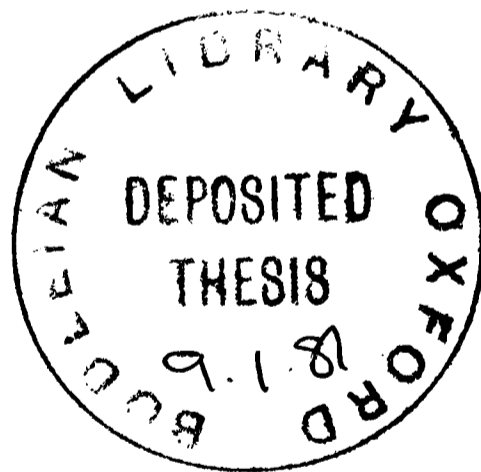


PARTIAL-VALUED LOGIC

STEPHEN RICHARD BLAMEY

St. Edmund Hall



This thesis is submitted in partial fulfilment
of the requirements for the degree of Doctor
of Philosophy in the University of Oxford.

Trinity, 1980^{MT}

CONTENTS

	<u>Page</u>
<u>GENERAL INTRODUCTION</u>	1
<u>CHAPTER I : PARTIAL-VALUED LANGUAGES (1) – SENTENCE COMPOSITION</u>	8
I.1 : INTRODUCTION	8
I.2 : TRUTH-FUNCTIONALITY	14
I.3 : EXHIBITING TRUTH-VALUE PRECONDITIONS	28
I.4 : LOGICALLY CONVEX SETS	47
<u>CHAPTER II : PARTIAL-VALUED LANGUAGES (2) – TERMS AND QUANTIFIERS</u>	57
II.1 : EXTENSIONALITY	57
II.2 : LANGUAGES AND MODELS	75
II.3 : DEFINITE DESCRIPTIONS	94
<u>CHAPTER III : PARTIAL-VALUED LOGIC (1) – SENTENCE COMPOSITION</u>	108
III.1 : INTRODUCTION	108
III.2 : THEORIES AND MODELS – PRELIMINARIES	115
III.3 : THEORIES AND SENTENCES	125
III.4 : THEORIES AND SEQUENTS	139
III.5 : LOGICALLY CONVEX SETS – THE REPRESENTATION THEOREM	143
<u>CHAPTER IV : PARTIAL-VALUED LOGIC (2) – TERMS AND QUANTIFIERS</u>	148
IV.1 : THEORIES AND MODELS – PRELIMINARIES	148
IV.2 : THEORIES AND FORMULAE	160
IV.3 : THEORIES AND SEQUENTS	184
IV.4 : COMPLETE THEORIES	196
<u>CHAPTER V : PRESUPPOSITION</u>	206
V.1 : INTRODUCTION : SEMANTICS AND LINGUISTIC PRACTICE	206
V.2 : PRESUPPOSITIONAL QUANTIFIERS	226
V.3 : THE 'PROJECTION PROBLEM'	236
V.4 : PRESUPPOSITION AND ASSERTION	248
V.5 : ANOTHER PRESUPPOSITIONAL QUANTIFIER	276
V.6 : ASSERTION AND PRESUPPOSITION	281
<u>REFERENCES</u>	296

for "time" and "times"

ACKNOWLEDGEMENTS

This thesis is a (very remote) descendent of my B. Phil. thesis, and I should like to thank Dana Scott, my supervisor throughout, for his continued help and guidance — not least for those mantic utterances whose significance I did not appreciate until months later. I am no less grateful for the tolerance and kindness he showed to such a recalcitrantly slow and disorganized pupil.

I have, of course, talked about my ideas with many people, and this has been useful. I wish I could say that I had substantially profitted from these conversations, but I fear it is not so. The fault is largely mine: I do not seem to have been very good at discussing my own work, and it has certainly suffered greatly as a result. However, those who made a valliant effort include (in rough chronological order of the first extended conversation): Paul Seymour, John Burgess, Martin Davies, Karen Green, Andrew Pigdon, Justin Gosling, Göran Sundholm, Jeremy Stone, Jan Crossthwaite, Matthias Schirn, Penelope Mackie, Marco Santambroggio, and David Holloway.

I feel I must also acknowledge a subtle influence from various seminars of Gareth Evans over the last few years. Now and again I fancy myself thinking about something in a way I picked up from him. However, inspiration from this source has been totally distorted, I suspect, by misunderstanding.

I wish I had had the opportunity of talking to Julianne Jack before writing Chapter V: unfortunately this did not happen until afterwards. However, I took a masochistic delight in being bullied into a coherence which I recognize is sadly lacking in the text here (not only in Chapter V).

Material comfort and intellectual stimulation of a more general kind has recently been provided by the Principal and Fellows of St. Edmund Hall, to whom I am most warmly grateful. I have been working in very congenial surroundings — though at time rather distracting ones.

The following pages are more of a pleasure to look at than to think about, and this is due to the exceptional skill and experience of my typist Ina Godwin. Indeed, she often took on the role of an editor, and, for the most part, this has been thoroughly beneficial. But at times I was pigheaded: I hope I have been forgiven for demanding that some of what I originally wrote be restored.

Finally, I should like to thank my parents for helping me to proof-read, and for making sure that I woke up in time when I was returning to Oxford.

WARNINGS

(1) The result of a complicated story is that there exists a distinction in style of quotation marks, but this is not one which corresponds to any distinction in significance which the logical eye might be expected to be attuned to. For there are single quotes throughout except in the case of complete English sentences (whatever is being done to or with them).

(2) The notation '□' and '□' is introduced initially for relations between truth-value classifications, but then it is extended to hold between many other kinds of thing (in particular models and formulae), with or without some kind of subscript for special uses. It seemed to me that this 'systematic ambiguity' helped rather than hindered understanding: I hope I was right. But there is also a non-systematic ambiguity, for we use '□' also as an end-of-proof symbol.

(3) Underlining is used not only as a punctuation device to signify 'emphasis', but also to introduce stipulative definitions in the more formally systematic sections.

PARTIAL-VALUED LOGIC

STEPHEN RICHARD BLAMEY — St. Edmund Hall

A thesis submitted for the degree of Doctor of Philosophy
in the University of Oxford, Trinity term 1980

ABSTRACT

PART I (Partial-valued Languages): In Chapter I we consider modes of sentence composition and ask what 'truth-functionality' is — functionality in two values, when there is also a third classification for meaningful sentences. An answer to this question might, we suggest, be seen as spelling out the very idea of the third classification as lack-of-a-value rather than a third value. We go on to ask how fully to exploit partial-valued semantics and motivate the need for modes of composition which themselves actually introduce non-trivial truth-value preconditions. There are two particularly interesting connectives which provide this expressive resourcefulness. The relation between partial- and total-valued languages is then considered. In Chapter II we consider sub-sentential modes of composition, and extensionality in general. Hence we come to see the shape of a partial-valued semantics which admits of undefinedness in a uniform way in all categories. But we restrict special attention to a simple kind of first-order language with terms and quantifiers and a discrete and determinate identity relation. There is, finally, a section devoted to definite descriptions.

PART II (Logic): In Chapter III we set up laws for sentential composition, and in Chapter IV for the simple languages mentioned above. A theory is defined to be a (consequence) relation satisfying these laws, and we investigate the connection between theories and their models. Thus we are able to answer (some) questions about the logic: familiar questions and also ones peculiar to our partial-valued framework.

PART III: In Chapter V we suggest that the features of our logic make it apt for deployment in a natural-language semantics which not only accommodates but actually gives a proper systematic treatment to 'presupposition'. The emphasis is on exhibiting presuppositions (as truth-value preconditions). We attempt, furthermore, to outline an account of linguistic practice which meshes with the semantic details in an illuminating way. 'Presupposing' is taken to be constitutive of assertion making, and what is presupposed constitutive of what is asserted.

GENERAL INTRODUCTION

My concern is to describe, investigate and apply a 'partial-valued' logic which admits of undefinedness in a uniform way in all semantical categories: in particular, sentences may be neither true nor false and singular terms may be without denotation.

Chapters I and II deal with the interpretation of modes of composition in partial-valued semantics; while in Chapters III and IV logic is set up for some simple kinds of language, whose precise expressive range has been motivated, and is investigated model-theoretically. In Chapter V we turn to the (familiar) idea of using partial-valued logic to cope with natural language 'presupposition' and urge that some of the (unfamiliar) features of our logic make this programme more workable and attractive than it might otherwise appear. The general point here is that we should not get bogged down over the relation of presupposition, but find a way systematically to exhibit presuppositions.

Presupposition is the only application we give any detailed consideration to, but it is enough to motivate the more rarefied themes of the first two chapters, and it can be illustrated well enough using the languages dealt with in Chapters III and IV. For it is not any philosophical perspective at odds with the realism of classical semantics that lies behind my advocacy of partial-valued logic. Rather, it is simply a dissatisfaction with such logically conservative manoeuvres as either dismissing from the scope of logic sentences with apparently no intrinsic defect, or 'idealizing' the meaning of sentences — often to the point of utter distortion — so that they fit the structure of a total-valued logic. The idea is not just to be able to 'accommodate' more, but also to provide a richer logical analysis: I wish to argue that going partial is a way in which greater subtlety of meaning may be revealed and systematically handled. This can

be done, I believe, without upsetting what is so simple and compelling about classical semantics.

Such were my original motivating thoughts. Subsequently, however, when the details had already taken shape, it became clear to me that if I were justified in taking seriously the terminology 'partial-valued logic' and 'truth-value gap', then I had a dispute on my hands with Dummett's forcefully argued views concerning truth values and, indeed, concerning also the interpretation of singular-terms: see Dummett (1973), e.g. Chapters 10 and 12. Though neither the formal details of the logic nor their proposed deployment in a natural language semantics would, I think, meet with any special objection from him, yet the understanding of those details and the envisaged point of their deployment undoubtedly would. There are, perhaps, two distinct points of disagreement, which, to avoid possible subsequent confusion, I shall specify and contrast immediately.

Dummett argues that we should appreciate the distinction between truth-values as evaluations of assertions, on the one hand, and as 'semantic roles' of sentences, on the other. The evaluation of an assertion is something independent of the structure of the sentence used to make the assertion; while a system of semantic roles is discerned precisely in order to take account of structure. Indeed, any linguistic item has a semantic role, that is to say the semantic contribution it makes to more complex items in which it may occur as a constituent. Within a systematic semantics the semantic roles of constituents determine the semantic role of a complex item (which might itself be further embedded) but the ultimate point of a system of semantic roles is to determine correct truth values in the first sense, viz. evaluations of assertions made using sentences. And so truth-values in the sense of semantic roles of sentences must correlate with truth-values as evaluations of assertions.

This general picture, which I have stated only very crudely, is not itself something I want to dispute. I have, of course, pushed into the background the question of the exact nature of the explanatory link-up between systematic semantics and linguistic practice, which this correlation of the two kinds of truth-values is supposed to play a part in. Nor have I stopped to question whether it is the linguistic act of assertion — or that linguistic act alone — which is to be taken as the basic one for the purposes of this explanatory link-up; nor, indeed, whether it is after all only complete sentential units of language *via* which systematic semantics and linguistic practice are to be linked.

But concerning the (structure-independent) evaluation of assertions, Dummett holds that there can be no rationale for anything but an exhaustive dichotomy into the TRUE ones and the FALSE ones — providing, at least, that these notions are objective. The important point is that there can, on his view, be no circumstances recognition of which would prompt us to say that an assertion was neither TRUE nor FALSE — unless perhaps this was a way of pointing out that the content of the assertion had not been fully specified. This is the first thesis of Dummett's that I would disagree with. It is discussed in Chapter V.

We can, however, proceed to formulate and discuss the application of our semantics before this issue is examined, at least as far as Dummett is concerned, since he readily entertains the possibility of a subtler system of truth-values as semantic roles: in particular a three-fold system of the kind partial-valued logic requires: 'true', 'false' and 'neither-true-nor-false'. This would be motivated for him by the need to get right the account of the systematic contribution of parts to whole — but only by that. The most obvious example of this need, and the one Dummett considers, arises from negation: if the natural negation construction of a language such as English is to be considered a genuine sentence-out-of-sentence functor,

for "on" and "in"

then a matrix

$$\frac{p/t \ n \ f}{\neg p/f \ n \ t}$$

is required to take account of how negation is understood when applied to certain sentences, such as ones containing definite descriptions (Dummett's standard example). In this case the matrix-entry t is correlated with TRUTH, and both f and n with FALSEHOOD: the three-fold system is invoked simply to make negation truth-functional, as it would not be on TRUE and FALSE.

In contrast to this, I wish to urge a rationale for the deployment of a three-fold evaluation scheme for assertions which correlates one-one with the semantical scheme. I seek a rationale which does not simply appeal to the semantic need for this, but one which might, conversely, help to explain why such a semantics is the one to adopt. The content of an assertion is seen as determined along with its evaluation conditions, and the 'semantic content', we might say, of a sentence along with its truth-value(= semantic role) conditions. So the dispute can be seen as whether or not it is right to discriminate between meanings in the sense of assertion content with the same subtlety that we discriminate between meanings in the sense of semantic content. The problem emerges in this form because, like Dummett, but unlike many gappy-logicians, I can see no attraction at all in the idea that semantic neither-true-nor-false conditions for sentences correspond to conditions under which no assertion can be made using those sentences.

I might add that the kinds of idiom that we shall be interested in in Chapter V, go far beyond sentences containing definite descriptions and their negations. The need for a presuppositional analysis has indeed been claimed for a great many kinds of sentence — especially in the linguistics literature — and we shall consider some of these. But we shall also

consider a class of idioms which have not, as far as I know, been considered presuppositional at all. These, I shall argue, in fact constitute something of a paradigm of the phenomenon. Indeed, the semantical importance of giving their semantic content in terms of a three-fold classification is not just that we can thereby explain how they are compounded with the familiar operators to form more complex sentences, but that we can explain how they themselves are compounded out of their own constituents — in unfamiliar ways. This points up the need for a similar explanation for all presuppositional idioms.

So much for assertion: the second disagreement with Dummett concerns truth-values as they are deployed in a systematic semantics. According to Dummett, whatever the system of truth-values as semantic roles (viz. matrix entires for modes of sentence composition), they are indistinguishable in semantical status, since we have no other conception of them but as semantic roles. In particular the gloss 'neither true nor false' cannot be taken to mean lack of either of the only two values but just means a third value. He would say that the temptation to think otherwise derives from a spurious Fregean analogy between sentences and names (i.e. singular terms generally):

sentence :: truth-value
name :: bearer.

According to Frege's unified theory of reference, truth-values are the references of sentences, just as bearers are the references of names. However, to get straight on the matter, Dummett again wields the notion of semantic role to draw a contrast — this time concerning names. We must, he says, prise apart Frege's notion of reference into

- (i) semantic role, and
- (ii) referent (bearer):

a name may lack a referent but must, if it makes any meaningful semantical

contribution at all, always have a semantic role — even if this is just lacking-any-referent; hence the two cannot be identified. The mistake arising from the indiscriminating notion of reference together with the assimilation of the categories of truth-value and name-bearer, is the idea that truth-values are like referents and so can fail to be present for a meaningful sentence.

Now, I do not wish to claim that the distinction between semantic roles and referents is misguided: on the contrary, I wish first to urge a further distinction: the notion of semantic role splits into two distinct ones, 'semantic classification' and 'semantic value'. Roughly, semantic classifications are to be extensionally speaking the same as Dummett's semantic roles (they comprise the necessary distinctions required of an adequate systematic semantics), whereas semantic values are what capture the idea of semantic contribution involved in the notion of semantic role. The semantic values of names and sentences may, if you like, always be taken to be semantic classifications, but there may be circumstances under which it is appropriate to say that only a proper subclass of semantic classifications are semantic values. In particular, I shall argue that the three-fold classification of sentences as 'true', 'false' and 'neither-true-nor-false' may be such that 'true' and 'false' are semantic values, while 'neither-true-nor-false' need not be taken as such, and that the classification of names as 'denoting-so-and-so' is a semantic value, while 'not-denoting-anything' need not be. Hence — in such circumstances — the actual identification of semantic values and referents (viz. bearers of names and, by analogy, reified truth-values) would after all be harmless and, indeed, very natural.

But, what are the circumstances under which it would be appropriate to say these things? It is, I think, a matter of the expressive range of the language in question; at bottom, the range of sentential, predicate,

and term-out-of-term(s) modes of composition. Of a given mode of one of these kinds, whose interpretation has been described by a function of the appropriate kind between semantic classifications, we can bring to bear a criterion of functional dependence to determine whether or not that mode of composition can in fact be fully specified as exhibiting dependence simply between the semantic classifications other than 'neither-true-nor-false' or 'not-denoting-anything'. (This is the subject of Sections I.2 and II.1.) And, if it can, then the interpretation of the mode can be seen equally well as a partial function (of the appropriate kind according to category) between just these classifications as a total function between all possible classifications. But then, if all the modes of a language can be seen in this way, the classifications 'neither-true-nor-false' and 'not-denoting-anything' are not needed at all as values between which the language exhibits semantical dependencies: they may be seen as indicating a lack of any semantic contribution.

According to this perspective, then, what makes a semantics partial valued is not some prior stipulation concerning the nature of the semantic classifications of names and sentences, but the fact that modes of composition can be seen as being interpreted by partial functions. To extend the classification/value terminology to modes of composition, we might say that the semantic classification of a mode is a total function between classifications, and its semantic value is the partial function between semantic values. Of course, such a semantics could still be seen as giving total-valued interpretations to the modes of composition, thereby forcing all classifications to be values; however, it need not be, and a general principal of semantical economy might dictate that where we need not count a classification as a value we should not — there would be no point. Furthermore, since our criterion for when a mode of composition can be seen as interpreted partially is non-trivial, we would have what was from the total-valued point of view an expressively highly incomplete language.

CHAPTER I

PARTIAL-VALUED LANGUAGES (1) - SENTENCE COMPOSITION

I.1 INTRODUCTION

We begin with propositional logic — with a consideration of partial-valued semantics as it affects sentences and modes of sentential composition. A lot of what comes up in this connection will serve as a paradigm for more general considerations: however, until Chapter II let us forget about the obligation to treat systematically of structure other than sentential structure. Life will appear rather more trivial than it really is until we do dig deeper, but this temporary self denial should prove therapeutic: it will direct attention towards important things about the sentential articulation of partial-valued languages, which could so easily get overlooked otherwise.

At sentence level, then, being partial-valued should mean that our semantics classifies items into three categories: the true ones, the false ones, and those that do not have a truth-value. (For this reason I eschew the term 'non-bivalent'). The idea is by no means novel, but there is a lot of confusion about it, and a lot of objection to it:

"If one is going to use undefined terms, why not undefined truth-values? Isn't that more natural? May be so, but I have yet to see a really workable three-valued logic. I know it can be defined, and at least four times a year someone comes up with the idea anew, but it has not really been developed to the point where one could say it is pleasant to work with. May be the day will come, but I have yet to be convinced."

(Dana Scott [1970])

Sometimes classifying sentences into true, false and gappy is identified with straight forward 3-valued logic, items in the neither-true-nor-false category being thought of as assigned a third semantic value on

a par with true and false; sometimes the category of gappies is taken to contain items that are meaningless or nonsignificant in some way or other; sometimes both identifications are made at once. On the other hand there are people who want to avoid either identification. Consult Haack (1974) for an interesting catalogue of various logical views of this kind, along with an interesting and often unsympathetic appraisal of them.

As far as I can see, I am one of the last mentioned: the logic I go on to develop I should like to see as two-valued, but partial-valued: sentential items may be undefined — a semantical status essentially distinct in character from having one of the positive truth-values true or false; yet no item within the scope of the semantics is to be meaningless, and each item is as meaningful as any other.

Of course there is ground-work to be done to try and see what sense there is in any claim that a particular triclassificatory logic is of one sort rather than another. My hope is that, as we pursue the question, possibilities and expectations concerning partial-valued semantics will take shape, and determine the character of an interesting and useful logic. I think it is fair to say that a lot of the trouble and confusion over truth-value gaps is due to the fact that even when the identification of the gap with a third value or with meaninglessness is not made, or even explicitly rejected, too strong an analogy with either or both of these kinds of triclassificatory logic nonetheless tends to dominate thinking.

Without going into questions of the envisaged linguistic practice to be analysed by a logic, it is not really clear, of course, what designating an item meaningless might be supposed to come to; and in any case it would seem to me that the more systematically the third category were integrated into the articulation of a semantics, the less happily chosen the term 'meaningless' would be. Nonetheless, for our present purposes it

will be sufficient to take it to be a consequence of this kind of way of thinking about the third classification that an item so classified is semantically deficient, or defective, and so, perhaps, can have no 'propositional content'. Typically a 'meaningless' item would, therefore infect any more complex item of which it was a constituent. What having a third value might mean is equally obscure — perhaps more so, though at least at the level of 'abstract semantics' it is clear enough what the idea is.

It is at this level that I propose to address myself to the matter in this chapter: in a style as far as possible divorced from any particular philosophical scenario or programme of application. It might appear that working in such rarefied abstraction is precisely what causes muddle: but, if we are subtle enough, I think considerations may be brought to bear to enable us to get a grip on the idea of a two-valued partial-valued semantics, as distinct from other triclassificatory ones.

To start with there are two kinds of questions we can ask:

- (1) What further characteristics of a logic might give a point to the claim that its canonical understanding was a partial-valued one?
- (2) What special features should we demand of partial-valued logic which would exploit its partial-valued character to the full?

Sections I.2 and I.3 are devoted to following through answers to these questions. In I.2 I propose a criterion for what it is to have a truth-functional partial-valued logic. The point is: what is functionality in the two values, true and false, when the semantics allows also for sentences that are undefined? This question, unlike the question in vacuo 'what is a genuinely gappy logic?', makes good sense, it seems to me: we have an interesting semantical property that can be brought to bear to fill out the bald question and would, indeed, give a point to saying that a logic were partial valued.

As far as I am aware, this issue has not been considered with the subtlety it deserves: the question of truth-functionality has either been ignored, or taken for granted, or given an immediate and trivializing answer. This, I conjecture, is an instance of the vicious influence from the two alternative paradigms of triclassificatory semantics. If too close an analogy is made with straight-forward 3-valued logic, then the question just does not arise: functionality in terms of the 3-fold classification is all that a theorist would bother to think about. On the other hand, if the third category is thought of as containing meaningless items, then a trivial answer is likely to be taken for granted: truth-functionality can just mean what it means in the total two-valued case, with a proviso not to take third-category items into consideration. Truth-functionality, of course, is just a special case of extensionality, and in Chapter II we generalize the question to cover extensionality for all semantic categories in partial-valued logic. The ideas deployed in Section II.1 will simply be a generalization of those in Section I.2.

If we adopt truth-functionality (as conceived of in I.2), then we have a constraining principle governing the expressive power of partial-valued languages. Against this, in I.3 I follow up the second motivating question, making it more precise by demanding special expressive richness: modes of sentence composition peculiar to partial-valued semantics, which yield something more useful than just a patchy replica of boolean logic. This is a liberating principle and what I find impressive about it is that it dovetails perfectly with the truth-functionality constraint: novel sentence connectives, whose introduction is independently motivated, are precisely what we require to add to the familiar logical vocabulary (suitably interpreted) to attain languages expressively complete for partial-truth-functional modes of composition.

This question of expressive richness – expressive richness at sentence level – is another issue which is generally not considered very seriously. An amazing conservatism seems to hamper investigations: the third category is rather grudgingly admitted, and so the starting point usually seems to be: "Sentences can, for one reason or another, be undefined: so we will have to loosen up logic to accommodate the fact – i.e. suitably reinterpret the existing logical vocabulary." And then, may be, a further thought: "Look, it is less trivial if we throw in some operators talking about the possibility of gaps, e.g., 'it is true that ... '." My approach is rather: "Let us investigate a logic whose modes of composition are from the very start conceived of as partial valued, and in particular whose logical vocabulary contains items which actually exhibit ways in which undefinedness can arise in sentences, in virtue of their logical structure." This should be the exciting thing to do: we should then have the resources for deeper semantical analysis, and might go on to provide non-trivial laws governing the occurrence of undefinedness. And the fact is that operators talking about gaps anyway get rejected by our truth-functionality criterion, while the operators that exhibit gap conditions are called in to achieve expressive completeness.

Again I suspect that lack of interest in this matter is the result of thinking by analogy with the rejected paradigms. If the third classification is glossed as 'meaningless', then the question of such novel connectives does not arise: who wants logical vocabulary specifically interpreted so as to introduce meaningless sentences? While if you think in terms of three values on a par, then the issue is easily overlooked: once there is a stock of connectives available which is known to be expressively complete for three-valued truth-functionality, there is no apparent need to think anymore about modes of composition.

There is, of course, a famous approach to triclassificatory semantics which I have failed so far to mention; one, furthermore, whose proponents claim that it yields the true gappy generalization of classical logic, viz. van Fraassen's supervaluational techniques. This kind of semantics does indeed eschew the idea of a third value or a category for meaningless sentences, nor is it at all plausible to see any influence from these paradigms. Nonetheless the claim that what we get really is a gappy logic has come in for some criticism — see, for example, Haack (1974), p.58.

The idea, as far as I understand the matter, is that the third classification is explicated as 'don't-care-which-truth-value': supervaluations are partial valuations which correspond to sets of ordinary total valuations in such a way that a sentential item is settled as gappy with respect to a supervaluation iff it is not the case that the corresponding set of total valuations agree on that item. The vast literature on the subject reveals that this idea is powerful and fruitful, but I shall proceed with my altogether more naive approach. The difference in perspective can already be seen clearly enough: looking for novel richness of vocabulary is not an interest that the supervaluational idea encourages, since the basic evaluation of a sentence goes on in terms of total two-valued semantics; and the way supervaluations are defined creates a kind of modal logic, which means that the question of partial-truth-functionality is of little interest: the logic is in any case non-extensional in the most obvious way, viz. 'intensional'.

In Section I.4 I consider the relation between (my proposed brand of) partial-valued semantics and classical total-valued semantics, and here the difference between me and van Fraassen is neatly pointed up. For him a sentence under a partial valuation corresponds to that same sentence's being evaluated with respect to a set of total-valued interpretations. For me a sentence in a partial-valued language is seen as corresponding to a

(special kind of) set of sentences from a total-valued language and a single total-valued interpretation induces a partial-valued one.

We shall find that systems of such sets of total-valued sentences can themselves be seen as playing the role of partial-valued languages and constitute a kind of algebraic analogue of partial-valued logic. They provide an interesting 'non-standard semantics', and at the same time can be considered as paradigm instances of the logic itself, since not only do such systems obey the laws of partial-valued logic but we shall have a representation theorem showing that any partial-valued system can be embedded in a structure preserving way into a system of such sets. The full details of all this are not described until Chapter III. However, partial-valued semantics can proceed without this apparatus: the connection with total valued procedures need not play the central role that it does in the case of supervaluations.

Considering languages and ways of interpreting them is, of course, only half the story: to do logic we must have a framework in which to state logical laws and in which to develop theories in these languages. Discussion of this is postponed until Chapter III.

I.2 TRUTH-FUNCTIONALITY

The question I want to ask in this opening section is:

- (1) What is it for a (triclassificatory) semantics to be partial-two-valued and truth-functional in these two values?

It was suggested that a convincing answer to this question should, if nothing else, enable us to get some grip on the idea of a partial-valued propositional language and help to guide the development of logic for such languages, without our having to rely at all on drawing analogies with either total 3-valued logic or with logic that admitted 'meaningless' items.

We should not think of (1) as containing two separate questions: (1a) "What is it to be a partial-two-valued language?"; and (1b) "What is it for such a language to be truth-functional in these two values?" since (1a) is, on its own, pretty inscrutable: we are at a loss what considerations to bring along in vacuo to make sense of distinguishing a semantics which discerns two values and a gap from any other kind of tri-classificatory semantics. It is by asking about truth-functionality in the same breath that we get an interesting and answerable question.

But can we, on the other hand, ask about truth-functionality with respect to two out of a total of three semantic classifications, independently of any explicit mention of any intended construal of the third category? viz.

- (2) What it is to be a triclassificatory semantics which is functional in a particular two of these classifications?

In contrast to (1a), I think this makes perfectly clear sense standing on its own: it is in fact the question that I shall be probing in this section. For (2), I believe, comes to much the same as the double-barrelled question (1) — at least, an answer to (2) is all we need as an answer to (1). The basic question is: what is it for a given mode of composition to be functional in two out of a possible three classifications? But then if we have a language all of whose modes of sentence composition meet some non-trivial constraint offered in answer to this question, it will be natural to think of items of the language assigned to the third category as thereby valueless — as we argued in the General Introduction.

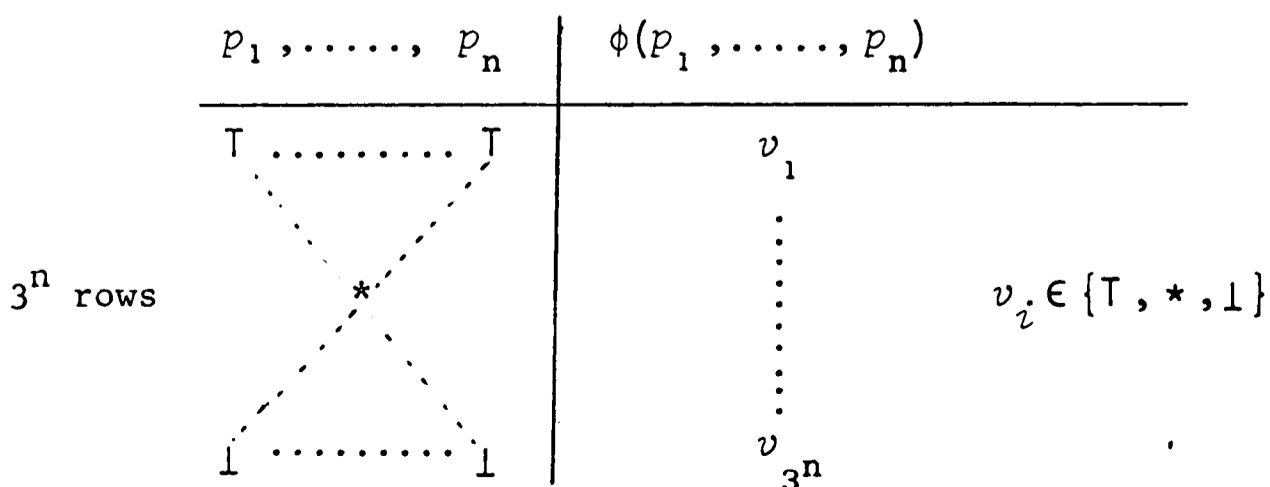
Hence, if question (1) is too mysterious for you as it stands, then I offer (2) as an immediate 'explication' of (1); the class of semantic values proper can just be that subclass of semantic classifications between the members of which a semantics is to discern functional dependencies. If, on the other hand, you think you have some sort of independent

intuition about what partial-valued propositional logic should be like – thinking, perhaps, by Fregean analogy (sentences: truth-values :: singular terms: objects denoted), then you do not have to agree right now to the assimilation of (1) and (2). The train of thought to be presented can be regarded as applying either directly to (1) or (2). And, in any case, if you follow it through in the second way, as applying directly to (2), it will become clear, I hope, that what we obtain is a semantical apparatus whose mechanisms are manifestly apt to have, as their canonical understanding, the status of representing a system of partial-valued semantics which meets any naive intuitions.

We must ask, then: what, in the context of a triclassificatory semantics, is it for a mode of sentence composition $\phi(p_1, \dots, p_n)$ to be value-functional in 'true' and 'false', where these are a particular two of the classifications?

We begin by approaching the question heuristically.

In the first place it would seem natural to exclude 'intensional' contexts by stipulating that the classification of a sentence $\phi(s_1, \dots, s_n)$ must not depend on anything over and above the 3-fold classification of the component sentences s_i . (I use p_i schematically and s_i for sentences with a fixed meaning – and so a determined classification.) Hence it would be a necessary condition for the two-valued-functionality of $\phi(p_1, \dots, p_n)$ that its semantical role as a functor – whether complex or not – can be represented by a 3-entry matrix:



Of course such a matrix directly describes a function $f_\phi: \{\top, *, \perp\}^n \rightarrow \{\top, *, \perp\}$.

If we accept this, then we have at our disposal the apparatus of total-valued, 3-valued, truth-functional semantics. This will be handy and it does not commit us to anything: we must expect our criterion of two-valued-functionality to yield a constraint on such 3-valued functions determining what subset of them may be taken to describe interpretations for modes of composition functional in 'true' and 'false'. We are hoping, furthermore, that the (3-valued) functions in this favoured subset are going to be apt as representations of partial functions $\{\top, \perp\}^n \rightarrow \{\top, \perp\}$ — functions partial, of course, not only in that they may be undefined for a given input, but also accommodating inputs where one or more (possibly all) argument places are 'empty'. For such must be the interpretation of functors in partial-valued (extensional) semantics. In anticipation I shall allow myself to talk of \top and \perp as positive values and $*$ as lack of a value, though for the moment this need be taken as no more than a façon de parler.

We have, then, a likely necessary condition for the two-valued-functionality of $\phi(p_1, \dots, p_n)$. But it could surely only be drawing too close an analogy with ordinary 3-value logic that would prompt us to say that it were also a sufficient condition: we would then be treating $*$ as a semantical value on a par with \top and \perp . But if it is not sufficient, then what kind of operator might we expect, intuitively, to be excluded?

There are plenty of authors who think of $*$ as having a semantical status different from \top and \perp (e.g. Woodruff (1970) — as a truth-value gap; or Hallden (1949) — as meaninglessness and a gap) who nonetheless include among their logical operators functor analogues of various metalinguistic notions. For example, we have $T(p)$ meaning 'it is true that p ', and $p \Rightarrow q$ meaning 'that p presupposes that q ', with the matrices

	$T(p)$	$p \quad q$	$p \Rightarrow q$
\top	\top	$\top \quad \top$	\top
*	\perp	$\top \quad *$	\perp
\perp	\perp	$\top \quad \perp$	\perp
		$* \quad \top$	\top
		$* \quad *$	\top
		$* \quad \perp$	\top
		$\perp \quad \top$	\top
		$\perp \quad *$	\perp
		$\perp \quad \perp$	\perp

Because of their metalinguistic character, such operators appear likely candidates for exclusion from the class of two-valued-functional ones.

Woodruff comments (page 135) that that 'distinctive feature' of these connectives is that they yield compounds which are defined under circumstances where every constituent is undefined (because of $T(p)$ at line 2 and $p \Rightarrow q$ at line 5), and hence they cannot be functional in \top and \perp . I disagree with this diagnosis: $T(p)$ and $p \Rightarrow q$ will in fact turn out to be not two-valued-functional by my criterion, but I think the argument just given is a bad one — especially so when seen as distinct from a more thorough-going view, of which it is just a half-hearted version. (I shall justify this remark later).

The thorough-going view, which we might label 'crude-truth-functionality', is that if any component of a complex item lacks a value, then (even if there are others around which do not) this complex item must lack a value too. So, for example, the crude truth-functionalist would have to reject 'Kleene's strong tables' for \wedge and \vee , and replace them with 'Bochvar's weak tables':

$p \quad q$	strong		weak	
	$p \wedge q$	$p \vee q$	$p \wedge q$	$p \vee q$
$\top \top$	\top	\top	\top	\top
$\top *$	$*$	\top	$*$	$*$
$\top \perp$	\perp	\top	\perp	\top
$* \top$	$*$	\top	$*$	$*$
$* *$	$*$	$*$	$*$	$*$
$* \perp$	\perp	$*$	$*$	$*$
$\perp \top$	\perp	\top	\perp	\top
$\perp *$	\perp	$*$	$*$	$*$
$\perp \perp$	\perp	\perp	\perp	\perp

It would certainly seem reasonable to take it that any mode of composition meeting this condition is to be counted as two-valued functional (assuming, of course, that it also meets the constraint ruling out intentionality), but I would be loath to demand this as a necessary condition. Temptation to do so would seem to come from the influence of the construal of $*$ as 'meaningless', where an item assigned $*$ would be defective in such a way that it would immediately reduce to $*$ any compound item in which it occurred. It could, for example, only be such a perspective that would disallow a constant operator, something like ' $t(p)$ ', interpreted so as to be just constantly true, regardless of p — i.e. regardless not only of what positive value (\top or \perp) p takes, if it does take one, but also of whether p takes a positive value or not. Similarly it would need such a perspective to rule out an operator independent of one given argument place, when there may be others, e.g. $O(p,q)$, defined just to be equivalent to p . And notice that the lines problematic in the strong tables for \wedge and \vee involve cases of 'relative independence': for example, given that p is assigned \top , then $p \vee q$ is \top , regardless of q .

It is interesting to see that Haack (1974), following her apt comments on supervaluations, seems to be hovering between the two extremes of

'crude-truth-functionality' and 3-valued functionality, as if they were the only options available for spelling out the idea of truth-functionality in a partial-valued context. She does not seem to be very happy with either of them; and quite rightly it seems to me, since the truth lies in between. In the next paragraph she goes on to say that she would like to be able to discriminate within 'truth-functional systems' (meaning 3-valued functional), and asks '... whether any particular 3-valued system is specially appropriate to truth-value gaps rather than a third truth-value.' This question, of course, is one we are trying to find an answer to. Her own subsequent speculations on the matter are inconclusive, though generally rather negative in character, and I shall make clear at the end of this section a point at which I object to what she says. One thing that I might mention immediately is that she thinks there may be Fregean motivation for the crude-truth-functionalist view. I disagree, but this is better discussed in the more general context of Section II.1.

How, then, should we proceed? Let us consider more carefully what truth-functionality is in the familiar total two-valued case. Maybe we can get some inspiration here, since, after all, what we want to do is not to abandon this notion, but simply extend its application. Most perspicuous is the substitutivity formulation:

$\phi(p_1, \dots, p_n)$ is truth-functional iff a sentence
 $\phi(s_1, \dots, s_i, \dots, s_n)$ takes the same value as
 $\phi(s_1, \dots, s'_i, \dots, s_n)$ whenever s_i and s'_i take the same value.

More informal, but conceptually more basic, perhaps, is the formulation in terms of 'dependency':

$\phi(p_1, \dots, p_n)$ is truth-functional iff the truth-value of a sentence $\phi(s_1, \dots, s_n)$ depends on (is determined by) the values of the component sentences s_i .

Notice that both formulations presuppose that all the sentences involved definitely take one of the values in question. What we must do is to try and find some acceptable reformulations in which this presupposition is not made (but which do come to the same thing when applied to the total-valued case).

Let us restrict attention, for the moment, to the 'dependency' criterion. We could regard those who consider that 3-valued functionality is all that there is to the matter as not actually cancelling the presupposition, but simply making it innocuous by taking 'value' to cover * as well as \top and \perp — or, if you like, putting 'classification' in its place. While crude-truth-functionalists could reach their conclusion by cancelling conditionally the presupposition that $\phi(s_1, \dots, s_n)$ has a value, and cancelling the presupposition that the s_i have values by building in the condition that they must: "If $\phi(s_1, \dots, s_n)$ has a value, then the s_i have values on which it depends." But such a condition is surely quite extraneous to the basic idea of functional dependence: can we not be more subtle? We might try cancelling the presuppositions uniformly and say: the value, if any, of $\phi(s_1, \dots, s_n)$ depends on the values, if any, of the s_i .

This is the kind of thing we want. But we must make it clear what role is being played by a given distribution of values and gaps: the value of $\phi(s_1, \dots, s_n)$ must depend on the values themselves of the s_i , as occupants of their particular argument place, and not on the overall distribution of values and gaps — on pain of degenerating simply into 3-valued functionality. Also, (though here there is no difference from the usual total-valued case), when we say that the value of a complex item 'depends' on a certain range of values of component items, this must not be taken to imply that each of those component values is essential in the sense that if any one were different, then the value of the whole would be affected. So let us formulate

our dependency criterion of functionality in the values \top and \perp as follows:

The value of $\phi(s_1, \dots, s_n)$, if it has one, depends at most on (is determined at most by) the values of those s_i which have values, if any of them do.

From this we shall be able to derive a precise substitutivity criterion, which will justify our working with 3-entry matrices and provide the required constraint determining exactly which are admissible. But first, to illustrate the ideas concerning dependency, we consider some examples.

What about the 'strong' tables for \wedge and \vee ? At line 2 $p \vee q$ takes the value \top , although q is $*$. The crude-truth-functionalist would object, but it is open to us to claim that the value of $p \vee q$ depends on (is determined by) the value \top assigned to p . We are justified in claiming this, because provided \top is assigned to p , the value of $p \vee q$ is determined, whatever happens in the q argument place — whether or not q takes on a value, and if it does, whether this value is \top or \perp . We can explain line 4 similarly. And parallel remarks apply to the 'strong' table for \wedge .

Contrast the case of \Rightarrow . At line 2 $p \Rightarrow q$ has a value, though q does not; but we cannot in this case say that that value (\perp) is determined by the assignment of \top to p , since if it were, then whenever p were \top , $p \Rightarrow q$ should have to be \perp . But this is not so: consider line 1. $p \Rightarrow q$, then, is not value-functional in \top and \perp .

What about $T(\)$? $T(p)$ is \perp when p is $*$, and so if $T(\)$ were two-valued functional, this value (\perp) would have to depend on the values in argument places which do take positive values; actually in this case there is none, but our criterion can still apply: it means that $T(p)$ would have to be constantly \perp — which it is not. This is a limiting case

which might look a little sophisticated, but the principle is exactly the same. Contrast these reasons for rejecting $T()$ and \Rightarrow with Woodruff's argument, mentioned above: that argument simply ignored the possibility of having constant operators.

Finally negation. We have not considered this so far, but a negation operator is something we are going to need. There are two possibilities on offer in the literature, known, I believe, as 'choice negation' and 'exclusion negation', given respectively by:

	$\neg p$
T	⊥
*	*
⊥	T

	$\sim p$
T	⊥
*	⊥
⊥	T

It should be clear that \neg meets our requirements — even the crude-truth-functionalists would have no qualms about this one; but \sim is ruled out by precisely the same reasoning that was deployed against $T()$.

But, now, how may we work towards a substitutivity criterion? To say, as we did, that if a sentence $\phi(s_1, \dots, s_n)$ has a value, then this depends on the values, if any, of the s_i that take values, can be spelt out by saying: if $\phi(s_1, \dots, s_n)$ takes a value α ($\alpha = T$ or $\alpha = \perp$), then for any s_i , if s_i has a particular value β ($\beta = T$ or $\beta = \perp$), then any s'_i which also takes the value β can be substituted for s_i and the value of the (modified) sentence remain α . Allowing ourselves to write 'is' instead of 'takes the value', we can set down:

$$\phi(s_1, \dots, s_i, \dots, s_n) \text{ is } \alpha \Rightarrow$$

$$\left(\forall \beta (s_i \text{ is } \beta \Rightarrow s'_i \text{ is } \beta) \Rightarrow \phi(s_1, \dots, s'_i, \dots, s_n) \text{ is } \alpha \right)$$

with tacit universal quantifiers binding $\alpha, s_1, \dots, s_n, s'_i$. Manipulating this we get

$$\forall \beta (s_i \text{ is } \beta \Rightarrow s'_i \text{ is } \beta) \Rightarrow$$

$$\forall \alpha \left(\phi(s_1, \dots, s_i, \dots, s_n) \text{ is } \alpha \Rightarrow (s_1, \dots, s'_i, \dots, s_n) \text{ is } \alpha \right)$$

Here in fact is our substitivity criterion. If we introduce $s \sqsubseteq s'$ as an abbreviation for

$$\forall \alpha (s \text{ is } \alpha \Rightarrow s' \text{ is } \alpha)$$

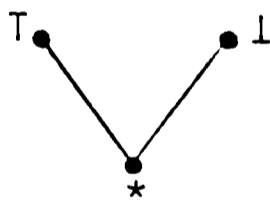
then we can write it:

$$s_i \sqsubseteq s'_i \Rightarrow \phi(s_1, \dots, s_i, \dots, s_n) \sqsubseteq \phi(s_1, \dots, s'_i, \dots, s_n).$$

Hence, as required, intensionality is ruled out — at most the \top -*- \perp classification of sentences is relevant. Given this, it is now a matter of considering the status of * as against \top and \perp , so that we can formulate our constraint on 3-entry matrices — on 3-valued functions — to pick out those that may represent bona fide modes of composition. As a relation between \top , \perp , and *, \sqsubseteq becomes

$$x \sqsubseteq y \iff \forall \alpha (x = \alpha \Rightarrow y = \alpha)$$

α ranging over $\{\top, \perp\}$ as before, x and y over $\{\top, *, \perp\}$. In other words \sqsubseteq is the partial ordering represented by the following diagram:



And so a 3-valued function f may represent a two-valued-functional mode of composition if and only if it satisfies:

$$x_i \sqsubseteq x'_i \Rightarrow f(x_1, \dots, x_i, \dots, x_n) \sqsubseteq f(x_1, \dots, x'_i, \dots, x_n)$$

In other words: iff f is monotonic with respect to the ordering \sqsubseteq .

Furthermore we may note, with pleasure, that it makes perfect sense to see such functions as representations of partial functions of the required kind. A partial function $\{\top, \perp\}^n \rightarrow \{\top, \perp\}$ may be represented by

a monotonic function $\{\top, *, \perp\}^n \rightarrow \{\top, *, \perp\}$, in just the same way that we take an ordered pair $\langle a, b \rangle$ to be represented by the set $\{\{a\}, \{a, b\}\}$ or a total function to be represented by a particular set of ordered pairs. On this understanding of the matter, we assign $*$ to truth-valueless items merely as part of the apparatus for representing the partial truth functions which themselves directly interpret partial-truth-functional modes of composition. Compare the very much more sophisticated use of monotonic functions to represent partial-valued ones, for example in Scott (1973a).

The upshot of all this, then, as far as the boolean-looking connectives is concerned, is that we can adopt the 'strong' tables for \wedge and \vee , and interpret \neg as 'choice negation'. Let us adopt, too, a constant true sentence, \top , and a constant false one, \perp . These operations are all monotonic; and clearly any complex mode of composition defined in terms of them will be monotonic also, since compounding monotonic functions never leads out of the class of monotonic functions. If this had not been the case, we should have had great cause for alarm.

But: notice that any $\phi(p_1, \dots, p_n)$ defined using only these resources will have the property that $\phi(p_1, \dots, p_n)$ fails to take on a positive value only if some p_i fails to — this is easy to check. Our criterion of truth-functionality does not, however, rule out the possibility of $\phi(p_1, \dots, p_n)$'s being undefined even on an assignment where all the p_i are defined. We need further vocabulary if we are to be able to express all monotonic modes of composition: this we shall find in the next section.

It is interesting to compare the way monotonicity turns up in our context with its use elsewhere. Consider, for example, its role in Fine (1975), and Kleene's constraint of 'regularity' — which is precisely the same as monotonicity — in Kleene (1952), § 64. In these cases what is

important about monotonicity is that it constitutes a stability condition for sentences under a 'dynamic' semantics — a semantics where stages of increasing information are involved. The requirement is that as a sentence gets evaluated at increasingly fuller stages, so it may take on a value if it had none before, but may not change its value or become valueless. In contrast, we have introduced monotonicity simply as a formal specification corresponding to our criterion of two-valued-functionality — in the context of an utterly 'static' semantics where individual interpretations are single fixed assignments of truth-values to basic sentences. It is true, however, that when we come to study logic proper, and want to generalize over interpretations, monotonicity as a stability condition will be technically important.

We mentioned Kleene: he is interested in monotonicity because the class of partial recursive predicates is closed under monotonic connectives. But Haack (Haack (1974)) considers Kleene's truth-tables, wondering whether they might not be ones peculiarly appropriate to gappy logic, as opposed to straight-forward 3-valued logic, because of the difference in status accorded to $*$ (Kleene and Haack use ' u ') as against \top and \perp (' t ' and ' f '). Of course this is exactly right! but she comes to the conclusion that the idea was no good after all — the alarming conclusion, that: "The principle Kleene is using is, in fact, precisely the one which justifies van Fraassen's semantics." Now, Kleene is dynamic and we are static, but there is formal similarity over the possibility of the (relative) independence of argument places — e.g. of q in $p \vee q$ if p is t . And it is over this that I think Haack goes wrong. She considers the example of disjunction: "Kleene argues that $|A \vee B|$ for $|A| = t$, $|B| = u$, should be t , for, since $|A| = t$, ' $A \vee B$ ' would be true whether B were true or false." She must have in mind the construal Kleene gives of u as 'unknown (or value immaterial)' when he considers disjunction on page 335 — this is the

constructive understanding of u offered alongside the classical construal 'undefined'. But she is surely wrong to have said simply '... whether B were true or false,' since what is unknown about B (or immaterial concerning the value of B), if it is assigned u , is not which of either t or f it is, but whether it is t or f or u for evermore (whether it is t or f or just u anyway). B need not be defined at all: it is not a matter of having a sentence, $Q(n)$ say, which may be algorithmically undecidable, though in fact we know (classically) that it must in fact be either true or false; $Q(n)$ might actually be undefined. Why is she ignoring the fact that partial recursive predicates can be actually undefined for some values?

Certainly if $Q(n)$ is true or false for a given n , then it is decidable which, but it is not in general decidable whether $Q(n)$ is true-or-false or not, hence not decidable in general whether $Q(n)$ is t or f or u . This means that it is surely a mistake for her to go on: "So the principle underlying this argument is, that if $F(A,B,\dots)$ would be $t(f)$ whether A,B,\dots were true or false, then it is to be $t(f)$ if A,B,\dots are u ", which, of course, as she points out, has the undesirable consequence that $|A| = u$ implies $|A \vee \neg A| = t$, and that, in general, what we get is van Fraassen's intensional, and at bottom total-valued, semantics.

Admittedly, Kleene also (p.336) considers applying his three-entry matrices to sentences formed from total predicates such as we would obtain by filling out partial recursive predicates: these would be decidable by algorithms for the original predicates on a given, possibly proper, subset of their domain but would in fact be defined everywhere. For this application Haack's remarks may be appropriate: but this application is not the basic one, and it is certainly not the idea in terms of which Kleene explains the use of his matrices to point up the distinction between t and f on the one hand, and u on the other.

I.3 EXHIBITING TRUTH-VALUE PRECONDITIONS

In partial-valued logic the 'semantic relation of presupposition', so called, becomes a non-trivial relation: ' ϕ presupposes ψ ' means that if ϕ is either true or false then ψ is true. It is furthermore taken to be an interesting relation in the writings of those who wish to offer something formally precise embodying ideas along the lines, say, of some 'Strawsonian' intuitions about 'the logic of ordinary language'. And this, indeed, is one of my aims too. Although, ultimately, we may be concerned with languages with a fixed interpretation, whose sentences therefore have a fixed truth-value classification, the relation defined as above over such sentences with respect to that single classification is not in itself going to be very revealing. To have a semantically interesting relation, there must be some kind of framework in operation which admits generalization, in some style or other, in terms of which we can take the definition to mean: whenever ϕ is either true or false, ψ is then true.

Drawing a parallel with the relation of consequence in ordinary total-valued logic should be instructive: when we say ' ϕ implies ψ ' we are usually interested not simply in the truth of the conditional 'if ϕ , then ψ ,' under a given interpretation, but rather in the fact that ψ is a logical consequence of ϕ , or, may be, that ψ is a consequence of ϕ in some particular theory in classical logic. And this logic makes its relation(s) of consequence workable and revealing because the languages considered contain two kinds of vocabulary:

- (i) items with a 'fixed' interpretation — the logical vocabulary,
- (ii) items which are sometimes (in pure logic) considered schematically and, at other's, at least considered amenable to having their interpretation varied.

It is by generalizing over instantiations, or interpretations, of type (ii) items, while keeping the interpretation of the type (i) items fixed, that

relations of consequence get a non-trivial semantical analysis. In logic we generalize over all possible instantiations, or all interpretations (models) for the language in question: while consequence-in-a-theory- T is given by generalizing over interpretations (models) of T . We have: $\phi \vdash \psi$ iff any model (of the relevant kind) which makes ϕ true, makes ψ true also.

Hence we are provided with an explanatory link-up between the vocabulary of type (i) and the notion of consequence in languages containing that vocabulary. Indeed, we are driven to want to characterize basic logical consequence directly in terms of the modes of occurrence of type (i) items in the structure of sentences: proof theory. There are lots of interesting laws concerning consequence, e.g.

$$\phi \vdash \phi \vee \psi \quad \phi \wedge \psi \vdash \phi \quad \neg \phi, \phi \vee \psi \vdash \psi$$

and so on.

Now back to partial valued languages and the relation of presupposition: what interesting instances of this relation can we find in sentential logic? Well, we can be sure of:

$$\phi \text{ presupposes } \top$$

but this is boring. Can we not reveal any rather more interesting examples? If we stick with the familiar classical connectives, \wedge, \vee, \neg , etc as the only vocabulary of type (i), then interesting cases do seem rather thin on the ground. For ' ϕ presupposes ψ ' is naturally spelt out: any interpretation which makes ϕ either true or false makes ψ true, where for present purposes interpretations are just assignments of \top, \perp or $*$ to basic sentence-symbols — our type (ii) vocabulary. As it stands this is 'logical presupposition', though we might want to constrain the range of interpretations to allow for a semantical characterization of presupposition in a given theory. We do have one moderately interesting — though still pretty unrevealing — instance of presupposition:

end of 1.5 : insert a comma

ϕ presupposes $\phi \vee \neg \phi$.

And notice that this looks more interesting on the naive approach than on the supervaluational one, since if you cobble up partial valuations in terms of classical ones, then the example just reduces to ' ϕ presupposes \top ' all over again. For us, ϕ is true or false if, as well as only if $\phi \vee \neg \phi$ is true.

But in virtue of what internal structure of a sentence might it interestingly presuppose something? Are we condemned to have nothing but trivial instances, other than any that may be imposed on the language by a particular theory constraining the interpretation of type (ii) vocabulary, or — what comes to the same thing from the point of view of sentential logic — resulting from non-sentential structure? There is a passage in van Fraassen (1971) (p.138), which seems to suggest that he was resigned to this. He contrasts non-classical logics which are non-classical because they contain novel connectives (such as modal operators) with logics such as the 'logic of presupposition', where, he says, 'one studies non-classical relations among (sets of) sentences'.

But we should not, I think, rest content with black-box semantics. The presupposition relation becomes non-trivial in partial-valued logic, and if it is furthermore considered an interesting logical relation, then we should surely attempt to reveal why it is in terms of structure discerned in the languages whose sentences are to come within the scope of that relation.

Certainly there are famous instances of the presupposition relation which can be explained in terms of exotic quantificational structure: consider, for example, the complex quantifiers, $\exists xFx$ and $\forall xFx$, meaning 'the F, x ' and 'every F, x ', with Strawsonian truth-falsehood conditions:

$\exists xFx \cdot Gx$ is true iff there is a unique F and it is G

is false iff there is a unique F and it is not G

$\forall xFx \cdot Gx$ is true iff there is at least one F and every one is G

is false iff there is at least one F and at least one is not G .

Although predicates and quantifiers do not come up officially until Chapter II, we can consider these cases by way of example, making the assumption that F and G are themselves totally-defined and unproblematic predicates. In their simplest occurrences (not within the scope of any further functors) these quantifiers provide the following instances of presupposition:

$\exists xFx \cdot Gx$ presupposes $\exists y\forall x(x = y \leftrightarrow Fx)$

$\forall xFx \cdot Gx$ presupposes $\exists xFx$.

However, it is, from the point of view of logical analysis at least, a pretty ad hoc manouevre just introducing primitive quantifiers to match the truth-falsehood conditions you want. Moreover they are somewhat cumbersome linguistic items. We should try to break them down and provide a paraphrase using simple quantifiers and sentential structure. This is what we do for 'any F is G ' when we analyse it $\forall x(Fx \rightarrow Gx)$, and this is what Russell attempted, in total-valued logic, for definite description idioms.

But clearly, if we want to do this and preserve the desired truth-falsehood conditions, then new sentence connectives are required: assuming that F and G are totally defined on the domain of quantification, it is easy to see that no sentence built up using $\neg, \wedge, \vee, \forall, \exists$, etc. could ever be undefined. What we need are operators that may exhibit conditions under which a sentence fails to take on a truth-value even when all its basic components are totally-defined — that can introduce gaps. In other words, we must be able to define sentential formulae, $\phi(p_1, \dots, p_n)$, which lack a truth-value under certain interpretations according to which all the p_i are defined. This, recall, was precisely the demand we made at the end of the last section.

The most natural connectives meeting this need are, I think, the following

\ast	$\top \ast \perp$
\top	$\top \ast \ast$
\ast	$\ast \ast \ast$
\perp	$\ast \ast \perp$

p/q	$\overbrace{\top \ast \perp}^q$
$\left\{ \begin{array}{l} \top \\ \ast \\ \perp \end{array} \right.$	$\top \ast \perp$
	$\ast \ast \ast$
	$\ast \ast \ast$

\ast , note, has the truth-conditions of \wedge and the falsehood-conditions of \vee . I call it 'squadge', but I will abandon my original mouthful 'squadgejunction' and instead shall call $\phi \ast \psi$ an interjunction — the interjunction of ϕ and ψ , which are its interjuncts. $/$ has the truth-conditions of \wedge but the falsehood-conditions of $\rightarrow (= (\neg \dots) \vee \dots)$. Clearly both $/$ and \ast have the required property of introducing truth-value-gaps under circumstances when all constituents are totally defined: and they are, of course, monotonic — hence truth-functional in the sense of the last section. Furthermore, if we add squadge and slash to our stock of connectives, then we shall be able to define any truth-functional mode of composition. In fact either one of them on its own would do. This we shall show very shortly.

These connectives may clearly be used to provide analyses for $\exists x Fx \cdot Gx$ and $\forall x Fx \cdot Gx$. $\exists x Fx \cdot Gx$ could be defined

$$\exists y \forall x (x = y \leftrightarrow Fx) / \forall x (Fx \rightarrow Gx)$$

or equivalently:

$$\exists y \forall x (x = y \leftrightarrow Fx) / \exists x (Fx \wedge Gx)$$

or even:

$$\exists y \forall x (x = y \leftrightarrow Fx) / \exists y (\forall x (x = y \leftrightarrow Fx) \wedge Gy)$$

or, using squadge:

$$\exists y (\forall x (x = y \leftrightarrow Fx) \wedge Gy) \ast \forall y (\forall x (x = y \leftrightarrow Fx) \rightarrow Gy) .$$

And for $\forall x Fx \cdot Gx$:

$$\exists x Fx / \forall x (Fx \rightarrow Gx) \text{ or } (\exists x Fx \wedge \forall x (Fx \rightarrow Gx)) \ast (\exists x Fx \rightarrow \forall x (Fx \rightarrow Gx)) .$$

It should be easy enough to convince yourself that these paraphrases have the right truth-conditions, since / and \times do not themselves occur within the scope of a quantifier: we shall in any case check up properly in Section II.2. These are stock examples of 'presuppositional quantifiers'. In Chapter V I shall be suggesting that there are others too whose analysis can illuminatingly be given in terms of our new expressive resourcefulness at sentence level.

I do not want to start writing Chapter V yet but it might be useful, just for a moment, to think of the presupposition of a logical formula — by which I mean its truth-value precondition — as playing the same kind of role as the presupposition of a presuppositional idiom in natural language, since we are presented with the problem: what, if any, are the natural language analogues of \times and /? Whatever complex idioms squadge and slash may contribute towards analysing, are there any idioms corresponding simply to $p \times q$ and p / q ?

At least in the case of squadge, I think that there are. First, consider what the usefulness of such an idiom might be. The presupposition of $p \times q$ is that p and q take on the same truth-value and the formula is true if that value is \top , false if it is \perp ; and so a corresponding natural language idiom would be a sentence with two immediate constituents, carrying the presupposition that those constituents stand or fall together. It seems likely that the typical point of asserting such a sentence would be to convey that the constituent sentences were being 'indiscriminately asserted' — asserted as coming to one and the same thing (for the purposes of that utterance at least). And this is something we might feel the need to do either because, for the sake of clarity, we want to make a point in alternative ways or because we feel the need to fill out or explain one way of saying something in terms of another.

It would, then, seem a good idea to look for modes of natural language composition associated with this kind of use, to see if we might plausibly discern \ast as their truth-functional ingredient. There are various kinds of idiom which are used in this way. Most frequently, perhaps, we just juxtapose two sentences with our intonation pattern indicating that there is an assertion of one complex sentence rather than two separate assertions; and this might be represented orthographically using a colon or using ' — '. Consider for example:

(1) John is a follower of Pythagoras — John eats no beans.

Often we use 'i.e.' or 'or' (= 'or you might just as well say') for a similar purpose:

(2) $a^2 = 1$ i.e. $a = 1$ or $a = -1$.

Now, I am tempted not merely to take such sentences as intuitive counterparts to formal interjunctions, but, conversely, to be a kind of idiom for which, as part of a programme of natural language semantics, we should propose an interjunctive analysis. Of course we should have to have a story to account for the way such modes typically involve more than the truth-functional role of \ast : indeed precisely those features which guided our search for a \ast -analogue are in question. But this is a familiar situation and it does not seem to me that our natural language interjunctions are any further away from \ast than, say, natural language conjunctions are from \wedge . One particular problem is that although \ast is commutative sentence (1) has a different feel about it from

(3) John eats no beans — John is a follower of Pythagoras.

But we have already considered a possible reason for this: the second interjunct in such an example is naturally taken to explain the first. Compare the notorious non-commutativity of 'and'.

A proper evaluation of the proposal to analyse these examples pre-suppositionally — with \ast — will have to be postposed, of course, until

we consider natural language presupposition more fully. But examples and theorising go hand-in-hand and, in my view, natural language interjunctions cast a very interesting perspective on the phenomenon of presupposition.

Before we leave the matter, we might consider some further examples, which are, perhaps, more compelling because they occur more often and with a more obvious grammatical manifestation. These are cases where I would say that we have interjunctions in reduced form, with a phrase in apposition. Such phrases may be written in brackets or introduced by 'viz.', by '-' or just by a comma. Compare the parallel grammatical behaviour of 'and':

(4) Cyril was serving drinks in the college garden – in the old church yard.

(5) Cyril was serving drinks in the college garden and in the front quadrangle.

And consider, finally, the following example of a reduced interjunction from the introduction of 'Truth and Meaning' (Eds. Evans and McDowell):

(6) Since it is undeniable that "snow is white" gives the meaning of, is a semantical representation of, "snow is white", ...

It is more of a problem to discern simple idioms corresponding to p/q . The point about a formula p/q is that it wears its presupposition – at least the presupposition introduced by that occurrence of / – on its sleeve, viz. p . Perhaps sentences of a form such as 'as well as the fact that p, q ' or 'not only p , but also q ' are the nearest we can get. May be we should not be surprised that we cannot do any better, since arguably there would not be much use for an explicit idiom corresponding to slash.

Before we proceed any further, we must pause to consider Belnap (1970). Here he introduces a conditional operator '/' whose features are in some respects like $out / :$ but the two should be thoroughly distinguished. (I had not read Belnap when I chose the symbol '/', so the coincidence of notation is either pure accident or has some subtle psycho-orthographical

explanation.) Belnap's use of $' / '$ is designed to cater for the kind of gap people have wanted to discern in conditionals 'if p, q ' when p is false and, along with this divergence of purpose, his account of its semantical role differs from our $/$ in some crucial ways. First, although he eschews the word 'meaningless' as a gloss for the classification of p/q when p is false, I think he would, in one respect at least, fall into the camp of those who work with the paradigm of the 'gap' as a kind of semantical deficiency, since his idea is, using a 'depragmatized' (*sic*) sense of 'what is asserted', that if p is false, then p/q asserts nothing — meaning expresses no proposition. He models propositions by sets of possible worlds, and with respect to a world in which p is false p/q is not assigned any set of possible worlds.

His semantics, then, consists of partial mappings into total-valued interpretations rather than straightforward partial-valued interpretations. If you wanted to extend my proposals to a possible world semantics, so as to make the contrast with Belnap clearer, then any sentence would have a propositional content with respect to any world. And the content of a sentence would be determined by the triple (T, U, F) , where T were the set of worlds where the sentence were true, U where valueless, and F where false. Such a triple would represent a single 'partial-valued proposition'. Hence gap conditions are determined by propositional content; they are not a precondition of there being any such content.

Furthermore, for Belnap if p is 'non-assertive' this does not necessarily make p/q 'non-assertive'.

Van Fraassen, however, when he takes over Belnap's $' / '$ (van Fraassen (1975)) wants to modify this last feature, and cast p/q non-assertive if p is. Disentangling it all from the possible worlds, proto-valuations, sets of possible worlds, valuations, etc., this means that $' / '$ would then be given the same truth-table as our $/$. I experienced some

personal chagrin when I discovered this article, but my spirits were largely restored when it turned out that this modification was not motivated by any of the features of this interpretation that I find dear. There seems to be no interest in monotonicity, and the nature of / 's gap introducing conditions are obscured in such a way that I could continue to see myself as an exhibitor of gaps in contrast to van Fraassen. For at bottom — even with / — his semantics contains 'proto-valuations', which are total assignments of sets of possible worlds (total-valued propositions) to sentence-world pairs. He goes partial by defining valuations in a supervaluational way out of the proto-valuations: valuations then turn out to be partial mappings into total-valued propositions in the spirit of Belnap's idea.

Van Fraassen's remarks here on linguistic practice are brief, but for him the truth-value preconditions of a sentence (equivalently for him and Belnap, proposition-expressing preconditions) would seem to correspond to preconditions for an assertion's having been made in uttering that sentence (though I am not clear whether he is thinking of the conditional kind of gap or the more familiar presuppositional kind or both). This will not be our way with presupposition: the difference in semantics will be paralleled by a corresponding difference in the account of assertion with presuppositional idioms. So far from presupposition being seen as a precondition of assertion, there will be no presupposition without a successful act of assertion: and what is presupposed will be directly determined by what is asserted, none of this depending on whether or not the presupposition holds.

Let us now set up some apparatus. Since we are just doing schematic semantics at the moment, we may consider a single propositional 'language' with a denumerable stock of atomic 'sentences' (which we talk about using 'p', 'q', etc.). Formulae (for which we use ' ϕ ', ' ψ ' etc.) are built up by means of \neg , \wedge , \vee , \ast and / from these atoms together

with the constant sentences \top and \perp ; and let us also have a constantly undefined sentence $*$. ' \rightarrow ' and ' \leftrightarrow ' can be defined in the usual way in terms of \neg , \wedge and \vee ; and note that $*$ could be taken to be defined also: for example by $\top * \perp$. A basic assignment v maps the atoms into $\{\top, *, \perp\}$ and can be extended to a valuation \bar{v} mapping any formula into $\{\top, *, \perp\}$. (Ambiguity with the constant sentences is harmless — we could even identify them.) \bar{v} is specified by the following clauses:

$$\bar{v}(\top) = \top \quad (\bar{v}(*)) = * \quad \bar{v}(\perp) = \perp$$

$$\bar{v}(p) = v(p) \quad \text{if } p \text{ is an atomic sentence}$$

$$\begin{aligned} \bar{v}(\neg\phi) = \top &\iff \bar{v}(\phi) = \perp \\ &= \perp \iff \bar{v}(\phi) = \top \end{aligned}$$

$$\begin{aligned} \bar{v}(\phi \wedge \psi) = \top &\iff \bar{v}(\phi) = \top \quad \text{and} \quad \bar{v}(\psi) = \top \\ &= \perp \iff \bar{v}(\phi) = \perp \quad \text{or} \quad \bar{v}(\psi) = \perp \end{aligned}$$

$$\begin{aligned} \bar{v}(\phi \vee \psi) = \top &\iff \bar{v}(\phi) = \top \quad \text{and} \quad \bar{v}(\psi) = \top \\ &= \perp \iff \bar{v}(\phi) = \perp \quad \text{or} \quad \bar{v}(\psi) = \perp \end{aligned}$$

$$\begin{aligned} \bar{v}(\phi * \psi) = \top &\iff \bar{v}(\phi) = \top \quad \text{and} \quad \bar{v}(\psi) = \top \\ &= \perp \iff \bar{v}(\phi) = \perp \quad \text{and} \quad \bar{v}(\psi) = \perp \end{aligned}$$

$$\begin{aligned} \bar{v}(\phi / \psi) = \top &\iff \bar{v}(\phi) = \top \quad \text{and} \quad \bar{v}(\psi) = \top \\ &= \perp \iff \bar{v}(\phi) = \top \quad \text{and} \quad \bar{v}(\psi) = \perp \end{aligned}$$

The positive truth-values \top and \perp get mentioned, ' $*$ otherwise' being taken for granted: so the semantics can be given in terms of the two positive truth-values, and in specifying truth-falsehood conditions the truth-value preconditions take care of themselves.

The assignments v are officially defined on all atomic sentences, though, of course, in consideration of a particular formula ϕ it is only what is assigned to the atomic sentences occurring in ϕ , which is relevant:

Lemma I.3.1: If v and w agree on all atoms occurring in ϕ , then

$$\bar{v}(\phi) = \bar{w}(\phi) .$$

Proof: An easy induction.

We shall subsequently invoke this lemma without special mention.

A relation of logical presupposition may now actually be defined:

ϕ presupposes ψ iff:

$$\bar{v}(\phi) = \top \text{ or } \bar{v}(\phi) = \perp \Rightarrow \bar{v}(\psi) = \top, \text{ for all } v .$$

And we have some interesting instances of this relation. For example for any formulae ϕ and ψ :

$$\phi/\psi \text{ presupposes } \phi .$$

Of course in virtue of internal complexity in ψ , ϕ may not represent 'the whole presupposition' of ϕ/ψ , e.g.

$$p/(q/r) \text{ presupposes } p$$

but also

$$p/(q/r) \text{ presupposes } q .$$

The important presuppositional property of squadge, on the other hand, is given by:

$$\phi \times \psi \text{ presupposes } \phi \leftrightarrow \psi .$$

And we could go on endlessly giving exciting examples, ones involving any degree of complexity: the relation of presupposition has become interesting. However, ironically — but natural enough, now that we have the resources adequately to exhibit presuppositions, the metalinguistic relation ceases to be of such theoretical interest; in just the same way that, since we can adopt a negation operator, the relation 'is the contradictory of' is not called upon to play any great role in the development of logic.

The ordering \sqsubseteq on $\{\top, *, \perp\}$ induces a degree-of-definedness ordering between assignments:

$$v \sqsubseteq w \Leftrightarrow v(p) \sqsubseteq w(p) \text{ for all atoms } p$$

and on valuations obtained from basic assignments :

$$\bar{v} \sqsubseteq \bar{w} \iff \bar{v}(\phi) \sqsubseteq \bar{w}(\phi) \quad \text{for all formulae } \phi$$

and so the monotonicity of all modes of composition in the language may be stated:

Lemma 1.3.2: $v \sqsubseteq w \Rightarrow \bar{v} \sqsubseteq \bar{w}$.

Proof: Induction: We have already observed that any mode of composition compounded out of basic monotonic ones must itself be monotonic.

Of course the converse of this lemma holds trivially.

A relation of logical equivalence may be defined in the obvious way:

$$\phi \simeq \psi \iff \bar{v}(\phi) = \bar{v}(\psi) \quad \text{for all } v .$$

Using this we can record some characteristic properties of formulae.

Squadge is 'half \wedge and half \vee '; and we may state this in terms of squadge itself:

$$\phi \times \psi \simeq (\phi \wedge \psi) \times (\phi \vee \psi)$$

Similarly slash is 'half \wedge and half \rightarrow ':

$$\phi / \psi \simeq (\phi \wedge \psi) \times (\phi \rightarrow \psi)$$

So slash is definable in terms of squadge. The converse is also true:

$$\phi \times \psi \simeq (\phi \leftrightarrow \psi) / \phi \simeq (\phi \leftrightarrow \psi) / \psi .$$

Compare this with our remarks about the presupposition of interjunctions.

The behaviour of negation is apt:

$$\neg(\phi \times \psi) \simeq \neg\phi \times \neg\psi \quad \text{and} \quad \neg(\phi / \psi) \simeq \phi / \neg\psi .$$

Squadge is self-dual, as it should be: an indiscriminating expression of two sentences ϕ and ψ should have a negation equivalent to the indiscriminating expression of the negations of ϕ and ψ . And the fact that negation just slips past slash — so slips past a presupposed component of a sentence — is a structural manifestation of the very definition of the semantical relation of presupposition.

A final glance back to our 'Strawsonian' quantifiers is to the point here. Taking it on trust that all goes well with quantifiers and predicates, these facts about negation show that

$$\neg \exists x Fx \cdot Gx \simeq \exists x Fx \cdot \neg Gx .$$

Contrast the scope problems of the Russellian paraphrase in total-valued logic. More on this however in Section II.3. And:

$$\neg \forall x Fx \cdot Gx \simeq \exists x Fx \cdot \neg Gx$$

where $\exists x Fx \cdot Gx$ is defined by $\exists x Fx / \exists x (Fx \wedge Gx)$. This is the obvious dual of 'every F is G ', viz. 'some F is G ', understood to be false only if there is an F .

Some basic facts about squadge :

$$\phi \times \psi \simeq \psi \times \phi \quad \text{and} \quad \phi \times (\psi \times \chi) \simeq (\phi \times \psi) \times \chi$$

so \times is commutative and associative. Also — there is no need to write it down — \times distributes both ways round with \wedge and with \vee . We have associativity of $/$ also:

$$\phi / (\psi / \chi) \simeq (\phi / \psi) / \chi.$$

$*$ is, of course, equivalent to $\top \times \perp$: and it is also equivalent to $\neg *$ and to the interjunction of any formula and its negation (i.e. $\phi \times \neg \phi$). We can complete the circle of inter-definability by showing that \times , and hence $/$ too, is definable in terms of $*$. Notice the self-dual patterns:

$$\begin{aligned} \phi \times \psi &\simeq (\phi \wedge \psi) \vee (\phi \wedge *) \vee (* \wedge \psi) \simeq (\phi \wedge (* \vee \psi)) \vee ((\phi \vee *) \wedge \psi) \\ &\simeq (\phi \vee \psi) \wedge (\phi \vee *) \wedge (* \vee \psi) \simeq (\phi \vee (* \wedge \psi)) \wedge ((\phi \wedge *) \vee \psi) \end{aligned}$$

Now, finally, we turn to the promised theorem on expressive completeness. First another definition: the relation of compatibility of truth-value classification. Where x and y range over $\{\top, *, \perp\}$, define

$$x \square y \iff \underline{\text{neither } x = \top \text{ and } y = \perp \text{ nor } x = \perp \text{ and } y = \top}$$

This induces a compatibility relation on assignments (cf. ' \sqsubseteq '):

$$v \square w \iff v(p) \square w(p) \text{ for all } p$$

We shall also use both ' \sqsubseteq ' and ' \square ' between n -tuples $\vec{x} \in \{\top, *, \perp\}^n$, so that

$$\vec{x} \sqsubseteq \vec{y} \iff x_i \sqsubseteq y_i \text{ for all } i \quad \vec{x} \square \vec{y} \iff x_i \square y_i \text{ for all } i$$

Now it is clear enough what it means to say that a formula $\phi(p_1, \dots, p_n)$ expresses a given function $f: \{\top, *, \perp\}^n \rightarrow \{\top, *, \perp\}$; but, so we can keep argument places straight, let us say: ϕ expresses f with respect to $\langle p_1, \dots, p_n \rangle$ iff:

$$f(v(p_1), \dots, v(p_n)) = \bar{v}(\phi) \text{ for all } v.$$

(Note: no p_i need be in ϕ , nor even need any of the atomic sentences in ϕ be among p_1, \dots, p_n for the definition to make sense.) Now given an n -place monotonic function f and an n -tuple $\langle p_1, \dots, p_n \rangle$, we can present an explicit receipt for constructing a formula $\phi_f(p_1, \dots, p_n)$ which expresses f with respect to $\langle p_1, \dots, p_n \rangle$. Let $\phi_f(p_1, \dots, p_n)$ be the following:

$$\left(\bigwedge_{\vec{x}} \bigvee_i \top(\vec{x}, i) \right) * \left(\bigvee_{\vec{x}} \bigwedge_i \perp(\vec{x}, i) \right)$$

where \vec{x} ranges over $\{\top, *, \perp\}^n$, i over $\{1, \dots, n\}$ and $\top(x, i)$ and $\perp(x, i)$ are defined by cases as follows:

$$\left. \begin{aligned} \top(\vec{x}, i) &= p_i \text{ if } x_i = \top \\ &= \neg p_i \text{ if } x_i = \perp \\ &= \top \text{ otherwise} \end{aligned} \right\} \text{--- if } f(\vec{x}) = \top$$

$$= \perp \text{ --- otherwise}$$

$$\left. \begin{aligned} \perp(\vec{x}, i) &= p_1 \quad \text{if } x_i = \perp \\ &= \neg p_1 \quad \text{if } x_i = \top \\ &= \perp \quad \text{otherwise} \end{aligned} \right\} \text{--- if } f(\vec{x}) = \perp$$

$$= \top \text{ --- - - - - - otherwise}$$

We write ϕ_f for short and refer to the left and right hand interjuncts respectively by ϕ_f^\top and ϕ_f^\perp , so $\phi_f = \phi_f^\top \ast \phi_f^\perp$.

Theorem I.3.3: If f is a monotonic function $\{\top, \ast, \perp\}^n \rightarrow \{\top, \ast, \perp\}$, then ϕ_f expresses f with respect to $\langle p_1, \dots, p_n \rangle$.

Proof: The first thing we show is that:

$$(i) \quad \text{for any } v, \bar{v}(\phi_f^\top) = \top \Rightarrow f(v(p_1), \dots, v(p_n)) = \top$$

Say $\bar{v}(\phi_f^\top) = \top$, then for some \vec{x} $\bar{v}(\bigwedge_i \top(\vec{x}, i)) = \top$. But from the definition of $\top(x, i)$ we can deduce that $f(\vec{x}) = \top$ and that $\vec{x} \sqsubseteq \langle v(p_1), \dots, v(p_n) \rangle$.

Hence, by the monotonicity of f , $f(v(p_1), \dots, v(p_n)) = \top$ as required.

From (i) we have: for any v ,

$$\bar{v}(\phi_f) = \top \Rightarrow f(v(p_1), \dots, v(p_n)) = \top.$$

Conversely we need:

$$(ii) \quad \text{for any } v, f(v(p_1), \dots, v(p_n)) = \top \Rightarrow \bar{v}(\phi_f) = \top$$

But if $f(v(p_1), \dots, v(p_n)) = \top$, then $\bar{v}(\top(\langle v(p_1), \dots, v(p_n) \rangle, i)) = \top$ for all i ; hence $\bar{v}(\phi_f^\top) = \top$. We must also show that $\bar{v}(\phi_f^\perp) = \top$. If this is not so, then there is an n -tuple \vec{x} such that for all i $\bar{v}(\perp(\vec{x}, i)) \neq \top$; hence $f(\vec{x}) = \perp$, and also $\vec{x} \sqsupset \langle v(p_1), \dots, v(p_n) \rangle$, and so we can find some $\vec{y} \in \{\top, \ast, \perp\}^n$ such that (a) $\vec{x} \sqsubseteq \vec{y}$, and (b) $\langle v(p_1), \dots, v(p_n) \rangle \sqsubseteq \vec{y}$.

But then, by monotonicity, (a) implies $f(\vec{y}) = \perp$ and (b) implies $f(\vec{y}) = \top$, which is impossible. So indeed $\bar{v}(\phi_f^\perp) = \top$. And $\bar{v}(\phi_f) = \top$. We have shown, therefore, that for any v :

$$\bar{v}(\phi_f) = \top \iff f(v(p_1), \dots, v(p_n)) = \top$$

end of 1.8 : insert "(and T or L)"

and by exactly parallel reasoning:

$$\bar{v}(\phi_f) = \perp \iff f(v(p_1), \dots, v(p_n)) = \perp$$

Hence $\bar{v}(\phi_f) = f(v(p_1), \dots, v(p_n)) \quad \square$

Corollary I.3.4: Any set of connectives in terms of which $\neg, \wedge, \vee, \times, \top$ and \perp may be defined are expressively adequate for partial truth-functional modes of composition.

Dana Scott originally proved the expressive completeness theorem for me, showing by induction on the number of variables that \neg, \wedge, \vee and $/$ were adequate. However it is nice to have this explicit proof, since the formulae ϕ_f provide pleasant 'normal forms'.

Corollary I.3.5: Any formula is equivalent to ϕ_f for some f .

Proof: If p_1, \dots, p_n are the atomic components of a formula ϕ (listing them any way you like), then ϕ expresses with respect to $\langle p_1, \dots, p_n \rangle$ the function $f: \langle x_1, \dots, x_n \rangle \mapsto \bar{v}_{\vec{x}}(\phi)$, where $v_{\vec{x}}(p_i) = x_i$ (and, for the sake of definiteness, say $v_{\vec{x}}(p) = *$ for any other p). And so $\phi_f \simeq \phi$. \square

What helps to make the ϕ_f pleasing is that their truth conditions and falsehood conditions are clearly exhibited: ϕ_f^\top determines the truth conditions and ϕ_f^\perp the falsehood conditions:

Corollary I.3.6: (a) $\bar{v}(\phi_f) = \top \iff \bar{v}(\phi_f^\top) = \top$,
 (b) $\bar{v}(\phi_f) = \perp \iff \bar{v}(\phi_f^\perp) = \perp$

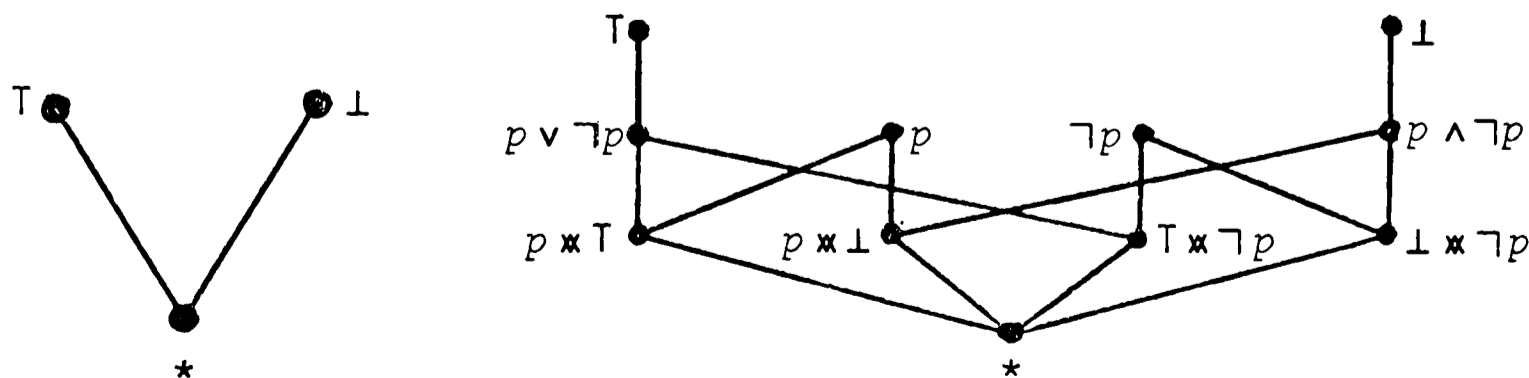
Proof: (a): \Rightarrow is trivial, while conversely if $\bar{v}(\phi_f^\top) = \top$, then by part (i) of the proof of the main theorem $f(v(p_1), \dots, v(p_n)) = \top$, and so by part (ii) $\bar{v}(\phi_f) = \top$. (b): Dual reasoning. \square

But ϕ_f is likely to contain many otiose occurrences of \perp and \top . We could obtain a more economical formula by first restricting the indexing

set for the disjunction ϕ_f^\top and the conjunction ϕ_f^\perp to those \vec{x} such that $f(\vec{x}) = \top$ and such that $f(\vec{x}) = \perp$, respectively, and then, for a given \vec{x} , restricting the indexing sets for the conjunctions in ϕ_f^\top and the disjunctions in ϕ_f^\perp to those i such that $x_i = \top$ or $x_i = \perp$. (Note: empty disjunctions are to be defined to be \perp and empty conjunctions to be \top .) What this comes to is simply ignoring the 'otherwise' cases in the definition of $\top(\vec{x}, i)$ and $\perp(\vec{x}, i)$. Further, we could weed out more disjuncts, respectively conjuncts, in ϕ_f^\top and ϕ_f^\perp by restricting the \vec{x} to those which are minimally necessary to determine the value \top , respectively \perp : i.e. the disjunction ϕ_f^\top could be indexed by $\{\vec{x} : f(\vec{x}) = \top \text{ and } \forall \vec{y} \sqsubseteq \vec{x} \cdot f(\vec{y}) \neq \top\}$ and the conjunction ϕ_f^\perp by $\{\vec{x} : f(\vec{x}) = \perp \text{ and } \forall \vec{y} \sqsubseteq \vec{x} \cdot f(\vec{y}) \neq \perp\}$. By monotonicity these \vec{x} would be sufficient to determine all cases.

In classical logic we have a neatly expressible function giving how many formulae there are, up to logical equivalence, in n atomic sentences. The question is equivalent to asking how many functions $\{\top, \perp\}^n \rightarrow \{\top, \perp\}$; and there are 2^{2^n} . What about our partial-valued languages? The question is: how many \sqsubseteq -monotonic functions are there $\{\top, *, \perp\}^n \rightarrow \{\top, *, \perp\}$? This looks an easy question but in fact it turns out to be a combinatorial problem of some complexity.

The function certainly grows very fast, but for 0 and 1 we can draw the following diagrams:



Here the ordering up and down corresponds to a degree-of-definedness relation between formulae, so we have yet another use for ' \sqsubseteq ':

$$\phi \sqsubseteq \psi \iff \bar{v}(\phi) \sqsubseteq \bar{v}(\psi) \quad \text{for all } v.$$

Similarly \square induces an important relation of logical compatibility between formulae:

$$\phi \square \psi \iff \bar{v}(\phi) \square \bar{v}(\psi) \quad \text{for all } v.$$

Notice that the meets (with respect to \sqsubseteq) on the diagram correspond to interjunctions. Have we got anything corresponding to the joins? The matter can be cast better in a more linguistic form: first, it is clear that there can be no mode of composition $\chi(p,q)$ such that for any ϕ and ψ , and any v :

$$\begin{aligned} \bar{v}(\chi(\phi, \psi)) = \top &\iff \bar{v}(\phi) = \top \text{ or } \bar{v}(\psi) = \top \\ &= \perp \iff \bar{v}(\phi) = \perp \text{ or } \bar{v}(\psi) = \perp \end{aligned}$$

since ϕ and ψ may not be compatible; and equally, since we are truth-functional, there cannot even be a mode of composition $\chi(p,q)$ such that if two formulae ϕ and ψ are logically compatible, then $\chi(\phi, \psi)$ has the truth-falsehood conditions exhibited. The point is that the formula $\chi(\phi, \psi)$ would have corresponded to the join of (the equivalence classes of) ϕ and ψ . All is not lost, however, since we can show that if ϕ and ψ are compatible, then at least there is always some formula with the desired truth-falsehood conditions. Let us call such a formula a 'joint': that is to say a formula is a joint of ϕ and ψ iff it satisfies the disjunctive truth-falsehood conditions exhibited above — for any v . (Joints will in fact be effectively obtainable, since everything is very finite.)

Theorem I.3.7: If $\phi \square \psi$, then there is a joint of ϕ and ψ .

for "11.4" read "1.4"

Proof: Let p_1, \dots, p_n be the atomic sentences occurring in either ϕ or ψ . For each $\vec{x} \in \{\top, \star, \perp\}^n$ pick a $v_{\vec{x}}$ such that $v_{\vec{x}}(p_i) = x_i$, and let $f_\phi: \vec{x} \mapsto \bar{v}_{\vec{x}}(\phi)$ and $f_\psi: \vec{x} \mapsto \bar{v}_{\vec{x}}(\psi)$. ϕ and ψ express f_ϕ and f_ψ , respectively, with respect to $\langle p_1, \dots, p_n \rangle$, and so, since $\phi \sqcap \psi$, $f_\phi(\vec{x}) \sqcap f_\psi(\vec{x})$ for all \vec{x} . Hence the following definition of a function f makes sense:

$$\begin{aligned} f(\vec{x}) &= \top \quad \text{if either } f_\phi(\vec{x}) = \top \quad \text{or } f_\psi(\vec{x}) = \top \\ &= \perp \quad \text{if either } f_\phi(\vec{x}) = \perp \quad \text{or } f_\psi(\vec{x}) = \perp \end{aligned}$$

Furthermore f is monotonic, and so, by Theorem (I.2.3), there is a formula, viz. $\phi_f(p_1, \dots, p_n)$, which has the right truth-falsehood conditions for being a joint of ϕ and ψ . \square

The converse of (I.3.7) of course holds trivially.

The matter of joints will be considered again in Chapters III and IV. An analogous result holds for logic with singular-terms and quantifiers but we need more powerful techniques to prove it; and we shall also be interested in generalizing the question to formulae compatible in a given theory.

II.4 LOGICALLY CONVEX SETS

Think of classical logic: think of any language which, whatever further structure it may possess, is at least closed under the boolean sentence-operators and obeys the laws of classical total-valued propositional logic. There will be a 'logical consequence' relation, \vdash , holding between arbitrary sets of sentences and sentences, which satisfies the obvious basic properties:

$$\Gamma, \phi \vdash \phi$$

$$\Gamma \vdash \phi \Rightarrow \Gamma, \Delta \vdash \phi$$

$$\Gamma, \psi \vdash \phi \quad \text{and} \quad \Gamma \vdash \psi \Rightarrow \Gamma \vdash \phi$$

which is compact:

$$\Gamma \vdash \phi \Rightarrow \Gamma_0 \vdash \phi \quad \text{for some finite subset } \Gamma_0 \subseteq \Gamma$$

and which obeys boolean logic. However, we need not assume that it is boolean structure alone which determines \vdash : for example \vdash may be determined by quantifier logic, or modal logic.

If such a language is considered with respect to a fixed interpretation, then we may, heuristically at least, think of the 'propositional content' of a given sentence. Even if the 'interpretation' in question is just given extensionally by a (set-theoretically constructed) model, the 'propositional content' of a sentence need not be taken to reduce simply to the resultant truth-value but can be understood as sensitive to the structure of the sentence. Now, for many purposes, we 'factor out modulo logical equivalence' and think of two, possibly distinct, sentences ϕ and ψ 'expressing', let us say, the same propositional content if $\vdash \phi \leftrightarrow \psi$. In this case it is equivalence-classes of sentences that capture propositional contents one-one. If we generalize over interpretations, then we can take these equivalence-classes as representing various possible propositional contents, corresponding to the various interpretations.

Now consider theories. A 'theory' T is to be understood to mean not a particular favoured set of axioms, but simply a set of sentences closed under logical consequence ($T \vdash \phi \Rightarrow \phi \in T$). In the context of a particular theory we are often interested in what a sentence expresses modulo not merely logical equivalence, but modulo equivalence in T . This is a stronger equivalence relation: $T \vdash \phi \leftrightarrow \psi$: and we may take the equivalence-classes to be representations of 'propositional content with respect to T ' — representing a fixed content under a given interpretation (a given model of T), or ranging over various propositional contents according as interpretations range through various models of T .

Let us use the notation $(T; \phi)$ for the equivalence-class of ϕ modulo T , i.e.:

$$(T; \phi) = \{ \psi \mid T \vdash \phi \leftrightarrow \psi \} .$$

for "proposition" and "propositional"

So, ϕ and ψ 'express the same thing in T ' iff $(T;\phi) = (T;\psi)$.

But now could it make sense to think of such sets representing proposition contents autonomously — independently of any specification of a theory? They would somehow themselves have to embody an equivalence relation modulo which they were an equivalence class. Do they? What are the sets $(T;\phi)$ like? The boolean properties of classical logic easily yield the following:

Lemma I.4.1: If $X = (T;\phi)$ for some theory T and sentence ϕ , then:

$$(1) \quad \psi, \chi \in X \Rightarrow \psi \wedge \chi \in X$$

$$(2) \quad \psi, \chi \in X \Rightarrow \psi \vee \chi \in X$$

$$(3) \quad \psi, \chi \in X \quad \text{and} \quad \vdash \psi \rightarrow \omega \quad \text{and} \quad \vdash \omega \rightarrow \chi \Rightarrow \omega \in X$$

Let us call any X which is non-empty and satisfies the closure conditions (1), (2) and (3) a logically convex set. It turns out that convexity is precisely the property we want in order to show that the equivalence classes we have been considering do indeed have a life of their own: we can show that any convex set is an equivalence-class modulo some theory T , furthermore that there is only one such, and that it is axiomatizable by the set of biconditionals between elements of the convex set. Bringing this together with the lemma just stated, we have:

Theorem I.4.2: X is logically convex iff there is a theory T and a sentence ϕ such that $X = (T;\phi)$. Given X , T is unique and can be axiomatized by the set $\{\psi \leftrightarrow \chi \mid \psi, \chi \in X\}$.

Proof: If $X = (T;\phi)$, we know that X is convex. Conversely if X is convex, let T be the theory described in the theorem and take $\phi \in X$. That $X \subseteq (T;\phi)$ is immediate. Now say $\psi \in (T;\phi)$: we must show that $\psi \in X$.

First note that for any $\chi \in T$,

$$X \vdash \chi \quad \text{and} \quad \neg X \vdash \neg \chi, \quad \text{where} \quad \neg X = \{\neg \omega \mid \omega \in X\} .$$

But $T \vdash \phi \rightarrow \psi$ and $T \vdash \psi \rightarrow \phi$. Hence:

$$X \vdash \phi \rightarrow \psi \quad \text{and} \quad \neg X \vdash \psi \rightarrow \phi .$$

And so by compactness:

$$\chi_1, \dots, \chi_m \vdash \phi \rightarrow \psi \quad \text{and} \quad \neg \omega_1, \dots, \neg \omega_n \vdash \psi \rightarrow \phi \quad \chi_i, \omega_j \in X .$$

Hence:

$$\vdash (\bigwedge_i \chi_i \wedge \phi) \rightarrow \psi \quad \text{and} \quad \vdash \psi \rightarrow (\bigvee_j \omega_j \vee \phi) .$$

From the conditions for convexity it follows that $\psi \in X$. Thus $X = (T; \phi)$.

To see that T is unique, first it is clear that if $X = (T'; \phi)$ for some ϕ , then $T \subseteq T'$. Conversely say $\psi \in T'$, and pick an element ϕ of X , then:

$$T' \vdash (\psi \wedge \phi) \leftrightarrow \phi \quad \text{and} \quad T' \vdash \phi \leftrightarrow (\psi \rightarrow \phi)$$

Hence $\psi \wedge \phi$ and $\psi \rightarrow \phi$ are both in X , and so:

$$T \vdash (\psi \wedge \phi) \leftrightarrow (\psi \rightarrow \phi) .$$

But then $T \vdash \psi$, and so we have shown that $T' \subseteq T$. Thus $T = T'$, and uniqueness has been proved. \square

This means that we may abandon relativization to particular theories and consider the whole class of convex sets of sentences from a particular language. These sets play an autonomous proposition-content-expressing role — a linguistic role just like sentences themselves. And their propositional content under a fixed interpretation of the underlying language, or their range of propositional contents as interpretations for the language vary, may be explained in the following way: a convex set expresses what an arbitrary element of the set would express under the assumption that all the elements in the set are in any case equivalent. Hence they are not to be thought of conjunctively, as we might think of the content of a theory, nor yet disjunctively, rather interjunctively: a convex set in this role is a kind of (except in special cases infinitary)

interjunction. When is a convex set true? When all its elements are true. When is a convex set false? When all its elements are false. So they may be neither — they are like sentences in a partial-valued language.

We can make this quite precise: consider a model theory for a language L of the kind we have been discussing, where models M assign the value \top or \perp to sentences of L in accordance with the boolean truth-tables (write $M(\phi)$ for this value), and furthermore where the relation, \vdash , is fully characterized model-theoretically by

$$\Gamma \vdash \phi \iff \begin{cases} \text{there is no } M \text{ such that} \\ M(\psi) = \top \text{ for all } \psi \in \Gamma \text{ and } M(\phi) = \perp . \end{cases}$$

How the models are actually defined may depend on the nature of the language L : we need know nothing more about the matter than what we have stated. We can now define the partial assignment that models for L make to convex sets of sentences of L :

$$M(X) = \top \iff M(\phi) = \top \text{ for all } \phi \in X$$

$$M(X) = \perp \iff M(\phi) = \perp \text{ for all } \phi \in X$$

This definition determines the semantical character of convex sets.

It should be clear now why I insisted on ruling out \emptyset in the definition of a convex set: this is a pathological case which would have to be both true and false with respect to any model. The smallest convex sets over L are the equivalence classes modulo logical equivalence, viz. the sets $(\bar{\phi}; \phi)$: these are always totally defined with respect to any M — let us call them boolean. There are two particularly interesting cases: the set of logical truths, and the set of logical falsehoods, which are both clearly convex and boolean and correspond to the constant sentences \top and \perp used in the previous section.

On the other hand there is a unique biggest convex set: the set of all sentences of L — this is the only convex set which ever contains any sentence and its negation, and it is of course undefined for all M . It is like our constant sentence $*$.

In between we may have convex sets of various size — the bigger, the less defined:

$$X \supseteq Y \iff M(X) \subseteq M(Y) \quad \text{for all } M.$$

These remarks are easy to check.

But what can we do with convex sets? By way of example, suppose that L contains first order quantifiers and that we are interested in Strawsonian truth-falsehood conditions for 'the F is G '. Classical logic does not provide sentences that are any good, but now we have convex sets to play with as well. Let $T_{E!F}$ be the theory with the single axiom $\exists x \forall y (x = y \leftrightarrow Fy)$, and consider the convex set $(T_{E!F}; \forall x (Fx \rightarrow Gx))$; we may note that this is the same set as $(T_{E!F}; \exists x (Fx \wedge Gx))$, and that there are many other specifications too. A general fact worth knowing in this connection is that

$$M((T; \phi)) \text{ is defined iff } M \text{ is a model of } T$$

while if M is a model of T , then

$$M((T; \phi)) = M(\phi) .$$

Hence the convex set $(T_{E!F}; \forall x (Fx \rightarrow Gx))$ has the right truth-falsehood conditions for a Strawsonian construal of 'the F is G ': it is just like the partial-valued sentence $\text{I}xFx \cdot Gx$; and we can easily make up more examples, e.g. a convex set corresponding to the quantifier $\forall x Fx \cdot Gx$.

It is all very well specifying convex sets in this kind of way — in terms of the structure of the underlying language, invoking the fact that they are all $(T; \phi)$ for some T and ϕ — but if we want to see them behaving as sentences do, then we should discern operations holding directly between convex sets themselves. Most obviously we should expect the boolean

structure of L to induce corresponding operations. Since a convex set expresses what its elements indiscriminately express, we can define negation, conjunction and disjunction simply by applying the operations of L 'element-wise' — however we must also ensure that the resulting set is logically convex. This is not difficult: the conditions for convexity yield a closure operation, $X \mapsto X^c$, where X^c is the smallest convex set containing X , i.e. the intersection of all convex sets containing X , which is itself clearly convex. We also have an explicit characterization of X^c :

$$\phi \in X^c \iff \begin{cases} \text{there exist } \psi_1, \dots, \psi_m, \chi_1, \dots, \chi_n \in X \\ \text{such that } \vdash \bigwedge \psi_i \rightarrow \phi \text{ and } \vdash \phi \rightarrow \bigvee \chi_j. \end{cases}$$

We may state the required definitions as follows. (The ambiguity with \neg , \wedge and \vee will be harmless.):

$$\neg X = \{\neg \phi \mid \phi \in X\}^c$$

$$X \wedge Y = \{\phi \wedge \psi \mid \phi \in X, \psi \in Y\}^c$$

$$X \vee Y = \{\phi \vee \psi \mid \phi \in X, \psi \in Y\}^c$$

It is now straightforward — but tedious — to check that we have the right truth-falsehood conditions. For any model M for L :

$$M(\neg X) = \top \iff M(X) = \perp$$

$$\perp \qquad \qquad \top$$

$$M(X \wedge Y) = \top \iff M(X) = \top \text{ and } M(Y) = \top$$

$$\perp \iff \perp \text{ or } \perp$$

$$M(X \vee Y) = \top \iff M(X) = \top \text{ or } M(Y) = \top$$

$$\perp \qquad \perp \text{ and } \perp$$

However there is still something missing — the easiest but most important operation of all:

$$X * Y = (X \cup Y)^c$$

X and Y are already the 'interjunction of their elements', so all we do here is throw everything in together to get a bigger 'interjunction': and we have:

$$M(X \times Y) = \top \iff M(X) = \top \text{ and } M(Y) = \top$$

$$\perp \iff M(X) = \perp \text{ and } M(Y) = \perp$$

The point of all this has been to see how we might get the effect of partial-valued logic using classical logic and a few definitions; though what arises is a very special kind of partial-valued system: one with totally defined elements, the boolean sets, at bottom. We shall return to the relation between total- and partial-valued systems in Chapter III when we produce the promised representation theorem showing how any partial-valued system (i.e. theory in partial-valued logic) determines a particular system of convex sets in which it may be interpreted 'faithfully' and 'conservatively'.

In preparation for this result we now consider a special — simple — kind of convex set. We know that a convex set X determines a unique theory, T_X say, modulo which it is an equivalence class: let us say that X is finitely specifiable iff T_X is finitely axiomatizable. This means, of course, that there is in fact a single sentence ψ such that $X = (\overline{\{\psi\}}; \phi)$ for some ϕ , where $\overline{\{\psi\}}$ is the theory axiomatized by ψ . Now, if ϕ and ψ are sentences of L , let $[\phi; \psi]$ be defined by:

$$[\phi; \psi] = \{\chi \mid \vdash \phi \rightarrow \chi \text{ and } \vdash \chi \rightarrow \psi\}$$

then it is easy to check that

$$(\overline{\{\psi\}}; \phi) = [\psi \wedge \phi; \psi \rightarrow \phi]$$

and, provided that $[\phi; \psi]$ is non-empty (i.e. $\vdash \phi \rightarrow \psi$), that:

$$[\phi; \psi] = (\overline{\{\phi \leftrightarrow \psi\}}; \psi) = (\overline{\{\phi \leftrightarrow \psi\}}; \phi) .$$

Hence the finitely specifiable convex sets are precisely the sets $[\phi; \psi]$

where $\vdash \phi \rightarrow \psi$. It is revealing to compare these two equations with the interdefinability of \ast and $/$ specified in the last section: $[\phi; \psi]$ is like a formula in the normal form we described, with ϕ giving truth conditions and ψ falsehood conditions, while $(\overline{\{\psi\}}; \phi)$ is ' ψ/ϕ '.

Now, the finitely specifiable convex sets constitute a genuine subsystem. Firstly, ' \top ', ' \perp ' and ' \ast ' are finitely specifiable, since we have $[\phi \vee \neg \phi; \phi \vee \neg \phi]$, $[\phi \wedge \neg \phi; \phi \wedge \neg \phi]$ and $[\phi \wedge \neg \phi; \phi \vee \neg \phi]$, where ϕ is any sentence of L ; and, further, there is closure under \neg, \wedge, \vee and \ast , as the following equations show:

$$\neg[\phi; \psi] = [\neg\psi; \neg\phi]$$

$$[\phi; \psi] \wedge [\chi; \omega] = [\phi \wedge \chi; \psi \wedge \omega]$$

$$[\phi; \psi] \vee [\chi; \omega] = [\phi \vee \chi; \psi \vee \omega]$$

$$[\phi; \psi] \ast [\chi; \omega] = [\phi \wedge \chi; \psi \vee \omega]$$

Moreover, it is interesting to see that finitely specifiable convex sets (unlike convex sets in general) are also closed under operations corresponding to ' $T(p)$ ' and ' $\neg T(\neg p)$ ' ('it is true that p ' and 'it is not false that p ') — modes of composition which are not monotonic. These operations are given by:

$$[\phi; \psi] \rightarrow [\phi; \phi] \quad \text{and} \quad [\phi; \psi] \rightarrow [\psi; \psi] .$$

In fact this property characterizes finite specifiability:

Lemma I.3.3: X is finitely specifiable iff there are convex sets Y and Z such that for any M for L :

$$\begin{array}{ll} M(X) = \top \iff M(Y) = \top & M(X) \neq \perp \iff M(Z) = \top \\ M(X) \neq \top \iff M(Y) = \perp & \text{and} \\ M(X) = \perp \iff M(Z) = \perp & \end{array}$$

Proof: 'only if': if $X = [\phi; \psi]$, then put $Y = [\phi; \phi]$ and $Z = [\psi; \psi]$.
 'if': note first that Y and Z must be boolean, hence identical to $[\phi; \phi]$ and $[\psi; \psi]$ for some ϕ and ψ . This means that X has to be $[\phi; \psi]$.

To conclude we give three examples of non-finitely-specifiable convex sets, lacking respectively a Y , a Z and either a Y or a Z with properties described above. Let L be a propositional language with atoms p_0, p_1, p_2, \dots and consider

$$\{p_0, p_0 \wedge p_1, p_0 \wedge p_1 \wedge p_2, \dots\}^c$$

$$\{p_0, p_0 \vee p_1, p_0 \vee p_1 \vee p_2, \dots\}^c$$

$$\{p_0, p_1, p_2, \dots\}^c.$$

CHAPTER II

PARTIAL-VALUED LANGUAGES (2) – TERMS AND QUANTIFIERS

II.1 EXTENSIONALITY

It has been interesting, I think, to restrict analytical attention to the sentential articulation of logic: we have seen how undefinedness can occur in virtue of modes of sentence composition. But in partial-valued languages – languages proper, in contrast to systems of convex sets – we have, of course, been allowing for undefinedness in sentential units at bottom, ones themselves unanalysed by propositional logic. In a schematic presentation of logic, where basic atoms are seen as amenable to arbitrary substitutions, this had to be the case given that there were complex sentences which could lack a value; but it also leaves the door open for the treatment of gaps arising in virtue of the internal analysis of basic sentences. Indeed, that gaps can occur in this way is a much more familiar idea.

Predicate/singular-term composition has been taken to give rise to truth-value gaps, and consideration in the literature has been given to two different ways in which this might happen:

- (i) because some constituent singular term lacks a denotation;
- (ii) because the predicate is not actually true or false of some individual or n -tuple of individuals.

Let us consider predicates $\phi(x_1, \dots, x_n)$ (simple or complex) defined (even if only partially) over a given domain D and singular-terms (simple or complex) denoting (or not denoting) elements of D . How might a thorough-going partial-valued semantics cater for these kinds of undefinedness?

We might ask about extensionality: when is a predicate $\phi(x_1, \dots, x_n)$ extensional? In a consideration of this question it is clear that the two

kinds of gap-causing phenomena cannot be treated independently, if we want a general answer, since partially defined predicates may be applied to denotationless terms. The question of the extensionality of a mode of predicate composition $\phi(x_1, \dots, x_n)$ is directly parallel to the question of the truth-functionality of $\phi(p_1, \dots, p_n)$, only now we have terms denoting or not denoting elements of D as the constituents to be compounded, instead of sentences bearing or lacking a truth-value.

We could run through the same line of thought as we did in section II.2. But, better, let us offer straightaway a dependency criterion of value (\top or \perp) on value (elements of D). This can be taken over mutatis mutandis from the formulation given for partial truth-functionality:

$\phi(x_1, \dots, x_n)$ is extensional iff the value (\top or \perp), if any, of a sentence $\phi(t_1, \dots, t_n)$ depends at most on (is determined at most by) the denotations of those t_i that denote, if any of them do.

As before, this yields a substitutivity criterion, which letting α range over \top, \perp and d over D , we can state:

$$\forall d (t_i \text{ denotes } d \Rightarrow t'_i \text{ denotes } d) \Rightarrow$$

$$\forall \alpha (\phi(t_1, \dots, t_i, \dots, t_n) \text{ is } \alpha \Rightarrow \phi(t_1, \dots, t'_i, \dots, t_n) \text{ is } \alpha).$$

Now, in the first place, this condition shows that the classification of $\phi(t_1, \dots, t_n)$ as \top or $*$ or \perp depends on nothing over and above the classification of the t_i as denoting a given object in D or as being denotationless. Hence, obvious kinds of 'intensionality' are ruled out. And, if we let \ominus be assigned to a denotationless term ($\ominus \notin D$), then the role of an extensional mode of composition $\phi(x_1, \dots, x_n)$ may be represented by a function

$$(D \cup \{\ominus\})^n \rightarrow \{\top, *, \perp\}$$

However, we wish to see the assignment of \ominus to a term as representing the lack of a denotation, rather than as being some special entity picked on

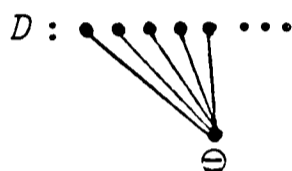
Frege-style to be the actual denotation of a term that would otherwise be denotationless.

On the other hand, assigning Θ to a term is not to mean that it is meaningless in any way. This can be seen to make sense, I think, when we look again at the substitutivity criterion and work out from it the constraint determining precisely which functions $(DU\{\Theta\})^n \rightarrow \{T, *, \perp\}$ are admissible as representations of the interpretation of an extensional predicate.

Let \sqsubseteq now be used also for the ordering on $DU\{\Theta\}$ defined by

$$x \sqsubseteq y \iff \forall d \in D \cdot x = d \Rightarrow y = d$$

and given by the diagram:



Then it is all and only

functions monotonic with

respect to \sqsubseteq in the new sense and \sqsubseteq in the old sense on $\{T, *, \perp\}$ that are admissible. We may take such functions to represent (be a set-theoretical representation of) the kind of function we need to provide extensions for n -place partial predicates. And so deploying Θ is nothing but part of this representation.

We can consider, too, another functional category which is made much use of in familiar first-order languages, viz. functors $t(x_1, \dots, x_n)$ combining singular terms to yield another singular term. What is extensionality for such modes of composition (simple or complex)? This time it is a matter of considering partial functional interpretations from D^n into D . The arguments do not bear repeating yet a third time: obviously such a mode of composition is extensional if and only if its behaviour can be specified by a function $(DU\{\Theta\})^n \rightarrow DU\{\Theta\}$ which is monotonic with respect to \sqsubseteq on $DU\{\Theta\}$. Then and only then can we say that the denotation of $t(t_1, \dots, t_n)$, if any, depends in the appropriate way on the denotations, if any, of the t_1 .

for "where" and "were"

Extending the terminology of section I.2, a 'crude extensionalist' would be someone whose criterion of functional dependency were less subtle than ours in virtue of a ruling that if t were denotationless then any sentence in which it occurred must lack a truth-value and any complex term in which it occurred must also lack a denotation. To pick up the matter ignored in section I.2: was Frege a crude extensionalist? He certainly wished to discern in natural language cases of singular terms which, while absolutely sense-full, lacked a denotation: and he considered sentences $\phi(t)$ which lacked a truth-value because t was non-denoting (though sense-full). He seems also to be credited with the general view that for any t and $\phi()$, if t is non-denoting, then $\phi(t)$ lacks a truth-value. Now if he did in fact hold this general view, then why did he, and should he have?

In her discussion of truth-value gaps in a Fregean context, Haack (1974) converts her more tentative remarks at the bottom of page 60, which deal just with truth-values, into the following argument:

"The consequences of the sense/reference theory for the question of non-denoting terms can be derived³ from the principles:

- (1) that all expressions, both sentences and their components, have both sense and reference,
- (2) the reference of a proper name being the object denoted, and the reference of a sentence being its truth-value, and,
- (3) that the reference of a compound expression depends on the references of its parts.

It follows from these principles that if a sentence contains a singular term which lacks a reference, then the sentence itself must lack reference, that is, must be without a truth-value.

On Frege's semantic theory, then, a sentence containing a non-denoting term, though having a perfectly good sense, lacks truth-value."

Let us accept (1), (2) and (3) — though (1) is a little oddly expressed, considering what the issue is. However, her conclusion surely does not

follow if we are sufficiently subtle over the understanding of (3), viz. concerning what functional dependency is between truth-values and the denotations of singular-terms.

Certainly, for most naturally occurring basic predicates $P(x)$, we will have $P(t)$ undefined if t is non-denoting, and, indeed, by monotonicity this could only fail to be the case if $P(x)$ took the same value for all elements of the domain; but we do not have to accept the crude-extensionalist's ruling as a general semantical principle, since a predicate $\phi(x)$ might be constantly \top or \perp — not merely 'de re' constant (constant on D), but absolutely constant. Put another way, $\phi(x)$ could be just totally independent of x . Consider how this might arise if we wanted to define a relation $R(x, y)$ by cases; say:

$R(x, y)$ is true if $P(y)$ is true
 is false if $P(y)$ is false and $Q(x)$ is false
 (and is undefined otherwise).

Clearly $R(x, y)$ is well defined, and the value of a sentence $R(t_1, t_2)$ will depend on nothing more than the values of $P(t_2)$ and $Q(t_1)$. We can assume that P and Q do themselves satisfy the crude extensionality condition. Yet say $P(t_2)$ is true, then $R(t_1, t_2)$ will be true, regardless of whether t_1 denotes or not — even if t_1 's being non-denoting means that $Q(t_1)$ is undefined. Of course, using the sentence connectives allowed in Chapter I, we could define $R(x, y)$ explicitly by:

$$(Q(x) \vee P(y)) * P(y) .$$

And, in general, since compounding monotonic functions invariably yields another monotonic function, we can use the truth-functional connectives defined in the last chapter to form complex predicates in such a way that, if the predicates we start off with are extensional, we shall never get

taken out of the class of extensional predicates. However, cases of absolute and relative independence of argument places will very easily and naturally arise, whether they are there at bottom or not, since the 'strong' tables for \wedge and \vee are allowed.

But to get back to Frege: the following footnote in Haack I think is telling:

"Frege himself does not use the sense/reference theory to establish this conclusion, but, rather, appeals to the intuitiveness of the claim that such a sentence as

'Odysseus was set ashore at Ithaca while sound asleep' while having a perfectly good sense, cannot be assigned either truth-value, in support of his thesis that the reference of a sentence is its truth-value. This difference of procedure is not, however, important for the present purpose."

I disagree with the last sentence: I think the difference in procedure is crucial. All he was considering was a particular case were $\phi(t)$ might lack a reference because t lacked one, and the general principle of crude extensionality does not follow from that. Admittedly, in this passage (in Frege (1892)) we find "Und Sätze, welche Eigennamen ohne Bedeutung enthalten, werden von der Art sein." — where the Art in question is having-no-Bedeutung — and this appears to be a statement of the general principle. However it is a rather casual one; and what is the argument for it?

If crude extensionality is to be justified by appeal to Fregean doctrine, perhaps we have to consider more than simply the reference (Bedeutung) of singular-terms and sentences. As Haack states under (1), all expressions have a reference according to Frege, including items of a functional category, viz. modes of composition themselves. Their references are functions. Now, in the discussion so far, we have simply been taking functions that interpret modes of composition to be what manifest dependence between resultant truth-values (or denotations) and constituent denotations

or truth-values; but if, according to Frege, the truth-value of a sentence 'depends' on the reference of all its parts, including items of a functional category, then talk of dependency must be 'pushed up a level'. And perhaps Fregean motivation for crude extensionality follows from views concerning what the functions interpreting modes of composition have to be like?

Dummett (1973, p.170) suggests that Frege would be unwilling to countenance the existence of partially defined functors — for example predicates that were true or false of certain objects, but undefined otherwise — because, thinking by analogy with the reference of names, (i.e. singular-terms in general), a functor would have either to refer to a totally defined function or refer to nothing at all. Certainly, then, if a predicate P 'had no reference', a sentence $P(t)$ would reasonably be taken to lack a truth-value; and equally if P did have a reference (a total function) but t had no reference, then, again, $P(t)$ would reasonably be taken to have no truth-value — because applying what had to be totally defined to nothing could yield nothing. Hence, if all functor categories had either to be totally undefined or have totally defined interpretations, then certainly crude extensionality would follow.

There are two difficulties, however, with this diagnosis. In the first place Dummett seems to be suggesting that it is because 'the presence of a name which lacks a reference deprives any sentence of a truth-value' that Frege was inclined to push the analogy between name-reference and functor-reference to the unnatural lengths of excluding partially defined predicates: we have gone in a circle. And secondly, the alternative view Dummett offers — one he says would be unfavourable to Frege — is not the alternative that has been offered here. For Dummett suggests that the natural response would be to say that partially defined functors had an 'incompletely specified reference' — one that could be filled out in various

ways to obtain various total-valued functions. However, our idea concerning the functional interpretation of modes of composition is not this: the extensional interpretation — reference, if you like — of a mode of composition is completely specified, but the specification may be of a partially-defined function. What is involved here is not merely a different way of looking at what is formally the same apparatus: the system of partially defined functions that we have defined — those representable by total monotonic functions — is a richer system than what would be obtained by taking partial interpretations to be like total interpretations, only possibly undefined for some arguments. And the difference is precisely that the system so obtained would in fact be crudely extensional.

This line of thought, then, would not seem to provide a knock-down Fregean rationale for crude extensionality.

Of course, when Frege came to set up his formal extensional language, he in any case eschewed undefinedness entirely: all basic names were denoting and he went to great — and often artificial — lengths to ensure that all functors were totally defined, so that every item in a name category (singular term or sentence), simple or complex, had to have a value. He did not see how logical laws might be given otherwise (Frege (1891)). It would be nice to think that, with this ruling in mind, Frege just did not address himself seriously to the question of functional dependence in general — in such a way as to include partial-valued contexts — and that if he had, he would have come up with an extensional semantics something like the one I am proposing. For this would seem to me to be the most satisfactory way to make systematically coherent a semantics for sense-without-denotation: a way that both preserves what is essential to Frege's systematic ideas and, at the same time, encompasses those gappy features of language which he took to be recalcitrant.

Dummett, however, seems to consider that crude-extensionality figured large in Frege's thought and appeals to this principle to explain why undefinedness was eschewed (p.185, op. cit.). I am insufficient of a Frege scholar to dispute that this is the best understanding of Frege, but there is a lot of Dummett's own thought in this passage that I would not go along with — if, that is, I have understood it correctly. Certainly I take the point that it would have been reasonable of Frege to despair of setting up a worth-while semantics if he were committed to a principle of crude extensionality: there would just be so many gaps. And I take the point, too, that Frege, faced with a sentence that was (intuitively) not either true or false because of the presence of a constituent singular-term which lacked a denotation, would say that the sentence lacked a truth-value, rather than that it had a third intermediate one. But I am sceptical of the importance of crude extensionality in a Fregean kind of semantics.

First a subsidiary worry. It would seem to me that, given a sentence which appeared to be neither true nor false, because it contained a denotationless name, the principle that truth-values were the reference of sentences just as denotations were the reference of singular terms, together with the general principle that compound reference depended on constituent reference, would alone have been sufficient to compel Frege into the no-value position rather than the intermediate-value one. Dummett, however, seems to suggest that the reason was the general principle of crude extensionality itself.

Secondly, a more important point. The major solution that Dummett has to offer, in order to allow us to work with denotationless names without being hampered by crude extensionality, is precisely the un-Fregean line that 'neither-true-nor-false' is in fact an intermediate truth-value. For then, he says, even if atomic sentences containing empty names have to be

neither true nor false (so take the third value), it does not follow that any complex sentence has to be too: arbitrary three-valued matrices are admissible, from among which we might choose plausible interpretations for sentence-connectives that issue in true or false sentences even when constituent sentences are neither.

In response to this, I would of course want to urge that such interpretations are available anyway, even if we follow Frege in taking there to be only two truth-values and so understand the classification 'neither-true-nor-false' as a lack of any truth-value; nor do we have to give up the Fregean idea of the reference of names (denotation), if any, determining the reference (truth-value), if either, of sentences. For, according to our principles of functional dependence, while not all three-entry matrices are admissible, still the supply is sufficiently rich to achieve the advantage that Dummett sees in the three-valued idea, viz. no general ruling that t non-denoting implies $\phi(t)$ neither true nor false. The third classification can remain a gap, but not a crippling one. And note that Fregean reasons for discerning a genuine gap are not undermined, if, as I suggested, these need not depend on the principle of crude extensionality itself.

In the following paragraph (p.186) Dummett runs over other possibilities, and the general conclusion is that 'the two guiding ideas of a theory of reference come into sharp conflict'. I take the supposed conflict to be between: (i) the notion of truth-value 'needed in a semantics of a standard kind' (i.e. 'semantic role'), according to which we could just have an intermediate truth-value; and (ii) truth-values as bearing the same relation to sentences as bearers do to names. I do not, however, see that there is any sharp conflict here at all: recall the discussion in the general introduction. We have seen that a system of three-valued matrices meeting the monotonicity constraint constitutes a system of

functional dependencies between two of the three matrix entries – 'true' and 'false'. If a sentence is true or false it makes a contribution towards determining the truth or falsehood, if either, of a compound sentence, whereas, if it is neither true nor false, it makes no contribution towards such a determination: this is because the semantic contribution of a mode of composition has been fully specified once the dependencies it exhibits between the two classifications 'true' and 'false' have been specified. Hence we are but one trivial step away from seeing such matrices as representations of genuine partial functions between just the two truth-values 'true' and 'false'. Hence, not only is the analogy

sentence : truth-value :: name : bearer

maintained, but, *via* predicate composition, the interaction of the system of truth-values and the domain of name-bearers is revealed – the truth-values 'true' and 'false' depending on bearers, if any, of names. The only thing we have given up is crude extensionality, and this is what we wanted to do.

But there is another problem: Dummett finally suggests that it was precisely because of the truth-value/name-bearer analogy that crude extensionality appeared compelling to Frege – because, that is, a compound name lacks a bearer if any constituent name does. Well, if intuitive examples of compound names do indeed all work in this way, then I suggest that it is this phenomenon that is most naturally ignored in a systematic partial-valued semantics: terms-out-of-term(s) forming functors should be treated more liberally, as we outlined earlier. And, of course, such a treatment does not involve abandoning the idea of dependency of denotation on denotations: the monotonicity constraint on admissible interpretations of such functors means that the classification 'non-denoting' need not itself be understood as a special kind of denotation: the contribution towards determining a compound denotation is made by constituent terms which have a

for "pradign" and "pradigne"

denotation, and any constituent term which does not simply makes no contribution. The semantic contribution of the functors themselves is completely specifiable in terms of the dependencies they reveal simply between objects denoted. Hence we are not, in offering this semantics, committed to a level of interpretation on which 'non-denoting' is seen as a 'semantic role' on a par with 'denoting-so-and-so' (just as a third truth-value would be on a par with truth and falsehood): this would be quite unnecessary; we can make do with objects themselves.

It is very important, I think, to contrast the discussion here with Woodruff (1970, pp.128-9). He seems to be trying to reconcile the use of 'strong' connectives, which violate crude truth-functionality, with a general Fregean way of thinking, not by arguing, as I am, that there is in fact no trouble over the functional dependence of reference on reference, but rather by arguing that although direct dependence may break down, still that does not matter. The idea seems to be that, providing the constituent items of a sentence all have a sense, including ones without a reference, then we at least have a compound sense for the whole sentence; and so this sense can be considered as determining the reference (truth-value). However this detour through sense is unnecessary: purely extensional functional interpretations, correctly understood as partial functions, show that dependency does not break down.

Fregean singular-terms, then, with a (descriptive kind of) sense but no denotation seem to be the paradigm examples of terms falling within our classification Θ — e.g. non-uniquely specifying description terms of the kind whose classification is determined simply by the meaning of the constituent vocabulary. And crude extensionality, I have argued, is not motivated by such cases. However, consideration of other kinds of singular terms that occur in natural language, and may be vacuous, might appear to suggest the cruder criterion governing gap-conditions. The most important

suggestion, I think, would come from the behaviour of singular terms whose function is taken to be to act as a medium for 'direct reference', we might say — terms which would be taken to have no semantical contribution at all unless they had a referent, this referent, if there is one, being determined by something external to the workings of a systematic semantics, i.e. neither determined directly by any associated Fregean sense, nor by being the resultant denotation of a compound of denotations.

Demonstratives (considered with suitable relativization to occurrences of use) might fall in this category, and so too, I take it, might proper names on the popular Mill-Kripke view. In such cases, if there were in fact no object referred to, then the term in question might plausibly be considered semantically defective in such a way as to infect similarly any sentence in which it had a straight-forward kind of occurrence: crudely, something would have gone wrong with the mechanisms of indexical reference, and an ingredient would be lacking in what would otherwise have contributed essentially to anything properly taken to be an understanding of the meaning of the sentence (relative to an occasion of use, if necessary). However, for this very reason, such cases should not, I think, prompt a reconsideration of our extensionality criterion: they simply fall outside our 'terms of reference', since Θ is not meant to signify 'semantic deficiency' of any kind.

When we come to set up our formal languages, there will certainly be non-complex items in the category of singular terms, which may be assigned the classification Θ in a model for the languages: if for no other reason we do this on the general principle that no basic item in a semantical category should have imposed on it, as a matter of logic, the arbitrary restriction that it be denied a semantical classification which a complex item of that category may take. But assigning Θ to these 'names' is in no way intended to represent the kind of radical failure of reference

which we have been considering. As it stands, our semantics does not touch on such problems of reference at all: all singular terms that come within the scope of our logic are such that their denotation, if any — and hence whether or not they have a denotation — is determined solely by the interpretation of their constituents, which is itself assumed to be referentially unproblematic. Admittedly this means that when we come to deploy our simple model theory the limiting case of unstructured singular terms may look odd, since models will provide only extensional 'interpretations', and so the classification '(meaningful but) non-denoting' will have to be directly stipulated, just like the classification 'denoting so-and-so'. However assigning \ominus to a term is not to be explained as representing failure of extra-systematic reference: if you want a picture, think rather of models as the extensional distillation of a full intensional semantics which would assign such terms a sense (intension) which is such that there is no denotation.

There is another point too concerning crude extensionality, this time involving predication itself: whatever we may want to say in general about the range of admissible extensions for predicates, still, might it not be the case that any naturally occurring atomic predicate has to be crudely extensional? Recall that we allowed for this possibility earlier in the discussion. Now, if to explain the evaluation of atomic constituents we say that $P(t)$ is true (false) iff $P(\)$ is true (false) of the object denoted by t , then this in itself does not seem to lead to crude extensionality, since we could argue that the relation ' X is true of x ' might itself for some values of X be constantly true regardless of x — so that putting a vacuous description 'the object denoted by t ' in for x might still yield a truth. Yet two things do seem to be plausible.

(i) At least over atomic predicates Frege must have been self-consciously crude: we have "Wer eine Bedeutung nicht anerkennt, der kann ihr ein Prädikat weder zu- noch absprechen." – a quotation from the same passage as before. As it stands, this could read as a reflection on arbitrary predicates, atomic or complex; but, evaluating a sentence Frege-style, it is only at bottom with atomic predicates that terms play a direct role, and so if the quote is to reflect on semantic evaluation, it would seem to be at the atomic stage that it is seriously relevant.

(ii) Constantly true (or false) atomic (one-place) predicates – extensionally speaking there are only two – would indeed seem unlikely to occur in a natural language: what possible point could they have in our vocabulary?

In spite of this I shall be allowing for atomic predicates taking these constant extensions, for the same reason that I want to be able to assign Θ to unstructured names: I want to be able to map out logical relationships schematically, so that all items in a given category satisfy the same principles. This does not mean to say that, when it comes to setting up particular theories governed by our logic, we shall not be able to specify constraints if we want to.

Analogous thoughts probably apply to ≥ 2 -place atomic predicates, though *a priori* it is less easy to see why it would be pointless to have, say, a two-place predicate which for some values to one of the argument places produced what was a constant one-place predicate with respect to the other. However, I can think of no convincing examples.

So far we have addressed ourselves only to extensionality of the most simple kind: concerning functors taking sentences into sentences, singular terms into sentences, and singular terms into singular terms. But what of higher level functional categories? Starting with sentences

and singular terms, we can define a hierarchy of categories: given any categories, K_0, K_1, \dots, K_n we have a category into which a mode of composition $F(X_1, \dots, X_n)$ would fall if it took items U_1, \dots, U_n of category K_1, \dots, K_n , respectively, to an item $F(U_1, \dots, U_n)$ of category K_0 . What in general is it for such a mode of composition to be extensional? Of course it depends on the vocabulary and construction of a particular language what categories it may be possible and worthwhile to discern, but we might think about the matter in general.

There are indeed languages which actually contain primitive 'logical' vocabulary from higher level categories — for example first-order quantifiers and term-forming definite description operators: these may, disentangling them from the apparatus of variable-binding, be considered as functors taking one-place predicates to sentences and to singular terms respectively. And consider: if we have primitive vocabulary in these categories and want to be sure that extensionality is preserved throughout, then there is not merely the question of asking what extensionality is in the higher level case, but also the demand that if complex functors of the basic kinds are formed using the higher level vocabulary, then these functors are guaranteed to be extensional in the ways already discussed. For example, with quantifiers and a definite-description operator we could construct such predicates in x as: $\forall y Fxy \quad \forall y Fx \lambda z Gxyz$. But if we get the right answer to the first question, then there should be no problem over this; and conversely we might expect there to be no problem only if we get the right answer.

To start with we can consider the simple case of a one-place mode of composition $\phi(X)$ which forms a sentence $\phi(P)$ when P is a one-place first level predicate (the category that \forall and \exists will fall in). Such a mode will have to be interpreted by a function taking one-place-predicate-

interpretations into truth-values. In total-valued logic $\phi(X)$ would be an extensional context iff the value of a sentence $\phi(P)$ were determined solely by the extension of P – i.e. $\phi(P)$ and $\phi(Q)$ would take the same value whenever P and Q were true and false of precisely the same things. Now in our case straight-forward identity of extension gives way to something more subtle. Let us pursue the matter in the linguistic and material modes at the same time and consider predicates P and Q , whose extensions are given by P' and Q' respectively – monotonic functions $D \cup \{\emptyset\} \rightarrow \{T, *, \perp\}$. We have, induced from the basic orderings \sqsubseteq on $D \cup \{\emptyset\}$ and $\{T, *, \perp\}$, a degree-of-definedness relation, which we can also call ' \sqsubseteq ':

$$P' \sqsubseteq Q' \iff P'(a) \sqsubseteq Q'(a) \text{ for all } a \in D \cup \{\emptyset\}$$

So, $P' \sqsubseteq Q'$ iff Q' is consistent with and no less defined than P' – for short we may say Q' is 'stronger than' or 'more defined than' P' .

Now to require simply that if $\phi(P)$ takes a value, then $\phi(Q)$ takes that value whenever $Q' = P'$ is too weak for extensionality in a partial-valued context. What is important about P , if it contributes to determining a value for $\phi(P)$ in virtue of its extension, is that it takes those elements of D which it takes to truth-values to the particular truth-values it does take them to: and so we must require that any other predicate Q which does this – whatever further elements of D it may also be defined on – can be substituted for P with the result that if $\phi(P)$ has a value, then $\phi(Q)$ has that same value: equivalently, if $P' \sqsubseteq Q'$, then if $\phi(P)$ has a value, $\phi(Q)$ has that value. In other words, we have a monotonicity condition again: a mode of composition $\phi(X)$ is extensional iff the way it determines truth-values is given by a functional

$$F : (D \cup \{\emptyset\} \xrightarrow{\text{monotonic}} \{T, *, \perp\}) \rightarrow \{T, *, \perp\}$$

which is itself monotonic, i.e.:

$$P' \sqsubseteq Q' \Rightarrow F(P') \sqsubseteq F(Q') .$$

We can now turn to the actual interpretation of the first-order quantifiers \forall and \exists ranging over a domain D . Happily all goes well treating them in the obvious way as the (possibly) infinitary analogues of \wedge and \vee , indexed by elements of D as a predicate is applied to them. Interpreting \forall and \exists respectively, we have F_{\forall} and F_{\exists} such that:

$$\begin{aligned}
 F_{\forall}(P') = \top &\iff P'(d) = \top \text{ for all } d \in D \\
 &= \perp \iff P'(d) = \perp \text{ for some } d \in D \\
 F_{\exists}(P') = \top &\iff P'(d) = \top \text{ for some } d \in D \\
 &= \perp \iff P'(d) = \perp \text{ for all } d \in D
 \end{aligned}$$

These functionals are clearly monotonic, as required. What we have are, in fact, precisely the classical conditions, but they are written in terms of \top and \perp so that they make sense when P' is properly partial.

An operator forming description terms falls in the category taking one-place predicates to terms, and, by parallel arguments, in an extensional language it will have to be interpreted over a given D by a monotonic functional $i: (D \cup \{\emptyset\} \rightarrow \{\top, *, \perp\}) \rightarrow D \cup \{\emptyset\}$. What this comes to is followed up in section II.3.

It should be clear that — as we hoped — if monotonic higher level operators are used to form complex first level predicates or first level functors taking terms to terms, then monotonicity for first level categories is preserved. And conversely, if we introduced a non-monotonic operator of higher level, then first level extensionality would break down; for example if O were a non-monotonic simple quantifier, we could easily find a 2-place predicate R interpreted monotonically, but in such a way as to yield a non-monotonic complex predicate $OyRxy$.

Of course we have still only scratched the bare roots of the whole hierarchy of categories. Nonetheless, the ideas just deployed should be

for "categorics" read "categories"

sufficient to see how the general criterion should go: we can define along with the categories themselves the range of extensional interpretations for items in those categories and also the degree-of-definedness ordering \sqsubseteq between such interpretations. Assuming this has been done for categories K_0, K_1, \dots, K_n , the extensionality criterion for an item in the category taking items from K_1, \dots, K_n to an item in the category K_0 is that its interpretation can be represented by a function monotonic with respect to the \sqsubseteq orderings for K_1, \dots, K_n and the \sqsubseteq ordering for K_0 . The new ordering for this category can then be specified by:

$$F' \sqsubseteq G' \iff F'(U'_1 \dots U'_n) \sqsubseteq G'(U'_1 \dots U'_n) \text{ for all } \langle U'_1 \dots U'_n \rangle$$

where the i^{th} coordinates of these n -tuples are (representations of) admissible interpretations for items of category K_i .

Of course, this covers more than the system of categories implicit in Frege, where all functional categories have to be ones forming sentences or singular terms (which he conflated anyway): here K_0 is allowed itself to be a functional category. We shall mention later (in passing only) an interesting instance from a functional category with output in the category of first level predicates.

II.2 LANGUAGES AND MODELS

To lighten the load later, we can now state some definitions. Unambitiously, I shall restrict myself to first-order languages with predicates, term-out-of-term forming functors, quantifiers and an identity predicate, and to the simplest kind of model for such languages. It is with these languages and models that I shall be doing logic in Chapter IV. However, some further ideas will be mentioned briefly at the end of this section. A discussion of description-term forming operators is postponed until the next section, where we shall discover that we in fact lose nothing by ignoring them — except the pleasure of actually stating laws for such terms, which can be done easily and unproblematically.

The only item of stock vocabulary which we have not yet considered is the identity predicate. In the last section predicates, in general, were taken to be defined over a given set D of individuals, and for the simplest kind of model such a set will constitute an underlying domain, with the identity predicate of the formal language interpreted to mean what we mean in our metalanguage by identity over this domain. The only complication is that singular terms may be classified as Θ , and so a model will have to interpret the identity predicate as a monotonic function

$$E : (D \cup \{\Theta\})^2 \rightarrow D \cup \{\Theta\} .$$

In models where D contains at least two elements we shall of course have to set $E(\Theta, a) = E(a, \Theta) = *$, on pain of getting E wrong on D itself; and so, if only for uniformity's sake, even when D has only one element, or none (for there is no need to exclude empty domains), this had better be the case in all models. E , then, would satisfy the crude extensionalist. One further thing: we shall have a constant undefined singular-term, which we write Θ — the ambiguity with ' Θ ' used in the apparatus of representing partial functions is harmless (cf. $*$ and $*$).

What we treat of, then, will at least have the appeal of being exactly like classical first-order logic except that undefinedness is catered for uniformly throughout. Languages, L , will consist of the following:

- (A):
- (1) Two constant sentences: \top and \perp ; and sentential operators: \neg, \wedge, \vee and \ast .
 - (2) Two quantifier symbols: \forall and \exists .
 - (3) A two-place identity predicate symbol: $=$
 - (4) A constant term: Θ .
 - (5) A set Var of denumerably many variables.
- (B):
- (1) A set $Prd(L)$ of predicate symbols
 - (2) A set $Func(L)$ of function symbols
 - (3) A set $Cns(L)$ of constant symbols.

Also we have functions λ and μ from $Prd(L)$ and $Fnc(L)$ into the natural numbers, giving the 'arity' $\lambda(P)$ and $\mu(f)$ of elements $P \in Prd(L)$ and $f \in Fnc(L)$.

All languages must have the vocabulary exactly as specified under (A), and when we consider different languages we actually take these items to be the same objects. On the other hand the vocabulary under (B) varies from language to language: these sets may any of them be empty or of any cardinality you care for. Let us also stipulate that items of vocabulary are all distinct from one another. We define: L' is an expansion of L iff $Prd(L') \supseteq Prd(L)$ $Fnc(L') \supseteq Fnc(L)$ and $Cns(L') \supseteq Cns(L)$.

Now a model for L is a structure M consisting of:

(1) A set D_M (which does not have to be non-empty)

(2) For each $P \in Prd(L)$ a monotonic function

$$P_M : (D_M \cup \{\Theta\})^{\lambda(P)} \rightarrow \{\top, *, \perp\} .$$

(3) For each $f \in Fnc(L)$ a monotonic function

$$f_M : (D_M \cup \{\Theta\})^{\mu(f)} \rightarrow D_M \cup \{\Theta\}$$

(4) For each $c \in Cns(L)$ an element $c_M \in D_M \cup \{\Theta\}$.

The orderings with respect to which the P_M and f_M are monotonic are the \sqsubseteq relations specified in section II.1. Θ must not be an element of D , but it does not matter what it is — it could even be the same thing as the constant undefined term Θ .

So much for the basic interpretations. Let us use the word expression for a finite sequence of items of vocabulary, and now specify how well-formed expressions can be built up and interpreted. First terms: Let $Trm(L)$ be the smallest set X of expressions such that:

(i) $Var \subseteq X$, $Cns(L) \subseteq X$, $\Theta \in X$; and

(ii) if $f \in Fnc(L)$ and $t_1, \dots, t_{\mu(f)} \in X$, then $f t_1 \dots t_{\mu(f)} \in X$

The classification of a term in a model for L must be with respect to an assignment s to the variables: assignments are just functions $s : Var \rightarrow D_M \cup \{\Theta\}$ (free variables are schematic and range over $D_M \cup \{\Theta\}$, not just D_M). We extend s to a function $Trm(L) \rightarrow D_M \cup \{\Theta\}$ in the obvious way:

$$\begin{aligned} \text{(i)} \quad M_s(x) &= s(x) \quad \text{if } x \in Var, \\ M_s(c) &= c_M \quad \text{if } c \in Cns(L), \\ M_s(\Theta) &= \Theta; \end{aligned}$$

and

$$\text{(ii)} \quad M_s(ft_1 \dots t_{\mu(f)}) = f_M(M_s(t_1), \dots, M_s(t_{\mu(f)})).$$

Now sentential formulae: let $Frm(L)$ be the smallest set X of expression such that:

$$\begin{aligned} \text{(i)} \quad \top \in X, \perp \in X, * \in X \quad \text{if } t_1, t_2 \in Trm(L) \text{ then } t_1 = t_2 \in X; \text{ and} \\ \text{if } P \in Prd(L) \text{ and } t_1, \dots, t_{\lambda(P)} \in Trm(L) \text{ then } Pt \dots t_{\lambda(P)} \in X, \end{aligned}$$

and

$$\begin{aligned} \text{(ii)} \quad \text{if } \phi, \psi \in X \text{ and } x \in Var, \text{ then } \neg\phi \in X, \phi \wedge \psi \in X, \phi \vee \psi \in X, \\ \phi * \psi \in X, \forall x\phi \in X \text{ and } \exists x\phi \in X. \end{aligned}$$

We have not mentioned brackets at all, but let us say that officially our language is defined with functor-first construction so that they are otiose. The signs I have written are all used simply to refer to expressions of L , which are themselves never used, only mentioned: and so if we had just replaced ' $t_1 = t_2$ ' by ' $= t_1 t_2$ ', ' $\phi \wedge \psi$ ' by ' $\wedge \phi \psi$ ' etc., then the metalinguistic referring convention would consistently have been to refer to an expression got by stringing together other expressions by the sign got by stringing together, in the right order, signs for those constituent expressions. But we have instead done the more natural thing and, as a result, we shall also be helping ourselves to brackets when necessary,

We can now use our definition of $M_s(t)$ to define the classification $M_s(\phi)$ of a formula ϕ in M with respect to s , though first we

need some more notation: if s is an assignment into $D_M \cup \{\Theta\}$, x is a variable and a is an element of $D_M \cup \{\Theta\}$, let $s(x|a)$ be the assignment defined as follows:

$$\begin{aligned} s(x|a)(y) &= a && \text{if } y = x \\ &= s(y) && \text{if } y \neq x \end{aligned}$$

Then, writing out explicit \top and \perp conditions, we will have the following:

$$M_s(\top) = \top, \quad M_s(\perp) = \perp$$

$$\begin{aligned} M_s(t_1 = t_2) = \top &\iff M_s(t_1) \in D_M, M_s(t_2) \in D_M \text{ and } M_s(t_1) = M_s(t_2) \\ &= \perp \iff M_s(t_1) \in D_M, M_s(t_2) \in D_M \text{ and } M_s(t_1) \neq M_s(t_2) \end{aligned}$$

$$\begin{aligned} M_s(Pt \dots t_{\lambda(P)}) = \top &\iff P_M(M_s(t_1), \dots, M_s(t_{\lambda(P)})) = \top \\ &= \perp \iff P_M(M_s(t_1), \dots, M_s(t_{\lambda(P)})) = \perp \end{aligned}$$

$$\begin{aligned} M_s(\neg\phi) = \top &\iff M_s(\phi) = \perp \\ &= \perp \iff M_s(\phi) = \top \end{aligned}$$

$$\begin{aligned} M_s(\phi \wedge \psi) = \top &\iff M_s(\phi) = \top \text{ and } M_s(\psi) = \top \\ &= \perp \iff M_s(\phi) = \perp \text{ or } M_s(\psi) = \perp \end{aligned}$$

$$\begin{aligned} M_s(\phi \vee \psi) = \top &\iff M_s(\phi) = \top \text{ or } M_s(\psi) = \top \\ &= \perp \iff M_s(\phi) = \perp \text{ and } M_s(\psi) = \perp \end{aligned}$$

$$\begin{aligned} M_s(\phi \times \psi) = \top &\iff M_s(\phi) = \top \text{ and } M_s(\psi) = \top \\ &= \perp \iff M_s(\phi) = \perp \text{ and } M_s(\psi) = \perp \end{aligned}$$

$$\begin{aligned} M_s(\forall x\phi) = \top &\iff M_{s(x|d)}(\phi) = \top \text{ for all } d \in D_M \\ &= \perp \iff M_{s(x|d)}(\phi) = \perp \text{ for some } d \in D_M \end{aligned}$$

$$\begin{aligned} M_s(\exists x\phi) = \top &\iff M_{s(x|d)}(\phi) = \top \text{ for some } d \in D_M \\ &= \perp \iff M_{s(x|d)}(\phi) = \perp \text{ for all } d \in D_M \end{aligned}$$

So note that once bound by a quantifier variables have no more to do with Θ : quantifiers range over the domain D_M only.

The languages specified do not contain $*$ or $/$, nor indeed \rightarrow or \leftrightarrow . These operators may be defined in the following ways:

$$* = \top \ \& \ \perp$$

$$\phi \rightarrow \psi = \neg\phi \vee \psi$$

$$\phi / \psi = (\phi \wedge \psi) \ \& \ (\phi \rightarrow \psi)$$

$$\phi \leftrightarrow \psi = (\phi \rightarrow \psi) \wedge (\psi \rightarrow \phi)$$

In the present context the status of these definitions is, of course, that of metalinguistic abbreviations.

The definition of what it is for an occurrence of a variable x in ϕ to be bound — otherwise free — is exactly as usual: free variables I shall often call parameters. The notation $u(t/x)$ will be used for the term obtained from the term u by substituting the term t uniformly for the variable x in u , and $\phi(t/x)$ for the formula obtained from ϕ by substituting t for all free occurrences of x in ϕ . ' $\phi(t/x)$ ' will be used only when t is substitutable for x in ϕ , i.e. when no occurrence of a variable in t becomes bound in $\phi(t/x)$. Clearly all these notions can be given precise recursive definitions using which we can easily check out some familiar looking semantical lemmas to show that the definitions of $M_s(t)$ and $M_s(\phi)$ are workable and reasonable. First:

Lemma II.2.1 (Relevant variables lemma)

$$(1) \quad \text{If } s_1(x) = s_2(x) \text{ for any } x \text{ in } t, \text{ then } M_{s_1}(t) = M_{s_2}(t)$$

$$(2) \quad \text{If } s_1(x) = s_2(x) \text{ for any } x \text{ free in } \phi, \text{ then } M_{s_1}(\phi) = M_{s_2}(\phi)$$

A sentence shall be a formula with no free variables, and for the set of sentences of a language L we write $Snt(L)$. As a corollary of the preceding lemma we know that if ϕ is a sentence, then $M_s(\phi)$ is totally independent of s and we can simply write $M(\phi)$. In fact we could explicitly define $M(\phi)$ for arbitrary formulae ϕ to be $M_s(\phi)$ where s is the

assignment such that $s(x) = \ominus$ for all x . Note that according to this definition $M(\phi) = \top$ iff $M_s(\phi) = \top$ for all s , and $M(\phi) = \perp$ iff $M_s(\phi) = \perp$ for all s .

We also have the following:

Lemma II.2.2 (Substitution for variables)

- (1) $M_s(u(t/x)) = M_s(x|M_s(t))(u)$
- (2) $M_s(\phi(t/x)) = M_s(x|M_s(t))(\phi)$ providing that t is substitutable for x in ϕ .

Now we can check up on monotonicity in various ways. In the first place the right substitutivity conditions for the two basic kinds of singular-term extensionality are forthcoming:

Lemma II.2.3: For any M and s , if $M_s(t_1) \sqsubseteq M_s(t_2)$ then:

$$M_s(u(t_1/x)) \sqsubseteq M_s(u(t_2/x)) \quad \text{and} \quad M_s(\phi(t_1/x)) \sqsubseteq M_s(\phi(t_2/x))$$

This is easy to state and check given the apparatus we have just introduced. (Note that setting it up this way allows – as we want – for non-uniform substitution of t_1 for t_2 .)

We could do similar things in terms of substitution for other linguistic constituents, but, rather, let us get the whole matter out of the way in one go in the material mode by defining a relation \sqsubseteq between models M and N for L so that $M \sqsubseteq N$ means that M and N have the same domain and for all P, f and c in $Prd(L)$, $Enc(L)$ and $Cns(L)$, respectively:

$$P_M \sqsubseteq P_N \quad f_M \sqsubseteq f_N \quad c_M \sqsubseteq c_N$$

in the appropriate senses of \sqsubseteq explained in section II.1. To recap, this means:

$$P_M(\vec{a}) \sqsubseteq P_N(\vec{a}) \quad \text{for all } \vec{a} \in (D \cup \{\ominus\})^{\lambda(P)}$$

$$f_M(\vec{a}) \sqsubseteq f_N(\vec{a}) \quad \text{for all } \vec{a} \in (D \cup \{\ominus\})^{\mu(f)}$$

where D is the common domain of M and N ; while for the constants the ordering is just the basic ordering on $D \cup \{\Theta\}$. Then defining \sqsubseteq also between assignments by :

$$s_1 \sqsubseteq s_2 \text{ iff } s_1(x) \sqsubseteq s_2(x) \text{ for all } x \in Var$$

we can state the following:

Lemma II.2.4 If $M \sqsubseteq N$ and $s_1 \sqsubseteq s_2$, then for any $t \in Trm(L)$ and $\phi \in Frm(L)$

$$M_{s_1}(t) \sqsubseteq N_{s_2}(t) \text{ and } M_{s_1}(\phi) \sqsubseteq N_{s_2}(\phi) .$$

The relation of logical equivalence and also that of degree-of-definedness between formulae can be defined with respect to a particular L as follows:

$$\phi \simeq \psi \text{ iff } M_s(\phi) = M_s(\psi) \text{ for any model } M \text{ for } L \text{ and any } s$$

$$\phi \sqsubseteq \psi \text{ iff } M_s(\phi) \sqsubseteq M_s(\psi) \text{ for any model } M \text{ for } L \text{ and any } s .$$

However, these relations are in fact absolute in the following sense: if \simeq_1 and \sqsubseteq_1 , and \simeq_2 and \sqsubseteq_2 , are defined for L_1 and L_2 , where L_2 is an expansion of L_1 , then if ϕ and ψ are formulae of L_1 , $\phi \simeq_1 \psi$ iff $\phi \simeq_2 \psi$ and $\phi \sqsubseteq_1 \psi$ iff $\phi \sqsubseteq_2 \psi$. So, in particular, to determine the equivalence or otherwise of ϕ and ψ we need look no further than the smallest language containing the vocabulary of ϕ and ψ . This is just like total-valued logic, of course, and it is easy enough to check once we have got the notion of a reduct: if L_2 is an expansion of L_1 and M is a model for L_2 , let $M|L_1$, the reduct of M to L_1 , be the model for L_1 which has the same domain as M and interprets the vocabulary of L_1 in precisely the same way as M does. Clearly, then, for $\phi \in Frm(L_1)$ ($\subseteq Frm(L_2)$), $M|L_1(\phi) = M_s(\phi)$ for any assignment s . If $N = M|L_1$ we say that M is an expansion of N to L_2 .

Now how do the quantifiers and the novel sentence connectives interreact? The following equivalences are interesting:

$$\forall x(\phi * \psi) \simeq \forall x(\phi \wedge \psi) * \forall x(\phi \vee \psi)$$

$$\exists x(\phi * \psi) \simeq \exists x(\phi \wedge \psi) * \exists x(\phi \vee \psi) .$$

This means that $*$ can always be dragged out of the scope of a quantifier — and hence that any formula is equivalent to one in which no occurrence of $*$ (nor $*$) comes within the scope of a quantifier. A similar result holds for $/$ too, though here we are officially taking ϕ/ψ to be defined.

But have we missed something out? Having got \forall and \exists parallel to \wedge and \vee , we might ask: should there not be a quantifier $\exists\forall$ parallel to $*$ — one with the truth-conditions of \forall and the falsehood-conditions of \exists ? Indeed there should, but we need not adopt new primitive vocabulary, since this can easily be defined:

$$\exists\forall x\phi = \forall x\phi * \exists x\phi .$$

It is immediate that if M is a model whose domain is non-empty, then for any s :

$$M_s(\exists\forall x\phi) = \top \iff M_{s(x|d)}(\phi) = \top \text{ for all } d \in D_M$$

$$M_s(\exists\forall x\phi) = \perp \iff M_{s(x|d)}(\phi) = \perp \text{ for all } d \in D_M$$

while if $D_M = \emptyset$, then invariably $M_s(\exists\forall x\phi) = *$. In Chapter V we shall be considering a complex-quantifier version of $\exists\forall$. For the present we may just note that $\exists\forall$ is self-dual ($\neg \exists\forall x\phi \simeq \exists\forall x \neg \phi$), and that $*$ may be pulled out of its scope ($\exists\forall x(\phi * \psi) \simeq \exists\forall x\phi * \exists\forall x\psi$). The 'presupposition' — i.e. truth-value precondition — of $\exists\forall x\phi$ is that ϕ is an all-or-nothing matter with respect to x — that ϕ is constant with respect to assignments to x from the domain (given a fixed assignment to the free variables if any of $\exists\forall x\phi$).

This is 'de re' constancy, and so $\exists\forall x\phi$ must be carefully distinguished from $\phi(\Theta/x)$, which presupposes absolute constancy. There is an

important difference here: the characteristic principle governing Θ is that

$$\phi(\Theta/x) \sqsubseteq \phi$$

Hence $\phi(\Theta/x)$ presupposes that, as we vary over assignments, the value, if any, of ϕ is quite independent of what is assigned to x . This is the (non-trivial) role that the logically non-denoting term plays in our logic — we shall see shortly how useful it is. Notice that while in an empty domain $\exists x\phi$ is always undefined, $\phi(\Theta/x)$ need not be. Of course Θ cannot be 'pulled out of the scope' of quantifiers, since it occurs as a singular term. It is an expressively very powerful item of vocabulary: given our interpretation of $=$, $*$ could be defined by $\Theta = \Theta$ and hence \times (and $/$) could be defined in terms of Θ too; while $,$ even with \times or $*$ at hand, we could not dispense with Θ without reducing expressive power. We shall explain this last remark in the next section.

The interpretation of the identity predicate is an unambitious one, but it will, even so, prove useful when we come to set up logic proper in Chapter IV. Apart from its own laws, it will play a role in connection with the quantifier principles. This is because our logic is one of those 'free logics' in which a self-identity statement is a kind of existence statement: $M_s(t=t)$ is true iff $M_s(t) \in D_M$. Note that unless $D_M = \emptyset$:

$$M_s(t=t) = M_s(\exists x(x=t)) \quad (\text{assuming } x \text{ does not occur in } t)$$

If t does not denote, then of course $t=t$ is truth-valueless: it can never be false — truth conditions and truth-value preconditions coincide. However this is all we could hope for — and all we should want — of a first order existence predicate in an extensional logic such as ours. A predicate $E(x)$ with truth/falshood conditions

$$\begin{aligned} M_s(E(x)) &= \top \quad \text{if } s(x) \in D_M \\ &= \perp \quad \text{otherwise} \end{aligned}$$

is often to be found in the literature, but it is clearly not extensional: it is a bit like the old 'it is true that ...' operator in character.

Though we have no such predicate, this does not mean to say that we shall be so resourceless when we come to set up theories in the logic, since a theory will be able to determine that a term t does not denote, just as well as that a term does denote (see Chapter IV).

Much as in total-valued logic, the identity predicate allows us to formulate finite cardinality conditions. For example, consider the formula $\forall x(\phi \leftrightarrow x=y)$, which we abbreviate $\phi(x!y)$. We may read $\phi(x!y)$ as 'y is the unique x such that ϕ ', assuming that y is a variable distinct from x and that y is not free in ϕ . For an element $d \in D_M$

$$M_s(y|d)(\phi(x!y)) = \top \iff \begin{cases} M_s(x|d)(\phi) = \top \text{ and} \\ M_s(x|e)(\phi) = \perp \text{ if } e \in D_M \text{ and } e \neq d \end{cases}$$

Hence d is the unique x such that ϕ only if it is determinately so.

Uniqueness has got to mean determinate uniqueness: there could be no formula $\psi(y)$ formalizing 'y is the unique x such that ϕ ' whose truth conditions have ' $=\perp$ ' replaced by ' $\neq\top$ ' in the above, since then we could easily find a formula ϕ and two models M and M' such that $M \sqsubseteq M'$ and which contained an element d in their domain for which $M_s(y|d)(\psi(y)) = \top$ but $M'_s(y|d)(\psi(y)) \neq \top$. Monotonicity would then have been violated.

$M_s(y|d)(\phi(x!y)) = \perp$, on the other hand, when either $M_s(x|d)(\phi) = \perp$ or $M_s(x|e)(\phi) = \top$ for some element e of D_M distinct from d . And, of course, $M_s(y|\Theta)(\phi(x!y)) = *$.

Pure cardinality sentences specifying a finite size for D_M work exactly like total-valued logic — they are always either true or false.

For any n we can put

for " $\exists x_1, \dots, \exists x_n$ " and " $\exists x_1 \dots \exists x_n$ "

$$\phi_{\geq n} = \exists x_1, \dots, \exists x_n \left(\bigwedge_{1 \leq i < j \leq n} x_i \neq x_j \right) \quad (x_i \neq x_j \text{ is short for } \neg x_i = x_j)$$

where the x_i are all distinct; so $\phi_{\geq n}$ means that there are at least n elements in D_M . Also, if we define

$$\phi_n = \phi_{\geq n} \wedge \neg \phi_{\geq n+1}$$

then $M(\phi_n) = \top$ iff D_M is of size n — and is false otherwise. In total-valued logic finite elementality equivalent models are isomorphic, and this is true for the present models too: we can prove it as a corollary to a subtler theorem which is equally interesting in itself.

Let us list three elementary relations between models for a given L :

- (1) $M \sim_e N \iff M(\phi) = N(\phi)$ for all ϕ
- (2) $M \sqsubseteq_e N \iff M(\phi) \sqsubseteq N(\phi)$ for all ϕ
- (3) $M \square_e N \iff M(\phi) \square N(\phi)$ for all ϕ

(Since we have the logically undefined constant, it is a matter of indifference whether we take 'all ϕ ' to mean all sentences of L or all formulae of L .) Also, if we have a one-one correspondence $\theta : D_M \rightarrow D_N$ between the domains of models M and N for L , then we can define the notion of an isomorphism with respect to θ and analogous notions for degree-of-defined-

ness and compatibility. We may define (1) $M \overset{\theta}{\sim} N$, (2) $M \overset{\theta}{\sqsubseteq} N$, and (3) $M \overset{\theta}{\square} N$ in terms of simple formulae, viz. ones of the form $Px_1 \dots x_{\lambda(P)}$, $y = fx_1 \dots x_{\mu(f)}$ or $y = c$:

- (1) $M \overset{\theta}{\sim} N \iff M_s(\phi) = N_{\theta \circ s}(\phi)$ for any s and simple ϕ
- (2) $M \overset{\theta}{\sqsubseteq} N \iff M_s(\phi) \sqsubseteq N_{\theta \circ s}(\phi)$ for any s and simple ϕ
- (3) $M \overset{\theta}{\square} N \iff M_s(\phi) \square N_{\theta \circ s}(\phi)$ for any s and simple ϕ

where $\theta \circ s$ is the compound function got by applying s first, then θ , and θ is taken to be extended to cover Θ by putting $\theta(\Theta) = \Theta$.

It should be clear that these relations could equivalently be defined as above in terms of arbitrary formulae ϕ . But the definitions in terms of simple formulae are nearer to the equivalent explicitly structural definitions. And notice that if θ is the identity function, then $M \overset{\theta}{\sim} N$ and $M \overset{\theta}{\sqsubseteq} N$ mean simply that $M = N$ and $M \sqsubseteq N$ respectively. We shall also be interested in compatibility on a common domain. This we write $M \square N$: it means that $M \overset{\theta}{\square} N$ where θ is the identity function. But we may extend the use of ' \square ' and, for a given D , define a notion of compatibility over $D \cup \{\Theta\}$ by:

$$a \square b \iff \text{not: } a \in D \text{ and } b \in D \text{ and } a \neq b.$$

Then we can define $M \square N$ explicitly by:

- (i) $P_M(\vec{a}) \square P_N(\vec{a})$ for all $\vec{a} \in (D \cup \{\Theta\})^{\lambda(P)}$
- (ii) $f_M(\vec{a}) \square f_N(\vec{a})$ for all $\vec{a} \in (D \cup \{\Theta\})^{\mu(f)}$
- (iii) $c_M \square c_N$

To return to the general case, we write $M \simeq N$, $M \overset{\theta}{\sqsubseteq} N$ and $M \overset{\theta}{\square} N$ if there is some correspondence θ , of the appropriate kind; and the first thing we show is:

Theorem II.2.5: If $M \overset{\theta}{\sqsubseteq}_e N$ and either D_M or D_N is finite, then $M \overset{\theta}{\square} N$.

Proof: First, since finite cardinality statements are determinate, if $M \overset{\theta}{\sqsubseteq}_e N$, then D_M and D_N are of the same finite size — n say. Let D_M be $\{d_1, \dots, d_n\}$ and pick n distinct variables x_1, \dots, x_n . If we put $s(x_i) = d_i$ for $i = 1, \dots, n$, while (for definiteness) say $s(x) = \Theta$ otherwise, it will then be sufficient to show that there is a one-one correspondence $\theta: D_M \rightarrow D_N$ such that for all simple formulae ϕ containing at

for " $\exists x_1 \dots x_n \psi$ " and " $\exists x_1 \dots \exists x_n \psi$ "
(L.9 and L.10)

for " $\text{and so } \theta'os = s'$ " and " $\text{and } \theta'os = s'$ "

most x_1, \dots, x_n free

$$M_s(\phi) \subseteq N_{\theta \circ s}(\phi)$$

For then, by basic semantic principles, $M \stackrel{\theta}{\subseteq} N$. But say this is not so; then for each correspondence θ there is a formula ϕ_θ containing at most x_1, \dots, x_n free such that

$$M_s(\phi_\theta) = \top \text{ but } N_{\theta \circ s}(\phi_\theta) \neq \top.$$

There are $n!$ such correspondences, so we can define the formula

$$\psi = \bigwedge_{\theta} \phi_{\theta} \wedge \bigwedge_{1 \leq i < j \leq n} x_i \neq x_j.$$

Clearly $M_s(\psi) = \top$ and hence $M(\exists x_1 \dots x_n \psi) = \top$. But $M \stackrel{e}{\subseteq} N$, and so $N(\exists x_1 \dots x_n \psi) = \top$ also. Since ψ contains all the $x_i \neq x_j$ as conjuncts, this means that there is an assignment s' which assigns distinct elements of D_N to x_1, \dots, x_n (and $s'(x) = \epsilon$ otherwise) such that $N_{s'}(\psi) = \top$. Now $\{s'(x_i) : 1 \leq i \leq n\} = D_N$, and so the function $\theta' : a_i \mapsto s'(x_i)$ is a one-one correspondence between D_M and D_N , and so $\theta' \circ s = s'$. Thus $N_{\theta' \circ s}(\psi) = \top$. But $\phi_{\theta'}$ is a conjunct of ψ , so then $N_{\theta' \circ s}(\phi_{\theta'}) = \top$ — which is a contradiction. \square

The case of equivalence and isomorphism now follows:

Theorem II.2.6: If $M \sim_e N$ and either D_M or D_N is finite, then $M \simeq N$.

Proof: $M \sim_e N$ implies that $M \stackrel{e}{\subseteq} N$ and $N \stackrel{e}{\subseteq} M$, and so by the last theorem there exist θ and η such that

$$M \stackrel{\theta}{\subseteq} N \stackrel{\eta}{\subseteq} M \stackrel{\theta}{\subseteq} N \stackrel{\eta}{\subseteq} \dots$$

If we compound these mappings and define:

$$\theta_1 = \theta \quad \theta_{k+1} = \theta \circ \eta \circ \theta_k$$

then clearly for all k $M \stackrel{\theta_k}{\subseteq} N$. Since there are only $n!$ possibilities,

$\theta_i = \theta_j$ for some $i < j$; hence for $d \in D_M$

$$\eta : \theta_i(d) \mapsto \eta \circ \theta_i(d) \quad \theta_{j-i} : \eta \circ \theta_i(d) \mapsto \theta_j(d) = \theta_i(d).$$

So $\theta_{j-i} = \eta^{-1}$ and $M \stackrel{\eta^{-1}}{\subseteq} N \stackrel{\eta}{\subseteq} M$, which means that $M \stackrel{\eta^{-1}}{\sim} N$.

for "section 1.2" and "section 1.3"

Notice that from the proof of this theorem we get the following corollary:

Corollary II.2.7: If $N \sqsubseteq_{\eta}^{\eta} M$ and $M \sqsubseteq_e N$, then $N \sim_{\eta} M$.

We can also state an analogous theorem for compatibility:

Theorem II.2.8: If $M \square_e N$ and either D_M or D_N is finite, then $M \square_{\leftrightarrow} N$.

There is a proof of this fact parallel to the proof given of II.2.5: in strategic places ' $\neq \top$ ' is replaced by ' $= \perp$ ' and ' $= \top$ ' by ' $\neq \perp$ '

Later, in Chapter IV, we shall find that (II.2.5) and (II.2.8) are in fact corollaries of a more powerful result.

Consider again the complex quantifiers $\forall x \dots x \dots$, $\exists x \dots x \dots$ and $\text{I}x \dots x \dots$ that were introduced in Section I.2. There we took it on faith that the definitions worked adequately; now we can check properly that they do — moreover not just for simple predicates which are assumed to be unstructured and totally defined on their domain, but in general. For arbitrary formulae ϕ and ψ we can put:

$$\forall x \phi \cdot \psi = \exists x \phi / \forall x (\phi \rightarrow \psi)$$

$$\exists x \phi \cdot \psi = \exists x \phi / \exists x (\phi \wedge \psi) .$$

The following conditions then hold:

$$M_S(\forall x \phi \cdot \psi) = \top \iff M_{S(x|d)}(\phi) = \top \text{ for some } d \in D_M; \text{ and} \\ \text{either } M_{S(x|d)}(\phi) = \perp \text{ or } M_{S(x|d)}(\psi) = \top \\ \text{for all } d \in D_M$$

$$M_S(\forall x \phi \cdot \psi) = \perp \iff M_{S(x|d)}(\phi) = \top \text{ and } M_{S(x|d)}(\psi) = \perp \text{ for some } d \in D_M.$$

Dual conditions hold for $\exists x \phi \cdot \psi$, and we have the following equivalence:

$$\exists x \phi \cdot \psi \simeq \neg \forall x \phi \cdot \neg \psi .$$

For $\text{I}x \dots x \dots$ we could have

$$\text{I}x\phi \cdot \psi = \exists y\phi(x!y) / \forall x(\phi \rightarrow \psi)$$

or

$$\text{I}x\phi \cdot \psi = \exists y(\phi(x!y) \wedge \psi(y/x)) \times \forall y(\phi(x!y) \rightarrow \psi(y/x))$$

or any of the other definitions we sketched before. These yield the following truth-falsehood conditions:

$$M_s(\text{I}x\phi \cdot \psi) = \top \iff M_s(y|d)(\phi(x!y)) = \top \text{ and } M_s(x|d)(\psi) = \top$$

for some $d \in D_M$

$$M_s(\text{I}x\phi \cdot \psi) = \perp \iff M_s(y|d)(\phi(x!y)) = \top \text{ and } M_s(x|d)(\psi) = \perp$$

for some $d \in D_M$.

This seems to be the most obvious generalization of what we were suggesting in the first chapter — but more on the matter in the next section.

To end up, some brief remarks on more exotic ideas. We have been dealing with identity as a determinate and discrete affair on the domains of our models, but there are two obvious ways in which we could be more liberal over the interpretation of equality formulae and yet still remain with a genuinely first-order relation. One concerns the relation itself, and the other the workings of denotation.

First: we could drop the assumption that identity on a domain D was determinate. This would involve having our interpretation of $=$ part company with the metalinguistic identity relation assumed when we present the set-theoretical definition of 'model for L ': we would not stipulate that necessarily $E(d, e) = \perp$ if d and e were distinct elements of D (where E is the function interpreting $=$). The fact that our model-theoretic language made determinate distinctions where the object language did not need not invalidate the role of such models as providing a representation of a relation of identity between objects that admitted cases of

genuine indeterminacy, wherever that idea itself made any sense. In this case $=$ would not have a uniformly stateable interpretation: there would be the constraint that $E(d, e) = \top$ iff $d = e \in D$, but within those limits there could be variation. There is, of course, a range of further stipulations that we might or might not want to make — e.g. symmetry of determinate distinctness.

The second liberalization would be to abandon the all-or-nothing character of denoting. Let us for the moment revert to the assumption that identity on D remains determinate, viz. that on D object-language identity coincides with metalinguistic identity. We could, however, modify the semantic classification of terms to accommodate 'partial denotation': a given term could indiscriminately denote items from a particular subset of D , while definitely not denoting the rest, though in the limiting case — to be identified with the classification \ominus ? — it need not definitely-not-denote anything at all. We would have a non-discrete system of classifications which could be represented by taking non-empty subsets of D to be assigned to terms: the ordering \supseteq would be the \sqsubseteq ordering on singular-term interpretations (and so a term would undividedly denote a single element d of D iff it were assigned $\{d\}$.) The interpretation of identity — and indeed of all other functors — would have to be monotonic functions from this set $\mathcal{P}(D) \setminus \{\emptyset\}$ into $\{\top, *, \perp\}$, and would admit this time of more cases of false equality formulae: $t_1 = t_2$ could be false even when either t_1 or t_2 or both did not undividedly denote a single element of D : the natural falsehood conditions for a and b ranging over $\mathcal{P}(D) \setminus \{\emptyset\}$ would be $E(a, b) = \perp$ iff $a \cap b = \emptyset$.

Both these liberalizations could be taken together, and then the system representing singular-term interpretations gets more complicated still. As we proceed in this kind of way the structures involved call

for "the functionals

$$(D \cup \{0\} \xrightarrow{\text{monotonic}} \{T, *, \perp\}) \xrightarrow{\text{monotonic}} \{T, *, \perp\} "$$

had "the functions

$$D \cup \{0\} \xrightarrow{\text{monotonic}} \{T, *, \perp\}$$

more and more on the richness of the whole 'partial power set' of D , viz the functionals

$$(D \cup \{\Theta\} \xrightarrow{\text{monotonic}} \{\top, *, \perp\}) \xrightarrow{\text{monotonic}} \{\top, *, \perp\}$$

and we find terms interpreted more like one-place predicates and predicates like second order functors. Abandoning the distinctions is one step away: we can obtain various kinds of language with type-free term-upon-term application, interpreted by models which make some coherent identification of elements of an underlying domain into the partial power set of this domain, along with a way of getting back from the power set to the domain-plus- Θ . Not only application but also abstraction will be interpretable. Various identifications meet various needs; and we have the possibility also of iterating the identification to provide embeddings of power sets into power sets, using devices like that employed in Scott (1973a).

The attraction of Scott's idea in this paper is that we may define a limit domain which constitutes a closed system of type-free functions. Quantifiers, however, would seem to present something of a stumbling block in an attempt to provide a full type-free logic. In the first place what exactly should they range over? And secondly, it would seem that on any reasonable answer to this question they would force us ever onwards and upwards without reaching a domain in which they were themselves actually contained. I do not, however, despair of at least finding some non-trivial and non-silly approximation to the ideal of a unified domain closed under its own quantifiers. A different approach would be to take the intensional way out: see for example Scott (1975).

Finally the example we promised of a non-Fregean extensional functor with output in a functional category. To keep things simple, think again of our simple models with a discrete and determinate domain, and consider how second-order logic might go. The monotonic character of our semantics

suggests that we might be interested in an inductive-definition operator $I(F, x)$ binding a predicate (for the moment say one-place) variable F and an individual variable x to form a (one-place) predicate expression $I(F, x) \phi(F, x)$ which would specify the 'minimal solution' for F in the equivalence:

$$Fx \simeq \phi(F, x) ,$$

i.e. $I(F, x) \phi(F, x)$ would be interpreted as the least defined extension satisfying the above equivalence. Familiar kinds of induction are in total-valued logic explicitly specifiable using ordinary second order quantifiers; however, it would not seem possible to do anything parallel for $I(F, x) \phi(F, x)$ in partial-valued logic; still, it is easy enough to take $I(F, x)$ as a primitive logical notion, whose interpretation we now explain. First, given two subsets X and Y of $D_M \cup \{\Theta\}$ such that $X \cap Y = \emptyset$ and such that if $\Theta \in X$, $X = D_M \cup \{\Theta\}$ and if $\Theta \in Y$, $Y = D_M \cup \{\Theta\}$, let (X, Y) be the function $D_M \cup \{\Theta\} \rightarrow \{\top, *, \perp\}$ given by

$$\begin{aligned} (X, Y)(a) = \top &\iff a \in X \\ &= \perp \iff a \in Y \end{aligned}$$

Now, we can define chains $\langle X_\alpha \rangle$ and $\langle Y_\alpha \rangle$ which stop at some stage and are such that $(\cup X_\alpha, \cup Y_\alpha)$ is the required extension in a model M for $I(F, x) \phi(F, x)$ with respect to some assignment s (to individual and predicate variables):

$$X_0 = \emptyset = Y_0$$

$$X_{\alpha+1} = \left\{ a \in D_M \cup \{\Theta\} \mid M_s(F \mid (X_\alpha, Y_\alpha), x \mid a) (\phi(F, x)) = \top \right\}$$

$$Y_{\alpha+1} = \left\{ a \in D_M \cup \{\Theta\} \mid M_s(F \mid (X_\alpha, Y_\alpha), x \mid a) (\phi(F, x)) = \perp \right\}$$

and for limit λ :

$$X_\lambda = \bigcup_{\alpha < \lambda} X_\alpha \quad \text{and} \quad Y_\lambda = \bigcup_{\alpha < \lambda} Y_\alpha$$

For the general theory behind this see e.g. Moschovakis (1974). Note that

(first-order) quantifiers in $\phi(F, x)$ are what can force us to go transfinite.

It is easy to see that if $M \sqsubseteq M'$, then the interpretation of $I(F, x) \phi(F, x)$ in M' is compatible with and no less defined than its interpretation in M , and so extensionality is preserved. Of course we can easily generalize all this to cover ≥ 1 -place variables F^n .

II.3 DEFINITE DESCRIPTIONS

This section is devoted to definite descriptions in extensional logic. The possibility of non-uniquely-specifying descriptions is what makes their treatment in total-valued logic problematic, so it would seem that our partial-valued semantics, catering in a uniform way for undefinedness in all semantical categories, is the ideal framework in which to handle them. Earlier, to illustrate the usefulness of \ast and $/$, we defined, among others, a complex quantifier, $Ix\phi(x)$, using which we might represent definite description idioms; now, in contrast, we must address ourselves to the semantics of languages with description terms. In philosophical and mathematical practice definite descriptions naturally arise in this form — and typically in contexts where theorising is taken to be governed by a simple extensional logic: so the first thing to investigate here is how such terms should go.

Of course, having determined this matter, the next question will be: can we deploy $Ix\phi(x)$ to define the terms away? If so, how and how successfully? If not, then can we nonetheless eliminate them in some other way — and, if so, how and how successfully?

Definite descriptions, and other similar grammatical constructions, occur in ordinary natural language of course, and you may wish to provide logical forms which use genuine terms — or equivalent devices — to represent

them. However, the relevance of the present material will depend on your view of the functioning of such phrases: whether, or to what extent, and in what way, you see them playing a role other than as a medium for indexical reference. The terms we shall be considering here are such that the denotation, if any, which the term-forming construction determines, depends on nothing over and above the interpretation of constituent vocabulary. The subject of indexical reference is ignored also in Chapter V. Historically the issues of indexical reference and natural language presupposition have been intertwined – Strawson (1950) – but the semantics and pragmatics of that phenomenon of presupposition which I want to consider in the final chapter is a matter which I believe is best discussed – to start with at any rate – quite independently.

What we must do, then, is extend our first-order languages by introducing a variable-binding operator, ι , to form terms $\iota x \phi(x)$. To interpret such terms in the simple kind of model that we have been considering, we need to decide on a class of functionals $i_D: (D \cup \{\Theta\} \rightarrow \{\top, *, \perp\}) \rightarrow D \cup \{\Theta\}$, so that we shall be able to put:

$$(1) \quad M_S(\iota x \phi) = i_{D_M}(P)$$

where P is the function such that $P(a) = M_S(x|a)(\phi)$ for $a \in D \cup \{\Theta\}$. In accordance with our extensionality principles, the i_D must be monotonic. There is, then, only one thing we can do: given a domain D we must specify that for $d \in D$

$$(2) \quad i_D(P) = d \iff P(d) = \top \text{ and } P(e) = \perp \text{ for all } e \in D \\ \text{such that } e \neq d.$$

This means that $i_D(P)$ is an element of D iff there is a d satisfying the right-hand side and otherwise $i_D(P)$ is Θ . It would not be sufficient to stipulate that $i_D(P) = d \in D$ provided merely that P is not true of any other element of D , since if P were not determinately false of everything

else, then we could find a strengthening P' of P , true of more than one element of D , and we should then want $i_D(P') = \Theta$; but in that case i_D would not be monotonic.

Hence we can see that $i_D(P)$ may be undefined for 'two different kinds of reasons': either because P is not sufficiently highly defined to determine a unique element of which it is true, or because it is sufficiently highly defined determinately to rule out there being any such element. The connection with uniqueness statements points this up. Let $\phi(x!y)$ be defined, as before, by $\forall x(x = y \leftrightarrow \phi)$, then, according to (1) and (2), if $d \in D_M$:

$$(3) \quad M_s(\iota x\phi) = d \iff M_s(y|d)(\phi(x!y)) = \top .$$

Of course we have yet to set up the systematic definition of $M_s(\)$ for the extended language and check that everything works; but, assuming it does, $\iota x\phi$ will denote an element d of D_M iff $\phi(x!y)$ is true of d , while if it is either undefined or false of d , then d cannot be the denotation of $\iota x\phi$.

Certainly, if we were interpreting $=$ more liberally, and allowing terms to range over a structure richer than just $D_M \cup \{\Theta\}$, then it would be open to us to make the classification of $\iota x\phi$ rather more subtle, so as to take into account the falsehood conditions of $\phi(x!y)$ as well. But we are here considering discrete and determinate models.

The literature on definite descriptions is vast and often interesting, but there has been no need for us to scout around in it for ideas what to do, since our course is completely determined by the general principles of extensionality we decided to adopt. It is interesting, however, to make some comparisons. The two papers in the more recent literature which I have found the most inspiring are Scott (1967) and Smiley (1960). In Scott's paper, though denotationless singular-terms are allowed, at sentence

level everything is totally defined. No such logic would suit us here, since if $\iota x\phi$ is denotationless, then an atomic formula, $P\iota x\phi$ say, could only be defined if P were itself constant — on pain of violating our ideas about functional dependency. But of course we want to allow for less trivial predicates, and so, if you take extensionality as expounded in Section II.1, then, even though this is a liberalization of the cruder idea attributed to Frege, it would still follow that partial-valued logic is not merely the ideal framework but in fact a necessary one for an extensional treatment of definite descriptions as terms.

Smiley's paper, in contrast, does allow for truth-value gaps, but, even so, when it comes to interpreting ι -terms, it is sufficient for $\iota x\phi$ to denote an object d that ϕ is true of d and not true of anything else — d does not have to be determinately unique: and so extensionality is violated in the way we explained earlier.

 Let us sketch some definitions. Languages are what they were before, with the addition of the symbol ι , and we can define the set $Wfe(L)$ of well formed expressions of a given language L all in one go, along with a specification of the kind of expression — term or sentential formula. Let W be the smallest set X such that:

If $t \in Var$ or $t \in Cns(L)$ or t is Θ , then $\langle t, 0 \rangle \in X$

If ϕ is \top or ϕ is \perp , then $\langle \phi, 1 \rangle \in X$

If $f \in Fnc(L)$ and $\langle t_1, 0 \rangle, \dots, \langle t_{\mu(f)}, 0 \rangle \in X$, then $\langle ft_1 \dots t_{\mu(f)}, 0 \rangle \in X$

If $P \in Prd(L)$ and $\langle t_1, 0 \rangle, \dots, \langle t_{\lambda(P)}, 0 \rangle \in X$, then $\langle Pt_1 \dots t_{\lambda(P)}, 1 \rangle \in X$

If $\langle t_1, 0 \rangle \in X$ and $\langle t_2, 0 \rangle \in X$, then $\langle t_1 = t_2, 1 \rangle \in X$

If $\langle \phi, 1 \rangle \in X$, then $\langle \neg\phi, 1 \rangle \in X$

If $\langle \phi, 1 \rangle \in X$ and $\langle \psi, 1 \rangle \in X$, then $\langle \phi \wedge \psi, 1 \rangle \in X$, $\langle \phi \vee \psi, 1 \rangle \in X$
 and $\langle \phi * \psi, 1 \rangle \in X$

If $\langle \phi, 1 \rangle \in X$ and $x \in Var$, then $\langle \forall x\phi, 1 \rangle \in X$, $\langle \exists x\phi, 1 \rangle \in X$
 and $\langle \iota x\phi, 0 \rangle \in X$.

Now we can define terms and formulae:

$$t \in \text{Trm}(L) \text{ iff } \langle t, 0 \rangle \in W \quad \text{and} \quad \phi \in \text{Frm}(L) \text{ iff } \langle \phi, 1 \rangle \in W$$

and put

$$\text{Wfe}(L) = \text{Trm}(L) \cup \text{Frm}(L) .$$

We have not identified $\text{Wfe}(L)$ with W because we want terms and formulae which do not contain ι to be terms and formulae in the old sense defined for the ι -free language with the same B -vocabulary (i.e. non-logical vocabulary), but $\text{Wfe}(L)$ and W correspond one-one and so we can happily induct on $\text{Wfe}(L)$ relying on the correspondence with W . For example, we can run the definition of $M_s(\alpha)$ – the classification of the well formed expression α in the model M with respect to the assignment s – on $\text{Wfe}(L)$ in such a way that $M_s(\alpha) \in \{\top, *, \perp\}$ if α is a formula and $M_s(\alpha) \in D_M \cup \{\ominus\}$ if α is a term. The clauses are just what they were before for $M_s(t)$ and $M_s(\phi)$ with the addition of one for $\iota x \phi$. We could use (1) and (2) above or write it out explicitly, as follows:

If $d \in D_M$:

$$M_s(\iota x \phi) = d \iff M_{s(x|d)}(\phi) = \top \quad \text{and} \quad M_{s(x|e)}(\phi) = \perp$$

for all $e \in D_M$ such that $e \neq d$

(and $M_s(\iota x \phi) = \ominus$ if there is no such d).

Variables can now of course be bound by ι as well as \forall and \exists ; and covering both terms and formulae the notation $\alpha(t/x)$ means the result of substituting the term t for all free occurrences of x in the well-formed expression α . As before we reserve this notation for when ' t is substitutable for x in α ', which now means that no free occurrence of a variable in t becomes bound in $\alpha(t/x)$.

We can state the two crucial semantical lemmas uniformly and check them by induction on $\text{Wfe}(L)$ (i.e. W):

Lemma II.3.1: If $s_1(x) = s_2(x)$ for all x free in α , then $M_{s_1}(\alpha) = M_{s_2}(\alpha)$.

Lemma II.3.2: If t is substitutable for x in α , then:

$$M_s(\alpha(t/x)) = M_s(x|M_s(t))(\alpha).$$

Now the connection described in (3) above, between description terms and uniqueness sentences, can be properly checked:

Lemma II.3.3: If $d \in D_M$, $M_s(\lambda x \phi) = d \iff M_s(y|d)(\phi(x!y)) = \top$

(where y is some variable occurring nowhere in $\lambda x \phi$).

The relation of (logical) equivalence between well formed expressions is given by:

$$\alpha \simeq \beta \text{ iff } M_s(\alpha) = M_s(\beta) \text{ for all } M \text{ and all } s$$

and note that if α and β are terms and x is a variable not occurring free in either α or β , then:

$$\alpha \simeq \beta \text{ iff } x = \alpha \simeq x = \beta.$$

Using this relation we can encapsulate the value conditions for $\lambda x \phi$ given in lemma II.3.3 by writing:

$$z = \lambda x \phi \simeq \exists y(\phi(x!y) \wedge z = y) \times \forall y(\phi(x!y) \rightarrow z = y).$$

This equivalence looks familiar: it is in fact $\lambda x \phi \cdot z = x$ according to one of the definitions we offered for this complex quantifier; and hence, of course, whatever equivalent definition we fixed on we would have $z = \lambda x \phi \simeq \lambda x \phi \cdot z = x$. In identity contexts, then, the quantifier would serve adequately for a contextual definition eliminating description terms: however does it always work? We must consider arbitrary formulae $\psi(\lambda x \phi)$ containing λ -terms. Recall that we saw an advantage in our definition of $\lambda x(\text{---}x\text{---}) \cdots x \cdots$ over Russell's paraphrase, couched in total-valued logic, because it did not yield unwanted scope distinctions in the case of negation — that is to say formulae with syntactical scope distinctions that

were also non-equivalent. It is easy to check that, in addition, $Ix\phi$ obeys some strong distributivity laws with respect to logical operations apart from negation:

$$\neg Ix\phi \cdot \psi \simeq Ix\phi \cdot \neg \psi$$

$$Ix\phi \cdot \psi \wedge Ix\phi \cdot \chi \simeq Ix\phi \cdot (\psi \wedge \chi)$$

and similarly for \wedge replaced by \vee or \times (or $/$). While for quantifiers, providing there is no free occurrence of y in ϕ :

$$\forall y (Ix\phi \cdot \psi) \simeq Ix\phi \cdot \forall y \psi$$

and similarly for \exists . Also, assuming that y does not occur free in ϕ , nor x in ψ :

$$Iy\psi \cdot (Ix\phi \cdot \chi) \simeq Ix\phi \cdot (Iy\psi \cdot \chi).$$

The interest of all this is that the 'more scopeless' a quantifier expression is, the more nearly it 'approximates to a singular term' for the language in question. However, even within the confines of our extensional language, $Ix\phi$ turns out not to be entirely scopeless: it cannot hop around a formula with total semantical freedom. For example it is easy to find instances for ϕ , ψ and χ such that:

$$Ix\phi \cdot \psi \vee \chi \not\equiv Ix\phi \cdot (\psi \vee \chi).$$

This is because the 'strong' table for \vee is allowed. χ may be true, making the left-hand side true, but there may be no determinately unique x such that ϕ , and this would leave the right-hand side undefined.

Hence, if we had in mind to attempt to use $Ix\phi \dots x \dots$ in a contextual definition for formulae $\psi(Ix\phi)$, then we should at very least have to stipulate that it were applied with narrow scope — to atomic formulae. But in fact this stipulation would be insufficient: we cannot use $Ix(\dots x \dots)$ as a means of defining away λ -terms completely. We have seen that it works for identity sentences, but it does not work for atomic

sentences in general — again because of the liberalness of our extensionality criterion: this time as regards the range of admissible interpretations for predicates. If a term $\lambda x\phi$ is classified as Θ , then any formula $\lambda x\phi \cdot \psi$ will be classified $*$, since $\exists y\phi(x!y)$ will not be true; however we may have an atomic formula $Pt_1 \dots x \dots t_{\lambda(P)}$ which takes a value when x is assigned Θ , and so $Pt_1 \dots \lambda x\phi \dots t_{\lambda(P)}$ and $\lambda x\phi \cdot Pt_1 \dots x \dots t_{\lambda(P)}$ would not receive the same classification if there were no determinately unique x such that ϕ .

We had better consider from scratch the question of eliminating λ . But what exactly is eliminability? In the first place we must distinguish the question of whether an item of vocabulary is eliminable in a particular theory from the question of eliminability as a matter of logic. We have not yet set up the apparatus for treating of theories, but when we do, in Chapter IV, questions of the first kind will be of great interest. However our present concern is with the eliminability of description terms in logic, where we assume no constraints on the interpretation of languages other than those specified by the interpretation rules for λ and the other logical vocabulary.

Secondly we must distinguish various modes of eliminability. λ would be eliminable in the weakest sense if for any formula of the language we were guaranteed the existence of some λ -free formula equivalent to it. But we might hope for something stronger than this: a uniform receipt for transforming a formula containing occurrences of λ into an equivalent λ -free formula. We have seen that our old paraphrase might be made to play a role here, if all atomic predicates were crudely extensional — but this is an assumption we do not want to make in pure logic. In any case, what we should really hope for is something stronger still: scopeless eliminability of λ — i.e. to be in possession of an elimination schema

$Ix(\phi(x), \psi(x))$ for $\psi(\ulcorner x\phi)$ which can be applied at any level of complexity in a given formula to yield an equivalent one. We have seen that $Ix\phi(x) \cdot \psi(x)$ could certainly not play this role.

We shall indeed produce such a schema — one which, as a complex quantifier $Ix(\phi(x), \dots x \dots)$, is as scope-free (semantically) as the term $\ulcorner x\phi$ it replaces. It will then follow that — with one obvious proviso — occurrences of \ulcorner in a formula may be eliminated by the schema in any order you like, regardless, even, of whether one occurrence is embedded within the scope of another. Also, we shall be able to eliminate two or more occurrences of the same term either altogether or one at a time.

To make this plain, we had better adopt a less perspicuous, but more precise notation, and state the basic result as:

$$\psi(\ulcorner x\phi/y) \simeq I(x, \phi, y, \psi)$$

where $I(x, \phi, y, \psi)$ is the elimination schema we are going to define. We might read this as 'the x such that ϕ is a y such that ψ '. $I(x, \phi, y, \psi)$ will then be seen to have the properties claimed for it, once we have checked that languages with \ulcorner admit of thorough-going substitutivity of equivalent formulae. The proviso I mentioned will result from an obvious restriction we must impose on $\psi(\ulcorner x\phi/y)$ for the above equivalence to hold: that $\ulcorner x\phi$ be substitutable for y in ψ . This means that if a variable occurring free in some constituent term $\ulcorner x\phi$ is bound in a formula by some occurrence of \forall or \exists or \ulcorner , then to eliminate $\ulcorner x\phi$ the schema must be applied within the scope of that occurrence. With respect to any other occurrence of \forall or \exists or \ulcorner , or any occurrence of a sentence connective, there will be total freedom.

How might we set about discovering $I(x, \phi, y, \psi)$? We know from the substitution-for-variables lemma (II.3.2) that $M_S(\psi(\ulcorner x\phi/y)) = M_S(y | M_S(\ulcorner x\phi))(\psi)$,

and from II.3.3 that this value is \top iff

$$\begin{aligned} &\underline{\text{either}} \quad M_{\mathcal{S}}(y|d)(\phi(x!y)) = \top \quad \text{and} \quad M_{\mathcal{S}}(y|d)(\psi) = \top \quad \text{for some } d \in D_M \\ &\quad \underline{\text{or}} \quad M_{\mathcal{S}}(y|\Theta)(\psi) = \top \end{aligned}$$

and is \perp iff

$$\begin{aligned} &\underline{\text{either}} \quad M_{\mathcal{S}}(y|d)(\phi(x!y)) = \top \quad \text{and} \quad M_{\mathcal{S}}(y|d)(\psi) = \perp \quad \text{for some } d \in D_M \\ &\quad \underline{\text{or}} \quad M_{\mathcal{S}}(y|\Theta)(\psi) = \perp . \end{aligned}$$

Hence it is easy enough to write down separate formulae which are respectively co-true and co-false with $\psi(\iota x \phi / y)$:

$$(i) \quad M_{\mathcal{S}}(\psi(\iota x \phi / y)) = \top \quad \text{iff} \quad M_{\mathcal{S}}(\exists y (\phi(x!y) \wedge \psi) \vee \psi(\Theta / y)) = \top$$

$$(ii) \quad M_{\mathcal{S}}(\psi(\iota x \phi / y)) = \perp \quad \text{iff} \quad M_{\mathcal{S}}(\forall y (\phi(x!y) \rightarrow \psi) \wedge \psi(\Theta / y)) = \perp .$$

Neither of these will do on its own, but notice that they are compatible — i.e. they can never take on opposing values for any model under any assignment. If we had the analogue of theorem I.2.7, then we would be able to deduce that there was a single formula with the truth-falsehood conditions of $\psi(\iota x \phi / y)$. The compatibility theorem does in fact hold (see Chapter IV), but we do not have it yet; and in any case it would only guarantee eliminability in the weakest sense, not the existence of a schema of elimination.

However we can interweave (i) and (ii) in either of two equivalent ways to yield a definition of $I(x, \phi, y, \psi)$:

$$(a) \quad \exists y (\phi(x!y) \wedge \psi) \vee (\forall y (\phi(x!y) \rightarrow \psi) \wedge \psi(\Theta / y))$$

$$(b) \quad \forall y (\phi(x!y) \rightarrow \psi) \wedge (\exists y (\phi(x!y) \wedge \psi) \vee \psi(\Theta / y)) .$$

If $\psi(\Theta / y)$ is true, then so is $\forall y (\phi(x!y) \rightarrow \psi)$, and if $\psi(\Theta / y)$ is false, then so is $\exists y (\phi(x!y) \wedge \psi)$: hence it is clear from (i) that (a) has the required truth conditions, and from (ii) that (b) has the required falsehood conditions. Furthermore by definition of $\phi(x!y)$ if $\exists y (\phi(x!y) \wedge \psi)$ is

true, $\forall y(\phi(x!y) \rightarrow \psi)$ is true, and if $\forall y(\phi(x!y) \rightarrow \psi)$ is false, $\exists y(\phi(x!y) \wedge \psi)$ is false; and so it is easy to see that (a) and (b) are in fact equivalent — empty domains not excluded.

So, if $I(x, \phi, y, \psi)$ stands for either (a) or (b), then, as required:

Theorem II.3.4: For any formulae ϕ and ψ , providing $!x\phi$ is free for y in ψ :

$$\psi(!x\phi/y) \simeq I(x, \phi, y, \psi).$$

It remains to show that we can interchange equivalent subformulae to yield equivalent formulae. To see this we have to take terms into consideration too: let us define between well formed expressions in general the relation of being a ϕ - ψ -variant, where ϕ and ψ are formulae. Writing this relation $\alpha(\phi, \psi) \beta$ we can use the following inductive specification:

- (i) $\alpha(\phi, \psi) \alpha$ for all $\alpha \in Wfe(L)$; $\phi(\phi, \psi)\psi$ and $\psi(\phi, \psi)\phi$
- (ii) If $t_1, t_2 \in Trm(L)$, $x \in Var$, $\alpha \in Wfe(L)$ and $t_1(\phi, \psi)t_2$, then:

$$\alpha(t_1/x)(\phi, \psi) \alpha(t_2/x).$$
- (iii) If $\alpha, \beta, \gamma, \delta \in Frm(L)$, $\alpha(\phi, \psi)\beta$ and $\gamma(\phi, \psi)\delta$, then:

$$\neg\alpha(\phi, \psi)\neg\beta, \alpha \wedge \gamma(\phi, \psi)\beta \wedge \delta, \alpha \vee \gamma(\phi, \psi)\beta \vee \delta, \alpha \times \gamma(\phi, \psi)\beta \times \delta,$$

$$\forall x\alpha(\phi, \psi)\forall x\beta, \exists x\alpha(\phi, \psi)\exists x\beta \text{ and } !x\alpha(\phi, \psi) !x\beta.$$

Having so laboured the definition of an ϕ - ψ -variant it is now an easy matter to check:

Lemma II.3.5: For any model M : if $M_s(\phi) = M_s(\psi)$ for all s and $\alpha(\phi, \psi)\beta$ then $M_s(\alpha) = M_s(\beta)$ for all s .

We are especially interested in the following corollary:

L.3 : "Hence" should begin a new line.

Theorem II.3.6: (Interchange of equivalent subformulae).

For any formulae ϕ, ψ, χ and ω : if $\phi \simeq \psi$ and $\chi(\phi, \psi)\omega$, then $\chi \simeq \omega$. Hence II.3.4 guarantees that our elimination schema had the properties we claimed.

This is a good point to mention another item of vocabulary that there may be temptation to introduce: a two-place term-out-of-terms forming operator $@$ which would serve as an analogue on $D_M \cup \{\ominus\}$ of squadge. We would have:

$$\text{If } d \in D : M_s(t_1 @ t_2) = d \iff M_s(t_1) = M_s(t_2) = d$$

$$(M_s(t_1 @ t_2) = \ominus \text{ if there is no such } d)$$

Now, $t_1 @ t_2$ would receive the same classification (in any M with respect to any s) as $\neg(x = t_1 * x = t_2)$, and so $@$ could be eliminated as satisfactorily as \neg — using $I(x, x = t_1 * x = t_2, y, \psi)$. Alternatively, we could use either of the following to eliminate $@$ from a formula $\psi(t_1 @ t_2/y)$:

$$\exists y((y = t_1 * y = t_2) \wedge \psi) \vee (\forall y((y = t_1 * y = t_2) \rightarrow \psi) \wedge \psi(\ominus/y))$$

$$\forall y((y = t_1 * y = t_2) \rightarrow \psi) \wedge (\exists y((y = t_1 * y = t_2) \wedge \psi) \vee \psi(\ominus/y)).$$

Notice that there are cases of natural language squadging which intuitively correspond to $@$, for example in idioms such as

Harold Macmillan, the Vice-Chancellor, attended the St. Edmund's day feast.

Returning to $I(x, \phi, y, \psi)$, it is interesting to note that this schema does not involve $*$ or $*$. However \ominus is required: we must have the resources to be able to represent independence of argument place. Conversely, it is easy to see that if we have \neg , then \ominus is definable, e.g. by $\neg(x \neq x)$, or $\neg x \perp$, or $\neg x *$, or the like; so the presence of either \neg or \ominus in the language provides equal expressive resourcefulness.

The question then arises: can we get rid of either and remain equally expressive? No. We saw in the last section how \ast and \ast could be defined in terms of \ominus , but announced that \ominus was strictly stronger: that is precisely what we now have to show. We can revert to the original ι -free languages and make the point by showing the following:

Lemma II.3.7: There is a formula which is not equivalent to any \ominus -free formula.

All we need to do is set up an example; take a language with a single monadic predicate P , and consider the sentence $P\ominus$. Let M and N be models with the same domain $\{d\}$, and such that

$$P_M(d) = P_M(\ominus) = P_N(d) = \top \quad \text{but} \quad P_N(\ominus) = \ast .$$

It is easy to check that for any formula ϕ not containing \ominus , $M_s(\phi) = N_s(\phi)$ where $s(x) = d$ for all variables x . But this means that no such formula can be equivalent to $P\ominus$, since $M_s(P\ominus) \neq N_s(P\ominus)$.

Of course there may be natural circumstances under which \ominus is eliminable, and in Chapter IV, when we consider eliminability results for theories, we shall have a model-theoretic criterion to offer determining exactly the conditions under which \ominus is eliminable.

Our elimination schema, viewed the other way about, provides a contextual definition for definite description terms, and because of its scope-free nature we can regard this contextual definition as coming to precisely the same thing as a syntactically genuine term. Hence in Chapter IV we shall be content to do logic and model theory for ι -free languages.

But there could be some confusion at this point. The following remark, though obvious, might be worth making: the 'scope-free elimination of description terms' which we have provided is a result whose importance is thoroughly semantic, not syntactic. The point is that we could quite

easily jigger about with syntax and introduce a primitive complex quantifier, $Dx\phi$ say, and interpret it directly so as to mimic $\iota x\phi$ — in particular to reflect semantically the scopelessness which $\iota x\phi$ has because of its syntax. $Dx\phi$ could be defined:

$M_s(Dx\phi \cdot \psi) = \top \Leftrightarrow$ either there is a $d \in D_M$ such that:

$M_{s(x|d)}(\phi) = \top$ and $M_{s(x|e)}(\phi) = \perp$ for all $e \in D_M$
such that $d \neq e$ and $M_{s(x|d)}(\psi) = \top$ or $M_{s(x|\ominus)}(\psi) = \top$

$M_s(Dx\phi \cdot \psi) = \perp \Leftrightarrow$ either there is a $d \in D_M$ such that:

$M_{s(x|d)}(\phi) = \top$ and $M_{s(x|e)}(\phi) = \perp$ for all $e \in D_M$
such that $d \neq e$ and $M_{s(x|d)}(\psi) = \perp$ or $M_{d(x|\ominus)}(\psi) = \perp$.

Any formula $--- Dx\phi(\dots x \dots)---$ would have the same value conditions as $---(\dots \iota x\phi \dots)---$. This would not however be an 'elimination' of description terms, since $--- Dx\phi(\dots x \dots)---$ would just be a fancy way of writing $---(\dots \iota x\phi \dots)---$. What is important about the scope-free elimination result is that it shows that this proxy quantifier $Dx\phi$ is definable in the simple language specified in the last section.

Of course, if we introduced modal operators or the like into our languages, then we should actually want to point up semantically non-equivalent scope distinctions involving definite descriptions, and this we could easily do: on any reasonable treatment of modal logic $\Box I(x, \phi, y, \psi)$ and $I(x, \phi, y, \Box \psi)$, for example, are not going to turn out equivalent. It is nice to be able to play scope tricks like this with precisely the contextual definition which does not yield any extraneous distinctions when they are not wanted. (In a possible worlds semantics for partial-valued modal logic, the natural thing to do would seem to be set $\Box\phi$ true in a world iff ϕ is true in all worlds accessible from it, and $\Box\phi$ false iff ϕ is false in some world accessible from it: dually for \Diamond .)

CHAPTER III

PARTIAL-VALUED LOGIC (1) – SENTENCE COMPOSITION

III.1 INTRODUCTION

To do logic we had better decide on a relation of 'consequence' in terms of which we can state laws for the logical vocabulary. We could in fact use some other relation for this purpose, for example the relation ' \sqsubseteq ' of degree-of-definedness between logical formulae, which we defined in I.3. But thoughts turn naturally to consequence relations, since they might be expected to exhibit some interesting connection with inference – inferential practice with languages to be analysed by the logic.

Considering, for the moment, the simple schematic language specified in I.3, we can address ourselves to the problem by asking under what circumstances we would want to say that ψ were not a consequence of ϕ . In other words, in terms of our truth-value semantics, when would an assignment ν constitute a counterexample to the claim that ψ were a consequence of ϕ ? Presumably, if ν makes ϕ true but does not make ψ true, then we would have a counterexample: consequence must at very least be truth preserving. But what if ν made ψ false, yet ϕ not false? In total-valued logic this would come to the same thing as the case just considered, but not so for us. It is easy to ignore the importance of this possibility, even in a partial-valued context, since when you think of defining consequence you naturally think of its role as warranting inferences from true premises to conclusions.

However, in a partial-valued logic which uses two positive values, true and false, it might be thought that a good inference pattern should not only provide a means of advancing the stock of established truths, but also warrant the refutability of a 'premise' – at least one of the 'premises' – on the grounds that a conclusion is false.

A proper discussion of inference is not included in this thesis, but we have to define a consequence relation, and, guided by these thoughts, I propose to take it that an alternative way in which an assignment ν can constitute a counterexample to ψ 's being a consequence of ϕ is by making ψ false and ϕ not false. There is, furthermore, a modest cluster of *desiderata* which will be satisfied if we adopt this approach, all revolving round the fact that symmetry between the true and the false is not destroyed.

ψ , then, will be a consequence of ϕ if there is no ν which is a counterexample in either of the ways we have mentioned. But we shall not always be interested only in logical consequence: we shall want to be able to impose constraints on the interpretation of a language to constitute a particular theory, in which we shall want ψ to be a consequence of ϕ just in case there are no counterexamples among interpretations admitted by the theory. Hence we shall want to work with a notion of semantical consequence which is relative to a given set V of basic assignments. Logical consequence will be the particular case where V is the set of all possible assignments.

We shall adopt another generalization, too, and employ relations between finite sets of formulae Gentzen-style (or better: Scott-style) — the set of premises to be understood conjunctively, as usual, and the set of conclusions disjunctively.

So, if Γ and Δ are sets (not necessarily finite), let us say:

ν is consistent with $\langle \Gamma, \Delta \rangle$ iff

neither: $\nu(\phi) = \top$ for all $\phi \in \Gamma$ and $\nu(\psi) \neq \top$ for all $\psi \in \Delta$

nor: $\nu(\phi) \neq \perp$ for all $\phi \in \Gamma$ and $\nu(\psi) = \perp$ for all $\psi \in \Delta$

And let us abbreviate this by ' $\Gamma \Vdash_{\nu} \Delta$ ' — it means that ν is not a counterexample. Hence we can define the required relation \vdash_V , of consequence with respect to V , to hold between finite sets Γ and Δ iff

$$\Gamma \Vdash_v \Delta \text{ for all } v \in V$$

Any relation \vdash which constitutes a consequence relation had better satisfy the following conditions, which Scott gives as a general characterization of such a relation — see, for example, Scott (1973b).

$$(R) \quad \Gamma \vdash \Delta \text{ if } \Gamma \cap \Delta \neq \emptyset$$

$$(M) \quad \frac{\Gamma \vdash \Delta}{\Gamma', \Gamma \vdash \Delta, \Delta'}$$

$$(T) \quad \frac{\Gamma \vdash \phi, \Delta \quad \Gamma, \phi \vdash \Delta}{\Gamma \vdash \Delta}$$

(M) and (T) are to be understood as conditional conditions: they require that, if what is above the line holds, then what is below the line holds too. As we shall throughout, we have dropped squiggly brackets and replaced union signs by commas. Happily, it is easy to check that for any $V \Vdash_v$ satisfies (R), (M) and (T).

For the sake of uniformity, and because this would be necessary if we were to generalize to infinitary consequence relations, we might replace (T) by:

$$\frac{\Gamma, \Xi \vdash T, \Delta \text{ for all } \Xi \text{ and } T \text{ such that } \Xi \cup T = \emptyset}{\Gamma \vdash \Delta}$$

However, we shall officially be working with (T).

Now, we may note with pleasure that the law of contraposition holds for \Vdash_v :

$$\phi \Vdash_v \psi \iff \neg \psi \Vdash_v \neg \phi$$

and that equivalence (with respect to V) can be defined as mutual consequence:

$$\phi \simeq_V \psi \iff \phi \Vdash_v \psi \text{ and } \psi \Vdash_v \phi$$

In particular, logical equivalence is mutual logical consequence. Also

$$\phi \wedge \psi \simeq_V \phi \iff \phi \vee \psi \simeq_V \psi \iff \phi \vdash_V \psi$$

which is the kind of way you want \wedge, \vee , equivalence and consequence to fit together.

If consequence were defined simply as truth preservation from left to right, then these facts would break down. Moreover, we have lost nothing, since, when we want to, we can easily define this half-way-round consequence relation in terms of \vdash_V by:

$$\Gamma \vdash_V *, \Delta$$

To get back to \vdash_V again, we may conjoin this with the other half:

$$\Gamma \vdash_V \Delta \iff \Gamma \vdash_V *, \Delta \text{ and } \Gamma, * \vdash_V \Delta .$$

Another important fact about the chosen definition of consequence is that no reliance has been made on any manoeuvre such as dividing up the classifications $\top, *$ and \perp between 'designated' and 'undesignated' values. This would be out of place, of course, since we want to regard $*$ as a 'value' only in our system of representing partial truth-functions on $\{\top, \perp\}$. If we had adopted one of the half-way-round relations — either truth preservation from left to right or falsehood preservation from right to left — then we might have been open to the charge of relying implicitly on such a conception, making $*$ an undesignated or designated value. As it is, we are manifestly invoking the two values in such a way that $*$ cannot be made to side with either of them as really a 'kind of truth' or 'kind of falsity'. So, at least there would appear to be nothing about our definition of consequence which would make the logic after all just 'three-valued'.

We can now write down some more principles concerning a relation \vdash , between finite sets of sentences, to add to the general conditions (R), (M) and (T).

for " $\phi, \neg\phi \vdash \neg\phi, \phi$ " and
" $\phi, \neg\phi \vdash \neg\psi, \psi$ "

Firstly, two 'axioms' - i.e. unconditional principles - for \top and \perp :

$$\top \vdash \quad \perp \vdash$$

For negation we state a rule - or, if you like, a pair of rules - which we shall call 'twist':

$$\frac{\Gamma \vdash \neg \Delta}{\Delta \vdash \neg \Gamma} \quad \frac{\neg \Gamma \vdash \Delta}{\neg \Delta \vdash \Gamma}$$

and also an axiom:

$$\phi, \neg \phi \vdash \neg \phi, \phi$$

Here, and throughout, if Θ is a set of formulae, we use ' $\neg \Theta$ ' to mean $\{\neg \theta \mid \theta \in \Theta\}$. $*$ is then governed by the following:

$$* \vdash \neg * \quad \neg * \vdash *$$

Conjunction and disjunction obey familiar looking principles:

$$\frac{\Gamma, \phi, \psi \vdash \Delta}{\Gamma, \phi \wedge \psi \vdash \Delta} \quad \frac{\Gamma \vdash \phi, \psi, \Delta}{\Gamma \vdash \phi \vee \psi, \Delta}$$

The double line means that these conditions go upwards as well as downwards. And, exploiting $*$ to split consequence into its two halves, for interjunction we have:

$$\frac{\Gamma, \phi, \psi \vdash *, \Delta}{\Gamma, \phi * \psi \vdash *, \Delta} \quad \frac{\Gamma, * \vdash \phi, \psi, \Delta}{\Gamma, * \vdash \phi * \psi, \Delta}$$

It is easy to check that, if $\vdash = \frac{\vdash}{V}$ for some V , then all these conditions are satisfied. Furthermore, it turns out that they constitute an adequate set of logical laws for the connectives. A 'theory' shall be any relation between finite sets of sentences satisfying (R), (M) and (T) and the principles listed above, and we shall be able to prove that \vdash is a theory iff $\vdash = \frac{\vdash}{V}$ for some V . This is one way of stating the completeness theorem, and the 'if' part (soundness) has now already been dealt with. (However these definitions remain provisional and will be generalized in Section 2.)

Note that / will not be taken as primitive in this chapter.

You may have noticed that logical truths – formulae true under all assignments – are rather thin on the ground. Though a given V may be such that $\vdash_V \phi$ (i.e. $\emptyset \vdash_V \phi$) for several ϕ , if V comprises all assignments, then there is going to be nothing very interesting. Indeed, no formula not containing \top or \perp can ever be a logical truth. However, this dearth of logical truths should not, it seems to me, cause any distress. There is surely no theoretical nor practical difficulty, since we are making consequence rather than truth the central notion to work with.

Obtaining a full stock of logical truths, and in particular preserving classically valid formulae, is quoted as a good feature of techniques such as the supervaluational one. However, in my view, this involves a misplaced conservatism. It might be too sweeping to say that logical truths are a superficial feature of classical logic, but at least it seems to me to be rather perverse to cling to a feature which classical logic possesses precisely in virtue of the fact that it is total-valued, when you describe yourself as attempting a partial-valued generalization of that logic – especially when this is done at the expense of features which surely are more essential to the character of classical semantics, most especially, of course, extensionality.

For a page or so of views opposed to mine, both over this issue and also others discussed in earlier chapters, see R.H. Thomason (1972). What is relevant here is the argument for preserving classically valid formulae on the grounds that "In so far as the logic is meant to account for valid reasoning, the valid formulas are crucial since they draw the distinction between good and bad reasoning." I would disagree: I see no reason why truths should be expected to play this role. If canons of reasoning are indeed what you want best preserved in a 'generalization' of total to partial-valued logic, then you should be interested rather in the laws of consequence – or, more generally, in the whole web of logical relationships between formulae.

In any case, there is a very natural way in which classical validity is exactly preserved in our logic; for if ϕ is a classical formula — i.e. contains no occurrence of \times (or / or $*$) — then it is classically valid if and only if

$$\{p_i \vee \neg p_i\} \vdash_V \phi$$

where the p_i are the atomic components of ϕ and V contains all possible assignments. Equivalently, this principle holds in all theories, i.e.

$$\{p_i \vee \neg p_i\} \vdash_V \phi \quad \text{for any set } V \text{ of assignments .}$$

This fact is non-trivial in our framework, just because we have not trivialized $\phi \vee \neg \phi$ into a logical truth. Is there not a certain charm in having $\phi \vee \neg \phi$ true (in a theory) iff ϕ is either true or false (in that theory)?

In our framework the importance of consequence over truth goes deeper than just pure logic, since, in general, theories will have to be determined directly in terms of their consequence relation rather than the set of 'theorems' — sentences true-in-the-theory. It is only a special subclass of theories that can be determined by truths, as we shall see in Section 4. The point is that we do not have any mode of composition ' $p \Rightarrow q$ ' such that in all theories \vdash

$$\phi \vdash \psi \iff \vdash \phi \Rightarrow \psi$$

Nor even, given ϕ and ψ in a particular theory, are we necessarily guaranteed the existence of any formula playing the role of $\phi \Rightarrow \psi$: as we shall also see, such a formula exists only under special circumstances. Hence, in general, we cannot encode consequence in terms of truth.

Introducing an operator to do this would not necessarily lead us to adopt a non-monotonic mode of composition; it might rather prompt the first step into a kind of intensional logic. However, I will not be going into this idea. Let us see how we get along — here and in Chapter IV — without

for "axioms which" read "axioms, which"

a consequence-encoding conditional of any sort.

III.2 THEORIES AND MODELS — PRELIMINARIES

For the purposes of this chapter, it will be convenient to liberalize our understanding of what a language is. By a language L I shall mean a set which contains two particular elements, \top and \perp , and is closed under a unary operation \neg and three binary operations \wedge , \vee and \ast . \ast can then be defined by $\top \ast \perp$. We are not assuming that L is built up from a stock of atomic sentences, nor anything about it, apart from what has been stated. Even so, elements of L will be granted the courtesy title 'sentence'. The language defined in I.3, and used in the introduction, is just one particular instance of a language in this abstract sense. We know another interesting kind of language too, viz. systems of convex sets — which were defined in Section I.4.

We can now define a theory in L to be a set Γ of pairs of finite sets of sentences of L satisfying (R), (M) and (T) and the laws for the connectives listed in the introduction. These conditions continue to make good sense for our generalized notion of language. To avoid writing vertically, we let ' \top ' rotate through 90° : $\top, \neg, \wedge, \vdash$. A relation \vdash , then, will be a theory if and only if it is a theory in some language: \vdash clearly determines a particular language and we shall denote it by $L(\vdash)$.

Given the general principles (R), (M) and (T), the rules for \wedge , \vee and \ast could equivalently be replaced by the following axioms which are sometimes easier to work with

$$\begin{array}{ll}
 \phi, \psi \vdash \phi \wedge \psi & \phi \vee \psi \vdash \phi, \psi \\
 \phi \wedge \psi \vdash \phi & \phi \vdash \phi \vee \psi \\
 \phi \wedge \psi \vdash \psi & \psi \vdash \phi \vee \psi \\
 \\
 \phi, \psi \vdash \phi \ast \psi & \phi \ast \psi \vdash \phi, \psi \\
 \phi \ast \psi \vdash \phi, \ast & \ast, \phi \vdash \phi \ast \psi \\
 \phi \ast \psi \vdash \psi, \ast & \ast, \psi \vdash \phi \ast \psi
 \end{array}$$

Furthermore it is not difficult to show that twist as stated is equivalent to a modified form of twist and 'double negation':

$$\begin{array}{ccc} \phi \vdash \neg\neg\phi & & \Gamma \vdash \Delta \\ \neg\neg\phi \vdash \phi & & \hline \neg\Delta \vdash \neg\Gamma \end{array}$$

while the axiom for negation is equivalent to either of the following:

$$* \vdash \phi, \neg\phi \qquad \phi, \neg\phi \vdash *$$

We can then deduce principles which specify the character of the binary connectives within the immediate scope of negation:

$$\begin{array}{ccc} \neg\phi, \neg\psi \vdash \neg(\phi \vee \psi) & & \neg(\phi \wedge \psi) \vdash \neg\psi, \neg\phi \\ \neg(\phi \vee \psi) \vdash \neg\phi & & \neg\phi \vdash \neg(\phi \vee \psi) \\ \neg(\phi \vee \psi) \vdash \neg\psi & & \neg\psi \vdash \neg(\phi \wedge \psi) \\ \\ \neg\phi, \neg\psi \vdash \neg(\phi * \psi) & & \neg(\phi * \psi) \vdash \neg\phi, \neg\psi \\ \neg(\phi * \psi) \vdash \neg\phi, * & & *, \neg\phi \vdash \neg(\phi * \psi) \\ \neg(\phi * \psi) \vdash \neg\psi, * & & *, \neg\psi \vdash \neg(\phi * \psi) \end{array}$$

In the presence of principles such as these, we could in fact dispense with twist, if we simply wanted to give a system of pure logic – or of certain particular theories. But we could not treat of theories in general. There is, moreover, a presentation of logic which is both twist-free and cut-free (i.e. (T)-free) and has 'introduction rules' only – rules for introducing connectives with widest scope and rules for introducing connectives within the immediate scope of negation. However, there is no space to go into this; and the fact remains that any theory governed by our logic must be twist-able and cut-able, however we may be able to specify it proof-theoretically.

Interesting examples of theories are not hard to find. For example, consider the system of convex sets on an underlying boolean language B , given in terms of a consequence relation \models for B , and define a theory \Vdash

for the language of convex sets by stipulating that $X_1, \dots, X_m \Vdash Y_1, \dots, Y_n$ iff

$$\exists j \forall b \in Y_j \exists a_1 \in X_1, \dots, \exists a_m \in X_m . a_1 \wedge \dots \wedge a_m \vdash b$$

and

$$\exists i \forall b \in X_i \exists b_1 \in Y_1, \dots, \exists b_n \in Y_n . a \vdash b_1 \vee \dots \vee b_n$$

This is not really as obscure as it looks. It is simply the required generalization of ' $X \Vdash Y$ ', which holds iff every element of Y is a consequence of (some element of) X , and, dually, every element of X yields some element of Y . The theory \Vdash thus defined we shall refer to as the canonical theory for a given system of convex sets.

It will be useful to have a notion of sublanguage: L_1 is a sublanguage of L_2 iff $L_1 \subseteq L_2$ and L_1 is a language — i.e. satisfies the required closure conditions. Note that if \vdash is a theory in L_2 and L_1 is a sublanguage of L_2 , then the relation obtained from \vdash by leaving in only those pairs of sets of sentences all of whose constituent sentences are in L_1 is still a theory. For example, the finitely specifiable convex sets constitute a sublanguage of a system of convex sets and the canonical theory restricted to them is a theory.

A theory has been defined as a relation between pairs of finite sets, which we can write ' $\Gamma \succ \Delta$ ' and call sequents (to contrast with the general notation ' $\langle \Gamma, \Delta \rangle$ '). Nonetheless, it will subsequently make life a lot easier if we extend the turnstile notation to hold between arbitrary pairs of sets of sentences and write ' $\Gamma \vdash \Delta$ ' to mean that $\Gamma_0 \succ \Delta_0 \in \vdash$ for some finite subsets Γ_0 and Δ_0 of Γ and Δ . It is easy to check that, if \vdash is a theory, then all the laws continue to hold as displayed, even when Γ and Δ are not finite. We shall use ' $\Gamma \not\vdash \Delta$ ' to mean 'not $\Gamma \vdash \Delta$ '.

Not all languages contain an interesting built-in structure, as convex sets do, so we might want to impose explicit conditions to determine a

theory, using 'non-logical axioms'. Let Σ be an arbitrary set of sequents, then we can define $\bar{\Sigma}$ — the theory axiomatized by Σ — to be the smallest theory containing Σ . This specification of $\bar{\Sigma}$ is, of course, relative to a particular language L , which we assume contains all sentences occurring in sequents in Σ , and can be given by:

$$\bar{\Sigma} = \bigcap \{ \vdash \mid \vdash \text{ is a theory in } L \text{ and } \Sigma \subseteq \vdash \}$$

We must check that this is actually a theory, but that is easy. In fact we can make the following two important statements:

Lemma III.2.1: Σ is a theory iff $\Sigma = \bar{\Sigma}$

Lemma III.2.2: $\Gamma \succ \Delta \in \bar{\Sigma}$ iff there is a finite subset Σ_0 of Σ such that $\Gamma \succ \Delta \in \bar{\Sigma}_0$.

There are two extreme kinds of theory. At one extreme a theory may contain all the sequents of its language; but such a theory is pretty useless. Let us say that a set Σ of sequents is consistent iff $\emptyset \succ \emptyset \notin \bar{\Sigma}$, and that Σ is inconsistent if it is not consistent. Then clearly, by (M), Σ is inconsistent iff $\bar{\Sigma}$ is the whole set of sequents.

The other extreme is the smallest theory for a given language L , viz. $\bar{\emptyset}$. This theory is logic for L . So, in particular, if L is a schematic atomic language, logic is just what you would expect it to be. But note that the canonical theory for a system of convex sets is not logic, unless \top , $*$ and \perp are the only sentences. This is because there are boolean sets X , not equivalent to \top or \perp , but such that $\Vdash X \vee \neg X$. These facts will be easy to check, once we have some semantical apparatus. Our definitions have in fact been sufficiently liberal to allow for languages in which $\bar{\emptyset}$ is obviously the inconsistent theory (if $\top = \perp$, say), but languages for which logic is inconsistent are not going to be of much interest.

Let us say 'model' instead of 'valuation', and write 'M' instead of 'v', to get in the mood for chapter IV. But a model M for L is just a function $L \rightarrow \{\top, \star, \perp\}$ which satisfies the required interpretation clauses, viz.

$$\begin{aligned}
 M(\top) = \top &\iff M(\perp) = \perp \\
 M(\neg\phi) = \top &\iff M(\phi) = \perp \\
 &= \perp \iff M(\phi) = \top \\
 M(\phi \wedge \psi) = \top &\iff M(\phi) = \top \text{ and } M(\psi) = \top \\
 &= \perp \iff M(\phi) = \perp \text{ or } M(\psi) = \perp \\
 M(\phi \vee \psi) = \top &\iff M(\phi) = \top \text{ or } M(\psi) = \top \\
 &= \perp \iff M(\phi) = \perp \text{ and } M(\psi) = \perp \\
 M(\phi \times \psi) = \top &\iff M(\phi) = \top \text{ and } M(\psi) = \top \\
 &= \perp \iff M(\phi) = \perp \text{ and } M(\psi) = \perp
 \end{aligned}$$

If L is an atomic language, with atoms L_0 , then the valuation \bar{v} defined from a basic assignment $v: L_0 \rightarrow \{\top, \star, \perp\}$ is a model for L . In fact:

Lemma III.2.3: There is a one-one correspondence between models for L and assignments to L_0 , which is given by

$$M \leftrightarrow M|L_0 \text{ (i.e. } M \text{ restricted to } L_0) \text{ and } \bar{v} \leftrightarrow v$$

Let us now generalize the definition given in the introduction of 'is consistent with' to hold between models for a language and (arbitrary) pairs of sets of sentences. We say that M is consistent with $\langle \Gamma, \Delta \rangle$, and write ' $\Gamma \Vdash_M \Delta$ ', iff

$$\begin{aligned}
 &\text{neither } M(\phi) = \top \text{ for all } \phi \in \Gamma \text{ and } M(\psi) \neq \top \text{ for all } \psi \in \Delta \\
 &\text{nor } M(\phi) \neq \perp \text{ for all } \phi \in \Gamma \text{ and } M(\psi) = \perp \text{ for all } \psi \in \Delta
 \end{aligned}$$

This is the fundamental definition connecting models and sequents, and we can exploit it in two ways.

Firstly, to give a definition of what it is to be a model of a set

of sequents — in particular a theory — we say that M is a model of Σ iff M is consistent with every sequent in Σ . We shall use ' $K(\Sigma)$ ' to denote the set of such models. (Note: 'model for L '; 'model of Σ — or \vdash ').

The soundness of our logical laws can now be stated in the following form:

Theorem III.2.4 (Soundness (1)): $K(\Sigma) = K(\bar{\Sigma})$

In other words, a model is consistent with every sequent in Σ iff it is consistent with every sequent which follows from Σ . That $K(\bar{\Sigma}) \subseteq K(\Sigma)$ is utterly trivial; the important fact is that $K(\Sigma) \subseteq K(\bar{\Sigma})$, but this is easy to check.

No model is consistent with $\emptyset \succ \emptyset$, and we know that if $\bar{\Sigma}$ has no model, nor does Σ ; so

Corollary III.2.5: If Σ has a model, then Σ is consistent

Happily logic for our schematic atomic languages is always consistent: there are plenty of assignments ν , and they all induce models $\bar{\nu}$ of $\bar{\emptyset}$.

$K(\Sigma)$ is the set of models determined by a set of sequents. The other way about, generalizing the definition we have of ' $\frac{\vdash}{V}$ ', we have a set of sequents determined by a set of models. Let K be any set of models (for some particular language) and define

$$\frac{\vdash}{K} = \{ \Gamma \succ \Delta \mid \Gamma \Vdash_M \Delta \text{ for all } M \in K \}$$

This we call the theory of K , since $\frac{\vdash}{K}$ is a theory: by (III.2.4), $K \subseteq K(\frac{\vdash}{K}) \subseteq K(\overline{\frac{\vdash}{K}})$, and so $\overline{\frac{\vdash}{K}} = \frac{\vdash}{K}$. This is another version of 'soundness':

Corollary III.2.6 (Soundness (2)): $\frac{\vdash}{K}$ is a theory.

Ring the changes in notation yet again:

Corollary III.2.7 (Soundness (3)): $\bar{\Sigma} \subseteq \frac{\vdash}{K(\Sigma)}$

But is the converse of (III.2.7) true? This is an obvious form of the completeness question: does a sequent follow from Σ only if any model of Σ is consistent with it? A positive answer to this question would yield the (required generalization of) the claim made in the introduction that any theory is determined by some set of models. This is because a theory \vdash would be determined by $K(\vdash)$ — the set of precisely those models which are models of \vdash . In fact we can attack the question simply by asking about $K(\vdash)$, and make our canonical statement of completeness the statement that a theory is determined by its models:

Theorem III.2.8 (Completeness): $\vdash = \overline{K(\vdash)}$ for any theory \vdash .

That $\vdash \subseteq \overline{K(\vdash)}$ is utterly trivial and does not even depend on soundness; what is important is that $\overline{K(\vdash)} \subseteq \vdash$.

Given completeness in this form, then, using (III.2.4), we can immediately deduce the desired strengthening of (III.2.7):

Corollary III.2.9: $\bar{\Sigma} = \overline{K(\Sigma)}$

The 'converse' of (III.2.6) holds too:

Corollary III.2.10: Σ is a theory iff $\Sigma = \overline{K}$ for some K .

And the consistency lemma (III.2.5) gets a converse: if there is a sequent not in $\bar{\Sigma}$, then there must be a model in $K(\Sigma)$ which is inconsistent with it.

Corollary III.2.11: Σ is consistent iff Σ has a model.

To prove (III.2.8) we must show that for any sequent not contained in a theory there is a model of that theory which is not consistent with it. In fact we shall show something stronger: that for arbitrary Γ and Δ , if $\Gamma \not\vdash \Delta$, then $\Gamma \not\vdash_M \Delta$ for some model M of \vdash . One final piece of new notation: let \Vdash_K be the relation

$$\{ \langle \Gamma, \Delta \rangle \mid \Gamma \Vdash_M \Delta \text{ for all } M \in K \}$$

So, note that \vdash_K is \Vdash_K restricted to sequents. What we shall show is

Theorem III.2.12 (Strong Completeness): $\Gamma \vdash \Delta$ iff $\Gamma \Vdash_{K(\vdash)} \Delta$.

'Only if' is trivial; the guts of the matter is that $\Gamma \not\vdash \Delta$ implies $\Gamma \not\Vdash_{K(\vdash)} \Delta$.

From this, bearing in mind (III.2.4), we shall be able to deduce a form of compactness:

Corollary III.2.13 (Sentence compactness): For any set Σ of sequents:

$$\Gamma \Vdash_{K(\Sigma)} \Delta \text{ iff } \Gamma \vdash_{K(\Sigma)} \Delta$$

A special case of this corollary is compactness for logical consequence, since, if K is the set of all models for a language, then \Vdash_K is logical consequence generalized to hold between arbitrary sets of sentences. But then $K = K(\emptyset)$, so

$$\Gamma \Vdash_K \Delta \text{ iff } \Gamma_0 \Vdash_K \Delta_0 \text{ for some finite subsets } \Gamma_0 \text{ and } \Delta_0 \text{ of } \Gamma \text{ and } \Delta.$$

Of course, given completeness and (III.2.2), we also have

Lemma III.2.14 (Sequent compactness): Σ has a model iff every finite subset of Σ has a model.

To return to the completeness problem, we must show that if $\Gamma \not\vdash \Delta$ then some model of \vdash is not consistent with $\langle \Gamma, \Delta \rangle$. If $\Gamma \not\vdash \Delta$ we say that $\langle \Gamma, \Delta \rangle$ is \vdash -rejected. The strategy will be the obvious one of extending a \vdash -rejected pair $\langle \Gamma, \Delta \rangle$ to a pair $\langle \Gamma', \Delta' \rangle$ from which we can deduce the existence of the required model. Let us say that $\langle \mathcal{E}, \mathcal{T} \rangle$ is exhaustive (in $L(\vdash)$) iff $\mathcal{E} \cup \mathcal{T} = L(\vdash)$. What we shall show is that any \vdash -rejected pair can be extended to one which is still \vdash -rejected but is exhaustive.

The connection with models is made by the following lemma. Call a pair $\langle \mathcal{E}, \mathcal{T} \rangle$ \ast -right iff $\ast \in \mathcal{T}$, and \ast -left iff $\ast \in \mathcal{E}$, then

Lemma III.2.15: There is a one-one correspondence between

- (i) \ast -right exhaustive \vdash -rejected pairs
- (ii) \ast -left exhaustive \vdash -rejected pairs
- (iii) models of \vdash

Between (i) and (iii) this is given by

$$\langle \Xi, \top \rangle \leftrightarrow M_r \quad \text{where} \quad M_r(\phi) = \top \quad \text{iff} \quad \phi \in \Xi \\ = \perp \quad \text{iff} \quad \neg\phi \in \Xi$$

$$\langle \{\phi | M(\phi) = \top\}, \{\phi | M(\phi) \neq \top\} \rangle \leftrightarrow M$$

and between (ii) and (iii) by

$$\langle \Xi, \top \rangle \leftrightarrow M_l \quad \text{where} \quad M_l(\phi) = \top \quad \text{iff} \quad \neg\phi \in \top \\ = \perp \quad \text{iff} \quad \phi \in \top$$

$$\langle \{\phi | M(\phi) \neq \perp\}, \{\phi | M(\phi) = \perp\} \rangle \leftrightarrow M$$

Proof: We consider a $*$ -right exhaustive \vdash -rejected pair and show that M_r as defined is indeed a model of \vdash . $*$ -left pairs work dually. That the correspondences are one-one, with the inverses given, is then easy to check.

Firstly, from the conditions for being a theory and the further principles we recorded earlier (on pages 115, 116) we can deduce that, if $\langle \Xi, \top \rangle$ is $*$ -right, \vdash -rejected and exhaustive, then for any ϕ and ψ

$$* \notin \Xi, \quad \neg * \notin \Xi, \quad \top \in \Xi, \quad \neg \perp \in \Xi$$

$$\neg\neg\phi \in \Xi \iff \phi \in \Xi$$

$$\phi \wedge \psi \in \Xi \iff \phi \in \Xi \text{ and } \psi \in \Xi$$

$$\neg(\phi \wedge \psi) \in \Xi \iff \neg\phi \in \Xi \text{ or } \neg\psi \in \Xi$$

$$\phi \vee \psi \in \Xi \iff \phi \in \Xi \text{ or } \psi \in \Xi$$

$$\neg(\phi \vee \psi) \in \Xi \iff \neg\phi \in \Xi \text{ and } \neg\psi \in \Xi$$

$$\phi \times \psi \in \Xi \iff \phi \in \Xi \text{ and } \psi \in \Xi$$

$$\neg(\phi \times \psi) \in \Xi \iff \neg\phi \in \Xi \text{ and } \neg\psi \in \Xi$$

Hence M_r is a model for $L(\vdash)$.

To see that it is, furthermore, a model of \vdash , assume that $\Gamma \not\vdash_{M_r} \Delta$ for some finite Γ and Δ : we shall show that $\Gamma \not\vdash \Delta$. There are two possibilities:

$$(a) \quad \Gamma \not\vdash_{M_r} *, \Delta \qquad (b) \quad \Gamma, * \not\vdash_{M_r} \Delta$$

In case (a) $\Gamma \subseteq \Xi$ and $\Delta \cup \{*\} \subseteq \top$. Hence $\Gamma \not\vdash *, \Delta$. Therefore $\Gamma \not\vdash \Delta$.

In case (b) $\neg\Delta \subseteq \exists$ and $\neg\Gamma \cup \{*\} \subseteq \top$. Hence $\neg\Delta \not\vdash *$, $\neg\Gamma$.

Therefore by our modified form of twist, $\Gamma, * \not\vdash \Delta$; and so $\Gamma \not\vdash \Delta$. \square

All that remains to prove strong completeness, and everything which follows, is to show

Lemma III.2.16: If $\Gamma \not\vdash \Delta$ then there exists an exhaustive \vdash -rejected pair $\langle \Gamma', \Delta' \rangle$ such that $\Gamma \subseteq \Gamma'$ and $\Delta \subseteq \Delta'$.

Once we have this, (III.2.12) is proved, since the model corresponding to $\langle \Gamma', \Delta' \rangle$, whose existence is guaranteed by (III.2.15), will not be consistent with $\langle \Gamma', \Delta' \rangle$ and hence not consistent with $\langle \Gamma, \Delta \rangle$.

We shall not stop to establish this: it will be corollary of the more general result which is central to the next section. This result will be used to provide 'model-theoretic' conditions which help us with interpolation, definability, and with further questions — peculiar to our partial-valued framework — concerning the expressive character of theories in general and logic in particular. Moreover, the methods will serve as an introduction to the analogous investigation into quantifier logic in Chapter IV. In fact, where extending a result to quantifier languages involves nothing essentially new, we shall not always repeat it.

However, there are some important divergencies between theories in the two kinds of language. We conclude with one of these. For propositional theories the following lemma provides alternative equivalent definitions of what it is for a theory to be 'complete' (assuming it is consistent):

Lemma III.2.17: The following are equivalent

- (i) \vdash is maximally consistent
- (iia) For all $\phi \in L(\vdash)$, $\vdash \phi$ or $\phi \vdash *$
- (iib) For all $\phi \in L(\vdash)$, $\phi \vdash$ or $* \vdash \phi$
- (iic) For all $\phi \in L(\vdash)$, $\vdash \phi$ or $\phi \vdash$ or both $* \vdash \phi$ and $\phi \vdash *$

In contrast, there will be two non-equivalent notions of 'complete theory' in Chapter IV. In the present context, however, the proof of (III.2.17) is easy, and we can make the connection with models also:

Lemma III.2.18: There is a one-one correspondence between models of \vdash and maximally consistent extensions of \vdash given by

$$M \leftrightarrow \overline{\vdash_{\{M\}}}$$

III.3 THEORIES AND SENTENCES

We begin by stating and proving the required generalization of (III.2.16). Given two theories \vdash_1 and \vdash_2 , whose respective languages — $L(\vdash_1)$ and $L(\vdash_2)$ — are, we may assume, sublanguages of some big language L , consider a set of sentences $\Lambda \subseteq L(\vdash_1) \cap L(\vdash_2)$ which contains \top and \perp and is closed under \wedge and \vee . We are interested in the following general interpolation condition:

$$\exists \lambda \in \Lambda. \Gamma \vdash_1 \lambda, \Delta \text{ and } \Xi, \lambda \vdash_2 \top$$

where Γ and Δ are subsets of $L(\vdash_1)$ and Ξ and \top subsets of $L(\vdash_2)$. Let us say that $\langle \Gamma, \Delta, \Xi, \top \rangle$ excludes Λ iff this condition fails. Our main lemma provides a necessary and sufficient condition for exclusion.

Lemma III.3.1 (Interpolant Exclusion): $\langle \Gamma, \Delta, \Xi, \top \rangle$ excludes Λ iff:

there exist Γ', Δ', Ξ' and \top' such that

$$\Gamma \subseteq \Gamma' \subseteq L(\vdash_1); \Delta \subseteq \Delta' \subseteq L(\vdash_1); \Xi \subseteq \Xi' \subseteq L(\vdash_2); \top \subseteq \top' \subseteq L(\vdash_2)$$

$\langle \Gamma', \Delta' \rangle$ is exhaustive in $L(\vdash_1)$ and \vdash_1 -rejected

$\langle \Xi', \top' \rangle$ is exhaustive in $L(\vdash_2)$ and \vdash_2 -rejected

and

$$\Gamma' \cap \Delta' \cap \top' = \emptyset$$

Proof: To prove the sufficiency of this condition, it is enough to observe that, since $\langle \Gamma', \Delta' \rangle$ and $\langle \Xi', \top' \rangle$ are exhaustive and rejected in their

respective theories, the fact that $\Gamma' \cap \Lambda \cap T' = \emptyset$ implies that $\langle \Gamma', \Delta', \Xi', T' \rangle$ excludes Λ ; and so $\langle \Gamma, \Delta, \Xi, T \rangle$ excludes Λ also. This depends simply on (R), (M) and (T).

Necessity is less trivial: we apply Zorn's Lemma to the set of quadruples extending $\langle \Gamma, \Delta, \Xi, T \rangle$ and excluding Λ . Let us define

$$\langle X, Y, U, V \rangle \leq \langle X', Y', U', V' \rangle \text{ iff } X \subseteq X', Y \subseteq Y', U \subseteq U' \text{ and } V \subseteq V'$$

Then, assuming that in all quadruples the first two coordinates are subsets of $L(\frac{_}{1})$ and the second two are subsets of $L(\frac{_}{2})$, the set S which we are interested in is

$$\{ \langle X, Y, U, V \rangle \mid \langle \Gamma, \Delta, \Xi, T \rangle \leq \langle X, Y, U, V \rangle \text{ and } \langle X, Y, U, V \rangle \text{ excludes } \Lambda \}$$

It is easy to check that, because theories are relations between finite sets, any \leq -chain in S has an upper bound in S . Hence S contains an element $\langle \Gamma', \Delta', \Xi', T' \rangle$ which is maximal with respect to \leq .

This quadruple will do. It excludes Λ ; hence, since \top and \perp are elements of Λ ,

$$\langle \Gamma', \Delta' \rangle \text{ is } \frac{_}{1}\text{-rejected} \quad \langle \Xi', T' \rangle \text{ is } \frac{_}{2}\text{-rejected}$$

Also, from principle (R),

$$\Xi' \cap \Lambda \cap T' = \emptyset.$$

It remains only to show that $\langle \Gamma', \Delta' \rangle$ and $\langle \Xi', T' \rangle$ are exhaustive in their respective languages. Consider $\langle \Gamma', \Delta' \rangle$ and suppose it were not exhaustive; then there would be a $\phi \in L(\frac{_}{1})$ such that $\phi \notin \Gamma' \cup \Delta'$. By the \leq -maximality of $\langle \Gamma', \Delta', \Xi', T' \rangle$, both $\langle \Gamma' \cup \{\phi\}, \Delta', \Xi', T' \rangle$ and $\langle \Gamma', \Delta' \cup \{\phi\}, \Xi', T' \rangle$ would fail to exclude Λ , and so there would be $\lambda_1, \lambda_2 \in \Lambda$ such that all the following held:

$$\Gamma', \phi \frac{_}{1} \lambda_1, \Delta' \quad \Xi', \lambda_1 \frac{_}{2} T'$$

$$\Gamma' \frac{_}{1} \phi, \lambda_2, \Delta' \quad \Xi', \lambda_2 \frac{_}{2} T'$$

But then, by (T) and the principles for disjunction,

$$\Gamma' \frac{_}{1} \lambda_1 \vee \lambda_2, \Delta' \quad \Xi', \lambda_1 \vee \lambda_2 \frac{_}{2} T'$$

However, Λ is closed under disjunction, and so $\lambda_1 \vee \lambda_2 \in \Lambda$: this contradicts the fact that $\langle \Gamma', \Delta', \Xi', T' \rangle$ excludes Λ . Hence $\langle \Gamma', \Delta' \rangle$ is exhaustive.

By exactly parallel reasoning, since Λ is closed under conjunction, $\langle \Xi', \mathcal{T}' \rangle$ is also exhaustive. \square

Corollary III.3.2: Lemma (III.2.16) holds

Proof: If $\Gamma \not\vdash \Delta$, put $\vdash = \vdash_1 = \vdash_2$ and $\Lambda = L(\vdash_1) = L(\vdash_2)$: then, by (T), $\langle \Gamma, \Delta, \emptyset, \emptyset \rangle$ excludes Λ . \square

Hence we have strong completeness and everything which follows.

Let us now introduce some obvious abbreviations for relations between models M_1 for L_1 and M_2 for L_2 , with $\Lambda \subseteq L_1 \cap L_2$:

$$\begin{aligned} M_1 \sqsubseteq_{\Lambda} M_2 &\iff \forall \lambda \in \Lambda. M_1(\lambda) \sqsubseteq M_2(\lambda) \\ M_1 \square_{\Lambda} M_2 &\iff \forall \lambda \in \Lambda. M_1(\lambda) \square M_2(\lambda) \\ M_1 \equiv_{\Lambda} M_2 &\iff \forall \lambda \in \Lambda. M_1(\lambda) = M_2(\lambda) \neq * \end{aligned}$$

When $\Lambda = L_1 = L_2$ we shall drop ' Λ '; and we shall allow ourselves to write $M_2 \supseteq_{\Lambda} M_1$ for $M_1 \sqsubseteq_{\Lambda} M_2$. Also, if L_1 and L_2 are atomic languages and Λ_0 is a set of atoms common to both L_1 and L_2 , then we define:

$$\begin{aligned} v_1 \sqsubseteq_{\Lambda_0} v_2 &\iff \forall \lambda \in \Lambda_0. v_1(\lambda) \sqsubseteq v_2(\lambda) \\ v_1 \square_{\Lambda_0} v_2 &\iff \forall \lambda \in \Lambda_0. v_1(\lambda) \square v_2(\lambda) \\ v_1 \equiv_{\Lambda_0} v_2 &\iff \forall \lambda \in \Lambda_0. v_1(\lambda) = v_2(\lambda) \neq * \end{aligned}$$

Distinguish \equiv_{Λ_0} from $=_{\Lambda_0}$:

$$v_1 =_{\Lambda_0} v_2 \iff \forall \lambda \in \Lambda_0. v_1(\lambda) = v_2(\lambda)$$

We shall be interested in two classes of sentences based on such a Λ_0 : the class of classical formulae — ones built up from Λ_0 using \neg , \wedge and \vee ; and the class of all formulae built up from Λ_0 , which is closed under $*$ as well. The following lemma then makes an important connection between the two lists of definitions.

Lemma III.3.3: If Λ is either of the above classes

$$\begin{aligned} v_1 \sqsubseteq_{\Lambda_0} v_2 &\iff \bar{v}_1 \sqsubseteq_{\Lambda} \bar{v}_2 \\ v_1 \square_{\Lambda_0} v_2 &\iff \bar{v}_1 \square_{\Lambda} \bar{v}_2 \end{aligned}$$

In addition, if Λ is the class of classical formulae

$$v_1 \equiv_{\Lambda_0} v_2 \iff \bar{v}_1 \equiv_{\Lambda} \bar{v}_2$$

If we now assume that Λ is closed under \neg , we may apply the correspondences described in (III.2.15) to yield the following model-theoretic conditions, which match the four possibilities concerning the occurrence of $*$ in Λ -excluding quadruples.

Corollary III.3.4 (Interpolant-Excluding Model Pairs):

If Λ satisfies the conditions for (III.3.1) and is also closed under \neg , then, corresponding to the four cases

- | | |
|-------------------------------------|-------------------------------------|
| (1) $* \in \Delta$ and $* \in \top$ | (3) $* \in \Delta$ and $* \in \Xi$ |
| (2) $* \in \Gamma$ and $* \in \Xi$ | (4) $* \in \Gamma$ and $* \in \top$ |

we have, as necessary and sufficient conditions for $\langle \Gamma, \Delta, \Xi, \top \rangle$'s excluding Λ , that there exist models M_1 of \vdash_1 and M_2 of \vdash_2 such that

$$\Gamma \not\vdash_{M_1} \Delta \quad \text{and} \quad \Xi \not\vdash_{M_2} \top$$

satisfying respectively

- | | |
|-------------------------------------|---------------------------------|
| (1) $M_1 \sqsubseteq_{\Lambda} M_2$ | (3) $M_1 \square_{\Lambda} M_2$ |
| (2) $M_1 \supseteq_{\Lambda} M_2$ | (4) $M_1 \equiv_{\Lambda} M_2$ |

Proof: Apply the details of (III.2.15) to (III.3.1). \square

Taken together, these four cases can provide useful information about interpolants in general, since we can reassemble them in the following way:

Lemma III.3.5 (Combination Lemma (1)):

$$\exists \lambda \in \Lambda. \quad \Gamma \vdash_1 \lambda, \Delta \quad \text{and} \quad \Xi, \lambda \vdash_2 \top$$

iff the following all hold:

- | | |
|---|---------------------------------------|
| (1) $\exists \lambda_1 \in \Lambda. \quad \Gamma \vdash_1 \lambda_1, \Delta, *$ | and $\Xi, \lambda_1 \vdash_2 \top, *$ |
| (2) $\exists \lambda_2 \in \Lambda. \quad *, \Gamma \vdash_1 \lambda_2, \Delta$ | and $*, \Xi, \lambda_2 \vdash_2 \top$ |
| (3) $\exists \lambda_3 \in \Lambda. \quad \Gamma \vdash_1 \lambda_3, \Delta, *$ | and $*, \Xi, \lambda_3 \vdash_2 \top$ |
| (4) $\exists \lambda_4 \in \Lambda. \quad *, \Gamma \vdash_1 \lambda_4, \Delta$ | and $\Xi, \lambda_4 \vdash_2 \top, *$ |

Proof: 'only if' is trivial. For 'if' put $\lambda = (\lambda_1 \wedge \lambda_3) \vee (\lambda_2 \wedge \lambda_4)$.

Since we now have completeness ($\vdash = \vdash_{K(\vdash)}$), we need only check truth-falseness conditions to show that this definition works. Note that we have only used the fact that Λ is closed under \wedge and \vee . \square

Bringing (III.3.4) and (III.3.5) together:

Corollary III.3.6 (Interpolant Existence):

$$\exists \lambda \in \Lambda. \Gamma \vdash_1 \lambda, \Delta \quad \text{and} \quad \Xi, \lambda \vdash_2 \top$$

iff between models M_1 of \vdash_1 and M_2 of \vdash_2 the following all hold:

- (1) $M_1 \sqsubseteq_{\Lambda} M_2 \Rightarrow$ either $\Gamma \Vdash_{M_1} \Delta, *$ or $\Xi \Vdash_{M_2} \top, *$
- (2) $M_1 \supseteq_{\Lambda} M_2 \Rightarrow$ either $*, \Gamma \Vdash_{M_1} \Delta$ or $*, \Xi \Vdash_{M_2} \top$
- (3) $M_1 \square_{\Lambda} M_2 \Rightarrow$ either $\Gamma \Vdash_{M_1} \Delta, *$ or $*, \Xi \Vdash_{M_2} \top$
- (4) $M_1 \equiv_{\Lambda} M_2 \Rightarrow$ either $*, \Gamma \Vdash_{M_1} \Delta$ or $\Xi \Vdash_{M_2} \top, *$

We should note that case four of (III.3.4) is an extreme case, which cannot arise if Λ contains a sentence undefined in either of the theories — in particular if Λ contains $*$. Hence we may record the following simplifications of (III.3.5) and (III.3.6).

Lemma III.3.7: If Λ is closed under $*$, then condition (4) in (III.3.5) and (III.3.6) is trivially true and may be ignored.

Let us turn now to relations of degree-of-definedness and compatibility between sentences and generalize the definitions given in I.3, not only so that they apply to our generalized notion of language, but also to allow for relativity to a particular theory \vdash . Given completeness, it is a matter of indifference whether we do this in terms of \vdash or of $K(\vdash)$:

$$\phi \sqsubseteq_{\vdash} \psi \iff \forall M \in K(\vdash). M(\phi) \sqsubseteq M(\psi), \text{ equivalently } \phi \vdash \psi, * \text{ and } *, \psi \vdash \phi$$

$$\phi \square_{\vdash} \psi \iff \forall M \in K(\vdash). M(\phi) \square M(\psi), \text{ equivalently } \vdash \phi, \neg \psi \text{ and } \vdash \neg \phi, \psi$$

And let us also define \vdash -equivalence:

$$\phi \approx_{\vdash} \psi \iff \forall M \in K(\vdash). M(\phi) = M(\psi), \text{ equivalently } \phi \vdash \psi \text{ and } \psi \vdash \phi$$

If \vdash is logic we shall simply write \sqsubseteq, \square and \simeq .

Our final preparatory result is a second combination lemma, which we state under the simplifying assumption that Λ is closed under \ast .

Lemma III.3.8 (Combination Lemma (2)):

$$\exists \lambda \in \Lambda. \phi \sqsubseteq_{\vdash_1} \lambda \text{ and } \lambda \sqsubseteq_{\vdash_2} \psi$$

iff the following all hold:

$$\exists \lambda_1 \in \Lambda. \phi \vdash_1 \lambda_1, \ast \text{ and } \lambda_1 \vdash_2 \psi, \ast$$

$$\exists \lambda_2 \in \Lambda. \ast, \psi \vdash_2 \lambda_2 \text{ and } \ast, \lambda_2 \vdash_1 \phi$$

$$\exists \lambda_3 \in \Lambda. \phi \vdash_1 \lambda_3, \ast \text{ and } \ast, \lambda_3 \vdash_1 \phi$$

Proof: 'Only if' is trivial. For 'if' put $\lambda = ((\lambda_1 \wedge \lambda_3) \vee \lambda_2) \ast (\lambda_1 \wedge (\lambda_3 \vee \lambda_2))$

Our first two results make use of the following general 'persistence' theorem, which provides a model-theoretic condition for when a sentence ϕ is equivalent, in a particular theory, to a sentence in a given set Λ , which we assume contains \top and \perp and is closed under \neg, \wedge and \vee .

Theorem III.3.9: There exists a $\lambda \in \Lambda$ such that $\phi \approx_{\vdash} \lambda$ iff between any pair of models M and N of \vdash :

$$(a) \ M \sqsubseteq_{\Lambda} N \Rightarrow M(\phi) \sqsubseteq N(\phi)$$

$$(b) \ M \square_{\Lambda} N \Rightarrow M(\phi) \square N(\phi)$$

$$(c) \ M \equiv_{\Lambda} N \Rightarrow M(\phi) = N(\phi) \neq \ast$$

Proof: Since $\phi \approx_{\vdash} \lambda$ iff $\phi \vdash \lambda$ and $\lambda \vdash \phi$, this theorem turns out to be nothing but a special case of (III.3.6), with $\vdash_1 = \vdash_2 = \vdash$, $\Gamma = \top = \{\phi\}$ and $\Delta = \Xi = \emptyset$. Note that conditions (1) and (2) together correspond to condition (a). \square

for "III.3.4." and "III.3.12"

--
for "any" and "why"

Corollary III.3.10: If Λ is closed under \ast as well, then (III.3.9) holds with (c) deleted.

Proof: III.3.7

Hence we might record a condition for when any sentence of a theory \vdash is equivalent in \vdash to one in a sublanguage Λ of $L(\vdash)$:

Corollary III.3.11: Any sentence of $L(\vdash)$ is \vdash -equivalent to a sentence in Λ iff between models M and N of \vdash

$$(a) \quad M \sqsubseteq_{\Lambda} N \Rightarrow M \sqsubseteq N$$

$$(b) \quad M \square_{\Lambda} N \Rightarrow M \square N$$

For our first application we do not, however, ignore (c). Given a set Λ_0 of sentences we might ask when a sentence ϕ is equivalent to a 'boolean' function of elements of Λ_0 - i.e. ones obtained from Λ_0 by applying \neg, \wedge and \vee , but not \ast . (III.3.9) can provide an answer. This is especially interesting if $L(\vdash)$ is an atomic language and Λ_0 is the set of atoms.

Theorem III.3.4: ϕ is \vdash -equivalent to a classical formula iff for any valuation ν , if ν is totally defined and $\bar{\nu}$ is a model of \vdash , then $\bar{\nu}(\phi)$ is defined.

Proof: Recall (III.2.3) and consider (III.3.9) with Λ the set of classical formulae. Then apply (III.3.3), to show that conditions (a) and (b) are trivially true, and to transform (c) into the required condition. \square

Corollary III.3.13: ϕ is logically equivalent to a classical formula iff $\bar{\nu}(\phi)$ is defined for any total assignment ν .

for "total-value" and "total-value"

Remaining still with atomic languages, we might consider the matter of 'definability'. The obvious notion of 'explicit definability' is that an atomic sentence p is explicitly definable in \vdash iff there exists a p -free formula ϕ such that $p \approx_{\vdash} \phi$. Let Λ_0 be the atomic sentences of $L(\vdash)$, then we can provide the following model-theoretic criterion:

Theorem III.3.14: p is explicitly definable in \vdash iff for any assignments v and w such that \bar{v} and \bar{w} are models of \vdash :

$$(a) \quad v \sqsubseteq_{\Lambda_0 \setminus \{p\}} w \Rightarrow v \sqsubseteq w$$

$$(b) \quad v \sqsupset_{\Lambda_0 \setminus \{p\}} w \Rightarrow v \sqsupset w$$

Proof: Consider (III.3.9) with Λ the set of p -free formulae. Note that simplification (III.3.10) applies, and fill in the details using (III.3.3). \square

But what of Beth's Theorem? In any case, how might we want to define 'implicit definability'? The most natural definition would seem to be to say that p is 'implicitly definable' in \vdash iff $p \succ p' \in \overline{\vdash \cup \vdash'}$, where \vdash' is as \vdash except for having p replaced by p' . This notion turns out to be strictly weaker than explicit definability, as the following model-theoretic condition will show.

Theorem III.3.15: p is 'implicitly definable' iff for any assignments v and w such that \bar{v} and \bar{w} are models of \vdash :

$$v \equiv_{\Lambda_0 \setminus \{p\}} w \Rightarrow v = w$$

(The natural proof of this fact follows exactly the same pattern as the usual proof of the analogous fact in total-value logic: we omit it.)

Clearly, then, explicit definability implies 'implicit definability', but the converse does not in general hold. For example, if we have a language with just two atomic sentences p and q , and if $v(p) = \top$, $v(q) = *$, $w(p) = *$ and $w(q) = \top$, then the theory of $\{\bar{v}, \bar{w}\}$ has \bar{v} and \bar{w} as its only models;

and so, trivially, p is 'implicitly definable'. However, p is not explicitly definable, since $v(q) \sqsubseteq w(q)$ but $v \not\sqsubseteq w$. There is, in fact, a notion of implicit definability which is equivalent to explicit definability, but it is unpleasantly messy. Whether something neater and more obvious is available, I do not know.

Now we pick up the question raised at the end of Section I.3 – 'joints'. Recall that we showed that for any two formulae which were logically compatible there was a formula which was true – on any given assignment – iff either of the two formulae were true, and false iff either were false. Now we generalize the question, and ask: under what conditions is it the case that, if any pair of formulae ϕ and ψ are compatible in a theory ($\phi \sqsubseteq_{\vdash} \psi$), then there is a sentence χ which is the joint of ϕ and ψ in that theory – i.e. a sentence χ such that for all models M of \vdash

$$M(\chi) = \top \iff M(\phi) = \top \text{ or } M(\psi) = \top$$

$$\perp \iff M(\phi) = \perp \text{ or } M(\psi) = \perp$$

or, equivalently, such the following all hold:

$$\chi \vdash \phi, \psi, * \qquad *, \psi, \phi \vdash \chi$$

$$\phi \vdash \chi, * \qquad *, \chi \vdash \phi$$

$$\psi \vdash \chi, * \qquad *, \chi \vdash \psi$$

If a theory has this property, we say that it has joints (sc. for any two compatible sentences).

First note that if χ is a joint of ϕ and ψ , then $\phi \sqsubseteq_{\vdash} \chi$ and $\psi \sqsubseteq_{\vdash} \chi$, while conversely, if there is any sentence ω such that $\phi \sqsubseteq_{\vdash} \omega$ and $\psi \sqsubseteq_{\vdash} \omega$, then we can define a joint for ϕ and ψ by:

$$((\phi \wedge \omega) \vee (\psi \wedge \omega)) \times ((\phi \vee \omega) \wedge (\psi \vee \omega))$$

So, a joint exists for any two given sentences iff there is some sentence more defined than both (in the theory). This second condition is easier to

for "no P such that" read
"no $P \in K(\text{---})$ such that"

work with: (III.3.4) provides a model-theoretic criterion.

Lemma III.3.16: $\phi \sqsubseteq_{\vdash} \chi$ and $\psi \sqsubseteq_{\vdash} \chi$ for some $\chi \in L(\vdash)$

iff for any models M and N of \vdash

$$M \sqsupset N \Rightarrow M(\phi) \sqsupset N(\psi)$$

Proof: Note first that, for any χ , $\phi \sqsubseteq_{\vdash} \chi$ and $\psi \sqsubseteq_{\vdash} \chi$ iff

$$\phi \vee \psi \vdash \chi, * \quad \text{and} \quad *, \chi \vdash \phi \wedge \psi$$

Hence, if we put $\vdash_1 = \vdash = \vdash_2$, $\Lambda = L(\vdash)$, $\Gamma = \{\phi \vee \psi\}$, $\Delta = \emptyset = \Xi$ and $\{\phi \wedge \psi\} = \top$, then, by the contraposition of case (3) of (III.3.4), such a χ exists iff for any $M, N \in K(\vdash)$

$$M \sqsupset N \Rightarrow M(\phi \vee \psi) \neq \top \text{ or } N(\phi \wedge \psi) \neq \perp$$

But this is clearly equivalent to what we want. \square

Using this lemma and an application of strong completeness we can bring everything together in the following way:

Theorem III.3.17 (Compatibility Theorem): The following are equivalent:

- (i) \vdash has joints
- (ii) $\forall \phi, \psi \in L(\vdash) : \phi \sqsupset_{\vdash} \psi \Rightarrow \exists \chi \in L(\vdash) . \phi \sqsubseteq_{\vdash} \chi \text{ and } \psi \sqsubseteq_{\vdash} \chi$
- (iii) $\forall M, N \in K(\vdash) : M \sqsupset N \Rightarrow \exists P \in K(\vdash) . M \sqsubseteq P \text{ and } N \sqsubseteq P$

Proof: The equivalence of (i) and (ii) has already been dealt with. To show the equivalence of (ii) and (iii), it is easiest to consider their negations. From (III.3.16) we know that (ii) fails to hold iff there exist ϕ and ψ such that

$$(a) \quad \phi \sqsupset_{\vdash} \psi$$

$$(b) \quad M \sqsupset N \text{ but } M(\phi) \not\sqsupset N(\psi) \text{ for some } M \text{ and } N \in K(\vdash)$$

Under such circumstances there can be no P such that $M \sqsubseteq P$ and $N \sqsubseteq P$, since then $P(\phi) \not\sqsupset P(\psi)$, which would contradict the fact that $\phi \sqsupset_{\vdash} \psi$.

Hence (iii) \Rightarrow (ii) is proved.

For the converse, first note that, for any models M and N , there is a $P \in K(\vdash)$ such that $M \sqsubseteq P$ and $N \sqsubseteq P$ iff

$$\mathbb{T}(M), \mathbb{T}(N) \not\vdash *$$

where $\mathbb{T}(M) = \{\phi \mid M(\phi) = \top\}$ and $\mathbb{T}(N) = \{\phi \mid N(\phi) = \top\}$. Hence, if (iii) fails to hold, then there are models M and $N \in K(\vdash)$ such that $M \sqcap N$ and $\mathbb{T}(M), \mathbb{T}(N) \vdash *$. But then, by the finiteness of \vdash and the fact that $\mathbb{T}(M)$ and $\mathbb{T}(N)$ are closed under \wedge , we know that there are sentences χ and ω such that

$$(c) \quad M(\chi) = \top \text{ and } N(\omega) = \top \qquad (d) \quad \chi, \omega \vdash *$$

Hence we can exhibit two sentences satisfying (a) and (b):

$$\chi \ast \top \qquad \neg \omega \ast \perp$$

It follows from (d) that $\chi \ast \top \sqsubseteq_{\vdash} \neg \omega \ast \perp$, and from (c) that $M(\chi \ast \top) = \top$ and $N(\neg \omega \ast \perp) = \perp$. Hence (ii) \Rightarrow (iii). \square

This theorem gives an alternative proof of (I.3.7): for, if \vdash is logic in an atomic language, then the model-theoretic condition is clearly satisfied, and so \vdash has joints. We may note, too, that the canonical theory for a system of convex sets invariably has joints: two convex sets X and Y are compatible iff $X \cap Y \neq \emptyset$, in which case $X \cap Y$ is itself precisely the joint of X and Y .

Not all theories have joints, however: let \vdash be the theory axiomatized by $* \succ p, q$ in the atomic language containing just p and q as atoms; then clearly both \bar{v} and \bar{w} are models of \vdash if $v(p) = \perp, v(q) = *, w(p) = *$ and $w(q) = \perp$, but no P such that $\bar{v} \sqsubseteq P$ and $\bar{w} \sqsubseteq P$ can be a model of \vdash .

So far we have not made use of the fact that in (III.3.1) and its corollaries \vdash_1 and \vdash_2 may be distinct theories. We shall now do so. Consider three theories with their respective languages L_1, L_2 and L , and a

further language Λ given by

$$\left. \begin{array}{l} \vdash_1 \subseteq \\ \vdash_2 \subseteq \end{array} \right\} \vdash \quad \Lambda \quad \left. \begin{array}{l} \subseteq L_1 \subseteq \\ \subseteq L_2 \subseteq \end{array} \right\} L$$

We can now state a general condition for the conditional existence of an interpolant :

Theorem III.3.18: For any sentences ϕ of L_1 and ψ of L_2

$$\phi \vdash \psi \Rightarrow \exists \lambda \in \Lambda. \phi \vdash_1 \lambda \text{ and } \lambda \vdash_2 \psi$$

iff for any models M_1 of \vdash_1 and M_2 of \vdash_2 the following all hold

$$(1) \quad M_1 \sqsubseteq_{\Lambda} M_2 \Rightarrow \exists M \in K(\vdash) . M_1 \sqsubseteq_{L_1} M \sqsubseteq_{L_2} M_2$$

$$(2) \quad M_1 \supseteq_{\Lambda} M_2 \Rightarrow \exists M \in K(\vdash) . M_1 \supseteq_{L_1} M \supseteq_{L_2} M_2$$

$$(3) \quad M_1 \square_{\Lambda} M_2 \Rightarrow \exists M \in K(\vdash) \left\{ \begin{array}{l} \text{either } M_1 \square_{L_1} M \supseteq_{L_2} M_2 \\ \text{or } M_1 \sqsubseteq_{L_1} M \square_{L_2} M_2 \end{array} \right.$$

To prove this we may appeal to (III.3.6) to obtain a condition for the existence of an interpolant for particular ϕ and ψ , and then deploy strong completeness, following the same pattern as in the proof of the Compatibility Theorem. We omit the details.

The 'Interpolation Theorem' for logic in an atomic language is a corollary of this result. For, if \vdash is a logic in such a language and $\phi \vdash \psi$, then we can take \vdash_1 and \vdash_2 to be logic in the languages which contain, respectively, nothing but the vocabulary of ϕ , and nothing but the vocabulary of ψ , and put $\Lambda = L_1 \cap L_2$. The theorem then implies that $\phi \vdash_1 \lambda$ and $\lambda \vdash_2 \psi$ for some $\lambda \in \Lambda$, since models correspond to atomic assignments and so conditions (1), (2) and (3) are easily seen to be satisfied. (Admittedly this is something of a detour — we could have appealed more directly to (III.3.6).)

Of course, for logic it does not matter what language we have in mind: logic in L is always a conservative extension of logic in any

sublanguage of L — because any model for the sublanguage can be extended to a model for L . Hence we lose nothing by stating the interpolation theorem in terms of logical consequence \vdash in a single (atomic) language:

Theorem III.3.19: If $\phi \vdash \psi$, then $\phi \vdash \lambda$ and $\lambda \vdash \psi$ for some λ which contains no vocabulary which does not occur in both ϕ and ψ .

We should now point out something which does not hold. In classical logic we have in general that, if $\Gamma, \Xi \vdash \Delta, \top$, then there is an interpolant λ such that $\Gamma \vdash \lambda, \Delta$ and $\Xi, \lambda \vdash \top$ with λ containing only vocabulary common to both $\Gamma \cup \Delta$ and $\Xi \cup \top$. However, in our case we cannot be assured the existence of an interpolant of any sort. For example, $\top, * \vdash *, \perp$; but there can be no λ such that $\top \vdash \lambda, *$ and $*, \lambda \vdash \perp$. This remark holds, in fact, for any consistent theory.

We have interpolation for \sqsubseteq as well. There is a neat general theorem, but let us turn straight to logical degree-of-definedness (in atomic languages).

Theorem III.3.20: If $\phi \sqsubseteq \psi$, then $\phi \sqsubseteq \lambda$ and $\lambda \sqsubseteq \psi$ for some λ which contains no vocabulary which does not occur in both ϕ and ψ .

Proof: Let \vdash_1 be logic in the language of ϕ and \vdash_2 logic in that of ψ , and put $\Lambda = L(\vdash_1) \cap L(\vdash_2)$. Consider the second combination lemma, (III.3.8): if any of the three conditions failed, we could deduce a contradiction from the relevant case of (III.3.4). \square

Finally, we consider a couple of unconditional 'expressive closure conditions' that a theory might possess.

Let us begin by asking when a theory has, for a given ϕ , a sentence ψ which is true (in the theory) iff ϕ is not true (in the theory).

Lemma III.3.21: $\phi, \psi \vdash \ast$ and $\vdash \phi, \psi$ for some ψ

(equivalently, $\exists \psi \in L(\vdash). \forall M \in K(\vdash). M(\psi) = \top \Leftrightarrow M(\phi) \neq \top$)

iff for all M and $N \in K(\vdash)$:

$$M \sqsubseteq N \Rightarrow M(\phi) \sqsubseteq N(\phi)$$

This can be proved using part (1) of (III.3.4). And we can now deduce when this condition obtains for any ϕ in $L(\vdash)$.

Theorem III.3.22: For any ϕ there is a ψ such that $\phi, \psi \vdash \ast$ and $\vdash \phi, \psi$ iff for all M and $N \in K(\vdash)$:

$$M \sqsubseteq N \Rightarrow M = N$$

There is another, stronger, condition to consider: that, given ϕ , there is a ψ which is true iff ϕ is not true and is false otherwise — i.e. if ϕ is true. This would mean that there was a sentence playing the role of the 'exclusion negation' of ϕ .

Lemma III.3.23: $\vdash \phi, \psi$ and $\ast, \psi \vdash \neg \phi$ for some ψ

(equivalently, $\exists \psi \in L(\vdash). \forall M \in K(\vdash). \left\{ \begin{array}{l} M(\psi) = \top \Leftrightarrow M(\phi) \neq \top \\ M(\psi) = \perp \Leftrightarrow M(\phi) = \top \end{array} \right\}$),

iff for all M and $N \in K(\vdash)$:

$$M \sqsupseteq N \Rightarrow M(\phi) = N(\phi)$$

This follows from part (3) of (III.3.4). A language in which every sentence satisfied this condition would be closed under 'exclusion negation', the negations of 'exclusion negations', 'exclusion negations' of negations, etc. This is a very strong property, which turns out to be equivalent to the conjunction of the property dealt with in (III.3.22) and the property of having joints.

Theorem III.3.24: The following are equivalent:

- (i) for any ϕ there is a ψ such that $\vdash \phi, \psi$ and $\ast, \psi \vdash \neg \phi$
- (ii) for all M and $N \in K(\vdash)$: $M \sqsupseteq N \Rightarrow M = N$
- (iii) \vdash has joints and satisfies the condition of (III.3.22).

Proof: The equivalence of (i) and (ii) follows immediately from the last lemma. For the equivalence of (ii) and (iii), look back to the Compatibility Theorem, and note that, of the following three conditions on models in $K(\vdash)$, the conjunction of the first two is equivalent to the third:

$$(1) \quad M \sqsupseteq N \Rightarrow \exists P \in K(\vdash). M \sqsubseteq P \text{ and } N \sqsubseteq P$$

$$(2) \quad M \sqsubseteq N \Rightarrow M = N$$

$$(3) \quad M \sqsupseteq N \Rightarrow M = N \quad \square$$

There is an interesting example of a kind of theory for which this condition holds, viz. systems of finitely specifiable convex sets. We saw this at the end of I.4. In general, of course, the condition fails for a canonical theory of convex sets; hence, since such a theory has joints, in general the condition of (III.3.22) fails also.

There are results analogous to (III.3.22) and (III.3.24) for the kind of theory we consider in Chapter IV. These can be proved from 'interpolant exclusion' in a similar way, but we shall not be presenting them.

III.4 THEORIES AND SEQUENTS

We now turn briefly to sequents and consider some ways of axiomatizing theories and some sequent-persistence results.

Let us call sequents of the form $\succ \Gamma$ (i.e. $\emptyset \succ \Gamma$) posits, and ones of the form $* \succ \Gamma$ contraposits. Let $Pos(\vdash)$ be the set of posits of the theory \vdash and $Con(\vdash)$ the set of contraposits of \vdash .

Theorem III.4.1:

- (1) M is a model of $Pos(\vdash)$ iff $M \sqsupseteq N$ for some model N of \vdash
- (2) M is a model of $Con(\vdash)$ iff $M \sqsubseteq N$ for some model N of \vdash

Proof: Consider (1): 'if' is trivial. For 'only if' assume that M is a

model of $Pos(\vdash)$. By strong completeness, it will be sufficient to show that $\not\vdash \Delta$, where $\Delta = \{\phi \mid M(\phi) \neq T\}$. But if $\vdash \Delta$, then $\supset \Delta_0 \in \vdash$ for some (finite) subset Δ_0 of Δ . However, this contradicts the assumption that M is a model of $Pos(\vdash)$. (2) is proved similarly. \square

Hence:

Theorem III.4.2: \vdash is axiomatizable by some set Σ of sequents such that, respectively,

- (1) all elements of Σ are posits
- (2) all elements of Σ are contraposits
- (3) all elements of Σ are either posits or contraposits

iff, respectively,

- (1) $M \in K(\vdash)$ and $M \sqsubseteq N \Rightarrow N \in K(\vdash)$
- (2) $M \in K(\vdash)$ and $M \sqsupseteq N \Rightarrow N \in K(\vdash)$
- (3) $M_1, M_2 \in K(\vdash)$ and $M_1 \sqsubseteq N \sqsubseteq M_2 \Rightarrow N \in K(\vdash)$

Proof: 'only if' is immediate. For 'if', first consider (1). It will be sufficient to show that $\overline{Pos(\vdash)} = \vdash$. So, say N is a model of $Pos(\vdash)$: by the last theorem, there is a model M of \vdash such that $M \sqsubseteq N$. But then N is a model of \vdash also. Hence $\overline{Pos(\vdash)} = \vdash$ by completeness. (2) is proved similarly; while for (3) we consider $Pos(\vdash) \cup Con(\vdash)$ and invoke both parts of the last theorem. \square

It is interesting to see that no theory may be axiomatized both by posits only and by contraposits only, unless it is logic — and so in fact has no non-trivial 'proper axioms' at all. This is because, by parts (1) and (2) of the last theorem, all models of the language would have to be models of the theory.

But, now, what of individual sequents in the context of a given theory? Under what conditions is a sequent 'equivalent' to a sequent of a

particular kind? Let us say that a sequent α is \vdash -equivalent to a sequent β iff

$$\alpha \in \overline{\vdash \cup \{\beta\}} \quad \text{and} \quad \beta \in \overline{\vdash \cup \{\alpha\}}$$

By completeness, an equivalent definition would be that for all $M \in K(\vdash)$

M is consistent with α iff M is consistent with β

Note also that $\Gamma \succ \Delta$ is \vdash -equivalent to $\Xi \succ \Upsilon$ iff all the following hold:

$$\begin{array}{cccc} \Xi \vdash \Gamma, \Upsilon & \Xi, \neg \Gamma \vdash \Upsilon & \Xi, \Delta \vdash \Upsilon & \Xi \vdash \neg \Delta, \Upsilon \\ \Gamma \vdash \Xi, \Delta & \Gamma, \neg \Xi \vdash \Delta & \Gamma, \Upsilon \vdash \Delta & \Gamma \vdash \neg \Upsilon, \Delta \end{array}$$

Using (III.4.1), we may show

Theorem III.4.3:

(1) α is \vdash -equivalent to a posit iff for all M and $N \in K(\vdash)$

such that $M \sqsubseteq N$:

M is consistent with $\alpha \Rightarrow N$ is consistent with α

(2) α is \vdash -equivalent to a contraposit iff for all M and $N \in K(\vdash)$ such that $M \sqsubseteq N$:

M is consistent with $\alpha \Rightarrow N$ is consistent with α

Proof: Consider (1): 'only if' is trivial. For 'if', let $\Sigma = \overline{\text{Pos}(\vdash \cup \{\alpha\})}$, and note that to show that α is equivalent to a posit, it is sufficient to show that $\alpha \in \overline{\vdash \cup \Sigma}$. For, in that case, $\alpha \in \overline{\vdash \cup \Sigma_0}$, where Σ_0 is some finite subset $\{\succ \Delta_i \mid 1 \leq i \leq n\}$ of Σ ; hence if $\phi = \bigwedge_i (M \Delta_i)$, $\alpha \in \overline{\vdash \cup \{\succ \phi\}}$. But $\succ \phi \in \Sigma$, and so α is \vdash -equivalent to $\succ \phi$.

So, assume that the model theoretic condition holds. Take a model M of $\vdash \cup \Sigma$. M is a model of Σ , and so, by part (1) of (III.4.1), there is a model N of $\vdash \cup \{\alpha\}$ such that $N \sqsubseteq M$. But both M and N are models of \vdash , and N is consistent with α ; so M is consistent with α . Hence, by completeness, $\alpha \in \overline{\vdash \cup \Sigma}$ as required. \square

The next question is: when is it the case that any sequent is \vdash -equivalent to a posit – or to a contraposit? Given the last theorem, it is not difficult to check:

Theorem III.4.4: The following are equivalent:

- (i) any α is a \vdash -equivalent to a posit
- (ii) any α is a \vdash -equivalent to a contraposit
- (iii) $M \sqsubseteq N \Rightarrow M=N$, for any models M and N of \vdash

Theorem (III.4.2) told us under what conditions a theory as a whole was determined by its truths, its non-truths, or by a combination of both; while (III.4.4) tells us when, in a theory, any sequent can be seen as represented by the truth (in the theory) of a single sentence – and this is precisely when they can be represented by non-truths. The model-theoretic condition shows, furthermore, that this is the case if and only if the conditions considered in (III.3.22) hold. In other words, we have a somewhat weaker property than the property that each sentence has another sentence playing the role of its 'exclusion negation' in the theory – although this is equivalent under the assumption that the theory has joints. Recall (III.3.24).

Finally, we might record the following variation on the theme of (III.4.1):

Theorem III.4.5: M is a model of $\{ * \succ \Gamma \mid \succ \Gamma \in \vdash \}$ iff $M \sqsubset N$ for some model of N of \vdash .

Proof: similar to (III.4.1)

(III.4.1) and (III.4.5) can be used to show that the messy notion of implicit definability I mentioned in the last section is equivalent to explicit definability. The proof is subtle, but, since the result itself is obscure, I shall resist the temptation to go into it.

in displayed mode Roman capitals should
be italic.

III.5 LOGICALLY CONVEX SETS – THE REPRESENTATION THEOREM

In preparation for the promised representation theorem, showing that any partial-valued theory can be 'embedded' (confusing only equivalent sentences) into a system of convex sets, we give a generalized definition of a 'boolean language' and 'boolean consequence' – along the lines of the one we gave for partial-valued languages. A boolean language shall be set closed under operations \neg , \wedge and \vee , and containing elements \top and \perp . A boolean consequence relation shall be a relation between finite subsets of a boolean language satisfying (R), (M) and (T), as specified in the introduction to this chapter, and also the following principles:

$$\begin{array}{ccc} \vdash \top & \top \vdash & \frac{X, \phi, \psi \vdash Y}{X, \phi \wedge \psi \vdash Y} \\ \\ \frac{X, \phi \vdash Y}{X \vdash \neg \phi, Y} & & \frac{X \vdash \phi, \psi, Y}{X \vdash \phi \vee \psi, Y} \end{array}$$

And a boolean valuation v for the boolean language B shall be a function from B into $\{\top, \perp\}$ satisfying

$$\begin{aligned} v(\top) = \top &\iff v(\perp) = \perp \\ v(\neg a) = \top &\iff v(a) = \perp \\ v(a \wedge b) = \top &\iff v(a) = \top \text{ and } v(b) = \top \\ v(a \vee b) = \top &\iff v(a) = \top \text{ or } v(b) = \top \end{aligned}$$

$X \not\vdash_v Y$ shall mean that it is not the case that $v(a) = \top$ for all $a \in X$ and $v(b) = \perp$ for all $b \in Y$. Then, if V is a set of boolean valuations, let \vdash_V be the set $\{X \succ Y \mid X \not\vdash_v Y \text{ for all } v \in V\}$, and it turns out that \vdash_V is a boolean consequence relation. (Note: $X \succ Y$ is used for pairs of finite sets, as before.) While, if \vdash is a boolean consequence relation and $V(\vdash) = \{v \mid X \not\vdash_v Y \text{ for all } X \succ Y \in \vdash\}$, then we have completeness: $\vdash = \vdash_{V(\vdash)}$. For details see Scott (1973b)

As we did for partial-valued logic, we shall extend the turnstile notation and write ' $X \models Y$ ' for arbitrary X and Y to mean that \models holds between some finite subsets of X and Y . Then, to define convex sets and their operations, we may put the following:

$$\underline{X \text{ is convex}} \iff X = \{a \mid X \models a \text{ and } a \models X\}$$

and

$$\begin{aligned} \neg X &= \{a \mid X, a \models \text{ and } \models a, X\} \\ X \wedge Y &= \{a \mid X, Y \models a \text{ and } a \models X \text{ and } a \models Y\} \\ X \vee Y &= \{a \mid X \models a \text{ and } Y \models a \text{ and } a \models X, Y\} \\ X \times Y &= \{a \mid X, Y \models a \text{ and } a \models X, Y\} \end{aligned}$$

We could easily define slash as well:

$$X/Y = \{a \mid X, Y \models a \text{ and } X \ a \models Y\}$$

Also

$$\begin{aligned} X \text{ is } \underline{\text{finitely specifiable}} \iff X &= \{a \mid b \models a \text{ and } a \models c\} \\ &\text{for some 'sentences' } b \text{ and } c \end{aligned}$$

These definitions are neater than the ones given in I.4, but they come to the same thing.

If \models is a boolean consequence relation, let $C(\models)$ be the system of convex sets defined as above on the language $B(\models)$ of \models . The canonical theory for $C(\models)$ can now be specified by saying that $X_1, \dots, X_m \Vdash Y_1, \dots, Y_n$ if and only if

$$\exists j \forall b \in Y_j \cup X_i \models b \text{ and } \exists i \forall a \in X_i \ a \models \cup_j Y_j$$

Now corresponding to a boolean valuation v on $B(\models)$ we have a partial-valuation v^0 on $C(\models)$ given by:

$$\begin{aligned} v^0(X) = \top &\iff v(a) = \top \text{ for all } a \in X \\ &= \perp \iff v(a) = \perp \text{ for all } a \in X \end{aligned}$$

v^0 is a model for $C(\models)$, and if $v \in V(\models)$, then v^0 is a model of the canonical theory \Vdash . Moreover any model M of the canonical theory is v^0 for some $v \in V(\models)$

To set up the representation theorem, we can employ the notion of a 'translation': let L_1 and L_2 be partial-valued languages, then $I: L_1 \rightarrow L_2$ is a translation iff

$$I(\top) = \top \quad I(\perp) = \perp$$

$$I(\neg \phi) = \neg I(\phi)$$

$$I(\phi \wedge \psi) = I(\phi) \wedge I(\psi)$$

$$I(\phi \vee \psi) = I(\phi) \vee I(\psi)$$

$$I(\phi \ast \psi) = I(\phi) \ast I(\psi)$$

If \vdash_1 and \vdash_2 are theories in the languages L_1 and L_2 respectively, we call I faithful (with respect to \vdash_1 and \vdash_2) iff for all ϕ and ψ

$$\phi \vdash_1 \psi \Rightarrow I(\phi) \vdash_2 I(\psi)$$

Note that checking that a translation is faithful is the analogue of checking that an algebraic homomorphism which is defined in terms of elements of equivalence classes is 'well-defined'. On the other hand we call I conservative (with respect to \vdash_1 and \vdash_2) iff for all ϕ and ψ

$$I(\phi) \vdash_2 I(\psi) \Rightarrow \phi \vdash_1 \psi$$

Checking that a translation is conservative is the analogue of checking that an algebraic homomorphism is one-one.

We can now state the representation theorem.

Theorem III.5.1: Given any partial-valued theory \vdash , there is a boolean consequence relation \models and a translation $I: L(\vdash) \rightarrow C(\models)$ such that:

- (i) I is faithful and conservative with respect to \vdash and the canonical theory for $C(\models)$.
- (ii) I maps sentences of $L(\vdash)$ to finitely specifiable elements of $C(\models)$.
- (iii) There is a one-one correspondence f from $K(\vdash)$ onto a set V of boolean valuations such that $\models = \models_V$ and $(f(M))^0(I(\phi)) = M(\phi)$ for all $\phi \in L(\vdash)$.

Proof (sketch): For the boolean language we take the power set $\mathcal{P}(K(\vdash))$ along with the obvious definitions:

$$\begin{aligned} \top &= K(\vdash) & \perp &= \emptyset & a \wedge b &= a \cap b \\ \neg a &= K(\vdash) \setminus a & & & a \vee b &= a \cup b \end{aligned}$$

The consequence relation is just (the multiary generalization of) set inclusion:

$$a_1, \dots, a_n \vDash b_1, \dots, b_m \iff \bigcap_i a_i \subseteq \bigcup_j b_j$$

The translation is defined by:

$$I(\phi) = \{a \mid \{M \mid M(\phi) = \top\} \subseteq a \subseteq \{M \mid M(\phi) \neq \perp\}\}$$

We define the function f by:

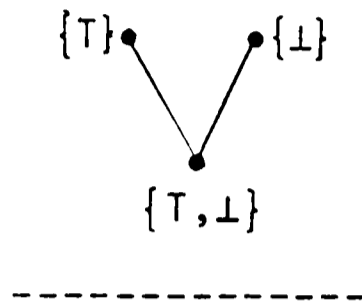
$$\begin{aligned} f(M)(a) &= \top \text{ if } M \in a \\ &= \perp \text{ otherwise} \end{aligned}$$

$f(M)$ is clearly a boolean valuation, and we simply put $V = \{f(M) \mid M \in K(\vdash)\}$.

Checking that I is indeed a translation, and that everything holds as stated, is now straight-forward enough. Note that the set V is the whole set $V(\vDash)$ iff $K(\vdash)$ is finite, since elements of $V(\vDash)$ correspond to ultrafilters on $\mathcal{P}(K(\vdash))$, and the valuations $f(M)$ to principal ultrafilters.

If we thought of systems $C(\vDash)$ and faithful translations into their canonical theories as semantical interpretations, rather than as interpretations of one theory in another, then we would have a non-standard semantics in which any consistent theory would be provided with a single 'model' determining that theory exactly. This is because our theorem guarantees the existence of a conservative interpretation. But we can also deduce the (weaker) fact that, for any theory \vdash , $\Gamma \vdash \Delta$ iff $I(\Gamma) \vDash I(\Delta)$ for all faithful translations I into canonical theories \vDash . This is 'completeness' with respect to such interpretations.

Of course, our standard semantics can be regarded as taking models of a theory to be faithful translations into one particular system of convex sets, viz. that based on the two-element boolean algebra $\{\top, \perp\}$, whose convex sets are



In contrast to the blank at the end of I.3, we can do an easy sum for the size of $C(\models)$, whenever $B(\models)$ is finite modulo \models -equivalence. If we factor out $B(\models)$, we have a boolean algebra with 2^n elements; $C(\models)$ will then have 3^n elements. To see this, let A be any set with n members. $C(\models)$ will be isomorphic to the system of convex sets on $\mathcal{P}(A)$, viz. the set

$$\{\{W \mid U \subseteq W \subseteq V\} \mid U \subseteq V \subseteq A\}$$

But there are $\binom{n}{k}$ subsets of A of size k , and 2^k subsets of each of these subsets. Thus the number of convex sets on $\mathcal{P}(A)$ — hence also on $B(\models)$ — is

$$\sum_{k=0}^n \binom{n}{k} 2^k = (2+1)^n = 3^n$$

CHAPTER IV

PARTIAL-VALUED LOGIC (2) – TERMS AND QUANTIFIERS

IV.1 THEORIES AND MODELS – PRELIMINARIES

In Section II.2 we described a class of languages, specified what a model M for a language L was to be, and set up some further definitions on this basis. All this we adopt. And we now present logical laws, as we did in the last chapter for propositional languages, by defining what a theory is.

However, to begin with, some more model-theoretic definitions. We must say what it is for a model M for L to be consistent with a pair $\langle \Gamma, \Delta \rangle$ of sets of formulae of L . As in Chapter III, this is to mean that M does not constitute a counter example to (the disjunction of) Δ 's being a consequence of Γ . The elements of Γ and Δ do not have to be sentences: they may contain parameters (free variables), whose role is to be schematic for singular terms, and so the definition naturally falls into two stages. Firstly, given a particular assignment s : (M, s) is consistent with $\langle \Gamma, \Delta \rangle$ – for which we write ' $\Gamma \Vdash_{M,s} \Delta$ ' – iff

neither $M_s(\phi) = \top$ for all $\phi \in \Gamma$ and $M_s(\psi) \neq \top$ for all $\psi \in \Delta$
nor $M_s(\phi) \neq \perp$ for all $\phi \in \Gamma$ and $M_s(\psi) = \perp$ for all $\psi \in \Delta$

And then: M is consistent with $\langle \Gamma, \Delta \rangle$ – ' $\Gamma \Vdash_M \Delta$ ' – iff

$$\Gamma \Vdash_{M,s} \Delta \text{ for all } s: \text{Var} \rightarrow D_M \cup \{\circ\}$$

If M , or (M, s) , is not consistent with $\langle \Gamma, \Delta \rangle$, we say it is inconsistent with it.

If K is any set of models for L , then the relation \Vdash_K is defined by

$$\{\langle \Gamma, \Delta \rangle \mid \Gamma \Vdash_M \Delta \text{ for all } M \in K\}.$$

But theories will be relations between sequents – pairs $\Gamma \succ \Delta$ of finite

sets – and so we shall be interested in the subrelation \vdash_K given by

$$\{\Gamma \succ \Delta \mid \Gamma \Vdash_M \Delta \text{ for all } M \in K\}$$

Our definition of 'theory' will be no good unless such relations are theories, and so, proleptically, we shall label \vdash_K the theory of K . If $K = \{M\}$, then we write \vdash_M for \vdash_K and call it the theory of M . Warning: in contrast to the propositional case, it is possible that $\Gamma \Vdash_M \Delta$, but not $\Gamma_0 \Vdash_M \Delta_0$ for any (finite) subsets of Γ and Δ ; also, we may have models M and N such that $\vdash_M = \vdash_N$ but $\Vdash_M \neq \Vdash_N$.

On the other hand, if Σ is any set of sequents of a language L , and M is a model for L , then M is a model of Σ iff M is consistent with every member of Σ , and $K(\Sigma)$ is the class of such models.

Some further notation: we shall write $\phi(x_1, \dots, x_n)$, or $\phi(\vec{x})$ for short, and $\Gamma(x_1, \dots, x_n)$, or $\Gamma(\vec{x})$, to denote a formula, or a set of formulae, all of whose parameters, if any, occur among x_1, \dots, x_n . Then, as we are justified in doing by (II.2.1), we can ignore assignments to other variables and write $M_{\vec{a}}(\phi(\vec{x}))$ for the value of $\phi(\vec{x})$ in M with respect to the assignment of a_i to x_i , and also $\Gamma(\vec{x}) \not\vdash_{M, \vec{a}} \Delta(\vec{x})$ to mean that M is inconsistent with $\langle \Gamma(\vec{x}), \Delta(\vec{x}) \rangle$ ($= \Gamma(\vec{x}) \succ \Delta(\vec{x})$, if $\Gamma(\vec{x})$ and $\Delta(\vec{x})$ are finite) under this assignment. Finally, to denote the formula, or set of formulae, got by substituting terms t_i for x_i we shall write $\phi(t_1, \dots, t_n)$ or $\phi(\vec{t})$, and $\Gamma(t_1, \dots, t_n)$ or $\Gamma(\vec{t})$.

Theories: a theory in L is a set \vdash of sequents of L which satisfies

- (1) (i) the rules (R), (M) and (T) of Chapter III
- (ii) the propositional rules of Chapter III
- and new closure conditions
- (2) (i) the rule (S)
- (ii) rules for the quantifiers and for identity.

Hence theories in the new sense will in fact be theories in the propositional sense.

(S) is a general rule of substitution:

$$(S) \quad \frac{\Gamma \vdash \Delta}{\Gamma(t/x) \vdash (t/x)}$$

This applies provided that the term t is substitutable for x in all the formulae in Γ and Δ . In the presence of this rule the remaining principles may be written with parameters throughout, instead of with schematic letters for arbitrary terms.

Before we turn to these, let us introduce the abbreviation ' $\Gamma \stackrel{\phi}{\vdash} \Delta$ ' for ' $\Gamma, \phi \vdash \Delta$ '. The force of ' $\Gamma \stackrel{\phi}{\vdash} \Delta$ ' is that under the assumption that ϕ is true (in the theory) Δ follows from Γ in the usual way. Though we have introduced this as an abbreviation, it is an important scheme of inference, which could be taken as the primitive one in a natural deduction system.

For the quantifiers we adopt the two-way principles

$$(V) \quad \frac{\Gamma \stackrel{x=x}{\vdash} \phi, \Delta}{\Gamma \vdash \forall x \phi, \Delta} \quad (\exists) \quad \frac{\Gamma, \phi \stackrel{x=x}{\vdash} \Delta}{\Gamma, \exists x \phi \vdash \Delta}$$

which go up as well as down and are subject to the proviso that x does not occur free in any formula in Γ or Δ . The proviso is only of importance for the downward rules, but, given (S), its presence does not hamper the force of the upward rules. Notice how $x=x$ is playing the role of an existence predicate — cf. the discussion in II.2. The upward rules are equivalent to the following unconditional principles, which we give special labels:

$$(V') \quad \forall x \phi \stackrel{x=x}{\vdash} \phi \quad (\exists') \quad \phi \stackrel{x=x}{\vdash} \exists x \phi$$

These 'axioms' will often be easier to work with.

For the 'substitutivity of identicals' we use the following

$$(SI) \quad \phi(x/u, y/v) \stackrel{x=y}{\vdash} \phi(y/u, x/v)$$

This means that if $x=y$ is true, then free occurrences of x and y may be shuffled around in a formula in any way you like, and these variants will all be equivalent. The axiom splits into two halves, either of which would in fact be sufficient on its own:

$$(1) \quad \phi(x/u, y/v), x=y \vdash \phi(y/u, x/v), *$$

and

$$(2) \quad *, \phi(x/u, y/v) \vdash x \neq y, \phi(y/u, x/v)$$

(SI) is not, however, sufficient to capture everything about substitutivity. If we were considering subtler theories of identity, we should have to adopt a general rule to cater for the range of conditions that monotonicity determines; however, for discrete domains (SI) will do, if we also include the principle

$$(SD) \quad \phi(x/z) \vdash x=x, \phi(y/z), *$$

which ensures that denoting is an all-or-nothing matter. The point of (SD) is that if a formula ϕ is true (or false) but ' x does not exist', then any term may be substituted for x *salva ueritate (aut falsitate)*.

Of course, $x=x$ cannot actually be false, though it may be undefined, and so we include an axiom

$$(Ext) \quad * \vdash x=x$$

Hence, notice that it is a matter of indifference whether or not we include ' $*$ ' on the right in (SD).

Our identity relation is to to be determinate, and we may formulate this with

$$(Det) \quad x=x, y=y \vdash x=y, x \neq y$$

Finally, for \ominus it is sufficient to have this axiom:

$$(\ominus) \quad x=\ominus \vdash *$$

From (\ominus) together with (SD) we may derive the characteristic principle

for " $\phi(\theta/z)$ " and " $\phi(\theta/y)$ "

$$(3) \quad \phi(\Theta/z) \vdash \phi(x/y), *$$

Again using (SD), we have $x \neq y \vdash x = x$, $y \neq y$, and so, by (Ext) and the propositional rule twist,

$$(4) \quad x \neq y \vdash x = x$$

and similarly

$$(5) \quad x \neq y \vdash y = y$$

while, from (SI),

$$(6) \quad x = y \vdash x = x$$

and

$$(7) \quad x = y \vdash y = y$$

According to the intended interpretation, (4) - (7) record the fact that any term in a true or false identity statement will denote. Also, from (7) and (\exists) we have

$$(8) \quad \exists x. x = y \vdash y = y$$

while, as an instance of (\exists') , $x = y \vdash \frac{x = x}{\exists x. x = y}$; and so, by (S),

$$(9) \quad \frac{y = y}{\exists x. x = y}$$

and, with precisely the same force,

$$(10) \quad y = y \vdash \exists x. x = y, *$$

We cannot, of course, omit $*$ on the right here, since $\exists x. x = y$ is false in empty domains.

Symmetry of identity follows immediately from (SI): we have

$$x = y \vdash x \neq y, y = x, \text{ and we can cut - i.e. use (T) - with } x = y, x \neq y \vdash *$$

to obtain

$$(11) \quad x = y \vdash y = x, *$$

insert "of" between "premise(s)" and "the"

However, to obtain symmetry of distinctness as well, we need to invoke more: from (4) and (5), switching x and y by (S), and cutting twice with (Det), we have $y \neq x \vdash x = y, x \neq y$, which, twisting, gives $x = y, x \neq y \vdash y = x$, and, cutting with what (SI) originally gave, yields

$$(12) \quad x = y \vdash y = x .$$

Similarly, transitivity of identity is immediate from (SI):

$$(13) \quad x = y, y = x \vdash x = z, *$$

However, in this case we should not try to get rid of the occurrence of $*$, since, if $x = z$ is false, then we are guaranteed that one of the identity formulae on the left is false only if ' y exists'. The general principle which we want — and which is derivable using (Det) — is most perspicuously formulated

$$(14) \quad x = y, y = z \vdash \frac{y = y}{x = z}$$

It is easy to check that any model is consistent with the axioms we have specified, and that if a model is consistent with the 'premise(s)' the conditional rules, then it is consistent also with the 'conclusion'. Hence, as we promised,

Theorem IV.1.1: If K is a class of models for L , then \vdash_K is a theory in L .

Now, relative to a particular language L , we define $\bar{\Sigma}$, the theory axiomatized by Σ , to be the smallest theory containing Σ :

$$\bar{\Sigma} = \bigcap \{ \vdash \mid \vdash \text{ is a theory in } L \text{ and } \Sigma \subseteq \vdash \}$$

Then

Lemma IV.1.2: Σ is a theory iff $\Sigma = \bar{\Sigma}$

Lemma IV.1.3: $\Gamma \succ \Delta \in \bar{\Sigma}$ iff there is a finite subset Σ_0 of Σ such that $\Gamma \succ \Delta \in \bar{\Sigma}_0$.

This is immediate: by the general theory governing such definitions, a sequent will be in $\bar{\Sigma}$ iff there is a (finite) 'proof' of it — taking our closure conditions as sequent rules — from a (finite) subset of Σ . And let us say that Σ is consistent iff $\emptyset \not\vdash \emptyset \in \bar{\Sigma}$. Inconsistent means not consistent. Hence Σ is inconsistent iff $\bar{\Sigma}$ is the set of all sequents.

The 'soundness' of our logical laws — guaranteed by (IV.1.1) — may be stated in various ways:

Corollary IV.1.4: $K(\Sigma) = K(\bar{\Sigma})$

Corollary IV.1.5: $\bar{\Sigma} \subseteq \frac{}{K(\Sigma)}$

Corollary IV.1.6: If $K(\Sigma) \neq \emptyset$, then Σ is consistent.

Logic in L is $\bar{\emptyset}$ — i.e. the smallest theory in L . Clearly, then, any model for L is a model of logic. Such models exist (in great abundance), and so logic is consistent.

As we did in Chapter III, we shall prove semantical completeness by showing that a theory is determined by its models:

Theorem IV.1.7 (Completeness): If \vdash is a theory, then $\vdash = \frac{}{K(\vdash)}$.

Then, from the previously stated facts,

Corollary IV.1.8: Σ is a theory iff $\Sigma = \frac{}{K}$ for some K

Corollary IV.1.9: $\bar{\Sigma} = \frac{}{K(\Sigma)}$

Corollary IV.1.10: $K(\Sigma) \neq \emptyset$ iff Σ is consistent.

But (IV.1.7) will be deduced from a stronger result, which treats not merely of sequents but of arbitrary pairs of sets of formulae. As before, we extend the turn-stile notation and write ' $\Gamma \vdash \Delta$ ' to mean that there are finite subsets Γ_0 and Δ_0 of Γ and Δ such that $\Gamma_0 \vdash \Delta_0 \in \vdash$. Note that the new laws, like the old, hold as stated for this extended use of ' \vdash '.

Theorem IV.1.11 (Strong Completeness): $\Gamma \vdash \Delta \Leftrightarrow \Gamma \Vdash_{K(\vdash)} \Delta$

Corollary IV.1.12 (Compactness): For any Σ : $\Gamma \Vdash_{K(\Sigma)} \Delta \Leftrightarrow \Gamma \Vdash_{K(\Sigma)} \Delta$

Furthermore, if L is countable, we shall want the obvious Löwenheim-Skolem theorem

Theorem IV.1.13: $\Gamma \Vdash_{K(\Sigma)} \Delta \Leftrightarrow \Gamma \Vdash_{K_c(\Sigma)} \Delta$

where $K_c(\Sigma)$ is the set of all models of Σ with a countable domain.

Assuming L is countable, we can establish all these results by showing

Lemma IV.1.14: If $\Gamma \not\vdash \Delta$ then $\Gamma \not\Vdash_M \Delta$ for some countable model M of \vdash .

This is what we shall do. We shall, however, indicate how to extend the construction used to establish (IV.1.14), so as to yield strong completeness in general. In any case, the actual construction we make will be more complicated than need be for (IV.1.14) alone, and it is postponed until the next section.

We conclude this section with some definitions and the required model-existence lemma.

Given a pair $\langle X, Y \rangle$ of sentences of a language L , let us say that $\langle X, Y \rangle$ is L -exhaustive iff $X \cup Y = \text{Snt}(L)$; that $\langle X, Y \rangle$ is \vdash -rejected iff $X \not\vdash Y$; and that $\langle X, Y \rangle$ is \ast -right (respectively \ast -left) iff $\ast \in Y$ (respectively $\ast \in X$).

Given a language L , let $L(C)$ be the language obtained from L by including as new constants the elements of C : these are assumed to be distinct from any existing vocabulary of L . And given a theory \vdash whose language is L , let \vdash^C be the theory in $L(C)$ axiomatized by \vdash .

Finally, if $\langle X, Y \rangle$ is a pair of sets of sentences of $L(C)$ and $D \subseteq C$, let us say that $\langle X, Y \rangle$ is D -witnessed iff

- (i) for all $d \in D$, $d = d \in X$ and $d \neq d \in Y$
- (ii) if $\exists x \phi(x) \in X$, then $\phi(d) \in X$ for some $d \in D$
- (iii) if $\forall x \phi(x) \in Y$, then $\phi(d) \in Y$ for some $d \in D$

Now, given a theory \vdash in L , a set of new constants C , and a subset D of C , we prove the following lemma concerning \vdash^C .

Lemma IV.1.15: Say $\langle X, Y \rangle$ is $L(C)$ -exhaustive, \vdash^C -rejected and D -witnessed, then

(A) the relation \sim defined on D by

$$d \sim e \iff d=e \in X \text{ and } d \neq e \in Y$$

is an equivalence relation.

(B)(i) if $\langle X, Y \rangle$ is $*$ -right, there is a model M of \vdash^C , whose domain is D/\sim , such that for all $\sigma \in Snt(L(C))$:

$$\begin{aligned} M(\sigma) = \top &\iff \sigma \in X \\ &= \perp \iff \neg \sigma \in X \end{aligned}$$

(ii) if $\langle X, Y \rangle$ is $*$ -left, there is a model M of \vdash^C , whose domain is D/\sim , such that for all $\sigma \in Snt(L(C))$:

$$\begin{aligned} M(\sigma) = \top &\iff \neg \sigma \in Y \\ &= \perp \iff \sigma \in Y \end{aligned}$$

Proof: (A) is easy. The reflexivity of \sim is immediate, since $\langle X, Y \rangle$ is D -witnessed; while symmetry and transitivity follow from the principles we established concerning ' $=$ '.

Consider (B). We shall sketch a proof of (i). Firstly from (7), (10), (11) and (13) it follows that

$$(a) \quad t=t \in X \iff d=t \in X \text{ for some } d \in D$$

$$(b) \quad t=u \in X \iff d=t \in X \text{ and } d=u \in X \text{ for some } d \in D$$

These facts will be useful in checking out the details of what follows.

Now define a function δ from the parameter-free terms of $L(C)$ onto $D/\sim \cup \{\Theta\}$ by

$$\delta(t) = \begin{cases} |d|_{\sim} & \text{if } d=t \in X \\ \Theta & \text{otherwise} \end{cases}$$

By (11) and (13), this definition makes sense, and δ is onto, since $\delta(d) = |d|_{\sim}$ if $d \in D$, and $\delta(\Theta) = \Theta$ (because, by (Θ) , $d = \Theta \in Y$ for all $d \in D$).

We now attempt to define the model M as follows:

$$D_M = D/\sim$$

if $P \in \text{Prd}(L(C))$ and $m = \lambda(P)$,

$$P_M(\delta(t_1), \dots, \delta(t_m)) = \begin{cases} \top & \text{if } Pt_1 \dots t_m \in X \\ \perp & \text{if } \neg Pt_1 \dots t_m \in X \end{cases}$$

if $f \in \text{Fnc}(L(C))$ and $n = \mu(f)$,

$$f_M(\delta(t_1), \dots, \delta(t_n)) = \delta(ft_1 \dots t_n)$$

if $c \in \text{Cns}(L(C))$,

$$c_M = \delta(c) .$$

Since δ is onto, our definitions do not ignore any elements of $D/\sim \cup \{\Theta\}$, but we must check that P_M and f_M are 'well defined', and that they are monotonic.

To this end, consider, first, terms $t_i, t'_i, ft_1 \dots t_i \dots t_n$ and $ft_1 \dots t'_i \dots t_n$, and invoke (SD) to show

(1) if $\delta(t_i) = \Theta$ and $\delta(ft_1 \dots t_i \dots t_n) \in D/\sim$, then

$$\delta(ft_1 \dots t_i \dots t_n) = \delta(ft_1 \dots t'_i \dots t_n)$$

Hence, switching t_i and t'_i ,

(1') if $\delta(t'_i) = \Theta$ and $\delta(ft_1 \dots t'_i \dots t_n) \in D/\sim$, then

$$\delta(ft_1 \dots t_i \dots t_n) = \delta(ft_1 \dots t'_i \dots t_n)$$

From (1) and (1') we can then deduce

(2) If $\delta(t_i) = \delta(t'_i) = \Theta$, then $\delta(ft_1 \dots t_i \dots t_n) = \delta(ft_1 \dots t'_i \dots t_n)$

(whether or not this value of δ is in D/\sim).

Also, invoking (SI), show

(3) if $\delta(t_i) = \delta(t'_i) \in D/\sim$, then $\delta(ft_1 \dots t_i \dots t_n) = \delta(ft_1 \dots t'_i \dots t_n)$

Then (2) and (3) imply that our specification of f_M is 'well-defined', and (1) and (3) imply that f_M is monotonic. By a similar – but more straight-forward – argument, considering atomic sentences $Pt_1 \dots t_i \dots t_m$ and their negations, we can show that P_M is 'well-defined' and monotonic.

Hence we do at least have a model M for $L(C)$. Clearly $M(t) = \delta(t)$ for all parameter-free terms t , and we can now check, by induction on the complexity of sentences, that their truth-falsity conditions are what we want.

Given an atomic sentence of the form $Pt_1 \dots t_m$, this is immediate from the definition of M . To check that

$$M(t=u) = \top \iff t=u \in X$$

we can make use of fact (b), above, and invoke principle (13); and to check that

$$M(t=u) = \perp \iff t \neq u \in X$$

we can use fact (a) and the principles (Det), (SI), (4) and (5). For the induction steps involving the connectives we may invoke the same properties of exhaustive and rejected pairs as we did in the case of (III.2.15). But note: this time we are given a model and we are inductively checking that its truth-falsity conditions yield the right values; whereas, previously, we were given the values and had to check that the truth-falsity conditions were right, in order to show that the valuation was a 'model'. The quantifier steps follow, given (II.3.1) and (II.3.2), from the fact that $\langle X, Y \rangle$ is D -witnessed.

Finally, we must show that M is a model of $\frac{C}{\perp}$. We argue by contraposition. Say $\Gamma(x_1, \dots, x_n) \not\vdash_M \Delta(x_1, \dots, x_n)$, so there exist $\delta(t_1), \dots, \delta(t_n)$ such that

$$\Gamma(x_1, \dots, x_n) \not\vdash_{M, \delta(t_1), \dots, \delta(t_n)} \Delta(x_1, \dots, x_n)$$

and so, by (II.3.2),

$$\Gamma(t_1, \dots, t_n) \not\equiv_M \Delta(t_1, \dots, t_n)$$

Hence

$$\text{either (i) } \Gamma(\vec{t}) \not\equiv_M^* \Delta(\vec{t}) \text{ or (ii) } \Gamma(\vec{t}), * \not\equiv_M \Delta(\vec{t}).$$

But (i) implies that $\Gamma(\vec{t}) \subseteq X$ and $\Delta(\vec{t}) \subseteq Y$, and so $\Gamma(\vec{t}) \not\equiv^C * \Delta(\vec{t})$.

Then, *a fortiori*,

$$\Gamma(t_1, \dots, t_n) \not\equiv^C \Delta(t_1, \dots, t_n)$$

and so, by (S),

$$\Gamma(x_1, \dots, x_n) \not\equiv^C \Delta(x_1, \dots, x_n)$$

From (ii) we can argue similarly, invoking the rule twist – as we did in the proof of (III.2.15).

Now we have shown how to prove part (i) of assertion (2). Part (ii) can be proved in an exactly parallel way. □

(IV.1.15) has an obvious converse: a model of \vdash^C with domain D/\sim , for some equivalence relation \sim on a subset D of C , determines a $*$ -right $L(C)$ -exhaustive D -witnessed pair (and also a $*$ -left one) which would, according to the definition we have provided, then determine that same model again.

The next obvious question is: what about complete theories? It is possible to bring 'complete theories' into the picture, but, as we stated at the end of III.2, that notion is not entirely straight-forward, and we shall not discuss it until section 4.

IV.2 THEORIES AND FORMULAE

The main result in this section is a direct analogue of the Interpolant Exclusion Lemma of Chapter III. Let \vdash_1 and \vdash_2 be theories in the (countable) languages L_1 and L_2 , and let $\Lambda \subseteq \text{Frm}(L_1) \cap \text{Frm}(L_2)$. Assume that Λ contains \top and \perp ; is closed under \wedge and \vee ; furthermore, is closed under uniform substitution of parameters for parameters; and also under the application of quantifiers. By this last closure condition we mean that if x is a variable and $\phi \in \Lambda$, then $\forall x\phi \in \Lambda$ and $\exists x\phi \in \Lambda$. Consider subsets Γ and Δ of $\text{Frm}(L_1)$ and Ξ and Υ of $\text{Frm}(L_2)$, and let C be a denumerable set of constants not already occurring in either language.

Lemma IV.2.1 (Interpolant Exclusion)

There is no $\lambda \in \Lambda$ such that $\Gamma \vdash_1 \lambda, \Delta$ and $\Xi, \lambda \vdash_2 \Upsilon$
if and only if

there exist

a subset $D \subseteq C$,

a substitution function $\pi: \text{Var} \rightarrow C$

subsets Γ' and Δ' of $\text{Snt}(L_1(C))$ and Ξ' and Υ' of $\text{Snt}(L_2(C))$

such that

$$\pi\Gamma \subseteq \Gamma', \quad \pi\Delta \subseteq \Delta', \quad \pi\Xi \subseteq \Xi', \quad \pi\Upsilon \subseteq \Upsilon'$$

and

$\langle \Gamma', \Delta' \rangle$ is $L_1(C)$ -exhaustive, \vdash_1^C -rejected and D -witnessed

$\langle \Xi', \Upsilon' \rangle$ is $L_2(C)$ -exhaustive, \vdash_2^C -rejected and D -witnessed

and also

$$\Gamma' \cap \Lambda[C] \cap \Upsilon' = \emptyset$$

where $\Lambda[C]$ is the following set of sentences common to both $L_1(C)$ and $L_2(C)$:

$$\{\lambda(c_1, \dots, c_n) \mid \lambda(x_1, \dots, x_n) \in \Lambda \text{ and } c_1, \dots, c_n \in C\}$$

Proof: 'If' is easy: because $\langle \Gamma', \Delta' \rangle$ and $\langle \Xi', \Upsilon' \rangle$ are exhaustive, given any $\lambda(\vec{c}) \in \Lambda[C]$,

for "L" read "H" (1.16)

$$\Gamma' \vdash_1^C \lambda(\vec{c}), \Delta' \iff \lambda(\vec{c}) \in \Gamma'$$

and

$$\Xi', \lambda(\vec{c}) \vdash_2^C \Upsilon' \iff \lambda(\vec{c}) \in \Upsilon'$$

This rules out an interpolant from Λ , since we have the rule (S).

'Only if' will take some time. First a few subsidiary lemmas:

Lemma IV.2.2: Given any theory \vdash , if $\{\langle X_i, Y_i \rangle\}_{i < \omega}$ is a chain of pairs of sets such that $X_i \not\vdash Y_i$ for all i , then $\bigcup_i X_i \not\vdash \bigcup_i Y_i$.

Proof: Easy.

Lemma IV.2.3: Given any theory \vdash , and any set C of new constants, if $X \not\vdash Y$ and π is a one-one substitution function: $Var \rightarrow C$, then $\pi X \not\vdash^C \pi Y$.

Proof: We argue by contraposition. If $\pi X \vdash^C \pi Y$ then there is a proof of $\pi X_0 \succ \pi Y_0$ (where X_0 and Y_0 finite subsets of X and Y) from sequents in \vdash . Let c_1, \dots, c_n be the elements of C which occur in this proof, and let x_1, \dots, x_n be distinct variables which occur nowhere in it (not even bound). If we replace c_i by x_i throughout, then we have a proof, entirely in the language of L , from sequents in \vdash , of $X'_0 \succ Y'_0$, where X'_0 and Y'_0 are obtained from X_0 and Y_0 by replacing any c_i which occurs by x_i . But then, by repeated applications of (S), replacing these x_i by the variable which π sent to c_i , we know that $X_0 \succ Y_0 \in \vdash$, and hence $X \vdash Y$. □

Lemma IV.2.4: Given a theory \vdash and a set of new constants C , if $X(c/x) \vdash^C Y(c/x)$, then $X \vdash^C Y$.

Proof: Observe that a proof of $X_0(c/x) \succ Y_0(c/x)$ can be converted into a proof of $X_0 \succ Y_0$, by replacing c by a totally extraneous variable y , and then using (S) with a substitution of x for y . □

Now, assuming that \vdash_1 , \vdash_2 , Λ and C meet the conditions for our

main lemma, let us reintroduce the abbreviation $\langle U, V, X, Y \rangle$ excludes $\Lambda[C]$ to mean that there is no $\lambda \in \Lambda[C]$ such that

$$U \vdash_1^C \lambda, V \quad \text{and} \quad X, \lambda \vdash_2^C Y$$

we shall need to know the following:

Lemma IV.2.5: Given that $\langle U, V, X, Y \rangle$ excludes $\Lambda[C]$,

- (1) If $\sigma \in Snt(L_1(C))$, then
either $\langle U \cup \{\sigma\}, V, X, Y \rangle$ excludes $\Lambda[C]$
or $\langle U, \{\sigma\} \cup V, X, Y \rangle$ excludes $\Lambda[C]$
- (2) If $\tau \in Snt(L_2(C))$, then
either $\langle U, V, X \cup \{\tau\}, Y \rangle$ excludes $\Lambda[C]$
or $\langle U, V, X, \{\tau\} \cup Y \rangle$ excludes $\Lambda[C]$

Proof: (1): Assume otherwise, then for some λ_1 and λ_2 in $\Lambda[C]$,

$$\begin{array}{ll} \text{(i)} & U, \sigma \vdash_1^C \lambda_1, V \\ \text{(iii)} & U \vdash_1^C \sigma, \lambda_2, V \end{array} \quad \begin{array}{ll} \text{(ii)} & X, \lambda_1 \vdash_2^C Y \\ \text{(iv)} & X, \lambda_2 \vdash_2^C Y \end{array}$$

By (T) and the rules for \vee , (i) and (iii) imply that

$$U \vdash_1^C \lambda_1 \vee \lambda_2, V$$

while (ii) and (iv) imply that

$$X, \lambda_1 \vee \lambda_2 \vdash_2^C Y$$

But this is a contradiction, since Λ is closed under disjunction and so $\lambda_1 \vee \lambda_2 \in \Lambda[C]$.

(2) is proved similarly, from the fact that Λ is closed under conjunction. □

Lemma IV.2.6: Assuming that c does not occur in any formula in U, V, X or Y , nor in $\phi(x)$,

- (1) If $\langle U \cup \{\exists x \phi(x)\}, V, X, Y \rangle$ excludes $\Lambda[C]$, then so does
 $\langle U \cup \{\exists x \phi(x), \phi(c), c=c\}, \{c \neq c\} \cup V, X \cup \{c=c\}, \{c \neq c\} \cup Y \rangle$

(2) If $\langle U, \{\forall x \phi(x)\} \cup V, X, Y \rangle$ excludes $\Lambda[C]$, then so does
 $\langle U \cup \{c=c\}, \{c \neq c, \phi(c), \forall x \phi(x)\} \cup V, X \cup \{c=c\}, \{c \neq c\} \cup Y \rangle$

(3) If $\langle U, V, X \cup \{\exists x \phi(x)\}, Y \rangle$ excludes $\Lambda[C]$, then so does
 $\langle U \cup \{c=c\}, \{c \neq c\} \cup V, X \cup \{\exists x \phi(x), \phi(c), c=c\}, \{c \neq c\} \cup Y \rangle$

(4) If $\langle U, V, X, \{\forall x \phi(x)\} \cup Y \rangle$ excludes $\Lambda[C]$, then so does
 $\langle U \cup \{c=c\}, \{c \neq c\} \cup V, X \cup \{c=c\}, \{c \neq c, \phi(c), \forall x \phi(x)\} \cup Y \rangle$

Proof: We argue by contraposition. So, say the consequent of (1) is false. Then there is a $\lambda(c) \in \Lambda[C]$ — $\lambda(c)$ might contain c , but it need not — such that

(i) $U, \exists x \phi(x), \phi(c), c=c \stackrel{C}{\vdash}_1 c \neq c, \lambda(c), V$

(ii) $X, \lambda(c), c=c \stackrel{C}{\vdash}_2 c \neq c, Y$

But, by (\exists') and (S),

$$\lambda(c), c=c \stackrel{C}{\vdash}_1 c \neq c, \exists y \lambda(y)$$

Hence, from (i), by (T),

(iii) $U, \exists x \phi(x), \phi(c), c=c \stackrel{C}{\vdash}_1 c \neq c, \exists y \lambda(y), V$

And, from (ii), by (IV.2.4),

$$X, \lambda(y), y=y \stackrel{C}{\vdash}_2 y \neq y, Y$$

Hence by (\exists) ,

(iv) $X, \exists y \lambda(y) \stackrel{C}{\vdash}_2 Y$

However, from (iii), by (IV.2.4),

$$U, \exists x \phi(x), \phi(x), x=x \stackrel{C}{\vdash}_1 x \neq x, \exists y \lambda(y), V$$

and so, by (\exists)

(v) $U, \exists x \phi(x) \stackrel{C}{\vdash}_1 \exists y \lambda(y), V$

But then, since Λ is closed under application of quantifiers, $\exists y \lambda(y) \in \Lambda[C]$; and so (iv) and (v) imply that the antecedent of (1) fails.

Part (2) may be proved similarly, applying the rules for \forall to $\phi(x)$.

after "enumerable" insert
" and $C_1 \cap C_2 = \emptyset$ "

While parts (3) and (4) follow the same patterns, this time relying on the fact that if $\lambda(c) \in \Lambda[C]$ then $\forall y \lambda(y) \in \Lambda[C]$ and applying the rules for \forall to $\lambda(y)$. □

We may now embark on the main construction required to prove (IV.2.1).

First, let $C = C_1 \cup C_2$ where both C_1 and C_2 are denumerable. We can then pick a substitution function straight away: let π be some one-one function: $Var \rightarrow C_1$. And, for future reference, let us index C_2 as $\{c_0, c_1, \dots\}$.

Now let $\{\sigma_i\}_{i < \omega}$ be an enumeration of $Snt(L_1(C))$, and $\{\tau_i\}_{i < \omega}$ an enumeration of $Snt(L_2(C))$. We shall define, by induction on i , a sequence $\{\langle U_i, V_i, X_i, Y_i, D_i \rangle\}_{i < \omega}$ where for all $i < j < \omega$,

$$\pi \Gamma \subseteq U_i \subseteq U_j \subseteq Snt(L_1(C))$$

$$\pi \Delta \subseteq V_i \subseteq V_j \subseteq Snt(L_1(C))$$

$$\pi \Xi \subseteq X_i \subseteq X_j \subseteq Snt(L_2(C))$$

$$\pi \Upsilon \subseteq Y_i \subseteq Y_j \subseteq Snt(L_2(C))$$

and

$$D_i \subseteq D_j \subseteq C_2$$

in such a way that we may then put

$$\Gamma' = \bigcup_i U_i \quad \Delta' = \bigcup_i V_i \quad \Xi' = \bigcup_i X_i \quad \Upsilon' = \bigcup_i Y_i$$

and

$$D = \bigcup_i D_i$$

We shall be assured that these sets satisfy the theorem, if we check — by induction along the way — that our definition guarantees that for all n :

$$(1) \quad \text{For all } i < n: \sigma_i \in U_{2n} \cup V_{2n} \text{ and } \tau_i \in X_{2n+1} \cup Y_{2n+1}$$

(2)(a) For all $i < n$:

if $\sigma_i = \exists x \phi(x) \in U_{2n}$, then $\phi(d) \in U_{2n}$ for some $d \in D_{2n}$

if $\sigma_i = \forall x \phi(x) \in V_{2n}$, then $\phi(d) \in V_{2n}$ for some $d \in D_{2n}$

if $\tau_i = \exists x \phi(x) \in X_{2n+1}$, then $\phi(d) \in X_{2n+1}$ for some $d \in D_{2n+1}$

if $\tau_i = \forall x \phi(x) \in Y_{2n+1}$, then $\phi(d) \in Y_{2n+1}$ for some $d \in D_{2n+1}$

(b) For all $d \in D_n$: $d = d \in U_n$, $d \neq d \in V_n$, $d = d \in X_n$, $d \neq d \in Y_n$

(3) $\langle U_n, V_n, X_n, Y_n \rangle$ excludes $\Lambda[C]$

Clearly (1) will imply exhaustiveness, and (2) will imply D -witnessedness. On the other hand, (3), together with (IV.2.2), will imply that $\langle \Gamma', \Delta', \Xi', T' \rangle$ excludes $\Lambda[C]$. From this we may then deduce, firstly, that $\Gamma' \cap \Lambda[C] \cap T' = \emptyset$, and, secondly, that $\langle \Gamma', \Delta' \rangle$ is $\frac{C}{1}$ -rejected (because $\Lambda[C]$ contains \perp) and that $\langle \Xi', T' \rangle$ is $\frac{C}{2}$ -rejected (because $\Lambda[C]$ contains \top).

Also, to make sure that the construction itself is good, we may check

(4) Only finitely many elements of C_2 occur in formulae in

U_n, V_n, X_n and Y_n .

This will ensure that there are infinitely many left over at each stage.

To begin the sequence we put

$U_0 = \pi\Gamma$, $X_0 = \pi\Delta$, $X_0 = \pi\Xi$, $Y_0 = \pi T$ and $D_0 = \emptyset$.

Note that conditions (1), (2) and (4) are trivially satisfied, for $n=0$; while condition (3) is satisfied in virtue of (IV.2.3).

Now, assuming that we have reached stage $2n$ and that (1) - (4) are satisfied, our definition of $\langle U_{2n+1}, V_{2n+1}, X_{2n+1}, Y_{2n+1} \rangle$ divides into cases, as follows:

Case 1: $\langle U_{2n}, V_{2n}, X_{2n} \cup \{\tau_n\}, Y_{2n} \rangle$ excludes $\Lambda[C]$

subcase (i): $\tau_n = \exists x \phi(x)$ for some $\phi(x)$

subcase (ii): not subcase (i).

Case 2: Not Case 1

subcase (i): $\tau_n = \forall x \phi(x)$ for some $\phi(x)$

subcase (ii): not subcase (i) .

Case 1:

subcase (i):

$$U_{2n+1} = U_{2n} \cup \{c=c\}$$

$$V_{2n+1} = V_{2n} \cup \{c \neq c\}$$

$$X_{2n+1} = X_{2n} \cup \{c=c, \phi(c), \tau_n\}$$

$$Y_{2n+1} = Y_{2n} \cup \{c \neq c\}$$

and

$$D_{2n+1} = D_{2n} \cup \{c\}$$

where c is the first element of C_2 (according to our enumeration) which does not occur in any formula in $U_{2n} \cup V_{2n} \cup X_{2n} \cup Y_{2n} \cup \{\phi(x)\}$. Such a c exists, since condition (4) holds for $2n$. Note that conditions (1), (2) and (4) are clearly satisfied for $2n+1$, and, by (IV.2.6), condition (3) is also satisfied.

subcase (ii):

$$U_{2n+1} = U_{2n}$$

$$V_{2n+1} = V_{2n}$$

$$X_{2n+1} = X_{2n} \cup \{\tau_n\}$$

$$Y_{2n+1} = Y_{2n}$$

and

$$D_{2n+1} = D_{2n}$$

And the conditions (1) - (4) are satisfied.

Case 2:

Note, first, that, by (IV.2.5), in this case $\langle U_{2n}, V_{2n}, X_{2n}, \{\tau_n\} \cup Y_{2n} \rangle$ excludes $\Lambda[C]$.

subcase (i):

$$U_{2n+1} = U_{2n} \cup \{c=c\}$$

$$V_{2n+1} = V_{2n} \cup \{c \neq c\}$$

$$X_{2n+1} = X_{2n} \cup \{c=c\}$$

$$Y_{2n+1} = Y_{2n} \cup \{c \neq c, \phi(c), \tau_n\}$$

and

$$D_{2n+1} = D_{2n} \cup \{c\}$$

where c is a new constant as in case 1(i). Conditions (1), (2) and (4) are clearly satisfied for $2n+1$, and, given our note above concerning case 2, then, by (IV.2.6), condition (3) is also satisfied.

subcase (ii):

$$U_{2n+1} = U_{2n}$$

$$V_{2n+1} = V_{2n}$$

$$X_{2n+1} = X_{2n}$$

$$Y_{2n+1} = Y_{2n} \cup \{\tau_n\}$$

and

$$D_{2n+1} = D_{2n}$$

And, again, conditions (1) - (4) are satisfied.

Having reached stage $2n+1$, the definition of

$\langle U_{2n+2}, V_{2n+2}, X_{2n+2}, Y_{2n+2} \rangle$ likewise divides into cases:

Case 1: $\langle U_{2n+1} \cup \{\sigma_n\}, V_{2n+1}, X_{2n+1}, Y_{2n+1} \rangle$ excludes $\Lambda[C]$

subcase (i): $\sigma_n = \exists x \phi(x)$ for some $\phi(x)$

subcase (ii): not subcase (i).

Case 2: Not Case 1

subcase (i): $\sigma_n = \forall x \phi(x)$ for some $\phi(x)$

subcase (ii): not subcase (i)

We shall not actually write out the definitions, since they are exactly parallel to what we had before, this time making special additions to the first two coordinates of the quadruple. And we check that (1) - (4) hold for $2n+2$ in a similar way. □

This completes the proof of (IV.2.1). The construction may readily be modified so as to cater for the case where either L_1 or L_2 is not countable. For, say that κ is the cardinality of the larger of the two languages, and pick C with κ elements. Then let C_1 be denumerable, as before, and index C_2 (the rest of C) by the least ordinal of size κ . $Snt(L_1(C))$ and $Snt(L_2(C))$ will also be of size κ , and can be indexed by the same ordinal as C_2 . Proceed with the construction as before, accumulating everything at limit stages: since theories are relations between finite sets, $\Lambda[C]$ -exclusion will be preserved.

We may now easily deduce completeness, compactness and our Löwenheim-Skolem theorem.

Corollary IV.2.7: (IV.1.14) holds.

Proof: Say \vdash is a theory in L and $\Gamma \not\vdash \Delta$. Let $\vdash_1 = \vdash$, and let \vdash_2 be logic in L . Put $\Lambda = Frm(L)$ and $\Xi = T = \emptyset$. There is no $\lambda \in \Lambda$ such that

$$\Gamma \vdash_1 \lambda, \Delta \text{ and } \lambda \vdash_2$$

since, otherwise, $\Gamma \vdash \Delta$, by (T). Hence we have an $L(C)$ -exhaustive \vdash^C -rejected, D -witnessed extension $\langle \Gamma', \Delta' \rangle$ of $\langle \pi\Gamma, \pi\Delta \rangle$, where $D \subseteq C$, C is a denumerable set, and π is a one-one substitution function $Var \rightarrow C$. $\langle \Gamma', \Delta' \rangle$ must either be \star -right or \star -left, and so, by (IV.1.15), we have a model M' of \vdash^C , with a domain D/\sim , such that $\Gamma' \not\vdash_{M'} \Delta'$. Hence, if M is the reduct of M' to the language L , M is a model of \vdash , and

$$\Gamma \not\vdash_{M,s} \Delta$$

where s is the assignment given by $s: x \mapsto M'(\pi x)$. □

In fact (IV.1.14) holds for uncountable L , if we delete 'countable', and so we have completeness and compactness for the general case. This follows from our remark on how to modify (IV.2.1).

We shall now apply (IV.2.1) to provide criteria for interpolant-exclusion in terms of models for the unexpanded languages L_1 and L_2 . We shall not, however, fully tap the range of results that (IV.2.1) has to offer, since we shall be making several further assumptions about Λ . Not only shall we assume (as we did at this stage in Chapter III) that Λ is closed under \neg , but also that it is closed under \forall . This means ignoring a treatment of classical formulae, but the analogues of (III.3.4) and (III.3.5) will be simpler. Moreover, we shall bypass possible subtleties concerning the relation between the domains of models, as a result of assuming that Λ contains the formulae $x=y$ for any variables x and y . Hence, but for the possible absence of Θ , Λ will be a full sublanguage of L_1 and L_2 .

First the definition of two relations between models M_1 of L_1 and M_2 of L_2 .

$$\left. \begin{array}{l} M_1 \sqsubseteq_{\Lambda} M_2 \\ M_1 \sqsupseteq_{\Lambda} M_2 \end{array} \right\} \iff \left\{ \begin{array}{l} M_1 \text{ and } M_2 \text{ share the} \\ \text{domain } D, \text{ and for all} \\ s: \text{Var} \rightarrow D, \text{ and all } \lambda \in \Lambda \end{array} \right\} \left\{ \begin{array}{l} M_{1_s}(\lambda) \sqsubseteq M_{2_s}(\lambda) \\ M_{1_s}(\lambda) \sqsupseteq M_{2_s}(\lambda) \end{array} \right.$$

These definitions are symmetrical in L_1 and L_2 , and, as before, we shall write ' $M_1 \supseteq_{\Lambda} M_2$ ' to mean that $M_2 \sqsubseteq_{\Lambda} M_1$. We have employed *de re* assignments, viz. assignments s such that $s(x) \neq \Theta$ for all x , but note that if Λ is closed under substitution of Θ for parameters, then these relations could equivalently have been defined in terms of arbitrary assignments $\text{Var} \rightarrow D \cup \{\Theta\}$.

Note, also, that, whatever set Λ may be, if, instead, we take its closure under all the conditions we shall be assuming for interpolant sets, then the relation is not affected.

Lemma IV.2.8: If Λ' is obtained from $\Lambda \cup \{\top, \perp, x=y\}$ (x and y distinct) by closure under the connectives, the quantifiers and uniform substitution of

parameters for parameters, then $\sqsubseteq_{\Lambda'}$ is the same relation as \sqsubseteq_{Λ} and $\square_{\Lambda'}$ is the same relation as \square_{Λ} .

Proof: Easy. Note that it is because the definitions stipulate a common domain that closure under quantifiers makes no difference.

There is another simple fact, along similar lines, which we shall subsequently find a use for:

Lemma IV.2.9: If Λ' is the result of closing Λ under the substitution of Θ for parameters, then

$$M_1 \square_{\Lambda'} M_2 \iff M_1 \square_{\Lambda} M_2$$

provided that the domain of M_1 and M_2 is non-empty.

Proof: Monotonicity.

Also:

Lemma IV.2.10: If $\Lambda = \text{Frm}(L_1) = \text{Frm}(L_2)$, then \sqsubseteq_{Λ} and \square_{Λ} are the same as \sqsubseteq and \square , where these unrelativized relations are those defined in section II.2.

In addition to all this, we need relations between assignments $\text{Var} \rightarrow D \cup \{\Theta\}$ for a given domain D :

$$s_1 \sqsubseteq s_2 \iff \text{for all } x: s_1(x) \sqsubseteq s_2(x)$$

$$s_1 \square s_2 \iff \text{for all } x: s_1(x) \square s_2(x)$$

And we shall often write ' $s_1 \supseteq s_2$ ' to mean that $s_2 \sqsubseteq s_1$.

Now we are in a position to consider 'interpolant-excluding model-pairs'. Assuming that $\vdash_1, \vdash_2, \Lambda, \Gamma, \Delta, \Xi$ and \top satisfy the conditions for (IV.2.1), also that Λ contains identity formulae $x=y$, and that Λ is closed under negation:

insert "(2)" at the beginning of l. 12
insert "(3)" at the beginning of l. 13

Corollary IV.2.11: In the three cases

$$(1) \quad * \in \Delta \quad \text{and} \quad * \in \Upsilon$$

$$(2) \quad * \in \Gamma \quad \text{and} \quad * \in \Xi$$

$$(3) \quad * \in \Delta \quad \text{and} \quad * \in \Xi$$

there exists no $\lambda \in \Lambda$ such that

$$\Gamma \vdash_1 \lambda, \Delta \quad \text{and} \quad \Xi, \lambda \vdash_2 \Upsilon$$

iff there exist models M_1 of \vdash_1 and M_2 of \vdash_2 which have a common domain D , and assignments s_1 and s_2 into $D \cup \{\Theta\}$, such that

$$\Gamma \not\vdash_{M_1, s_1} \Delta \quad \text{and} \quad \Xi \not\vdash_{M_2, s_2} \Upsilon$$

and such that, respectively,

$$(1) \quad \begin{array}{lll} (a) \quad M_1 \sqsubseteq_{\Lambda} M_2 & (b) \quad s_1 \sqsubseteq s_2 & (c) \quad \forall \lambda \in \Lambda. M_{1, s_1}(\lambda) \sqsubseteq M_{2, s_2}(\lambda) \\ (a) \quad M_1 \supseteq_{\Lambda} M_2 & (b) \quad s_1 \supseteq s_2 & (c) \quad \forall \lambda \in \Lambda. M_{1, s_1}(\lambda) \supseteq M_{2, s_2}(\lambda) \\ (a) \quad M_1 \square_{\Lambda} M_2 & (b) \quad s_1 \square s_2 & (c) \quad \forall \lambda \in \Lambda. M_{1, s_1}(\lambda) \square M_{2, s_2}(\lambda) \end{array}$$

Proof: 'Only if' may be checked directly. For 'if', if we apply (IV.1.15) to (IV.2.1), then we know that there exist models M'_1 of \vdash_1^C and M'_2 of \vdash_2^C such that

$$\Gamma' \not\vdash_{M'_1} \Delta' \quad \text{and} \quad \Xi' \not\vdash_{M'_2} \Upsilon'$$

where for some one-one substitution function $\pi: Var \rightarrow C$, $\pi\Gamma \subseteq \Gamma'$, $\pi\Delta \subseteq \Delta'$, $\pi\Xi \subseteq \Xi'$ and $\pi\Upsilon \subseteq \Upsilon'$, and the domains of M'_1 and M'_2 are obtained by factoring out some subset D of C modulo the equivalence relations defined, respectively, by

$$d=e \in \Gamma' \quad \text{and} \quad d \neq e \in \Delta'$$

and by

$$d=e \in \Xi' \quad \text{and} \quad d \neq e \in \Upsilon'$$

However, given that Λ contains atomic identity formulae and is closed under negation, it is easy to check that, since $\Gamma' \cap \Lambda[C] \cap \Upsilon' = \emptyset$, these relations in fact coincide. Use (Det). Hence M'_1 and M'_2 have the same domain D/\sim , where \sim is the equivalence relation in question. (Note also that for any $d \in D$ $M'_1(d) = M'_2(d)$.)

But now, if we put

$$M_1 = M'_1 |_{L_1} \quad s_1 : x \mapsto M'_1(\pi x)$$

$$M_2 = M'_2 |_{L_2} \quad s_2 : x \mapsto M'_2(\pi x)$$

then M_1 is a model of $\frac{\Gamma}{1}$ and M_2 is a model of $\frac{\Xi}{2}$, and

$$\Gamma \not\equiv_{M_1, s_1} \Delta \quad \text{and} \quad \Xi \not\equiv_{M_2, s_2} \top$$

It remains to check (a) (b) and (c) in the three different cases.

Case (1): Say $\lambda \in \Lambda[C]$, then

$$\begin{aligned} M'_1(\lambda) = \top &\iff \lambda \in \Gamma' \\ &\implies \lambda \notin \top' \\ &\iff \lambda \in \Xi' \iff M'_2(\lambda) = \top \end{aligned}$$

Hence, since Λ is closed under negation, $M'_1(\lambda) \sqsubseteq M'_2(\lambda)$ for any $\lambda \in \Lambda[C]$.

From this we can deduce (a) (b) and (c) :

(a): Say $\lambda(x_1, \dots, x_n) \in \Lambda$. Then, for any $d_1, \dots, d_n \in D$,

$$M'_1(\lambda(d_1, \dots, d_n)) \sqsubseteq M'_2(\lambda(d_1, \dots, d_n))$$

and so, for any $s : \text{Var} \rightarrow D/\sim$, $M_{1_s}(\lambda(x_1, \dots, x_n)) \sqsubseteq M_{2_s}(\lambda(x_1, \dots, x_n))$.

(b): Observe that if $c \in C$ and $d \in D$, $d=c \in \Lambda[C]$, and so for any variable x and any $d \in D$,

$$M'_1(d=\pi x) \sqsubseteq M'_2(d=\pi x)$$

Hence, it is easy to check that for any x , $M'_1(\pi x) \sqsubseteq M'_2(\pi x)$ — i.e.

$$s_1(x) \sqsubseteq s_2(x).$$

(c): Immediate.

Case (2): Similarly.

Case (3): Say $\lambda \in \Lambda[C]$, then

$$\begin{aligned} M'_1(\lambda) = \top &\iff \lambda \in \Gamma' \\ &\implies \lambda \notin \top' \\ &\iff \lambda \in \Xi' \iff M'_2(\lambda) \neq \perp \end{aligned}$$

for "in cases (1) and (2)" read "in all cases" ; and delete "Moreover,.... also."

And, similarly, $M'_1(\lambda) = \perp \Rightarrow M'_2(\lambda) \neq \top$ Hence, $M'_1(\lambda) \sqsubset M'_2(\lambda)$ for any $\lambda \in \Lambda[C]$.

From this we can deduce (a) (b) and (c):

(a): Say $\lambda(x_1, \dots, x_n) \in \Lambda$. Then, for any $d_1, \dots, d_n \in D$,

$$M'_1(\lambda(d_1, \dots, d_n)) \sqsubset M'_2(\lambda(d_1, \dots, d_n))$$

and so, for any $s: \text{Var} \rightarrow D/\sim$, $M_{1_s}(\lambda(x_1, \dots, x_n)) \sqsubset M_{2_s}(\lambda(x_1, \dots, x_n))$

(b): Observe that if $c \in C$ and $d \in D$, $d=c \in \Lambda[C]$, and so for any variable x and any $d \in D$,

$$M'_1(d=\pi x) \sqsubset M'_2(d=\pi x)$$

Hence, it is easy to check that for any x , $M'_1(\pi x) \sqsubset M'_2(\pi x)$ — i.e.

$s_1(x) \sqsubset s_2(x)$.

(c): Immediate. □

Simplification IV.2.12: If $\Delta = \emptyset = \Xi$, then we may take $s_1 = s_2$.

Proof: In cases (1) and (2), either s_1 or s_2 , as defined, would serve both roles. In case (3), take the assignment s such that, if $d \in D$, then $s(x) = d$ iff either $s_1(x) = d$ or $s_2(x) = d$ (and $s(x) = \ominus$ otherwise). □

Simplification IV.2.13: If Λ is closed under substitution of \ominus for parameters, then condition (c) is trivial — follows trivially from (a) — in cases (1) and (2). Moreover, if $\Delta = \emptyset = \top$, then condition (c) is similarly trivial in case (3) also.

To make use of these general results we shall need some theory-relative notions of equivalence, degree-of-definedness and compatibility between formulae. These generalize the relations considered in section II.2. Given completeness, the following pairs of definitions are obviously equivalent:

\vdash -equivalence:

$$\begin{aligned} \phi \approx_{\vdash} \psi &\iff \phi \vdash \psi \text{ and } \psi \vdash \phi \\ &\iff M_s(\phi) = M_s(\psi) \text{ for any } M \in K(\vdash) \text{ and any } s \end{aligned}$$

degree-of-definedness in \vdash :

$$\begin{aligned} \phi \sqsubseteq_{\vdash} \psi &\iff \phi \vdash \psi, * \text{ and } *, \psi \vdash \phi \\ &\iff M_s(\phi) \sqsubseteq M_s(\psi) \text{ for any } M \in K(\vdash) \text{ and any } s \end{aligned}$$

\vdash -compatibility:

$$\begin{aligned} \phi \square_{\vdash} \psi &\iff * \vdash \phi, \neg \psi \text{ and } * \vdash \neg \phi, \psi \\ &\iff M_s(\phi) \square M_s(\psi) \text{ for any } M \in K(\vdash) \text{ and any } s \end{aligned}$$

Finally, we state the 'combination lemmas', which bring the three separate cases of (IV.2.11) back together again.

Lemma IV.2.14: If Λ is closed under \wedge, \vee and \ast ,

$$\exists \lambda \in \Lambda. \quad \Gamma \vdash_1 \lambda, \Delta \text{ and } \Xi, \lambda \vdash_2 \top$$

iff the following all hold

- (1) $\exists \lambda_1 \in \Lambda. \quad \Gamma \vdash_1 \lambda_1, \Delta, *$ and $\Xi, \lambda_1 \vdash_2 \top, *$
- (2) $\exists \lambda_2 \in \Lambda. \quad *, \Gamma \vdash_1 \lambda_2, \Delta$ and $*, \Xi, \lambda_2 \vdash_2 \top$
- (3) $\exists \lambda_3 \in \Lambda. \quad \Gamma \vdash_1 \lambda_3, \Delta, *$ and $*, \Xi, \lambda_3 \vdash_2 \top$

Proof: 'Only if' is trivial: put $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$. For 'if' put $\lambda = ((\lambda_1 \wedge \lambda_3) \vee \lambda_2) \ast (\lambda_1 \wedge (\lambda_3 \vee \lambda_2))$. Now we have completeness, we need only check truth-falsity conditions to show that this definition works. \square

Lemma IV.2.15: If Λ is closed under \wedge, \vee and \ast ,

$$\exists \lambda \in \Lambda. \quad \phi \sqsubseteq_{\vdash} \lambda \text{ and } \lambda \sqsubseteq_{\vdash} \psi$$

iff the following all hold

- (1) $\exists \lambda_1 \in \Lambda. \quad \phi \vdash_1 \lambda_1, *$ and $\lambda_1 \vdash_2 \psi, *$
- (2) $\exists \lambda_2 \in \Lambda. \quad *, \psi \vdash_2 \lambda_2$ and $*, \lambda_2 \vdash_1 \phi$
- (3) $\exists \lambda_3 \in \Lambda. \quad \phi \vdash_1 \lambda_3, *$ and $*, \lambda_3 \vdash_1 \phi$

Proof: 'Only if' trivial again, and for 'if' $((\lambda_1 \wedge \lambda_3) \vee \lambda_2) \ast (\lambda_1 \wedge (\lambda_3 \vee \lambda_2))$ will again do. \square

We may now begin to apply these results. Firstly, a general persistence lemma for a theory \vdash , a formula ϕ of $L(\vdash)$, and a set Λ of formulae of $L(\vdash)$ which contains \top, \perp and identity formulae $x=y$, and is closed under all the connectives, under application of quantifiers and under substitution of parameters for parameters.

Lemma IV.2.16: ϕ is \vdash -equivalent to an element of Λ iff the following conditions hold for models M and N of \vdash :

$$(A): \text{ if } M \sqsubseteq_{\Lambda} N \text{ and } s \text{ is any assignment,}$$

$$M_s(\lambda) \sqsubseteq N_s(\lambda) \text{ for all } \lambda \in \Lambda \Rightarrow M_s(\phi) \sqsubseteq N_s(\phi)$$

$$(B): \text{ if } M \square_{\Lambda} N \text{ and } s \text{ is any assignment,}$$

$$M_s(\lambda) \square N_s(\lambda) \text{ for all } \lambda \in \Lambda \Rightarrow M_s(\phi) \square N_s(\phi)$$

Proof: Since $\phi \simeq_{\vdash} \lambda$ iff $\phi \sqsubseteq_{\vdash} \lambda \sqsubseteq_{\vdash} \phi$, this follows from the second combination lemma (IV.2.15) together with the contraposition of (IV.2.11), subject to the simplification guaranteed by (IV.2.12). For we may put $\Gamma = \{\phi\} = \top$ and $\Delta = \emptyset = \exists$, with $\vdash_1 = \vdash = \vdash_2$. Cases (1) and (2) yield (A), and case (3) yields (B). \square

Simplification IV.2.17: If Λ is closed under the substitution of Θ for parameters, then ϕ is \vdash -equivalent to an element of Λ iff the following conditions hold for models M and N of \vdash .

$$(A): M \sqsubseteq_{\Lambda} N \Rightarrow M_s(\phi) \sqsubseteq N_s(\phi) \text{ for any } s$$

$$(B): M \square_{\Lambda} N \Rightarrow M_s(\phi) \square N_s(\phi) \text{ for any } s$$

Proof: Simplification (IV.2.13).

As promised in Chapter II, we shall provide a model-theoretic criterion for when Θ is eliminable in a theory \vdash — that is to say, for when any ϕ in the language of \vdash is \vdash -equivalent to some Θ -free formula.

This involves a new model-theoretic relation $\sqsubseteq_{\mathbf{r}}$: $M \sqsubseteq_{\mathbf{r}} N$ means that N is more defined than M 'de re'. More precisely, $M \sqsubseteq_{\mathbf{r}} N$ iff M and N share

the same domain D , and for all $P \in \text{Prd}(L(\vdash))$, $f \in \text{Fnc}(L(\vdash))$ and $c \in \text{Cns}(L(\vdash))$:

$$\begin{aligned} P_M(\vec{d}) &\sqsubseteq P(\vec{d}) \text{ for all } \vec{d} \in D^m \\ f_M(\vec{d}) &\sqsubseteq f_N(\vec{d}) \text{ for all } \vec{d} \in D^n \\ c_M &\sqsubseteq c_N \end{aligned}$$

The theorem is

Theorem IV.2.18: Θ is eliminable in the theory \vdash iff, if M and N are non-empty models of \vdash , then

$$M \sqsubseteq_{\mathbf{r}} N \Rightarrow M \sqsubseteq N .$$

First we prove a lemma. Let ϕ be some formula of $L(\vdash)$ and let Λ be the set of Θ -free formulae of $L(\vdash)$, then:

Lemma IV.2.19: ϕ is \vdash -equivalent to an element of Λ iff, if M and N are non-empty models of \vdash such that $M \sqsubseteq_{\Lambda} N$, and s is any assignment, then

$$M_s(\lambda) \sqsubseteq N_s(\lambda) \text{ for all } \lambda \in \Lambda \Rightarrow M_s(\phi) \sqsubseteq N_s(\phi)$$

Proof: We shall show that this is really just a special case of (IV.2.16).

For, in the first place, if $D_M = D_N = \emptyset$, then conditions (A) and (B) of that lemma must invariably hold for our present choice of Λ . This is because there is only one assignment possible, viz. the s such that $s(x) = \Theta$ for all x , and so, if $M_s(\lambda) \sqsubseteq N_s(\lambda)$ for all $\lambda \in \Lambda$, then

$$M_s(\phi) = M_s(\phi') \sqsubseteq N_s(\phi') = N_s(\phi)$$

and if $M_s(\lambda) \sqsupset N_s(\lambda)$ for all $\lambda \in \Lambda$, then

$$M_s(\phi) = M_s(\phi') \sqsupset N_s(\phi') = N_s(\phi)$$

where ϕ' is obtained from ϕ by replacing any occurrence of Θ by a variable. While, secondly, if $M \sqsupset_{\Lambda} N$ and the domain is non-empty, then, by (IV.2.9) and (IV.2.10), $M \sqsupset N$, and so, again, condition (B) automatically holds. Our present lemma describes what is left over of (IV.2.16). \square

Proof of IV.2.18: From (IV.2.8) it is easy to check that the relation \sqsubseteq_r is none other than \sqsubseteq_Λ where Λ is the set of Θ -free formulae. So 'if' follows immediately from the preceding lemma, and, for 'only if', we may argue by contraposition in the following way. If our model-theoretic condition did not hold, there would be non-empty models M and N of \vdash such that $M \sqsubseteq_\Lambda N$ but $M \not\sqsubseteq N$. In that case $M_s(\phi) \not\sqsubseteq N_s(\phi)$ for some formula ϕ and some assignment s . But let ϕ' be obtained from ϕ by substituting Θ for all the parameters x of ϕ for which $s(x) = \Theta$, and let s' be as s except that it assigns some element of the domain of M and N to those variables which s assigns Θ . Then

$$M_{s'}(\phi') = M_s(\phi) \not\sqsubseteq N_s(\phi) = N_{s'}(\phi').$$

Since s' is *de re*, $M_{s'}(\lambda) \sqsubseteq N_{s'}(\lambda)$ for all $\lambda \in \Lambda$, and so, by the previous lemma, ϕ' is not equivalent to any $\lambda \in \Lambda$ — in other words, not equivalent to any Θ -free formula. □

Let us define explicit definability in a theory \vdash , for a predicate symbol P , a function symbol f , or a constant symbol c , by saying that, (respectively)

$$Px_1 \dots x_m \simeq_{\vdash} \phi$$

$$y = fx_1 \dots x_n \simeq_{\vdash} \phi$$

$$y = c \simeq_{\vdash} \phi$$

for some ϕ which does not contain P , f , or c , (respectively). Then simplification (IV.2.17) of (IV.2.16) provides a model theoretic criterion for explicit definability. Let us write $M \sqsubseteq_{(P)} N$ and $M \square_{(P)} N$, $M \sqsubseteq_{(f)} N$ and $M \square_{(f)} N$, and $M \sqsubseteq_{(c)} N$ and $M \square_{(c)} N$ to mean that M and N share the same domain, on which the relations of degree-of-definedness, respectively compatibility, hold (in the appropriate category) for the interpretation

of all items of non-logical vocabulary, except possibly P , f , or c , (respectively). Then

Theorem IV.2.19: Necessary and sufficient conditions for the explicit definability of P , f , and c , are, respectively, that all models M and N of \vdash

$$M \sqsubseteq_{(P)} N \Rightarrow M \sqsubseteq N$$

$$M \square_{(P)} N \Rightarrow M \square N$$

$$M \sqsubseteq_{(f)} N \Rightarrow M \sqsubseteq N$$

$$M \square_{(f)} N \Rightarrow M \square N$$

$$M \sqsubseteq_{(c)} N \Rightarrow M \sqsubseteq N$$

$$M \square_{(c)} N \Rightarrow M \square N$$

Proof: If Λ is (respectively) the set of P -free, f -free, or c -free formula, we may apply (IV.2.8) and (IV.2.17).

It is worth mentioning that, given a 'definition' of a function symbol f , a scheme of uniform elimination is available (cf. section II.3).

Theorem IV.2.20: If $y = fx_1 \dots x_n \simeq_{\vdash} \phi$, then, providing $fx_1 \dots x_n$ is substitutable for y in ψ , the following are \vdash -equivalent:

$$(1) \quad \psi(fx_1 \dots x_n / y)$$

$$(2) \quad \forall y (\phi \rightarrow \psi) \wedge (\exists y (\phi \wedge \psi) \vee \psi(\Theta / y))$$

$$(3) \quad \exists y (\phi \wedge \psi) \vee (\forall y (\phi \rightarrow \psi) \wedge \psi(\Theta / y))$$

And clearly we can do the same for a definable constant symbol — with no need of any proviso.

As in chapter III, explicit definability is not equivalent to the obvious notion of 'implicit definability' — and for exactly parallel reasons. However, as before, there is a messy notion of 'implicit definability' which is equivalent.

The Interpolation Theorem for logical consequence:

Theorem IV.2.21: If \vdash is logic (in some language L) and $\Gamma \vdash \Delta$, then there is a formula λ which contains no non-logical vocabulary not occurring in both Γ and Δ such that

$$\Gamma \vdash \lambda \quad \text{and} \quad \lambda \vdash \Delta$$

Proof: Let \vdash_1 be logic in the smallest language containing all of Γ as formulae, let \vdash_2 be logic in the smallest language containing all of Δ as formulae, and let $\Lambda = \text{Frm}(L_1) \cap \text{Frm}(L_2)$, where L_1 and L_2 are the languages just specified. It will be sufficient to show that there is a $\lambda \in \Lambda$ such that

$$\Gamma \vdash_1 \lambda \quad \text{and} \quad \lambda \vdash_2 \Delta$$

since then $\Gamma \vdash \lambda$ and $\lambda \vdash \Delta$. But say there is no such λ . Then, by the first combination lemma (IV.2.14), there are three alternatives to consider.

Case (1): There is no $\lambda \in \Lambda$ such that

$$\Gamma \vdash_1 \lambda, * \quad \text{and} \quad \lambda \vdash_2 \Delta, *$$

In this case, by (IV.2.11) (case (1)), and simplification (IV.2.12), there exist models M_1 for L_1 and M_2 for L_2 such that $M_1 \sqsubseteq_{\Lambda} M_2$ and, for some s ,

$$M_{1s}(\phi) = \top \quad \text{for all } \phi \in \Gamma$$

but

$$M_{1s}(\psi) \neq \top \quad \text{for all } \psi \in \Delta$$

But now let M be an expansion of M_2 to L which interprets the vocabulary which is in L_1 but not in L_2 in the same way that M_1 does. Then

$$M_s(\phi) = \top \quad \text{for all } \phi \in \Gamma \quad (\text{since } M_1 \sqsubseteq M|_{L_1})$$

$$M_s(\psi) \neq \top \quad \text{for all } \psi \in \Delta$$

This, however, contradicts the fact that $\Gamma \vdash \Delta$.

Case (2): There is no $\lambda \in \Lambda$ such that

$$*, \Gamma \vdash_1 \lambda \quad \text{and} \quad *, \lambda \vdash_2 \Delta.$$

Argument parallel to Case (1).

Case (3): There is no $\lambda \in \Lambda$ such that

$$\Gamma \vdash_1 \lambda, * \quad \text{and} \quad *, \lambda \vdash_2 \Delta$$

In this case, by (IV.2.11) (case (3)), and simplification (IV.2.12), there exist models M_1 for L_1 and M_2 for L_2 such that $M_1 \sqcap_{\Lambda} M_2$ and, for some s ,

$$M_{2_s}(\phi) = \top \quad \text{for all } \phi \in \Gamma$$

but

$$M_{2_s}(\psi) = \perp \quad \text{for all } \psi \in \Delta$$

But now let M be a model for L with the same domain D as M_1 and M_2 whose interpretations are defined

(i) on the vocabulary of Λ by putting

$$\begin{aligned} P_M(\vec{a}) = \top &\iff \text{either } P_{M_1}(\vec{a}) = \top \text{ or } P_{M_2}(\vec{a}) = \top \\ &= \perp \iff \text{either } P_{M_1}(\vec{a}) = \perp \text{ or } P_{M_2}(\vec{a}) = \perp \end{aligned}$$

if $d \in D$,

$$\begin{aligned} f_M(\vec{a}) = d &\iff \text{either } f_{M_1}(\vec{a}) = d \text{ or } f_{M_2}(\vec{a}) = d \\ & (= \emptyset \text{ if there is no such } d) \end{aligned}$$

if $d \in D$,

$$\begin{aligned} c_M = d &\iff \text{either } c_{M_1} = d \text{ or } c_{M_2} = d \\ & (= \emptyset \text{ if there is no such } d) \end{aligned}$$

(ii) on the remaining vocabulary of Γ in the same way as M_1

(iii) on the remaining vocabulary of Δ in the same way as M_2

(iv) on any other vocabulary *ad lib.*

Then

$$M_s(\phi) = \top \quad \text{for all } \phi \in \Gamma$$

$$M_s(\psi) = \perp \quad \text{for all } \psi \in \Delta$$

This, however, contradicts the fact that $\Gamma \vdash \Delta$.

All three possibilities are therefore ruled out. □

The proof of this theorem reveals that we could in fact state a *prima facie* stronger result: that λ follows from Γ and λ yields Δ in logic in the smallest possible languages. But, of course, relatively to a particular language is irrelevant for logical consequence, since if L' is an expansion of L , then logic in L' is a conservative extension of logic in L . This is because any model for L can be expanded to a model for L' . Cf. our remarks in II.2 about \simeq and \sqsubseteq .

There is interpolation for degree-of-definedness too.

Theorem IV.2.22: If $\phi \sqsubseteq \psi$, then there is a formula λ which contains no non-logical vocabulary not occurring in both ϕ and ψ such that

$$\phi \sqsubseteq \lambda \quad \text{and} \quad \lambda \sqsubseteq \psi$$

Proof: Analogous to the proof of (IV.2.21), this time using the second combination lemma (IV.2.15).

We end this section with some preliminaries for the compatibility theorem, which itself will be postponed until after the necessary 'diagram' techniques have been introduced. What we shall want to provide is a model-theoretic criterion for when a theory has joints — i.e. is such that for any two formulae ϕ and ψ which are compatible in the theory ($\phi \sqcap_{\vdash} \psi$) there exists a formula χ which is a 'joint' of ϕ and ψ . The definition of a joint may be given either in terms of sequents of the theory or model-theoretically: χ is a joint of ϕ and ψ in \vdash iff all the following hold:

$$\begin{array}{ll} \chi \vdash \phi, \psi, * & *, \psi, \phi \vdash \chi \\ \phi \vdash \chi, * & *, \chi \vdash \phi \\ \psi \vdash \chi, * & *, \chi \vdash \psi \end{array}$$

iff for any model M of \vdash and any $s: Var \rightarrow DU\{\Theta\}$:

for " - in other words " read " - and so
for any such M and N "

$$M_s(\chi) = \top \iff M_s(\phi) = \top \text{ or } M_s(\psi) = \top$$

$$M_s(\chi) = \perp \iff M_s(\phi) = \perp \text{ or } M_s(\psi) = \perp$$

As in Chapter III, we work *via* a simpler notion: let us say that χ covers ϕ and ψ in \vdash iff $\phi \sqsubseteq_{\vdash} \chi$ and $\psi \sqsubseteq_{\vdash} \chi$. What we now establish is

Lemma IV.2.23: The following are equivalent

- (1) ϕ and ψ have a joint in \vdash
- (2) ϕ and ψ can be covered in \vdash
- (3) for any models M and N of \vdash :

$$M \sqcap N \Rightarrow M_s(\phi) \sqcap N_s(\psi) \text{ for any } s.$$

Proof: (1) \Rightarrow (2) is immediate, and, to show that (2) \Rightarrow (1), observe that (just as in Chapter III) if χ covers ϕ and ψ in \vdash , then a joint for ϕ and ψ may be defined by

$$((\phi \wedge \chi) \vee (\psi \wedge \chi)) * ((\phi \vee \chi) \wedge (\psi \vee \chi))$$

That (2) \iff (3) follows from (IV.2.11), together with its two simplifications (IV.2.12) and (IV.2.13). This is because a formula λ covers ϕ and ψ in \vdash iff

$$\phi \vee \psi \vdash \lambda, * \text{ and } *, \lambda \vdash \phi \wedge \psi$$

Hence, if we put $\Lambda = \text{Frm}(L(\vdash))$, $\vdash_1 = \vdash_2 = \vdash$, $\Gamma = \{\phi \vee \psi\}$, $\top = \{\phi \wedge \psi\}$ and $\Delta = \emptyset = \Xi$, then, from case (3) of (IV.2.11), and (IV.2.10), we may deduce that such a formula exists iff, if M and N are models of \vdash and $M \sqcap N$, then for any s

$$\text{either } M_s(\phi \vee \psi) \neq \top \text{ or } N_s(\phi \wedge \psi) \neq \perp$$

— in other words: $M_s(\phi) \sqcap N_s(\psi)$. □

We could continue turning the handle to churn out more facts following from the Interpolant Exclusion Lemma (IV.2.1) — with or without our various simplifying assumptions. But we shall not. It would, in fact, be

equally interesting to look back to the construction in the proof of (IV.2.1) and consider ways of modifying it. For example, if we dropped the assumption that Λ was closed under the application of quantifiers, then we could proceed with a simpler construction which ignored witnesses, and yet, given completeness, we could still obtain pairs of interpolant-excluding models. The relations between these models would be more complicated, but (assuming that Λ contained $x=y$ and were closed under negation) they could be characterized in terms of 'overlapping' domains. In this way we could, for example, provide a model-theoretic criterion for quantifier elimination.

However we now turn to sequents. I am afraid that the model theory will continue to be somewhat abstract. We should, perhaps, rather be considering particular examples of theories and models, or, at least, more specific kinds of theory (for example equational theories, which would seem to be amenable to an interesting treatment in our framework). However, the abstract investigation is not without a point. In the first place, we are providing ways of finding out things about the logic — things which need not themselves be characterized semantically. But, in any case, even those results which are explicitly model-theoretic have an importance, since they show (I hope) that the logic makes good sense — that the way theories and their models fit together is not outrageously peculiar. This is crucial, since our semantics — like classical semantics, but, perhaps, unlike certain more exotic model-theoretic enterprises — can be seen as providing quite direct and literal interpretations for the simple formal languages under consideration.

IV.3 THEORIES AND SEQUENTS

Most of the basic results in this section will be instances of the following theorem schema:

Schema (1): $M \in K(\vdash \cap \Sigma) \iff MR_{\Sigma} N$ for some $N \in K(\vdash)$

where Σ is a specified class of sequents, and R_{Σ} is an associated relation between models. From Schema (1) we can deduce corresponding instances of

Schema (2): \vdash is axiomatizable by sequents in Σ iff

$$MR_{\Sigma} N \text{ and } N \in K(\vdash) \implies M \in K(\vdash)$$

Theorem IV.3.1: If Σ and R_{Σ} satisfy Schema (1), then they satisfy Schema (2).

Proof: It is sufficient for (2) that $\vdash = \overline{\vdash \cap \Sigma}$ iff the model-theoretic condition holds. But, given (1), this is easy to check.

We shall be making use of a richer stock of relations between models than we have hitherto. Firstly, recall the purely elementary relations \sim_e and \sqsubseteq_e (whose inverse we shall write ' \supseteq_e '), which we specified in II.2. If we have a model M for L , let us define

$$\top(M) = \{\sigma \in Snt(L) \mid M(\sigma) = \top\}$$

$$\perp(M) = \{\sigma \in Snt(L) \mid M(\sigma) = \perp\}$$

$$\not\top(M) = \{\sigma \in Snt(L) \mid M(\sigma) \neq \top\}$$

$$\not\perp(M) = \{\sigma \in Snt(L) \mid M(\sigma) \neq \perp\}$$

Then the following equivalences are easy to check.

Lemma IV.3.2:

$$(1) \quad M \sim_e N \iff \top(M) \not\equiv_N \not\top(M), * \iff *, \not\perp(M) \not\equiv_N \perp(M)$$

$$M \sqsubseteq_e N \iff \top(M) \not\equiv_N * \iff * \not\equiv_N \perp(M)$$

$$M \supseteq_e N \iff \not\top(M) \not\equiv_N \not\top(M) \iff \not\perp(M) \not\equiv_N \perp(M)$$

Now, consider models M and N , for a particular language L , a monotonic function $\theta: D_M \cup \{\Theta\} \rightarrow D_N \cup \{\Theta\}$, and a set Λ of formulae of L , and define,

$$\begin{array}{l}
 (1) \quad M \overset{\theta}{\sim}_{\Lambda} N \\
 (2) \quad M \sqsubseteq_{\Lambda} N \\
 (3) \quad M \supseteq_{\Lambda} N
 \end{array}
 \left. \vphantom{\begin{array}{l} (1) \\ (2) \\ (3) \end{array}} \right\} \iff \begin{array}{l} \text{for all } s: \text{Var} \rightarrow D_M \\ \text{and all } \lambda \in \Lambda \end{array} \left\{ \begin{array}{l} M_s(\lambda) = N_{\theta \circ s}(\lambda) \\ M_s(\lambda) \sqsubseteq N_{\theta \circ s}(\lambda) \\ M_s(\lambda) \supseteq N_{\theta \circ s}(\lambda) \end{array} \right.$$

(Observe that if Λ is closed under substitution of Θ for parameters, then the restriction to *de re* assignments is otiose.) If we vary Λ let us say that we have different 'forms' for the function θ , while (1) - (3) are the different 'modes' for θ and Λ . (There is, of course, a fourth mode, defined in terms of \square , but we shall not be using that here.)

All the forms that we consider will be such that there is no restriction on the non-logical vocabulary of Λ , and we shall give three particular forms a special name:

- (i) simple: if Λ contains the quantifier free formulae of L .
- (ii) elementary: if Λ contains all formulae of L
- (iii) Θ -variant: if Λ contains the Θ -free formulae of L

We shall mostly be interested in embeddings — i.e. θ 's which are one-one. Observe that in modes (1) and (2) θ is necessarily an embedding, if Λ contains identity and distinctness formulae, as it does in the case of our three special forms. In mode (3), however, non-one-one θ 's arise more easily: the possibly conflicting behaviour of two elements of D_M may be resolved into indeterminacy with respect to a common image in D_N . Observe also, concerning embeddings of mode (2), that we may have a term t of L such that $M_s(t) = \Theta$, for some s , while $N_{\theta \circ s}(t)$ is in D_N and yet not in $\theta(D_M)$.

I leave it as an 'exercise' to formulate equivalent definitions, in explicitly structural terms, for the various modes of simple embedding, and

then to add further conditions definitive of elementary embeddings. We may take the same route in a definition of Θ -variant embeddings also.

When no specific qualification is made as to mode, 'embedding' shall mean embedding of mode (1). For the existence of some embedding or other, we use the following notation:

$$\begin{array}{l}
 (1) \quad M \xrightarrow{\Lambda} N \\
 (2) \quad M \xrightarrow[\rightarrow \Lambda]{\sqsubseteq} N \\
 (3) \quad M \xrightarrow[\rightarrow \Lambda]{\supseteq} N
 \end{array}
 \left. \vphantom{\begin{array}{l} (1) \\ (2) \\ (3) \end{array}} \right\} \iff \text{there is an embedding } \theta \text{ such that}
 \left\{ \begin{array}{l} M \xrightarrow[\Lambda]{\theta} N \\ M \xrightarrow[\Lambda]{\sqsubseteq} N \\ M \xrightarrow[\Lambda]{\supseteq} N \end{array} \right.$$

But, in place of ' Λ ', we put 'o', 'e' and 'r' for the forms simple, elementary, and Θ -variant, respectively.

If θ is the identity function, we have kinds of substructure, or, the other way about, extension

To prove instances of Schema (1), we use strong completeness and 'diagrams'. Given a model M for L , let M^+ be the expansion of M to the language $L(\underline{D}_M)$, where \underline{D}_M is a set of distinct new constants \underline{d} for elements d of D_M . $\top(M^+)$, $\perp(M^+)$, $\mathcal{X}(M^+)$ and $\mathcal{X}'(M^+)$ are defined as above — in terms of sentences of the expanded language, of course — and if $\Lambda \subseteq \text{Frm}(L)$ let $\Lambda[\underline{D}_M]$ be the following set of sentences of $L(\underline{D}_M)$:

$$\{ \lambda(\underline{d}_1, \dots, \underline{d}_n) \mid \lambda(x_1, \dots, x_n) \in \Lambda \text{ and } d_1, \dots, d_n \in D_M \}$$

If we assume that Λ is closed under negation, then the following lemma is easy, but very tedious, to prove.

Lemma IV.3.3 (Diagram Lemma): If Γ and Δ are sets of sentences of $L(\underline{D}_M)$ and N is inconsistent with $\langle \Gamma, \{*\} \cup \Delta \rangle$, then, if the function $\theta: D_M \cup \{\Theta\} \rightarrow D_N \cup \{\Theta\}$ is given by $\theta: d \mapsto N(\underline{d})$ for $d \in D_M$ and $\theta: \Theta \mapsto \Theta (= N(\Theta))$, the following hold:

- (1) If $\Gamma(M^+) \cap \Lambda[\underline{D}_M] \subseteq \Gamma$ and $\mathcal{X}(M^+) \cap \Lambda[\underline{D}_M] \subseteq \Delta$, then $M \overset{\theta}{\sim}_{\Lambda} N|L$
- (2) If $\Gamma(M^+) \cap \Lambda[\underline{D}_M] \subseteq \Gamma$, then $M \overset{\theta}{\sqsubseteq}_{\Lambda} N|L$
- (3) If $\mathcal{X}(M^+) \cap \Lambda[\underline{D}_M] \subseteq \Delta$, then $M \overset{\theta}{\supseteq}_{\Lambda} N|L$

Also, if either $\underline{d \neq e} \in \Gamma$ for all distinct d and e in D_M , or $\underline{d = d} \in \Gamma$ for all d in D_M and $\underline{d = e} \in \Delta$ for all distinct d and e in D_M , then θ is one-one.

Corollary IV.3.4: If N is inconsistent with $\langle \Gamma \cup \{*\}, \Delta \rangle$ then (1) - (3) hold with 'T' and 'L' and 'Γ' and 'Δ' switched. Also, if either $\underline{d = e} \in \Delta$ for all distinct d and e in D_M , or $\underline{d \neq d} \in \Delta$ for all d in D_M and $\underline{d \neq e} \in \Gamma$ for all distinct d and e in D_M , then θ is one-one.

The first lemma is our version of the 'Embedding Theorem'.

Lemma IV.3.5: M is consistent with the quantifier-free sequents of \vdash iff M can be simply embedded in some model of \vdash .

Proof: 'If' is easy: observe that if $M \rightarrow_o N$, by the embedding θ , then, for any s , $(N, \theta \circ s)$ will be inconsistent with any quantifier-free sequent that (M, s) is inconsistent with.

'Only if': assume that M is consistent with the quantifier-free sequents of \vdash , and recall that $\vdash^{\underline{D}_M}$ is the theory in $L(\vdash)$ (\underline{D}_M) axiomatized by \vdash . It will be sufficient to show that

$$\Gamma \not\vdash^{\underline{D}_M} *, \Delta$$

where Γ contains precisely the quantifier-free sentences of $\Gamma(M^+)$ and Δ the quantifier-free sentences of $\mathcal{X}(M^+)$. For then, by strong completeness, there will be a model N of $\vdash^{\underline{D}_M}$ which is inconsistent with $\langle \Gamma, \{*\} \cup \Delta \rangle$, and hence, by the diagram lemma, $M \rightarrow_o N|L$. While clearly $N|L \in K(\vdash)$.

However, say that $\Gamma \vdash^{\underline{D}_M} \Delta, *$, then for some finite $\Gamma_0(\vec{d})$ and $\Delta_0(\vec{d})$,

subsets of Γ and Δ ,

$$\Gamma_0(\vec{d}) \vdash_{D_M} \Delta_0(\vec{d}), *$$

Hence, by lemma (IV.2.3), replacing distinct d_i by distinct x_i ,

$$\Gamma_0(\vec{x}) \succ \Delta_0(\vec{x}), * \in \vdash$$

This is a quantifier-free sequent; yet, by definition of Γ and Δ , M must be inconsistent with it. This is a contradiction, and so $\Gamma \not\vdash_{D_M} \Delta$, as required. □

Theorem IV.3.6: A theory can be axiomatized by quantifier-free sequents iff the class of its models is closed under simple substructures.

Proof: (IV.3.1). Observe that there will always be a substructure in the isomorphism class of an embedded model — whatever the form and whatever the mode.

Another example: let a sequent be called '*de re*' iff it is of the form $\Gamma \succ_{\{x_i=x_i\}} \Delta$, in other words

$$\Gamma, \{x_i=x_i\}_{1 \leq i \leq n} \succ \{x_i \neq x_i\}_{1 \leq i \leq n}, \Delta,$$

where all the parameters of Γ and Δ are among the x_i , and Θ occurs nowhere in Γ or Δ .

Lemma IV.3.7: M is consistent with the *de re* sequents of \vdash iff there is a Θ -variant embedding of M into some model of \vdash .

Proof: 'If' is trivial again, and for 'only if' the proof follows the same pattern as before. This time it will be sufficient to show that, assuming M is consistent with the Θ -free sequents of \vdash , $\Gamma \not\vdash_{D_M} \Delta, *$, where Γ contains the Θ -free elements of $\Gamma(M^+)$ and Δ the Θ -free elements of $\Delta(M^+)$.

But if not, then for some $d_1, \dots, d_n \in D_M$

$$\Gamma_0(\underline{d}_1, \dots, \underline{d}_n) \stackrel{D}{\vdash} \Delta_0(\underline{d}_1, \dots, \underline{d}_n), *$$

where these sets are finite subsets of Γ and Δ . Hence

$$\{\underline{d}_i = \underline{d}_i\}_{1 \leq i \leq n}, \Gamma_0(\underline{d}_1, \dots, \underline{d}_n) \stackrel{D}{\vdash} \Delta_0(\underline{d}_1, \dots, \underline{d}_n), \{\underline{d}_i \neq \underline{d}_i\}_{1 \leq i \leq n}, *$$

and so, by Lemma (IV.2.3), for some distinct variables x_i ,

$$\Gamma_0(\vec{x}) \stackrel{\{x_i = x_i\}}{\succ} \Delta_0(\vec{x}), * \in \vdash$$

This sequent is *de re*, yet M is inconsistent with it; and so we have a contradiction. □

Theorem IV.3.8: A theory can be axiomatized by *de re* sequents iff the class of its models is closed under Θ -variant substructures.

Observe that our definition of 'de re sequent' is stronger than simply ' Θ -free'. Indeed, any theory can be axiomatized by Θ -free sequents, since a model is consistent with $\Gamma(\Theta/x) \succ \Delta(\Theta/x)$ iff it is consistent with both $\Gamma \succ \Delta, x=x$ and $x \neq x, \Gamma \succ \Delta$.

We appeal simply to (IV.3.2) for the next result. It raises an interesting question.

Lemma IV.3.9: M is consistent with the parameter-free sequents of \vdash iff M is elementarily equivalent (\sim_e) to some model of \vdash .

Proof: 'If' is trivial. For 'only if' we need only show that, assuming M is consistent with the parameter-free sequents of \vdash , $\Gamma(M) \not\vdash \mathcal{A}(M), *$, since then, by strong completeness and (IV.3.2), there will be a model N of \vdash such that $M \sim_e N$. But, if $\Gamma(M) \vdash \mathcal{A}(M), *$, then there would be finite subsets Γ and Δ of $\Gamma(M)$ and $\mathcal{A}(M)$ such that $\Gamma \succ \Delta, * \in \vdash$: a contradiction. □

Theorem IV.3.10: A theory can be axiomatized by parameter-free sequents iff the class of its models is closed under elementary equivalence (\sim_e).

But: is not the class of models of any theory closed under \sim_e ?
 No. We shall see some counterexamples in Section 4. So not all theories are axiomatizable by parameter-free sequents. The relation of 'elementary' equivalence under which any $K(\vdash)$ is closed is the strictly stronger one of being consistent with the same sequents. If we had a consequence-encoding conditional, then, given the quantifier principles, this distinction between kinds of elementary equivalence would not arise, but we do not have such a thing.

However, the following is true:

Theorem IV.3.11: The class of models of any theory is closed under elementary substructures (equivalently, elementary embeddings (\rightarrow_e)).

Proof: Observe that if $M \rightarrow_e N$ under the embedding θ , then, if (M, s) is inconsistent with a sequent, so is $(N, \theta \circ s)$. \square

So far we have restricted ourselves to embeddings of mode (1) and elementary equivalence. Clearly there are further analogous results to be obtained in this mode by combining in different ways the various features which have figured. But we now turn to modes (1) and (2). It is interesting to see that in these modes distinct combinations of features collapse to give equivalent conditions on a theory.

Let us revive the terminology of Chapter III and call sequents of the form $\succ \Gamma$ posits and sequents of the form $* \succ \Gamma$ contraposits.

Lemma IV.3.12: The following are equivalent:

- (1) (a) M is consistent with the contraposits of \vdash
- (1) (b) M is consistent with the parameter-free contraposits of \vdash
- (2) (a) $M \sqsubseteq_{\rightarrow_e} N$ for some model N of \vdash
- (2) (b) $M \sqsubseteq_{\leftarrow_e} N$ for some model N of \vdash

Proof: That (1a) \Rightarrow (1b) is immediate. To see that (1b) \Rightarrow (1a), observe that a model is consistent with the sequent $* \succ \Gamma(x_1, \dots, x_n)$ iff it is consistent with $* \succ \Pi x_1 \dots \Pi x_n \wedge \Gamma(x_1, \dots, x_n)$, where $\Pi x\phi$ is an abbreviation for $\forall x\phi \wedge \phi(\Theta/x)$. Note that this quantifier ' Πx ' commutes with itself.

That (2a) \Rightarrow (1a) and (2b) \Rightarrow (1b) is straight-forward. To show the converses, we use the now familiar technique appealing to strong completeness: for (1a) \Rightarrow (2a) we invoke corollary (IV.3.4) of the diagram lemma with the pair $\langle \{*\}, \perp(M^+) \rangle$; for (1b) \Rightarrow (2b) we invoke lemma (IV.3.2). \square

Theorem IV.3.13: The following are equivalent:

- (1) (a) \vdash is axiomatizable by contraposits
- (b) \vdash is axiomatizable by parameter-free contraposits
- (2) (a) If $M \sqsubseteq_{\exists e} N$ and $N \in K(\vdash)$, then $M \in K(\vdash)$
- (b) If $M \sqsubseteq_e N$ and $N \in K(\vdash)$, then $M \in K(\vdash)$

For the case of posits we can add an extra condition:

Lemma IV.3.14: The following are equivalent:

- (1) (a) M is consistent with the posits of \vdash
- (b) M is consistent with the parameter-free posits of \vdash
- (2) (a) $M \sqsupset_{\exists e} N$ for some model N of \vdash
- (a') $M \sqsupset_e^\theta N$ for some model N of \vdash , for some θ (not necessary one-one)
- (b) $M \sqsupset_e N$ for some model N of \vdash

Proof: That (1a) \Rightarrow (1b) is immediate, while (1b) \Rightarrow (1a) because a model is consistent with the sequent $\succ \Gamma(\vec{x})$ iff it is consistent with $\succ \Gamma(\vec{\Theta})$.

(2a) \Rightarrow (1a), (2a') \Rightarrow (1a) and (2b) \Rightarrow (1b) are all straight-forward.

To show that (1a) \Rightarrow (2a) and (1a) \Rightarrow (2a') we use the diagram lemma (IV.3.3) and strong completeness with the pairs $\langle \{ \underline{d} = \underline{d} \mid d \in D_M \}, \mathcal{X}(M^+) \cup \{*\} \rangle$ and

$\langle \emptyset, \mathcal{X}(M^+) \cup \{*\} \rangle$, respectively; and to show that (1b) \Rightarrow (2b) we use lemma (IV.3.2).

There is a subtlety in the proof of (1a) \Rightarrow (2a), so we shall run through it. Assuming M is consistent with the posits of \vdash , we have to show that

$$\{ \underline{d} = \underline{d} \}_{d \in D_M} \not\vdash^{\underline{D}_M} \mathcal{X}(M^+), *$$

But if this were not the case, then

$$\{ \underline{d}_i = \underline{d}_i \}_{1 \leq i \leq n} \vdash^{\underline{D}_M} \Delta_0(\underline{d}_1, \dots, \underline{d}_n, \underline{e}_1, \dots, \underline{e}_m), *$$

for some finite $\Delta_0(\underline{d}_1, \dots, \underline{d}_n, \underline{e}_1, \dots, \underline{e}_m)$ which is a subset of $\mathcal{X}(M^+)$.

Hence, replacing constants by distinct variables,

$$\{ x_i = x_i \}_{1 \leq i \leq n} \succ \Delta_0(x_1, \dots, x_n, y_1, \dots, y_m), * \in \vdash$$

And so, by the rule (\forall) (and (Ext)),

$$\succ \forall x_1 \dots \forall x_n \ M \Delta_0(x_1, \dots, x_n, y_1, \dots, y_m), * \in \vdash$$

Now we can argue to a contradiction. □

Theorem IV.3.15: The following are equivalent:

- (1) (a) \vdash can be axiomatized by posits
- (b) \vdash can be axiomatized by parameter-free posits
- (2) (a) If $M \sqsupseteq_e N$ and $N \in K(\vdash)$, then $M \in K(\vdash)$
- (a') If $M \sqsupseteq_e^\theta N$, for some θ , and $N \in K(\vdash)$, then $M \in K(\vdash)$
- (b) If $M \sqsupseteq_e N$ and $N \in K(\vdash)$, then $M \in K(\vdash)$

Also, given (IV.3.12) and (IV.3.14), we could easily deduce the analogues of (III.4.2) part three.

Sequent-persistence results along the lines of those in III.4 are available also; however we need to be careful about the notion of

'equivalence' for sequents. We must distinguish between two kinds of \vdash -consequence, and hence between two kinds of \vdash -equivalence, which we define as bi-consequence. The senses in which α can be a consequence, in \vdash , of β are

$$(i) \quad \alpha \in \overline{\vdash \cup \{\beta\}}$$

$$(ii) \quad \alpha' \in \overline{\vdash \cup \{\beta'\}}, \text{ where } \alpha' \text{ and } \beta' \text{ are obtained from } \alpha \text{ and } \beta \text{ by replacing parameters, uniformly, by distinct new constants.}$$

The corresponding semantic definitions are:

$$(i) \quad M \text{ is consistent with } \beta \Rightarrow M \text{ is consistent with } \alpha, \\ \text{for all } M \in K(\vdash)$$

$$(ii) \quad (M, s) \text{ is consistent with } \beta \Rightarrow (M, s) \text{ is consistent with } \alpha, \\ \text{for all } M \in K(\vdash) \text{ and all } s$$

It is equivalence defined as bi-consequence in sense (ii) for which the persistence results follow the same pattern as in Chapter III.

Observe that consequence in this stronger sense can be 'axiomatized'. That is to say given any α and β , there is a set Σ of sequents such that α is a consequence of β in a theory \vdash iff $\Sigma \subseteq \vdash$. If $\alpha = \Gamma \succ \Delta$ and $\beta = \Xi \succ \Upsilon$, then we may take Σ to contain the sequents

$$\begin{array}{ll} \Xi \succ \Gamma, \Upsilon & \Xi, \Delta \succ \Upsilon \\ \Xi, \neg \Gamma \succ \Upsilon & \Xi \succ \neg \Delta, \Upsilon \end{array}$$

In contrast, consequence in sense (i) cannot in general be axiomatized. This is another penalty — if it is to be considered such — that we pay for having no conditional connective to encode consequence between formulae. For consider the simple rule

$$\frac{Px \succ Qx, *}{\succ}$$

We may note that this rule holds in a theory \vdash iff $\{Px \succ Qx, *\} \cup \vdash$ is inconsistent. If it could be axiomatized, then there would be a set Σ such that

$$M \in K(\Sigma) \iff Px \not\equiv_M Qx, *$$

for all models $M \in K(\vdash)$, where \vdash is any theory. In particular, if \vdash is logic, this condition would have to hold for any model M for the language of P and Q . However, we shall show at the end of the next section that there can be no such Σ .

Finally, we prove the Compatibility Theorem. Recall that a theory 'has joints' iff any two formulae compatible in the theory have a joint in the theory, as this was defined at the end of Section 2. (Of course, conversely, a pair of formulae with a joint must inevitably be compatible.)

Theorem IV.3.16: A theory has joints iff any two compatible models of the theory are elementarily weaker substructures of some model of the theory.

(In terms of embeddings, this model-theoretic condition on a theory \vdash is: if $M, N \in K(\vdash)$ and $M \sqsubset N$, then there exists a $P \in K(\vdash)$ and an embedding θ such that $M \sqsubseteq_{\theta}^e P$ and $N \sqsubseteq_{\theta}^e P$.)

Proof: By (IV.2.23), we know that a theory \vdash fails to have joints iff there exist ϕ and ψ such that

$$(1) \quad \phi \sqsubset_{\vdash} \psi$$

$$(2) \quad M_s(\phi) \not\sqsupset N_s(\psi) \text{ for some } M \text{ and } N \text{ in } M(\vdash) \text{ such that } M \sqsubset N, \text{ and some assignment } s.$$

Firstly, (1) and (2) imply that our model-theoretic condition cannot hold. This is because, taking ϕ, ψ, M, N and s as specified, if there were a model P of \vdash such that $M \sqsubseteq_{\theta}^e P$ and $N \sqsubseteq_{\theta}^e P$ for some embedding θ , then, by (2), $P_{\theta \circ s}(\phi) \not\sqsupset P_{\theta \circ s}(\psi)$, which would contradict (1). Hence 'if' is proved by contraposition.

Conversely, by contraposition again, if the model theoretic condition failed to hold, then, by the diagram lemma (IV.3.3) and strong completeness, we would have models M and N of \vdash with a common domain, D say, such

that $M \sqsubset N$ but

$$\mathbb{T}(M^+) , \mathbb{T}(N^+) \stackrel{D}{\not\vdash} *$$

(where $\mathbb{T}(M^+)$ and $\mathbb{T}(N^+)$ contain the same constants \underline{d} for $d \in D$). Hence, since these sets of sentences are closed under conjunction,

$$\chi(\underline{d}) , \omega(\underline{d}) \stackrel{D}{\not\vdash} *$$

for some $\chi(\vec{x})$ and $\omega(\vec{x})$ such that $M_{\vec{d}}^{\rightarrow}(\chi(\vec{x})) = N_{\vec{d}}^{\rightarrow}(\omega(\vec{x})) = \top$. So, by lemma (IV.2.3),

$$\chi(\vec{x}) , \omega(\vec{x}) \not\vdash *$$

But then, if we put $\phi = \chi(\vec{x}) \ast \top$ and $\psi = \neg\omega(\vec{x}) \ast \perp$,

$$(i) \quad \phi \sqsubset \not\vdash \psi$$

$$(ii) \quad M_{\vec{d}}^{\rightarrow}(\phi) = \top \text{ and } N_{\vec{d}}^{\rightarrow}(\psi) = \perp$$

Therefore (1) and (2) hold, and so $\not\vdash$ does not have joints. \square

Hence we may deduce the following interesting and important corollary

Theorem IV.3.17: ϕ and ψ are logically compatible iff they have a joint in logic.

Proof: 'If': trivial. 'Only if': if $M \sqsubset N$, then clearly we can specify a model P such that $M \sqsubseteq P$ and $N \sqsubseteq P$. This is more than sufficient to show that logic has joints. \square

Joints in logic will, of course, be joints in any theory.

for "modles" read "models"

IV.4 COMPLETE THEORIES

This section contains many questions, but regrettably few answers.

Let us say that a theory \vdash is sentence complete iff for all $\sigma \in \text{Snt}(L(\vdash))$ $\vdash \sigma$ or $\sigma \vdash *$. Clearly this is equivalent to saying that for all such σ either $\succ \sigma \in \vdash$ or $\{\succ \sigma\} \cup \vdash$ is inconsistent (or, indeed, that either $\sigma \succ * \in \vdash$ or $\{\sigma \succ *\} \cup \vdash$ is inconsistent). On the other hand, let us say that \vdash is sequent complete iff for any sequent α either $\alpha \in \vdash$ or $\{\alpha\} \cup \vdash$ is inconsistent. Hence, assuming that \vdash is consistent, sequent completeness is equivalent to maximal consistency. In this section 'theory' will always mean consistent theory, and so we may think directly in terms of maximal consistency.

We saw at the end of III.1 that for propositional logic these two kinds of completeness coincided – and so they would if we were to set up classical total-valued logic in the style adopted in this chapter. However, for our present kind of theory sentence completeness is a strictly weaker notion than maximal consistency. To show this it is sufficient to exhibit a pair of models M and N which are elementarily equivalent – in the sense \sim_e , which in this section we label e-equivalence – but whose theories \vdash_M and \vdash_N are such that $\vdash_M \subseteq \vdash_N$ but $\vdash_N \not\subseteq \vdash_M$. Then \vdash_M will be sentence complete but, because \vdash_N is a larger consistent theory, not maximally consistent.

Exhibiting two such models will show that the relation \leq , defined by

$$M \leq N \iff \vdash_M \subseteq \vdash_N$$

(i.e. N is consistent with all the sequents that M is consistent with) is equivalent neither to the relation \sim_e nor to the relation \sim_E , defined by

$$M \sim_E N \iff M \leq N \text{ and } N \leq M$$

We shall call \sim_E E-equivalence. Clearly, $M \sim_E N$ implies $M \leq N$ and $M \leq N$ implies $M \sim_e N$, but we shall have counterexamples to the converses.

A given e-class (an equivalence class of models modulo \sim_e) will then be partitioned into smaller E-classes (equivalence classes of models modulo \sim_E), and we shall want to know how this goes — how the relation \leq behaves on a given e-class. By the Löwenheim-Skolem Theorem (IV.1.13), countable theories are determined by their countable models, so we shall restrict attention to countable languages and keep models small. Henceforth by 'e-class' and 'E-class' we shall mean equivalence classes restricted to countable models.

Before we turn to a simple example of an e-class, note the following reformulation of (IV.3.11).

Lemma IV.4.1: $M \rightarrow_e N \Rightarrow N \leq M$

Consider models $M_{\alpha\beta}$ $\alpha, \beta \leq \omega$, for a language L containing two monadic predicates P and Q , which are specified as follows:

$$D_{M_{\alpha\beta}} = U_\alpha \cup V_\beta \cup X \cup Y$$

where

$$U_\alpha = \{a_i\}_{i < \alpha} \quad V_\beta = \{b_i\}_{i < \beta} \quad X = \{c_i\}_{i < \omega} \quad Y = \{d_i\}_{i < \omega}$$

and

$$P_{M_{\alpha\beta}}(e) = \top \iff e \in V_\beta \cup X \quad Q_{M_{\alpha\beta}}(e) = \top \iff e \in U_\alpha \cup X$$

$$= * \text{ otherwise} \quad \quad \quad = * \text{ otherwise}$$

All these models are models of the theory \vdash axiomatized by

- (i) $* \vDash Px$ (equivalently, $* \vDash \forall x Px \wedge P\Theta$)
- (ii) $* \vDash Qx$ (equivalently, $* \vDash \forall x Qx \wedge Q\Theta$)
- (iii) $\vDash \exists x_1 \dots \exists x_n \left(\bigwedge_{1 \leq i < j \leq n} x_i \neq x_j \wedge \bigwedge_{1 \leq k \leq n} (Px_k \wedge Qx_k) \right)$, for all n

$$(iv) \quad * \succ \exists x_1 \dots \exists x_n \left(\bigwedge_{1 \leq i < j \leq n} x_i \neq x_j \wedge \bigwedge_{1 \leq k \leq n} (\neg Px_k \wedge \neg Qx_k) \right), \text{ for all } n$$

Observe that (iii) means that there are infinitely many things of which both P and Q are true, and (iv) means that there are infinitely many things of which neither P nor Q is true.

Now, all the models $M_{\alpha\beta}$ are models of \vdash , and any countable model of \vdash , will be isomorphic to some $M_{\alpha\beta}$. These facts are easy to check. Furthermore

Lemma IV.4.2: $M_{\alpha\beta} \rightarrow_e M_{\alpha'\beta'}$ if $\alpha \leq \alpha'$ and $\beta \leq \beta'$ (where, according to our specification of the models, the embedding may be taken to be the identity embedding). (Proof omitted).

From this we may deduce, firstly, that the $M_{\alpha\beta}$ are all e-equivalent – and so, since \vdash is determined by its countable models, \vdash is sentence complete – and, secondly, that $M_{\alpha\beta} \leq M_{\alpha'\beta'}$ if $\alpha \geq \alpha'$ and $\beta \geq \beta'$.

None of the $M_{\alpha\beta}$, however, are E-equivalent. Consider the following axioms, which discriminate between them:

$$\zeta_0 : Qx \succ Px, *$$

$$\eta_0 : Px \succ Qx, *$$

.
.
.
.
.

$$\zeta_n : \{x_i \neq x_j\}_{1 \leq i < j \leq n}, Px_1, \dots, Px_n \succ Qx_1, \dots, Qx_n, *$$

$$\eta_n : \{x_i \neq x_j\}_{1 \leq i < j \leq n}, Qx_1, \dots, Qx_n \succ Px_1, \dots, Px_n, *$$

Clearly $M_{\alpha\beta}$ satisfies ζ_i iff $\alpha \leq i$, and satisfies η_j iff $\beta \leq j$. Hence we have a great array of models providing examples to show that our two notions of completeness do not coincide.

The class $K_e(\vdash)$ is indeed an e-class. It contains the $M_{\alpha\beta}$, and any countable model e-equivalent to some $M_{\alpha\beta}$, since any such model is a model of \vdash , by (IV.3.10), (and, hence, in fact isomorphic to some $M_{\alpha\beta}$). And so $\vdash (= \overline{K_e(\vdash)})$ is the smallest sentence complete theory settling sentences in the way the $M_{\alpha\beta}$ determine.

Let us say that a model M is characteristic for a theory iff that theory is precisely \overline{M} , then it is easy to check that $M_{\alpha\beta}$ is characteristic for

$$\overline{\vdash \cup \{ \zeta_i \}_{\alpha \leq i < \omega} \cup \{ \eta_j \}_{\beta \leq j < \omega}}$$

In particular $M_{\omega\omega}$ is characteristic for \vdash . At the other extreme, M_{00} is characteristic for $\overline{\vdash \cup \{ \zeta_i \}_{i < \omega} \cup \{ \eta_j \}_{j < \omega}}$. This theory is maximally consistent, since it is \aleph_0 -categorical. It is, furthermore, the only maximally consistent extension of \vdash .

The e-class we have been considering is obviously a very simple one.

Question: Just how typical is it of e-classes in general?

First of all we consider small sentence complete theories. Given a model M for L , let $\Theta(M)$ be the theory axiomatized by

$$\{ \succ \sigma \mid \sigma \in T(M) \} \cup \{ * \succ \sigma \mid \sigma \in \mathcal{A}(M) \}$$

We call $\Theta(M)$ the e-theory of M , in contrast to \overline{M} , which we can now label the E-theory of M .

Lemma IV.4.3: $N \sim_e M \iff N \in K(\Theta(M))$

Clearly, if M is in the e-class K , then $\Theta(M) = \overline{K}$, and this is the smallest theory settling sentences the way M - equivalently K - determines.

Let us call such theories minimally sentence complete. Observe that in our example $\vdash = \overline{M_{\omega\omega}} = \Theta(M_{\omega\omega}) = \Theta(M_{\alpha\beta})$ for any α and β . Hence $M_{\omega\omega}$ is

l.3 : for "any" read "wmy"
l.10 : for "any" read "wmy"

line 3 up : "I" should begin the following line.

is characteristic for the minimally sentence complete theory \vdash . But in general:

Question 1a: Does any minimally sentence complete theory have a characteristic model?

Equivalently:

Question 1b: Is there invariably a \leq -minimum E-class in any e-class?

I don't know.

Indeed, any consistent extension of our example theory had a characteristic model. In general:

Question 2: Does any sentence complete theory have a characteristic model?

No: at least I can find a counterexample to this — but it is not good enough to settle the first question.

We turn now to large theories: By Zorn's Lemma (or, better, since we are countable, by some enumeration construction), we can show

Lemma IV.4.4: Any consistent theory has a maximally consistent extension.

Hence in particular, a minimally sentence complete theory will have a maximally consistent — in other words, a sequent complete — extension. Clearly any such extension will have a characteristic model, and the E-class of this model will be \leq -maximal in its e-class.

In our example there was only one maximally consistent extension.

But, in general:

Question 3a: Does a minimally sentence complete theory (or any sentence complete theory) invariably have a unique maximally consistent extension? equivalently,

Question 3b: Is there invariably a \leq -maximum E-class in any e-class? I don't know, but I doubt it.

Given (IV.4.4), the following question is obviously equivalent:

Question 3c: If $M \sim_e N$, does there necessarily exist a model P such that $M \leq P$ and $N \leq P$?

Let us call a model of a maximally consistent theory a complete model. Our example shows that we cannot make do with complete models — that is to say, we do not have semantic completeness with respect to them. But what are complete models like? I know no illuminating independent characterization. However, one typical kind is those models which have, for every element in their domain, some parameter-free term denoting it. (Easy to check). In particular the models defined in (IV.1.15) are of this kind, and so are the models M^+ . Hence observe that \vdash_{M^+} is always maximally consistent.

Look back to our example models. Clearly, given any $M_{\alpha\beta}$ and $M_{\alpha'\beta'}$, we can find a one-one correspondence θ between their domains such that

$$M_{\alpha\beta} \stackrel{\theta}{\sqsubseteq}_e M_{\alpha'\beta'}$$

How typical is this? Of course, up to isomorphism, we have only one item in each E-class — and, certainly, this feature is quite untypical, as easy examples show. But, at least:

Question 4: If $M \sim_e N$, can we always find M' and N' such that $M' \sim_E M$, $N' \sim_E N$ and $M' \sqsubseteq N'$?

The answer is 'yes' (and, apart from any intrinsic interest this answer may have, it might help us to find answers to our other questions).

In preparation for this answer, we consider new kinds of 'diagram'. These are actually theories: the theories $\Theta(M^+)$ and \vdash_{M^+} , which are the e-theory and the E-theory, respectively, of the expansion M^+ to $L(\underline{D}_M)$ of a given model M for L . First observe

Lemma IV.4.5: $N \in K(\Theta(M^+)) \iff M \rightarrow_e N|L$

for " $M^* \subseteq N^*$ " and " $M^* \subseteq_L N^*$ "

This suggests an analogous notion of E-embedding, which we might define by

$$M \rightarrow_{\mathbf{E}} N \iff N = N' \upharpoonright L \text{ for some } N' \in K(\frac{\perp}{M^+})$$

An alternative characterization would then be derivable:

Lemma IV.4.6: $M \rightarrow_{\mathbf{E}} N$ iff there is an embedding $\theta: D_M \cup \{\Theta\} \rightarrow D_N \cup \{\Theta\}$ such that for any assignment $s: Var \rightarrow D_M \cup \{\Theta\}$ and any sequent α , (M, s) is consistent with α iff $(N, s \circ \theta)$ is consistent with α .

Notice that e-embeddings are characterized by the 'if' part of this condition. However, they do not necessarily satisfy the 'only if'. Of course, if $M \rightarrow_{\mathbf{E}} N$, then $M \sim_{\mathbf{E}} N$.

Now the interesting result is

Lemma IV.4.7: If $M \sqsubseteq_e N$, then there exist models M' and N' such that $M \rightarrow_{\mathbf{E}} M'$, $N \rightarrow_{\mathbf{E}} N'$ and $M' \sqsubseteq N'$.

Proof: Let us call the language for which M and N are models L .

Consider the two theories $\frac{\perp}{M^+}$ and $\frac{\perp}{N^+}$ in the expanded languages $L(D_M)$ and $L(D_N)$. Clearly there is no $\lambda \in Frm(L)$ such that

$$\frac{\perp}{M^+} \lambda \quad \text{and} \quad \lambda \frac{\perp}{N^+} *$$

And so, by (IV.2.11) case (1), there exist models M^* of $\frac{\perp}{M^+}$ and N^* of $\frac{\perp}{N^+}$ such that $M^* \sqsubseteq N^*$. Hence it will do to put $M' = M^* \upharpoonright L$ and $N' = N^* \upharpoonright L$.

Corollary IV.4.8: If $M \sim_e N$, then there exist models M' and N' such that $M \rightarrow_{\mathbf{E}} M'$, $N \rightarrow_{\mathbf{E}} N'$ and $M' \sqsubseteq N'$.

Hence we have, in fact, obtained something interestingly stronger than a simple 'yes' to question 4.

(No doubt you have been wondering when I was going to provide the alternative proof for (II.2.5). Well, this is it. Moreover, if we change ' \sqsubseteq_e ' to ' \square_e ' and ' \sqsubseteq ' to ' \square ' in the statement of lemma (IV.4.7), then the result is true (and can be proved along similar lines), and this

provides an alternative proof of (II.2.8).)

We cannot, of course, strengthen (IV.4.8) by changing ' \sqsubseteq ' to '=', on pain of collapsing \sim_e and \sim_E . However, could we do this, if, instead of E-embeddings, we considered e-embeddings? In other words:

Question 5a: If $M \sim_e N$, does there necessarily exist a model P such that $M \rightarrow_e P$ and $N \rightarrow_e P$?

Now we shall shortly prove

Lemma IV.4.9: There exists a P such that $M \rightarrow_e P$ and $N \rightarrow_e P$ iff there exists a Q such that $Q \leq M$ and $Q \leq N$.

Hence, the following question is equivalent:

Question 5b: If $M \sim_e N$, does there necessarily exist a model P such that $P \leq M$ and $P \leq N$?

We should note that 'yes' to (1a) and (1b) would clearly imply 'yes' to (5b) (and, hence, (5a)). Furthermore, these questions would all be equivalent, if we had at least \leq -minimal E-classes in an e-class. But do we? I don't know.

In connection with (IV.4.9):

Lemma IV.4.10: The following are equivalent ways of saying that M is a model of \vdash :

- (1) $M \in K(\vdash)$
- (2) $\vdash \subseteq \vdash_M$
- (3) $\vdash \subseteq \vdash_{M^+}$
- (4) $\vdash \cup \vdash_{M^+}$ is consistent
- (5) $\vdash \cup \Theta(M^+)$ is consistent
- (6) $M \rightarrow_e P$ for some $P \in K(\vdash)$

Proof: (1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (5) \Rightarrow (6) \Rightarrow (1).

Lemma IV.4.11: The following are equivalent ways of saying that $M \leq N$:

(1) $\vdash_M \subseteq \vdash_N$ (the 'definition')

(2) $\vdash_M \subseteq \vdash_{N^+}$

(3) $\vdash_M \cup \vdash_{N^+}$ is consistent

(4) $\vdash_M \cup \Theta(N^+)$ is consistent

(5) $\vdash_M \cup \Theta(M^+) \cup \Theta(N^+)$ is consistent

(6) $\vdash_{M^+} \cup \Theta(N^+)$ is consistent

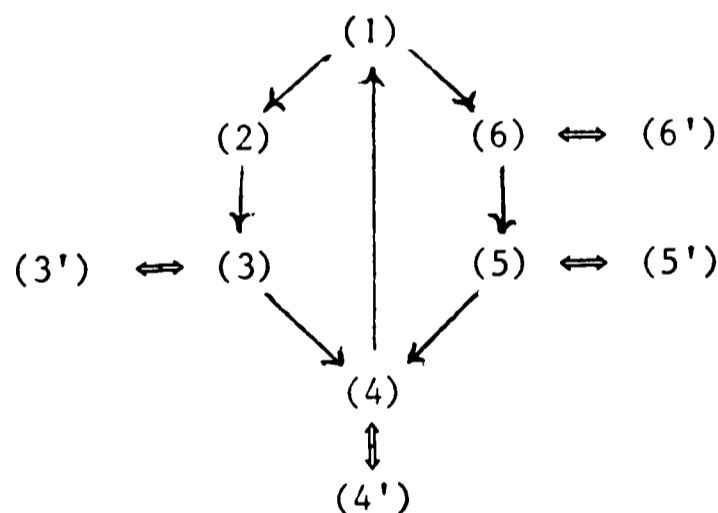
(3') there is a P such that $M \leq P$ and $N \rightarrow_E P$

(4') there is a P such that $M \leq P$ and $N \rightarrow_e P$

(5') there is a P such that $M \leq P$, $M \rightarrow_e P$ and $N \rightarrow_e P$

(6') there is a P such that $M \rightarrow_E P$ and $N \rightarrow_e P$

Proof:



The only difficult arrows are (4) \Rightarrow (1), which the last lemma fills in, and (1) \Rightarrow (6), which we shall consider.

First observe that $\vdash_{M^+} \cup \Theta(N^+)$ is consistent iff

(i) $\mathcal{T}(N^+) \not\perp \mathcal{X}(N^+)$

where \vdash is the theory in the doubly expanded language $L(D_M)(D_N)$ axiomatized by \vdash_{M^+} . Say that $M \leq N$ but (i) failed to hold. Then there exist

for "(Iv. 4.1)" and "(IV. 4.1)"

finite $\Gamma(\vec{x})$ and $\Delta(\vec{x})$ such that, for some $d_1, \dots, d_n \in D_N$,

$$(ii) \quad \Gamma(\vec{d}) \vdash \Delta(\vec{d})$$

and

$$(iii) \quad \Gamma(\vec{x}) \not\vdash_N \Delta(\vec{x})$$

From (iii), since $M \leq N$, $\Gamma(\vec{x}) \not\vdash_M \Delta(\vec{x})$, and hence

$$\Gamma(\vec{x}) \not\vdash_{M^+} \Delta(\vec{x})$$

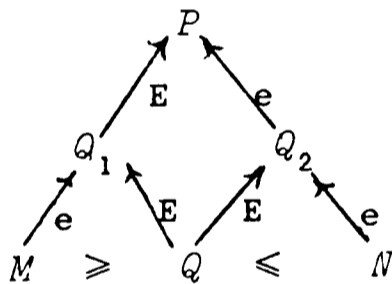
But, from (ii), by lemma (IV.2.3),

$$\Gamma(\vec{x}) \vdash_{M^+} \Delta(\vec{x})$$

which is a contradiction. □

Now we may return to (IV.4.9).

Proof of (IV.4.9): 'Only if': (IV.4.1). For 'if', consider the following relationships:



Given models satisfying the bottom row, we can find Q_1 and Q_2 , by condition (6') of the last lemma. Since Q_1 and Q_2 must be E-equivalent, by condition (6') again, we can find P . This P will do: $M \rightarrow_e P$ and $N \rightarrow_e P$.

Finally, there is an item of unfinished business from the end of the last section. We have to show that there is no set Σ such that $M \in K(\Sigma)$ iff $Px \not\vdash_M Qx, *$. Consider our example models M_{00} and $M_{\omega\omega}$. $Px \not\vdash_{M_{\omega\omega}} Qx, *$, and so, if such a Σ existed, $M_{\omega\omega} \in K(\Sigma)$. But $M_{00} \rightarrow_e M_{\omega\omega}$; hence, by (IV.4.1), $M_{00} \in K(\Sigma)$. However, $Px \vdash_{M_{00}} Qx, *$. Therefore no such Σ exists.

for " $(1a) - (2a)$ " read " $(1a) - (4a)$ "
and for " $(1) - (2)$ " read " $(1) - (4)$ "

CHAPTER V
PRESUPPOSITION

V.1 INTRODUCTION : SEMANTICS AND LINGUISTIC PRACTICE

Philosophers and linguists sometimes think about a class of idioms occurring in natural language which we might label 'presuppositional' and specify pretheoretically by a few paradigm examples from English:

sentences with a definite description in subject position, such as

- (1) The wisest philosopher in Oxford beats Mary.

sentences containing quantifier phrases, such as

- (2) All Jack's children beat Mary.

cleft sentences, such as

- (3) It's Jack who beats Mary.

sentences with aspectual verbs, such as

- (4) Jack has stopped beating Mary.

It is often said that each of these sentences has associated with it a certain presupposition, which we might specify respectively as

- (1a) that there is a wisest philosopher in Oxford
(2a) that Jack has children
(3a) that someone beats Mary
(4a) that Jack has been beating Mary

The point of saying this would be, roughly, that when someone sets out to make an assertion using (1) - (4) (in circumstances which are apparently typical and straightforward, and hence can be taken as theoretically paradigm) then he does not thereby assert (1a) - (2a) as a (conjunctive) constituent of his overall assertion, but, let us say, 'presupposes' these presuppositions - which is something different. The contrast would be with assertions made using what we might call the 'conjunctive counterparts' of (1) - (2):

(1b) There is a wisest Philosopher in Oxford and he beats Mary.

(2b) Jack has children and they beat Mary.

(3b) Someone beats Mary and Jack is he.

(4b) Jack has been beating mary but is'nt beating her now.

Now, the connection between assertions made using these sentences and (1a) - (4a), as conjunctive constituents of such assertions, seems relatively clear; and we can wield classical total-valued logic to provide analyses of (1b) - (4b) which reveal the connection reasonably well. But what have (1a) - (4a) to do with the assertoric use of (1) - (4)? What sort of thing are (1a) - (4a) as 'presuppositions'? What is there relationship with the sentences (1) - (4), and how might it be revealed?

There would seem to be a great variety of views on what the phenomenon of presupposition is supposed to come to, and hence, too, concerning exactly what kinds of idiom we would be justified in classifying together as presuppositional. Paradigms lead to theorizing and theorizing suggests further examples and ... But this final chapter is very far from being a survey of the literature: I shall follow a very narrow path and encounter only a very few authors who I shall want to invoke as partial allies or opponents - or both.

As a starting point, it is surely uncontentious to claim that we have some sort of intuition of something in common between our example sentences concerning their 'presuppositions', and that we have this intuition in virtue of some kind of linguistic regularity that speakers of English abide by: they are inclined to use presuppositional idioms and expect them to be used by others differently from their conjunctive counterparts. This is to say nothing yet concerning the nature of the difference nor how serious it is for an understanding of linguistic practice. Nonetheless it does not seem too unreasonable to think that there is something we might want a coherent account of.

The proposal offered in this chapter is that we should use the logic developed earlier to analyse presuppositional idioms. We shall then, I argue, be able systematically to reveal the connection between such idioms and their presuppositions. Of course, this will still leave open the question what it is to presuppose something when using a presuppositional idiom. Providing analyses in our logic makes this question more pressing, not less, since we should expect an answer which fits in a mutually illuminating way with the proposed analyses. What is the point of having presuppositional idioms in a language at all? What is going on when people use them? Even if we cannot manage an explicit definition for ' x presupposes that p in uttering s ', still we shall want to have some account of the connection between presupposition and important concepts in the theory of linguistic practice — in particular assertion.

We might mention immediately that, among our example sentences, (4) stands out in contrast to (1) - (3), because the presuppositional character of the sentence apparently arises from the lexically full-blooded item 'stop'. We shall be more interested in cases such as (1) - (3) where our simple partial-valued logic can, as it stands, provide a revealing analysis: but we do not want to ignore the possibility of treating the presupposition of sentences such as (4) in a uniform way. The only difference is that in the case of the more 'structural' examples we can apply our logic and rely on the fact that the relevant syntactic composition and its interpretation is familiar, whereas in the case of (4) we have a strange mode of composition — verb modifying verb — and, without lexical analysis at least, that part of the meaning of the sentence which determines its presuppositional character cannot be revealed as a matter of logic. However from the point of view of natural language this is probably not a very important difference.

However we are jumping ahead. In what sense are presuppositions to have anything to do with meaning? If we were forced at gun-point to paraphrase (1) - (4) in total-valued logic, then, with the possible exception of (2), we would be likely to provide the same - or at least equivalent - logical forms for these sentences as for their conjunctive counterparts (1b) - (4b). And a hard-line conservative might well claim that such a paraphrase for presuppositional idioms is the right thing to propose. Anything further that is to be said on the matter of presupposition would then have to be on this basis. Why complicate logic?

In the hope of finding some common ground with the conservative theorist on which to argue, or at least to air prejudices, something needs to be said about why sentences are to be 'analysed' by logical forms at all. I shall assume that the point is to provide semantic representations which are to go proxy for surface sentences in a systematic semantic theory for the language in question, and furthermore that this theory is, or is something like, a truth theory: such a theory can be seen as giving the 'literal meaning' of sentences of the language. This idea is certainly not incompatible with the assumption that the governing logic is other than classical total-valued logic, and if we talk for the moment - as the conservative theorist would - as if the governing logic were in fact total, generalizations will be easy to make.

Among the important *desiderata* of such a semantic theory is that it be 'revelatory' rather than 'relational' in at least the following two ways. In the first place it must be 'contributational' in that it shows how the meaning of a sentence (or indeed any subsentential constituent) is determined by the meaning of its parts by giving the meaning of a sentence in terms of the meaning of its parts. This would contrast with a theory which just specified logical and/or semantical relationships between sentences as

wholes. In the second place it must be 'interpretative': the meaning of the linguistic items mentioned must be given using the language of the theory in such a way that, ultimately, the meaning of an object language item is completely determined by the meaning which the language of the theory itself exhibits. 'Ultimately' is included here to allow for the possibility of incorporating model theory, say, or any other mechanisms more complicated than mere disquotational clauses, into the semantic theory. The contrast with a relational approach remains: we are not to make do simply with mentioning relations involving the object language — either logical and/or semantical relations holding between object language items themselves, or a translation relation between object language items and some further language or system of semantic representations. See the introduction of Evans and McDowell (1976) on this.

If a semantic theory is revelatory in these ways, then the logical relations that govern the theory itself will be more important than those that govern the object language: the latter should indeed be determined by the former. In this context, then, the point of giving sentences of a language 'logical forms' is to provide the theory with items whose 'syntactic' structure is semantically perspicuous, so that it may systematically treat of sentences *via* their logical forms. What logical forms should actually look like may be a problem: there is likely to be some tension between the systematic requirements of the theory on the one hand, and the desire to preserve appropriate features of grammatical structure on the other. For example, should complex quantifier idioms be 'analysed out' or not? In logic it is important to see how this may be done, and as an heuristic aid it is certainly useful to make truth conditions perspicuous in this way; but is it ultimately out of place in a semantic theory for natural language sentences? This question need not, however, detain us: our discussion here and later will mostly be sufficiently coarse-grained to allow us to

'factor out modulo (logical) equivalence'. In particular we might characterize the conservative theorist as one who, using classical logic, would provide the 'same' logical forms for sentences (1) - (4) and (1b) - (4b).

Finally, a good semantic theory must be 'linkable': the theory will have to be hedged around in some way or other to take account of linguistic practice in such a way as to reveal why it is the right semantic theory. Itself the theory just gives the meaning of sentences, while the surrounding account would — if ever it could be spelt out fully — actually tell us 'what meaning is'. Let us assume that the revelatory character of a semantical theory provides, directly or indirectly, a semantical ascent predicate T in terms of which it gives the literal meaning of a sentence s by stating the equivalence of $T(s)$ and p_s , where p_s is a sentence of the language of the theory. We must then have some way of linking T with what speakers of the language are doing. If T is actually taken to be a truth predicate, then the account would have to be one which catered for the idea that determining the truth conditions of a sentence and determining its literal meaning came to the same thing; and this would presumably involve some way of relating the notion of truth deployed in the treatment of sentences by the semantic theory, to the notion of truth we deploy in evaluating assertions made using those sentences. At any rate, from such an account it would seem reasonable to demand the license to talk in one breath — as inevitably we are tempted — about sentence truth and the truth of assertions. A link principle must be forthcoming to the effect that a sentence s is true if and only if what is asserted using s is true; and, if our semantic theory is right, what is asserted using s is that p_s .

Of course making assertions is not the only kind of linguistic activity, but we shall be restricting attention to this, and so these vague remarks will do. It might, however, be worth pointing out that, as far as

our interest in presupposition is concerned, this restriction perhaps makes life more difficult than it need be. For if only we were to consider questions too and urge that this speech act has as much right, if not more, to be linked directly with a truth theory, or urge that questioning should be linked correlatively with assertion making, then many of our points about presupposition would be easier to make: roughly, the phenomenon is less easy to ignore in the case of questions.

But to return to assertions, we should in any case note that the link formula as stated is unbelievably crude. Not only is it silent on the exact nature of the explanatory link-up between systematic semantics and linguist practice, but (i) it ignores the possibility of indexical expressions and other kinds of contextual dependence; and (ii) it ignores the subtlety of what speakers can be taken to be doing with innocent looking sentences. And, of course, it has been stated under the assumption that the governing logic is classical and total-valued: how might the link formula go if our partial-valued logic, for example, were used?

Nevertheless, postponing this final question, and bearing in mind the qualifications that are going to be called for, let us consider what the conservative theorist might do about presupposition. When a semantic theory is itself too austere to point up a particular aspect of meaning, a theorist might naturally attempt to invoke the Gricean notion of conversational implicature to boost up the account. This idea can serve to explain how what a speaker means may be at variance with what he (strictly and literally) says, where what is said corresponds with what the semantic theory gives as the literal meaning of the sentence used, while what is meant may have more to it and may even be incompatible with what is said. In its purest form the beauty of this idea is that, given just what the semantic theory specifies as literal meaning, then, on quite general

principles of communication, we have an account determining what may be meant in given kinds of context. So, for example, we might have an account of irony or metaphor, and so too we might have a way of accounting for certain aspects of the extra richness of natural language modes of composition that their logical analogues do not exhibit. (Consider the discussion of * in section I.3 in this connection). In fact the mechanisms of implicature give rise to one of the ways in which our link principle was stated too crudely, viz. qualification (ii); and so for 'asserted' we should understand 'said', not 'meant'.

Could, then, the conservative theorist deploy this notion to explain presupposition, if (1) - (4) are given the same (or equivalent) logical forms as (1b) - (4b)? Could he deploy it in such a way that the description of what is usually meant in an assertoric utterance of (1) - (4) accommodates our intuitions about presuppositions, though what is strictly and literally said is the same as - or equivalent to - what would have been said (and meant) using (1b) - (4b)? I am sceptical. I would grant that there are sentences which are sometimes, with good reason, considered to be presuppositional which may, nonetheless, be amenable to this kind of treatment - perhaps the presupposition that p in utterances made using sentences of the form ' x knows that p ' could be explained in this way; but, without some contamination of the purity of the idea, I do not see how it could be applied to our examples. The point is this: if (1) - (4) are given the same or equivalent literal meanings as (1b) - (4b), then these conjunctive counterparts should behave in exactly the same way as the presuppositional idioms and hence exhibit the same presuppositional behaviour - but they do not. It was precisely the difference in use of the two kinds of idiom that was invoked to point up the idea of presupposition in the first place.

for "hand-cord" read "hand-cove"

How, then, could there be an account which appealed only to the literal meaning of a presuppositional sentence, if this is taken to be what the conservative theorist would stipulate? We should need also to appeal to characteristics of the surface sentence which distinguished it from its conjunctive counterpart. And these are not superficial characteristics, such as length or complicatedness, but they involve particular words or particular forms of grammatical construction. But if we allow this into the account, then purity would be lost: we should be invoking a conventional distinction between presuppositional idioms and their conjunctive counterparts. For I am, surely, entitled to see these distinctions as conventional — however exactly you want to spell out that notion — since presuppositional idioms might have been such that they were used in the same way as their conjunctive counterparts, whereas in fact they are not.

On the conservative view, then, it would appear that we would have to have presupposition as an aspect of meaning which was not an aspect of hard-core literal meaning, but was stuck on alongside, so to speak, and ignored as far as treatment by the semantic theory was concerned. We might crudely describe this approach as assigning sentences pairs $\langle \phi, \pi \rangle$ where ϕ is the logical form proper and π some tag specifying the presupposition.

This I think would be a pity: it is time to air one of my guiding prejudices in a very general form. If there is any inclination to discern an aspect of meaning which is pointed up by some contrast in the use of different (classes of) surface sentences, then this aspect of meaning and this difference should, if possible, be catered for as an aspect of literal meaning in as hard-core a sense as possible. No doubt tensions may arise between this principle and *a priori* considerations motivating a particular

kind of semantic theory. But conservatism alone should not be given any weight: a lot may be done without disturbing important features desired of a theory, or modifications may be adopted which preserve what is taken to be important about those features. Any specification of what a semantic theory looks like will presumably be committed to idealizing matters, but the more subtleties a theory can be seen to handle, the less idealization is involved, and hence the more compelling such a theory will look.

To return to presupposition, the conservative theorist might be prompted to respond at this stage by pointing out that it is open to him to be subtle about the particular logical forms he assigns to (1) - (4) and (1b) - (4b). They may in fact look very different, though they are equivalent; and he would say that it was wrong of me to be so cavalier earlier, when I claimed that we could factor out modulo equivalence and talk of equivalent logical forms as the same logical form. For the semantic theory, if it is a truth theory say, would in fact give different truth conditions for the two kinds of sentence: 'truth conditions' are to be sensitive to the details of structural complexity - they are not merely the states of affairs specified. Hence a difference in meaning would have been revealed.

However, if presuppositional idioms and their conjunctive counterparts are taken to have materially the same truth conditions (whether or not the semantic theory itself actually has this as a theorem) and differ only in their structure, then the theory could still not tell us which was which. Certainly the right s may get paired with the right p_s - the mechanisms of the theory would preserve particular distinctions; but the theory would provide nothing systematic which we could refer to if we wanted to attempt a general explanation of the distinction between presuppositional idioms and their conjunctive counterparts. The proposal would have little edge over the crude tagging device using pairs $\langle \phi, \pi \rangle$, only now π would be

implicit in the form of ϕ . However my prejudice leads me to demand more explanatory potential of the semantic theory.

Let me draw an analogy. Ignoring the semantical importance of the distinction between a presuppositional idiom and its conjunctive counterpart would seem to me to be as bad as ignoring the distinction between a pair of contradictory sentences, claiming, say, that really "It is raining" and "It is not raining" have the same, or equivalent, literal meanings, while the distinction is to be explained peripherally: sometimes we make assertions, and sometimes we make denials, and it just so happens that we use sentences like "It is raining" to make assertions and sentences like "It is not raining" to make denials.

In a good truth theory, however, we should expect the nature of the semantical ascent predicate and an understanding of the governing logic of the theory to provide a general framework to invoke when we wanted to talk about contradictoriness. Note: I have said nothing here about negation — this would take us a step further.

These remarks explain why I am happy to count myself among those who would deploy a partial-valued semantics to explain presupposition — or at least to provide the necessary explanatory potential. The idea is familiar: and it has some intuitive plausibility. If, then, we want to adopt a revelatory semantic theory whose governing logic is partial-valued, this might prompt the following refinement of the principle that literal meaning is to be specified in terms of truth conditions: a determination of the literal meaning of a sentence is to come to the same thing as a determination of its truth-falsehood conditions (as these notions are deployed in the theory). Then presupposition as an aspect of literal meaning would fall neatly into place: a determination of the presupposition associated with

a sentence comes to the same thing as a determination of the truth-value preconditions of that sentence, i.e. preconditions for the sentence's being either true or false. Clearly literal sentence-presupposition would then actually be a function of overall literal sentence-meaning; and it is allowed to be non-trivial, since truth conditions and falsehood conditions are not assumed to be exhaustive.

Of course the appeal of all this depends on the plausibility that such a theory would be 'linkable' in an appropriate way with linguistic practice. We shall return to this.

But perhaps this sketch already takes us further than some partial-valued logicians would bargain for? It would, however, seem to me to be the right path to follow if we are to take seriously the requirement that our semantic theory be 'revelatory' rather than 'relational'. And this requirement would seem to me to strengthen the case against the conservative theorist; for we can point out that we have not merely a regular distinction between presuppositional idioms and their conjunctive counterparts, but that the behaviour of presuppositional idioms is systematically articulated in the language.

The most obvious example of this is negation: the 'natural negation' of a presuppositional sentence appears to be associated with the same presupposition as the original sentence; and this, of course, is in no way equivalent to anything that would count as the natural negation of the corresponding conjunctive counterpart. Cf. Frege (1892). By the natural negation of a sentence I mean what is grammatically natural and is, furthermore, the idiom naturally used to contradict what someone says using the original idiom — or at least, to do what intuitively appears to be contradicting what someone says. For example, the natural negations of (1) - (4) are:

- (1c) The wisest philosopher in Oxford does not beat Mary.
- (2c) Some of Jack's children do not beat Mary.
- (3c) It is'nt Jack who beats Mary.
- (4c) Jack hasn't stopped beating Mary.

This negation phenomenon is often invoked as the inroad into the very idea of presupposition – and it is sometimes even used to provide a 'definition'. However there is more to the structural involvement of presupposition than simply its behaviour under negation: there are various systematic phenomena, which the linguists worry over under the general heading 'the projection problem for presupposition'. This is what we discuss in Section 3 of this chapter.

Of course, in saying that natural negations are grammatically natural I do not mean to imply that we have any straightforward grammatical criterion. Examples (1c), (3c) and (4c) might have been taken as suggesting that natural negations are obtained by applying a grammatical operation of negation directly to the main verb, but example (2c) shows this to be false: "All Jack's children are not bald" is not the natural negation of (2). So there is certainly no simple argument that we should have a logical operation of natural negation because it corresponds directly with a grammatical operation.

Such an operator would force us out of total two-value logic, since it could not be taken to yield a sentence equivalent to the classical negation of a conjunctive counterpart; but the conservative theorist could say, that we simply have to be sensitive about scope: in the case of (1c) for example, negation (ordinary classical negation) would come within the scope of the definite description thought of as a complex quantifier. Example (2c) could provide no evidence against this manoeuvre. On the other hand example (4c) might do so, since without lexical analysis there could be

nothing for negation to be appropriately within the scope of; but then examples such as (4) and (4c) would be ones whose presuppositional behaviour may be more easily handled in terms of implicature.

The fact remains, however, that speakers have these intuitions about natural negation, which mesh with the distinction felt between presuppositional idioms and their conjunctive counterparts. And, however complicated the mechanisms may be, we have morphological distinctions systematically coinciding with inclinations and expectations concerning the use of sentences. How, then, can we resist the temptation to imbue this grammatical articulation with a semantical point, and, for one thing, to discern a logical operation of negation which applied to (the logical forms of) (1) - (4) yield sentences giving, if only up to logical equivalence, their natural negations (1c) - (4c) ?

The discussion has now reached the point where it is not so much the conservative theorist who is being cajoled but the triclassificatory semanticist who, while emphasizing the need to take presupposition as a serious semantical phenomenon, would not take the proper steps to see it revealed but would be content merely to specify semantical relationships. This remark should be taken to apply to all cases of presupposition that we may be motivated to discern in the literal meaning of sentences, but it is particularly pertinent to the kind of example where the presupposition may be supposed to arise in virtue of some readily schematizable structure. For to treat of these examples we might be expected to need basic vocabulary in our language of logical forms which not only takes but also makes presupposition: that is to say, iterable items of vocabulary a specification of whose meaning by the semantic theory will reveal a contribution to the presuppositional character of sentences in which they occur — and moreover reveal what contribution they make.

To return to the analogy with contradictory pairs of sentences: ignoring this requirement would be like recognizing that "It is raining" and "It is not raining" are contradictory, but not bothering to discern an operation of negation in terms of which this fact could be explained.

We have, of course, returned to something like the theme of Section I.3 which motivated the introduction of our novel sentence connectives. There we called for logical analysis in terms of these connectives, and showed how it could be done in some particular cases: in particular, we defined some presuppositional quantifiers. Now the point is that since we have got these quantifiers then let us deploy them, analysed out or not, in a revelatory semantics governed by our partial-valued logic. We discuss this in Section 2. In Section 5 we consider another presuppositional quantifier, which can also be analysed in our logic, and we argue that this should be used to provide an account of the meaning of a class of idioms which, as far as I know, have not actually been thought to be presuppositional before.

But now, how would a semantic theory governed by our partial-valued logic fit with linguistic practice? The naive approach would be to see if we could work with the obvious modification of the classical link principle, and, invoking two positive truth-values true and false, say: a sentence *s* is true (respectively false) if and only if what is asserted using *s* is true (respectively false). If this made sense, then we would be able not only to see a determination of the literal meaning of a sentence as the same thing as the determination of its truth-falsehood conditions as specified by the semantic theory, but correspondingly to see a determination of the content of an assertion — what is asserted — as coming to the same thing as a determination of its evaluation conditions in terms of 'true' and 'false' as applied to assertions. This is the course I want to take (modulo

the same kind of qualifications as before concerning contextual dependence, implicature, etc.). What a speaker is presupposing in virtue of his use of a presuppositional idiom on a given occasion will then be seen as a function of what is asserted — just as literal sentence presupposition is a function of literal sentence meaning; and we will therefore be seeking an account of the relevant notion of speaker presupposition — what it is to presuppose that p in uttering s — which reveals it to be a linguistic sub-act of the act of assertion.

But this naive link principle is likely to meet with objections — even from those who might be prepared to go along with what I have said up till now. I know of two (incompatible) lines of objection. The first would come from theorists who held the link formula to be ill-formulated (or rather truth-value-less!) on the grounds that there could be no assertion made with a sentence which was neither true nor false: only if a presupposition held would an attempted act of assertion be fully successful — only then would any assertion actually be made. Recall our discussion of Belnap and van Fraassen in Section I.3. I see no attraction in this idea. Perhaps it is the historical connection between presupposition failure and the failure of indexical reference that has given this view currency? However that might be, the rival picture which I wish to urge, with initial examples such as (1) – (4) in mind, is outlined in section 4. And I hope that the account we give there will shed illuminating light also on the novel examples considered in Section 5.

The second objection is quite different, and, given my rejection of the first, it becomes very disquieting. It is Dummett's. Granted the use of a triclassificatory semantic theory to account for structural features of languages such as English, and granted too that a corresponding triclassificatory evaluation of assertions may be made, simply by taking

the link principle as a stipulation of how to do it, this does not mean to say that the practice of making assertions itself warrants such a classification. Dummett thinks that it does not. Correspondingly, though a notion of literal meaning may be forthcoming from a triclassificatory semantic theory, this does not mean to say that the content of assertions is correctly characterizable in such a way as to yield a system of assertion-contents directly isomorphic to the system of literal meanings. This has nothing to do with a Gricean distinction between 'meaning' and 'saying'; there is to be a disparity between what is strictly and literally said and the literal meaning, in the semantical sense, of a sentence.

The objection is not this time that the third classification means no assertion, but rather that as an evaluation of assertions it is nothing more than a kind of falsity — or rather just is falsity *simpliciter*. There is, on this view, a rationale only for a two-fold evaluation of assertions, and so the notion of assertable content determined in this way is a coarser way of discerning meaning distinctions than the, now somewhat demoted, notion of literal meaning. In Section 6 I shall attempt a response to this view by bringing to bear the discussion in Section 4 to suggest that a three-fold classification of assertions is non-trivial.

This issue is the issue of how serious presupposition is — how serious in linguistic practice the distinction is between a presuppositional idiom and its conjunctive counterpart. For on the Dummett view all our inclinations and expectations concerning the distinction are to be explained solely in terms of the structure of our language — details which are seen as having no independently recognizable point to them.

I find this view implausible: when, earlier, I was urging that we should give in to the natural temptation to imbue grammatical articulation with a semantical point, I was taking the semantical point to be something

we should see further explained in terms of linguistic practice; I did not think that grammatical articulation was to be its own semantical point. However this would appear to be what the Dummett view comes to.

I should like to consider this matter immediately. For in the first place it is something of a puzzle to me why Dummett is so ready to consider a triclassificatory semantics at all, given his views on assertion. Should he not rather be a hard-line conservative? Perhaps his discussion of three-valued semantics is nothing more than a desire to provide a rational reconstruction of a Fregean heritage, and a technical interest in many-valued logics as they are traditionally conceived (with a dichotomy of 'values' into the designated and the undesignated)?

Let us consider negation again. As we outlined in the general introduction, Dummett envisages using a three-entry matrix

<i>s</i>	<i>t u f</i>
<i>not-s</i>	<i>f u t</i>

This would provide the interpretation for a genuine univocal sentence functor taking a sentence to its 'natural negation', and, if it were adopted, it would alone be sufficient to force us into a triclassificatory semantics. But why might Dummett be motivated to adopt such a thing? My motivation to do so is that I have intuitions about natural negation: intuitions which go hand-in-hand with general intuitions concerning presupposition and which I would hope to see further illuminated by an account of linguistic practice. However what is Dummett allowed to make of any such intuitions? For his account of assertion might be taken to show that no further illumination is possible — that in fact these intuitions are positively misleading: while *t* is correlated with TRUTH, *u*, as well as *f*, is to be correlated with FALSITY. What then would be the point of

considering a univocal sentence functor? As we have seen, and as Dummett himself points out, there is no simple grammatical mechanism that such a functor would be directly imitating; and the conservative theorist can, without much difficulty, mess around with scope to make what he says formally consistent.

However it would appear that Dummett does in fact take intuitions about natural negation seriously. He would, for example, wish to explain our temptation to deploy a three-fold evaluation of assertions in terms of natural negation — by saying that this temptation arises because of our inclination to call a sentence s false just in case $not-s$ is true. In other words, we are being invited to define t , u and f in terms of TRUE, FALSE and an operation of natural negation. But then do we not have something of a problem? For, while intuitions concerning negation are apparently themselves inexplicable, they are nonetheless being called upon to explain other presuppositional intuitions. (Dummett even proposes the rather round-about route of explaining the presupposition of questions in terms of the natural negation of sentences used to make assertions.)

Perhaps Dummett's reply to this would be that of course we are not just pulling a notion of negation out of the air and expecting it to back up all our intuitions about presupposition; rather, natural negation is just one of an interconnected web of linguistic phenomena which support one another. Moreover, this web of phenomena need not be restricted to how we see the relationship between sentences with given structures, but we may say that it includes such things as styles of response. For example, the conversational manifestation of the triclassificatory idea is itself such a phenomenon, viz. the inclination not only to say "Yes, you're right" and "No, you're wrong" to presuppositional idioms but, under certain conditions, also, "Oh but, ...". And clearly features of the linguistic

context of an utterance are important too: for example, if something has just been asserted, then this might determine the use of a presuppositional idiom rather than its conjunctive counterpart, if the conjunctive counterpart would involve explicitly reasserting the same thing. All these features of our language, it would be said, are so systematic and widespread that our intuitions concerning any one of them can be quite adequately explained away in terms of the others without recourse to anything outside this web.

As a way of accounting for any particular intuition these remarks may carry some force, but I would still hanker after something more substantial: some light to be shed on the whole web of phenomena. We shall be attempting to do this in Section 4. Yet the difficulty will be to know whether or not we have really disentangled ourself from the web; for it may be that even within it there is quite a lot to say about presupposition in terms of what might quite appropriately be called 'linguistic practice'—things concerning when and why speakers use presuppositional idioms rather than conjunctive counterparts, and concerning what an audience might be expected to make of the choice of one idiom rather than the other, etc. etc. The point remaining would presumably be whether the way things are described really breaks far enough out of a linguistic scenario: whether the phenomenon of presupposition has been shown to be important in accounting for how languages fit onto the world — in accounting for how people's cognitive state, their view of how things are, their expectations and behaviour, are affected by understanding what is said to them.

This is a difficult question, but my remarks will be aimed at the thesis that presupposition is important in this way.

V.2 PRESUPPOSITIONAL QUANTIFIERS

The 'semantic notion of presupposition' is usually introduced in terms of a relation ' s_1 presupposes s_2 ' between sentences. However, in accordance with the change of emphasis that we have been urging throughout, let us rather introduce the idea by saying that the presupposition of a sentence is (determined by) its truth-value preconditions — that is to say conditions for being either true or false. This formula is to be understood in the same way as the familiar formula "the meaning of a sentence is (determined by) its truth conditions". It is not particularly relevant, just at the moment, exactly what way this is — whether, for example, you favour thinking in terms of sets of possible worlds or in terms of mention-use clauses. Nor are we to enquire yet what the three-fold semantic classification 'true-false-neither' really comes to: until Section 4 we must rely on intuition (or prejudice).

But if we adopt a modification of the general principle we have quoted concerning meaning and truth conditions, and say that the meaning of a sentence is (determined by) its truth-falsehood conditions, then our way of introducing sentence presupposition will show clearly that it is constitutive of sentence meaning; for given the truth-falsehood conditions of a sentence the truth-value preconditions are thereby given also. From all this a relation of presupposition between sentences could easily be defined, and the role of a partial-valued logic such as ours in explaining certain instances of it would fall into place. However we do not need to appeal to any such relation to say what a presuppositional idiom is; rather let us say, with calculated looseness, that a presuppositional idiom is a sentence with a non-trivial truth-value precondition.

Our first example sentence in the last section involved a definite description; and contemporary writers on presupposition often like to trace

their intellectual pedigree back to Strawson's consideration of definite description idioms. In Strawson (1950) he is taken to have popularized the originally Fregean notion (Voraussetzung) in his attack on the analysis of definite descriptions in Russell (1905). Ironically the word 'presupposition' does not actually occur anywhere in this article, and, in any case, the debate over the truth-falsity conditions of definite description idioms was only one, and not perhaps the most important, strand in the encounter. Frege had considered the phenomenon of Voraussetzung in connection with the recalcitrance, as he saw it, of natural language denoting expressions which, while fully senseful, might lack a denotation and hence give rise to truth-value gaps. Recall the discussion in section II.2. Strawson went a stage further and entangled the question of truth-value gaps with a particular theory of reference, according to which the functioning of definite description phrases was taken to be similar to that of demonstrative expressions.

To disentangle matters we ought first, I think, to discern (at least) two ways in which an account of definite descriptions could be Strawsonian as opposed to Russellian. Firstly, do definite description expressions function — as Strawson held — like demonstratives? (A further question would be: do they function in the particular way Strawson took demonstratives to function?) Or do they — as Russell held — function not as singular terms at all, but rather in the same way as complex quantifier expressions such as 'every ϕ ', 'some ϕ '? (Of course Russell's proposal went further than this, since he offered a paraphrase of the complex quantifier 'the ϕ ' in terms of the simple quantifiers $\forall x$ and $\exists x$ and logical connectives — a paraphrase, of course, in classical total-valued logic.) Secondly, there is the question whether, if there is no unique ϕ , neither a true nor a false assertion can be made using a sentence of the form 'the ϕ is ψ ' (Strawson), or whether a false one would be made (Russell).

Now, a Strawsonian answer to the first question might encourage a Strawsonian answer to the second; however, the other way about, a Russellian view on the first issue could fit perfectly well with a Strawsonian view over the truth-falsehood conditions. Of course it would be rash to talk of 'natural language definite descriptions' as if we had a clearly demarkated totality of nominal phrases for which we are seeking a uniform account. But let us for the moment take phrases of the form 'the ϕ ' as they occur in idioms of the form 'the ϕ is ψ ' as paradigms in English of what we mean. And for such phrases I would in fact be inclined to seek a uniform account — one along the lines of the Russell-Strawson combination I mentioned. Note that, as I stated it, the 'Strawsonian' view on truth-falsehood conditions covers both the no-assertion thesis and the truth-valueless-assertion thesis: but, as indicated in the introduction, I shall be interested in the second version. I do not propose to argue properly for a uniform treatment of definite descriptions; my present concern is simply to urge that the phenomenon of presupposition, as it may be taken to arise in the case of definite description idioms, need not be taken to have anything to do with special problems concerning singular reference, but may be aligned rather with the presuppositional behaviour of the other example sentences mentioned in the introduction: we have a quite general phenomenon on our hands.

Over the first issue (the issue concerning how definite descriptions function as linguistic items) we might record an intermediate position — Frege's, in fact — viz. the construal of definite description expressions as singular denoting expressions, but ones whose function is quite unlike that of demonstratives in that their denotation, if any, is determined solely by the meaning of the constituent vocabulary of the description. It would seem to me that this proposal is really more Russellian in spirit than Strawsonian. We saw in Section II.3 that any Fregean ι -term

for "expressions" read "descriptions" (1.27)

could be perfectly mimicked by a quantifier — one, furthermore, which was definable in terms of our basic vocabulary. Hence, in partial-valued logic at least, the Fregean approach is subsumed under the Russellian one; at least this is so modulo logical equivalence. Admittedly, the converse is not the case, since, as we also saw in II.3, if we start with the idea that definite descriptions are quantifiers, we might choose an interpretation which could not mimic terms properly even in an extensional context. But the important point is that the contrast between Strawson and Russell is a contrast between Strawson and Frege too: on a Strawsonian account of the functioning of definite description expressions, the ϕ in 'the ϕ ' is merely one of the things contributing to a determination of the object (if any) for 'the ϕ ' to refer to. There could, as one result of this, be a temptation to countenance misdescribed objects as nonetheless *bona fide* referents, though such an object could never be the Fregean denotation of 'the ϕ '.

This is not to say that a Frege-Russell approach is committed to divorcing definite description terms completely from contextual dependence. On this analysis there remains the problem of 'incomplete descriptions' — I mean cases where contextual relativization such as 'in this room' or 'in Oxford' would have to be supplied for a description uniquely to determine what it is (correctly) understood to denote. To say that strictly and literally such descriptions are vacuous does not seem satisfactory, not only because so many every-day descriptions are like this, but also because grasping what relativization is made would seem sufficiently crucial for the understanding of an utterance to motivate an account of this relativization as part of what is literally said. However there is surely no reason to suppose that we should be driven back to aligning expressions with a simple demonstrative paradigm of singular identifying reference, since we have what is surely the same phenomenon with most quantifier expressions in natural

language. For example, if I say "Every undergraduate sits two public examinations", I mean every undergraduate in Oxford.

Some uniform account of the mechanisms of reletivization in all these cases is required: there always seems to be some indexical tag on 'variable-binding operators' — and not necessarily the same one for all such operators in a single sentence. Consider for example: "Everyone (*sc.* in this meeting) knows that the oldest head of house (*sc.* in Oxford) admires almost every young woman (*sc.* anywhere)". I have nothing intelligible to say on this matter, however, and I shall be abstracting away from the problem in what follows. (Nonetheless, it is tempting to speculate whether, once a theory of quantifier indexicality has got going, we might not after all be able to see demonstratives and definite descriptions in a uniform light: not because descriptions are forced into a special theory of singular demonstratives, but rather because demonstrative expressions are tantamount to a special case of indexically tagged quantifiers, viz. certain kinds of definite description.)

We might, then, consider deploying our complex quantifiers $\exists x \dots x \dots$ and $\forall x \dots x \dots$ to give the logical form of idioms such as sentences (1) and (2) in the last section. There is no need to go into details, since we have already discussed these quantifiers in Chapters I and II, and their use is obvious. But we can note that example (3) could be given the analysis

$\exists!x. x \text{ beats Mary} / \text{John beats Mary}$

or equivalently

$\exists x (x \text{ beats Mary}). \text{John} = x.$

To compensate for this cavalier attitude towards the details of the analysis of (3), let us now consider how well $\exists x \dots x \dots$ and $\forall x \dots x \dots$ can be deployed in a direct representation of complex sentences in which quantifiers are embedded in the scope of one another. The question to hold

in mind is whether the truth-falsehood conditions are plausible — whether they yield a plausible presupposition for the compound sentence. It is likely to be psychologically difficult to 'work out' the compound presupposition of a formula if this is attempted directly *via* truth-falsehood clauses, but there will always be a formula we can write down which actually displays the presupposition. For we know how $\exists x \dots x \dots$ and $\forall x \dots x \dots$ may be analysed in terms of / and we can show that any formula of our logic is equivalent to one in the form ϕ/ψ where ϕ and ψ contain only the classical quantifiers and connectives: hence ϕ will represent the presupposition of the original formula — providing, at least, that no basic predicates make any presuppositional contribution, or, if they do, providing we ignore it. Furthermore there is an effective procedure for transforming a formula into such a normal form: we may use the following rules:

$$\phi \simeq \top/\phi$$

$$\neg(\phi/\psi) \simeq \phi/\neg\psi$$

$$(\phi/\psi) \wedge (\chi/\omega) \simeq (\phi \wedge \chi) \vee (\phi \wedge \neg\psi) \vee (\chi \wedge \neg\omega) / \psi \wedge \omega$$

$$(\phi/\psi) \vee (\chi/\omega) \simeq (\phi \wedge \chi) \vee (\phi \wedge \psi) \vee (\chi \wedge \omega) / \psi \vee \omega$$

$$(\phi/\psi) \rightarrow (\chi/\omega) \simeq (\phi \wedge \chi) \vee (\phi \wedge \neg\psi) \vee (\chi \wedge \omega) / \psi \rightarrow \omega$$

$$(\phi/\psi) / (\chi/\omega) \simeq \phi \wedge \psi \wedge \chi / \omega$$

$$\forall x(\phi/\psi) \simeq \forall x(\phi \wedge \psi) \vee \exists x(\phi \wedge \neg\psi) / \forall x \psi$$

$$\exists x(\phi/\psi) \simeq \forall x(\phi \wedge \neg\psi) \vee \exists x(\phi \wedge \psi) / \exists x \psi$$

The left-hand side of / may be pretty horrendous, but we can subsequently simplify it. If we make the assumption that the quantifiers do in fact have something to range over and that basic predicates are not presuppositional (i.e. that they are totally defined), then this simplification procedure is no different from what it would be in classical logic.

As a simple illustration, consider a sentence such as

(5) The King's uncle is bald.

to be analysed in the form:

$$(5a) \quad \exists x(\exists!y Fy . Gxy) . Hx$$

Equivalent formulae are

$$(5b) \quad \exists!y Fy \wedge \exists!x(\exists y(Fy \wedge Gxy)) / \exists x \exists y(Fy \wedge Gxy \wedge Hx)$$

and

$$(5c) \quad \exists!y Fy \wedge \exists!x(\forall y(Fy \rightarrow Gxy)) / \forall x \forall y(Fy \rightarrow (Gxy \rightarrow Hx))$$

The left-hand side of either formula can be paraphrased by "There is a unique king and he has a unique uncle". This seems plausible.

A more complicated example would be:

$$(6) \quad \text{Everyone hurts the one he loves.}$$

Let us use $\forall x$ for 'everyone', to make things easier; then we have something of the form :

$$(6a) \quad \forall x(\exists!y Fxy . Gxy)$$

This is equivalent to:

$$(6b) \quad \left\{ \begin{array}{l} \forall x(\exists!y Fxy \wedge \forall y(Fxy \rightarrow Gxy)) \vee \\ \exists x(\exists!y Fxy \wedge \forall y(Fxy \rightarrow \neg Gxy)) \end{array} \right\} / \forall x \forall y(Fxy \rightarrow Gxy)$$

$$(6c) \quad \left\{ \begin{array}{l} \forall x(\exists!y Fxy \wedge \exists y(Fxy \wedge Gxy)) \vee \\ \exists x(\exists!y Fxy \wedge \exists y(Fxy \wedge \neg Gxy)) \end{array} \right\} / \forall x \exists y(Fxy \wedge Gxy)$$

In other words the presupposition of (6) is that either everyone loves some (unique) person whom he hurts, or there is someone who loves a (unique) person whom he does not hurt. Now at first sight this may seem inordinately complex: why does hurting (G) enter into the presupposition? A first attempt at specifying the presupposition of this example might have been simply $\forall x \exists!y Fxy$. But this will not do, because it is only a condition for truth, whereas what we want is a condition for truth or falsity. To falsify (6) we do not require that for anyone there is a (unique) person whom he loves, but merely that there is someone who loves a (unique) person — if that is a person whom he does not hurt. Hence hurting (G) has to be brought into a specification of the presupposition.

Perhaps the matter becomes a little clearer if we consider the natural negation of (6), viz.

(7) Someone does not hurt the one he loves.

This may be analysed by:

(7a) $\exists x(IyFxy.\neg Gxy)$.

Happily everything fits, since (7a) is equivalent to the negation of (6a), and so it has the same presupposition.

In the next section we shall consider the related problem of how natural language modes of sentence composition determine a presupposition for complex sentences in terms of the presupposition of their parts. This is really a much simpler issue, since it concerns only complete sentential units; but we shall see that different theoretical approaches to semantic presupposition can encourage different answers to the questions that arise. There will, however, be no disagreement over one mode of composition, viz. 'natural negation': the presupposition of a sentence determines the same presupposition for its natural negation. Consider, for example, (1) - (4) vis à vis (1c) - (4c).

But, it might be said, however important natural negation may be, am I not ignoring a use of such negative sentences according to which the negative construction works like the eschewed operator of 'exclusion negation', so that, for example, a sentence such as (1c) would be true provided only that (1) is not true? There are two questions here: firstly are sentences of the form 'the ϕ is not ψ ' ambiguous? Secondly, if and when some such construal were required in a semantic representation, would this force us beyond the expressive range of our logic?

Let me be brief about this. The answer to the second question is no: the conservative theorist would analyse this supposed reading of the

negative sentence using a Russellian total-valued quantifier with negation enjoying widest scope, and we could do this too. We can do anything he can do; it is the converse which is false.

However, we might wish to avoid having to discern an ambiguity. A negative reason might be that, though the alternative reading would not push us beyond the expressive resources of our logic, it could not be so neatly explained as it would be on the conservative approach, where it is simply a matter of ambiguity of scope with respect to a quantifier. But there are positive reasons too: contexts in which the exclusion-negation construal of an assertion is appropriate really do seem to be non-standard. Typically they are ones in which the speaker is being rhetorical or clever, and in which we might expect a special, or at least exaggerated, intonation pattern to indicate that 'something funny is going on'. For example: "The oldest philosopher in Patagonia is not bald! (because there is no such person)".

But is this not suggestive? Is it not in precisely such cases that we can, with a clear theoretical conscience, appeal to the standard technique for dealing with fringe examples which do not fit a semantical account: the notion of conversational implicature? There would seem to be some incomprehensible prejudice against accounting for special speaker meanings in terms of a presuppositional specification of literal meaning: proposals abound for accounting for the presence of presupposition in terms of a classical analysis, but proceeding the other way about — as we would be doing here — is never suggested. However, if we are really serious about making 'semantic presupposition' a hard-core feature of literal meaning, then what could be more natural? Indeed, our remarks in Section 1 should suggest that it is easier to climb down from presupposition than to climb up to it; for, in climbing down, we do not have the problem of explaining

why there are contexts in which another sentence, with supposedly equivalent literal meaning, does not behave in the same way.

Admittedly, before we could give a detailed account of particular cases we should first have to have some idea of how presuppositions entered into assertion making in the standard cases — i.e. of what it was for a speaker to presuppose a sentence's semantic presupposition in asserting that sentence; and we have not said anything about that yet.

We have, in this section, merely scratched the surface of many interesting issues — issues which arise even if we restrict ourselves to straight-forward sentences which do not involve any complicated embeddings in that-clauses or the like. For example, there is going to be further need to appeal to implicature to account for contexts in which a presuppositional idiom (positive or negative) is used to mean something like what its conjunctive counterpart would mean. Furthermore, we have not dealt with the grammatical complexities of presuppositional expressions: the question arises whether the-phrases, every-phrases, etc. give rise to a presupposition in all their occurrences. I believe that in fact they almost always do, but that there is an interesting class of exceptions. These exceptions then raise the question how we might do justice to the very natural idea that we should, nonetheless, give such phrases a uniform interpretation. The solution, I believe, is to become less rigid about 'syntactical categories' in our language of logical forms. It is possible to give a liberalized language a quite coherent semantics within partial-valued logic.

V.3 THE 'PROJECTION PROBLEM'

What, then, has our approach got to offer in connection with the 'projection problem' for presupposition, as this arises in the case of modes of sentence composition? The problem is to specify for a given mode of composition $\phi(p_1, \dots, p_n)$, the presupposition of a sentence $\phi(s_1, \dots, s_n)$ in terms of the presupposition of the constituent sentences s_i . One simple mode of composition is negation, and here the answer is familiar: *not-s* has the same presupposition as *s*. The general question would be solved once we had a systematic way of answering the question for each basic mode of composition, and this is the approach taken in Karttunen (1973). He considers a wide range of categories, and in particular classical kinds of sentence composition – not only negation but also 'if ... then ...', '... and ...' and '... or ...'. We shall concentrate on these.

In the first place it is important to realize that there is, on our approach, no 'projection problem' as such: the way any mode of composition affects the presupposition of a sentence in which it occurs is implicit in, and fully determined by, a single uniform specification of how that mode of composition contributes to the meaning of sentences. Projection rules could be stated – rules giving the presupposition of a compound in terms of the presuppositions of its parts – but it is not necessary to state them explicitly as part of a semantical account, since they are automatically determined.

Karttunen, on the other hand, uses a relation 'presupposes' – in this article at least – and would specify the presupposition of a sentence by specifying what sentences that sentence presupposes. Hence he requires explicit projection rules as part of this specification. Furthermore he has a general critique of logical (*sic*) approaches, which, he argues, could never hope to be subtle enough. We shall return to this later: I fail to

understand the argument and I believe that it must result from a muddle.

Concerning the projection problem, Karttunen begins by pointing out that no simple cumulative principle can be adopted: it is not correct to say that a sentence $\phi(s_1, \dots, s_n)$ presupposes a sentence s iff some s_i presupposes s . For example

(8) All Jack's children are bald.

presupposes that Jack has children, though the following sentence appears to lack any non-trivial presupposition:

(9) If Jack has children, then all Jack's children are bald.

This is, of course, no surprise to us: if we construe 'if ... then ...' as ' $\dots \rightarrow \dots$ ' (which is perhaps questionable but will do for now), and if we bear in mind that a sentence $\forall xFx.Gx$ is analysable as $\exists xFx / \forall x(Fx \rightarrow Gx)$, so that (9) can be represented in the form $\exists xFx \rightarrow (\exists xFx / \forall x(Fx \rightarrow Gx))$, then we can actually see how the presupposition disappears. For in our logic:

$$\phi \rightarrow (\phi / \psi) \simeq \phi \rightarrow \psi .$$

However there are more interesting cases. Karttunen notes that if a sentence s_1 semantically entails (*sic*) what a sentence s_2 presupposes, then, again, the conditional 'if s_1 then s_2 ' loses this presupposition. Let us consider two examples — neither Karttunen's own, but both, I think, ones he would count as examples meeting the description just given:

(10) If Jack has a wife and children, then all his children are bald.

(11) If Jack is a father, then all his children are bald.

To accommodate the behaviour not only of (9) but also of (10) and (11), Karttunen proposes the following projection rule:

"Let S stand for any sentence of the form "If A then B ".

(a) If A presupposes C , then S presupposes C .

(b) If B presupposes C , then S presupposes C unless A semantically entails C ."

Notice that this rule not only involves the relation of presupposition but also a relation of semantical entailment — another relation we might wish to avoid actually having to mention in a semantic theory. It is worth noting, too, that, although the rule is a modification of the simple cumulative idea, it is not a very radical modification of it. According to this and similar rules proposed for other modes of composition, although a sentence $\phi(s_1, \dots, s_n)$ does not have to presuppose everything presupposed by some s_i , still, it can presuppose only what some s_i presupposes. The rules merely delete presuppositions: they cannot modify the totality of sentences presupposed — which, on a relational account, determine 'the presupposition' of a sentence — in any more subtle way,

However that may be, let us now compare what our semantics dictates concerning (10) and (11). Firstly (10): if we analyse this into the form $\exists x(Fx \wedge Hx) \rightarrow (\exists x Fx / \forall x(Fx \rightarrow Gx))$, then, according to our logic, the presupposition that $\exists x Fx$ disappears just as well as it did in example (9). For, given any three formulae ϕ, ψ and χ , if $\phi \vdash \psi$, then

$$\phi \rightarrow (\psi / \chi) \simeq \phi \rightarrow \chi .$$

So, in particular, if $\phi = \exists x(Fx \wedge Hx)$, $\psi = \exists x Fx$ and $\chi = \forall x(Fx \rightarrow Gx)$, then this equivalence holds, since ψ is a consequence of ϕ . Of course example (11) cannot be explained like this: we may take (11) to be of the form $\phi \rightarrow (\psi / \chi)$, with ψ and χ as before and ϕ this time meaning "Jack is a father"; but in this case we do not have $\phi \vdash \psi$ as a matter of logic. To remedy this we might envisage adding meaning postulates in our semantic theory to the effect that fathers have children, so that $\phi \vdash \psi$ would hold. However this is not, I think, something we have to do in order to show that our treatment of presupposition is viable. And it is not necessarily something we should want to do in a revelatory semantics.

To see this, first recall that given any formula in our logic we can find an equivalent one in the form ϕ / ψ , where the only occurrence of $/$ is

the one exhibited (and there are no occurrences of \ast either). Then, assuming, of course, that no basic mode of predicate-singular-term composition gives rise to presupposition, the whole presupposition of the original formula is given, up to equivalence, by ϕ . Indeed, the reduction rules given in Section 2 for transforming a formula into this kind of normal form could be taken as our 'projection rules'. For the sake of comparison, note that our ϕ would correspond to the conjunction of the set of sentences presupposed according to Karttunen's rules. In particular, for our analysis of (11) we have :

$$\phi \rightarrow (\psi / \chi) \simeq (\phi \rightarrow \psi) / (\phi \rightarrow \chi) .$$

The presupposition of (11) can then be paraphrased by "If Jack is a father, then Jack has children"; and if this is not a logically trivial presupposition, it is at any rate sufficiently trivial – trivially true, that is – to explain our intuition that the presupposition of "All Jack's children are bald" has 'disappeared' in (11).

Now, this discussion might be taken to point up that we have been somewhat careless in our talk of 'having a presupposition' or 'not having a presupposition', or, slightly less carelessly but still rather vaguely, of having or not having a 'non-trivial presupposition': for when is a presupposition trivial or non-trivial? I feel no compulsion, however, to answer this question in general. I can tell you about the presupposition that a sentence exhibits as a result of its 'logical structure', because I have a tidy account of the truth-falsity conditions of logical modes of composition; but once the problem has been reduced to semantical relationships between lexical items, it is a different matter. May be we should require that our semantic theory say something about such relationships, but may be not. If our semantics is revelatory rather than relational, not only in that it accounts subtly enough for the contribution constituents make to compounds in which they occur, but also in that this is done by

using the language of the semantic theory to give the meaning of the object language, then it is no more a basic condition of the theory's being a semantic theory that it explicitly specify when a presupposition is trivial than that it specify explicitly when a sentence is trivially true. Indeed, as we have seen, in certain cases these two things go together.

Karttunen offered his projection rules in the first place as semantic rules, but subsequently he considers examples which he thinks show that, after all, an account of the presupposition of compounds cannot be handled in a 'purely semantic' way. However, I think that this conclusion — which Karttunen himself says he regrets — is in fact unwarranted.

The examples in question are sentences such as conditionals 'if s_1 then s_2 ' (or sentences of the form 'either s_1 or s_2 ') where s_1 does not in any sense itself entail the presupposition of s_2 , but where, according to Karttunen, the presupposition of s_2 'disappears' from the compound sentence in virtue of the fact that s_1 together with 'background assumptions' entails the presupposition of s_2 — disappears, that is, for a particular utterance. He sees this as the same kind of thing as the disappearance we had in the earlier examples; and he concludes that presupposition must be accounted for using a three-place relation ' s_1 presupposes s_2 in X ', where X is a set of (sentences representing) background assumptions, whose role has to be spelt out pragmatically. Hence he modifies condition (b) of the previous projection rule as follows:

"If B presupposes C , then S presupposes C unless there is some (possibly null) set of assumed facts such that $X \cup \{A\}$ entail C . (Constraints on X : X does not entail *not-A* and X does not entail C .)"

To illustrate the situation Karttunen uses the following (highly unsavoury) sentences:

(12) If Geraldine is a Mormon, she has given up wearing her holy underwear.

or equivalently

(12a) Either Geraldine is not a Mormon or she has given up wearing her holy underwear.

Let us consider (12) and agree with Karttunen that the consequent presupposes what could be paraphrased by

(13) Geraldine has worn holy underwear.

Now, in the example situation, our speaker is supposed to be a peeping Tom making assertions to no one in particular about what he has just seen, and so the 'background assumptions' are supposed to be nothing but the speaker's (previous) beliefs. We are encouraged to think ourselves in and out of two possible cases:

(i) The speaker has beliefs which taken together with the antecedent, viz

(14) Geraldine is a Mormon.

entail (13).

(ii) The speaker has no such beliefs. To make the contrast clearer we are asked to suppose that he believes that Mormons are in fact not supposed to wear 'holy underwear'.

In case (i), according to Karttunen, (12) carries no presupposition, while in case (ii) it carries the same presupposition as the consequent, viz. (13).

Is this how we should explain the matter? Karttunen is surely right to claim that the simple cumulative idea is in need of greater modification than he originally gave it; but he is wrong, I think, to conclude that we have to give up the idea of deriving the presupposition of a compound solely from its constituents: we do not need to introduce the parameter X , we just need to be subtle about semantics. The contextual phenomena that Karttunen is striving to explain can then be quite adequately accounted for *via* whatever general link is envisaged between sentence presupposition and linguistic practice.

But, before we consider the matter from our point of view, we might ask what general link it is that Karttunen envisages. He opened his paper by offering a notion of 'pragmatic presupposition', which he attributes to Stalnaker, and formulates by saying that to presuppose something as a speaker is to take its truth for granted and to assume that the audience does the same. The idea was then that, if a sentence *A* presupposes *B*, then in any 'sincere' utterance of *A*, *B* must be – or be entailed by – a pragmatic presupposition: hence sentence presuppositions are to correspond to sincerity conditions. This will not be our way with linguistic practice, but, for the present discussion, we might try and think in terms of it.

The question now immediately arises: are the sentences in Karttunen's set *X* actually supposed to be pragmatic presuppositions of a given context? It appears from his later article Karttunen (1974) that this is so. In this later article the whole account is in fact much smoother; for, instead of first relativising the presuppositions of a sentence *A* to a set of pragmatic presuppositions *X* and then considering whether *A* is sincerely utterable in *X*, i.e. considering whether the presuppositions of *A* (relative to *X*) are in, or follow from, *X*, he works with a basic relation '*X* satisfies-the-presuppositions-of *A*'. This approach is equivalent to the more clumsy one, since the presuppositions of *A* relative to *X* follow from *X* if and only if *X* satisfies-the-presuppositions-of *A*.

Now, it would seem to me that the smooth way of describing Karttunen's idea shows clearly that it does not in fact involve him in anything less 'semantic' than he had before: the new relation is surely no less a semantical relation than the original relation 'presupposes'. Both relations would be specified, without direct reference to linguistic practice, by induction on the complexity of sentences, and both relations alike would call for some further explanation in terms of linguistic practice. Hence,

pace Karttunen's way of describing it, what I believe our account can save him from is not a departure from a 'purely semantical' theory, but simply a departure from straight-forward projection rules for presupposition. Of course, such rules are, on a revelatory account, of secondary importance in any case: the real advantage that our approach has to offer is the possibility of avoiding explicit mention of any semantical relations at all.

Let us return to the first article and the example given there. According to our account, where Karttunen went wrong was in transforming a conditional presupposition for the compound sentence into a contextual cancelling condition for the presupposition of the consequent. For (12) is just like (11): we may analyse it as $\phi \rightarrow (\psi / \chi)$, where $\phi = (14)$, $\psi = (13)$ and χ means "Geraldine does not now wear holy underwear". But this formula is equivalent to $(\phi \rightarrow \psi) / (\phi \rightarrow \chi)$, and so we can paraphrase the presupposition of (13) by

(15) If Geraldine is a Mormon, then she has worn holy underwear.

When Karttunen considered a context of kind (i) for an utterance of (12), it was, according to him, because background assumptions together with (14) entailed (13) that presuppositions associated with the consequent disappeared — in particular (13) itself. This is the same as saying that the reason was that the background assumptions entailed "if (14) then (13)" — i.e. (15). And what it meant to say that the presupposition disappeared was that (13) ceased to be a sincerity condition. Indeed, Karttunen's revised account was motivated precisely by having to explain how, in case (i), (12) was sincerely utterable without (13) as an assumption.

But why, then, not simply say that (15) gives the (total) presupposition of (12) and that it was because (15) was, or followed from, a background assumption that the utterance of (12) was sincere? The way he set up the example distorted the situation: presuppositions seemed to disappear altogether because we were given the truth of (15) and asked what

further assumptions would then be required of a sincere utterance of (12). Clearly none. The presupposition seemed to 'disappear' simply because it was held true — not logically true, nor even 'semantically true', but true by *fiat*.

What, however, of case (ii), where Karttunen says that the presupposition of (12) is (13)? He does not invent a complete story for this case, but we can surely argue that, since we have one kind of context in which the presupposition is as weak as (15), then, whatever stronger pragmatic presupposition or the like there might have been temptation to associate with the use of (12) in other kinds of context, only the minimum presupposition (15) need actually be associated with the sentence itself as a condition for its 'sincere utterance'. The only way Karttunen could object to this would be to find a context in which there was the pragmatic presupposition that (15) held, yet in which an utterance of (14) was insincere because (13) was not a pragmatic presupposition. Otherwise, according to Karttunen's idea of presuppositions as sincerity conditions, our specification of the sentence presupposition would be quite unexceptionable. This would be somewhat difficult, however, since, if (15) is a pragmatic presupposition but (13) is not, then, according to his cancelling conditions, (12) would be totally presuppositionless. At least this would be so unless the negation of (14) were a pragmatic presupposition: however, such an extreme case was not what Karttunen seemed to have in mind for case (ii). We shall not stop to consider what might in fact be said about this extreme case — where a speaker is pragmatically presupposing the negation of the antecedent of a sentence he is asserting — because we shall not in fact be adopting Karttunen's envisaged link between sentence presupposition and linguistic practice.

To sum up the difference between Karttunen and me, we might say that Karttunen draws a sharp contrast between sentences (9), (10) and (11), on the one hand, whose constituent presuppositions are taken to disappear as

a matter of semantics, and (12), on the other, where they do not; whereas I would draw a contrast — but a much less sharp one — between (9) and (10) where the constituent presupposition disappears as a matter of logic and (11) and (12) where it does not disappear but is modified (weakened).

These differences stem from the general difference of approach. Let us note first that Karttunen is running his definitions directly on surface sentences. At least, sentences of the left of 'presupposes' are surface sentences; though it becomes clear in the second article that the sentences presupposed and the sentences which specify background assumptions are 'logical forms'. On our approach, in contrast, the presupposition of a sentence is conceived of as an aspect of literal meaning which is to be embodied within a single all-encompassing semantic representation; and our logic shows how this may be done. It does not, of course, show how to relate surface sentences with these presuppositional semantic representations, and I do not know enough about grammar to comment on the matter; nonetheless, it would not seem to me to be too much of an intrusion into the linguist's territory to suggest that this is what he should try to find rules to do. That is to say, instead of having to provide sentences with a total-valued logical form, he has to provide them with a partial-valued one.

Apart from any grand views about what semantics should look like, I believe that our approach has the advantage of enabling us to avoid adhocery, since it provides something systematic to appeal to when intuitions are unclear concerning particularly complicated examples. Of course this remark presupposes that there is some intuitively clear fit between a logical form and a surface sentence that we wish to do justice to in the final account. But, as an example of what I mean, consider the problem — which we touched on in the last section — of accounting for the presupposition of a compound sentence which contains presuppositional quantifiers not

simply compounded with modes of sentence composition but actually embedded within the scope of one another. How would the Karttunen approach handle these cases?

Finally, we might note a minor difference of detail which is connected with the difference of approach. Karttunen gives 'and' and 'or' an asymmetrical treatment, whereas we do not. However Karttunen himself dismisses the importance of this, and I think we may do so too.

Let us in fact consider 'and', for we have not done so yet, and it is in connection with conjunctions that Karttunen makes his puzzling criticism of the 'logical approach'. First, observe that the projection problem for conjunctions is directly parallel with that of conditionals. It is so on Karttunen's account, since he proposes precisely the same cancelling conditions and subsequently modifies them in precisely the same way; and it is so on our account, since we have the following pair of equivalences:

$$\phi \wedge (\psi / \chi) \simeq (\phi \rightarrow \psi) / (\phi \wedge \chi)$$

$$\phi \rightarrow (\psi / \chi) \simeq (\phi \rightarrow \psi) / (\phi \rightarrow \chi)$$

Now, two sentences which could be analysed in the form $\phi \wedge (\psi / \chi)$ are:

(16) Paris is the capital of France, and the King of France is bald.

(17) Marseilles is the capital of France, and the King of France is bald.

Karttunen considers Lukasiewicz' truth-table for \wedge (which is like ours) and a 'quasi-truth-table' derived from van Fraassen. He goes on to say, concerning (16) and (17):

"Assuming the facts as we know them to be, in Lukasiewicz' and van Fraassen's system (16) presupposes that France has a King, since the sentence is neither true nor false in the case the

King does not exist. On the other hand, given the actual state of affairs, (17) in their logic does not presuppose the existence of the King, since the falsehood of the first conjunct is sufficient to make the conjunction bivalent. From the point of view of ordinary language, this outcome is definitely unacceptable. Relative to our actual world, where the form of government a country has is not determined by the choice of capital, both sentences surely presuppose that France has a King.

In general, whether or not a presupposition of a particular constituent gets filtered out in (16) and (17) depends [i.e. on the logicians' account — supposedly] on the truth value of the other constituent, not on the semantic relation between them as the case seems to be in ordinary language."

There would, however, appear to be some muddle here: perhaps a muddle between, on the one hand, the use of truth-tables (or other semantic devices) to work out the truth-value classification of a sentence or formula under a particular interpretation, and, on the other, the use of truth-tables (or other semantic devices) to provide constraints on the range of possible interpretations of sentences or formulae. Generalizing over interpretations subject to such constraints can clearly provide us with non-trivial semantic relations. One such relation is 'presupposes'; and its interrelationship with other semantic relations can be systematically and non-trivially revealed. It may indeed be precisely in virtue of interpretations provided by truth-tables that semantic relations are what they are: for example the truth-table for \neg explains why $\neg\phi$ and ϕ are contradictory. Ignoring, or muddling, this role of truth-tables is especially exasperating from our particular point of view, since the semantic relations that can be defined — or better: the semantical relationships that can be revealed — in virtue of truth-table interpretations is so vastly enriched when we have \times and / at our disposal.

What presuppositional logician would ever determine a notion of presupposition relative to one particular interpretation or assignment of truth-values, let alone relative to one interpretation of one conjunct of a conjunction but not the other (whatever that could mean)? Yet this is what Kattrunen appears to have in mind. He surely has no right to criticise presuppositional logic for not doing adequately what it does not set out to do.

The only reasonable objection he could make would be an objection to the appropriateness of the presupposition which a given logic determines, in its proper way, for given conjunctions. No doubt he would in fact object to what our logic dictates concerning (16) and (17); for, according to him, the presupposition of both (16) and (17) is that there is a King of France (though in saying that he seems to have forgotten that it follows from his previous thesis that this presupposition might disappear altogether in certain contexts); whereas, according to our logic, the presupposition is that if Paris (respectively Marseilles) is the capital of France, then there is a King of France. However the debate about the right thing to say here is exactly parallel to the case of conditionals, and there is no need to run through it all again.

V.4 PRESUPPOSITION AND ASSERTION

What then of speaker presupposition? Recall that we are hoping for some account of speaker presupposition as a linguistic sub-act of the act of assertion, so that we can say that, in straight-forward cases, in using s to make an assertion a speaker presupposes that p , where 'that p ' is a report of the semantic presupposition — i.e. truth-value preconditions — of the sentence s . Of course we do not want simply to appeal to the proposed semantic details. We cannot say that a speaker presupposes that p in making an assertion with s if and only if the truth-value preconditions of s are that p . We want to have a picture of how presupposing that p

enters into asserting that q which is independent of semantic details; though, of course, we want one which fits with those details — in particular one which fits in accordance with the naive link formula described in the introductory section.

Now such a notion of speaker presupposition would be very deeply rooted in assertion making. Should we work with such a notion? And could it be so deep without disappearing altogether?

The first question first. Recall that Karttunen, not an enemy (initially at least) to the idea of seeing presuppositional phenomena catered for in a semantical account, envisaged a link with linguistic practice by invoking Stalnaker's general notion of 'pragmatic presupposition'. Stalnaker's own initial formulation in Stalnaker (1974) is:

"A proposition P is a pragmatic presupposition of a speaker in a given context just in case the speaker assumes or believes that P , assumes or believes that his addressee assumes or believes that P , and assumes or believes that his addressee recognizes that he is making these assumptions, or has these beliefs."

This definition is quite autonomous — independent of any kind of linguistic act. But Karttunen's idea was that a specification of the semantic presupposition(s) of a sentence was a specification of sincerity conditions; conditions such that only if a speaker were pragmatically presupposing the semantic presupposition would he be sincerely uttering that sentence. Stalnaker himself considers such a possibility, though in place of sincerity conditions we have the more neutral idea of 'conversational acceptability'. But note that these accounts are very far from providing a notion of speaker presupposition which is a part of assertion-making, even if — and if so, in whatever way — a speaker might, in making an assertoric attempt with a given sentence, be described as doing something which an audience could recognize as presupposing that p .

In any case, this link with sentence presupposition is not the primary role of Stalnaker's notion of pragmatic presupposition. Moreover, he does not favour accounting for sentence presupposition in terms of a triclassificatory semantics: he thinks that this approach might cause unnecessary complications; and anyway it would obscure the explanatory uniformity he seeks from pragmatic presupposition. If there is a case to be made out for given presupposition's being conventionally associated with a given word or kind of idiom, then he would seem rather to favour the idea of keeping a specification of this presupposition distinct from the semantic representation proper of the word or idiom. He would distinguish between the 'content' of a sentence (given by total two-valued truth conditions) and its overall 'meaning' which is sensitive to any conventionally associated presupposition (given by context-conditions using his definition of pragmatic presupposition). Hence he would seem to fall into the category of theorists I characterized in Section 1 as using some kind of *ad hoc* tagging.

This account is seen as running alongside other cases of presupposition which arise in context, not in virtue of the 'meaning' of any linguistic item, but by mechanisms of pure implicature: content and context might alone be such that it would be reasonable to infer that the speaker were pragmatically presupposing so-and-so.

However even such cases are not really near the heart of the basic notion of pragmatic presupposition: in itself it is not something which is conveyed in any way, either in virtue of the meaning of sentences or by implicature: it is something which is seen as providing the background to a linguistic transaction rather than being a part of it. Admittedly, to accommodate situations (such as those that commonly arise because of sentence meaning or implicature) where a speaker is acting as if he believed

that certain assumptions were a common background, though obviously they are not in fact, Stalnaker offers a modified definition of pragmatic presupposition: it becomes a linguistic disposition — a disposition to behave in one's use of language as if one were pragmatically presupposing in the original sense. But this is still a back-drop to the linguistic stage, and its link with the use of presuppositional idioms must be described in terms of conversational acceptability.

This is not the end of the story, however, since the notion of conversational acceptability is itself to be spelt out in terms of the general notion of pragmatic presupposition together with an account of particular speech acts. Stalnaker is intentionally vague about this, but we might consider what this idea implies in the particular case of assertion. He would seem to take the use of a sentence to make an assertion to be conversationally acceptable just in case it can reasonably be expected to accomplish its 'normal purposes' in a 'normal way'; and pragmatic presupposition is supposed to enter into an account of this — presumably in the form of conditions stating that only if such-and-such is being pragmatically presupposed can such-and-such assertoric attempt be expected to accomplish its normal purposes in a normal way. It might, then, be a consequence of this kind of account that the use of a presuppositional idiom is such that unless its presupposition is pragmatically presupposed, it does not, or might not, constitute a totally successful act of assertion — because conditions standardly required for it to accomplish its purposes are absent. If this is not a consequence, then the account is just too complicated as well as too vague to help in answering straight-forward questions concerning how the use of presuppositional idioms is constrained; but if it is, then I have a worry: will not the condition that a sentence's presupposition be pragmatically presupposed turn out to be a trivial success condition? For will it not be the case that if an assertoric attempt with a presuppositional idiom has got off the ground at all, then a pragmatic presupposition

of the sentence presupposition — or something no less effective — will automatically be present; while such a pragmatic presupposition surely cannot be any help in actually getting off the ground in the first place.

To see why this may be so, we might do well to concentrate more on the audience than on the speaker — for I take it that being assertorically effective must have something to do with an actual or possible audience. To be brief: firstly, how could a speaker, provided simply that he was recognisably attempting to make a straight-forward assertion with an identifiable sentence, fail to be taken to be pragmatically presupposing any sentence presupposition carried by his sentence — taken to be, that is, by any linguistically competent audience? For if he were not so doing, then, according to the account we have extrapolated from Stalnaker, this would conflict with the hypothesis that he was (recognisably) attempting a straight-forward assertion, which, presumably, he would want to be totally effective. A competent audience would appreciate this, even if they did not think it out explicitly. But then, secondly, how could this be any less effective to the proper accomplishment of his assertion than actually pragmatically presupposing? The effectiveness of an (attempt at) assertion — which pragmatic presupposition is supposed to facilitate — must surely depend on an audience's grasp of what the speaker is about. But pragmatic presupposition is merely a disposition; and it cannot but be more important for a speaker to be taken to have a disposition than simply to have it. Indeed, as far as an audience's grasp of the speaker's immediate activity is concerned, actually having it is irrelevant, unless it is manifest.

We have still to consider the possibility that a pragmatic presupposition of a sentence's presupposition is required before an utterance of the string of words that make up a sentence could be taken to be an assertoric attempt at all; but this is surely incredibly implausible. What contribution, for example, could your pragmatic presupposition that Jack

had children possibly make to my appreciation of the fact that you were trying to make an assertion with the words "All Jack's children are bald"?

This Stalnakerish tangle would seem to give support to the idea that we should, as far as an account of the use of presuppositional idioms is concerned, look directly for a notion of speaker presupposition which is, or is part of, a particular linguistic act. Such a notion might indeed turn out to be in some way a species of a more general notion of speaker presupposition such as Stalnaker's; but our initial concern is narrower. And we can leave till the distant future the question whether Stalnaker is right to suppose that a general notion could ultimately explain all presuppositional phenomena. There would be a lot of explaining to do.

The recognition of a presupposition in an assertoric attempt with a presuppositional idiom is, then, something that we wish to take for granted in our account of speaker presupposition — the recognition, that is, that something has been presupposed and of what has been presupposed. Our job is to say what role in an understanding of the speaker that recognition plays. So, might an audience's beliefs concerning whether or not the presupposition holds be relevant to the effect an (attempt at) assertion has on him? I believe so; but I do not believe that this actually affects the accomplishment of an act of assertion. That is to say, it is not, I think, fruitful to work with a definition of assertion which implies that such a linguistic act, performed using a presuppositional idiom, is in any sense not fully accomplished unless the audience believes that the associated presupposition holds. For it is not clear to me how any attitudes towards the content of a presupposition could play an additional role in an understanding of what the speaker is about, once what the presupposition is has been recognized.

Admittedly an audience with strong convictions that a presupposition fails would be inclined to react "Oh but, you can't say that because ...", a response which contrasts with a simple "No, you're wrong" as markedly as it does from "Yes, you're right"; and this is an important piece of linguistic data. But, even so, should we not take such a response to be an evaluation of a fully accomplished assertion, rather than as an expression of frustration at an incompetent performance: an expression of frustration it is, let us call it the 'third response', but it is a frustration which the audience experiences because they fully understand the speaker, not because they do not understand. Contrast, perhaps, the case when an audience fails to pick up an indexical reference — or, indeed, an indexical relativization of a quantifier expression.

What then of the 'no assertion' view of presupposition failure, which we mentioned in Section 1? Well, if the accomplishment of an assertion is unaffected by the convictions of the audience concerning the presupposition, then *a fortiori* it should be unaffected by the objective truth or otherwise of the presupposition.

But what is the relevant kind of understanding that an audience must have? Clearly it is not enough that it is understood that an assertion has been made; the audience must understand what has been asserted — recognize the content of the assertion. And, to take the previous ideas a stage further, I think we should insist that what is asserted using a presuppositional idiom is not in any way affected either by the convictions of the audience or by objective matters of fact concerning what is presupposed. It is not as if your assertion using the words "The oldest philosopher in Patagonia is bald" is reduced to a trivial blank if I do not believe in such a person, or if such a person does not in fact exist.

We might pause to notice that these views fit very happily with our semantic proposals: * is not to be glossed as (strictly speaking) meaningless;

and truth-value preconditions are not preconditions for a sentence's taking on a 'proper' propositional content. Contrast the ideas of Belnap and van Fraassen which we discussed in Section I.3. Contrast, also, the informal gloss on presupposition failure provided in Keenan (1971): this is in spite of the fact that Keenan's semantic proposals — e.g. in Keenan (1973) — are formally speaking more like ours.

However semantics is to be kept in its proper place. We opened this sub-section by saying that recognition of a speaker's presupposition by the audience was something that our account was to take for granted. As a result of the foregoing discussion, we can close by saying that, as far as presupposition is concerned, the recognition of an assertion is also something we can take for granted. Indeed, the claim which is to follow will entail that it is in virtue of the recognition of an assertion that a presupposition is recognized.

So, then, we must ask: what effect does a presupposition have on an assertion? The suggestion is that the relevance of presupposition is sufficiently deep-rooted that talk of 'affecting' assertions is quite misplaced. The linguistic act of speaker presupposition is to be seen as constitutive of assertion making, and what is presupposed is to be seen as constitutive of what is asserted. Presupposition does not run alongside assertion any more than it is prior to it; it is part of it.

In the first place, then, I see what is presupposed (the content of the presupposition) as a function of what is asserted (the content of the assertion). We may, modulo equivalence, take two different assertions to involve the same presupposition; for example assertions made using "All Jack's children are bald" and "All Jack's children play poker" would both involve the presupposition that Jack has children, but, I argue, you cannot

have the same assertion along with different presuppositions. Furthermore there is a partial converse to this: though a specification of what has been presupposed does not uniquely determine an assertion — since, for example, the two assertions we have just considered are different — nonetheless, the range of assertions which can give rise to a given presupposition is restricted: that Jack has children clearly cannot be what is presupposed in an assertion using "It's Jack who beats Mary".

How, then, should we expect to specify assertions so as to reveal their non-independence from presupposition? If we have already specified a presupposition, for example the presuppositions (1a) - (4a) associated with (1) - (4), then it is natural to say that 'under' these presuppositions the assertions made are that he beats Mary, that they beat Mary, etc. Clearly these specifications are parasitic on the prior specification of the presuppositions, but if we want to specify an assertion all in one go, we can do nothing better than put 'that' in front of the sentence used (and make any necessary grammatical modifications). This seems obvious, but, perhaps in an effort to avoid the look of triviality, presuppositionalists often seem tempted to go in for something more complicated. Consider for example the possibility of taking the following sentences to be presuppositional:

(18) It is still raining.

(19) It is already raining.

The proposal might be, pushing the necessary tense logic under the carpet, that an assertion of (18) would involve the presupposition that it has been raining, and an assertion of (19) that it is about to rain. These are different presuppositions, but it might be said that what is asserted is the same, viz. that it is now raining. Now, if the presupposition has already been reported, then this way of reporting an assertion of (18) or (19) is appropriate enough, but, according to our account, this will not do as a

complete and independent specification of the assertions made. What is asserted is that it is still raining, or that it is already raining — or, if we are not in the speakers contextual shoes, what was asserted is that it was still raining at the time and place of utterance, or that it was already raining at the time and place of utterance.

This is the framework. So we must ask in what way presupposing makes assertion a more complicated activity than it would otherwise be, and in what way the content of a presupposition enriches the content of an assertion which it is part of. Most of our previous remarks are consistent with the idea that the presupposition of an assertion is really nothing more than a conjunctive constituent; and so it is now going to be an up-hill struggle to distinguish assertions with a (non-trivial) presupposition, such as we claim are made using presuppositional idioms, from assertions such as those made using their conjunctive counterparts.

Of course the way we wish to see the content of an assertion specified is in terms of a three-fold scheme of evaluation which is to correspond — in straight-forward cases — directly with the three-fold semantic scheme. These would be objective classifications used to determine what an audience understands when an assertion is made; but an account of the point of specifying assertion content in this way should involve no more than an account of why an audience, having understood an assertion, would, depending on their convictions concerning how things objectively are, respond sometimes 'yes', sometimes 'no', but at other times use the third response indicating frustration. Of course these responses themselves are not the only datum we have concerning what an audience understands by an assertion: there are more complicated things to say, which may shed light on the presuppositional richness of assertions and show what it is we are capturing by a scheme of evaluation that mimics a three-fold style of response.

But before we turn finally to this we might ask how exactly the abstract semantical features of our partial-valued logic may be relevant to the envisaged account. In Chapter I our approach was contrasted on the one hand with those according to which the third classification was construed as 'meaningless', and on the other with those according to which the third classification was taken to be a semantic value on a par with 'true' or 'false'. Now, we have already seen that the first contrast is appropriate for an analysis of presuppositional idioms, since, on our view of assertion making at least, the failure of a presupposition does not in any way affect the proper accomplishment of an assertion nor an understanding of what is said. What, however, of the second contrast? This is more difficult. But in the first place recall that the weight of this claim — that we had two values and a gap — was put on the expressive range of our language (*via* the idea of functional value-dependency). Hence one way to see the appropriateness of this second contrast is simply to observe that, at least as far as accounting for presupposition is concerned, we do not need to invoke any analyses which would take us out of an expressive range within which we can remain partial-two-valued rather than three-valued. (Indeed, I think it quite plausible to claim that no mode of natural language composition forces us out of two-valued logic. Modes such as 'it is true that ...' and 'it is false that ...', which might be taken to require a non-monotonic interpretation in terms of a three-entry matrix, and hence force recognition of the third classification as a semantic value, are perhaps better accounted for as intensional modes. These would require a quite different kind of interpretation — in the same framework in which a consequence-encoding conditional would have to be interpreted.)

Nonetheless we might ask why this is so. And we might ask for some distinction between the evaluations 'true' and 'false' and the third one which would further spell out the idea of values versus gap. However I am not sure that I can offer anything very substantial to meet these demands.

Certainly, I hope that the remarks which follow will point up some distinction between YES/TRUE and NO/FALSE on the one hand, and the third response/classification on the other; but then there is clearly going to be a distinction with YES/TRUE on one side, and both NO/FALSE and the third response/classification on the other. Why the first distinction should be taken as one corresponding to presence versus lack of truth-value and the second not, is a question which I feel may have a deep answer, but, if so, I do not know how to describe it. Officially, then, the idea of values versus gap is to remain a purely semantical distinction, and its relevance to natural language is *via* expressive requirements. Why we are inclined to use the labels 'true' and 'false' for two of the classifications but dismiss the third as just being neither of the other two, is another matter — and a more trivial one.

How, then, are presuppositional idioms used? What does an audience make of their use and what should the theorist make of it? Let us enter the discussion in a leisurely way.

In the first place it would seem that, since we are seeking an account of speaker presupposition which is a sub-linguistic-act of assertion, there is no reason not to accommodate within the basic account all those intuitively quite standard cases in which, in virtue of making an assertion with a presuppositional idiom, a speaker might be said to convey the information that p , where 'that p ' is a specification of the presupposition of his sentence. Or, if this turns out to be too strong a way of describing such a situation, let us at least observe that as a result of an assertion an audience may come by the belief that the associated presupposition holds, though they did not previously have any such belief, and that this is something a speaker might be taken to expect, or even intend. For example, I might announce:

(20) The undergraduate living above me keeps me awake at night. before an audience who has no prior beliefs concerning my domestic situation and who I know has no such beliefs. As a result of my assertion they would presumably come by the belief that I had an undergraduate living above me, and this is something I would expect.

What, then, is the difference between assertions made using presuppositional idioms and ones made using conjunctive counterparts? To redress the balance, we might consider what limits there are on the use of presuppositions to convey beliefs. Examples are not hard to find. Say I make an assertion to an Oxford audience — who know, let us suppose, that I live in college, but know nothing further. First consider the case where I use the sentence:

(21) The Cowley-worker living above me keeps me awake at night. To speak very vaguely, this utterance would surely seem odd — indeed 'conversationally unacceptable'. Let us assume that there is in fact a Cowley worker living above me — on a conference, say. But this makes no difference. There would, in contrast, be no such oddity if I used the sentence:

(21a) There is a Cowley-worker living above me and he keeps me awake at night.

It would seem that using (21a) would give due assertoric weight to an unusual fact, whereas using (21) would be to sneak in a surprising fact in an underhand way.

In the case of (21), before my audience were prepared to listen to what I had to say about the Cowley-worker, they would require some more to be said about his very existence. At least, this would be so if they took me seriously: a sophisticated audience would perhaps be more likely to take me to be implicating something — something rude(?) — about my upstairs neighbour, in virtue of having used a description they assume to be literally

speaking inaccurate. The moral seems to be: do not presuppose surprising facts, otherwise your linguistic intentions are likely to be thwarted or to go astray.

Perhaps a different example will suggest something more substantial. Consider a maths lecturer who is not actually going into systematic proofs but providing an informal survey. He asserts:

(22) The l.u.b. of X is less than a .

Now it is surely only if he has previously demonstrated, or even just asserted, the existence of a l.u.b. for the set X , or if its existence follows directly from a theorem familiar to his audience or is trivially obvious, that he can properly state his result in this way: otherwise a member of the audience might rightly interject " X does have a l.u.b., does it?". We may suppose that there is no reason to disbelieve in its existence, and that the lecturer is known to be a sufficiently competent mathematician not to go around presupposing things that do not obtain, but still he would not be entitled to utter (22) under conditions other than those we have specified. On the other hand it would be quite alright to utter:

(22a) X has a l.u.b. and it is less than a .

According to this story the existence of a l.u.b. is not even a surprising fact, but it is one which is at least felt to require justification. It would seem that a fact like this ought not to be presupposed, though it can quite naturally be asserted — on its own, or as a conjunctive constituent.

To account for these examples and to explain the point of having in a language the resources to make presuppositional assertions, we might be tempted to see things in the following kind of way. Think first just about assertion. Making assertions is a risky business: an audience might disagree, and if that happens a debate might ensue. In fact, should we not say

that the practice of assertion making is — is known by competent language users to be — an activity which by its very nature might, under certain circumstances, be expected to give rise to a debate or to an investigation, viz. when an audience is sceptical or has contrary convictions? And should we not say that involved in making assertions is a commitment to be able to provide justification for what is asserted, or at least to put up some kind of case? Indeed the very fact that this is so would explain why so often no debate is in fact entered into: serious asserters are taken to be able to justify what they assert.

Now in the context of this feature of assertion, might we not see what is presupposed as a fact — or a supposed fact — which the speaker puts forward as an assumption which may be taken for granted in any ensuing debate or investigation concerning what he has asserted: a fact, or supposed fact, which both he and his audience may appeal to? An asserter, then, would certainly be seen as committing himself to the truth of what he presupposes, and may be indeed to the fact that the presupposition is justified, but he is not committing himself to the expectation of having, if required, to provide justification. We might go further and say that he is in fact to be taken to be getting his audience to accept what he presupposes as not itself requiring a justification or any supporting argument. Hence a speaker's commitment to the presupposition of an assertion might in fact be regarded as a stronger commitment than his commitment to the overall assertion, but this 'commitment' is not a 'claim'.

We might attempt to encapsulate the preceding rather unnaturally vivid and dramatic ideas precisely by invoking a distinction between committing yourself to a proposition and claiming it: what is asserted is actually claimed; what is presupposed is not claimed, though it is certainly something to which a speaker would be committed. Of course, according to

our thesis that a presupposition is constitutive of the content of an assertion, the claim made by a speaker must itself be dependent upon the commitment to the presupposition. But this fits perfectly with our more dramatic account, since an assertion is a claim backed up by the possibility of an argument or investigation in which the presupposition is accepted as a fact. (And, though there is no space to go into the matter here, it is worth mentioning that the most natural approach to a natural deduction system for our logic fits very well with these ideas.)

However, we have been concentrating too much on the speaker: what about the audience? In the first place, a three-fold style of response might now be seen to fall into place. It would be natural to take any debate provoked by an utterance as a debate about the 'truth' or 'falsity' of what has been asserted; in the limiting case this might just be a childish expression of conviction: 'isn't', 'is', 'isn't, This is how 'no, wrong' and 'yes, right' are opposed. If, however, an audience is presented with facts, or supposed facts, as premises for rational debate, or even just rational comment, which they believe are not facts, then there is not disagreement but frustration, and the third response is appropriate. This informal explanation of course also accommodates the subtler cases we have considered, where, even with no contrary convictions, an audience might want prior assurance concerning the presupposition before they consider the assertion — either to argue about or just to store away as information received. We shall turn shortly to what this might mean.

But to sum up so far: as a (rough) account of when, in making an assertion, a speaker is thereby presupposing that p , I should like to propose that this is the case when his assertion is such that he is to be taken to be in a position to provide justification, given that p , for what he said; we should not require that the speaker is to be taken to be assuming that there is already anything like a 'mutual belief' that p between him and his

audience, though, of course, he is to be taken to expect that, having made his assertion, if not before, there will be a mutual belief that p ; but then we must also stipulate that he is to be taken not to expect that he will be called upon to justify this belief that p .

This is how I should like to see presupposition entering into assertion making, and our semantics is then precisely what we need to show how the content of an assertion may be represented.

I have recently been tempted to take these ideas a stage further. The inspiration for this final move comes from the view – compellingly urged by Gareth Evans in seminars – that the institution of assertion making is a vehicle for transmitting knowledge, not just beliefs. So it might be said that, in virtue of recognizing that someone had asserted that p , an audience may thereby come to know that p and, on the authority of the original speaker, go on to tell someone else that p . Of course, actually to come to know that p , the audience must take the assertion to be serious and reliable, and they must be correct in so doing. For, indeed, it must in fact be the case that p ; and the audience must be convinced of this (which they would not be if they had countervailing beliefs which were sufficiently strong).

This general view is one I find attractive, and it certainly fits with what I have urged concerning assertion making and the speaker's responsibility to be in a position to justify what he says. However, presupposition complicates the picture; but in what I believe is a very interesting way.

The first thing I want to urge is that knowledge that p cannot be transmitted by an assertion whose presupposition is that p , and that here we have a stark contrast between presuppositional idioms and their conjunctive counterparts. This is to say that knowledge cannot be acquired simply

in virtue of recognizing a presupposition, as it can be, under favourable circumstances, simply in virtue of recognizing an assertion or a conjunctive part of an assertion. Consider, for example, assertions made using

(23) Our woman fellow enjoys bananas.

(23a) We have a woman fellow and she enjoys bananas.

Let us assume that a hearer has no convictions about anything relevant but is trusting and eager for knowledge. I claim that if (23) were uttered he would be wrong to take himself thereby to have picked up the knowledge that we had a woman fellow, though he could do so if (23a) had been used. This distinction is in need of careful qualification, but it might be seen as lying behind the idea that what is presupposed does not have the proper assertoric weight of an assertion.

A more revealing example should be the story we considered before concerning the maths lecturer. Here the audience was passively accumulating knowledge, from a totally reliable source, only, when a presuppositional idiom – (22) – was used, they demanded a justification of the presupposition before they were prepared to add the content of the assertion to their stock of facts about the subject. This seems to be exactly in accord with the idea that a presuppositional assertion does not in itself provide for the transmission of knowledge of the presupposition.

This example takes us on to the next – and more interesting – question: for what are we to say about knowledge of the content of assertions themselves, when they are presuppositional assertions? Is there an interesting difference between the kinds of condition under which we would want to say that in virtue of recognizing a presuppositional assertion that p , an audience thereby came to know that p , and the kind of conditions under which we would want to say this concerning an assertion with no non-trivial presupposition?

Here we must tread carefully, I think, for there is something not

entirely straight-forward about making knowledge attributions using 'that p ' where p is presuppositional — there is something strangely 'incomplete'. In the first place, to dismiss an irrelevancy, we might note that, in ordinary usage, a sentence such as "John knows that the oldest philosopher in Oxford is bald" might often best be construed with the definite description having wide scope, so that no presupposition remains within the scope of 'that'. But if we ignore such readings and take all of 'the oldest philosopher in Oxford is bald' to be part of what knowledge is attributed, then the peculiarity has not disappeared. Perhaps we should again appeal to an influence from ordinary usage and say that sentences of the form ' x knows that p ' presuppose that x knows that q where q is the presupposition of p . (Note: this has nothing to do with the claim that such a sentence carries the presupposition that p .) Having observed this, we might then, as theorists considering knowledge attributions, want to avoid a problem by stipulating that ' x knows that p ' is to mean that x knows that- p -as-a-truth — i.e. something tantamount to the claim that x knows that p_{cc} , where 'that p_{cc} ' is a specification of the content of the conjunctive counterpart of a sentence whose content 'that p ' specifies.

Let us, boring though it may be, in fact adopt this meaning for ' x knows that p ' for the time being. However I think we are still left with something deeper than an influence from natural language sentences containing the word 'know', if we are considering those knowledge attributions which we might be tempted to make as resulting from a recognition of an assertion that p . Indeed, adopting this artificial use of 'knows that p ' points up the issue more sharply. To explain this, and at the same time to allow presupposition to fall into place, I suggest that we should say the

following: unless an audience already knows that q , where q is the presupposition of an assertion that p , they cannot acquire the knowledge that p as a (direct) result of the assertion; though if they do know that q , then knowledge acquisition may result from an assertion that p under the same conditions that it would if it had been a non-presuppositional assertion. (Of course 'already' should not be taken too temporally, since if an audience came to know that q later than an assertion that p , they could still at that later time come by the knowledge that p as a result of the assertion, providing only that they remembered it.)

Presuppositions would then be conditions knowledge of which would allow assertions to transmit knowledge which they would not otherwise transmit. Perhaps, as a necessary and sufficient condition for when a speaker S presupposes that q in asserting that p , we might say: 'an audience is in a position to acquire the knowledge that p in virtue of S 's assertion if and only if they know that q '. By an audience's being in a position to acquire an item of knowledge in virtue of an assertion, I mean that any claim they made to have acquired that knowledge in virtue of that assertion would stand or fall on precisely the same grounds as the claim that the assertion was serious and reliable. As it stands this formulation makes q 'the whole presupposition' of p : if 'if and only if' were changed to 'only if', then q would cover any conjunctive part of the whole presupposition. This is not of course a proper definition of presupposition, but rather a constraint relating presupposition and assertion; and in any case it is rather vague. But we should not, I think, be scared to relate presupposition and knowledge in some such way as this.

Finally we might now notice how appropriate the non-artificial use of 'know' becomes as a word for reporting knowledge picked up from a presuppositional assertion — I mean the use which presupposes knowledge of any

presupposition in the scope of 'know'. If, for example, I say that someone knows that all his tutor's children are allergic to buttercups, on the grounds that he was present and attending when his tutor said "All my children are allergic to buttercups", then my attribution presupposes precisely what, according to our suggestion, is the precondition for his having acquired that knowledge from his tutor's assertion.

But does the account really work? Recall that earlier we said that quite standardly a presupposition that q could convey the belief that p . In fact 'conveying information' were the words that first sprang to mind, and they were only subsequently modified through a sense of caution. Hence, it might be argued that, even if we were to follow the letter of the present account, and agree to say, first, that knowledge that q cannot be picked up directly from an assertion that p whose presupposition was that q , and, further, that it is only once there is knowledge that q that knowledge that p can be acquired in virtue of the assertion that p , then even so, this would be nothing but an otiose theoretical elaboration of the situation; since, if conditions for knowledge acquisition are otherwise favourable, then how could an audience fail to acquire the knowledge that q , if only 'indirectly', and hence be in a position immediately to come by the knowledge that p ? The indirect knowledge would result from appreciating that the speaker was serious and well informed (as ex hypothesi he must be) and hence would not be presupposing something without grounds for doing so.

As an account of what happens in very many cases where an audience does not already know that a presupposition holds, I suspect that this is the right kind of story to tell: consider the first of the example contexts we invented earlier — for sentence (20). The subsequent examples do more to back up our present claim, but even so, they do not perhaps take us far

enough. In the case of the maths lecturer the audience was totally at the speaker's disposal, trusting every word, and what I found so suggestive about the example was that, in spite of this, the audience still demanded justification for the presupposition before they would accept the assertion: in other words, they wanted to come by the knowledge that there was a l.u.b. for the set X (the presupposition) in a proper way, before they would add the fact that it was greater than a (the assertion) to their stock of knowledge. Nonetheless, this example has its limitations, since, if the lecturer dropped dead and so was unable subsequently to assert, let alone demonstrate, that X had a l.u.b., even so, would we want to deny that the audience left the lecture-room knowing that in fact it did? Perhaps not. Hence, while the example is most certainly suggestive of at least a distinction between direct and indirect knowledge acquisition from linguistic activity, it does not provide a convincing example to show that something can be presupposed and not be conveyed as knowledge at all, even though, in the same context, it would have been conveyed as knowledge, had it been asserted. An example of this is something we might feel obliged to provide: a really clear example would be one where the audience did not even pick up a belief that the presupposition held.

We shall have to consider a case where the existing beliefs and convictions of the audience play a crucial role. We shall require that their view of how things are is balanced between, on the one hand, a state in which, whether or not they already believe the presupposition, or even if they are inclined to disbelieve it, they would, at any rate, be (justifiably) swung into accepting it as a fact once the utterance has been made, and, on the other hand, a state in which they are in any case sufficiently strongly convinced that the presupposition does not hold for them not to accept it as a fact even if it were asserted. And, of course, to make such an example work, we need a speaker who is not taken to be such an ideal of reliability.

Now, rather than simply consider an assertion which presupposes what another assertion directly asserts, we might take a comparison between an assertion made using a presuppositional idiom and one made using its conjunctive counterpart: the latter would be an assertion containing as a conjunctive part what the former presupposes, and so it would suit us equally well. We shall then be able to appeal to the same example in Section 6.

So, let us say that at its first meeting the *ad hoc* room-allocation committee, which had been convened to consider the new female presence in college, allocated room 2 to a prospective woman undergraduate. At its second meeting, however, some of the more gallant members of the committee expressed doubts about this, on the grounds that the advantage of saving her dainty legs from the effort of climbing too many stairs was more than outweighed by the assault she would suffer on her delicate ears from the noise in the bar directly beneath. As a result of discussion, both inside and outside the committee, at the third meeting the Bursar grudgingly undertook to waste his secretary's time in trying to reorganize allocation, so that the girl could be moved from room 2.

Let us say that I know all about this, and a few weeks into Michaelmas a friend, who likes to complain about everything on everyone's behalf, makes a typical angry utterance (and then rushes out of the room). There are two cases:

(24) The girl in room 2 has complained about noise.

(24a) There's a girl in room 2 and she's complained about noise.

Now it would seem to me that if sentence (24) were used, then I should be likely to believe he had made a mistake about the occupant of room 2: perhaps he had forgotten about the change of plan, or attended the first meeting but had been away during subsequent ones. At any rate, I would think that there was some muddle — he had either got the room wrong or the occupant wrong — and my belief that room 2 did not contain a woman would not be

overturned — at most some uncertainty might set in. On the other hand, if he had used sentence (24a), I think I would immediately accept as a fact that there was a girl in room 2. I would take him to be complaining precisely because, contrary to the committee's final view, by some negligence or stubbornness, a woman had in fact been put there and she was indeed being disturbed. Of course in neither case is it likely that I would actually indulge in the rationalizations described: I would be immediately affected one way or the other in virtue of my understanding of English and my recollection of the room-debate. But, if challenged, I might explain myself like this, and it would surely be reasonable.

So far this just illustrates the different potential that the two sentences have for affecting my view of how things are. This is indeed very important: it is a feature of the difference between presuppositional idioms and their conjunctive counterparts which we shall want to emphasize in Section 6. However our present concern is with knowledge transmission. Let us, therefore, further suppose that room 2 does in fact contain a woman and that the speaker knows this (and knows also that she has complained). It would then seem to me that if the idea of assertion making as a channel of knowledge has any application at all, then the use of sentence (24a) in these circumstances would be such a case. I would be left knowing that there was a girl in room 2. On the other hand suppose he had used sentence (24). He might well have done, for it is easy to imagine circumstances in which it would have been reasonable — though false — for him to assume that either I already knew there was a girl in room 2 or would be willing to accept this as a fact, without his having actually to assert it. In this case, according to the story, I cannot have come by the knowledge of this fact, because I do not even believe it.

Perhaps we should now turn back to considering the speaker more carefully, since there may be some new puzzlement, raised by this last

example — if not before — concerning what kinds of intentions, or at least expectations, the utterer of a presuppositional idiom may be supposed to have concerning his audience's attitude towards the content of the presupposition and its relation to the overall assertion. The case where the speaker believes that his audience already knows what he presupposes are unproblematic, and they are probably the most common. But what are we to say in general, so as to cover both these cases and ones where the speaker does not believe that the audience already knows what he is presupposing? It seems difficult to deny that the speaker is to (be taken to) expect that his audience will, once he has made his assertion, if not before, accept the presupposition as a fact — as something (mutually) known by him and his audience: indeed, once knowledge has been brought into the picture at all, how could it be otherwise, if presuppositions are to serve a role as fixed background 'premises' assuming which any debate, or just discussion, is to proceed?

If this expectation is not fulfilled, then a speaker's assertion is ill-suited to his audience. We do not, of course, want to deny that assertions are often ill-suited in this way, any more than we want to deny that assertions often do not convince an audience: we have considered examples of just this situation. But there remains the question of how to describe the difference between what the speaker expects concerning his presupposition and what he expects as a result of his assertion. A speaker's particular expectations may of course be different in different contexts; but it is, I claim, no part of what a speaker is to be taken to expect *qua* asserter of a presuppositional idiom that he will be cited as an authority for the truth of his presupposition. For he is certainly not offering any guarantee: if his audience act, verbally or otherwise, on the truth of what he presupposes, they are doing so off their own bat. In contrast to this, a speaker is to be taken to expect that what he actually asserts may be picked up by

an audience, taken as knowledge, and then possibly repeated to a third person on his authority. But, of course, if we have a presuppositional assertion, then this knowledge is dependent on the presupposition: the speaker is only a purveyor of potential or conditional knowledge: if you grant him the presupposition, he will guarantee you the assertion.

A speaker who asserts that p is telling someone (or anyone) that p : he is providing the information that p . Hence he is required, as a serious language user, to have some reasonable answer to the question "How do you know that p ?". If he has no reasonable answer, then he was being irresponsible in asserting that p . Not so with presupposition. A speaker is not, in presupposing that q , telling anyone that q or providing the information that q , and he would not be revealed to be irresponsible if he had no satisfactory answer to the question "How do you know that q ?". He is to be taken to expect that a hearer will accept, as a fact, that q , without his having to tell them, let alone actually explain why he would be justified in telling them. This is because the role of a presupposition that q in an assertion that p is precisely to be the (inseparable) assumption with which the information that p is conveyed. This is revealed in the very question "How do you know that p ?" (which, as we have said, it is appropriate to ask someone who has asserted that p): this question, construed in the natural way, presupposes that the addressee — the original speaker — knows that q and, in the presence of this presupposition, calls for an explanation of his supposed knowledge of what he asserted.

We might now consider why, because presupposition is not an official channel of knowledge, it is in fact a less reliable channel than assertion itself. Firstly, since a speaker is not informing you of a presupposition, it is not something he is necessarily very interested in, or even very aware of, and so he is less likely to be in a position to provide an explanation of why he takes himself to know it, and he is more likely to have made a

mistake. What someone tells you is inspired by having seen or heard such and such, or having read so and so, or having found a nice *a priori* argument — well remembered routes to the source of knowledge. On the other hand, presuppositions may spring from distorted old memories, from prejudices or from uncritically accepted common beliefs. Secondly, even if a presupposition is in fact well founded, it is less likely to convince an audience. If there are any doubts about a presupposition, the speaker's utterance is not going to quell those doubts: the audience is being expected to accept the presupposition, when they need to be told that it is so. They may reason that this supposed fact is in need of justification; the speaker is not, however, representing himself as able to provide such a justification; on the contrary, he is assuming it needs none: hence he has gone very far astray indeed and is not to be relied upon.

We might, finally, turn back to Stalnaker's definition of 'pragmatic presupposition', and to the idea that our assertion-dependent notion of presupposition is somehow a species of this. Pragmatic presupposition was eventually turned into a linguistic disposition, but it would not seem too far off the mark to take it that, on our view, if an assertion has been made along with the presupposition that *p*, then, in making that assertion, the speaker was thereby exhibiting the behaviour a disposition for which Stalnaker defines as pragmatic presupposition, viz. acting in his use of language as if

We might, in fact, be inclined to suggest a reformulation of what is inside the scope of 'as if' by replacing 'believes' by 'knows'. This would fit in very happily with our sub-speech-act view of presupposition and it would not obviously make the definition 'too strong' for general purposes, since, after all, we have altered only what is entirely within the scope of 'as if'.

The point of mentioning the connection between our account and Stalnaker's definition is so that, without having to think any more for myself, I can point out a rather interesting connection between two *prima facie* rather remote things, viz. the role of pragmatic presupposition as a background to linguistic transaction, and our partial-valued semantics for literal meaning.

In connection with the first thing, Stalnaker comments that the 'increment of information' conveyed by two different statements may, in the presence of a given pragmatic presupposition, be the same. For example, presupposing that my neighbor (*sic*) is a male, "My neighbor is a bachelor" and "My neighbor is unmarried", convey the same information (assuming we agree to take 'bachelor' to mean 'unmarried man'). The connection between this phenomenon (which I discussed, in my own way, in my B.Phil. thesis) and partial-valued semantics should now be obvious: logically convex sets. Stalnaker does not suggest any way of specifying what the information, or increment of information, in such cases actually is, but we have the apparatus to do that. Idealizing, perhaps, as Stalnaker does himself, we may assume that a body of pragmatic presuppositions is deductively closed — is a theory T . Then, the information content in T of a sentence ϕ may be represented as $(T; \phi)$; and so ψ will provide the same information in T if and only if $(T; \phi) = (T; \psi)$: that is, if and only if $T \vdash \phi \leftrightarrow \psi$. See Section I.4. It all fits.

Furthermore, we know how partial-valued languages fit with convex sets. Hence we can now very satisfactorily relate the literal meaning of presuppositional idioms with this contextual phenomenon. To be brief: presuppositional idioms are idioms which do indeed carry along their own presupposition modulo which they provide a given increment of information. This is their assertoric content, and we want to represent this content by

a logical form in a partial-valued language. But, of course, the results of Sections I.4 and III.5 show that (roughly) this could equivalently be done in terms of convex sets, which, as we have seen, can be used to represent 'increments of information'.

Finally, we can point out that convex sets representing the 'literal increment of information' of a presuppositional idiom may of course be contextually enlarged to specify an overall contextual increment of information. This enlarged convex set would contain not only sentences equivalent modulo the literal presupposition of the sentence, but, in addition, all sentences equivalent to those modulo the literal presupposition together with the background presuppositions.

V.5 ANOTHER PRESUPPOSITIONAL QUANTIFIER

Recall the kind of sentence which we proposed in Section I.3 as the 'natural language analogues' of interjunctions: sentences containing two component sentences (possibly joined by 'i.e.' or 'or in other words' or something of the sort) whose interpretation was taken to be such that an assertion had the effect of asserting the components as coming to the same thing. These components were taken to be asserted neither as conjunctive parts nor as disjunctive parts, but rather asserted together — as one. And we suggested that such a sentence carried the presupposition that the two components were equivalent — materially equivalent that is. We did not suggest that the presupposition itself involved anything stronger, though, certainly, it would be natural to use such an idiom in cases where the presupposition was taken to hold because of some stronger equivalence.

It would seem that this idea — taken seriously as a proposal for how to analyse the natural language idioms, rather than merely as a way to read formal interjunctions — is in accordance with the account given in the last section. Consider the examples we gave (and recall that we dismissed the

difference between the two sentences as no more serious than the similar assymetry of 'and' and hence as one we could justifiably abstract away from):

(25) John eats no beans — he is a follower of Pythagoras.

(25a) John is a follower of Pythagoras — he eats no beans.

Someone uttering either of these sentences would be presupposing that John eats no beans if and only if he is a follower of Pythagoras; and this would mean that the speaker would be offering this as a fact to be accepted by his audience and used by them if they wished to attempt to refute his claim about John — and, equally, to be used by himself if he wished to attempt to justify it. From this it would follow that either component sentence could be argued about, and, once that issue had been settled, the issue concerning the whole sentence would thereby be settled also. But this is what we have already suggested about these cases: that the two components are offered as standing or falling together. Our detour through the ideas of the last section has brought us back to the original intuition about these cases.

We suggested that interjunctions were more usual in reduced form, for example

(26) He is walking north — towards Carfax.

which would be true or false of a pedestrian in St. Aldates but truth-valueless of one in St. Giles.

But the question now arises whether there are any other idioms which are like interjunctions in having a presupposition which is paraphrasable in terms of a biconditional. I think that there are; and I think that they have probably been overlooked in the literature precisely because their presuppositions would have this complicated looking form if they were written down explicitly. And so overlooking them might in fact be traced back to too much emphasis on specifying a relation ' s_1 presupposes s_2 ', rather than simply specifying the content of sentences in such a way that their

presuppositions are thereby manifest. The idioms in question are, I think, as basic and paradigm as interjunctions themselves; they are, indeed, ones whose grammatical form is of the simplest: for example:

- (27) Trees are wooden.
- (28) Women are feeble.
- (29) The elephant is wise.
- (30) The rich are wicked.

The claim is that such sentences are standardly used to make assertions which involve the presuppositions

- (27a) that either all trees are wooden or none is
- (28a) that either all women are feeble or none is
- (29a) that either every elephant is wise or none is
- (30a) that either all rich people are wicked or none is

These are 'biconditional' presuppositions because we could have said 'that all trees are wooden if and only if some tree is wooden', etc.

Given these presuppositions, what is asserted could then be paraphrased by 'that they are in fact wooden', 'that they are in fact feeble', etc. But this is unrevealing. Let us introduce a new quantifier $\exists/x \dots x \dots$, binding x , so that we can say that these sentences may be analysed in the form $\exists/x Fx \cdot Gx$. Using $\forall x \dots x \dots$ and $\exists x \dots x \dots$, we could define this new quantifier — the 'squadgifier' we might say — by stipulating that $\exists/x Fx \cdot Gx$ is equivalent to

$$\forall x Fx \cdot Gx \ast \exists x Fx \cdot Gx .$$

Notice that the existential presupposition of $\forall x \dots x \dots$ and $\exists x \dots x \dots$ gets incorporated. But this is not what we are interested in at the moment, rather it is the presupposition introduced by \ast .

Is this presupposition a plausible proposal? The use of these simple idioms which most readily springs to mind is one where, though what

is grammatically speaking the predicate would apply to objects satisfying what is grammatically speaking the subject term, still, in some intuitive sense, what is being 'talked about' is the kind, or class, specified by the subject term. Accordingly, the presupposition is that to settle the truth or falsity of what is asserted it is a matter of indifference which particular instance of the kind (trees, women, elephants, rich people, etc.) is considered. Thus, for example, we can draw a theoretical distinction between the 'crude feminist' who would straightforwardly deny sentence (28) and the 'subtle feminist' who would claim that male chauvenism resided in the unwarranted presupposition that all women were alike in respect of being feeble or not.

Furthermore we may note that the crude feminist might assert "No, women are not feeble". Here, and in the case of other such idioms, the behaviour of negation and straightforward denial seems to accord with our presuppositional analysis. The natural negation

28b) Women are not feeble.

should be equivalent to something of the form $\exists x Fx \cdot \neg Gx$ — since 'women' in (28b) seems to be playing the same semantic role vis-à-vis 'not feeble' as it does in (28) vis-à-vis 'feeble'. And yet (28b) is the contradictory of (28). However there is no tension between these two things, since

$$\neg \exists x Fx \cdot Gx \simeq \exists x Fx \cdot \neg Gx .$$

I am not sure what serious rival analysis might be proposed for these all-or-nothing idioms: a plurality quantifier such as 'most' or 'almost all', or perhaps a concealed 'typically' inside or outside the scope of a universal quantifier? But where would this interpretation come from? I think that to propose such an analysis would be to read the result of a general phenomenon into the particular sentences: the general phenomenon being that 'precise statements' are frequently used 'vaguely'. We might

for example say "All Australians are heavy drinkers" for rhetorical effect, when we mean, and are taken to mean, nothing more than that almost all Australians are heavy drinkers. The mistake would then be to take the literal meaning of our kind of sentence, which has no explicit quantifier expression, to be the same as this vague use of universal sentences. But such an interpretation would be plucked out of thin air; and, in any case, I find it implausible, because, if it were correct, then we might expect sentences of the form ' ϕ 's are ψ ' etc. to give rise to the implicature that not all ϕ 's are ψ . However they never seem to.

I suppose that most often there would be a lazy man's analysis for (27) - (30) - a construal under which these sentences were taken simply to express universal claims. But this is surely a crude analysis; though, without partial-valued logic, it is perhaps the best we could do. The crudity would, of course, according to our analysis, lie in the identification of 'not-truth conditions' with 'falsity conditions' and it would result in complications concerning negation and denial. This is precisely the crudity which, in general, presuppositional analysis can resolve, and which * can reveal to be resolved.

It is interesting to see how many of the scope-independent features of $Ix \dots x \dots$ the complex squadgifier exhibits too. Indeed in examples (29) and (30) 'the' is used in the English idiom; and in some languages a definite article may be used with plural nouns too, for example:

(28c) αἱ γυναῖκες εἰσιν ἀσθενεῖς

This is intuitively very satisfactory in all cases, and it is particularly suggestive in the case of mass nouns: an analysis of mass nouns as 'predicates' may escape those objections which stem from their singular-term-like features, if we take sentences in which they occur in subject position as squadgifications; for example:

- (31) Snow is white.
(32) Grass is green.
(33) Porridge is revolting.

Of course this comment about mass terms is made with a formal language of logical forms in mind which makes basic sentences by means of predicate/singular-term composition — a language in which these are two syntactically distinct categories. What we have said may turn out to be just a first stage towards a more radical proposal for analyses in a language where such a syntactic distinction does not exist.

There is more to say about natural language and squadge. This must wait, however, until the next section.

V.6 ASSERTION AND PRESUPPOSITION

Dummett asks:

"Suppose a language of which we know nothing: it is intelligible to us, antecedently to any knowledge of the mode of composition of some sentence of the language, or any other, to be told that, by means of it, a speaker asserts that some specified condition holds good. Is it equally intelligible to be told that, by means of some sentence, a speaker asserts that a certain condition obtains, presupposing that a certain other prior condition obtains? Or would this information make sense only on the assumption of certain features of the inner composition of the sentences of the language?"

(Dummett (1978), p.xv)

We might object to the way he seem to suppose that presupposing is presupposing a condition 'prior' to the condition asserted, since, on our view, it is an inseparable part of it. But, waiving this complaint, our answer would be yes — see Section 4. Dummett's answer is no:

"Suppose one were told, of a language of which one otherwise knew nothing, that by uttering a certain sentence A , a speaker asserted that a certain condition C' obtained, presupposing the fulfilment of a prior condition C , while another sentence B carried no presupposition, but was used to assert that C and C' both obtained: thus, if C and C' both held, A and B were both true, if C held but C' failed, both were false, but, if C failed, then, while B was false, A was neither true nor false. From either characterisation, one would be supposed to derive a knowledge of the meanings of the two sentences, and the difference in meaning between them, via one's knowledge of the linguistic act of presupposing one thing and asserting another, or of the standard connection between meaning and conditions from truth and falsity. I did not, and still do not, see that one could derive anything of the sort. What difference in use would reflect the differences in meaning of the two sentences? In what circumstances would a speaker use the one and in which the other? In what different ways would a hearer react to the one utterance or to the other?"

(Dummett (1978) p.xvi)

In response to these rhetorical questions, I would again refer the reader to the discussion in Section 4. But Dummett has a general argument to back up his negative answers to these questions:

"The roots of the notions of truth and falsity lie in the distinction between a speaker's being, objectively, right or wrong in what he says when he makes an assertion. ...

Suppose that an assertoric utterance is such that it is possible, within a finite time, effectively to discover whether or not the speaker was right in what he said: and suppose that it is found that he was not right, so that he is compelled to withdraw his statement. What possible content could there be to the supposition that, nevertheless, the conventions governing that utterance were such that, in the case in question, he was not actually wrong? How could he have gone further astray than by saying something in saying which he was conclusively shown not to have been right, than by being forced to take back what he

said? What would be the point of introducing a distinction between his being wrong and its merely being ruled out that he was right? Or, conversely, The question is whether there is a place for a convention that determines, just by the meaning of an assertoric utterance of a certain form, that, when all the relevant information is known, the speaker must be said neither to have been right nor to have been wrong: and it seems clear that there is no such place."

(Dummett (1978) pp.xvii-xviii)

This is the negative view I want to question — it is elaborated further in Dummett (1959) and, here and there, in Dummett (1973).

To justify our proposed three-fold classification of assertions, we might be tempted just to reply that there is a place for an intermediate evaluation of assertions, and appeal once more to the account offered in Section 4. But I think it will be better to grant Dummett a dichotomy along roughly the lines he draws and then argue that presuppositional assertions are richer than he makes out because there is an important distinction to make between two different ways of being 'wrong'. This is perhaps no more than a terminological decision concerning the word 'wrong', but it is a tactically important one to adopt for some preliminary remarks.

In the first place it makes clear that I am not one of those whom Dummett censures because they come along armed with the labels 'true' and 'false', supposing them to be terms whose application is already well understood, and then claim that some assertions are neither true nor false. Of such people Dummett says:

"... the characterisation of a sentence as true in certain cases, false in others, and neither true nor false in yet a third kind of case, is not yet sufficient to determine the content of an assertion effected by the utterance of that sentence: it remains to be discovered whether the speaker does or does not rule out the possibility that the sentence is neither true nor false."

(Dummett (1973), p.347)

We might take the point that there is a general sense of 'rule out' in which, if we are not given whether or not the speaker is (or better, perhaps: it to be taken to be) ruling out the third set of circumstances, then we cannot be in possession of the content of what he asserted. The point about presuppositional idioms is not to deny this, but rather to discern some distinction between ways of ruling out. For it is not that we can make do with less than what Dummett requires, but rather that we need to know more about what the speaker is to be taken to be doing, if we are to get a complete specification of the content of his assertion.

Secondly we do not want our proposal that Dummett's dichotomy should be complicated to be taken to have anything to do with an anti-realist's gap, which Dummett himself explicitly allows for. As far as presupposition is concerned we may talk indifferently about 'circumstances' or 'recognisable circumstances', since, clearly, the problematic case of presupposition failure may arise under recognisable circumstances.

The following passage touches on this point, and it also raises a third, and rather more relevant, matter:

"In order to grasp the content of an assertion, we have to know in what circumstances the assertion is to be judged correct and in what incorrect. If the assertoric sentence is neither ambiguous nor vague, then these sets of circumstances must be disjoint and exhaustive. They must be exhaustive, at least, in the sense that there are not circumstances the recognition of which would entitle us to say that no further information would determine the assertion as correct or incorrect: if there were, the assertoric sentence could have had only an indefinite, i.e. partially specified, sense."

(Dummett (1973), pp.417-418)

Dummett, then, explicitly excludes 'ambiguous' and 'vague' assertions from falling within the range of his rationale for dividing assertions exhaustively into the right and the wrong: he would allow us to call an assertion

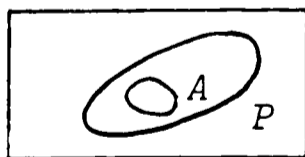
neither correct nor incorrect, if this were a way of pointing out that it did not have a 'determinate content'. We shall shortly consider to what extent he is right to exclude such cases. Incidentally, they do not have anything to do with out first remark: it does not appear from the context that Dummett was thinking about such cases in the passage quoted, and they are not within the scope of our concessive remarks about it.

However, accepting all Dummett's caveats and qualifications, we would still seem to have something to disagree about. For recall that we wish to see a three-fold evaluation scheme for assertions as a way of giving their content which is independent of, but will illuminate, the details of a tri-classificatory semantics for sentences used to make those assertions. Dummett, on the other hand, would see such a classification as stemming from nothing but details of linguistic structure which it was found convenient to incorporate into a semantic thoery. We discussed this earlier. The present concern is not with how exactly he would view the enterprise of using a tri-classificatory semantics to account for details of structure, but rather with the claim that such a semantics would have no other point to it than this — the claim that there could be no structure-independent rationale for a non-trivial three-fold classification of assertions.

According to Dummett, circumstances, or states of affairs, are exhaustively divided into those ruled in and those ruled out by an assertion — by what a speaker is (to be taken to be) doing when he makes an assertion; and this determines the content of the assertion. I take it that it is because he sees his dichotomy as grand and fundamental that he sees assertion content determined in this way. I take it, also, that it is because he sees assertion content as determinable in terms of such a dichotomy that he takes it to be grand and fundamental, overshadowing any

other discriminations. This is not a charge of circularity: rather I want to reveal what I take to be common ground: the idea that the question of assertion content and the mode of evaluation of assertions according to how things are go hand in hand.

Before we consider how Dummett himself elaborates the idea of ruling in versus ruling out, let us briefly rehearse the view which Section 4 suggests. We would say that there is a serious equivocality in the notion of 'ruling in' and, correspondingly, in that of 'ruling out'. Let us talk of states of affairs, as Dummett does, but also allow ourselves the heuristic benefit of schematic letters for classes of states of affairs. Let V be the class of all states of affairs. Then, we would say that, in one sense of 'rule in', a speaker rules in all those states of affairs, P say, which are compatible with a commitment to what he presupposes; but, in a second sense, subject to this commitment, he discriminates among P and rules in a subclass A of P , which, by his assertion, he is claiming contains the actual state of affairs. Hence, in one sense, he is ruling out those states of affairs which are excluded by his commitment — the states of affairs under which his presupposition would be said to have failed — viz. $V \setminus P$; and, in another sense, subject to his commitment, he is ruling out states of affairs incompatible with his claim, viz. $P \setminus A$.



He is certainly expecting a hearer to take him to be dismissing any state of affairs in $V \setminus P$, but he is not offering any assurance about this: he is only out to be reliably informative to hearers who will go along with him in dismissing $V \setminus P$. In contrast, he is offering a guarantee that, among P , $P \setminus A$ may be dismissed: he is providing information to warrant

dismissing states of affairs in $P \setminus A$, so that, always assuming that a hearer has agreed to ignore $V \setminus P$, A is left as the class of possibilities.

This is all part of assertion making. The semantical features of a partial-valued language then fall into place — consider in particular negation — but we do not have to appeal to them in our description of P and A . Moreover, the labels 'true' corresponding to A , and 'false' corresponding to $P \setminus A$ seem appropriate: A and $P \setminus A$ are alike distinct from $V \setminus P$ in an important way. Nothing should hang on these labels, however, since, of course, A is distinct in an equally important way from both $P \setminus A$ and $V \setminus P$. Let us, then, say that both $P \setminus A$ and $V \setminus P$ correspond to falsity: I would nonetheless urge that there are two different kinds of falsity which are to be distinguished not just for semantical reasons.

But how, then, does Dummett argue that the distinction between A and $V \setminus A$ ($= (P \setminus A) \cup (V \setminus P)$) overshadows everything else? For this is where he would discern the great divide. $V \setminus A$ — states of affairs that are 'ruled out' — are characterized as those conditions under which the speaker would be forced to withdraw his assertion. He appears to be impressed by the fact that, faced with any state of affairs in $V \setminus A$, a speaker would, if he is honest and serious, have to withdraw. Let us accept this: however, I still do not see why, having drawn a line in this way, such a line is the one and only important line for determining assertion content. It would, admittedly, seem the obvious line to draw if we were allowed only one: but why, for determining assertion content, should we suppose that we are allowed to draw only one line? Or, having drawn one line, why should we then suppose that any further discriminations are to be accounted for in some totally different kind of way — in terms of nothing but linguistic structure, or inclinations and expectations based on nothing but this? To quote Dummett's own words for a final time:

"A statement, so long as it is not ambiguous or vague, divides all possible states of affairs into just two classes. For a given state of affairs, either the statement is used in such a way that a man who asserted it but envisaged that state of affairs as a possibility would be held to have spoken misleadingly, or the assertion of the statement would not be taken as expressing the speaker's exclusion of that possibility. If a state of affairs of the first kind obtains, the statement is false; if all actual states of affairs are of the second kind, it is true."

(Dummett (1959))

Here we have the idea of speaking misleadingly. It is worth considering this formulation, for it might be expected to have some connection with Dummett's challenge to those who disagree with him to provide an example where making an assertion with a presuppositional idiom would have a different gross communication effect from an assertion made with its conjunctive counterpart. It would presumably be Dummett's position that a presuppositional idiom and its conjunctive counterpart would mislead in precisely the same circumstances. However I do not believe that this is the case.

We have only to consider reactions to assertions on the part of people who are not totally without existing beliefs of a relevant kind. Take the example in Section 4 concerning the girl in room 2. If we change the story by assuming that there is in fact no girl in room 2 (as the committee desired), then we have a situation in which the presuppositional idiom — sentence (24) — would not mislead me, but the conjunctive counterpart — sentence (24a) — would do. This is because in the first case there would be no assertoric weight behind the supposed fact that there is a girl in room 2, whereas in the second case there would be. Admittedly my natural response in either case might be to check up on the story in more detail; but if I had no means of doing so, but was left merely to my own reflections, then it would seem to me that such reflections would do nothing but back up my

immediate inclinations — in the first case to remain of the opinion that there was no girl in room 2, and in the second case to have been jolted out of this belief into the opinion that there was. In any case, as far as determining the content of an assertion is concerned, how my opinions eventually sort themselves out is surely somewhat irrelevant.

It is, I believe, possible to provide many cases of this sort — for example ones which involve an actual or supposed variation in a commonly accepted pattern of events. If you assert something unusual or unexpected there is a very good chance that an audience will believe you, but if you presuppose something of this sort an audience is very much more likely to believe that you have made a mistake — or that you are being clever and non-literal in your use of language. This is because an assertion actually offers information, whereas a presupposition is what a speaker is taken to expect his audience to agree to. Hence it is easier to mislead people by asserting something than by presupposing it.

In an attempt nonetheless to ground presuppositional phenomena in nothing but structural intricacies, it might be argued that at least no totally unstructured language could ever have presuppositional sentences for making assertions. In response to this I would say, first, that even if this were so, it would not show that presuppositional phenomena arise in virtue of nothing but structure: even if presuppositional assertions could only be made in a structured language, still, this does not show that presupposition is nothing but a phenomenon of structure. Putting it in a quasi-historical way, there is surely no reason to suppose that presupposition arose from the way structures independently developed, rather than that it was the urge to make presuppositional assertions that forced language to manifest presuppositional structures.

However, I am not convinced that unstructured sentences could not give rise to presupposition. For example, a cave-man language may contain a word-sentence 'groom' for asserting the presence of a certain kind of weather. May be the application of this word presupposes that a certain cluster of criteria go together: in other words we might have to explain the meaning of 'groom' as something like

rain * black-clouds * wind.

Climatic conditions, we may suppose, are such that rain and black clouds and wind typically go together, so that 'groom' provides a useful univocal sentence. On freak occasions, however, or if a speaker goes into the next valley, things may be different: an utterance of 'groom' might suffer from presupposition failure.

But would such an example be admissible as a case of presupposition? Clearly, exactly parallel situations arise in interpreting one known language into another, when a simple predicate of the one language embodies a different cluster of criteria of application from any simple predicate in the other. For example, parallel to what we said about 'groom', we should have to say something like

x is a πόλις iff x is a city * x is a state

Indeed, the phenomenon is recognisable within a single language. For example, perhaps:

x is breakfast

≈

x is a meal at breakfast-time * x is the first meal after waking.

Now it is often agreed that a simple subject-predicate sentence of the form Fa may be said to be neither true nor false, if there is some conflict of criteria concerning whether or not F is correctly predicated of a , but is this a case of presupposition failure? That is to say, are such sentences used to make presuppositional assertions in the first place? I

would argue that they are. We have been ignoring this kind of example just because our interest has been with structure, but I do not see why our account of presuppositional assertion — or rather our account of assertion simpliciter, with presupposition an inescapable, but sometimes trivial, constituent — does not fit perfectly with these cases. In making an assertion with a sentence of the form Fa , a speaker is expecting his audience to go along with him that F definitely does or does not apply to a , i.e. that the cluster of criteria stand or fall together, and, subject to this presupposition, he is claiming that F does in fact apply (according to all relevant criteria) to a .

I suspect that Dummett might indeed be one of those to say that Fa were neither true nor false if criteria of application do not determine a determinate answer; but then he might account for this by saying that it is a way of pointing out that an assertion made using Fa did not enjoy a completely definite content. Recall that 'ambiguous' and 'vague' sentences were excused on these grounds from being an exception to his general principle. Something similar might be said in the present case also.

We might, in fact, note that predicates which exhibit multidimensionality in their criteria of application often suffer from an open-endedness — a kind of vagueness — concerning exactly what the dimensions are. But this aspect of multi-dimensionality has nothing to do with our preceding remarks, which simply concern how given dimensions are welded together.

Let us now turn (briefly) to vague sentences, that is to say sentences containing vague words: for example a basic sentence of the form Fa where F is vague. I take it that the kind of vagueness that Dummett has in mind has less to do with any multi-dimensionality of criteria of application than with border-line problems. So, for example, if F were

'is bald' and a were a boarder-line case of baldness, then there is temptation to say that Fa is neither true nor false.

But now, why does Dummett think that an assertion made using such a sentence constitutes a *bona fide* exception to his general rationale for the exclusive dichomy between ruling in and ruling out? I owe to John Burgess an argument to the effect that if they are not exceptions, then Dummett's argument against presupposition fails. For, if we consider the withdrawal test as a way of providing a notion of wrong, then it is surely the case that if a speaker has asserted Fa but is subsequently persuaded that a is a boarder-line case of being F , then the speaker must withdraw the assertion — hence it is wrong. However this conflicts with our inclination to say that Fa is neither wrong nor right because a is a boarder-line case. But then this would show that we have not provided an adequate general rationale for dividing assertions exhaustively into the right and the wrong. Hence why, in particular, should we suppose that it can be applied in the case of presuppositional idioms?

Dummett, however, has the idea that indeterminacy due to vagueness arises out of the 'partially specified' nature of the sense of vague words, and hence of sentences in which they occur. But I have difficulty in appreciating what this is supposed to come to: clearly there are boarder-line cases of baldness, but if saying that the sense of 'bald' is only partially specified, is no more than saying just that, then I do not know what it is. And presumably, to avoid saying that these cases are to be excluded from the scope of the rationale simply because they do not work, we must suppose that, according to Dummett, there is something independently important in saying that the sense of vague words is only 'partially specified'. As a form of words, I might myself prefer to say that their sense is 'totally specified' but that their sense is partial — i.e. determined by application and non-application conditions which are not exhaustive. (Cf. the discussion in II.1).

This is not to ignore the problem of 2nd, 3rd, ..., n^{th} , ... order vagueness, nor the threat that vagueness makes to our way of thinking about semantics. It is certainly too difficult a matter to begin to discuss properly. However, from our present point of view, I am impressed most strongly by the fact that, in spite of all these problems, we do seem to understand — in an apparently straight-forward kind of way — assertions made using sentences which contain vague words. Such assertions seem to be 'completely' understood, without having to ask the speaker for a more precise or more complete formulation — even when the assertion in fact turns out to be neither true nor false due to a border-line case. This is why I want to see the sense of vague words as 'completely specified'; and this is why I am interested in John Burgess's argument, for surely all that should matter, to make Dummett's proposed rationale relevant, is that the speaker has been — or could be — completely understood.

Of course, if we do apply Dummett's rationale to assertions made using vague sentences, then it might, in certain cases, be indeterminate whether the speaker was right or wrong. This is the problem about vagueness: even if a speaker is definitely prepared to withdraw a statement that *Fa*, once convinced that *a* was a genuine boarder-line case, we also have border-line border-line-cases, over which he might hesitate. But then, if we were now to take this to indicate that an exclusive dichotomy is not appropriate for determining the content of vague assertions, then, as before, how can we be sure, without further argument, that we can exclude the possibility that the rationale is not vulnerable for other reasons — for example because of our thesis that 'wrong' is interestingly ambivalent? Of course, our challenge is not so metaphysically deep, but it is, perhaps, as important as far as understanding assertions is concerned.

Recall that as well as vague sentences Dummett also excludes ambiguous ones from the sope of his rationale. Now, this is *prima facie* much

more reasonable: if we have a sentence which contains a lexical item which is an obvious homonym or a sentence which is blatantly structurally ambiguous, then, if an assertion is made in a context where the usual disambiguation mechanisms — whatever they are exactly — are not operative, then we would not know what to take the speaker to be asserting. However, we are not always faced with obvious or blatant ambiguities. There are subtle ones too, which, though unresolved on a given occasion of assertion, may nonetheless go entirely unnoticed. They may be carried along without causing a hitch in communication, and so, for the same reason that I am unconvinced that vagueness should be treated as a special case in a discussion of the activity of assertion making, I am disinclined to make a special case of such occurrences of subtle ambiguity.

Cases I have in mind are ones where speaker and hearer alike would accuse a logician of pedantry, or talking nonsense, if he stepped in and pointed out where the ambiguity lay. Clearly logicians may often read ambiguities into natural language sentences, because their notation generates them, though there is really no good reason to discern them. But again, we may be justified in discerning an ambiguity in a sentence — two readings, let us say — because there are clear circumstances under which one of the readings would be what was intended and understood and the other clearly ruled out as either intended or understood, and because there are also clear circumstances under which it would be the other way round, while there is in fact also a third kind of context in which there is no decisive case to be made out either way. Consider, for example, sentences such as:

(34) John walked quickly to the dentist's.

which is ambiguous between a quick walking and a quick getting-there. (This example comes from P.P.E. prelims 1975.) Or:

(35) John believes that the Subrector is a Roman.

which is ambiguous between a *de re* and a *de dicto* belief. Or:

(36) John recommended her to try several different kinds of beer. which is ambiguous between one recommendation to try several kinds and several recommendations to try a particular kind.

In many cases of communication with sentences such as these it would appear that the distinction between the two alternative readings is patently irrelevant to the linguistic transaction: something is asserted — some one thing is asserted — and the assertion is understood. But exactly what, then, is the content of the assertion? There is no puzzle: the answer should be obvious. Let ϕ and ψ be specifications of the alternative readings of the sentence in question, then what is asserted is that $\phi \times \psi$. That ϕ and ψ come to the same thing (in that particular case) is being presupposed, and the assertion is the indiscriminate assertion of these two things. Whether this account — if it is accepted — should prompt us to discern a third semantical reading of such sentences, viz. $\phi \times \psi$, I am unclear: perhaps it does not matter very much. But what is important, I think, is to ensure that a theory of ambiguity, or rather of disambiguating mechanisms, take into account this kind of situation. We have no ambiguity in the assertion itself; nor is it a confused piece of linguistic behaviour, in the sense of being muddled or inaccurate; nor is the content of the assertion only 'partially specified'. Certainly we have a conflation of the two readings, but we have a coherent and orderly one, which serves to facilitate communication, not to debilitate it.

REFERENCES

- BELNAP, N.D., Jn., (1970), 'Conditional assertion and restricted quantification', *Nous*, (4).
- DUMMETT, M.A.E., (1959), 'Truth', *Proceedings of the Aristotelian Society*, (59).
(1973), *Frege*, (Duckworth).
(1978), *Truth and Other Enigmas*, (Duckworth).
- EVANS and McDOWELL, (Eds.), (1976), *Truth and Meaning*, (Oxford University Press).
- FINE, K., (1975), 'Vagueness, truth and logic', *Synthese*, (30).
- van FRASSEN, B.C., (1971), *Formal Semantics and Logic*, (Macmillan).
(1975), 'Incomplete assertion and Belnap connectives', In *Contemporary Research in Philosophical Logic and Linguistic Semantics*, Hockney *et al.* (Eds.) (Reidel).
- FREGE, G., (1891), 'Funktion und Begriff', In *Philosophical Writings of Gotlob Frege*, Geach and Black (Eds.), (Blackwell, 1960).
(1892), 'Ueber Sinn und Bedeutung', In *op. cit.*
- HAACK, S., (1974), *Deviant Logic*, (Cambridge University Press).
- HALLDÉN, S., (1949), *The Logic of Nonsense*, (Uppsala Universitets Årsskrift).
- KARTTUNEN, L., (1973), 'Presuppositions and compound sentences', *Linguistic Enquiry*, (4).
(1974), 'Presupposition and linguistic context', *Theoretical Linguistics*, (1).
- KEENAN, E.L., (1971), 'Two kinds of presupposition in natural languages', In *Studies in Linguistic Semantics*, Fillmore and Langendoen (Eds.), (Holt, Rinehart and Winston).
(1973), 'Presupposition in natural logic', *Monist*, (57).
- KLEENE, S.C., (1952), *Introduction to Metamathematics*, (North-Holland).
- MOSCHOVAKIS, Y.N., (1974), *Elementary Induction on Abstract Structures*, (North-Holland).
- RUSSELL, B., (1905), 'On denotation', *Mind*, (14).

- SCOTT, D.S., (1967), 'Existence and description in formal logic', In Bertrand Russell, *Philosopher of the Century*, Schoenman (Ed.), (Allen and Unwin).
- (1970) 'Advice on modal logic', In *Philosophical Problems in Logic*, Lambert (Ed.), (Reidel).
- (1973a), 'Models of various type-free calculi', In *Logic, Methodology and Philosophy of Science IV*, Suppes *et al.* (Eds.), (North-Holland).
- (1973b), 'Background to formalization', In *Truth Modality and Syntax*, Leblanc (Ed.), (North-Holland).
- (1975), 'Combinators and classes', In *λ -calculus and Computer Science*, Böhm (Ed.), (Springer).
- SMILEY, T.J., (1960), 'Sense without denotation', *Analysis*, (20).
- STALNAKER, R.C., (1974), 'Pragmatic presupposition', In *Semantics and Philosophy*, Munitz and Unger, (Eds.), (New York University Press).
- STRAWSON, P.F., (1950), 'On referring', *Mind*, (59).
- THOMASON, R.H., (1972), 'A semantic theory of sortal incorrectness', *Journal of Philosophical Logic*, (1).
- WOODRUFF, P. (1970), 'Logic and truth-value gaps', In *Philosophical Problems in Logic*, Lambert (Ed.), (Reidel).