I think they do so quite independently of the question whether they are trustworthy or not. No doubt some of our theories are, from a methodological point of view, better justified than others. By all means, there are favourite horses in our race, and it is better to lay your bet on them than on the crippled ones. I even admit that within science a way of gambling has been developed that, if not profitable, at least reduces the loss to a minimum. Nevertheless, all kinds of bets are being laid out, both stupid and clever ones, and I would not give an adequate picture of this betting if I were to restrict myself to the clever bets or, even worse, to the winning ones. And that is what my opponent is doing: he is taking the winning bets to be the only real ones.

Let me leave this matter by giving my opponent a little puzzle in return: what are the truth conditions for statements like 'It should have been raining now' or 'They should have been arrived by now' or in general for judgments $\langle$*it should have been that* $A,\langle s,w\rangle,w'\rangle$. To my mind these kinds of judgments are often true in contexts where the speaker is perfectly aware of *the fact* that $w' \notin [\![A]\!]_{\langle s,w\rangle}$. And in the spirit of definition 3 I can define:
$w' \in [\![\textit{it should have been that } A]\!]_{\langle s,w\rangle}$ iff $|\mathbb{O}^{w'}_{\langle s,w''\rangle}| \subset [\![A]\!]_{\langle s,w\rangle}$ for some w'' earlier than w.[14]

However, if I place myself in my opponent's position, then I see no way of defining truth conditions for the judgments in question, and the reader does not have to go into the details of my proposal to see the problem: if not opinions that *did* not in fact hold, what else can determine the truth value of the statement 'It should have been raining now' at a moment when the speaker in question is perfectly aware of the fact that it does not rain? In my analysis this speaker, who truthfully states 'It should have been raining now' in the afternoon, is the same as the one who truthfully stated 'It must be raining this afternoon' in the morning. According to my opponent this speaker was wrong in the morning;

and it seems to me that my opponent must on the same ground - viz., that a judgment cannot owe its *truth* to an erroneous belief - conclude that this speaker is wrong in the afternoon as well. In that case I see no way in which judgments $\langle$*it should have been that* $A, \langle \mathbf{s},\mathbf{w}\rangle, \mathbf{w}'\rangle$ can ever be true if they are made in the circumstances for which they seem to have been invented.

Meanwhile, the reader may disagree both with me and with my opponent. Perhaps he will blame me for letting truth depend on the beliefs a speaker holds at a certain time, whereas he will blame my opponent for letting truth depend on a speaker's knowledge: 'Why don't both of you just stick to the facts?'

I think my first opponent can meet this demand up to a certain point. He may admit that an assertion can express a law of nature, even it that law has not been discovered as yet; he may be willing to repair his truth definition such that laws would be laws independently of the question whether they are known to be laws or not: $\mathbf{w}' \in [\![\boxtimes A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$ iff $\mathbb{H} \subset [\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$; likewise $\mathbf{w}' \in [\![\oplus A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$ iff $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle} \cap \mathbb{H} \neq \emptyset$.[15]

For my part such a 'solution' is out of the question, as I do not believe that there exist such things as laws which are discovered first, and known afterwards. Laws are hypotheses. They are *made* and not discovered: lawfulness is a matter of *imputation*[16]. I have already criticized my opponent for underestimating the fact that we often impute this lawfulness to a proposition which does not deserve it. Now I must add that this second manoeuvre shows that he does not realize that even the propositions which deserve it, must get it in the first place. To continue our metaphor: There are winning horses in the race, but you are not going to be paid if you have not laid your bet on them.

In any case, there are no other concessions to be made. Not even for my opponent. For on which facts should the truth of statements 'it may be the case that A' and 'it must be the case that A' depend? On facts in other possible worlds? Which worlds? Worlds that resemble the actual world in some important respects? Which respects?

§ 3 Counterfactual 'Truth'

Let $\mathbb{F}$ be a frame and $\mathcal{L}$ be a language. Take $\langle \mathbf{s},\mathbf{w}\rangle \in \mathbb{C}$. Take $A, B \in AT$. Let $\mathbf{w}' \notin [\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$ and $\mathbf{w}' \notin [\![B]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$. Moreover, let $[\![\neg A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle} \in \mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$ and $[\![\neg B]\!]_{\langle \mathbf{s},\mathbf{w}\rangle} \in \mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$. Assume $|\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}| \neq \emptyset$.

I create this situation only to avoid any questions which may arise on presuppositions, now that I have this speaker $\mathbf{s}$ at this time $\mathbf{w}$ stating that B would have been the case at $\mathbf{w}'$ if A had been the case there: what are the sufficient and necessary conditions for this counterfactual judgment $\langle (A \boxed{c}\!\!\rightarrow B), \langle \mathbf{s},\mathbf{w}\rangle, \mathbf{w}'\rangle$ to be true?

Let the following serve to guide us: $\langle (A \boxed{c}\!\!\rightarrow B), \langle \mathbf{s},\mathbf{w}\rangle, \mathbf{w}'\rangle$ is true iff the judgments $\langle B, \langle \mathbf{s},\mathbf{w}\rangle, \mathbf{w}''\rangle$ are true at all worlds $\mathbf{w}''$ at which $\langle A, \langle \mathbf{s},\mathbf{w}\rangle, \mathbf{w}''\rangle$ is true, and which *in other respects resemble* $\mathbf{w}'$ *as closely as possible in the opinion of* $\mathbf{s}$ *at* $\mathbf{w}$.

And this is our goal: to clarify the italicized phrase; how do things *become* similar for someone?[17].

We first define in general:

<u>Def.4</u> Let $\mathbb{F}$ be a frame and $\mathbf{p} \subset \mathbb{W}$. A $\mathbf{p}$-*accommodating subset* of $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$ is a set $\mathbb{Q}$ such that (i) $\mathbb{P}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle} \subset \mathbb{Q} \subset \mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$ and (ii) $|\mathbb{Q}| \cap \mathbf{p} \neq \emptyset$.

Imagine what an $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$-accommodating subset $\mathbb{Q}$ of the particular $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$ of our example would be: a subset $\mathbb{Q}$ of the opinions of $\mathbf{s}$ in $\langle \mathbf{s},\mathbf{w}\rangle$ on $\mathbf{w}'$ that includes all prejudices. Moreover, there are worlds $\mathbf{w}''$ among the elements of $|\mathbb{Q}|$ such that $\langle A, \langle \mathbf{s},\mathbf{w}\rangle, \mathbf{w}''\rangle$ is true. Hence the worlds $\mathbf{w}'' \in ([\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle} \cap |\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}|)$ are worlds at which $\langle A, \langle \mathbf{s},\mathbf{w}\rangle, \mathbf{w}''\rangle$ is true and which resemble $\mathbf{w}'$ at least in *some* other respects in the opinion of $\mathbf{s}$ at $\mathbf{w}$.

Notice that in general $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$ has no $\mathbf{p}$-accommodating subsets iff $|\mathbb{P}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}| \cap \mathbf{p} = \emptyset$, that is to say that in our example there are no worlds $\mathbf{w}'' \in [\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$ which are a bit like $\mathbf{w}'$, if $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$ is incompatible with the prejudices of $\mathbf{s}$ in $\langle \mathbf{s},\mathbf{w}\rangle$ on $\mathbf{w}'$[18].

But let our example be such that $|\mathbb{P}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}| \cap [\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle} \neq \emptyset$.

<u>Def.5</u> Let $\mathbb{F}$ be a frame and $\mathbf{p} \subset \mathbb{W}$. A $\mathbf{p}$-accommodating subset $\mathbb{M}$ of $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$ is *maximal* iff there is no $\mathbf{p}$-accommodating subset $\mathbb{Q}$ of $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$ such that $\mathbb{M} \subset \mathbb{Q}$ and not $\mathbb{Q} \subset \mathbb{M}$.

Imagine, with respect to our example, a maximal $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$-accommodating subset $\mathbb{M}$ of $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$. What else can the worlds in $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle} \cap |\mathbb{M}|$ be but worlds at which $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$ holds and which in other respects resemble $\mathbf{w}'$ *as closely as possible* in the opinion of $\mathbf{s}$ at $\mathbf{w}$? Every set $\mathbb{M} \cup \{\mathbf{q}\}$ - $\mathbf{q} \in \mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$, $\mathbf{q} \notin \mathbb{M}$ - is not an $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$-accommodating subset of $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$ any longer: the worlds in $|\mathbb{M} \cup \{\mathbf{q}\}|$ are too much like $\mathbf{w}'$.

This, however, is not all there is to it. Firstly, it is very probable that exist *several* maximal $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$-accommodating subsets of $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$. The reader may easily verify this by taking $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle} = \{\mathbf{p},\mathbf{q},\mathbf{r}\}$; $\mathbb{P}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle} = \emptyset$; $\mathbf{p} \cap \mathbf{q} \cap \mathbf{r} \neq \emptyset$; $(\mathbb{W} \setminus \mathbf{p}) \cap \mathbf{q} \cap \mathbf{r} = \emptyset$; $(\mathbb{W} \setminus \mathbf{p}) \cap \mathbf{q} \neq \emptyset$; $(\mathbb{W} \setminus \mathbf{p}) \cap \mathbf{r} \neq \emptyset$; $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle} = \mathbb{W} \setminus \mathbf{p}$. Then both $\{\mathbf{q}\}$ and $\{\mathbf{r}\}$ are maximal $[\![A]\!]_{\langle \mathbf{s},\mathbf{w}\rangle}$-accommodating subsets of $\mathbb{O}^{\mathbf{w}'}_{\langle \mathbf{s},\mathbf{w}\rangle}$. But on a more informal level it must be clear, too, that this is very likely to happen. There may be several alternative ways in which to add more assumptions to a non-maximal $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset of $\mathbb{O}^{w'}_{\langle s,w\rangle}$. Thus one may obtain several different maximal $[\![A]\!]_{\langle s,w\rangle}$-accommodating subsets of $\mathbb{O}^{w'}_{\langle s,w\rangle}$. The members of $|\mathbb{M}|$, for each of these maximal $[\![A]\!]_{\langle s,w\rangle}$-accommodating subsets $\mathbb{M}$, all 'in other respects' resemble w' as closely as possible, but different members of different $|\mathbb{M}|$'s resemble w' 'in other respects' as closely as possible 'in different respects'. At this point every attempt to arrive at a comparative *overall* similarity relation in Lewis's sense must be abandoned: we should have to slur over the fact that a possible world w'' may in some respects be *more*, in some other respects *less* similar to a world w' than a world w''' ; they can be alike in that both are p-worlds resembling w' as closely as possible, and differ in that w'' is a q-world much more similar to w' than w''' .

Secondly[19], the possibility is not excluded that there be *no* maximal $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset of $\mathbb{O}^{w'}_{\langle s,w\rangle}$, even if we assume that there is some non-maximal $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{Q}$. It might be possible that one can add assumption after assumption to $\mathbb{Q}$ without ever reaching a limit. Imagine a speaker s without prejudices who explicitly assumes for each natural number n that there exist more than n grains of sand. Let p be the proposition that there are finitely many grains of sand. This proposition can be accommodated within every finite subset of s's assumptions, for there is bound to be a largest number n, such that the assumption that there are more than n grains of sand is an element of $\mathbb{Q}$. On the other hand p cannot be accommodated within an infinite number of s's assumptions; Then which set should be a maximal p-accommodating subset of s's set of opinions?

If the reader wants to check formally that a p-accommodating subset does not always have a maximal extension, then he may choose a frame $\mathbb{F}$ such that $W = \omega+1$; $\mathbb{O}^{w'}_{\langle s,w\rangle} = \{p_i : p_i \in pow(W)$ and $i<\omega$ and $p_i = \{j : i<j<\omega\}\}$; $\mathbb{P}^{w'}_{\langle s,w\rangle} = \emptyset$; $p = \{i : i<\omega\} = \omega$. It is easy to prove that $\mathbb{Q} \subset \mathbb{O}^{w'}_{\langle s,w\rangle}$ accommodates p iff $\mathbb{Q}$ is a finite subset of $\mathbb{O}^{w'}_{\langle s,w\rangle}$.

There is a very natural way of avoiding the difficulties involved in this phenomenon: we can resign ourselves to what we shall call the *Finiteness Assumption*: every speaker in every context holds only a finite number of opinions on each world. It is obvious that this is a sufficient condition in order for every p-accommodating subset of every set of opinions to have a maximal extension.

We are now ready to give truth conditions for the counterfactual operators $\boxed{c}\!\!\rightarrow$ and $\langle c\rangle\!\!\rightarrow$. In def. 6^F we restrict ourselves to the case that we can freely

utilize maximal accommodating subsets, i.e. to frames that satisfy the Finiteness Assumption:

<u>Def. 6^F</u> $w' \in [\![(A \boxed{c}\!\!\rightarrow B)]\!]_{\langle s,w\rangle}$ iff *every* maximal $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{M}$ of $\mathbb{O}^{w'}_{\langle s,w\rangle}$ is such that $([\![A]\!]_{\langle s,w\rangle} \cap |\mathbb{M}|) \subset [\![B]\!]_{\langle s,w\rangle}$

$w' \in [\![(A \langle c\rangle\!\!\rightarrow B)]\!]_{\langle s,w\rangle}$ iff *some* maximal $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{M}$ of $\mathbb{O}^{w'}_{\langle s,w\rangle}$ is such that $([\![A]\!]_{\langle s,w\rangle} \cap |\mathbb{M}|) \cap [\![B]\!]_{\langle s,w\rangle} \neq \emptyset$

Amplification: Substituting 'some' for 'every' in the first line of this definition would allow for the truth of both $\langle (A \boxed{c}\!\!\rightarrow \neg B), \langle s,w\rangle, w'\rangle$ and $\langle (A \boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$, even in case $[\![A]\!]_{\langle s,w\rangle} \cap |\mathbb{P}^{w'}_{\langle s,w\rangle}| \neq \emptyset$. Substituting 'every' for 'some' in the third line would destroy the equivalence of $\langle (A \boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ and $\langle \neg(A \langle c\rangle\!\!\rightarrow \neg B), \langle s,w\rangle, w'\rangle$.

Without the Finiteness Assumption, the formulation of the truth conditions becomes more complicated, but the idea remains the same: $w' \in [\![(A \boxed{c}\!\!\rightarrow B)]\!]_{\langle s,w\rangle}$ if one can add assumptions to every $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{Q}$ of $\mathbb{O}^{w'}_{\langle s,w\rangle}$ up to a point where one gets an $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{N}$ such that $([\![A]\!]_{\langle s,w\rangle} \cap |\mathbb{N}|) \subset [\![B]\!]_{\langle s,w\rangle}$. That one may go on adding more and more assumptions to $\mathbb{N}$, still keeping $[\![A]\!]_{\langle s,w\rangle}$-accommodating subsets $\mathbb{N}'$, is of minor importance since every such $|\mathbb{N}'|$ is a subset of $|\mathbb{N}|$.

If, on the other hand, there exists some $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{Q}$ of $\mathbb{O}^{w'}_{\langle s,w\rangle}$ such that you never reach that point, whatever assumptions you add to $\mathbb{Q}$, and notwithstanding the fact that you can go on adding more and more assumptions and still keep $[\![A]\!]_{\langle s,w\rangle}$-accommodating subsets of $\mathbb{O}^{w'}_{\langle s,w\rangle}$, then $w' \notin [\![(A \boxed{c}\!\!\rightarrow B)]\!]_{\langle s,w\rangle}$. Again it is of minor importance that with the Finiteness Assumption you need only concern yourself with the maximal $[\![A]\!]_{\langle s,w\rangle}$-accommodating subsets of $\mathbb{O}^{w'}_{\langle s,w\rangle}$ which you must necessarily get in this way.

We define with respect to the general case:

<u>Def. 6^G</u> $w' \in [\![(A \boxed{c}\!\!\rightarrow B)]\!]_{\langle s,w\rangle}$ iff every $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{Q}$ of $\mathbb{O}^{w'}_{\langle s,w\rangle}$ is a subset of some $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{N}$ of $\mathbb{O}^{w'}_{\langle s,w\rangle}$ such that $([\![A]\!]_{\langle s,w\rangle} \cap |\mathbb{N}|) \subset [\![B]\!]_{\langle s,w\rangle}$

$w' \in [\![(A \langle c\rangle\!\!\rightarrow B)]\!]_{\langle s,w\rangle}$ iff some $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{Q}$ of $\mathbb{O}^{w'}_{\langle s,w\rangle}$ is such that $([\![A]\!]_{\langle s,w\rangle} \cap |\mathbb{N}|) \cap [\![B]\!]_{\langle s,w\rangle} \neq \emptyset$ for *every* $[\![A]\!]_{\langle s,w\rangle}$-accommodating subset $\mathbb{N}$ of $\mathbb{O}^{w'}_{\langle s,w\rangle}$ of which $\mathbb{Q}$ is a subset.

A final explanatory remark: Compare the wording of the truth condition for judgments $\langle (A \langle c\rangle\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ with the phrase '$w' \in [\![(A \langle c\rangle\!\!\rightarrow B)]\!]_{\langle s,w\rangle}$ iff some

$[\![A]\!]_{<s,w>}$-accommodating subset $\mathbb{Q}$ of $\mathbb{O}^{w'}_{<s,w>}$ is such that $([\![A]\!]_{<s,w>} \cap |\mathbb{Q}|) \cap [\![B]\!]_{<s,w>} \neq \emptyset$. This would be to weak: adding more assumptions to such a $\mathbb{Q}$ might give an $[\![A]\!]_{<s,w>}$-accommodating subset $\mathbb{N}$ such that $([\![A]\!]_{<s,w>} \cap |\mathbb{N}|) \subset (\mathbb{W} \sim [\![B]\!]_{<s,w>})$. This would allow for the truth of both $<(A \diamond\!\!\rightarrow B),<s,w>,w'>$ and $<(A \boxed{c}\!\!\rightarrow \neg B),<s,w>,w'>$.

I cannot hope that my final explanatory remark will have been a final one in defence of these truth conditions as well. Indeed, I may expect the same kind of severe criticism I met in § 2. Once again this opponent is proposing to replace, throughout def. 4 - def. 6^G, the occurrences of '$\mathbb{O}^{w'}_{<s,w>}$' by occurrences of '${}^*\mathbb{O}^{w'}_{<s,w>}$'; the occurrences of '$\mathbb{P}^{w'}_{<s,w>}$' by occurrences of '${}^*\mathbb{P}^{w'}_{<s,w>}$'; and the occurrences of '$[\![\;]\!]$' by occurrences of '$[\![\;]\!]^*$'. He is still worried about the fact that prejudices and assumptions which in fact do not hold play a role in our truth conditions. And what a role! Any phantasm, utterly unfounded but nevertheless figuring in a speaker's stock of opinions, may make a difference to the truth value of his counterfactual judgments. To give a simple example: a statement $<(A \boxed{c}\!\!\rightarrow B),<s,w>,w>$ may owe its truth simply and solely to the prejudice $[\![(A \supset B)]\!]_{<s,w>}$, which in fact does not hold for some $w' \in \mathbb{H}$. Once more my opponent argues that even the speaker s of this example would himself conclude, after a falsification of this theory $[\![(A \supset B)]\!]_{<s,w>}$ of his, that the statement at issue was false. And again the reader finds me replying that the only conclusion to draw for s, after a falsification, is that he must drop this theory $[\![(A \supset B)]\!]_{<s,w>}$. Then a judgment $<(A \boxed{c}\!\!\rightarrow B),<s,w'>,w>$, made next moment w', will probably become false - only probably: s may in the meantime have gotten, rightly or wrongly some further prejudices and assumptions about w -, but there is nothing paradoxical in still reckoning $<(A \boxed{c}\!\!\rightarrow B),<s,w>,w>$ among the true statements that have been made in that particular context $<s,w>$.

Both this opponent, to whom I shall henceforth refer as my *opponent with his *definitions, and myself may expect to encounter a second opponent - we have already met him at the end of § 2 in the person of the reader - who will accuse both of us of advocating a kind of coherence theory of truth for counterfactual conditionals; and that, according to him, is no theory of truth at all. In any adequate (correspondence) theory of truth a true counterfactual statement would owe its truth to the facts, and not to the accidental knowledge - or even worse, to the accidental beliefs - that the utterer of that statement happens to have about these facts.

(It is worth noting how seriously the three of us differ in opinion, not only in theory, but also in practice. The following puzzle may serve as an illustration:

Three persons are involved: Mr. A, Mr. B and a conjurer. Not such a very good conjurer actually, for he accidently murdered his wife on the stage, sawing her in two. Mr. A has just heard this on the radio.

Mr. B attended the show, but he seems not to have noticed what happened. 'This final trick, marvellous!' he says to Mr. A and (1) *'If you had seen it, you would not have believed your eyes'*.

Mr. A is puzzled by this statement; he has been in the juggling business himself, you know, and he knows all about the sawing trick, so that he *agrees* with Mr. B. On the other hand, if he tells Mr. B about this juggling career of his, then Mr. B will probably change his mind and say: *'O, well, then you might have believed your eyes of course'*. And then they would *disagree*, until the moment when he tells Mr. B what he has heard on the radio. Then Mr. B must change his mind again, and say: (3) *'If you had seen it, you would not have believed your eyes'*.

Question: Assume that both my *opponent and myself, as well as our common adversary, agree that statement (3), if stated in the circumstances as described, would be true. What, then, would be the truth value of the statements (1) and (2) according to each of us?)

How do we answer this second opponent? How do we convince him that this wish for a correspondence theory of truth for counterfactuals cannot be realized? Neither I nor my *opponent (if I may speak on his behalf) know of any decisive argument in our favour. Nevertheless, our adversary's position is all but enviable. For what we can point out is that the correspondence theory which he is looking for cannot in any case be based on a comparative similarity relation of worlds. However, his only alternative seems to be to explain counterfactual truth in terms of such a relation, and to deny that this relation is linked up with either the knowledge or the beliefs of any speaker whatsoever. In other words, he seems to be obliged to revert to a pre-established comparative similarity-relation fixed by factual similarities and dissimilarities.

Any attempt to develop a correspondence theory along these lines is doomed to failure. This can be shown by my *opponent. After all, within his theory counterfactual truth, and with it comparative similarity, are being judged in the light of facts, though it be only in the light of the facts known to a speaker. As a consequence our common adversary can only reproach him for not taking unknown facts into account. Another consequence is that this adversary must be able to indicate which *additional* facts should be reckoned with. Are all additional facts relevant?

Let me illustrate this question by means of an example. Suppose a particular judgment $\langle (A \boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ is false according to my *opponent, but true by the standards of our adversary (compare statement (1) of the above puzzle). Here my *opponent will be able to select an $[\![A]\!]^*_{\langle s,w\rangle}$-accommodating subset $\mathbb{Q}$ of $^*\mathbb{O}^{w'}_{\langle s,w\rangle}$ which cannot be extended (by facts known to s) to an $[\![A]\!]^*$-accommodating subset $\mathbb{N}$ of $^*\mathbb{O}^{w'}_{\langle s,w\rangle}$ such that $([\![A]\!]^*_{\langle s,w\rangle} \cap |\mathbb{N}|) \subset [\![B]\!]^*_{\langle s,w\rangle}$. That is, there are, for all s knows about w', no $[\![B]\!]^*_{\langle s,w\rangle}$-worlds in $([\![A]\!]^*_{\langle s,w\rangle} \cap |\mathbb{Q}|)$ which resemble w' more than the $[\![\neg B]\!]^*_{\langle s,w\rangle}$-worlds in it do.

Our common adversary, for his part, cannot deny that the worlds of $([\![A]\!]^*_{\langle s,w\rangle} \cap |\mathbb{Q}|)$ all are worlds which in fact resemble w' in some respects: the propositions $p \in \mathbb{Q}$, which hold for w', hold also for these worlds. Furthermore, he must admit that all the $[\![A]\!]^*_{\langle s,w\rangle}$-worlds which in fact resemble w' in more respects than these $\mathbb{Q}$-respects are elements of $([\![A]\!]^*_{\langle s,w\rangle} \cap |\mathbb{Q}|)$ as well. So we must hope with him that there are at least some $[\![B]\!]^*_{\langle s,w\rangle}$-worlds among the elements of $([\![A]\!]_{\langle s,w\rangle} \cap |\mathbb{Q}|)$; otherwise he cannot defend the truth of $\langle (A \boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ at all. And if there are $[\![B]\!]^*_{\langle s,w\rangle}$-worlds in $([\![A]\!]^*_{\langle s,w\rangle} \cap |\mathbb{Q}|)$, then he must be able to defend the proposition that some of these $[\![B]\!]^*_{\langle s,w\rangle}$-worlds do in fact resemble w' more than the $[\![\neg B]\!]^*_{\langle s,w\rangle}$-worlds in $([\![A]\!]^*_{\langle s,w\rangle} \cap |\mathbb{Q}|)$. In this case we may expect him to tell us that we should see past the end of s's nose, and he, in turn, may expect to be asked *how far* we should see past s's nose. Has he himself arrived at his conclusion by taking *all* facts in w' into account and by *exhaustively* comparing the $[\![B]\!]^*_{\langle s,w\rangle}$-worlds in $([\![A]\!]^*_{\langle s,w\rangle} \cap |\mathbb{Q}|)$ with the $[\![\neg B]\!]^*_{\langle s,w\rangle}$-worlds in it? To put it differently: if our adversary wants his fixed comparative similarity relation to be more than a hollow phrase, then he must be able to point out, in advance, the respects in which the possible worlds are to be compared with each other, and give some reason why the respects which he neglects are to be neglected.

Let us look at the consequences of the most natural answer to this question: compare the worlds in all respects; all facts are equally relevant. The point is that my *opponent can easily simulate a situation in which all facts are taken into account. We may consider an omniscient speaker s who at a certain moment w happens to know everything there is to know about all worlds $w' \in \mathbb{W}$:

Let $\mathbb{F}$ be such that $\mathbb{O}^{w'}_{\langle s,w\rangle} = \{p : p \in pow(\mathbb{W}) \text{ and } w' \in p\}$ for all $w' \in \mathbb{W}$; Take $\mathbb{P}^{w'}_{\langle s,w\rangle} = \emptyset$ iff $w' \notin \mathbb{H}$, and $\mathbb{P}^{w'}_{\langle s,w\rangle} = \{p : p \in pow(\mathbb{W}) \text{ and } \mathbb{H} \subset p\}$ iff $w' \in \mathbb{H}$.[20] It is easy to prove that on this condition:

$$w' \in [\![(A \boxed{c}\!\!\rightarrow B)]\!]^*_{\langle s,w\rangle} \text{ iff } w' \in [\![((A \supset B) \;\&\; (\neg A \supset \boxtimes(A \supset B)))]\!]^*_{\langle s,w\rangle}.$$

And it is easy to see why this result entails that the most natural answer

to our question is wrong. If it were right, then the logic of counterfactual statements $\langle (A \boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w\rangle$ would boil down to the logic of statements $\langle ((A \supset B) \;\&\; (\neg A \supset \boxtimes(A \supset B))), \langle s,w\rangle, w\rangle$. It would be valid, for example, to conclude that a statement $\langle ((A \;\&\; C) \boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w\rangle$ is true, given the truth of $\langle \neg A, \langle s,w\rangle, w\rangle$ and $\langle (A \boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w\rangle$. This, however, cannot be our adversary's intention.[21]

It should be clear now that our adversary finds himself in a rather precarious situation. On the one hand he reproaches my *opponent for only reckoning with the facts known to a speaker. On the other hand he must admit that taking all facts into account would cause our cartoon to give a wrong picture of the logic of counterfactual conditionals. Hence he must neglect some facts himself, and draw a line somewhere in between. But whether he draws this line vaguely or precisely, he cannot exclude the possibility of a speaker's wanting, on the grounds of his knowledge, to take more facts into account than he is actually supposed to do.

It is time to assess the pros and cons. I hope that the above argument has convinced the reader that any correspondence theory of truth for counterfactuals based on a pre-established comparative similarity relation of worlds is an absurdity, however vaguely one may define that relation. I doubt whether I have succeeded in persuading him to move directly into accepting the point of view recorded in the definitions 4-6G. He may prefer the less risky step to the standpoint of my *opponent and refuse to make any further concessions until some convincing argument forces him to do so. I must admit that such an argument has not been given. What makes me incline to my "own" point of view is the belief that counterfactuals fall into the category of *may-*, *might-*, *might have been-* and *must-*, *should-*, *should have been*-sentences. And I still flatter myself with the hope that I have refuted my *opponent at least when we discussed the semantic properties of these sentences.

§ 4 'Counterfactual' Presuppositions

At the outset of the preceding section I called upon the reader not to raise questions concerning presuppositional matters until further notice. I doubt whether he has managed to comply with that request, since both my *opponent and myself may certainly have provoked such questions through having intimated how they are *not* to be answered. Our truth conditions leave no room for a counterfactual judgment to be neither true nor false. Hence we cannot trace back the oddity of, for example, the sentence *'It would not have been for the common good, if I had resigned'*, stated by Nixon the day *after* his resignation, to something like a truth value gap. And we have no intention of doing so: questions of presupposition are not to be answered within semantics, but within pragmatics; their solution does not lie in a theory of truth, but in a theory of conversation. The oddity of the above statement must be attributed to the fact that Nixon, in stating this sentence, would have broken a general rule of conversation.

One cannot utter any sentence at any time about any situation; judgments can be *conversationally out of place*, whether they are true or not. Having made this trifling observation - after all 7+5=12 - we might ask for the criteria by which it can be determined whether or not a judgment is conversationally out of place. We might try to find these criteria in some *general rules of conversation* - better: some principles of rational co-operative behaviour - which the participants in a conversation should observe in order that their conversation be for all of them as fruitful and to the point as possible. Then, having found these criteria, we might carry on and try to stake out for every kind of sentence the circumstances in which it can be used correctly.

Recently both linguists and logicians have begun to fill in the details of this program, and in particular Kempson[22] and Stalnaker[23], both inspired by Grice's 'Logic and Conversation'[24], have broken a lance for the idea that "*the phenomena often incorporated into semantics under an umbrella label of presupposition can naturally be explained within* (such F.V.) *a framework which depends on a logically prior system of linguistic conventions*"[25].

I hope to illustrate the fertility of this idea below. More specifically: I shall curtail the speakers figuring in our formal cartoon in their liberty of using $(A \,\boxed{c}\!\!\rightarrow B)$- and $(A \diamond\!\!\!\rightarrow B)$-judgments. I shall supply some evidence in favour of the claim that the restrictions imposed on these speakers apply to ourselves as well. And I shall show - or at least try to - how these restrictions can be justified in view of some general rules of conversation.

Remember the definition of an *implicit opinion*: a proposition $\mathbf{p}$ is an implicit opinion of s at w on w' if and only if $|\mathbf{O}^{w'}_{\langle s,w\rangle}| \subseteq \mathbf{p}$. Using this concept we can distinguish nine kinds of circumstances - they are mutually exclusive

iff $|\mathbf{O}^{w'}_{\langle s,w\rangle}| \neq \emptyset$ - in which a judgment $\langle (A \,\boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ or $\langle (A \diamond\!\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ can be made (It is optional to read '$[\![\;]\!]^{(*)}$' as either $[\![\;]\!]$ or $[\![\;]\!]^{*}$):

C1. $[\![A]\!]^{(*)}_{\langle s,w\rangle}$ and $[\![B]\!]^{(*)}_{\langle s,w\rangle}$ are implicit opinions of s at w on w'.

C2. $[\![A]\!]^{(*)}_{\langle s,w\rangle}$ and $[\![\neg B]\!]^{(*)}_{\langle s,w\rangle}$ are implicit opinions of s at w on w'.

C3. $[\![A]\!]^{(*)}_{\langle s,w\rangle}$ is an implicit opinion of s at w on w', and neither $[\![B]\!]^{(*)}_{\langle s,w\rangle}$ nor $[\![\neg B]\!]^{(*)}_{\langle s,w\rangle}$ are implicit opinions of s at w on w'.

C4. $[\![\neg A]\!]^{(*)}_{\langle s,w\rangle}$ and $[\![B]\!]^{(*)}_{\langle s,w\rangle}$ are implicit opinions of s at w on w'.

C5. $[\![\neg A]\!]^{(*)}_{\langle s,w\rangle}$ and $[\![\neg B]\!]^{(*)}_{\langle s,w\rangle}$ are implicit opinions of s at w on w'.

C6. $[\![\neg A]\!]^{(*)}_{\langle s,w\rangle}$ is an implicit opinion of s at w on w', and neither $[\![B]\!]^{(*)}_{\langle s,w\rangle}$ nor $[\![\neg B]\!]^{(*)}_{\langle s,w\rangle}$ are implicit opinions of s at w on w'.

C7. Neither $[\![A]\!]^{(*)}_{\langle s,w\rangle}$ nor $[\![\neg A]\!]^{(*)}_{\langle s,w\rangle}$ are implicit opinions of s at w on w', and $[\![B]\!]^{(*)}_{\langle s,w\rangle}$ is an implicit opinion of s at w on w'.

C8. Neither $[\![A]\!]^{(*)}_{\langle s,w\rangle}$ nor $[\![\neg A]\!]^{(*)}_{\langle s,w\rangle}$ are implicit opinions of s at w on w', and $[\![B]\!]^{(*)}_{\langle s,w\rangle}$ is an implicit opinion of s at w on w'.

C9. Neither $[\![A]\!]^{(*)}_{\langle s,w\rangle}$ nor $[\![\neg A]\!]^{(*)}_{\langle s,w\rangle}$, and neither $[\![B]\!]^{(*)}_{\langle s,w\rangle}$ nor $[\![\neg B]\!]^{(*)}_{\langle s,w\rangle}$ are implicit opinions of s at w on w'.

The remainder of this section, then, will be devoted to a discussion of the following stipulation:

- A judgment $\langle (A \,\boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ or $\langle (A \diamond\!\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ is conversationally correct only if it is made in the circumstances described in C4, C5 and C6.

Let it be understood that the above stipulation is meant to frame only a *necessary* condition for the correctness of the conditional judgments at issue. We shall call every necessary condition for the correctness of a judgment a *presupposition* of that judgment. Thus, a counterfactual judgment $\langle (A \,\boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ carries *at least* the presupposition that $[\![\neg A]\!]^{(*)}_{\langle s,w\rangle}$ is an implicit opinion of s at w on w'. But this is not the only condition which must be satisfied: the judgment $\langle (A \,\boxed{c}\!\!\rightarrow B), \langle s,w\rangle, w'\rangle$ must also be regarded as incorrect if, for instance, even the speaker himself does not believe that he is speaking the truth.[26]

Is there any evidence in favour of the claim that a judgment '*if it were* *the case that A, then it would be the case that B'* is conversationally out of place if made in the circumstances C1, C2, C3, C7, C8 and C9? I hope the reader will answer this question in the affirmative after he has allowed himself to take part in the following experiment. First I should like him to examine what his reactions would be to each of the following sequences of statements

(C1) *I am in love with you, darling, and I want to marry you. And if I were in love with you, then I would like to marry you.*

(C2) *I visited Paris last week, but I did not see the Eiffel Tower. And I would have seen the Eiffel Tower, if I had visited Paris last week.*

(C3) *Maybe I passed; maybe not. I haven't got the result yet. Anyway, I have done my utmost. And if I had done my utmost, then I would have passed.*

(C7) *Maybe we have met before; maybe not. But I know your husband quite well. And if we had met before, then I would know your husband quite well.*

(C8) *We may have met before; I don't remember. And if we had met before, I would remember.*

(C9) *I may have borrowed your car; I don't remember. And I may have crashed it too. I really don't know. But if I had not borrowed it then I would not have crashed it.*

The reader, playing the part of the addressee, may have listened very willingly to the first part of each of these sequences (trusting that the utterer is telling at least something that he believes himself). But then, I hope, the subsequent counterfactual must have confused him and made him doubtful of either his own eyes or the sanity of the utterer. If I am right here, then the first part of our experiment has been successful.

The utterers of the above statements have been so kind as to indicate the relevant circumstances previous to uttering the relevant - irrelevant actually - counterfactual. In such cases the alleged incorrectness is quite obvious. However, I must show, too, that the counterfactual statements in question would have been incorrect if the speakers had simply concealed these circumstances, or if the relevant beliefs had not been explicit ones. To settle this, I must ask the reader to fill the part of the speaker: just believe what is stated in the first part of each of these sequences and try to utter, then, the subsequent counterfactual; or utter the counterfactual in question and "realize", then,

what you are actually believing.

These examples in favour of our claim that a counterfactual judgment is incorrect if it is made in circumstances C1, C2, C3, C7, C8, C9 - I assume that the second part of our experiment has been successful too - can of course at most challenge the reader to search for counter-examples. I have some arguments in store, however, which are bound to thwart his attempt to find these. But first I shall show that a counterfactual judgment can be correct if it is made in the circumstances C4, C5 and C6.

It has been argued, traditionally, that counterfactuals presuppose the falsehood of both the antecedent and the consequent. The following examples serve the purpose of showing that this view is based on a confusion of the notions of truth and falsity with the notions of correctness and incorrectness.

(C5) *If Santa Claus did not exist, we would get no Christmas presents.*

To my mind, this sentence can be perfectly correct, if at least it is uttered by a child which happens to believe that Santa Claus exists. Things are different, however, if its father utters the same sentence. Then the resulting statement is conversationally incorrect indeed. The difference is that the child in question is honest, whereas its father is only pretending that he believes in the existence of Santa Claus.

Of course, there still is *something* wrong with the statement of this child, since in uttering this statement correctly it is giving evidence of holding an erroneous belief. This, however, is quite another thing. You can only conclude that this child actually holds this belief if you *first* take for granted that it is not trying to cheat you.

Supporters of the traditional view do not draw the above distinctions. They accuse both this mistaken child and its dishonest father of the same offense. Consequently, they cannot accuse anyone dissimulating an erroneous belief of any offense whatsoever.[27]

That a counterfactual judgment does not presuppose the falsity of its consequent - and not even that its consequent is believed to be false - may appear from the following example:

(C6) *Do you love me? You would love me, wouldn't you, if I only were a bit like Elvis Presly. But do you love me, anyhow?*

I think there is nothing wrong with this sequence of sentences if indeed the utterer is not merely asking for the sake of asking.

The case C4 requires some special attention.

(C4) *John did not attend the show; nor did Mary or Peter. John would have gone if Mary had gone. But if Mary had taken Peter along, then he would not have gone.*

The third statement in the above sequence is a counterfactual judgment made in circumstances C4 - if indeed the utterer believes what he has stated in the first sentence of this sequence - and there is no fault to be found with its correctness. Hence a counterfactual judgment made in the circumstances C4 can be correct. The reader will have noticed, however, that our example is rather complicated. Indeed, we get into trouble if we try a simpler one. Consider:

Neither John nor Mary attended the show. And if Mary had gone to it, then John would not have gone to it.

To my mind, the second statement in this sequence is incorrect. It is not so much a counterfactual conditional as a so-called semifactual conditional that is due here:

Neither John nor Mary attended the show. And if Mary had gone to it, then John would still not have gone to it.

A semifactual conditional can be obtained from a counterfactual by inserting 'still' in the consequent and (sometimes) 'even' in the antecedent. Semifactuals are incorrect in the circumstances C1, C2, C3, C7, C8, C9, just as counterfactuals are. It should be noted that these semifactual conditionals can be used appropriately also in other circumstances than C4:

Neither John nor Mary nor Peter attended the show. John would have gone to it if Mary had gone. And he would still have gone to it, even if Mary had taken Peter along.

I consider semifactuals to form a special class of counterfactuals: in matters of truth they behave like counterfactuals, and in the first instance they behave like counterfactuals in matters of presupposition too. That a more subtle look will allow us to articulate some differences is beyond doubt.

Evidence in favour of our claim that a might-counterfactual is always conversationally out of place in the circumstances C1, C2, C3, C7, C8, C9, and sometimes conversationally correct in the circumstances C4, C5 and C6 can be furnished in the same way. We leave it to the reader to adapt the examples given for (would-)counterfactuals.

We now pass on to the last subject of this section: An attempt to show that the incorrectness of judgments $\langle (A \boxed{c}\!\!\rightarrow B), \langle \mathfrak{s},\mathfrak{w}\rangle, \mathfrak{w}'\rangle$ and $\langle (A \diamond\!\!\!c\!\!\rightarrow B), \langle \mathfrak{s},\mathfrak{w}\rangle, \mathfrak{w}'\rangle$ made in the circumstances C1, C2, C3, C7, C8 and C9 can be explained in view of some maxims of conversation (To prevent disappointment, I must ask the reader to read the foregoing sentence once again, emphasizing the word 'attempt').

Consider a judgment $\langle (A \boxed{c}\!\!\rightarrow B), \langle \mathfrak{s},\mathfrak{w}\rangle, \mathfrak{w}'\rangle$ or $\langle A \diamond\!\!\!c\!\!\rightarrow B, \langle \mathfrak{s},\mathfrak{w}\rangle, \mathfrak{w}'\rangle$ made in circumstances such that $[\![A]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle} \cap |\mathbb{O}^{\mathfrak{w}'}_{\langle \mathfrak{s},\mathfrak{w}\rangle}| \neq \emptyset$ - this applies to all the circumstanc C1, C2, C3, C7, C8 and C9 (if $|\mathbb{O}^{\mathfrak{w}'}_{\langle \mathfrak{s},\mathfrak{w}\rangle}| \neq \emptyset$[28]). Notice that under this condition $\mathbb{O}^{\mathfrak{w}'}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$ - or ${}^{*}\mathbb{O}^{\mathfrak{w}'}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$ if you prefer the *definitions - is an extension of every $[\![A]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$-accommodating subset $\mathcal{Q}$ of ${}^{(*)}\mathbb{O}^{\mathfrak{w}'}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$. Therefore $\mathfrak{w}' \in [\![(A \boxed{c}\!\!\rightarrow B)]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$ iff $([\![A]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle} \cap |{}^{(*)}\mathbb{O}^{\mathfrak{w}'}_{\langle \mathfrak{s},\mathfrak{w}\rangle}|) \subset [\![B]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$, and $\mathfrak{w}' \in [\![(A \diamond\!\!\!c\!\!\rightarrow B)]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$ iff $([\![A]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle} \cap |{}^{(*)}\mathbb{O}^{\mathfrak{w}'}_{\langle \mathfrak{s},\mathfrak{w}\rangle}|) \cap [\![B]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle} \neq \emptyset$. This means that $\mathfrak{w}' \in [\![(A \boxed{c}\!\!\rightarrow B)]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$ iff $\mathfrak{w}' \in [\![\mathit{must}(A \supset B)]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$ and $\mathfrak{w}' \in [\![(A \diamond\!\!\!c\!\!\rightarrow B)]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$ iff $\mathfrak{w}' \in [\![\mathit{may}(A \;\&\; B)]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$.

In other words, if a *counterfactual* conditional '*If it were the case that A, then it would be the case that B*' is stated in circumstances in which its antecedent is not believed to be false, then it is true iff the corresponding *indicative* conditional '*If it is the case that A, then it must be the case that B*' is true. Likewise: A subjunctive conditional '*If it were the case that A, then it might be the case that B*' is true in these circumstances iff the corresponding *indicative* statement '*maybe it is the case that A and B*' is true. Hence, a speaker who has uttered such a subjunctive sentence in such circumstances, *might* just as well have uttered the corresponding indicative sentence. And he *should* have uttered this corresponding indicative sentence, if indeed it was not his aim to mislead his discussion-partner.[29]

There is no need for a speaker $\mathfrak{s}$ to make a judgment $\langle (A \boxed{c}\!\!\rightarrow B), \langle \mathfrak{s},\mathfrak{w}\rangle, \mathfrak{w}'\rangle$ if $[\![A]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle} \cap |\mathbb{O}^{\mathfrak{w}'}_{\langle \mathfrak{s},\mathfrak{w}\rangle}| \neq \emptyset$. There is no need for him to use the subjunctive mood if the proposition expressed in the antecedent of this counterfactual is compatible with his opinions on $\mathfrak{w}'$. For in this case he does not have to give up any of his opinions on $\mathfrak{w}'$ in order to enable himself to imagine that $[\![A]\!]^{(*)}_{\langle \mathfrak{s},\mathfrak{w}\rangle}$ holds for $\mathfrak{w}'$. The addressee, however, hearing the phrase 'If it were the case that ....', will assume that the speaker does have to give up some of his opinions. For if he hears the subjunctive mood instead of the indicative one, then he will think that the speaker in order to make his point needs to use that mood. Therefore, it is misleading to use the subjunctive mood when the indicative one would serve just as well.

I think it is a rule for rational co-operative behaviour that we should only make use of a certain provision in our language if its application is indeed really necessary for our purpose. That is why I think that a counterfactual judgment is only correct if it is made in circumstances in which the antecedent is believed to be false. Granted, my formulation of this rule leaves much to be desired where clarity is concerned. Nevertheless, we would do better to search for a more precise formulation than to deny that we should take heed of it.

One additional observation pertaining to subjunctive conditionals should be made. It concerns the question whether there exists a second kind of subjunctive conditionals which behaves differently in matters of truth from the counterfactuals we have studied. It has been argued several times that this question must be answered in the affirmative. Lewis, for instance, entitled his book 'Counterfactuals' rather than 'Subjunctive Conditionals' for the following reason:

'The title 'Subjunctive Conditionals' would not have delineated my subject properly ..... There are subjunctive conditionals pertaining to the future like '*If our ground troops entered Laos next year, there would be trouble*' that appear to have the truth condition of indicative conditionals rather than of the counterfactual conditionals I shall be considering.'[30]

Two remarks are in order: Firstly, we agree with Lewis that a subjunctive conditional may appear to have the truth condition of an indicative conditional. We have seen above that a conversationally incorrect counterfactual judgment has this property.

Secondly, we would disagree with Lewis, however, if he were to claim that such a subjunctive conditional might be conversationally correct. We would argue that even a subjunctive conditional pertaining to the future carries the presupposition that its antecedent is believed to be false. Compare:

- *Our ground troops may enter Laos next year; if they do, there will be trouble.*
- *Our ground troops will not enter Laos next year; if they did, there would be trouble.*
- *Our ground troops may enter Laos next year; if they did, there would be trouble.*

Once again I must ask the reader to fill the part of the speaker: believe what is stated in the first sentence of each of the above sequences and utter the second one. If in doing so you meet some difficulties as to the third sequence, then you may endorse our conclusion that those 'other' subjunctive conditional judgments which appear to have the truth condition of indicative conditionals are

just counterfactual conditional judgments, made in circumstances appropriate to the use of only indicative ones.

§ 5 Appendix

Let $\mathcal{L}$ be a language answering to def. 2. If matters of logic are of interest to us here, then the first thing for us to do is to define what it means for a sentence C of $\mathcal{L}$ to be valid. We are faced with several alternatives. In the first place we must distinguish between a notion of validity*, based on the *definitions of my *opponent, and a notion of validity, based on my 'own' definition.[31] Secondly, we must realize that there is a distinction to be made between the logic of judgments and the logic of statements, and between the logic of assertions about historical worlds and the logic of assertions about 'just' imaginary worlds.

Def.7 Let C be a sentence of .

C is valid_1 iff $w' \in [\![C]\!]_{<s,w>}$ for every $\mathbb{F}$, $[\![\ ]\!]$, $<s,w> \in \mathbb{C}$, and $w' \in \mathbb{W}$

C is valid_2 iff $w' \in [\![C]\!]_{<s,w>}$ for every $\mathbb{F}$, $[\![\ ]\!]$, $<s,w> \in \mathbb{C}$, and $w' \in \mathbb{H}$

C is valid_3 iff $w' \in [\![C]\!]_{<s,w>}$ for every $\mathbb{F}$, $[\![\ ]\!]$, $<s,w> \in \mathbb{C}$, and $w' \in \mathbb{W} \sim \mathbb{H}$

C is valid_4 iff $w \in [\![C]\!]_{<s,w>}$ for every $\mathbb{F}$, $[\![\ ]\!]$, and $<s,w> \in \mathbb{C}$

C is valid_1^* iff $w' \in [\![C]\!]^*_{<s,w>}$ for every $\mathbb{F}$, $[\![\ ]\!]^*$, $<s,w> \in \mathbb{C}$, and $w' \in \mathbb{W}$

C is valid_2^* iff $w' \in [\![C]\!]^*_{<s,w>}$ for every $\mathbb{F}$, $[\![\ ]\!]^*$, $<s,w> \in \mathbb{C}$, and $w' \in \mathbb{H}$

C is valid_3^* iff $w' \in [\![C]\!]^*_{<s,w>}$ for every $\mathbb{F}$, $[\![\ ]\!]^*$, $<s,w> \in \mathbb{C}$, and $w' \in \mathbb{W} \sim \mathbb{H}$

C is valid_4^* iff $w \in [\![C]\!]_{<s,w>}$ for every $\mathbb{F}$, $[\![\ ]\!]^*$, and $<s,w> \in \mathbb{C}$

The following observations should be made:

- If a sentence is valid_i then it is valid_i^* $(i \in \{1,2,3,4\})$
- There are sentences which are valid_i^* , but not valid_i $(i \in \{1,2,3,4\})$
  Example: Take $A \in AT$, and consider the sentence $(\boxtimes A \supset A)$.
- A sentence is valid_2 iff valid_4, and valid_2^* iff valid_4^*
- If a sentence is valid_1, then it is valid_2 and valid_3
  If a sentence is valid_1^*, then it is valid_2^* and valid_3^*
- There are sentences which are valid_2 and valid_2^*, but neither valid_3 nor valid_3^*
  Example: Take $A \in AT$, and consider $(\oplus A \supset \Box \oplus A)$.
- There are sentences which are valid_3 and valid_3^*, but neither valid_2 nor valid_2^*
  Example: Take $A \in AT$ and consider $(\boxtimes A \supset \Box A)$.

It may have occurred to the reader that the notions of validity_1, validity_2

and validity$_3$ - and the notions of validity$^*_1$, validity$^*_2$ and validity$^*_3$ - yield different logics simply as a result of the requirement, recorded in definition 1, that within every frame $\mathbb{F}$, $\mathbb{P}^{w'}_{<s,w>} = \emptyset$ if $w' \notin \mathbb{H}$, and $\mathbb{P}^{w'}_{<s,w>} = \mathbb{P}^{w'}_{<s,w>}$ if $w' \in \mathbb{H}$. Let us for a moment drop this requirement and call every sextuple answering to this new definition 1 an *imperfect frame*. Let us define eight new notions of validity - ^{0}validity$_1$, ^{0}validity$^*_1$, ^{0}validity$_2$, etc. - just as we did the old notions, but then relative to the class of imperfect frames. Indeed, it appears that a sentence is ^{0}valid$_i$ iff ^{0}valid$_j$, and ^{0}valid$^*_i$ iff ^{0}valid$^*_j$ $(i,j \in \{1,2,3,4\})$. Notice furthermore that every ^{0}valid$_i$ sentence is valid$_i$ and that every ^{0}valid$^*_i$ sentence is valid$^*_i$ $(i \in \{1,2,3,4\})$.

We shall now specify a logic L_o for $\mathcal{L}$ which is correct and complete with respect to the notions of ^{0}validity$_i$ $(i \in \{1,2,3,4\})$.
L_o is the smallest subset of SEN which meets the following conditions:

i) L_o contains the following *axioms*:
- All truth functional tautologies.
- All sentences $(\Diamond A \equiv \neg\Box\neg A)$; $(\Box(A \supset B) \supset (\Box A \supset \Box B))$; $(\Box A \supset A)$; $(\Diamond A \supset \Box\Diamond A)$.
- All sentences $(\Box A \supset \boxtimes A)$; $(\oplus A \equiv \neg\boxtimes\neg A)$; $(\boxtimes(A \supset B) \supset (\boxtimes A \supset \boxtimes B))$.
- All sentences $(\boxtimes A \supset must A)$; $(may A \equiv \neg must \neg A)$; $(must(A \supset B) \supset (must A \supset must B))$.
- All sentences $((A \Diamond\!\!\rightarrow B) \equiv \neg(A \boxed{c}\!\!\rightarrow \neg B))$;
$((A \boxed{c}\!\!\rightarrow (B \supset C)) \supset ((A \boxed{c}\!\!\rightarrow B) \supset (A \boxed{c}\!\!\rightarrow C)))$;
$(((A \boxed{c}\!\!\rightarrow B) \ \& \ (B \boxed{c}\!\!\rightarrow A)) \supset ((A \boxed{c}\!\!\rightarrow C) \supset (B \boxed{c}\!\!\rightarrow C)))$;
$(\oplus A \supset ((A \boxed{c}\!\!\rightarrow B) \supset (A \Diamond\!\!\rightarrow B)))$;
$(\boxtimes(A \supset B) \supset (A \boxed{c}\!\!\rightarrow B))$;
$((A \boxed{c}\!\!\rightarrow B) \supset must(A \supset B))$;
$(may A \supset (must(A \supset B) \supset (A \boxed{c}\!\!\rightarrow B)))$.

ii) if $A \in L_o$ and $(A \supset B) \in L_o$, then $B \in L_o$. (I.o.w., L_o is closed under *Modus Ponens*.)

iii) if $A \in L_o$ then $\Box A$ L_o. (I.o.w., L_o is closed under *the Rule of Necessitation*.)

We omit the proof of the *theorem* that $A \in L_o$ iff A is ^{0}valid$_i$ (i {1,2,3,4}).

It is easy to extend L_o to a logic L^*_o which is correct and complete with respect to the notions of ^{0}validity$^*_i$ $(i \in \{1,2,3,4\})$.
L^*_o is the smallest subset of SEN which meets the following conditions:

i) L^*_o contains the following *axioms*:
- All axioms of L_o.
- All sentences $(\boxtimes A \supset A)$; $(must A \supset A)$.

ii) if $A \in L^*_o$ and $(A \supset B)$ L^*_o, then $B \in L^*_o$.

iii) if $A \in L^*_o$ then $\Box A$ L^*_o.

We omit the proof of the *theorem* that $A \in L^*_o$ iff A is ^{0}valid$^*_i$.

Let us compare L_o and L^*_o, and especially their counterfactual fragments, with Lewis's 'official logic of counterfactuals', his system VC.[32]

A cursory look yields the following differences:

- The sentences $((A \ \& \ B) \supset (A \boxed{c}\!\!\rightarrow B))$, which figure as axioms within VC, are not all provable within L^*_o. The sentences $(must(A \ \& \ B) \supset (A \boxed{c}\!\!\rightarrow B))$, all provable within L^*_o, can be regarded as their (weaker) substitutes. All other axioms of VC are provable within L^*_o.
- Some of the VC-axioms $((A \boxed{c}\!\!\rightarrow B) \supset (A \supset B))$ are not provable within L_o. The L_o-axioms $((A \boxed{c}\!\!\rightarrow B) \supset must(A \supset B))$ can be regarded as their (weaker) substitutes. Besides, within L_o not all sentences $must(A \ \& \ B) \supset (A \boxed{c}\!\!\rightarrow B)$ are provable; here we can only offer the still weaker $(may(C \vee \neg C) \supset (must(A \ \& \ B) \supset (A \boxed{c}\!\!\rightarrow B)))$ as alternatives to the VC-axioms $((A \ \& \ B) \supset (A \boxed{c}\!\!\rightarrow B))$. The remaining axioms of VC are all provable within L_o.

Problems arise if we try to extend L_o and L^*_o to logics L_1, L_2, L_3 and L^*_1, L^*_2, L^*_3 correct and complete with respect to the notions validity$_1$, validity$_2$, validity$_3$ and validity$^*_1$, validity$^*_2$, validity$^*_3$ respectively. To mention some of them: All sentences $(\oplus A \supset \Box\oplus A)$ are valid* and valid$_2$, but such is not the case with all sentences $\Box(\oplus A \supset \Box\oplus A)$. Hence L^*_2 and L_2 are not closed under the Rule of Necessitation. Likewise; all sentences $(\boxtimes A \supset \Box A)$ are valid$_3$ and valid$^*_3$, but such is not the case with all sentences $\Box(\boxtimes A \supset \Box A)$. Hence L^*_3 and L_3 are not closed under the Rule of Necessitation either.

I have not yet overcome all the difficulties arising from these phenomena. Therefore I cannot but conclude this article with some conjectures.

Let L^* be the smallest subset of SEN which meets the following conditions:

i) $L^{(*)}_1$ contains the following *axioms*:
- All axioms of $L^{(*)}_o$
- All sentences $((\oplus A \supset \Box\oplus A) \vee (\boxtimes B \supset \Box B))$

ii) if $A \in L^{(*)}_1$ and $(A \supset B) \in L^{(*)}_1$, then $B \in L^{(*)}_1$

iii) if $A \in L^{(*)}_1$, then $\Box A \in L^{(*)}_1$

(The reader will have understood that he is supposed to read the foregoing definition at least twice: reading $L^{(*)}_1$ and $L^{(*)}_o$ as L_1 and L_o the first time, and as L^*_1 and L^*_o the second time.)

Conjecture: $A \in L_1^{(*)}$ iff A is valid$_1^{(*)}$.

Let $L_2^{(*)}$ be the smallest subset of SEN which meets the following conditions:

i) $L_2^{(*)}$ contains the following *axioms*:

- All axioms of $L_1^{(*)}$
- All sentences $(\oplus A \supset \Box \oplus A)$

ii) if $A \in L_2^{(*)}$ and $(A \supset B) \in L_2^{(*)}$, then $B \in L_2^{(*)}$

iii) if $A \in L_1^{(*)}$, then $\Box A \in L_2^{*}$

Conjecture: $A \in L_2^{(*)}$ iff A is valid$_2^{(*)}$ ($A \in L_2^{(*)}$ iff A is valid$_4^{(*)}$).

Let $L_3^{(*)}$ be the smallest subset of SEN which meets the following conditions:

i) $L_3^{(*)}$ contains the following *axioms*:

- All axioms of $L_1^{(*)}$
- All sentences $(\boxtimes A \supset \Box A)$

ii) if $A \in L_3^{(*)}$ and $(A \supset B) \in L_3^{(*)}$, then $B \in L_3^{(*)}$

iii) if $A \in L_1^{(*)}$, then $\Box A \in L_3^{(*)}$

Conjecture: $A \in L_3^{(*)}$ iff A is valid$_3^{(*)}$.

Notes

0 I am indebted to E.C.W. Krabbe and J.B.M. van Rijen for their helpful criticism, to M.J. Petry and J. Vrieze for correcting the English, and to C.J.J. de Ruiter for typing the manuscript.

1 The reader will notice, however, that my approach is akin to Rescher's. Some of his views, expressed by himself in proof-theoretical terms, will re-emerge here, but then framed in the language of model theory. See especially Rescher (1964).

2 See Åqvist (1973), Gabbay (1972), Lewis (1971), Lewis (1973), Nute (1975a), Nute (1975b), Stalnaker (1968), and Stalnaker & Thomason (1970).

3 I disregard divergencies as to details (reckoning the Limit Assumption among them).

4 Stalnaker (1968), p. 109.

5 Lewis (1972), p. 91.

6 Popper (1972), p. 421.

7 Of course, it is nothing but a working hypothesis that such a truth definition is possible at all. Cf. note 14.

8 Thomason (1974), p. 67.

9 I am dividing each $\mathbb{O}^{\mathbb{w}'}_{<\mathbb{s},\mathbb{w}>}$ into two: prejudices, i.e. opinions a speaker is not willing to abandon, come what may, and assumptions, i.e. opinions a speaker is willing to give up if indeed he is forced to do so. Some refinements come to mind: we might introduce a partial ordering of each set of assumptions in order to account for the fact that a speaker may be less willing to give up his assumption $\mathbb{p}$ than his assumption $\mathbb{q}$. Introducing these partial orderings of each $\mathbb{O}^{\mathbb{w}'}_{<\mathbb{s},\mathbb{w}>} \sim \mathbb{P}^{\mathbb{w}'}_{<\mathbb{s},\mathbb{w}>}$, however, would only complicate the discussion in §3, whereas it would in the end make no difference whatever to the logic of counterfactuals.

10 It must be clear, however, that our cartoon does not distort reality to such an extent that $[\![A]\!]_{<\mathbb{s},\mathbb{w}>} = [\![A]\!]_{<\mathbb{s}',\mathbb{w}'>}$ for *every* $A \in$ SEN and $<\mathbb{s},\mathbb{w}>, <\mathbb{s}',\mathbb{w}'> \in \mathbb{C}$ (*must*, *may*, $\boxtimes$ and $\oplus$ behave like indexical expressions).

11 See Groenendijk & Stokhof (1975). Their account differs from mine in two respects: $\mathbb{O}^{\mathbb{w}'}_{<\mathbb{s},\mathbb{w}>}$ is only defined for $\mathbb{w}'=\mathbb{w}$, and $\mathbb{O}^{\mathbb{w}}_{<\mathbb{s},\mathbb{w}>}$ is a set of sentences (rather than propositions). Consequently, they can only attribute truth values to those *may*- and *must-statements*, in which *may* and *must* do not occur in iteration.

12 Groenendijk & Stokhof (1975), p. 69.

13 Take note of the fact that there is nothing strange in the following cases:

$<$*must*A, $<\mathbb{s},\mathbb{w}>, \mathbb{w}'>$ is false, whereas in fact $\mathbb{w}' \in [\![A]\!]_{<\mathbb{s},\mathbb{w}>}$

$<$*may*A, $<\mathbb{s},\mathbb{w}>, \mathbb{w}'>$ is true, whereas in fact $\mathbb{w}' \notin [\![A]\!]_{<\mathbb{s},\mathbb{w}>}$

$<$*must*A, $<\mathbb{s},\mathbb{w}>, \mathbb{w}'>$ is true, and in fact $\mathbb{w}' \in [\![A]\!]_{<\mathbb{s},\mathbb{w}>}$

$<$*may*A, $<\mathbb{s},\mathbb{w}>, \mathbb{w}'>$ is true, and in fact $\mathbb{w}' \in [\![A]\!]_{<\mathbb{s},\mathbb{w}>}$

$<$*must*A, $<\mathbb{s},\mathbb{w}>, \mathbb{w}'>$ is false, and in fact $\mathbb{w}' \notin [\![A]\!]_{<\mathbb{s},\mathbb{w}>}$

$<$*may*A, $<\mathbb{s},\mathbb{w}>, \mathbb{w}'>$ is false, and in fact $\mathbb{w}' \notin [\![A]\!]_{<\mathbb{s},\mathbb{w}>}$

14 Notice the ambiguities in 'It should have been raining yesterday'. Notice furthermore that the proposed truth condition for *should have been* is not in accord with our working hypothesis that the proposition expressed by a sentence A/uttered in a context $<\mathbb{s},\mathbb{w}>$ can be defined in terms of the propositions expressed by the component sentences B/uttered in the *same* context $<\mathbb{s},\mathbb{w}>$.

15 This, presumably, would be the strategy of Åqvist, Stalnaker and Thomason. Cf. Åqvist (1973), p. 4.

16 For further discussion see Rescher (1962).

17 I owe this formulation of the problem to a conversation with N. Stemmer.

18 Reactions to counterfactual statements form good clues to determine someone's prejudices. If the addressee replies with something like 'If ifs and ans were pots and pans, there would be no use of tinkers', then you have detected one.

19 Readers familiar with Lewis's theory may compare the subsequent remarks with his observations in connection with the Limit Assumption. See Lewis (1973), p. 19-21. The examples were suggested by E.C.W. Krabbe.

20 The particular choice of $\mathbf{P}^{w'}_{<s,w>}$ is not essential to our argument.

21 This inference pattern is discussed more fully in e.g. Lewis (1973). Its invalidity should be obvious from the following example (borrowed from Lewis): *If the USA threw its weapons into the sea tomorrow, there would be war. Hence, if the USA and the other nuclear powers all threw their weapons into the sea tomorrow, there would be war.*

22 See Kempson (1976).

23 See Stalnaker (1973).

24 See Grice (1975). This is the first publication of his William James Lectures at Harvard 1968.

25 Kempson (1975), p. 25.

26 This in view of the following maxim of conversation: '*Be sincere! Do not try to convince your partner of something which you do not believe yourself.*' Formally: a judgment $<A,<s,w>,w'>$ is conversationally incorrect if $[\![A]\!]^{(*)}_{<s,w>} \not\subseteq \mathbf{O}^{w'}_{<s,w>}$: Let us furthermore agree that every judgment is conversationally incorrect if $|\mathbf{O}^{w'}_{<s,w>}| = \emptyset$. I think we are entitled to stipulate this in virtue of the following maxim: *See to it that your opinions are mutually compatible!*

27 Consider a sentence 'If John had not killed him, who else would have done it?' stated by someone who, though he erroneously believes that John is not the murderer, wants nevertheless to witness against him.

28 $<(A \boxright B),<s,w>,w'>$ and $<(A \Diamond\!\!\rightarrow B),<s,w>,w'>$ are in any case incorrect if $|\mathbf{O}^{w'}_{<s,w>}| = \emptyset$. We assume therefore that $|\mathbf{O}^{w'}_{<s,w>}| \neq \emptyset$.

29 I do not want to say that such an incorrect counterfactual judgment becomes automatically correct if it is changed into the corresponding indicative judgment. Indeed, I think that even the judgments $<must(A\supset B),<s,w>,w'>$ are always incorrect if stated in the circumstances C2, C3 and C4. My *opponent can prove this on account of the maxims mentioned in footnote 26. We leave it to the reader to check this.

30 Lewis (1974), p. 4.

31 Throughout this section I shall refer to def. 6^G and *def. 6^G, and not to def. 6^F and *def. 6^F.

32 See Lewis (1973), p. 132-134.

References

- ÅQVIST, L. (1973), Modal Logic with Subjunctive Conditionals and Dispositional Predicates. Journal of Philosophical Logic 2, 1-76
- GABBAY, D.M. (1972), A General Theory of the Conditional in Terms of a Ternary Operator. Theoria 38, 98-104
- GRICE, H.P. (1975), Logic and Conversation. pp 64-75 in: Davidson, D. and G. Harman (Eds.), The Logic of Grammar, Encino, Cal.: Dickenson
- GROENENDIJK, J. and M. STOKHOF (1975), Modality and Conversational Information. Theoretical Linguistics 2, 61-112
- KEMPSON, R.M. (1975), Presupposition and the Delimination of Semantics. Cambridge: The Cambridge University Press
- LEWIS, D. (1971), Completeness and Decidability of Three Logics of Counterfactual Conditionals. Theoria 37, 74-85
- LEWIS, D. (1973), Counterfactuals. Oxford: Blackwell
- NUTE, D. (1975a), Counterfactuals. Notre Dame Journal of Formal Logic 16, 476-482
- NUTE, D. (1975b), Counterfactuals and the Similarity of Worlds. The Journal of Philosophy 72, 773-778
- POPPER, K.R. (1972), The Logic of Scientific Discovery. Sixth impression (revised). London: Hutchinson
- RESCHER, N. (1964), Hypothetical Reasoning. Amsterdam: North-Holland Publishing Company
- RESCHER, N. (1962), Lawfulness as Mind-Dependent, pp 178-197 in: Rescher, N. et al. (Eds.), Essays in Honor of Carl G. Hempel, Dordrecht: Reidel
- STALNAKER, R.C. (1968), A Theory of Conditionals. pp 98-112 in: Rescher, N. (Eds.), Studies in Logical Theory, Oxford: Blackwell
- STALNAKER, R.C. (1973), Presuppositions. Journal of Philosophical Logic 2, 447-457
- STALNAKER, R.C. and R.H. THOMASON (1970), A Semantical Analysis of Conditional Logic. Theoria 36, 23-42
- THOMASON, R.H. (1974), Introduction. pp 1-69 in: Thomason, R.H. (Ed.), Formal Philosophy. Selected Papers of Richard Montague, New Haven, Conn: Yale University Press