Syllabus Semantics II

General outline

Part I: Intensionality

a. Modal, Temporal and Intensional Logic
b. Intensionality in Montague Grammar

Part II: Dynamic Semantics (Discourse Representation Theory)

Themes:
(i) Anaphora
(ii) Conditionals and Quantification
(iii) Tense and Aspect
(iv) Lexical Semantics

Brief Description of Issues and Problems

I. Intensionality.

Classical logic - the First Order Predicate Calculus, Higher Order Predicate Logic and the Typed Lambda Calculus - provides us with the formal means to adequately describe *extensional* aspects of natural language meaning: reference, predication, boolean operations and quantification. But often the semantic contributions that the constituents of natural language sentences make to the meanings of those sentences are non-extensional and these classical logic is not designed to describe; to the extent that it can be made to describe them at all, it can do so only in an indirect, round-about way.

This difficulty was already known to Gottlob Frege (1848-1925), one of the two architects of classical predicate logic.[1] In fact, the problem preoccupied Frege for more or less his entire scientific career. The

[1] The other ‚inventor' of the predicate calculus was the American mathematician and philosopher Charles Sanders Peirce (1839-1914). Peirce's results in formal logic were, like much of his other work, not easily accessible during his lifetime, and historically his influence on the development of logic during its decisive period (roughly, the last quarter of the 19-th and the first third of the 20-th century) has been almost negligeable. A crucial difference was that Frege's work was continued in Russell and Whitehead's *Principia Mathematica*, the monumental work in which the authors develop a variant of Frege's formal system and apply it to the formalisation of a large part of mathematics.

most famous of all his attempts to come to grips with it is his essay *Über Sinn und Bedeutung* (1892). As Frege explains there, each expression of a natural language L has a sense (*Sinn*, as he calls it) as well as a denotation (or *Bedeutung*, in his terms). The formal logic that he had developed[2] – a system of higher order logic of which the core has survived as first order predicate logic as we know it today – could deal directly only with *Bedeutungen* – with "extensional constructions", as we would say today. It is suited for describing how the *Bedeutungen* of complex expressions depend on the *Bedeutungen* of their grammatical constituents. But natural languages, Frege was aware, are rife with sentences in which the *Bedeutung* of the whole does not depend on the *Bedeutungen* of its parts, but on their *Sinn*, and to deal with that kind of dependencies the logic of Frege's *Begriffsschrift* is not the right instrument. Frege appears to have made certain efforts towards the development of a formal "Logic of Sense and Denotation", but as far as we know he never really succeeded with this to his own satisfaction. Several proposals for such a logic were worked out in the course of the 20-th century, but none of them was wholly successful on all fronts.

However, among these systems there is one that has proved particularly useful for purposes of natural language semantics. This is the system of *Higher Order Intensional Logic* (HOIL) developed by Richard Montague (1928- 1971) in the mid-sixties. Montague's HOIL combines valmost unlimited expressive power with comparative notational simplicity. It is these virtues, as well as the effective use that Montague himself made of HOIL in the systematic descriptions that he gave of the semantics and logic of substantial fragments of English, that have made HOIL and systems closely related to it into the standard tools for doing semantics that they still are now.

Since non-extensionality - or *intensionality*, as we will henceforth also call it - is so wide-spread in natural language, and a semantics of natural language that cannot deal with it isn't worth its mettle, it is essential that we acquaint ourselves with the tools that HOIL and related logical systems provide. To do this, however, it is natural to begin by studying the logic which provided the inspiration for HOIL and which can still be considered its foundation. This is *Modal Logic*, as it had been shaped in the late fifties and early sixties, especially through the work of Saul Kripke. Since the standard systems of Modal Logic are a good deal simpler than HOIL, we will begin by studying them. And we will use the opportunity to also briefly explore systems of

2 Mainly in his *Begriffsschrift* of 1879

Temporal Logic (also known as Tense Logics) which are formally quite similar to the standard systems of Modal Logics, and which have come to play their own role in the history of the Formal Semantics of Natural Language.

##

In the last sentence of the previous paragraph I advisedly spoke of the history of "Formal Semantics of Natural Language". That history isn't very old; by the reckoning of many it dates from the second half of the sixties, when Montague did his path breaking work. But the subject of Semantics is much, much older than that. The earliest suggestions why certain words have the meanings they do have go back to antiquity. But then, and from then on for nearly the entire period that stretches up to the quite recent developments mentioned above, Semantics was confined almost entirely to the meanings of words. How word meanings get integrated into sentence meanings – how sentences "derive" their meanings from the meanings of the words from which they are built – wasn't perceived as a problem that needed investigation or explanation. In fact, it needed the development of formal logic, and not just of its syntax, but also of its model-theoretic semantics, to open our eyes to the true nature of the full range of problems that any theory of meaning in natural language will have to address.

We recall here that the syntax and semantics of predicate logic are perfectly attuned to each other. The syntax of the predicate calculus is given by a recursive definition of "well-formed term" and a similarly recursive definition of "well-formed formula". The model-theoretic semantics for this calculus then makes use of precisely these definitions: given a model M, it (i) specifies the values in M for all atomic terms, then (ii) defines the values in M for complex terms *following the recursive definition of terms*, (iii) then defines the truth values in M for atomic formulas and finally (iv) defines the truth values in M of complex formulas *following the recursive definition of formulas*. These model-theoretic definitions show that the syntax of a predicate logic term or formula gives a faithful representation of its meaning: the syntax structure tells us, as it were, directly how the meaning of the complex expression is obtained by stepwise integration form its ultimate constituents (the symbols of predicate logic). It is in this way that the denotations in any model M of the infinitely many terms and the truth values in M of the infinitely different formulas of predicate-logical systems are fixed by the interpretations that M assigns to the non-logical constants of the predicate calculus.

Doesn't something like this also apply to natural language? For many years since the time when the syntax and semantics of formal languages like the first order predicate calculus and the intimate relationship between them were first well understood, there were serious doubts that that would be possible. Natural languages, it was believed, were far too unsystematic, far too much infected by the accidents of historical developments and the whims of those whose "creative" use of language keep rocking their syntactic and semantic stability, to permit semantic analyses along such neat and stringent logical lines. We see this judgement reflected in the spirit in which the formal systems of mathematical logic were developed in the late 19-th and early 20-th century. These systems were intended to *replace* natural language, where it really mattered: in mathematics, science and philosophy, where the need for precise formulations, in an unambiguous and logically transparent vernacular, had become an urgent, and in some cases a vital, necessity.[3] Natural languages, it was felt, were all right so long as what you said, or how you said it, doesn't matter too much. That they are ultimately subject to the same kinds of general principles that govern the relationship between meaning and syntactic form in the languages of formal logic was a thought that none of the great logicians of the late nineteenth and the first half of the 20-th century (with the exception perhaps of Frege) seems to have seriously entertained.

For a linguist, however, the natural question that needs asking points in the exact opposite direction: How could a natural language NOT be governed by such general principles? How could we ever learn a language if it wasn't governed by such principles? And how could we ever use it, once we had learned it, in the way we do use the languages we speak. if that weren't the case? When someone learns a language, he learns its vocabulary – the words and their meanings – and its grammar – the rules that enable him to produce grammatical sentences which express what he wants to say, and to interpret the sentences that reach him from others. It is by applying the rules to the words that speakers can arrive at the sentences that correctly express their thoughts and interpreters recover the thoughts those sentences express; and this is surely the only way in which we can produce and understand sentences that we have never seen or heard before.

[3] This was so especially in mathematics, where in the course of the second half of the 19-th century paradoxes had been discovered that not only provoked a crisis in the foundations of mathematics itself but shook confidence in the trustworthiness of analytical method in all realms of thought. The imprecisions and ambiguities of natural language, as it was used in mathematics and science, were held at least in part responsible for this crisis, and it was hoped that formal languages would provide the badly needed sanitation.

These considerations are by now so obvious and familiar that it is difficult to imagine a situation in which the weren't obvious to everyone. But that was the situation at the time when Montague initiated the discipline which we now know as the "Formal Semantics of Natural Language". (The more specific form in which he pursued the aims of formal semantics is also known as "Montague Grammar".) The reason why Montague's work is so exceptionally important is that it went against the prejudices of his time and that it was done so well that many who thought that natural languages could not be analysed in rigorous terms became persuaded that they could.

Montague was convinced that the meanings of English sentences could be computed from the meanings of their words on the basis of syntactic composition rules that were close to those that had been assumed by traditional grammar. And while it is true that some of the syntactic rules he adopted were motivated by the need to make the semantics work smoothly, he managed by and large to stick to this guiding principle. But most importantly, the syntactico-semantic treatments of fragments of English he formulated conform to the *Principle of Strict Compositionality* – the principle that for each syntactic rule that can be used to form complex expressions out of smaller constituents there is a corresponding semantic composition rule, which forms the meaning of the complex expression out of the meanings of those constituents. The Principle of Strict Compositionality is one of the central methodological assumptions in Montague Grammar. As we will briefly see below, and much more extensively in the second part of this Syllabus, the Principle has to be modified in certain ways if certain phenomena are to be covered that Montague Grammar is not able to handle. But even if the approach we will study in Part II, which is designed to deal with those phenomena, does not adhere to the Principle in its strictest form, the Principle continues to function, there and elsewhere within Formal Semantics, as a kind of informal standard or ideal.)

One important ingredient to the success that Montague's work had in persuading the community that a systematic formal semantics of natural language is possible after all, was his treatment of *non-extensional constructions*. In particular, he offered treatments of those constructions that had had much (if mostly negative) attention from logicians and philosophers at the time, those involving propositional attitude verbs - verbs like *believe*, *want*, *intend* and the like – and verbs like *seek*, *be looking for* or *imagine*, whose direct objects often describe the *kinds of things* that the subject is said to seek, look for or imagine. Here are a couple of illustrations:

(1) a. Fred believes that Mary is in Paris.
b. Fred believes that your cousin is in Paris.
c. Fred believes that Eva is in Berlin.

(2) a. Mary is looking for a secretary.
b. Mary is looking for her secretary.
c. Mary is looking for your cousin.
d. There is a secretary Mary is looking for.

The first point of the examples in (1) is that the second argument of an attitude verb like *believe* – which in (1.a) is given by the *that*-clause *that Mary is in Paris* - is something like a sentence – an expression that, in the extensional semantics that has come down to us from Frege, is treated as "denoting a truth value", where the truth values are "True" and "False" (or, as we will assume following the usual convention, 1 and 0, for "True" and "False", respectively). However, it cannot be that the *that*-clauses in (1.a-c) do no more than contribute their Fregean denotations (i.e. their truth values) to the denotations (i.e. truth values) of the sentences (1.a-c) which contain them as constituents. For suppose that Mary is your cousin, that she is in Paris and the Eva is in Berlin. Then all three *that*-clauses are true. If the truth values of the sentences (1.a-c) were determined just by the denotations of their constituents, then we could conclude from this that they too must all have the same truth value. (For apart from their *that*-clauses the constituents of those sentences – the name *Fred* and the verb *believe* – are identical. So these constituents will contribute to the truth values of the sentences (1.a-c) in precisely the same way.) But this conclusion is clearly absurd. Even if the *that*-clauses of (1.a-c) are all true, the truth values of the sentences themselves may obviously vary. Fred may believe that Mary is in Paris, yet refuse to accept the *that*-clause of (1.b) because he doesn't know that Mary is your cousin, and has no opinion of any kind about the whereabouts of your cousin, who he assumes he doesn't know. And he may refuse to believe that Eva is in Berlin, because he has no opinion about Eva's whereabouts either, even though that is just as true as that Mary is in Paris.

What this somewhat tiresome argument shows is something we all know: what determines the truth value of a sentence like those in (1) is not the truth value of its *that*-clause but what that *that*-clause *says*. It is the *content* of the *that*-clause that matters to belief, and not whether that content happens to be actually true or false.[4] This implies, however, that an adequate semantics for sentences with the verb *believe* must have a way of dealing with propositional content. And that is something that classical logic doesn't do and for which we need an intensional logic like HOIL.

The sentences in (2) are non-extensional in a slightly different way than those in (1). In (2) it is the NP governed by the preposition *for* that contributes (or at least can contribute) more than its ordinary denotation. In extensional semantics the denotation of an NP is the object that the NP refers to (if the NP is a referential NP, e.g. a name or a definite description), and the value(s) of the bound variable corresponding to the NP, in case the NP is quantificational. Thus the NP *her secretary* in (2.b) denotes the person who is Mary's secretary and the NP *your cousin* in (2.c) denotes your cousin. And the denotation of the indefinite NP *a secretary* in (2.d) is a value for the variable, ranging over secretaries, that is introduced by this existentially quantifying NP.

If we look just at (2.b-d), we could get the impression that *look for* semantically behaves just like an ordinary transitive verb, which expresses a 2-place relation between individuals. In (2.b) and (2.c) it expresses a relation between Mary on the one hand and her secretary, or your cousin, on the other; and the natural reading of (2.d) is that there is some secretary x such that Mary stands in the look-for-relation to x. But when we turn to (2.a) we see that such an analysis cannot be right in general. What (2.a) is most naturally taken to express is that Mary is looking for someone who fulfils her criteria for a (good) secretary. Presumably she doesn't have any one particular person in mind from the start, she is just looking of suitable candidates. On this reading of (2.a) the object NP *a secretary* contributes to the truth value of the sentence something like the *concept* of being a secretary, and not some particular individual that instantiates the concept.

To deal with such "concept-oriented" interpretations of *look for* we again need to go beyond classical logic; and here too HOIL provides a framework that can provide a far better analysis of the non-extensionality involved than is possible with the inadequate means that are provided by extensional logics.

Note by the way that non-extensional aspects not only affect (2.a), but also sentences like (2.b) and (2.c). This time, let us assume that your

---

[4] Instead of the content of a sentence or clause one also speaks of its *propositional content* or of the *proposition* expressed; we will use these terms interchangeably.

cousin is as a matter of fact the same person as Mary's secretary, so that the NPs *her secretary* and *your cousin* have the same denotation. Even so the sentences (2.b,c) may differ in truth value. (Or more accurately, they allow readings on which their truth values may be different.) Mary may be looking for your cousin, because you, who do not know any more than she does that your cousin is her secretary, have told her that she should meet your cousin, since the two of them would (you said) get on famously with each other. In that case (2.c) would be true and (2.b) false. The converse possibilty – (2.b) true and (2.c) false – could easily arise as well.[5]

We will return at length to the examples in (1) and (2) in the second half of Part I, in which we will present a couple of explicit syntactico-semantic treatments of fragments of German that will be close in spirit to the work of Montague and that will serve as illustrations of the methods of Montague Grammar.

Part II.

The spirit of Montague's work has been the main guiding force in Formal Semantics of Natural Language from his days to our own. But nevertheless it became increasingly clear as time went on that natural languages cannot be analysed in quite the way he had proposed. There are, formal semanticists came to realise more and more, bigger differences between natural languages and the languages of formal logic than Montague had allowed for. (So, as far as these discrepancies are is concerned, the sceptics from Montague´s own days have been proved right to a certain extent, but only in the sense that formal semantics of natural language is different from and more complicated than what Montague's own work suggests, not in the sense that formal semantics is fundamentally impossible.) Perhaps the most dramatic departure from the architecture of Montague Grammar to date is found in what has come to be known as *Dynamic Semantics* (a term cast in the late eighties by the Dutch philosopher-logicians Jeroen Groenendijk and Martin Stokhof (Groenendijk & Stokhof, 1989?)n (Groenendijk & Stokhof, 1990?)).

The main difference between Dynamic Semantics and classical Montague Grammar is that in Dynamic semantics the principal "unit of semantic analysis" is not the sentence but the discourse (which typically will consist of more than one sentence). Connected with that is

---

5 Exercise 1: Think of a natural scenario in which (2.b) is true and (2.c) false.

that the meaning of a sentence is no longer treated as that which determines its truth value, but rather as that which enables the sentence to make a meaningful contribution to the discourse in which it occurs: Sentence meanings are “update potentials” - functions from given information packages (given by the antecedent part of the discourse) to new, extended information packages, in which the contribution of the given sentence has been incorporated.

The version of Dynamic Semantics we will present here is *Discourse Representation Theory* (or “DRT” for short; see (Kamp,1981), (Kamp & Reyle, 1993)). There are two connected reasons for doing so. First, DRT is more detailed when it comes to the actual “logical forms” (or “semantic representations” of natural language sentences and discourses and texts. Second, - and this is a more local and pragmatic reason – much of the research in semantics that is going on at the IMS of the University of Stuttgart, for which this Syllabus is written, uses DRT as general framework (not only in dealing with multi-sentence discourse, but also with individual sentences and with the semantics of words). Participants in the course in which the Syllabus is used should acquire some familiarity with the basic ideas and methods of DRT so that it is easier for them to become part of the local research community.

Here we give just two examples of the kind of phenomena that have motivated the move to Dynamic Semantics. First consider the two sentence “discourse” in (3). This discourse is made into a semantic whole by the pronoun *it* in the second sentence, which cn only be understood as referring back to the indefinite NP *a donkey* in the first sentence. Because of this anaphoric connection it is only the two sentences together that can be described a clear interpretation. Together they mean something like “Pedro owns a donkey that is unhappy”. We will argue in Part II that there is no real hope of analysing a discourse like (3) as the conjunction of two independent propositions, one expressed by the first sentence and the other by the second sentence. (The argument is not straightforward. But it is important and we will devote the attention to it that it needs and deserves.)

(3) Pedro has a donkey. It is not happy.

(4) If Pedro has a donkey, it won’t be happy.

That the two sentences of (3) form a single, indivisible semantic unit is an example of an extremely general and essential feature of natural language. To account for this feature one must explain how each new sentence of a discourse gets interpreted in the context presented by the preceding sentences. Such an account must incremental: it must first assign an interpretation/meaning to the first sentence of a discourse or text, then use this interpretation as context for the interpretation of the second sentence. That interpretation must take the form of integrating the contribution made by the second sentence into the context presented by the first, thus providing a context for the third sentence (or the final discourse interpretation, if the second sentence is were the discourse ends); and so on.

The phenomenon illustrated in (3) has its repercussions within single sentences like, for instance, (4). The difficulty represented by the anaphoric connection between the pronoun *it* in the consequent of the conditional (4) and the indefinite NP *a donkey* in the antecedent of the conditional that is the only NP we can construe as *it* 's anaphoric antecedent is not quite the same as that presented by the anaphoric connection in (3). This time the problem is not that the anaphoric relationship forges two sentences into a single semantic whole, but rather that the usual canon for translating natural language sentences into logical forms breaks down unexpectedly. One of the fist rules for translating natural language into predicate logic is that indefinite NPs should be translated as existential quantifiers. When we apply this rule to (4) the version that suggests itself most strongly is that in which the existential quantifier introduced by a donkey is given narrow scope, restricted to the antecedent of the conditional, as in (5.i). But of course this won't do, as the occurrence of x in the consequent of this formula t bound. (It can't be bound by the quantifier (∃x), since its scope is restricted to the conditional's antecedent.) Perhaps, it might be suggested in response to this, this is a case where the quantifier should be given wide scope, as a way to accommodate the need for the last occurrence of y to be bound as well. This would lead to the logical form on (5.ii). Now all occurrences of x are bound, but the truth conditions that (5.ii) assigns to (4) ar all wrong. (Note that (5.ii) is true as long as there is any thing that isn't both a donkey and owned by Pedro. Even Pedro himself presumable satisfies this condition.)

There is a predicate logical formula that does capture the truth conditions correctly (or close enough to correctly to satisfy us on this score.) This is (5.iii), in which a donkey has given rise to a universal quantifier with wide scope. The problem with this logical form is that it is hard to see how it could be generated from (4) by principles that aren't completely *ad hoc*.

(5) i. $(\exists x)(\text{donkey}(x) \,\&\, \text{own}(p,x)) \rightarrow \neg\text{happy}(x)$
ii. $(\exists x)(\text{donkey}(x) \,\&\, \text{own}(p,x) \rightarrow \neg\text{happy}(x))$
iii. $(\forall x)(\text{donkey}(x) \,\&\, \text{own}(p,x) \rightarrow \neg\text{happy}(x))$

In Part II we will look closely the incremental dimension of meaning and interpretation and the ways in which DRT deals with it. This will, we will see, also give us a natural way of explaining the semantics of (4).

Part I: Intensionality in Formal Semantics.

I.1 Modal Logic, Tense logic and Intensional Logic.

As noted in the introduction, the tool we will need for the analysis of intensional constructions of natural language that we will present in the second half of Part I is Montague's Higher order Intensional Logic. But as we also noted there, the central idea behind the treatment of intensionality in HOIL is one that comes from Modal Logic. Modal Logic is also of importance for Semantics in its own right. And moreover, it plays an important role in other parts of computational linguistics and of computer science more generally. This, and also the fact that Modal Logic is also simpler than HOIL are reasons for presenting it separately, partly as a preliminary to HOIL, but also for its own sake.

I.1.1 Modal Logic.

The central subject matter of Modal Logic are the notions of necessity and possibility. To study the *logic* of these notions we must look at the ways in which they interact with other logical notions. In this respect the modal notions are no different from those of quantification theory. There are some aspects of the logic of quantifiers that can be discovered by studying just them. But by and large, the logical properties of quantifiers reveal themselves only in their interactions with other logical notions, in particular the familiar sentence connectives *not*, *and*, *or* and *if ... then*. It is these interactions that are studied in predicate logic as we know it today. It shouldn't be taken for granted that these interactions – between the quantifiers and the sentence connectives are all that there is to say about the quantifiers; interactions with other notions might be important as well. But on the whole these interactions have proved to be especially important, and the immense usefulness of the predicate calculus as we have it today, which makes these interactions explicit, is a testimony to that.

The simplest systems of modal logic that have been used to study the logic of necessity and possibility embody the same choice as standard quantification theory: Look first at the interactions between these notions and the truth-functional connectives. Such a system is obtained by taking classical propositional logic and adding necessity and possibility to it in the form of two 1-place sentence connectives, (functioning syntactically just like the negation operator $\neg$). The most commonly used symbols for these connectives are $\Box$ for "it is necessary that" and ◊ for "it is possible that". These are the ones we will use too. Such a system is specified in Def. 1.

Def. 1 (Syntax of Standard Propositional Modal Logic)

1. Vocabulary:

   i. Propositional constants: q_1, q_2, q_3, ...
   ii. Connectives: ¬, &, v, →, ↔
   iii. Modal operators: □, ◊.
   iv. Parentheses: (,)

We define the *formulas* of Propositional Modal Logic in Backus-Naur form:

2. Formulas

   Form ::= q_i | ¬ Form | (Form & Form) | (Form v Form) | (Form → Form) | (Form ↔ Form) | □ Form | ◊ Form

Modal Propositional Logic allows us to study the logical properties of possibility and necessity along the same lines as that is done in formal logic generally: The logic of □ and ◊, in their interaction with the truth functional connectives ¬, &, v, →, ↔, is revealed by *logical validity* and *logical consequence*, i.e. by which formulas of the system are logically valid and which formulas follow logically from which others.

Nowadays the central method for investigating these questions is that of model-theoretic semantics: One defines a class of models for the given formal system and then defines:

(i) the logically valid formulas as those formulas that are true in all models; and

(ii) the relation of logical consequence as the semantic relation which holds between a premise set Γ and a putative conclusion A iff A is true in every model in which all the formulas of Γ are true.

The model-theoretic method, however, is a comparativels recent development in formal logic. This is also true for non-modal logic, both propositional logic and, especially, predicate logic. For instance, Frege, in his *Begriffsschrift* from 1879, characterised logical validity and logical consequence for predicate logic in the form of inference rules. Ot wasn't until 1929, when Gödel gave a precise semantic characterisation of validity and logical consequence for first order predicate logic. And it was only through the work of Tarski (1903-1982?) after the World War II that validity and logical consequence were given the model-theoretic definitions that have since become the standard.

On this respect the history of Modal Logic resembles that of Quantification Theory. Modal Logic (which, by the way, has, like non-modal logic, its roots in the work of Aristotle and his contemporaries), was cast in a formally precise form in the second decade of the 20-th century. maimly through the work of C.I. Lewis (1883-1964). Lewis studied systems of modal propositional logic in much the same way that systems of non-modal logic had been studied for several decades previously, defining the syntax of systems of modal logic essentially as we just did in Def. 1. He then set about to characterise logical validity of formulas and arguments by the method which was the only one that was known at the time: Lay down a number of intuitively valid axioms and/or inference rules and then define validity as follows.

(i) an argument $< \Gamma, A>$, with premise set Γ and conclusion A, is *valid* iff there is a proof of A from Γ using the given axioms and rules;
and

(ii) a formila A iss valied if there is a proof of A from the empty premise set $\varnothing$.

In the case of modal logic, however, this approach proved to be much more problematic than it is for non-modal logic (propositional or quantificational). Even in the case of non-modal logic the method isn't without pitfalls. For instance, some of Frege's inference principles were challenged at the time and later, and some debates over what is to count as valid continue to this day.[6] But the disputes surrounding proof-theoretic characterisations of validity in non-modal propositional and predicate logic are nothing when compared to those that have surrounded Lewis' axiomatic characterisations of the logical properties of necessity and possibility. This is indicated by the quite bewildering variety of different axiomatisations that Lewis himself came up with. Most of these axiomatisations impose different logics on these notions; they single out different sets of formulas and different arguments

[6] For instance, there are those who advocate a constructive conception of valid proof in mathematics and who have challenged the principles in his system that are responsible for the general validity of formulas of the form $A \vee \neg A$.

$<\Gamma,A>$ as "valid". But neither Lewis himself nor his comtemporaries were able to decide which of those axiomatisations captures the logic of necessity and possibility correctly.

There could be two reasons for this quandary that Lewis found himself in: either (i) there are different notions of possibility and necessity that play a role in our use of these notions and our intuitions of what is valid and what is not waver between these alternatives; or (ii) our understanding of modal notions is simply not precise enough to enable us to choose between the options Lewis offers. We will see below that it is not easy to determine whether it is (i) or (ii) that is responsible for the difficulty that people have in choosing between Lewis's options, and that presumably both play a role. But before we can address the matter we should present at least a small selection from the axiomatic characterisations that Lewis proposed. The choice we have made is motivated on the one hand by the model theory for the system of Def, 1 that we will present following our presentation of these Lewis axiomatisations, and on the other by the prominence that these systems have gained in current work on Modal Logic.

<u>Def. 2</u> (Some Axiom Systems for Standard Propositional Modal Logic)

1. *The system K*:

i. A complete axiomatization of classical propositional logic with Modus Ponens as only inference rule.

Modus Ponens:
$$\frac{A,\ A \to B}{B} \qquad \text{(M.P.)}$$

(N.B. the axioms specified by these two schemata include all substitutions involving formulas of modal logic for schematic letters in the chosen axiom schemata. For instance, if one of the axioms is $A \to (B \to A)$, then among its instantiations are not only, say, $\neg q_1 \to ((q_1 \to \neg q_2) \to \neg q_1)$, but also formulas like $\neg\Box q_1 \to (\Diamond(q_1 \to \neg q_2) \to \neg\Box q_1$. This remark also applies to the schemata mentioned below.)

ii. The axiom schemata:

$$\Box(A \to B) \to (\Box A \to \Box B) \qquad \text{(Distr.)}$$
$$\Diamond A \leftrightarrow \neg\Box\neg A \qquad \text{(Dual)}$$

iii. The inference rule of *Necessitation*:

$$\frac{\vdash A}{\vdash \Box A} \qquad \text{(Nec)}$$

2. *The system T*:

The system K, together with the axiom schema:

$$\Box A \rightarrow A \qquad \text{(T)}$$

3. *The system S4*:

The system T, together with the axiom schema:

$$\Box A \rightarrow \Box\Box A \qquad \text{(S4)}$$

4. *The system S5*:

The system S4, together with the axiom schema:

$$\neg \Box A \rightarrow \Box \neg \Box A \qquad \text{(S5)}$$

For each of these axiom systems a *proof* of a formula A from a premise set Γ is any finite sequence C_1, ... C_n, such that $C_n = A$ and each C_i (i= 1, .. n) is either (a) a member of Γ, (b) an instance of one of the axiom schemata of the system in question, (c) follows from two earlier lines in the proof by M.P. or (d) follows from one earlier line in the proof by the rule of Necessitation.

Which of these axiom systems correctly captures the logic of necessity and possibility? Or is that logic captured by none of them and do we need some other system? As mentioned above, it is difficult to answer this question. In fact, it is difficult for two different reasons. First, it is hard to see just by inspecting any axiom system if it generates proofs for all formulas and arguments that we would intuitively consider valid, and also that it doesn't generate too much. To find out whether a formula can be proved in such a system, or can be proved from some given set of premises, is in principle a matter of trial and error. (For some axiomatic systems proof search can be formulated as an algorithm, which answers each question “Is A provable from Γ?” after a finite number of steps with either a “yes” or a “no”. But establishing such algorithmic proof search methods can be quite difficult, and is often is. And even when this has been established and a proof search algorithm has been defined, applying the algorithm to particular questions of the form “Is A provable from Γ?” can still be quite elaborate and time consuming.)

The second reason is the specific difficulty with modal logic that we alluded to above. Our intuitions about what the logical properties of necessity and possibility really are seem to be so fuzzy that we don't even know how to decide for particular formulas of the system of Def. 1 whether to count them as logically valid (or as following logically from certain others), and this even for formulas that are quite simple. So we can’t even tell whether we would *want* them to be provable (or provable from given other formulas), quite apart from the question whether they can be proved within a given axiomatic system.

The range of axiomatic systems which Lewis bequeathed upon us - there are many, many more than the four displayed here – thus presents us with a true *embarras du choix*: We simply cannot tell which one should be selected. Perhaps there are different notions of necessity and/or possibility that surreptiously compete in our judgements and thus confuse them. In that case there wouldn’t be just one correct system of modal logic, but several, each capturing one of those competing notions. But that too is something that we cannot say for sure on the basis of Lewis’ proof-theoretic method alone.

In order to make progress with these questions it is natural to try to be more articulate about our pretheoretic conceptions of what these terms mean, or can mean, and only then to relate this our improved understanding to the problem of validity for formulas and arguments of the system of Def. 1. It is here that the model-theoretic method proves helpful, just as it has helped to sharpen our understanding of validity in non-modal logic. It wasn’t until the second half of the nineteen fiftes that this approach to modal logic got properly under way. It found its first culmination in the work of Saul Kripke (1940 -), (who at the time of his decisive break through was still in high school)[7]. The leading idea behind the model-theoretic approach to modal logic on which Kripke and others fastened was a refinement of an idea that goes back to Leibniz (1646-1716). According to Leibniz being

---

[7] Other important contributors to these early developments of the model theory for modal logic were Jaakko Hintikka (1929 -), Stig Kanger (1924-1988) and Richard Montague.

necessary, or being necessarily true, amounts to being true in all possible worlds. Leibniz believed that the actual world is only one among many possible worlds - worlds that could have been if God had decided to create them instead of the actual world in which we exist. in creating the world that is ours God simply chose one among the set of possible worlds and actualised that one.[8]

We can turn Leibniz' proposal for the meaning of (*it is*) *necessary that/necessarily* into a formal semantics for the formal system of Def. 1 by defining the models for that system as follows.

Def. 3 A *model* for the modal propositional logic of Def. 1 is a pair $\mathbb{M} = \langle W, F \rangle$, where W is a non-empty set (the set of "possible worlds" of $\mathbb{M}$) and F is an interpretation function. F assigns each propositional constant q_i a truth value at each world w from W. (Thus $F_w(q_i)$ is either 0 or 1 for all q_i and all $w \in W$.)

For the models $\mathbb{M}$ of Def. 3 we can define the *truthvalue of* a formula A of our modal logic *in* $\mathbb{M}$ *at* any world w, $[A]_{\mathbb{M},w}$, by the following clauses:

Def. 4

(i) $[q_i]_{\mathbb{M},w} = F_w(q_i)$

(ii) $[\neg A]_{\mathbb{M},w} = 1$ iff $[A]_{\mathbb{M},w} = 0$

(iii) $[A \,\&\, B]_{\mathbb{M},w} = 1$ iff $[A]_{\mathbb{M},w} = 1$ and $[B]_{\mathbb{M},w} = 1$, and similarly for the other truthfunctional connectives

(iv) $[\Box A]_{\mathbb{M},w} = 1$ iff for all $w' \varepsilon W$, $[A]_{\mathbb{M},w'} = 1$

(v) $[\Diamond A]_{\mathbb{M},w} = 1$ iff for some $w' \varepsilon W$, $[A]_{\mathbb{M},w'} = 1$

8 Leibniz also thought that God, being supremely rational, supremely knowledgeable and supremely good, chose the world that, all things considered, is the best possible one. If it does not always look like that to us, this is just because of our shortsightedness. (cf. the parody of this idea in Voltaire’s *Candide*.) This notion, of our world being the best of all possible worlds, plays no part in contemporary modal logic.

One of the basic formal results in modal logic is that the logic generated by this "Leibnizian" model theory is captured by the axiomatic system S5. That is: for any formula A of the sytem of Def. 1 and any set of formulas Γ we have:

(i) A is true in every model $\mathbb{M}$ at every world w of $\mathbb{M}$ iff A is provable in S5.

(ii) A is true in every model $\mathbb{M}$ at every world w of M such that all members of Γ are true in $\mathbb{M}$ at w iff A is provable from Γ in S5.[9]

The formal result that S5 captures the logic of the Leibnizian semantics might be regarded as speaking in favour of that system. But how conclusive is Leibniz' idea as an analysis of the concepts of necessity and possibility as we use them? Are there perhaps other ways of understanding possibility or necessity, which determine other logics than S5 and are reflected in other model theories than the one of Def. 3? In particular, could it be the case that there are other axiomatic systems within the multiplicity that Lewis defined and explored that capture such alternative conceptions of necessity and possibility?

One way in which we can look for alternatives to the model theory of Def. 3 is by asking whether the truth of "It is necessary that A" should always involve *all* possible worlds. Couldn't it be that at least on some occasions where we use "It is necessary that A" (or praphrases thereof) it is not all worlds that are envisaged, but only some subset – for instance the set of those worlds that are *plausible* alternatives to the actual world from some particular point of view? It is in this direction that the logicians of the late fifties and early sixties who were concerned with the semantics of modal logic, refines the Leibnizian analysis of necessity and possibility that is represented by the model theory of Definitions 3 and 4.

There are various ways in which the idea that not all possible worlds are always relevant to the evaluation of a necessity or possibility claim, and that the question which worlds are relevant may depend on context or point of view, can be made more precise. The way chosen by Kripke and others in the fifties and sixties is still fairly non-committal, at least

9 Neither this result nor the other formal results about modal logic we mention below will be proved here. Proofs of these results can be found in many introductory textbooks to modal logic. Useful and reliable for such questions is G. Hughes and M. Cresswell: *A New Introduction to Modal Logic*. Routledge, 1996

for a start. It simply assumes a relation R of *accessibility* between possible worlds – wRw' means that w' is among the worlds relevant (from the point of view represented by w) for the evaluation of modal statements at w. This leads to the following modification of the Leibnizian model theory of Definitions 3 and 4. Models are now triples of the form <W, R, F>, where W and F are as in Def. 3 and R is the acccessibility relation. And the clauses (iv) and (v) of Def. 4 are now replaced by clauses that make the truth values of formulas $\Box A$ and $\Diamond A$ dependent on R.

<u>**Def. 5**</u> 1. (Models for the system of Def. 1)

A *model* for the system of Def. 1 is a triple $\mathbb{M} = \langle W, R, F \rangle$, where W and F are as in Def. 3 and R, the so-called *accessibility relation*, is a binary relation between worlds in W (i.e. $R \subseteq W \otimes W$).

(N.B. models of this form are now generally referred to as *Kripke models*.)

2. (Truth definition for the modal system of Def. 3 for Kripke models)

Clauses (i)- (iii) of Def. 4 together with:

(iv') $[\Box A]_{\mathbb{M},w} = 1$ iff for all w' ε W such that wRw', $[A]_{\mathbb{M},w'} = 1$.
(v') $[\Diamond A]_{\mathbb{M},w} = 1$ iff for some w' ε W such that wRw', $[A]_{\mathbb{M},w'} = 1$.

But what logic is generated by this semantics? That depends on what properties that we attribute to the relation R? And to determone what those proerties should be proves to be quite hard – often, one feels, it is just as hard as making up whether a given formula should be regarded as valid. The following discussion should make this difficulty visible. We start with a comparatively uncontroversial property, viz. reflexivity. If we assume that, as a matter if general principle, R is reflexive, i.e. that (6) holds, then all instances of the schema in (7) will come out as valid.

(6) for all models M = <W,R,F> and all worlds $w \in W_M$, wRw.

(7) $\Box A \rightarrow A$

In fact, the following stronger result can be proved:

(8) (i) A formula A of the system defined in Def. 3 is provable (from the empty set of premises) in the axiomatic system T defined above iff $[A]_{M,w} = 1$ for all models satisfying (6) and all $w \in W_M$.

(ii) A formula A of the system defined in Def. 3 is provable from the set of premises Γ in T iff $[A]_{M,w} = 1$ for all models M satisfying (6) and all $w \in W_M$ such that $[C]_{M,w} = 1$ for all $C \in \Gamma$.

(N.B. Clearly (ii) includes (i) as the special case in which $\Gamma = \varnothing$. Conversely, for each statement about validity of the kind exemplified in (i) there is a generalisation about derivability from some arbitrary given premise set Γ. Below I will only mention statements about validity that are like (i); but in all these cases the corresponding generalisation in the sense of (ii) will hold as well.)

Should we adopt the general principle that the alternative relation R is reflexive? The assumption seems compelling: Surely the world w in which a statement of the form $\Box A$ is claimed and/or evaluated as true or false should itself be among those that count among the possible worlds that are relevant to the question whether $\Box A$ is true in w. If A isn't even true in the world in which this question is asked, how could it be *necessarily* true? Necessary truth surely entails, or presupposes, actual truth.

This sounds quite persuasive. If we are persuaded by it and adopt the reflexicity of R as a matter if general principle, then, as (8) tells us, the schema $\Box A \rightarrow A$ is adopted as a logically valid statement form. But do we really need the route via the new model theory, involving the accessibility relation R and its structural properties, to arrive at this conclusion? Can't we argue for the logical validity of $\Box A \rightarrow A$ simply by observing that necessary truth must entail actual truth? It very much looks like that, for the two considerations are virtually identical.

So much for the reflexivity of R and the validity of the distinctive axiom schema of T. But how do things stand with the axiom schemata that are characteristic for the other two Lewis systems we mentioned above, S4 and S5?

Let us, before addressing the conceptual issues that these schemata involve, first state the formal results about S4 and S5 that parallel result (8) for the schema $\Box A \rightarrow A$.[10]

(9) A formula A of the system defined in Def. 3 is provable (from the empty set of premises) in the axiomatic system S4 iff $[A]_{M,w} = 1$ for all models M in which R is a transitive relation and all $w \in W_M$.

(10) A formula A of the system defined in Def. 3 is provable (from the empty set of premises) in the axiomatic system S5 iff $[A]_{M,w} = 1$ for all models M in which R is an equivalence relation and all $w \in W_M$.

Exercise 2: 1. Show that all instances of $\Box A \rightarrow \Box\Box A$ are true in all transitive models at all worlds of those models.

2. Show that all instances of $\Box A \rightarrow A$, $\Box A \rightarrow \Box\Box A$ and $\neg\Box A \rightarrow \Box\neg\Box A$ are true at all worlds of all models $\langle W,R,F\rangle$ in which R is an equivalence relation.

(9) tells us - among other things - that the characteristic schema of S4, $\Box A \rightarrow \Box\Box A$, is valid if it is assumed as a matter of general principle that R is transitive. But what reasons could we have for adopting either the transitivity principle or the validity of $\Box A \rightarrow \Box\Box A$?

Can we argue that R must have the property of transitivity?

Here is an attempt to argue for the transitivity of R: What worlds are relevant for the evaluation of a given modal statement C at a world w may vary with the kind of statement C is, and perhaps also with other aspects of the context in which the statement is made. But once these factors have been taken into account, the set of worlds that are relevant to the evaluation has thereby been fixed: these are the worlds that are the relevant alternatives to w for the given evaluation of C and thus the ones that stand, in the model determined by the given parameters, in the relation R to w. Since it is just these worlds that are relevant, all evaluations of modal parts of C, which are required as part of the evaluation of C, are to be evaluated with respect to that set of worlds. This means in particular that if a subformula B of C has to be evaluated, as part of the evaluation of C, at some other world w' to which earlier steps in the evaluation of C have led, then it should still be the same set of worlds that

[10] As noted above, both (9) and (10) can be generalised in the same way that (8.i) can be generalised to (8.ii). (8), (9) and (10) are among the formal results that were proved by Kripke in the late fifties.

should be taken into account. Put differently, if a world w'' is relevant (under the given conditions) to the evaluation of B at some world w' in this set, then it is again part of the given set of relevant worlds and thus accessible from w: a world w'' that is accessible from a world w' that is accessible from w must be itself accessible from w. Conclusion: R is transitive.

We can also try to argue directly for the validity of $\Box A \rightarrow \Box\Box A$. Suppose that some instance of $\Box A \rightarrow \Box\Box A$ were not true. Then its antecedent $\Box A$ would be true, while its consequent $\Box\Box A$ would be false. If $\Box\Box A$ is to be false, then there must be the possibility for the statement in the scope of the outer $\Box$ of $\Box\Box A$, i.e. for $\Box A$, to be false. This is the possibiltiy of there being a possibility for A to be false. But isn't that just saying that there exists a possibility that A might be false? And that is just what the truth of the antecedent of $\Box A \rightarrow \Box\Box A$ – that is, the truth of $\Box A$, would seem to exclude. Conclusion: the antecedent of $\Box A \rightarrow \Box\Box A$ cannot be true without its consequent being true as well. So $\Box A \rightarrow \Box\Box A$ must be true.

Both of these arguments may seem plausible, but neither need be seen as conclusive. The first rests on certain assumptions about the way in which the set if relevant worlds is determined by the setting and form of a given statement – but why couldn't it be the case that one part of a given statement requires different possibile worlds for its evaluation than some other part? And the second argument assumes that what counts as a possibility from some possible perspective thereby counts as a possibility from the actual perspective from which we must understand the statement as a whole; but why, really, should that be so? Assuming that it must be so seems, when we think about it, pretty much like begging the very question that is posed by the validity of $\Box A \rightarrow \Box\Box A$.

The first argument involves also an additional difficulty. It argues for more than we asked for. For it argues not only for the validity of $\Box A \rightarrow \Box\Box A$, but also for the validity of the characteristic schema of S5, $\neg \Box A \rightarrow \Box \neg \Box A$. That this is so can be seen as follows. Suppose that w' is a world that is relevant to the evaluation of C at w and that w'' is a world relevant to the evaluation of parts of C at w', i.e. that w'Rw''. Then, according to our argument above, w'' also belongs to the set of worlds relevant to the evaluation of C and its parts. So evaluations of parts of C at w'' will once again involve just those worlds that belong to original set of worlds accessible from R. Thus w', which belongs to that set, will be relevant to evaluations at w'' and thus be accessible from w'', i.e. w''Rw'. This shows that T is not only transitive but also symmetric and thus, given the already estalbished assumption that R is reflexive, that it is an equivalence relation. So in the light of (10) above (and also in the light of Exercise 2) it would follow from the argument that the system we should adopt isn't S4, but at the very least S5.

The arguments that have been sketched above are just two among many attempts that have been made since the time when Kripke models were first proposed to determine what would be the “true” modal logic (or the different true modal logics). They illustrate the difficulties that these arguments tend to run into if they start from some pretheoretic notions of modality; and that this remains to be so when we have semantic as well as the syntactic intuitions that we can draw from. Seen in this light the model theory for modal logic we have presented doesn’t seem to help all that much.

Nevertheless it is only the development of its model-theoretic semantics that has allowed modal logic to become the extraordinarily useful tool that it is today. There are two reasons for this, neither of which is visible from what we said so far. First, there are many applications of modal logic in which the properties of the relation R can be determined much more clearly and precisely than seems possible for its original application, as the logic of necessity and possibility. Second, the analysis of modal notions themselves, such as subjunctive and counterfactual conditionals[11], has led to other relations between possible worlds than the accessibility relation R. This has led to a much richer spectrum of semantic analyses of modal notions than the models with a single relation R that we have considered so far. We will see examples of both these aspects of modal logic below during our brief foray into tense logic.

First, however, something that we need more directly on our way to HOIL. So far we have looked at the modal operators only in the context of propositional logic, which permits us to study their interactions with the truth functional connectives. But what we will need is a logic that enables us to analyse modal notions as they apply to actual preidcations, involving predicates and terms denoting individuals, and in their interaction with operators that bind individual variables, most of all the quantifiers. As a first step in this direction we now proceed to forumulate a system of *modal predicate logic*.

---

11 Counterfactual conditionals are conditionals that imply the faslity of their antecdents. An example is: “ if he hadn’t come, the party wouldn’t have been fun.” Subjunctive conditionls leave open whether their antecedent is true with perhaps a tendency towards falsehood. Example: “If he would come, it would certainly be more fun”.

I.1.2 Modal Predicate Logic.

The syntax of the system of modal predicate logic we will consider is obtained in the same way as the system of modal propositional logic specified in Def. 3. This time we start from a standard system for predicate logic and add the modal operators $\Box$ and $\Diamond$ to that. The syntax of a system of modal predocate logic of this sort is given in Def. 6.

<u>**Def. 6**</u> (Syntax of Standard Modal Predicate Logic)

1. <u>Vocabulary</u>:

   i. Individual Variables: v_1, v_2, v_3, ...
   ii. n-place Predicates: $P^n{}_i$, for all n,i ε ω
   iii. Individual Constants: c_1, c_2, c_3, ...
   iv. Connectives: $\neg$, &, v, $\rightarrow$, $\leftrightarrow$
   v. Quantifiers: $\forall$, $\exists$
   vi. Indentity: $=$
   vii. Modal operators: $\Box$, $\Diamond$
   viii. Existence predicate: E
   ix. Parentheses: (,)

2. <u>Terms</u>

   Term ::= v_i | c_i

3. <u>Formulas</u>

   Form ::= $P^n{}_i$ (Term$_1$, ..,Term$_n$) | Term$_1$ =Term$_2$| $\neg$ Form |
   (Form & Form) | (Form v Form) | (Form $\rightarrow$Form) |
   (Form $\leftrightarrow$ Form) |($\forall v_i$)Form | ($\exists v_i$)Form | $\Box$ Form |
   $\Diamond$ Form

There are problems connected with systems of modal predicate logic like that defined in Def. 6, which do not arise for systems for modal propostional logic. They become evident as soon as we turn to the semantics of modal predicate logic. Our models for the propositional language of Def. 3 were structures of the form <W,R,F>, with W a set of possible worlds, R a relation on W and F a function which assigns a truth value to each combination (w, q_i), with $w \in W$ and a q_i propositional constant. We can think of the third component of such a model as a function F which associates with each $w \in W$ an assignment F_w which gives a truth value for each propositional constant. Such an assignment is nothing but an extensional model for propositional logic. Thus a model <W,R,F> for modal propositional logic can be seen as providing an extensional model F_w for classical propositional logic for each $w \in W$. The natural analogue of this notion of a model for the case of modal predicate logic would be models of the form <W,R,M> in which M is a function that associates with each $w \in W$ an extensional model M_w for classical predicate logic. With respect to such models we can give a truth definition of essentially the same form as that given in Def. 5 for modal propositional logic, in which once again the modal operators are the only devices for building larger formulas out of smaller ones that lead from evluations at one world to evaluations at other worlds.

Models of this form were adopted by Kripke and others in the late fifties, when a model theory for modal logic was developed, and they are the ones that we still use today. This is so in spite of the fact that they are infected with the problem alluded to above, and that they are still in use is a reflection of the fact that the dificulty is intrinsic to the interaction between modality and quantification and arises for any plausible model theory for a system like that oif Def. 6. difficulty connected with the that we do neither find with the models for propositional modal logic of Def. 3 nor with those of Def. 5. In connection with models <W,R,F> the difficulty manifests itself in the relations between the universes of the models M_w. An extensional model for predicate logic has the form <U,F>, where U is the universe of of the model and F a function that assigns suitable extensions, relative to U, to the non-logical constants of the predicate-logical language in question: elements of U to individual constants, subsets of U to 1-place predicate constants and so on. It seems intuitively plausible that the universes U_w of the different extensional models M_w that are involved in a given model $\mathbb{M}$ = <W,R,M> will in general not be the same. I, for example, exist in the actual world, but surely there are countless other possible worlds in which I do not exist; and there seems no reason why worlds in which I do exist and worlds in which I don't should't occur as members of the world set W of a single model $\mathbb{M}$.

However, if we allow for this possibility – that different extensional models M_w belonging to the same model $\mathbb{M}$ may have distinct universes U_w – then trouble brews. The problem that comes with different universes for different possible worlds is known in the philosophical literature as the problem of "quantifying in". It confronts us when a modal operator is applied to a formula which contains free occurrences of variables which are then subsequently bound outside the scope of the operator. Consider for instance the formula

(11) $(\exists x)(Q(x,c) \;\&\; \Diamond\neg Q(x,c)) \;\&\; \Box(\neg(\exists x)Q(x,c) \rightarrow P(c))$

(11) says (a) that there exists an object u that stands in the relation Q to the individual c but of which it is possible that it would not have stood in that relation to c; and (b) that it is necessarily the case that if there hadn't been anything/one standing in the relation Q to c, then c would have had the property P. (For concreteness, suppose that c is a dachshund, that Q is the relation that holds between two persons or animals if they are playmates and P is the property of being bored. Then (11) says that c has a playmate but that the two might not have been playmates, and, further, that necessarily, if c didn't have any playmates, she would be bored.) The question that this and similar formulas provoke is: Which possible worlds are brought into play by the modal operators that the formula contains?

Suppose we want to evaliuate (11) in a given model $\mathbb{M} = \langle W,R,M\rangle$ at some world w in W. How does this evaluation work? In order for (11) to be true in $\mathbb{M}$ at w both its conjuncts must be true in $\mathbb{M}$ at w, so in particular its first conjunct. That is the case provided there is some individual x, presumably belonging to the universe U_w, so that u satisfies the formula $Q(x,c) \;\&\; \Diamond\neg Q(x,c)$ in $\mathbb{M}$ at w. This means (i) that u and the denotation of c stand in the relation Q in $\mathbb{M}$ at w - this is decided by the model M_w in the usual manner – and (ii) that $\Diamond\neg Q(x,c)$ is satisfied by u in $\mathbb{M}$ at w. It is this second requirement that is potentially problematic. The difficulty is that there may be many worlds in W in which either u or c or both do not exist. And we can have no a priori grounds for assuming that not some such worlds w' are accessiblre from w, i.e. that wRw'. According to the semantics for $\Diamond$ that we have adopted in our discussion of modal propositional logic $\Diamond\neg Q(x,c)$ is true in $\mathbb{M}$ at w (with u as value for x) iff there is some world w' such that wRw' and $\neg Q(x,c)$ is true $\mathbb{M}$ at w' (with u as value for x). But what if w' is a world in which u doesn't exist (i.e. u is not a member of U_w)? Presumably in such a world the relation Q doesn't hold between u and c – if u doesn't even exist in w', how could it be a playmate of c in w' (even assuming that c does exist in w') Surely you can't have non-existent individuals for playmates!

Such a world w' would, you might say, verify $\Diamond\neg Q(x,c)$ at w on the cheap. It would be a world in which $\neg Q(x,c)$ is true, but intuitively it is not the kind of world we want for the verification of $\Diamond\neg Q(x,c)$ at w. What we want for a proper verification is an accessible world w' from w in which u and c do both exist, but in which they nevertheless are not playmates.

Much the same point can be made in connection with the second conjunct of (11), $\Box(\neg(\exists x)Q(x,c) \rightarrow P(c))$. For this conjunct to be true in $\mathbb{M}$ at w it must be the case at all worlds w' that are accessible from w that if $\neg(\exists x)Q(x,c)$ is true in $\mathbb{M}$ at w' then so is P(c). This time the problematic worlds are accessible worlds w' in which c does not exist. By the same reasoning as above, we may assume that there is no playmate of c in such a world w'. For how could anything – be it man or beast - be the playmate of something non-existent? So $\neg(\exists x)Q(x,c)$ should be true in $\mathbb{M}$ at such a w'. On the other hand P(c) would presumably be false at w'. For how can you be bored if you don't even exist.[12] An accessible world w' in which c doesn't exist would thus falsify the claim $\Box(\neg(\exists x)Q(x,c) \rightarrow P(c))$ at w. But in this case it would seem thatthe claim is falsified for the wrong reason. What $\Box(\neg(\exists x)Q(x,c) \rightarrow P(c))$ seems to claim intuitively is that our dachshund c is bored in all relevant possible worlds in which c *exists* and is nevertheless without playmates.

The general moral is thus that when we evaluate modal claims about existing inviduals we should look only at possible worlds in which these individuals also exist.[13] But how can we guarantee that only such worlds are brought into play for the evaluation of a modal formula at some world w? Note that it cannot be right to restrict R to the point that the problem of accessible worlds with problematic existence failures doesn't arise any more. For instance, it might be that among the worlds that are intuitively accessible from the world w there are some in which the individual denoted by c exists but that denoted by some other constant c' does not. Such worlds should be relevant to evaluation at w of a claim like $\Diamond P(c)$, while not relevant to evaluation at w of, say, $\Diamond P(c')$; on the other hand there might be intuitively accessible worlds where c' exists but c does not. Those worlds could be relevant to the evaluation of $\Diamond P(c')$, but would not be relevant to the evaluation of $\Diamond P(c)$. If we restrict R in such a way that we exclude worlds of both these kinds,

---

12 Non-existence may seem rather boring to us; but that doesn't justify the assumption that it would be boring to the non-existent. That assumption just doesn't make sense.

13 If there are any modal claims for which this is not so, then they must be of a rather "philisophical", speculative sort. We will leave such claims, if any there be, to the metaphysicians.

then we might throw away both the worlds we need for an intuitvely correct evaluation of $\Diamond P(c)$ and the ones we need for a correct evaluation of $\Diamond P(c')$.

One way out of this difficulty is to make use of an *existence predicate* E, whose extension in each model M_w of a model $\mathbb{M}$ is the set of things existing in w, and to "restrict" both quantifiers and modal operators to E by adding atomic formulas involving E in the scopes of quantifiers and operators. This procedure will for instance turn (11), the formula we used as formalisation of our modal claim about the dachshund c, into the formula in (12).

(12) $(\exists x)(E(x) \,\&\, Q(x,c) \,\&\, \Diamond(E(x) \,\&\, E(c) \,\&\, \neg Q(x,c))) \,\&$
$\Box((E(c) \,\&\, \neg(\exists x)(E(x) \,\&\, Q(x,c))) \rightarrow P(c))$

With (12) the problems we ran into in connection with (11) no longer arise; all evaluations of predications at w or worlds accessible from w that arise in the evaluation of (12) at w now take place in worlds in which the arguments of the predications exist (in the sense of satisfying E); and that, we saw above, is just the way that things should be.

It might be objected that this solution is not optimal in that it still leaves many formulas of the system of modal predicate logic of Def. 7 with evaluations that are often counterintuitive in that they involve predications of non-existants. For instance nothing has changed as regards (11). That of course is true. But that need not be too much of a problem so long as we restrict our use of the formalism to formulas like (12), in which the E-predications make sure that no ecaluations of other predications will ever involve non-existent arguments.

If it is only the "E-protected" formulas whose semantics is of direct interest to us, we need not be too concerned about what our model theory does to the others. In particular, there is less now that hangs on how we define the ranges of the quantifiers $\exists$ and $\forall$: So long as the range of a quantifier evaluated at w includes the things existing at w, the evaluation of any E-protected formula $(\exists x)A$ or $(\forall x)A$ at w will come out the same irrespective of what else may be included in its range. So we have a certain freedom here.

One of the options that are compatible with the requirement that the range of a quantifier at w include all entities that exist at w is to take the range to exist just of the things existing at w. A second option, which is common in the literature and which we will again encounter below when we get to HOIL, is as follows. One assumes that each model

$\mathbb{M}$ comes with a fixed set U of "possible individuals" or "possibilia", and treats the quantifiers as ranging over this same set at each world. However, in each world w normally only some of the possibilia in U will exist. These are the elements of U that belong to the extension of E at w.

The simplest way to formalise this idea is to make the set U into the universe of each model M_w of $\mathbb{M}$ and let E carve out at each world w the set of entities existing at w as its extension at w. This is the solution we adopt here. It leads us to the following notion of a *model* for modal predcate logic. Our models are quadruples $\mathbb{M} = \langle W, R, U, F \rangle$, where W and R are as before, U is the set of *possibilia of* $\mathbb{M}$ and F is a function which assigns at every $w \in W$ an interpretation, relative to U, to each non-logical constant from the vocabulary in Def. 6. These models are given in Def. 7. All that we need by way of motivation for them has been said, but there remains one point that should perhaps be stressed. In classical logic we assume that all models have non-empty universes. That assumption isn't strictly forced upon us, but it makes the logic simpler – it gives us classical logic as we know it – and for that reason it has by now become the standard. We want to maintain this convention here too, so that the non-modal formulas of our modal predicate logic retain the classical logic that also results from the standard extensional model theory for predicate logic in which universes are always non-empty. For our models for modal predicate logic this amounts to the assumption that at every world w of any model $\mathbb{M}$ the set of entities existing at w is non-empty.

At last we can proceed to the formal definition of the class of models for modal predicate logic.

Def. 7 A *model* for modal predicate logic is a quadruple $\mathbb{M} = \langle W, R, U, F \rangle$, where

- (i) W is a non-empty set (the set of "possible worlds" of $\mathbb{M}$);
- (ii) R is a 2-place relation on W;
- (iii) U is a non-empty set (the set of possibilia of $\mathbb{M}$);
- (iv) F is an interpretation function, which assigns to each $w \in W$ suitable extensions at w for all the non-logical constants; that is, for each individual constant c_i $F_w(c_i) \in U$ and for each n-place predicate P^n_i, $F_w(P^n_i) \subseteq U^n$.
  The extensions of the special 1-place predicate E are never empty: For all $w \in W$, $F_w(E) \neq \emptyset$.

If $\mathbb{M}$ = <W, R, U, F>, is a model for modal prediate logic, then for any $w \in W$ we denote as "M_w" the model < U,F_w> for classical predicate logic that $\mathbb{M}$ specifies for the world w.

In order to define the truth values of formulas of modal predicate logic in these models, one need - as always when quantifiers are involved – assignments of entities to variables. In the case of modal predicate logic it isn't immediately clear how assignments should be defined, and – somewhat surprisingly perhaps – much hangs on the precise in which we define them. To get a sense of the options between which we have to choose, consider first the way in which the models of Def. 7 treat individual constants. Def. 7 allows individual constants to vary their denotations from world to world; it is possible that $F_w(c_i) \neq F_{w'}(c_i)$. From a conceptual point of view this possibility seems reasonable, as we find something similar with certain definite descriptions in natural language, such as e.g. *the president*: As a rule the person who actually *is* the president need not have been the president; i.e. in another possible world someone else would have been elected in which case that person would be the denotation of *the president* in that world. If we want to use individual constants to represent such descriptions, then individual constants too should be allowed to vary their denotations between worlds.

But what about assignments to variables? Should the values they assign also be allowed to vary between worlds? In the fifties and sixties this question was the subject of intensive debate. The decision which emerged from that debate, and that was to be one of the important features of Montague's HOIL, is now more or less the standard. According to this decision assignments to variables do *not* vary: An assignment in a model $\mathbb{M}$ assigns each variable a single value (an element from the set U of $\mathbb{M}$'s possibilia) which will then count as the value of that variable at all the worlds of $\mathbb{M}$.

To appreciate the merits of this decision let us consider once more the first conjunct of formula (11):

(13) $(\exists x)(Q(x,c)\ \&\ \Diamond \neg Q(x,c))$

As we saw, the intuitive content of this formula can be informally expressed as: "There is something that stands to c in the relation Q but which could very well not have stood to c in this relation". The second part of this paraphrase apparently says that the very thing which stands to c in the relation Q in the actual world might not have stood in that relation to c, i. e. that there is some possible world accessible from the actual world in which *that very same thing* does not stand to c in the relation Q. A notion of assignment which allows assignments to specify different values for the same variable at different worlds could not do justice to this interpretation of (13). For it would then be possible for the assignment used to evaluate the existential quantifier (∃x) in (13) to assign one value to the variable x in the actual world and another values to x in some possible world in which Q(x,c) is false. Such an evaluation would determine (13) as true, but it would do that for the wrong reason, since the individual that fails to satisfy Q(x,c) in the possible world wouldn't be the same as the one that does satisfy Q(x,c) in the actual world.[14]

The notion of assignment for which we have argued is given in Def. 8. The notion is used in the truth defintion Def. 9.

<u>**Def. 8**</u> Let $\mathbb{M}$ = <W, R, U, F> be a model for modal predicate logic. An *assignment in* $\mathbb{M}$ is a function from the set of variables to elements of U.

<u>**Def. 9**</u> Let $\mathbb{M}$ = <W, R, U, F> be a model for modal predicate logic And let a be an assignment in $\mathbb{M}$. The *value of* an expression A *in* $\mathbb{M}$ *at* w *under* a, $[A]_{\mathbb{M},w,a}$, is defined as follows:

A. <u>Terms</u>

i. $[v_i]_{\mathbb{M},w,a} = a(v_i)$
ii. $[c_i]_{\mathbb{M},w,a} = F(w)(c_i)$

---

14 Note that when we focussing on (11) in arguing for the models of Def. 7, we assumed implicitly that the constant c in (11) denoted the same individual in the different possible worlds we considered. As we have just seen, the models of Def. 7 do not guanrantee that individual constants behave this way. We could, however, designate certain individual constants for this special role – we could treat them as *rigid designators,* in the terminology of modal logic: An individual constant c behaves as a rigid designator in a model $\mathbb{M}$ if $F_w(c) = F_{w'}(c)$ for all w, w' ∈ W.
There are additional reasons for wanting non-varying variable assignments; but as far as I can see, they all turn in some way on the point that is central to this example. The full range of reasons for wanting to define assignments the way we do here is quite complex and will not be discussed. Other solutions to the assignment issue in modal predicate logic (and more generally to the treatment of variables in modal and intensional logic, can be found in David Kaplan's Doctoral Dissertation *Foundations of Intensional Logic* and in Church's Logic of Sense and Denotation (Church. ??)

B. Formulas

i. $[P^n{}_i(t_1, ..., t_n)]^{\mathbb{M},w,a} = 1$ iff $<[t_1]^{\mathbb{M},w,a},..,[t_n]^{\mathbb{M},w,a}> \in F_w(P^n{}_i)$

ii. $[t_1 = t_2]^{M,w,a} = 1$ iff $[t_1]^{\mathbb{M},w,a} = [t_2]^{\mathbb{M},w,a}$

iii. $[\neg A]^{\mathbb{M},w,a} = 1$ iff $[A]^{\mathbb{M},w,a} = 0$

iv. $[A \,\&\, B]^{\mathbb{M},w,a} = 1$ iff $[A]^{\mathbb{M},w,a} = 1$ and $[B]^{\mathbb{M},w,a} = 1$ and similarly for the other truthfunctional connectives

v. $[(\forall v_i)A]^{\mathbb{M},w,a} = 1$ iff for all $u \in U$, $[A]^{\mathbb{M},w,a[u/v_i]} = 1$ (and similarly for $)

vi. $[\Box A]^{\mathbb{M},w,a} = 1$ iff for all $w' \in W$ such that wRw', $[A]^{\mathbb{M},w',a} = 1$

vii. $[\Diamond A]^{\mathbb{M},w,a} = 1$ iff for some $w' \in W$ such that wRw', $[A]^{\mathbb{M},w',a} = 1$

As usual, the truth values of sentences, or closed formulas (= formulas without free variable occurrences) do not depend on the choice of assignment. So for a sentence A we can simply speak of its truth value in a model $\mathbb{M}$ at a world w, suppressing reference to assignments. Often "Sentence A is true in $\mathbb{M}$ at w" is denoted as "$\mathbb{M},w \models A$".

Validity is defined as for modal propositional logic. This time we give the definition explicitly:

Def. 10 Let A be a sentence of modal predicate logic and Γ a set of such sentences. Then:

A *follows logically from* Γ - in symbols $\Gamma \models A$ - iff for every model $\mathbb{M}$ and every world w from $\mathbb{M}$:

if for all $C \in \Gamma$, $\mathbb{M},w \models C$, then $\mathbb{M},w \models A$.

A is *logically valid* iff $\varnothing \models A$.

Exercises

1. Universal Instantiation in Modal Predicate Logic

One curiosity of the logic generated by the semantics of modal predicate logic that is given by Definitions 7 – 10 is that the rule of Universal Instantiation (UI) is no longer valid. (The rule UI says (among other things) that for any universal formula $(\forall x)A$ and any closed term t $A[t/x]$ can be derived from $(\forall x)A$.) An example where UI fails is the inference of (15) from (14): (15) does not follow logically from (14).

(14) $(\forall x)P(x) \rightarrow \Box P(x))$
(15) $P(c) \rightarrow \Box P(c)$

Construct a model $\mathbb{M}$ with a world w such that
$\mathbb{M},w \models (\forall x)P(x) \rightarrow \Box P(x))$, but not $\mathbb{M},w \models P(c) \rightarrow \Box P(c)$.

Hint: UI fails because assignments are "rigid", while individual constants are in general not rigid.

For an intuitive counterexample, suppose (contrary to modern conviction, medical knowledge and surgical expertise) that the property of being male is an essential property of people: that whoever has this property has it necessarily. This claim is expressed by (14) if P is interpreted as denoting the property "male". So according to our assumption (14) is true. If on the other hand the constant c stands for "the president", then the "instance" (15) of (14) need not be true. It won't be true if the president happens to be a man, since the president could very well have been a woman. One way to obtain a formal counterexample is to define an $\mathbb{M}$ and w which reflect this intuition.

2. The Barcan Formulas.

Many of the disputes over the treatment of quantification in modal logic during the fifities and sixties focussed on the validity or invalidity of the so-called *Barcan formulas*, which turn on the interaction between quantifiers and modal operators.[15] There are two Barcan formulas, given below in (16) and (17)

(16) $(\forall x)\Box P(x) \rightarrow \Box(\forall x)P(x)$
(17) $\Box(\forall x)P(x) \rightarrow (\forall x)\Box P(x)$

---

15 After the American philosopher Ruth Barcan-Marcus, who was the first to draw attention to the importance to these formulas.

Show: (i) On the semantics of Def's 7 – 10 both (16) and (17) are logically valid.

(ii) The "E-protected" versions (18) and (19) of (16) and (17) are not logically valid.

(18) $(\forall x)(E(x) \rightarrow \Box(E(x) \rightarrow P(x))) \rightarrow \Box(\forall x)(E(x) \rightarrow P(x))$

(19) $\Box(\forall x)(E(x) \rightarrow P(x)) \rightarrow (\forall x)(E(x) \rightarrow \Box(E(x) \rightarrow P(x)))$

(iii) If we drop the assumption from Def. 7 that for all models M_W, $M_{W'}$ involved in a model $\mathbb{M}$ $F_W = F_{W'}$, then (16) and (17) are not logically valid.

3. Protecting E-predications

In the text we looked at one example of how a formula can be "protected" from the need to evaluate predications with non-existent arguments by adding predications involving the existence predicate E. In our example the unprotected formula was (11) and the result of introducing the protections was (12). We repeat the two formulas.

(11) $(\exists x)(Q(x,c) \,\&\, \Diamond\neg Q(x,c)) \,\&\, \Box(\neg(\exists x)Q(x,c) \rightarrow P(c))$

(12) $(\exists x)(E(x) \,\&\, Q(x,c) \,\&\, \Diamond(E(x) \,\&\, E(c) \,\&\, \neg Q(x,c))) \,\&$
$\Box((E(c) \,\&\, \neg(\exists x)(E(x) \,\&\, Q(x,c))) \rightarrow P(c))$

From this one example it is not completely clear how protecting E-predications should be introduced in general; and in fact, it is not obvious that there is just one recipe for introducing them that is optimal in all cases where we use formulas to represent given propositions or sentences from some natural language. But here is one recipe, which seems to do what we want for many such applications. It can be described as follows:

By the *scope* of a quantifier of modal operator understand the formula to which the quantifier or operator is applied. For instance, A is the scope of the outer universal quantifier $(\forall x)$ in the formula $(\forall x)A$ and B is the scope of $\Box$ in the formula $\Box B$.

The recipe for introducing protecting E-predications is now:

Let A be the formula that is to be protected, let B be any subformula of A beginning with a quantifier or modal operator (B is either a proper subformula of A or A itself) and let C be the scope of the outer quantifier or operator of B. Let α_1, ... α_k be a list of all the individual constants occurring in C and all the variables that have free occurrences in C. Form the conjunction $E(\alpha_1)$ & ... & $E(\alpha_k)$. (We refer to this conjunction below as "$E(\alpha_1 .. \alpha_k)$".) In case B begins with either ∀ or □, replace C by ($E(\alpha_1 .. \alpha_k)$ → C) and in case B begins with either ∃ or ◊, replace C by ($E(\alpha_1 .. \alpha_k)$ & C).

It can be seen that (12) does not strictly conform to this recipe, since the conjunct E(c) is missing from the scope of either occurence of the quantifier (∃x). A strict application of the recipe to (11) produces the formula (20).

(20) (∃x)(E(x) & E(c) & Q(x,c) & ◊(E(x) & E(c) & ¬Q(x,c))) &
□((E(c) & ¬(∃x)(E(x) & E(c) & Q(x,c))) → P(c))

It is easy to verify that the conjunct E(c) in the scope of the second occurrence of in (20) is redundant: (20) is logically equivalent to the formula we get when we omit this conjunct. (Task: show this!) But this is not so for the first new conjunct E(c) in (20). We did not bother to invlude this conjunct when discussing (12) since we assumed implicitly that c denotes an existing entity in the "acstual" world in which we assumed (11) and (12) were being evaluated. but at worlds where the denotation of c does not exist the presence of E(c) as a conjunct in the scope of the first quantifier (∃x) clearly makes a difference.

<u>Task:</u> Apply the recipe described above to introduce E-protections in the followlwing formulas:

(i) ◊P(c) & ◊¬P(c))
(ii) (P(c) v ¬P(c)) & ◊P(c) & ◊¬P(c))
(iii) (∃x)(∃y)(Q(x,y) & □((∃z)Q(x,z) → (∃z)Q(z,y))
(iv) □ P(c) → □ □ P(c)
(v) P(c) → □ ◊ P(c)

I.1.3 Tense Logic.

Our discussions of modal propositional logic in Section I.1.1 led us to the conclusion that it is very difficult to determine which system gives us the “right” modal logic. And that problem, we found, arises not only for attempts to characterise validity in proof-theoretic terms along the lines pursued by Lewis, but also for semantic characterisations in terms of a model theory à la Kripke. When we proceed model-theoretically, what logic we get depends on what general properties we are prepared to attribute to the accesibility relation R. But what properties we should attribute to it seems hard to decide on the basis of our nebulous intuitions about necessity and possibility. Hoever, as we remarked already, this has not prevented the Kripke Semantics for modal logic from becoming immensely useful in a wide and still growing spectrum of applications.

The reason for this success is that the applications of modal logic are almost without exception outside the domain of the possible and necessary as such, and typically enable us to make assumptions about the properties of R that are much more solidly grounded than is possible on the basis of our poor pretheoretic intuitions about necessity and possibility. The one example of such an application outside the realm of modality proper we will discuss here is *Tense Logic*. Just as the modal systems we discussed in I.1.1 were developed in order to study the properties of the operators $\Box$ and $\Diamond$ which enable us to speak not just of what is actually the case but also about what could or would have been, so systems of Tense Logic were designed to study temporal operators - operators which enable us to speak not just of what is the case roght now, but also of what was the case or happened in the past and of what be the case, or will happen, in the future.

A further feature that the first systems of Tense Logic we willl look at share with the systems of modal logic of the preceding sections is that in both cases the concepts at issue are treated as (1-place) propositional operators. And in both cases this choice was motivated by the ways in which we tend to express these concepts in language. The modalities are typically expressed as sentence adverbs – “possibly”, “necessarily” – or as phrases like “It is necessary” or “It is necessarily the case”, which require *that*-complements to become complete sentences, and all of these seem to function semantically as operators that turn propositions into other propositions. Likewise, it was thought by those who developed these systems that the tenses of the verb can be seen as operators that turn propositions into others, and that therefore sentence operators were the natural way to formalise them. For instance, the past tense can be seen as an operator which turns the proposition expressed by a present tense sentence into the proposition expressed by the corresponding past tense sentence. For example, it will turn the proposition that it is raining into the proposition that it was raining. Analogously the future tense will turn the proposition that it is raining into the proposition that it will be raining.

We note already at this point that there are many ways to study the "logic of time", or even to understand what the "logic of time" means and that even among the logic-based approaches towards the study of time, operator-based tense logics represent just one of several possibilities. In fact, it was realised fairly soon that they are not particularly useful in the study of time as it is expressed in natural languages. We will discuss a number of reasons for this in Part II, in which we will adopt a quite different approach. (Some of these systems, however, have proved to be of great value in certain areas of computer science, in particular for program verfication and the testing needed in chip design, but that is of no concern to us here. See also the remarks on this in I.1.3.3.)

The simplest of the systems of tense logic we will consider here contain 1-place "tense operators" P and F, correspondoing to the simple past and the simple future tense, respectively. These two operators stand in a very different relation to each other, however, than the operators $\Box$ and $\Diamond$ of our systems of modal logic. $\Box$ and $\Diamond$ are duals in the same way that $\forall$ and $\exists$ are duals. This is shown by the "modal law" $\Diamond A \leftrightarrow \neg\Box\neg A$ (see the axiom system T in Def. 2), which is the exact formal analogue of the duality law $(\exists x)A \leftrightarrow \neg(\forall x)\neg A$ of classical quantification theory. (The similarity of $\Diamond$ to the existential and of $\Box$ to the universal quantifier are of course also directly visible in the clauses for $\Diamond$ and $\Box$ in the truth definitions for modal logic, e.g. Def. 4.) P and F are not duals in this sense. In fact, they can both be seen as being like $\Diamond$, as oppopsed to $\Box$. Take P. If q stands for the proposition that it is raining, then Pq will stand for the proposition that it was, or has been raining – a proposition that is true provided q was true at *some* time in the past. Thus P is like $\Diamond$ in that the relation between Pq and q involves, like that between $\Diamond$q and q, an existential quantifier: Pq is true if there exists some time t in the past such that q was true at t just as $\Diamond$q is true if there exists an accessible world where q is true. F resembles $\Diamond$ in the same way that P does.

Connected with this difference is that, unlike $\Diamond$ and $\Box$ P and F cannot be defined in terms of each other. The "duality law" $\Diamond A \leftrightarrow \neg\Box\neg A$ entails that we could dispense with ◊ without actual loss in expressive power, since each occurrence of ◊ can be replaced by the combination ¬□¬, with as result a formula that is logically equivalent to the original one. Moreover, K (the weakest of our modal systems) also verifies the reverse duality law □A ↔ ¬◊¬A. This means that we could also eliminate □ while retaining ◊, using the combination ¬◊¬ instead of □. But in a system with P and F neither can be defined (or "simulated") with the help of the other. There is a close relationship between them – P and F can be seen as temporal "mirror images" of each other – but that does not enable us to define one from the other.

In fact, systems of tense logic are often equipped with duals to P and F (in the sense in which □ is the dual of ◊ in modal logic). There are even standard names for these dual operators, just as P and F have become the standard designations for the operators they represent. For the operator that is "dually" related to P we use "H", and for the operator that is dual to F we use "G". Thus systems of tense logic all verify the duality laws PA ↔ ¬H¬A, HA ↔ ¬P¬A, FA ↔ ¬G¬A and GA ↔ ¬F¬A. From these relationships it is easy to see that HA can be paraphrased as "It has always been the case that A" and GA as "It will always be the case that A".

The first tense logical systems we present have the four tense operators P, F, H and G. They are due to the father of Tense Logic, Arthur Prior (1914 -1969). Their syntax is virtually identical with that of the modal propositional logic of Def. 1. (Just replace ◊ and □ by P, F, H and G.) The full definition is given in Def. 11.

Def. 11 (Syntax of Standard Priorean Tense Logic)

1. Vocabulary:

   i. Propositional constants: q_1, q_2, q_3, ...
   ii. Connectives: ¬, &, v, →, ↔
   iii. Modal operators: P, F, H, G .
   iv. Parentheses: (,)

2. Formulas

   Form ::= q_i | ¬ Form | (Form & Form) | (Form v Form) |
   (Form → Form) | (Form ↔ Form) | P Form | F Form|
   H Form | G Form

We also adopt essentially the same model theory for this formalism. Models will once again be triples <W,R,F>, but the difference lies in the intuitive meaning of their components. The first set, W, is now not the set of possible worlds, but the set of instants of time. And R is now the "earlier-later" relation between temporal instants. Formally F does still the same thing as it does in the models for modal logic – it assigns a truth value to each constant q_i at each element of W; but of course, in view of the new interpretation of W the meaning of these assignments also changes.

Because our models now represent instants of time rather than possible worlds, we will use "T" instead of "W". Also we will use "<" to denote the earlier-later relation, and thus talk about models as triples <T,<,F>.

Def. 12 gives the truth definition for the system of Def. 11.

__Def. 12__ Assume that $\mathbb{M}$ = <T,<,F> is a model for Priorean Tense Logic and that t is an element of T.

(i) $[qi]_{\mathbb{M},t} = F_t(qi)$

(ii) $[\neg A]_{\mathbb{M},t} = 1$ iff $[A]_{\mathbb{M},t} = 0$

(iii) $[A \& B]_{\mathbb{M},t} = 1$ iff $[A]_{\mathbb{M},t} = 1$ and $[B]_{\mathbb{M},t} = 1$, and similarly for the other truthfunctional connectives

(iv) $[HA]_{\mathbb{M},t} = 1$ iff for all t' ε T such that t' < t, $[A]_{\mathbb{M},t'} = 1$

(v) $[PA]_{\mathbb{M},t} = 1$ iff for some t' ε T such that t' < t, $[A]_{\mathbb{M},t'} = 1$

(vi) $[GA]_{\mathbb{M},t} = 1$ iff for all t' ε T such that t < t', $[A]_{\mathbb{M},t'} = 1$

(vii) $[FA]_{\mathbb{M},t} = 1$ iff for some t' ε T such that t < t', $[A]_{\mathbb{M},t'} = 1$

Note that the clauses for H and P make a different use of the relation < than the clauses for G and F. The clauses for G and F directly correspond to those for □ and ◊ in modal logic. (E.g. the clause for G turns literally into the clause for □ of Def. 4 when we replace G by □, < by R and t and t' by w and w'. respectively.) We can also turn the clauses for H and P into exact copies of the clauses for □ and ◊ if we replace the formulas "t' < t" in them by the formulas "t > t' ", using now as counterpart of the modal accessibility relation R not the earlier-later relation <, but its converse, the "later-earlier" relation >. Trivial though this point may seem, it is important to draw attention to it. The real point is that when compared to the semantics of I.1.1 for the formalism of Def. 1 the present model theory makes use of two different accesibility relations, < and >, with two sets of modal operators ({G,F} and {H,P}) whose semantics is specified in terms of those respective relations.

This makes our system of Tense Logic into what is called a system of *multi-modal* logic. Systems of multi-modal logic are systems whose models involve two or more acccessibility relations and which have operators whose semantics is given with the help of these different relations. Often there exist systematic relations between the different accessibility relations, which can, and usually will, manifest themselves in the form of logical relations between the corresponding operators (i.e. in the validity of formulas in which operators corresponding to different accesssibility relations occur together). The tense logic of Def. 11 and Def. 12 is a particularly simple example of a multi-modal logic: Its semantics involves just two accessibility relations, < and >, each with a pair of operators corresponding to it, in the way that $\Box$ and $\Diamond$ correspond to the single accessibility relation R of Def. 7. Moreover, the relations < and > stand in a very simple relation to each other in that they are each other's converse. This relation is reflected by the validity in the schemata $A \rightarrow GPA$ and $A \rightarrow HFA$, among others.

The validity of these schemata brings us to the general question what formulas and arguments of the formalism of Def. 11 are valid. This is the same question that confronted us in connection with modal logic, but the situation is now quite different, for our conception of the structure of time are much more articulate than our ideas about the structure of the "space" of possible worlds.   At the very least, the earlier-later relation is asymmetric and transitive, and there are good reasons for thinking also that it is linear. (Linearity has been disputed, a point that will be briefly discussed in Section I.1.3.3below.) Usually, it is assumed that it is part of our conception of time that it has these properties, and thus that in all models $\mathbb{M} = <T,<,F>$ for the formalism of Def. 11 < must be a strict linear order of T.[16] Often this assumption -

16 We recall that a relation R on a set X is a *strict linear order of* X iff R has the following properties:

(i)	$(\forall x \in X)(\forall y \in X)(x < y \rightarrow \neg\ y < x)$	(asymmetry)
(ii)	$(\forall x \in X)(\forall y \in X)(\forall z \in X)(\ (x < y \ \&\ y < z \rightarrow x < z)$	(transitivity)

and the same goes for other such assumptions, in tense logic but also in other applications of modal logic – is stated in terms of *frames*. A *frame* in modal logic is simply that part of a model that consists of its “world” set and the accessibility relation (or relations) on that set. Frames <T,<> can be turned into models $\mathbb{M}$ = <T,<,F> by adding an interpretation function F to them. A single frame can be turned into different - and usually into many different - models by combining it with different F’s. We say that the model <T,<,F> is *based on* the frame <T,<>.

The structural properties of accessibility relations can obviously be described as the properties of frames. This is true for the case of tense logic as it is in modal logic generally. In the case of tense logic the frames are time structures $\mathbb{T}$ = <T,<>. The claim we made above that being lineraly ordered is part of the conception of time is thus the claim that the frames $\mathbb{T}$ = <T,<> of models for tense logic are linearly ordered structures, or linear orderings for short.

Def. 13 gives an axiomatisation of the sets of formulas and arguments of the formalism of Def. 11 that are valid on the assumption that the model frames for this formalism are all linear orderings and that any linear ordering can serve as frame.

<u>Def. 13</u> (Axiomatisation of validity for Priorean Tense Logic on the assumption that time is linear)

(i) A complete axiomatisation of classical propositional logic with M.P. as inference rule

(ii) The axiom schemata:[17]

(a) $H(A \rightarrow B) \rightarrow (HA \rightarrow HB)$
$G(A \rightarrow B) \rightarrow (GA \rightarrow GB)$

(b) $PA \leftrightarrow \neg H \neg A$
$FA \leftrightarrow \neg G \neg A$

(c) $A \rightarrow HFA$
$A \rightarrow GPA$

(iii) $(\forall x \in X)(\forall y \in X)(x < y \vee x = y \vee y < x)$ (linearity)

[17] As specified the axioms contain some redundancies (schemata that can be derived from the others. Thus one of the schemata in (c) and one of those in (d) can be omitted. We have given the axiom system in this redundant form because it brings out the symmetries between past and future more strongly.

(d) $HA \rightarrow HHA$
$GA \rightarrow GGA$

(e) $PFA \rightarrow (PA \vee A \vee FA)$
$FPA \rightarrow (PA \vee A \vee FA)$

(iii) The following two inference rules

$$\frac{\vdash A}{\vdash HA} \qquad \frac{\vdash A}{\vdash GA}$$

That this system gives us precisely the formulas and arguments that are valid when time is assumed to be a linear order was proved in the mid-sixties, not long after Prior had formulated the present system and conjectured the completeness of these axioms. But, as in the case of modal logic, to the *conceptual* question which tense-logical formulas and arguments should be considered valid this isn't a conclusive answer. That the earlier-later relation s a strict partial order – that it is asymmetric and transitive – is probably uncontroversial. But that time is linear is less obvious and is something that has been contested from the days of Aristotle to the present. (For more see Section I.1.3.3.) On the other hand, there are a number of further properties of which one could argue that time has them as a matter of conceptual necessity – properties of which it seems reasonable to ask: "Is this property part of our conception of time?".

Here is a list of the most salient of these properties:

(21)

(i) (DEN(sity))

$\langle T,<\rangle$ is a *dense* ordering iff
$(\forall t \in T)(\forall t' \in T)(t < t' \rightarrow (\exists t'' \in T)(t < t'' \ \&\ t'' < t'))$

(ii) (DIS(creteness))

$\langle T,<\rangle$ is a *discrete* ordering iff
(a) $(\forall t \in T)(\exists t' \in T)(t < t' \ \&\ \neg(\exists t'' \in T)(t < t'' \ \&\ t'' < t'))$ and
(b) $(\forall t \in T)(\exists t' \in T)(t' < t \ \&\ \neg(\exists t'' \in T)(t' < t'' \ \&\ t'' < t))$

(iii) (BEG(inning))

$\langle T,<\rangle$ is an ordering *with a first point* iff
$(\exists t \in T)(\forall t' \in T)(t' \neq t \rightarrow t < t')$

(iv) (END)

$\langle T,<\rangle$ is an ordering *with a last point* iff
$(\exists t \in T)(\forall t' \in T)(t' \neq t \rightarrow t' < t)$

(v) (NOBEG(inning))

$\langle T,<\rangle$ is an ordering *without a first point* iff
$(\forall t \in T)(\exists t' \in T)\ t' < t$

(v) (NOEND)

$\langle T,<\rangle$ is an ordering *without a last point* iff
$(\forall t \in T)(\exists t' \in T)\ t < t'$

(vi) (O(rder) C(omplete))

$\langle T,<\rangle$ is *order complete* iff
$(\forall X \subseteq T)(\forall Y \subseteq T)(X \neq \emptyset\ \&\ Y \neq \emptyset\ \&\ X < Y \rightarrow (\exists t \in T)(X \leq t\ \&\ t \leq Y))$

Here: "X < Y" is short for: $(\forall t \in X)(\forall t' \in Y) t < t'$,
"X ≤ t" is short for: $(\forall t' \in X)(\ t' < t \lor t' = t)$,
"t ≤ Y" is short for: $(\forall t' \in Y)(\ t < t' \lor t' = t)$

Each of these properties makes a difference to validity. Below we give for each of them one or two formula schemata that is/are characteristic for the propety in that it is/they are valid iff the property is assumed to hold for all models.

(22)

DEN: $PA \rightarrow PPA,\ FA \rightarrow FFA$

DIS: $((A\ \&\ HA\ \&\ F(A \lor \neg A)) \rightarrow FHA)\ \&$
$((A\ \&\ GA\ \&\ P(A \lor \neg A)) \rightarrow PGA)$

BEG: $\neg P(A \lor \neg A) \lor P\neg P(A \lor \neg A)$

END: $\neg F(A \vee \neg A) \vee F\neg F(A \vee \neg A)$

NOBEG: $P(A \vee \neg A)$

NOEND: $F(A \vee \neg A)$

OC: $(FA \,\&\, FG\neg A) \rightarrow F(G\neg A \,\&\, \neg PG\neg A) \,\&$
$(PA \,\&\, PH\neg A) \rightarrow P(H\neg A \,\&\, \neg FH\neg A)$

<u>Exercise</u>: Show that the schemata in (22) are characteristic for the properties of the relation < that are cited to their left.
More prescisely, show:

> If P is one of these properties and S is the schema (or one of the schemata) appearing to its right, then:
>
> (i) If $\mathbb{T} = \langle T,<\rangle$ is a linear frame which has P, then for every model $\mathbb{M} = \langle T,<,F\rangle$ based on $\mathbb{T}$, every $t \in T$, and every formula B which instantiates S, $[B]_{\mathbb{M},t} = 1$;
>
> (ii) If $\mathbb{T} = \langle T,<\rangle$ is a linear frame which does not have P, then there exists a model $\mathbb{M} = \langle T,<,F\rangle$ based on $\mathbb{T}$, some $t \in T$, and some formula B instantiating S such that $[B]_{\mathbb{M},t} = 0$.

N.B. In all these cases adding the characteristic formula (or one of the characteristic formulas) for one of the properties listed under (21) to the axiom system of Def. 13 results in an axiom system that is complete for the class of models whose frames have that property (or, to be pedantic, whose frames are linear orderings with that property). This is also true for any conjunction of properties in (21) and their characteristic formulas.[18]

---

[18] These facts, whch again will not be proved here, have been known since the late sixties. Proofs can be found in various publications. See in particular the contribution of J. Burgess to Vol II of the (old) Handbook of Philosophical Logic (ed. D. Gabbay & F. Guenthner; Reidel, 1984) and Vol. VII of the revised edition of the Handbook of Philosophical Logic (ed. D. Gabbay, Kluwer, 2002), as well as: D. Gabbay, I. Hodkinson & M. Reynolds (eds.) Temporal Logic, Vol. I (Oxford, The Clarendon Press, 1994).

What properties does time really have?

What can we say about the properties listed in (21)? Can our intuitions about time tell us something about these properties?

Time is a prime concern to pretty much all of us. It is almost perpetually present to us, whether we are planning for what lays ahead or reflecting on what has been. Moreover, it has been a central concept in branches of science as far apart as physics and psychology. Those different sciences have developed very different ways of thinking about time, and these have led to quite different, and sometimes contradictory, assumptions about properties like those in (21). The most dramatic example of this are the different views that have been expressed about the conflicting properties of density and discreteness. Psychology has stressed the essentially discrete nature of time as it manifests itself in our experiences, whereas in physics the dominating view has been, at least since the days of Galileo, that time is a continuous, and therefore densely ordered, medium.

Within philosophy the conception of time as discretely ordered is usually associated with Kant (1724-1804). For Kant the concept of time aqnd the concept of counting (and with that our understanding of arithmetic) were intimately connected. Our ability to count, adding one unit after another indefinitely, and our ability to experience time as the passing of successive events, were, as he saw it, manifestations of the one and the same fundamental feature of our mind. This idea - that time is given to us as the "ticking" of a succession of separate events – has often been seen as entailing that the order of time is discrete. it is a conception of time that still has its advocates today, especially among philosophers with idealist or phenominological leanings.

Kant's observations of time touch upon what seems to be an essential feature of the way we experience time: it is only by registering successive events *as* successive, in the sense that the one is already past when the other starts, that time exists in our awareness. But it is not clear why this should lead to the conclusion that the structure of time itself is discrete. After all, nobody would assume that the events of which he becomes aware and that make him aware of the passage of time, are the only events there are; what is noticed by anyone of us is only a minute fraction of all that happens around us, let alone of all that happens in all those places to which we have no direct access. In the gaps between successive events that imprint themselves upon me there may be countless others that escape my notice.

It does not follow from this consideration that time must be dense rather than discrete. But what it does imply is that, even if discreteness is an essential feature of our own experience of time, that doesn't entail discreteness of time as the temporal order of everything that happens.

It is true of course that modern physics tells us that we live in a finite universe. And if that is so, then presumably there are between any two events only finitely many others, even if their number may be "atronomical". If that is so, then, strictly speaking, time might be discrete after all. But of course that conclusion is valid only if we assume that time depends on what happens – that it is the product of the succession of events. This is an assumption we have been making implicitly in the last few paragraphs. It seems a quite natural assumption, and it is often made, in various forms and for various reasons (e.g. Leibniz in the Correspondence with Clarke), (Russell, in his Lectures on logical Atomism). But it has not gone unchallenged. One of its main critics in the history of western thought was Isaac Newton, who saw time and space as forming a fixed, immutable structure in which the events of our world unfold. On this conception of time there can be time without change. And it might well be that among the accidental properties of the actual physical world that is ours there are periods of time when, as it happens, nothing happens.

How can we know anything about the structure of a time that is so loosely connected with the events we can observe? If our conception of time is this abstract, then arguments that certain properties are intrinsic properties of time will have to be far more indirect than they can be when time is seen as generated by real or mental events. There is no direct relying here on our own experience or understanding of time, nor on our observations of physical events. Rather, the only thing we can rely on are reflections on what is needed in order to describe the laws that govern the physical world and the preconditions for stating and applying those laws. It is the calculus, the theory and practice of integration and differentiation, that shows us what structure time and space must have in order that physical objects can move in and through them according to the laws of Newtonian Mechanics, which require the calculus for their formulation.

It was only in the nineteenth century, through the work of mathematicians like Cantor, Dedekind, Weierstrass and others, that the structural implications of this calculus-based perspective became fully explicit. In particular, it was only then that the difference between (merely) dense orders and continuous orders – orders that are both dense and order-complete –was clearly articulated. (It was only then that the difference between the rational numbers, which are densely but not continuously ordered, and the real numbers, whose order is continuous as well as dense, was made fully transparent and explicit).[19] With the hindsight of those later developments we can now say with some confidence that the Newtonian conception of time was that of a linear order that is isomorphic to the reals. For a formalism like the tense logic of this section that would mean validity of the both the two

---

[19] One of the most remarkable features of the real numbers is the relation it embodies between points and intervals (or, when we think of the real numbers as time, between instants and periods): That it takes a non-denumerably infinity of points to get an interval of any non-zero length is something that boggles the mind of anyone who is confronted with this property of the reals for the first time. That the reals have this property has to do with the fact that they come not only with an order but also with a metric. (When we think of the rela numbers as instants of time, then the metric manifests itself in the lengths – or "temporal durations" – of the interevals by which two intervals are separated.) It is because of the way in which order and metric are connected that the assumption that time (as well as space) has the structure of the reals provides an solution to the paradoxes of Zeno. Recall Zeno's best-known paradox, that of Achilles and the tortoise. Achilles, a notoriously fast runner is to run against the tortoise, who is generally known as slow and plodding. To give the tortoise a fair chance Achilles generously offers it a head start (of, say, a couple of hundred yards, though as the story will show, the amount is immaterial). The race begins and Achilles has soon covered the distance between the point where he started and the strarting point of the tortoise. But when he gets to the point where the tortoise started the tortoise has moved a little too, even if it isn't very much. So, in order to overtake her, Achilles will have to also cover this bit that the tortoise has just covered. But by the time he gets to the point where the tortoise was when he reached the tortoise's starting point, the tortoise will have moved again. So Achilles will have to cover that bit as well; and so on.In thi way we can argue that before Achilles can overtake the tortoise, he will have to do, one after another, an infinite number of things, and those will involve an infinite number of times. Thus, Zeno concludes, Achilles would need *an infimite amount of time* to overtake the tortoise. Zeno's final conclusion is that, contrary to what Achilles himself and those watching the race might have expected from the start, and thought they saw as they were watching, he and we might have thought, Achilles never did vertake the tortoise.

I have put the phrase which is repsonsible for this story getting off the rails in Italics. An infinite amount of time is not the same thing as an infinite number of times. That there is a difference, and what that diference is, is the profound philosophical achievement of the theory of the real number continuum, with the subtle analysis it makes possible of the concept of a limit.

In the light of this success it might seem curious hat even today Zeno's paradoxes are often cited as a neat dmonstration that there is more to the concept of time than mathematics and physics can explain. The common denominator to the thoughts of those who are involved in this seems to be the intuition that fundamental to our experience of time - and, therefore, it is implied, also to our conception of time - is an element of discrete, step-wise progression, which the identification of time with the real number continuum fails to capture. For some Kant's conception of time is still very much alive.

schemata that are caharatreristic for DEN and the one that is characteristic for OC.

It might be felt that there is something incongruous about appealing to a system as complex and powerful as the calculus and its application to the theory of physical motion in order to determine the logic of a formalism that is as simple as the propositional tense logic of this section. There seems to be an immense gap between the expressive power of the language of mathematical analysis and that of the language of propositional tense logic that has been the subject of this section. In fact, one might be incined to ask, are there really any useful applications for such a limited system at all? For us this question is important primarily in relation to the semantics of natural language. That is of course a very diffferent enterprise from the integral and differential calculus. But almost as soon as the H,G,P,F-system was given its definitive formulation, one realised that it falls well short of what is needed to represent the temporal information found in sentences of natural language. Presently this led to more expressive tense logics, in which more temporal relations between propositions can be represented than in the system of Def. 11. We will look at one such system in Section I.1.3.2.

I.1.3.1 Tense Predicate Logic.

Just as we can define a system for modal predicate logic by adding $\Box$ and $\Diamond$ to a sytem of classical predicate logic (see Def. 6), so we can define a system of tense predicate logic by adding to classical predicate logic the operators P, F, H and G. Since the syntax of the system that results from this substitution is obvious, we will spare ourselves the explicit definition of it and refer to the reader to Def. 6. He will be able to make the necessary substitutions himself.

Much the same goes for the models for tense predicate logic. Here too it is natural to follow the semantcs for modal predicate logic formulated in I.1.2. For our system of tense predicate logic this means that models are now quadruples $<T,<,U,F>$, where $<T,<>$ is a temporal frame, U is a non-empty set of “possibilia” and F assigns for each $t \in T$ suitable interpretations, relative to U, to the system’s non-logical constants.

After what has been said about the semantics of modal predicate logic there is not much that we need to add about this similarly structured semantics for Tense Predcate Logic. But there is one new consideration that deserves comment. The set U is now to be understood as a set which includes all individuals that exist at at least one $t \in T$; but once again we can rely on the predicate E to select from these possibilia for each time t the set E_t of entities existing at t. This enables us to express quantification at t over just the entities existing at t, just as E enables us to express the effect of quantifying at a world w just over the entities existing in w.

But there is a difference between the behaviour of entities through time and their behaviour "across" possible worlds. It has to do with what might be termed the "reality of history". Historiography, the discipline which studies history, deals with things past and past events. And many of the things it deals with are truly past in that they exist no longer. It is a fundamental, unquestioned assumption that historians share with pretty much everybody else that even if most the facts about the past are lost, they are nevertheless *facts*. The past must have been one way or the other; and it is among the basic tasks of historiography to establish as well one can, which way it actually was. Often the answer to the question: "Did such and such a thing happen?" or to the question: "Did a thing with such and such properties exist?" may be irretrievably lost, but even in such cases we are firmly convinced that there is a true answer, and see it as our loss that it is beyond us to find out what that answer is.

There are two ways in which we talk about past things in which this conviction – that the past is fully definite, although only partially known to us – shows up. First, there are cases where an entity has disappeared but has left us its name. In such cases we feel no compunction about using the name, and we happily in statements that explicitly or implicitly assert that the name's bearer is no more: - statements like: "Carthage once existed, but exists no longer.", or "Caesar was killed on the ides of March of 44 B.C." We are convinced that by using these names we refer to the bearers of those names – a former city and a citizen of ancient Rome – and that what we say about them are literal, facual truths.

But it is not just in the way we talk about things that we still know by their names, or even about things of whose past existence we have clear proof such as dynosaurs of which we have discovered the fossile remains) in which we manifest our conviction that our past is definite and unique. We assume unhesitatingly that been myriads of things that existed in the past which have since disappeared without a trace. And with that assumption goes the conviction that questions like: "Did such and such an entity exist once?" or "Did such and such an entity exist at such and such a time?” have a definite answer one way or the other, although in most cases the answer is irretrievably lost.

Our attitude towards future existents seems quite different. First, how could we make claims to the effect that some particular entity will exist, but does not exist yet? In referring to the entity by name? But where would the name come from, or how would it get its reference? If the entity that we want to speak of exists only in the future, how could we have attached the name to it already, so that we can now use the name to refer to the entity? Sure, there are some cases when something like this is possible. The capital of Brazil was known as “Brasilia” before it was built, even before all the plans were drawn up or its exact location fixed. But such cases, where a future entity has a name before it comes into existence for real, are comparatively rare, and they confirm the general principle: a future entity can have a name now only if there is a well-defined plan for it (or, aternatively, a clear conception of it, combined with an expectation that an entity fitting that conception will come about).

It is not just the wider availability of names that distinguishes past entities from future ones. The real difference goes deeper. We are convinced that the past was just the way it was. But our conception of the future is different. We see the future as something that could turn out this way or that; if it will unfold one way, then entities with such and such properties will come into being, but if it unfolds the other way, then that won’t be so. From this perspective, many future entities – presumably the vast majority – are merely possible, and there is thus an important difference in status between them and the entities of our past. This difference is sometimes described as that between entities that *subsist* – the past entities (as well as, perhaps, a small number of future entities that are already “in the making”) - and the merely possible ones. It is convenient to think of subsistence as a kind of “weak” existence, so that true existents also count as subsisting, but are distinguished from the merely subsisting by being full-fledged existents on top of that. This leads us, for each time t in the course of history, to a pair of two-way distinctions:

(i) the difference between (a) subsisting entities - those presently existing at t, those which existed at some time preceding t as well as, perhaps, a small number of future entities in the planning or making stage at t – and (b) the merely possible entities that may become existent at some time in the future of t; and

(ii) the difference between (a) the entities that merely subsist at t – the past entities and those in the planning or making stage - and (b) those that currently exist at t.

To capture these distinctions between the merely possible, the merely subsistent and the genuinely existent we adopt besides the predicate E a second 1-place predicate S (for "subsistence") which singles out the subsistent entities from the totality of possibilia.

Given the way we have defined subsistence, the store of subsisting entities steadily grows as time goes on: New entities come into existence, while others go out of existence. But those which go out of existence still remain as subsistent entities, while those that come into being reinforce the ranks of the existent, and therewith of the subsistent entities. For the models $\mathbb{M}$ of our system of tense predicate logic this means that in each model the extension of S can only grow as time progresses: $F_{\mathbb{M},t}(S) \subseteq F_{\mathbb{M},t'}(S)$, if $t < t'$. This doesn't fully determine what we should say about the universes of the models M_t that a model should provide. But now that we have introduced S as well as E, we are free to resort to the same convenient solution that we also adopted for the case of modal predicate logic in section I.1.2: We assume a single fixed universe of possible entities and let S and E select from U at each time t the set of entities subsisting at t and the set of things existing at t.[20]

These considerations and choices lead to the following notion of a model for tense predicate logic:

Def. 14 A *model for tense predicate logic* is a 4-tuple $\langle T,<,U,F\rangle$, such that

(i) $\langle T,<\rangle$ is a linear frame
(ii) U is a non-empty set
(iii) F is a function which assigns for each $t \in T$ a suitable interpretation at t, relative to U, to each non-logical constant
(iv) for all $t \in T$, $F_t(E) \neq \emptyset$
(v) for all $t \in T$, $\bigcup_{t' \leq t} F_{t'}(E) \subseteq F_t(S)$

[20] We will have more to say about future existents in Section I.1.3.4 below.

As in the case of modal logic we do not go into the question which formulas and arguments are valid according to the model theory whose models are those given in Def. 14. As in the case of propositional tense logic, the question depends in large part on the strucutural properties that we are prepared to ascribe to time. But in addition there are also formulas and arguments whose validity depends on the interaction between time and quantification. This matter isn't particularly complicated, but given our ultimate aims there would be little point in pursuing it here.[21]

I.1.3.2 More Powerful Tense Logics.

We concluded Section I.1.3 with the remark that the need for more powerful tense logics was felt not long after the P,F,H,G-system had been given its definitive formulation. One way to appreciate the expressive limitations of that system is to compare the following examples.

(22) (i) After Fred visited us, he sent us a postcard.
(ii) Since Fred visited us, he hasn't given a sign of life.

Abbreviating "Fred visits us" as q and "Fred sends us a postcard" as r, we can symbolise (22.i) in the P,F,H,G-system as (23).

(32) P(r & Pq)

---

21 We note in passing that our semantics renders the following formula valid.

(i) $(\forall x)((E(x) \vee PE(x)) \rightarrow S(x))$

An alternative to Def. 14 would be to restrict the entities subsisting at t to those exisiting at or before t: $F_t(S) = \bigcup_{t' \leq t} F_{t'}(E)$

This alternative semantics verifies not only (i) but also (ii) (which logically entails (i)) and (iii). (ii) just says that to subsist is to exist now or else to have existed at some previous time. (iii) says that once something subsists it will continue to subsist forever.

(ii) $(\forall x)(S(x) \leftrightarrow (E(x) \vee PE(x)))$
(iii) $(\forall x)(S(x) \rightarrow GS(x))$

Exercise: Show that (i) is valid on the semantics described in Def. 14 and that (ii) and (iii) are valid on the modified definition just described in this footnote.

But it is not clear how we could symbolise (22.ii). You will find that out when you try to find a symbolisation yourself. You won't find the right formula no matter how hard you try. And that won't be because you aren't clever enough; it can be proved rigourously that no formula of the P,F,H,G-system captures the truth conditions of this sentence. The problem is that what (22.ii) says is that there is a time in the past when Fred visited us and that between that time and now there was *no* time at which he gave a sign of life (or, put slightly differently, at all times between that time and now the proposition "Fred gives a sign of life" is false). This is a temporal relation between two propositions – the proposiiton that Fred visits us and the proposition that he is giving a sign of life – which the tense operators P and F are unable to express. In fact, something much stronger holds: no combination of 1-place tense operators is able to express this relation. Adding further 1-place operators to the formalism of Def. 11 is therefore not going to help with this problem; what we need is a tense operator of more than one place.

The simplest solution to the expressibility problem that (22.ii) poses is to add a new 2-place operator whose semantics is given by the very propositional relation that (22.ii) expresses. The operator we introduce to this end is suitably referred to as "Since" and represented by the letter "S". Syntactically S behaves like a 2-place sentence connective; in other words, its syntax is the same as thst of the 2-place truth functional connectives &, v,→ and ↔. But its semantics is more like that of the tense operators P and F. The clause that defines its truth conditions is given in (24.i). It is easily verified from this clause that if we abbreviate "Fred visits us" as q and "Fred gives a sign of life" as s, then (22.ii) can be correctly represented by the formula S(q,¬s).

(24) (i) $[S(A,B)]_{\mathbb{M},t} = 1$ iff for some $t' \varepsilon T$ such that $t' < t$, $[A]_{\mathbb{M},t'} = 1$ and for all $t'' \varepsilon T$ such that $t' < t'' < t$, $[B]_{\mathbb{M},t'} = 1$

(ii) $[U(A,B)]_{\mathbb{M},t} = 1$ iff for some $t' \varepsilon T$ such that $t < t'$, $[A]_{\mathbb{M},t'} = 1$ and for all $t'' \varepsilon T$ such that $t < t'' < t'$, $[B]_{\mathbb{M},t'} = 1$

Just as P has a future counterpart in the operator F, S has a future counterpart in the 2-place operator U ("Until") whose semantics is given in (24.ii).

The system of propositional tense logic whose tense operators are S and U proves to be very powerful. In fact, as propositional tense logics go, it is almost as powerful as such a system could be. And if we assume that time is order-complete (e.g. that it is like the integers or the reals), then it is *literally* as powerful as any such system could be. For a start we can express the operators P and F with the help of S and U. For PA is logically equivalent to S(A, A∨¬A) and FA to U(A, A∨¬A). But that is just the beginning. When time is order-complete, any operator that you can think of and whose truth conditions can be defined along the lines of the truth condition clauses for P, F, S and U can be expressed by means of S and U (in the sense in which we have just expressed P and F by means of S and U).[22]

We can add S and U, with the truth clauses for them that are given in (24) either to the tense propositional logic of Section I.1.3 or to the tense predicate logic of Section I.1.1.2. In either case we get a system that is (on the assumption that time is order-complete) *functionally complete* in the sense that all tense operators that can be defined in terms of temporal order can be expressed in it. At the same time the propositonal system with S and U shares some of the attractive properties of weaker systems like the system with P and F. For instance, the sets of valid formulas and arguments for the different notions of time considered in Section I.1.1.3 are axiomatisable and also decidable. This is one of the reasons why the S,U system has become an important tool in computer science, where it is used to describe certain properties of program executions; in these applications the "instants" of time are identified with the periods between two successive "ticks" of the internal clock of the computer (which is assumed to have the traditional Von Neumann architecture) on which the program is being

[22] The S,U-system arose out of a puzzle that Prior presented in a seminar he gave at UCLA in the fall of 1965. As it stands the P,F,H,G-system is strictly "topological": the operators P, F, H and G are sensitive to the order of time, but not to its metric. For instance, PA says that A was true at some time in the past, but it says nothing about how far in the past it was true. Can we, Prior asked, use the P,F system nevertheless as the basis for introducing some kind of temporal metric? A possible first step in that direction might be this: assume that there are particular "clocktime" propositions, such as "It is midnight", or "The sun is just appearing above the horizon", which are true at regular intervals; or better: which are true at certain instants of time, which we see as separated by intervals of (roughly) the same duration during which they are not true. (It is not an unreasonable assumption that this idea - that propositions or events can be temporally located with respect to other propositions or events that we come to see as demarcating equal portions of time - is at the root of our conception of time as having a metric as well as an order.) Let p be such a clock time proposition and suppose I want to say that q was the case since the last time it was the case that p. How do we express that in the P,F.H,G-system? It turns out that this can't be done, and the proof that this is so was the main impetus towards the design of the S,U-system, in which such statements can be expressed. (Exercise: How can we express the given proposition in the S,U-system?) The system, together with its functional completeness (on the assumption that time is order-complete) was in place by the end of 1966. See J.A.W. Kamp, *Tense Logic and the theory of Linear Order*, Ph.D Dissertation, UCLA, 1968

run; and the propositions are properties of the internal states that the computer is in at such “instants” (i.e. between two successive ticks).

In spite of the remarkable expressive power of the S,U-system, it has proved to be not what we want for dealing with the semantics of time in natural language. That is a matter which, as I noted before, will be taken up at length in Part II. For now let me mention just one reason why the system isn’t really suitable as a tool for dealing with the temporal aspects of meaning in natural language. While the expressive power of the S,U-system leaves nothing to be desired as such, the *way* in which it represents temporal relations is often very different from the way these are expressed in natural language. So formulas from the S,U-system that correctly capture the truth conditions of certain natural language sentences do so at the price of throwing most of the structure of the symbolised sentences overboard and replacinf it by the baroque and seemingly unrelated structure of representing S,U-formula. The S,U-system makes for very odd logical forms. General principles that translate sentences of natural language into such “logical forms” are very difficult to state; and to the extent that they can be stated, they do not seem to give us any insight into how temporal reference in natural languages really works.

As an example of this consider the sentence in (25.i). Using the abbreviations given in (25.iii) this sentence can be represented with S and U as in (25.ii). But note how far the form of (25.ii) has strayed from that of (25.i). From the perspective of (25.i), (25.ii) seems quite convoluted. And its convolutedness isn’t just an accidental feature of this particular attempt to get the truth conditions of (25.i) right. To get a sense that this is an intrinsic problem with the S,U-system, just try if you can find a more congenial, less convoluted formalisation of (25.i).

(25) (i) Since the last time Fred visited us he has called us just once.

(ii) $S(p \,\&\, \neg q \,\&\, S(q,\neg q), \neg p \,\&\, \neg q)$

(iii) Abbreviations:

q:= Fred visits us;
p:= Fred calls us

Exercise: Give symbolisations in the S,U-system of the sentences and sentence sequences below, using the given abbreviation schemes.

Remark 1. In natural language temporal relations between propositions are often expressed "across sentence boundaries"; i.e. one of the related propositions is found in one sentence and the other proposition in the next sentence. The temporal relation is then often conveyed by the choice of tenses in the two sentences, sometimes in combination with adverbs like *then*, *afterwards* and the like. Since this way of expressing temporal relations in natural language is so extremely common, some examples involving cross-sentential temporal relations have been added here. A special reason for adding such examples is that they give us a first taste of what we will be the central focus in Part II, where we will be primarily concerned with the meaning of multi-sentence discourse.

Note well: within the S,U-system the cross-sentential temporal relations can be captured only by choosing a *single* formula of that system to represent the content of the *entire sentence sequence* of natural language that is to be represented

Remark 2. What you are asked to do in this exercise is very old-fashioned: You are asked to make use of your understanding of English to grasp the truth conditions of the given English sentences and discourse and then to find a formula of the S,U-system with the same truth conditions. To ascertain that the formula you come up with has the truth conditions of the sentence or discourse bit for whch it is intended as symbolisation, you have t rely on the one hand on the intuitive undestanding of English already mentioned and on the other on the truth definition for the S,U-system, which assigns each formula of the system (and thus in particular the formula you have chosen) definite truth conditions (in that it specifies for each model $\mathbb{M}$ and each $t \in T_{\mathbb{M}}$ the truth value of the formula in $\mathbb{M}$ at t).

This is the traditional way in which formal logic has been used to clarify certain aspects of meaning in natural language. It is still practiced in most introductory texts and classes to formal logic for philosophers and other non-mathematicians. It might be called the "pre-Montague" way of using logic in the study of natural language meaning. This is not how we want to do semantics here. (It is a way of doing semantics that is not very useful for Computational Linguistics, for you can't expect a program or computer to have the intuitive understanding of English(or whatever other natural language that is being studied or to which the given semantic theory is to be applied). The point of a theory of natural language semantics useful to CL is precisely that it should help us to *teach* the program or computer the kind of semantic competence that the present exercise presupposes: the competence to compute the interpretations of sentences or sentence sequences from their syntactic structure and the known meanings of the words (often in combination with further contextual information). We will leave the "prehistorical" method practiced in this exercise behind us when we get to Montague Grammar in the second part of Part I.

(1) He has been here since he arrived and will be here until he leaves.

Scheme of abbreviations:

q:= he is here
r:= he arrives
s:= he leaves

(2) Since Mary's friend has been staying until breakfast, when she takes him home with her, her father has been morose.

Scheme of abbreviations:

q:= Mary takes her friend home with her
r:= Mary's friend stays
s:= breakfast is happening
t:= Mary's father is morose

(3) First Fred arrived. Then Charles arrived. And since then nobody else arrived.

Scheme of abbreviations:

q:= Fred arrives
r:= Charles arrives
s:= someone arrives

(4) Since Charles and Fred have met, Fred always brings Fred along. And that will continue to be so until they split.

Scheme of abbreviations:

q:= Charles sand Fred meet
r:= Fred brings Charles along
s:= Charles and Fred split

(5) Mary will come, when Charles has his brithday. We will meet at least twice before she comes again.

.

Scheme of abbreviations:

q:= Maria comes
r:= Charles has his birthday
s:= we meet

I.1.3.4 The Open Future: Branching Time or Branching Worlds?

When discussing the semantics for Tense Predicate Logic I mentioned the deep sense we have that past and future differ from each other in that the future is “open”, while the past is “closed”: when we go back in time there is only one immutable past that we could encounter; but if we move forward, we could find ourselves in any one of a great many different futures, all of which are still possible given what thngs have been like and are like now. In Section I.1.3.2 this issue was raised in connection with the difference in status between what exists no longer and what exists not yet. This is a difference that could also be put in a somwhat different way: statements about future existents have a different status from statements about past existents. The typical statement that there will be things or people with certain properties –a building more than half a mile high or a clone of George W. Bush – have, most of us feel, an element of indeterminism about them; they might be false, but they might also be true; which will be the case depends on how things will shape up. Not so for statements about the past existence of people or things; such statements are simply true or false. For there is only one past, and it is about that past that the statement makes a claim. We may not be able to *tell* whether the statement is true or false, because there isn’t enough of the past left for us to see. But that doesn’t alter our conviction that burried in that past there is a definite answer to the question: true or false? - even if the answer is irrecoverably lost to us.

The idea that many statements about the future are neither definitely true nor definitely false goes back to Aristotle. In Aristotle’s *De Interpretatione* there is a discussion of the statement in (26):

(26) There will be a sea battle tomorrow.

One possibility, Aristotle notes, is that at the point when (26) is uttered, the question whether there will be a sea battle has been decided one way or the other. Either the wheels which lead inexorably to the battle taking have already been set in motion; or steps have already been taken by one or both sides of the conflic that have made the occurrence of a battle tomorrow already impossible (e.g. the ships have been pulled back so far that the two sides couldn't join in battle tomorrow, even if they)changed their minds). Under such conditions (26) will have a definite truth value at the time when it is uttered – it is true in the first case and false in the second. But if no things have yet taken place that force or prevent the occurrence of tomorrow's battle, then the statement is neither true nor false.

The intuition that Aristotle expresses in this discussion has a good deal of initial plausibility. But various problems arise when we try to make that intuition more precise, and especially when we try to determine what its consequences are for the concept of truth, and for logic in general. One of these problems has to do with the notion of "definite truth". In the last paragraph I once used the words "definitely true/false" and shortly after that I spoke simply of "true" and "false". How, you may have wondered, is that to be understood? Are these just stylistic variants, different ways of referring to the same concept? Or are the phrases to be understood as referring to distinct concepts, with "definite truth" a special variety of truth, like "necessary truth"?

Both these possibilities – (i) there is just one concept of truth and alsity and that is definite truth/falsity and (ii) there are two notions, "plain" truth and falsity and definite truth and falsity - have been explored and formalised. In fact, there is hardly any way one could think of dealing with the Sea Battle problem that hasn't been investigated and, if coherent at all, also formalised, something that isn't really surprising, given that the solution we adopt is to have consequences for a host of central philosophical concepts and concerns, from truth and logical validity themselves to determinism, responsibility and freedom of the will.

This is not the place to engage in a comprehensive review of the various solutions that have been proposed. In particular, we will pass by those solutions that come closest to what seems to have been Aristotle's own position, viz. that most statements about the future are without a truth value at the time when they are made. A direct formalisation of this idea involves either admitting "truth value gaps" – a truth value gap arises whenever a statement lacks a truth value, i.e. is neither true nor false – or, what comes to much the same thing, though not quite - admitting one or more additional truth values besides the two standard values “true” and “false”. Multi-valued logics (logics with more than two truth values) and partial logics (logics admitting truth value gaps) present many problems of their own, and to address those would involve a good part of this course. Since these questions are not germane to the central goals of this course, we let this matter rest. We will stick throughout to the classical paradigm in which every well-formed statement always has one of the two truth values “true” and “false”.

But even this restriction leaves us with a number of options for capturing something of Aristotle’s distinction between future-directed statements that are (definitely) true and those that are not. The first is to stick to the principle that only definite truth is truth and that everything that isn’t true in this sense is thereby false. For Aristotle’s sea battle example this means that so long as nothing has been done to rule out one of the possible outcomes, both (26) and its apparent negation (27) will be false:

(27) There won’t be a sea battle tomorrow.

A second option is to distinguish between “mere” truth and “definite” truth (also called “determinate truth). Once the things have happened that make tomorrow’s battle either inevitable or impossible, (26) acquires the status of definite/determinate truth or definite/determinate falsity. But that doesn’t mean that it didn’t have a truth value until then. Eventually things will turn out one way or another, whether they were bound to come about the way they eventually do already at the time when the statement was made, or came about that way because of what happened later, or by mere chance. By the end of tomorrow there either will have been a sea battle or there won’t have been. By that time the truth value of the statement which I make when I utter (26) today will have been decided; and if we just wait until tomorrow night, then we will also know what that truth value is. Moreover, since there is only one way that things will turn out in the future that, by chance or necessity, will become our actual future, what grounds could there be for denying that that truth value is already the truth value of my assertion of (26) now, at the time when I am making it?[23] The statement won’t be deteminately true in the sense that nothing

[23] This is, it seems to me, the most natural way to understand claims that we make about the future as part of placing bets. I say to you “Liverpool will beat Chelsea tomorrow. I bet you 10 £.” In such a situation I am simply making a statement about how things will turn out. It is that, and nothing more or less, that I am putting my

could still happen to make it come out false. But it will be true in virtue of what will actually happen, in virtue of our actual future, it still qualifies as “plainly” (or non-determinately) true.

The distinction between plain and detereminate truth/falsity presupposes that among all the possible contiunations of the world as it is now there is one that will reveal itself as the actual continuation. This question – whether there is among all the possible future continuations of the world as we have it now one that is singled out as the future that will actually be ours – constitutes another important division between formalisations of the open future. And it is bound up with another, somewhat more technical issue. We can think of the open future as a branching tree: the world as it is right now is a node of this tree with as many branches sprouting from that node as there are possible ways for the world to go on from now; likewise the later points along any of these possible continuations are branching points too; the same also goes for stages in the unfolding of our world that are already behind us.

But how is this branching structure to be represented? One option is to represent it as the branching of time itself: Time is not a linear order but a certain kind of partial order, viz. a “tree” in the mathematical sense of the term. In such a tree-like time structure $<T,<>$ the elements t of T are to be thought of as “worlds as existing at a given time”, and for any such t the $t' \in T$ such that $t < t'$ are the ways in which it is possible for the “world stage” t to continue. When the open future is modelled in this way it is still possible in principle to assume that one branch of the tree $<T,<>$ (i.e. one maximal linearly ordered subset of T) represents the actual world as it develops through time. If such a branch B is singled out, then it is possible to define at least for the points lying on this branch B what it is for a statement of the form FA to be “actually true at any of its points: FA is true at $t \in B$ iff there is a $t' \in B$ such that $t < t'$ and A is true at t’. But as a matter of fact, open futures modelled as tree-like time structures $<T,<>$ are normally not enriched by identifying some branch of $<T,<>$ with the actual world, so that there is no way of captuzring the notion of actual truth as distinct from definite or inevitable truth on the one hand and merely possible truth on the other.

One disadvantage of the account indicated in the last paragraph is that it conflates the notions of times and worlds in a way that becomes an

money on. doing that I stick my neck out, insofar as that which I am affirming and putting my money on.

impediment especially in applications to semantics. This difficulty is absent from the following way in which the open future can be modelled, and this is the one we will look at a little more closely. This way is based on the idea that each possible world extends through time from its earliest past to its most distant future. (Roughly eac h such worlds corresponds to a model $\mathbb{M} = <T,<,F>$, with $<T,<>$ a linear ordering, that we have been defined as part of the semantics for systems of propositional tense logic.) However, two such worlds may remain indistinguishable from each other up to some point in time, at which they unfold in different directions. Suppose that a bundle W of such worlds is given. Let w be one of them and let t be an instant from the time structure of T. Consider all the worlds $w' \in W$ which coincide with w up to t. These worlds represent, together with w itself, all the possible ways in which w might continue after t, given how w is at t and has been before t.

Formalising actual truth in this setting is straightforward: That only has to do with what happens at various times in the world in which the given statement is being evaluated. In particular, the truth of FA in w at t depends on what is the case at times later than t in w. But we do not just want to be able to account for actual truth of future tense statements, but also for what it means for such statements to be inevitably true. Given the present set-up it is quite natural to think of inevitability as a form of modality, which can be formalised with the help of a modal operator: Just as we have analysed A's being necessarily true as the truth of $\Box$A, so it seems reasonable here to analyse A's being inevitably true as the truth of DA, where D is the operator representing "inevitably". (The letter "D" which is used to represent this operator is derived from "determinately", where "determinately" is to be understood in the sense of in which we have been using it, viz. that of "determinately in view of the present and past".) A formula DA is then true in w at t iff A is true at t in all worlds w' which coincide with w up to t. So the determinate truth of a proposition of the form "it will be the case that q" will be represented as DFq, a formula that is true in w at t iff for each w' that coincides with w up to t there is a $t' > t$ such that q is true in w' at t'.

In this way we are led to a multi-modal system in which the tense operators H, G, P and F are combined with the modal operator D. This system is not only important because it provides a logical basis for dealing with the central philosophical concepts that were mentioned earlier (concepts like determinism, knowledge, free will, moral responsibility), but also because the open future plays an essential part in the analysis of meaning in natural language. The main reason for this is that so much of what people talk about are their plans and actions. Our understanding of what is involved in the design and execution of plans presupposes a conception of ourselves as agents, who by their actions can shape the future to greater or lesser extent, ruling out certain possible continuations and thus making their contribution towards the future that will actually emerge.

The system we present here provides no more than a quite basic formal skeleton for formalisations of the many important notions that require as foundation. But even though it is quite simple, it has a number of interesting properties as it stands. Moreover, at the present stage of our explorations in these notes the system is also of interest for the new twists it provides to some logical patterns that we have already encountered when looking into earlier systems.

The syntax of the system hardly needs comment. We obtain it by adding D either to the P,F,H,G-system of propositional tense logic that we defined in Section I.1.3 or to the corresponding sytem of tense predcate logic of Section I.1.3.2. (We could of course also combine D with S and U, but we won't bother to do that.) As far as the semantics is concerned, most of what is needed has already been said too, but there remain a few points to be settled. First, there is a question of uniformity. Our models, we said, are to take the form of bundles of worlds, where each world has its own history through time. There is no a priori reason that the time structures of different worlds should be the same. In fact, if we think of the time of a world as generated by the events that happen in that world, then we should expect that there can be worlds within the same bundle whose time structures do not coincide. (Recall in this connection the discussion at the end of Section I.1.3). On the other hand, the assumption that the worlds that make up a given model do not differ in their time structures has a certain appeal, since models become easier to handle, when they satisfy this condition. The situation is thus somewhat comparable to the one we encountered in connection with the semantics for modal and tense predicate logic. There we observed that there was no reason to assume that the universes of the different worlds in a model are the same, but adopted a construction which nevertheless enabled us to get by with that assumption. Here the problem is less dramatic: the technical inconveniences that come with assuming that time may vary from world to world aren't as big as they are in connection with quantification, while on the other hand adopting a single time structure per model doesn't seem to lead to much of a conceptual distortion. (Not at least in connection with applications to the analysis of meaning, which are our primary concern here.) We will assume therefore that each of our models $\mathbb{M}$ comes with a single time structure $\mathbb{T}$ and that every world of $\mathbb{M}$ represents the unfolding of a complete history through $\mathbb{T}$.

The second point we need to clear has to do with the notion of two worlds coinciding up to a certain time. We will treat this notion as a primitive and represent it as a 3-place relation, between two worlds and a time. We denote this relation as "$\equiv$", writing "$w \equiv_t w'$" to say that the relation holds between w, w' and t, and pronouncing this relation as: "worlds w and w' coincide (at least) up to time t". For given t, $\equiv_t$ is an equivalence relation between worlds (as the paraphrase: "coincides" implies) and this relation becomes stricter as time goes on: when w and w' coincide up to t and t' < t, then they also coincide up to t'.

There is one further issue connected with coincidence. Coincidence between w and w' up to t means, we said, that at t happen and obtain in w (and conversely). We can see this as carrying certain implications for the interpretation function F of the model. Let us restrict attention just to the propositional case. (We won't give a formal definition of the model theory for a predicate logical version of the present system, although this wouldn't involve much additional effort.) One position that would seem natural to adopt is that atomic expressions never pertain to the past or future but only to the present. More specifically, the atomic expressions of a propositional system, the propositional constants q_i, should be seen as making claims about what is currently the case, and not about what was the case or what will be the case. On this assumption the interpretation function should assign each q_i the same value at any time $t' \leq t$ in two worlds w and w' which coincide up to t.

I have incorporated this extra assumption about the interpretations of atomic sentences in coinciding worlds as condition (v) into Def, 15 below, which specifies the models for the P,F.H,G,D-system. But there is no need to insist on this: Whoever prefers this, can drop condition (v) from the definition without problematic side-effects.

Def. 15 A *model for* the propositional system with the operators H, G, P, F and D is a 4-tuple $<W,<T,<>,\equiv,F>$, such that

(i) $<T,<>$ is a linear frame
(ii) W is a non-empty set
(iii) $\equiv$ is a 3-place relation between worlds, worlds and times ($\equiv \subseteq W \otimes W \otimes T$);

for each $t \in T$, $\equiv_t$ is an equivalence relation on W;
for all t, $t' \in T$, if $t < t'$, then $\equiv_{t'} \subseteq \equiv_t$.

(iv) F is a function which assigns, for each $w \in W$, each $t \in T$ and each propositional constant q_i, a truth value $F_{w,t}(q_i)$ in w at t

(v) for all w, $w' \in W$, $t \in T$ and propositonal constant q_i, if $w \equiv_t w'$, then $F_{w,t}(q_i) = F_{w',t}(q_i)$.

The truth definition for the (H,G,P,F,D)-system is what the reader presumably expects, given all that has already been said on this score in the remarks leading up to Def. 15. The truth value of a formula now depends on both w and t. As always the truth functional connectives are interpreted "locally"; here that means that the truth value of, say, $\neg A$ in w at t depends only on the truth value of A in w at t, and so on. Formulas whose outer operator is a tense operator depend for their truth value in w at t only on the truthvalues of the argument formula at other times in the same world w; and finally, formulas of the form DA will be true in w at t iff A is true in w' at t for all worlds w' which coincide with w at t. We just give the clauses of the truth definition for these three cases, trusting that the reader will be able to write down the remaining clauses himself.

Def. 16 (Fragments of the truth definition for the (H,G,P,F,D)-system)

(i) $[\neg A]_{\mathbb{M},w,t} = 1$ iff $[A]_{\mathbb{M},w,t} = 0$

(ii) $[PA]_{\mathbb{M},w,t} = 1$ iff for some $t' \varepsilon T$ such that $t' < t$, $[A]_{\mathbb{M},w,t'} = 1$

(iii) $[DA]_{\mathbb{M},w,t} = 1$ iff for all $w' \varepsilon W$ such that $w \equiv_t w'$, $[A]_{\mathbb{M},w',t} = 1$

One point of interest connected with the (H,G,P,F,D)-system is that here we have a multi-modal system, with temporal operators pointing in two different directions and in addition a modal operator D, whose models do not seem to have the form that might have been expected for such a system, given what we have said about multi-modal systems in I.1.3. The models described there are structures with a single "index set" I and accessibility relations on I for each of the system's operators. (Recall that truth values in such models are dependent on the "indices" in the set I.) It is not difficult, however, to recast the models of Def. 15 as models of such a form. Let $\mathbb{M}$ be a model in the sense of Def. 15 and let $\mathbb{M}'$ be a structure derived from $\mathbb{M}$ as follows:

(i) We choose as index set I of $\mathbb{M}'$ the cross product $W \otimes T$;

(ii) The accessibility relation R_F for the operators G and F is that relation on I defined by:

$$\langle w,t\rangle\ R_F\ \langle w',t'\rangle \text{ iff } w = w' \text{ and } t < t';$$

similarly the accessibility relation R_P for the operators H and P is the relation defined by:

$$\langle w,t\rangle\ R_P\ \langle w',t'\rangle \text{ iff } w = w' \text{ and } t > t';$$

(iii) The accessibility relation R_D for the operator D is the relation on I defined by:

$$\langle w,t\rangle\ R_D\ \langle w',t'\rangle \text{ iff } t = t' \text{ and } w \equiv_t w';$$

It is easily verified that the truth clauses for formulas of the forms GA, HA and DA can then be given the same form as the clause for $\Box$A in our formulation of Kripkean semantics for modal logic in Section I.1.1. (Likewise, the clauses for P and F can, as we saw before, be given in the same form as the clause for $\Diamond$). It is also easy to see that these truth clauses assign to any formula A the same truth value in $\mathbb{M}'$ at <w.t> that it gets in $\mathbb{M}$ in w at t.

The point illustrated by the models for the (H,G,P,F,D)-system that are specified in Def. 15 is that it is often desirable to give the semantics for multi-modal systems by defining models from which accessibility relations for the different operators of the system can be reconstructed, but which at the same time impose certain constraints on those relations that are difficult or less natural to state when the accessibility relations are specified directly. One indication of the connections between the accessibility relations for H and D that are built into the models defined in Def. 15 is that formulas of the form HDA → DHA are valid on our semantics, while formulas of the form DHA → HDA are not. (Exercise: show this)

I.1.4 2-place Operators outside Tense Logic

One of the earliest applications of modal logic outside the conceptual domain of necessity and possibility was that to *Deontic Logic* (Von Wright, Hintikka, Anderson, Hansson, Aquist and others). Deontic logic is the logic of "ought": of what ought to be the case, from either a moral or legal point of view, and also of what may be the case, in the sense of being permissible and of what ought not to be the case, in the sense of being morally wrong, or forbidden by law.

Applying modal logic in this direction seems natural because the deontic notions just mentioned have a distinct flavour of necessity or possibility: what ought to be the case for moral or legal reasons is, one could also say, what is necessitated by moral or legal principles; and what is permissible can be seen as what is possible as far as such principles are concerned. In fact, given these suggestive parallels, it might be thought that all we need in order to apply the modal systems of sections I.1.1 and I.1.2 to these deontic notions is a special interpretation of the necessity and possibility operators $\Box$ and $\Diamond$. We will follow tradition in using the symbol "O" for the deontic necessity operator. ("O" stands for "ought". Sometimes "P" is used for the deontic possibility operator, but we won't do this here because of the potential confusion with the use of P as past tense operator.)

The main task that the application presents us with is the semantics for O. One obvious difference with necessity in the sense of Sections I.1.1 and I.1.2 is that what ought to be the case isn't thereby true, i.e. the schema $OA \rightarrow A$ should not come out as valid. ("You cannot derive "is" from "ought", as the Humean slogan has it.[24]) This means that if we interpret O via an accessibility relation R_O, then R_O should in general not be reflexive. The simplest characterisation of R_O – and as far as I know also the first to be proposed – satisfies this requirement. It rests on the following idea: what ought to be is what is true in those worlds in which everything is as it ought to be – in those worlds that are morally or legally perfect in that everything in them is in accordance with the relevant moral or legal principles. More formally, a model for deontic logic should now not just provide a set W of possible worlds, but also a subset MP of W, consisting of all those worlds in which everything that ought to be the case is the case. ("MP" stands for "morally perfect".) MP gives us an accessibility relation R_O via the obvious and simple definition:

---

24 See David Hume (1711-1776), *A Treatise of Human Nature.*

(28) for all $w, w' \in W$, $w\ R_O\ w'$ iff $w' \in MP$

It is easily verified from this definition that provided MP is a proper subset of W, R_O is not reflexive on W. But otherwise the structure of R_O is very simple: on its own co-domain (that is, on the set of those worlds w such that for some w', $w'\ R_O\ w$) R_O is the universal relation, and so *a fortiori* an equivalence relation.[25]

Unfirtunately this semantics is too simple. Problems with deontic logic in this form were realised more or less instantly, and in fact go back to discussions that antedate the attempt to develop a logic of “ought” along the lines of modal logic. They arise as soon as one tries to use the system just described to formalise general principles, such as “Someone who has committed a serious crime should be punished”, or “If someone is in trouble, you should help him”. Let us focus on two instantiations of these principles, given in (29).

(29) i. If Fred has committed a murder, he ought to go to jail.
ii. If Fred has a tumor, his tumor ought to be removed.

How should such sentences, which involve both an “ought” and a conditional, be formalised in our system? The central issue here, it might be thought, is that of scope: Should “ought” be treated as having scope over the entire conditional or should its scope be restricted to the main clause (i.e. to the consequent of the conditional)? From what is known about the interaction between auxiliaries and *if*-clauses neither possibility can be excluded *a priori*, so both schemata in (30) are potential candidates:

(30) i. $O(A \rightarrow B)$
ii. $A \rightarrow OB$

But neither of these candidates stands up to scrutiny. The inadequacy of (30.i) is easiest to demonstrate. Consider (29.ii). The *if*-clause of (29.i) is something that itself ought not to have been the case. That is, if we abbreviate “Fred has committed a murder” as p_1, then $O \neg p_1$ should be considered true. However, it is easy to verify from the semantics of O that if A logically entails B, then OA logically entails OB. Note, however that because of the truth-functional properties of $\rightarrow$, $\neg p_1$ logically

25 This means that if S is any valid schema of S5, then OS is a valid schema of Deontic Logic with the given MP-based semantics; and conversely. (Exercise: Show this, using the fact that the valid schemata of S5 are those which are valid in all models in which the accessibility relation R is an equivalence relation)

entails $p_1 \rightarrow B$ for any B whatever. So $O \neg p_1$ entails $O(p_1 \rightarrow B)$ for any B whatever. Suppose now that we abbreviate the main clause of (29.i) without "ought", i.e. "Fred goes to jail", as q_1 and formalise (29.i) as an instance of (30.i), i.e. as $O(p_1 \rightarrow q_1)$. This clearly doesn't catch what (29.i) wants to express. For given that $O \neg p_1$, the formula $O(p_1 \rightarrow \neg q_1)$, which says that if Fred has committed a murder, he should not go to jail, will be just as true as $O(p_1 \rightarrow q_1)$. Obviously this is not what we want.

In the light of this argument we may conclude that (30.i) is in general too weak to capture correctly what is expressed by conditionals like those in (29). With (30.ii) we have a different problem – in a certain sense it is too strong. We can see what this problem is by having a closer look at (29.ii). In (29.ii) the proposition q_2 that is expressed by the main clause without "ought" (i.e. the proposition that Fred's tumor is removed) entails the proposition p_2 expressed by the *if*-clause (i.e. the proposition that Fred has a tumor): you can only remove a tumor from someone if he has one. Note that it folllows straightforwardly from the semantics we have chosen for O that for arbitrary A and B, if A entails B, then OA entails OB.[26] So we can conclude that Op_2 entails Oq_2. But that means that if we adopt the instance $p_2 \rightarrow Oq_2$ of (30.ii) as formalisation of (29.ii), then we also get $p_2 \rightarrow Op_2$, which says that if Fred has a tumor, then he ought to have a tumor. This statement seems to express some sort of moral fatalism: everything that happens is morally right, because otherwise it wouldn't have happened. (For instance, because otherwise God, who knows everything and decides everything, would not have allowed it to happen). This position may not be exactly inconsistent – there appear to be people who hold such beliefs about right and wrong – but it certainly isn't something that should be imposed upon us by mere logic.

We have only looked at the two most obvious candidates for the formalisation of conditionals like those in (29). So, just showing both of these to be unsatisfactory is no proof that the present system of deontic logic is altogether incapable of providing adequate formalisations for such conditionals. But in fact, the difficulties we have discovered in

---

26 Recall that in the weakest of the modal systems that were listed in Section I.1, viz. the system K, the following schema is valid: (i) $\Box (A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$: moreover, from the validity of $A \rightarrow B$ we can infer in K the validity of $\Box (A \rightarrow B)$. And finally, it follows from the fact that A entails B that $A \rightarrow B$ is valid. But, obviously, everything that holds in K also holds for our present semantics for O, since the valid principles of K are entirely neutral with regard to the structure of the accessibilty relation R.

discussing these two options point us to the real source of the trouble. Morality and the law aren't only about the distinction between the worlds that are morally or legally perfect and all the other, more or less imperfect worlds; they are also, and crucially, about what can and should be done in worlds that are corrupted by earlier transgressions of moral or legal principles, but which we should nevertheless endeavour to make or keep as good as circumstances permit. In particular, we want a semantics that allows for the comparison – from a moral or a legal point of view - of worlds w and w' both of which are tainted with things forbidden or reprehensible. For even for two such worlds it is often intuitively clear which one is the morally or legally better one. In a nutshell, two sins *are* worse than one. The semantics for O should be sensitive to such distinctions.

One way in which we can formulate such a more sensitive semantics for O makes use of the model theory for the multi-modal system of the last section. Note that both conditionals in (29) talk about what should be done once something else has occurred. That something else may be deplorable, but it cannot be undone - it is, in the terminology of the last section, "determinate". But that doesn't make the question how the world should be made to continue now, at the time t that that something else has become a fact, any less urgent. And as a rule there is some answer to that question: Some continuations of the world after t are morally or legally preferable to certain others.

One way to model this distinction between good and bad continuations of a given world from a certain time t onwards is to enrich the models of Def. 15 with a function PREF, which maps each pair consisting of a world w and time t onto some subset PREF (w,t) of the set PC(w,t) = {w': w $\equiv_t$ w'} of *all* possible continutations of w after t. (PREF should of course be invariant with respect to the relation $\equiv_t$; that is, we have in general: if w $\equiv_t$ w', then PREF (w,t) = PREF (w',t).) We can then define the semantics of O by means of the clause (31):

(31) $[OA]_{\mathbb{M},w,t} = 1$ iff for all w' ε PREF(w,t), $[A]_{\mathbb{M},w',t} = 1$

In this system the conditionals in (29) can be formalised as in (32):

(32) i. $Dp_1 \rightarrow OFq_1$
ii. $Dp_2 \rightarrow OFq_2$

In plain words (32.ii) says that if p_1 is determinately true in w at t, then the morally/legally right continuations of w after t are those in which

q_1 is true at some time t' > t. For instance, if the murder has been committed at t, then it is morally/legally necessary for the murderer to go to jail at some time t' after t. Likewise for (29,i).[27]

The embedding of the "ought" operator O within the multi-modal (P,F,H,G,D)-system requires us to be specific about the temporal aspects of the *ought*-sentences which we want to represent. To some this might seem an unnecessary complication. But the temporal dimension is crucial to our understanding of many deontic and other conditionals. Therefore, being forced to represent this aspect of conditionals explicitly often helps us to understand better what they really mean, and what seemed to be a mere nuisance at first proves to be a blessing in disguise in the end.

Although this approach, in which deontic operators are added to the multi-modal system of the last section, appears to be very natural, we will not pursue this possibility any further here.

27

General principles, such as that murderers must go to jail and tumors must be removed can then be represented by formulas in which the schema of (32) is embedded within the scope of an operator saying "it is always the case that". To formulate general principles like "If someone commits a murder he should go to jail." we need more than just a system of propositional multi-modal logic, however. For such principles typically involve not only universal quantification over times (which we can express with the help of tense operators) but also over entities of other sorts. For instance, the general principle just mentioned intuitively involves universal quantification over whoever committed a murder and should be sent to jail. (That universal quantification can be expressed with the help of *someone* in such sentences is an intriguing fact of natural languages like English, to which we drew attention in the introduction and which we will consider in detail in Part II. Here we will just assume that *someone* acts as a universal qunatifier with scope over the entire conditional which binds *he* as a kind of bound variable.) So we would need a predicate logic version of the propositional system with P,F,H,G,D and O defined in the last section. Given what has been said in earlier sections about the "upscaling" of propositional modal systems to systems of predicate logic, defining such a version would at this point not pose any particular problems. We won't go into the details of this, however, and simply assume that a predicate logic version of the (P,F,H,G,D,O)-system is in place.
In such a ssytem we can represent the general principle mentioned above as:

$$A(\forall x)(D(P(x) \rightarrow OFQ(x)),$$

where P(x) stands for "x has committed a murder", Q(x) for "x goes to jail" and for any B "AB" is an abbreviation for "HB & B & GB". Thus "AB" can be read as "It is, always has been and always will be the case that B".

Those involved in the early development of deontic logic had a different reaction to the problems connected with the 1-place operator O that we have just drawn attention to. It is this reaction that motivates including this discussion of deontic logic in these notes.

The reaction can be summarised as follows: The problem with using the 1-place operator O in representing deontic conditionals like those in (29) is that it presupposes that the deontic dimension and the conditional aspect of such sentences can be separated. But that is not so. Sentences like those im (29) express connections between *if*-clause proposition and main clause proposition in which the conditional and the modal aspect form a single, indivisible semantic whole. To do justice to this it is necessary to represent these two meaning aspects with the help of a single 2-place operator - a "deontic conditional connective" - whose semantics captures the way in which conditionality and deontic modality are bound together. (The graphics of the symbol "O⇒" that is commonly used to represent this operator are designed to bring out the inseparability of these two aspects.)

Using the operator O⇒, the sentences in (29) can be represented as p_1 O⇒ q_1 and p_2 O⇒ q_2, respectively. But what should we take these formulas to mean? The answer is contained in the following semantic intuition:

> A O⇒ B should count as true provided among the worlds in which A is true, those in which B is true are on the whole preferable to those in which B is false.

But this is still quite informal an vague. Puzzling may seem in particular what could be meant by the phrase "on the whole preferable" – why add this qualification "on the whole""?.

The reason for adding the qualification is the following. It would be nice if we could state the truth consitions of A O⇒ B in as simple a form as that in (33).

(33) A O⇒ B is true iff any A-world (i.e. world in which A is true) that is also a B-world is preferable to any A-world which is a ¬B-world.

But it is easy to see that (33) cannot be right. A holocaust world in which our murderer Fred is sent to jail is surely not to be preferred to a world where Fred doesn't go to jail and where there is no holocaust.

One way to deal with this complication is to assume that it is possible to compare worlds according to their moral or legal soundness over all. Suppose that we have a relation $\geq$ which does this – i.e. $w \geq w'$ means that w is (morally, or legally) preferable to w' over-all. We can then define A O⇒ B as true if, roughly, for every A world that is a ¬B-world there is an A world that is a B-world and that is preferable to it. There can still be some argument over the exact formulation of this last condition. But a widely used form is that in (34), which articulates the truth conditions for A O⇒ B in the model-theoretic formal that we have been using throughout.

(34) $[A\ O\!\Rightarrow B]_{\mathbb{M},w} = 1$ iff (i) $(\forall w' \in W)(([A]_{\mathbb{M},w'} = 1\ \&\ [B]_{\mathbb{M},w'} = 0) \rightarrow$
$(\exists w'')(w'' \geq w'\ \&\ [A]_{\mathbb{M},w'} = 1\ \&\ [B]_{\mathbb{M},w'} = 1))$
&

(ii) $(\exists w'' \in W)([A]_{\mathbb{M},w'} = 1\ \&\ [B]_{\mathbb{M},w'} = 1\ \&$
$\neg(\exists w'')(w' \geq w''\ \&\ [A]_{\mathbb{M},w'} = 1\ \&\ [B]_{\mathbb{M},w'} = 0))$

(In words: A O⇒ B is true in $\mathbb{M}$ at w iff (i) for every A-world that is not a B-world there is an A-world that is a B-world and that is at least as good as the first, and (ii) there are A-worlds that are B-worlds and that are better than any A-world that is not a B-world.)

Where, you may ask, does the relation $\geq$ come from? Is there any way in which it can be defined or explained in terms of notions more immediately connected with morality or the law? This is a difficult question to which a fully satisfactory answer is still outstanding. Consider over-all legal preferability. If w is preferable to w' in this sense then this should have to do with there being fewer violations of the law in w than there are in w'. But it cannot be just the number of violations, the gravity of the individual transgressions must count for something too. Even a hundred parking violations cannot outweigh a single case of causing death through reckless driving. And other factors may play a part in the comparison too. How is one to compare a world in which there is one brief explove spell of violations of the law with one with a modest but steady trickle of violations? And should the distribution of violations over the population at large also count for something? The more you think of it, the more difficult it seems to come up with some kind of formula for relating over-all preferability to the way in and degree to which the law is being observed.

The notion of moral over-all preferability is even more problematic, for here there is an additional problem with the notion of a fixed well-defined set of moral rules. That moral perfection consists merely in "never doing anything wrong" is one conception of morality, but for many of us it is neither very plausible nor very "likable". The history of ehtics is marked by a remarkable variety of quite different conceptions of what constitutes morality. One that is very different from the rule-based conception, according to which moral behaviour is just a matter of staing within a well-defined moral code, is Benthamist *Utilitarianism*.[28] According to this conception a course of action is moral provided it produces the greatest amount of happiness and least amount of pain over-all ("the greatest happiness for the largest number"). Bentham thought of ethics as a kind of calculus which would enable us to compute the happiness resulting for each of us from a given course of action, and actions could then be compared according to how much happiness (and how little pain) they would cause. To the extent that such calculations are possible at all, one could imagine it to then also be possible to calculate the aggregate happiness of an entire world.´, and them compare worlds in terms of their over-all happiuness values. But, of course, except for some applications in the domain of public policy (some of which were of the first importance in his own day), Bentham's idea of putting numbers on states or feelings of happiness are really quite cookie, and it has been a long time since anyone took it seriously.

But then, what measure should be applied in comparing the over-all moral goodness of two worlds? Nobody really knows. However, in spite of this deontic logicians have been ready to assume that relations ≥ do exist, in terms of which the truth conditions for A O⇒ B can be given along the lines of a condition like (34).

(34) is one way to spell out the semantics of O⇒. But it is only one of several variations that can be found in the literature. (Different variations typically give rise to somewhat different logics for O⇒, but that need not be our concern here.) Moreover, the literature contains many other examples of 2-place modal operators (not necessarily deontic), whose semantics is roughly comparable to that of O⇒. Among those to whom we owe much of our knowledge and understanding of such operators should be mentioned in particular David K. Lewis (Lewis, 1973) and Angelika Kratzer (Kratzer, 1978,1979,1981). Kratzer developed in her doctoral dissertation a general account of the semantic relations which in languages like English and her native German (on which her early work is strongly focused) are expressed by combinations of modal auxiliaries and *if*-clauses (as in the sentences in (29)). Different auxiliaries, as well as different interpretations for one and the same auxiliary, give rise to different operators.

Kratzer's semantics for such operators involves two separate components:

(i) the *modal base*. This is the set of possible worlds that are consistent with all that is taken for granted in the context in which a given conditional is being used; worlds not belonging to this set are worlds that are simply not considered in the given context. (Examples of worlds that are excluded from most contexts would be "outlandish"

---

28 Jeremy Bentham (17??, 18??) English philosopher and the father of so-called *Utilitarianism* in ethics, in which the moral *optimum bonum* is equated with "the greatest amount of happiness for the largest number".

worlds in which, say, the laws of nature fail, or fairy tale worlds in which there are talking cats or gigantic caterpillars, or worlds in which Paris is situated in the English Midlands, or (pointing ahead to the next section) in which Bamberg is part of Saxony. But often the context will exclude many more worlds, so that the modal base may be quite restricted.)

(ii) the *ordering source*. The ordering source determines which worlds count as preferable to which others for the evaluation of the particular operator in question. This means that the ordering source varies as a function of the kind of modality that is involved, even within one and the same speech context. For example, legal "ought" and moral "ought" correspond to different ordering sources. (The first, as we saw, compares worlds according to their moral soundness, while the second compares them in terms of number and seriousness of violations of the law.) But Kratzer's theory is designed to cover not only moral and legal modalities, but also modalities of altogether different types, such as for instance *epistemic* modalities, illustrated by a sentence like the following:

(35) "If Fred invested heavily in London real estate in the nineties, he ought to have made a lot of money."

In (35) "ought" doesn't have an ethical or legal connotation, but serves merely to indicate that the consequent may be inferred from the antecedent on the basis of general knowledge the speaker presupposes - here: knowledge about the development of the real estate market in London since the nineties. (Issues related to the epistemic use of modal auxiliaries will be discussed in the next section.)

Modal base and ordering source can be used to state the truth conditions for the different modalised conditional operators, using definitions of the form (34) (or some variant of it). In (34) the modal base was not mentioned explicitly, but we can think of it as present implicitly, viz. as the field of the relation $\geq$.[29] However, as Kratzer has argued convincingly, there are good reasons for keeping modal base and ordering source apart. For instance, speakers often use deontic and epistemic conditionals in the same context; in such situations the two conditionals will typically involve the same modal base, but different ordering sources (a deontic and an epistemic ordering source, respectively).

[29] The field of a 2-place relation R is the set of all those entities which occur either as first or as second terms of R: $field(R) = \{u: (\exists u')(uRu' \vee u'Ru)\}$. So the field of $\geq$ is the set $field(\geq) = \{w \in W: (\exists w' \in W)(w \geq w' \vee w' \geq w)\}$.

I.1.4.1 Alethic Conditionals

In this final section of our survey of structure and applications of modal logic we return to its original application domain, that of the notions of necessity and possibility, in the more specific sense of what is necessarily or possibly true. In other words, the concepts at issue are once more those that occupied us in Sections I.1.1 and I.1.2.

The sytems of modal logic we looked at in Sections I.1.1 and I.1.2 had 1-place operators, $\Box$ for "necessarily" and $\Diamond$ for "possibly". In Section I.1.3 we moved to a different application domain for modal logic systems, in which times play the part that was previously played by possible worlds. Instead of the modal operators of the first two sections the operators we were now dealing with and the logical systems of which those operators were the distinctive constituents were suitably called "tense operators " and "tense logics". We noted that the sytems of tense logic which could be obtained in this way cannot express many temporal relations between propositions that one would want to be within the scope of such a system, and that - we simply stated but did not prove this - no addition of further 1-place operators can close this gap. This led us to the introduction of the 2-place tense operators S and U. In the last section we then reported on the use of 2-place operators in another application domain than that of tense logic, that of the deontic modalities.

In this section we will be looking at some of the arguments that suggest that 2-place operators are needed also in the domain of the *alethic* modalities, i.e. the modalities of being possibly and necessarily true. In fact we will be doing more than that. In introductions to formal logic it is not uncommon for students to be brow-beaten into accepting that the *if.., then..*-sentences of natural language can, without serious loss or danger, be treated as material conditionals. We will begin by looking at some of the pros and cons of this claim. After concluding that the "material conditional" analysis of natural language conditionals isn't tenable after all, we will then move on to an argument that the conditionals of natural languages cannot be analysed either as strict conditionals (i.e. as having the logical form $\Box(A \rightarrow B)$), but that their logical form requires a 2-place modal operator, which we will represent as $\Box\!\!\Rightarrow$.

There is a long-standing debate, going back to antiquity, whether natural language conditionals – and prominently among them those epxressed by *if.., then..*-sentences - are truth-functional, i.e. whether their truth conditions are those of the material conditional, stated once more explicitly in (36)

(36) an *if.., then..*-sentence is false when its antecedent is true and its consequent is false, and it is true in all other cases; and that is all there is to its truth conditions.

On the face of it (36) seems very implausible. Consider for instance the conditionals in (37).

(37) i. If Mary is in Bamberg, then she is in Bavaria.
ii. If Mary is in Bamberg, then she is in Saxony.
iii. If Mary is in Bamberg, then Liverpool just won against Chelsea.

(37.i) seems plainly true. Bamberg is part of Bavaria, So if anyone is in Bamberg, she is *ipso facto* in Bavaria, she can't possibly fail to be in Bavaria. And this consideration is independent of whether Mary is actually in Bamberg or not. It only excludes the possibility that she would be Bamberg without being in Bavaria. By much the same token (37.ii) seems false. Bamberg is not in Saxony, so by being in Bamberg Mary would *not* be in Saxony. If as it happens, Mary is not in Bamberg, that, we feel, doesn't make (37.ii) any better. It still seems just as false, even though as a material conditional it ought to be true in this case, given that its antecedent is false.

(37.iii) illustrates the same point from a slightly different angle. If you hear me say this sentence out of the blue, you may wonder what I am talking about. You may be groping for a story that could account for my making such a statement: Perhaps Mary belongs to a German fan club of Liverpool FC which has made it a habit to congregate in Bamberg for a joint celebration each time their club has scored an important win. Not frightfully plausible perhaps, but when you hear something like (37.iii) you feel some pressure to look for sómething that enables you to make sense of it. Or perhaps you won't come up with anything plausible and you suspend judgement, hoping that some explanation will be forthcoming.

Suppose now that the day after my statement of (37.iii), when the facts are in – Mary wasn't in Bamberg the day before, and Liverpool lost – you ask me what I really had in mind when I made my statement the previous day and suppose I reply by saying: "Well, I wanted to tell you something that was true and that you didn't know. And, indeed, that is what I did. I did tell you something that you didn't know at the time and what I said wás true, wasn't it. For Mary wasn't in Bamberg yesterday." That wouldn't be a very satisfactory answer, and you could rightly accuse me of avoiding the issue.

Such considerations seem to suggest strongly that the content of conditionals like those in (37) is not that of the material conditional. However, the matter is not quite as straightforward as that. That it isn't was pointed out in the nineteen fifties by Paul Grice (), in one of his first attempts at what developed gradually into his *Theory of Conversation* (the name under which Grice's theory is now generally known). The argumentation strategy of Grice's Theory of Conversation can be applied to a broad spectrum of issues in the theory of meaning. But here we will only be concerned with the possibility of applying it to the semantics of conditionals.

The argument begins by conceding that, typically, when a speaker makes a statement of the form "if A, then B", she does so because she perceives a certain systematic connection between the truth conditions of A and those of B. That connection excludes the possibility that A would be true without B being true as well. The connection should be one that is independent of the actual truth values of A and B; it should not only exclude the possibility that A is true and B false for the actual world, but for a whole range of alternative possible worlds. Moreover, it is not just that the speaker herself will perceive such a connection, by stating "if A, then B", she will convey that there is such a connection also to her audience.

This much is in agreement with the observations we made in relation to our conditionals in (37). But now comes Grice's point. We can explain all this, he argues, on the simple assumption that the truth conditions of "if A, then B" are those of the material conditional $A \rightarrow B$, provided we also take into account certain "principles of conversation"– principles that govern the proper use of language in communication, or "conversation". Suppose that in a given conversational situation a speaker knows the truth value of either A or B, and that this truth value is compatible with the material conditional $A \rightarrow B$; that is, the speaker knows that A is false or she knows that B is true, or perhaps she knows even more – that A and B are both false, or that A and B are both true, or that B is true while A is false. In each of these situations the speaker could have made a stronger statement by just asserting what is the case concerning A and/or B – e.g. by saying "not-A", or by saying "B", and so on – than she would have been making had she uttered the conditional "if A, then B". (At least this is so on the assumption that the truth conditions of "if A, then B" are the same as those of $A \rightarrow B$.) So, according to one of the principles of conversation, according to which a speaker who makes an assertion should be as informative as she can be, the speaker should have made a direct statement about the truth value of either A or B (or both), rather than using the less informative conditional.

That leaves as the only occasions on which the conditional can be properly used those in which the speaker doesn't know of either A or B whether or not it is actually true. But on such an occasion the only grounds she can possibly have for assuming the conditional to be true, is that there is some systematic connection between A and B, which excludes the possibility that the second is false while the first is true. Only when all these conditions – no knowledge of the actual truth values of A and B combined with knowledge of some systematic connection between their possible truth values - are fulfilled, the argument concludes, will the speaker be entitled to assert the conditional.[30] This explains on the one hand why those who assert conditionals must assume there to be some systematic connection between antecedent and consequent. And it also explains why the recipient of a statement of the form "if A, then B" will typically infer that there must be a systematic connection between A and B. For normally recipients assume that speakers abide by the principles of proper conversation. And if the speaker who stated "if A, then B" has done that, then it follows that (at least as far as the speaker can see) such a connection exists.

The general upshot of this argument is methodological: Even on the simple assumption that *if.., then..*-statements have the content of material conditionals, the same observations about the implications carried by such statements can be accounted for. But then it is preferable to stick to this simple hypothesis about the meaning of "*if* A,

30 One implication of this is that when a speaker makes a legitimate utterance of a conditional, she will typically say less than she could say so long as she doesn't make the systematic connection she perceives between A and B explicit. But that is a quite general feature of verbal communication: We make an additional verbal effort to convey additional information only when we think that that additional information is relevant to what we are saying, or useful to the recipient, and we have reason to think that the recipient wouldn't be able to infer that information from what we have said already. Note that this observation does not conradict the Gricean principle invoked in the text. There the issue is between the assertion of the weaker and at the same time longer and syntactically more complex conditional "*if* A, *then* B" on the one hand and the simpler and logically stronger "not-A" (or "B", as the case may be) on the other.

then B". A particularly attractive feature of this hypothesis, moreover, is that it narrows the gap between natural language and classical logic.[31]

Grice's strategy is ingenious and appealing (especially to those who believe that there is an important place for formal logic in the analysis of meaning), and in the course of the fifty years that have passed since his first proposals along these lines, it has become an essential tool in the theorising about meaning in natural language and in particular about the relation between semantics and pragmatics. But, alas, its application to the semantics of *if.., then..*-statements is one that does not stand up to closer scrutiny - not at least when is taken as a general account of such statements. Among the cases that tell most strongly against the "material conditional" analysis of *if.., then..*-statements are so-called *counterfactual* conditionals – conditional sentences like those in (38), whose *if*-clause is in the past perfect and main clause in the future perfect of the past (the form *would have* + past participle).

(38) i. If Maria had been in Bamberg, she would have been in Bavaria.
ii. If Maria had been in Bamberg, she would have been in Saxony.
iii. If Fred had taken a train after two o' clock, he wouldn't have been in time for the meeting.

Such conditionals are called "counterfactuals" because they imply that their antecedent and consequent are both false: You are not supposed to use this form unless you know this to be so.[32] Moreover, in many

---

31 Grice's Theory of Conversation, of which this application to the semantics and pragmatics of conditionals was one of the first manifestations, arose within a philosophical climate (that of Oxford University in the nineteen fifties and sixties) which was on the whole quite hostile to formal logic as a tool of philosophical analysis and skeptical about its usefulnesss as a tool in the analysis of meaning. Grice's theory aimed at debunking some of the arguments that were put forward to show that meaning in natural language is organised along lines that are different from those that govern the semantics of formal languages like the predicate calculus. One of the arguments was that conditionals in natural language function in an entirely different way from the material conditional in classical logic.

32 There are some variants of this sentence pattern which only entail falsity of the antecedent. An example is the sentence: "Even if Fred had taken a train between one and two, he wouldn't have been on time.", which could be said by someone who knows that Fred wasn't on time in actual fact. The "even" of this sentence turns the implication concerning the main clause around: The sentence implies that it is being used in a context in which it is assumed that the main clause is true under conditions which differ from those that are described by the sentence's *if*-clause. (These may be

situations in which a counterfactual is used, the speaker not only knows about the falsity of antecedent and consequent herself, but also knows or assumes that this information is available to the addressee. Given that this is so, the "Gricean" argument we have run through cannot be right. For if it were, then counterfactual conditionals of the form "if it had been the case that A, then it would have been the case that B" could never be used correctly. As we have just seen, their correct use requires the speaker to be aware that A and B are false. But then, according to the Gricean argument, the speaker would have been in a position to make the "stronger" assertion "not A and not B" (or simply "not A"), and so she should have made that statement - rather than the conditional, which according to the Gricean assumption expresses the weaker proposition "not A or B" (i.e. the one expressed by the material conditional). In fact, in those cases where the speaker also assumes that the addressee knows that A and B are false, the use of the counterfactual would be even more absurd: If all the counterfactual meant was that not A or B and if the addressee already knows that not A, then the statement of the counterfactual conditional would not provide him with any new infromation at all, and so would have been pointless.

To summarise: on the Gricean account of conditonals sketched out above there could be no occasions for the poper use of counterfactual conditionals. But that is of course nonsense. People do use counterfactuals and normally our uses of them seem perfectly legitimate and above board. Moreover, when we reflect on the typical uses of counterfactuals the impression that they do express some systematic connections between antecedent and consequent becomes compelling.

Perhaps the most prominent use for conditionals like those in (38) is that where we look back in time to a point t when a certain decision was still to be made that could have been made differently, or where something was about to happen which might also not have happened, and where we contemplate what would have followed in case things had been different in a certain way from that time onwards. In such a situation it is not just that we must, in order to be in a position to make the counterfactual claim "if it had been the case that A, then it would have been the case that B", know something about the power of A to carry with it the truth of B (given what had been the case in the given world w up to time t); the claim we make can only be understood as a

---

either the actual conditions, or else conditions that were just mentioned by another speaker, as when the present sentence is uttered in reaction to (38.iii).)

claim about worlds other than w; for it is on those non-actual worlds that our attention is focussed and about which the counterfactual claim is meant to make a statement.

Once we have become persuaded that counterfactual conditionals cannot be analysed as material conditionals, the Gricean story looks more suspect also in relation to non-counterfactual conditionals, such as the indicative conditionals in (37) or subjunctive conditionals like those in (39).[33]

(39) i. If Mary were in Bamberg, then she would be in Bavaria.
ii. If Mary were in Bamberg, then she would be in Saxony.
iii. If Fred were to take a train after two o' clock, he wouldn't be in time for the meeting.

Arguing that such conditionals must also be understood as claims that the truth of A carries with it the truth of B in a range of different possible worlds isn't quite as straightforward as it is for counterfactuals. But once the point has been made for counterfactuals, maintaining that conditionals that are not strictly counterfactual have the semantics of material conditionals becomes an uphill battle. What speaks for a uniform (and thus, given what has already been establsihed at this point, non-truthfunctional) analysis of counterfactual and non-counterfactual conditionals is that by and large both kinds of conditional – the non-counterfactual as well as the counterfactual - are used to more or less the same ends. Consider for instance the conditionals in (39). One of their main functions is to give expression to our thoughts when we are making plans, in order to realise the goals we want to pursue. When we are engaged in planning, we rely on our kowledge of what kinds of effects are likely to be brought about by what kinds of actions, and under what circumstances. Such connections between actions and their effects are naturally expressed by conditionals in which the effect is stated in the consequent while the action, possibly in conjunction with some of the relevant circumstances, is stated in the antecedent. The conditionals used in the context of planning aren't – and shouldn't be – counterfactual, since the question whether the contemplated action will actually be performed is typically still undecided. That is what distinguishes the conditonals in (37) and

33 Other combinations of tenses in *if*-clause and main clause are possible as well, but here I have only listed those that seem most relevant in connection with our present concerns.

(39) from those in (38). None of the conditionals in (37) and (39) entails the falsity of either antecedent or consequent.[34]

Assuming that the truth conditions of sentences of the form *if* A, *then* B - all of them or most of them - are not those of the material conditional A → B, then what are they? To explore this question a bit more closely, let us return to what we said about the claim that a counterfactual can be used to make about non-actual contininuations of the world w after some past time t. (What we will be saying about this case can be easily adapted without so that it applies to other uses of counterfactual and non-counterfactual conditionals, but we will not go into that matter here, and simply proceed on the assumption that such an adaptation is possible.) At least for uses of this kind we can restate what we said about the content of the counterfactual *if it had been the case that* A, *then it would have been the case that* B in a way that still allows the material conditional A → B to play a certain role:

(40) *if it had been the case that* A, *then it would have been the case that* B is true in w iff for all the continuations w' of w after t, A → B is true in w'.

(The relation between (40) and our earlier description of the semantic contribution made by the counterfactual should be obvious: There are two kinds of continuations of w after t, those in which A is false, among which is w itself, and those in which A is true. It is about the latter that the counterfactual makes a claim directly, viz. that they also verify B. The material conditional A → B sets the continuations of the first kind aside, while securing the substantial claim that the conditional makes about the worlds of the second kind.)

In the light of (40) it might be thought that all that we need in order to represent counterfactuals and other non-material conditionals is already present in the modal systems of Sections I.1.1 and I.1.2. We can represent the counterfactual conditional *if it had been the case that* A, *then it would have been the case that* B as in (41) and our only remaining task will be to make sure that the necessity operator gets the intuitively correct interpretation.

---

[34] There seems to be a certain tendency for subjunctive conditionals like those in (39) to give more prominence to the possibility that the antecedent is false, and to some extent also to the falsity of the consequent being false more prominent than in the case of indicative conditionals. But the difference between such conditionals and indicative conditionals of the sort found in (37) is subtle, and we won't try to get to the bottom of what if any the difference is.

(41) $\Box(A \to B)$

Indeed, once the right set W' of relevant worlds has been determined, we can cast our analysis in the exact form of a necessity operator in the sense of Section I.1.1 be defining the accessibility relation R for the $\Box$ of (41) by:

(42) For any $w, w' \in W$, $w\ R\ w'$ iff $w' \in W'$.

(For the case we have been discussing, where W' is the set of continuations of w after t, this would make R into the relation $\leq_t$.)[35]

Conditionals whose logical form is that in (41) are often referred to as *strict conditionals*. Thus the claim that natural language conditionals can be represented in this from can be rephrased as "natural language conditionals are strcit conditionals". If this was all that needed to be said, then we could stop here: Conditionals in natural language are strict conditionals and the logical machinery to deal with those is already in place; end of story.

But *can* this be the end of the story? At the very least one would expect from such an account something about how the world set W' throughout which $A \to B$ must be true is determined. Or, to put it in slightly different terms: Suppose that C is the conditional which is to be analysed as a strict conditional $\Box_C(A \to B)$, where $\Box_C$ is the necessity operator that must hold throughout a world set W_C in order that C be true. How does the set W_C depend on C (by itself or in combination with certain contextual factors)? If the backtracking uses of counterfactual conditionals and the corresponding uses of non-counterfactuals in planning contexts were the only ones that we needed to worry about, it might perhaps be possible to come up with a systematic answer to this question. Consider again the case of a counterfactual conditional *if it had been the case that* A, *then it would have been the case that* B, where it is known that A is false. Often A will make it possible to determine roughly how far back in time we must go to reach a point t where A was still undecided. To the extent that t can be determined, that then also determines, we have seen, the set W_C (viz. as the set $\{w' \in W: w \leq_t w'\}$).

35 An account of the meaning of counterfactuals would have to include also the preecondtiions that A, and for certain counterfactuals also B, be false. There is general agreement that these conditions should be seen as *presuppositions* for the use of counterfactuals. The treatment of presuppositions is a story in its own right, which cannot be addressed here. A few words will be said about it in Part II.

For planning-related uses of non-counterfactuals an interpreter is likely to need more contextual information to determine W_C: He has to have an idea what the plan is in relation to which the conditional is being asserted and what assumptions are being made about the circumstances in which the plan is to be carried out. On the face of it it looks like more parameters are involved than in the typical backtracking cases, and often interpretation will be a matter of guess work. Still, in these cases too we can perceive some systematicity in the determination of W_C.

But these are only some of the uses of conditionals and in general the question how W_C can be determined seems much harder. The following two conditionals, thought up by Quine[36] for the very purpose of showing this, are a case in point.

(43) i. If Caesar had been made commander in chief in the Korean war, he would have used catapults.

ii. If Caesar had been made commander in chief in the Korean war, he would have used the atom bomb.

Both these conditionals seem reasonable, but each in its own way - or ‘in its own setting’, we should perhaps say. The first is naturally understood as saying something about what would have happened if someone with the military knowledge of the first century B.C. had been given the command over the American troops in a war during the nineteen fifties. Here the relevant worlds are those in which Caesar, while being given the supreme command in Korea, nevertheless operates with a knowledge of warfare as it was practiced in his own day. The second conditional is different. Here the relevant possible worlds are worlds in which Caesar has been informed about modern weaponry, while having retained the single-minded ruthlessness that many associate with him on the basis of what is known about him through his own writings and those of his contemporaries. Here it is certain character traits that are being kept constant, not military expertise. Apparently, then, the two conditionals in (38) require different sets W_C. Either conditional sounds reasonable and potentially true only if the right set W_C is assumed.

[36] Willard van Orman Quine (1908-2000), one of the leading analytical philosophers and philosophical logicians of the 20th Century.

The moral of such examples is double-edged. On the one hand they seem to show that a simple recipe for determining the sets W_C in terms of which they are to be evaluated is a remote possibility. But on the other they might also be seen as indicating that in actual practice that isn't as much of a problem as might have been expected. Many conditionals, those in (43) among them, will sound reasonable only on the assumption that they are evaluated in terms of certain sets W_C. A speaker who makes a reasonable use of such a conditional C must assume a context that specifies a set W_C on which C is reasonable. And an audience that takes a speaker to say reasonable things will, when it hears or reads C assume that the context assumed by the speaker is such a context; and if all goes well, it will arrive in this way at roughly the same set W_C that the speaker had in mind. (How interpreters manage to do this is is another matter; but somehow interpreters seem to manage quite well.) In this way conditionals will select their own contexts, as it were. When they are used the user is under a general obligation to assume a context in which they make sense and the interpreter can infer this context from the assumption that the speaker is making sense.

But that isn't quite right. Conditionals aren't that free in selecting their own contexts. Once a context has been settled on – in whatever way or by whatever means – then it cannot be changed at will. This appears to be a quite general principle, and it appears to be true in particular for the use of conditionals. We can appreciate this by seeing what happens when, say, two conditionals of the kind given in (43) are juxtaposed in one and the same context, as in (44).

(44) If Caesar had been made commander in chief in the Korean war, he would have used catapults. But if he had been made chief commander in the Vietnam war, he would have used the atom bomb.

You just cannot interpret this two sentence discourse as making the following conjunction of claims: (i) in the worlds in which Caesar was put in command in Korea and in which his knowledge of warfare was that of a Roman general in the 1st Century B.C., he deployed the catapult; and (ii) in the worlds in which he was put in command in Vietnam with his knowledge of weaponry updated to post World War II standards, he deployed the atom bomb. You cannot, in other words, evaluate the first conditonal on the assumption that its context is one where Caesar retains his 1st century B.C. knowledge of warfare and the second conditional on the assumption that its context is one in which the knowledge is not preserved.[37]

The difficulty one finds in making sense of (44) supports the more general observation that contexts cannot be changed without explicit indication that one is doing so. The speaker may have some freedom in choosing the context so that her claims are sensible and to expect her audience to follow her in this. But once the choice has been made, then she is bound by it; fickleness – flitting from one context to another at will – is something that the principles of conversation do not allow.

This means in particular that conditionals do not themselves have the power to select their own contexts. Their evaluation depends on the context insofar as it depends on the set W_C, and it is the context that provides that set. Moreover, conditionals may provide clues as to what the context is. But when two conditionals provide contradictory clues, then we are in trouble, for a single context cannot accommodate them both by providing two different sets W_C at once.

We can take the formal implications of these observations to be the following. Utterances are always made in a context and often the context is needed in some way or other to determine whether the utterance was true. In particular, a context c in which a conditional "*if* A, *then* B" is uttered will provide a world set W_c all worlds of which must verify the material conditional $A \rightarrow B$. We can model this with the formal tools available to us if we assume that each utterance context determines a model $\mathbb{M}_c = \langle W, R_c, F \rangle$, where R_c can be thought of as obtained from a world set W_c via (41), and where, conversely, W_c can be recovered as the field of R_c. In this way we can hold on to the proposal that natural language conditionals have the form of strict conditionals $\Box_c(A \rightarrow B)$, but whose necessity operator $\Box_c$ is context-dependent insofar as its evaluation depends on the contextually determined R_c (or, equivalently, on W_c).

---

[37] It seems that for me personally the most plausible interpretation of (43) that comes to mind involves alternative worlds in which Caesar is projected into the 20th Century to take over the supreme command – either in Korea or in Vietnam - while on the one hand still much given to the 1st Century B.C. style of campaigning that he was so good at in his own day, but with on the other hand some vague knowledge of the atom bomb as a newly invented "wonder weapon", that is to be used only as a last resort. Vietnam proved to be such a last resort case; Korea wan't a catastrophe of quite that magnitude. But that is presumably only one of the various fantasies which an interpreter of (43) may be encouraged to engage in when struggling to make sense out of this conjunction.

If we could stop here, that would leave us with an account of conditionals that may not seem very satisfactory, but that might arguably be the best we can hope for. We summarise the position we have reached once more in terms of the following three points:

(i) Natural language conditionals are strict conditionals;
(ii) The set of worlds throughout which the corresponding material conditional must be true varies from context to context, with in some cases a possibility of determining this set in terms of the conditional itself (such as in the backtracking case).
(iii) However, contexts cannot be changed at will, without explicit warning to the audience.

But this too is a position that has been challenged. The challenge comes from conjunctions of conditionals which suggest that the context permits some flexibility in the choice of the set W_c after all. (45) is one of the examples that are often used to argue this.

(45) If Fred were to come to the party, the party would be a success. But if had brought Susan along, the party would be a desaster.

It is plain that the proposition "the party is a success" and the proposition "the party is a desaster" contradict each other: no party can be a success and a desaster all at once. We take account of this, while at the same time simplifying things slightly and inessentially, by symbolising the first proposition as q and the second as ¬q. Further, note that the *if*-clause of the second conditional is naturally understood as elliptic for "Fred comes to the party and brings Mary". So, abbreviating "Fred comes to the party" as p and "Fred brings Mary to the party" as r, the full representation of the second *if*-clause becomes "p & r". If we now represent the two conditionals in (45) according to the "strict conditional recipe" in (41), we get as logical form for (45) the formula in (46). (We have added the sibscript $_c$ to indicate that the necessity operator is supposed to depend on the context c.)

(46) $\Box_c(p \rightarrow q) \,\&\, \Box_c((p \,\&\, r) \rightarrow \neg q)$

But given our current assumptions this cannot be right. For note, first, that if there is any point in mentioning the second conditional of (45), the speaker must assume that there are possible worlds in which Fred comes to the party and brings Mary. For otherwise the second conditional would be vacuous, and there is a general conversational prohibition against making vacuous statements. So we may assume that there are some worlds in the set W_c in which p and r are both true. But then (46) must be false. For on the one hand the first conjunct of (46) requires that in all p-worlds of W_c it is the case that q, while on the other hand its second conjunct requires that in all (p&r)-worlds q is false. But those worlds are *ipso facto* p-worlds, so they are worlds in whih q is both false and true. So (46) will be false in any context which satisfies the preconditions for its being used reasonably.

This blatantly contradicts our intuitions. Surely (46) can be used to make statements that are both reasonable and, to the best of our knowledge, true. To see what went wrong we must bring to the surface what this intuition rests on.

That isn't too difficult. Let us relct on what would be a natural context for uttering (45). In such a context the truth requirement imposed by its first condtional should come to something like this: Its consequent q (= the party was a success) should be true in the most salient (or most 'plausible') worlds in which its antecedent p (= Fred comes to the party) is true, but not necessarily in all such worlds that the context admits; and the implication is that those most salient worlds in which Fred comes to the party are the ones in which he comes by himself. (Perhaps these are the most salient worlds precisely because people who invite Fred to their parties don't encourage him to bring Susan, as she is such a notorious killjoy; so worlds in which he brings her are easily dismissed as unlikely or implausible, and thus as "non-salient"). On this assumption, then, the worlds in which Fred comes together with Susan will be ignored in the evaluation of the first conditional, which will qualify as true so long as the party is a success in the worlds in which Fred comes by himself. It is only when the antecedent of the second conditional forces us to face the non-salient possibility that Fred might turn up with Susan in his wake that it is no longer possible to ignore such worlds. And it is consistent with wat has been claimed up to that point (i.e. by the first conditional) that in those worlds the opposite is true of what holds of the more salient worlds in which Fred comes alone: In the salient worlds in which Fred comes to the party, i.e. in the worlds in which he comes on his own, we have q; but that is perfectly compatible with the assumption that in certain non-salient worlds in which he comes to the party, viz. those in which he brings Susan, q is false.

If this analysis is on the right track, then the context doesn't rigidly fix the set of worlds in terms of which conditionals are evaluated once and for all. Rather, the context presents a range of different worlds some of which are more plausible, or salient, than others. The antecedent of a conditional can then select from this range the most salient worlds among those in which it is true.

If we want to allow conditionals this much freedom to select, within the bounds set by the contet, the sets of worls in which the corresponding material conditionals are to be true, then we need an account of saliency:
What exactly is saliency, in the sense relevant here, and how could it be captured formally? One popular proposal, which goes back to David Lewis and Howard Sobel[38], is that saliency is a matter of how much the worlds in the given set - the world set W_c determined by the context c - resemble (in contextually relevant ways) the world in which the conditional is being evaluated. Lewis proposes to formalise this idea in terms of a 3-place relation $\leq$ between worlds. Predications involving this relation are usually represented in the form "$w_1 \leq_w w_2$", which is to be read as "w_1 is a world that is more similar to w than w_2 is". For given w the relation $\leq_w$ is assumed to be a weak linear ordering. Its field (here, since $\leq_w$ is assumed to be reflexive, this is simply the set $\{w': w' \leq_w w'\}$) corresponds roughly to the set W_c, which we assumed as part of the analysis of natural language conditionals as context-dependent strict conditionals.

To formalise this idea within the general settting of our modal logic-based analysis, we need, first of all, a more refined notion of model. The model $\mathbb{M}_c$ determined by a context c must now be assumed to provide us not just with a relation R_c, or with corresponding world set W_c, but with a pair $< W_c, \leq_w>$, consisting of a world set W_c and a 3-place relation $\leq_w$ on W_c of the kind just described. But that is not enough. At this point we can no longer assume that natural language conditionals have the logical form of strict conditionals. The reason is that the antecedent A of a conditional "*if* A, *then* B" must now be able to play the double role that we have already informally described: (i) the role of selecting from the set W_c the worlds in which the corresponding material conditional A $\rightarrow$ B is to be evaluated, and (ii)

38 The proposal to formalise natural langauge conditonals in this way was formulated independently and more or less simultaneously by David Lewis and Howard Sobel - at the time both assistant professors at UCLA - in the spring of 1967. The findings were reported in the same research seminar conducted by Montague, who summarised the results of that seminar in his paper "Pragmatics". (See Montague, 1974)

the role it plays in virtue of being the antecedent of that material conditional. Lewis and Sobel proposed to do justice to this double role of the antecedents of alethic conditionals by attributing them the same kind of logical form as that (A O⇒ B) which we introduced in the last section for deontic conditionals: alethic conditionals are represented with the help of a 2-place connective □⇒, whose first argument position covers both roles that the present analysis assigns to the antecedents of such conditionals. The double role of the antecedent A in formulas A □⇒ B is put in evidence by the truth clause for □⇒ proposed by these authors, a version of which is given in (47).

(47) $[A \Box\!\!\Rightarrow B]_{\mathbb{M},w} = 1$ iff (i) $(\forall w' \in W_c)\ [A]_{\mathbb{M},w'} = 0$

v

(ii) $(\exists w' \in W_c)([A]_{\mathbb{M},w'} = 1\ \&$

$(\forall w'' \in W_c)(w'' \leq_w w'\ \&$

$[A]_{\mathbb{M},w''} = 1 \rightarrow [B]_{\mathbb{M},w''} = 1))$

N.B. (47) articulates the truth conditions for A□⇒B in the form of a disjunction. The first disjunct has been added to deal with a kind of degenerate case, which arises when there is no relevant world in which the antecedent A is true; how we handle this case is somewhat arbitrary, but we get a nicer logic if we assume that the conditional is true under those conditions, and that is the effect that the first disjunct of (47) achieves. But it is the second disjunct that really matters. It says that when there are relevant worlds in which A is true, then there is a relevant A-world w' so that all A-worlds at least as close to w as w' are also B-worlds.

The route by which we have arrived at this final proposal, according to which alethic conditionals are to represented with the help of the operator □⇒, has been long and circuitous. If all this effort has been made just in order to present one further example of a 2-place modal operator, was that really worth it? Probably not. But just adding annother 2-place operator to our inventory wasn't the only point of this excursion. Conditionals are a topic of prime importance in their own right, and they will play a crucial role in Part II. When we return to the topic there, that which has been said in this section will serve us as a useful starting platform.

It might be added that there are few topics in the theory of language and logic that have had as much attention, and from as great a diversity of different perspectives as conditionals. On the one hand this is because conditionals play a central role in the theory of inference, and at the same time a role that is highly controversial. Those who investigate conditionals from this angle aren't even agreed on what sorts of things they really are: Some think of them as statements, which have a propositional content in virtue of which they are either true or false (much as we have assumed in what has been said in this section). Others think of them instead as inference rules, which may be used to deduce conclusions from premises in languages that themselves may lack conditionals among their grammatical constructs.

These concerns can be distinguished from the principal interest that the topic of conditionals holds for linguists. Linguistics is concerned with the form and meaning of conditionals in natural languages. One part of the problem here is to determine the meanings of conditionals expressed by different linguistic forms: what the truth conditionals of those condtionals are, and perhaps also, what the logic is that those truth conditions define. In this section we have almost exclusively been concerned with this second problem. But the first problem, what linguistic forms are used in different languages to express conditionals, is non-trivial too, and more interesting than might have been thought. For instance, contrary to what one could easily be inclined to think, the presence of an *if*-clause in a sentence is by itself no guarantee that the sentence expresses a conditional, in which the *if*-clause plays the part of antecedent. Here is an example of a sentence for which this is not so: "Sometimes, if a farmer buys a donkey, he pays for it in cash." This sentence just says that there are occasions when a farmer buys a donkey and pays for it in cash. (Lewis, 1979). In no logical representation of this sentence is there any use for the material conditional $\rightarrow$, the strict conditional $\Box(..\rightarrow..)$ or even, as we have defined its semantics in (47), for the 2-place operator $\Box\!\!\Rightarrow$. On the other hand, conditional operators are often needed in the representaetion of natural language sentences whose conditionality is not overtly manifest through the presence of particles such as *if* or *when*, which are good clues to conditionality (even if, as we have just seen, they do not do so invariably).