On the

Interpretation

of the

Propositional Calculus

Tristan Grøtvedt Haze
The University of Sydney
tristanhaze@gmail.com
draft of April 20, 2011

The question considered is 'How can formulae of the propositional calculus be brought into a representational relation with the world?'. Four approaches are discussed: (1) the denotational approach, on which formulae are taken to denote objects, (2) the abbreviational approach, on which formulae and connectives are taken to abbreviate natural-language expressions, (3) the truth-conditional approach, on which truth-conditions are stipulated for formulae, and (4) the modelling approach, on which formulae, together with either valuation- or proof-theory, are regarded as an abstract structure capable of bearing (via stipulation) a representational relation to the world.

The modelling approach is developed here for the first time. The simple technical apparatus used for this is then applied to two issues in the philosophy of logic. (1) I demonstrate a corollary or converse to Carnap's result that certain 'non-normal' valuation-functions can be added to the set of admissible valuations of formulae without destroying the soundness and completeness of standard proof-theories. This sheds considerable light on a recent thread of the inferentialism debate which involves dialectical use of Carnap's result. (2) I show how the approach can be extended to quantification theory, by defining a model-theoretic notion of validity equivalent to the usual one, but making use of a proof-theoretic apparatus in place of the device of assigning values to formulae. This sheds light on the close relationship between proof- and valuation-theory.

Introduction

Here is a first pass at expressing the question I wish to investigate:

How can formulae of the propositional calculus be given a meaning?

However, the use of the term 'meaning' may have the unwanted effect of suggesting word-language as a paradigm. More carefully expressed, the question is:

How can formulae of the propositional calculus be brought into a representational relation to the world?

There has been a fair amount of discussion at least highly relevant to our question, but mostly before 1955. I make sense of this as follows: in the first half of the 20th century, philosophy of logic was focused on foundational issues, and on working out a stable, coherent view of the interpretation of modern logical systems. Eventually, a good deal of order was imposed on this domain by Quine, Tarski and many others. Since then, a different question has taken centre stage: that of the relation between formal and natural languages. In these investigations, there is a tendency to take the representational function of formal languages as being completely clear-cut and perspicuous - as it were, the one thing we've got a handle on - with the semantics of natural language, and its relation to that of formal languages, being the problematic elements.

The present-day student of logic is likely to come to the subject with some idea of a mathematized subject which is somehow about reasoning. They will be introduced to formal "languages", their "semantics", and their "proof theory". Somewhere along the line, they may encounter some kind of general answer to the question I pursue here with respect to the propositional calculus (hereafter, PC). However, this answer may never come in for any serious testing. It is perfectly natural for an intelligent student to carry on with a provisional answer. and to go on prove some theorems in a system, learn about metalogical results, relationships to the theory of computation etc. At this point, our now proficient student may now be able to *do* a lot - they may even contribute new results or invent new systems. The original philosophical question about how logical systems can be brought into representational relations with the world *may* remain quite unexamined, without impeding such technical development.

Although the idea that PC can be brought into a representational relation with the world is extremely widespread, the question of how this can be done has not been systematically examined in our era. Bothered by my own lack of understanding in this regard, I took up the question, and was surprised at how far it led me.

Here I distinguish four kinds of answer. Different answers need not mutually exclude one another; perhaps there are several quite different ways to make PC formulae represent. In addition to the four kinds of answer I distinguish, there are many possible points of difference between versions of each kind. I aim to survey these possibilities as comprehensively as is reasonable. This makes some of the discussion rather complicated. I have tried to make these parts as readable as possible without sacrificing accurateness or thoroughness, but I have not done so well that I can afford to omit this excuse.

To be clear: I take PC syntax, semantics and proof theory as given. The inclusion of semantics here might seem odd, given our question. However, what is called 'semantics' in formal logic need not be regarded as *intrinsically* semantic in the non-mathematical sense, i.e. having to do with meaning in language. To avoid confusion, I will use the term 'valuation theory' for what logicians more customarily call the 'semantics' of PC. Since the term 'proof theory' is not much used outside of logic, I will not make any corresponding adjustment. Questions could be raised about the appropriateness of this term too, but it is at least more *historically* warranted than 'semantics' for valuation-theory; in early modern logistic, proof-theory really was about proving truths, but valuation-theory was commonly seen as a purely technical development.[1]

Now, to introduce briefly the four approaches:

The denotational approach: Formulae are regarded as denoting (designating,

[1] Witness this passage in Russell's *Principles of Mathematics*: Treated as a "calculus", the rules of deduction are capable of many other interpretations. But all other interpretations depend upon the one here considered, since in all of them we deduce consequences from our rules, and thus presuppose the theory of deduction. One very simple interpretation of the "calculus" is as follows: The entities considered are to be numbers which are all either 0 or 1; "$p \supset q$" is to have the value of 0 if p is 1 and q is 0; otherwise it is to have the value 1; $\sim p$ is to be 1 if p is 0, and 0 if p is 1; $p \,.\, q$ is to be 1 if p and q are both 1, and is to be 0 in any other case; $p \vee q$ is to be 0 if p and q are both 0, and is to be 1 in any other case; and the assertion-sign is to mean that what follows has the value 1. (p. 183-184). The passage is repeated, *verbatim*, in *Principia* (p. 115). (These references from Korhonen (2007, pp. 425-6)).

referring to) propositions, sentences or truth-values.

The abbreviational approach: Formulae and connectives are regarded as abbreviating certain natural language expressions.

The truth-conditional approach: Truth-conditions are stipulated for formulae, or at least complex formulae.

The modelling approach: Formulae, *together with* either valuation- or proof-theory, are regarded as an abstract structure capable of bearing (via stipulation) a representational relation to the world.

These categories are not entirely separate; for example, one might regard formulae as denoting, and connectives as abbreviating - a view which I consider under the denotational heading (in accord with the descriptions above). Nor are they completely exhaustive: the thoughts expressed in the *Tractatus*, for example, cannot properly be fit under any of these headings. It could be said that the *Tractatus* gives no *answer* to our question, but takes a path on which it cannot arise. Similar considerations perhaps apply to Frege's notion of elucidation, as opposed to *definition*, of primitive formal logical vocabulary.[2]

Furthermore, our taxonomy makes no room for the view that PC *cannot* be brought into a representational relation. I do not consider this view explicitly, but will show that it is false, by considering *at least* one "successful" approach. (I think all four approaches are successful to some degree, and personally find the abbreviational and the modelling approaches to be the most *evidently* successful at bringing about representation.)

For all that, this taxonomy is, I think, suitable with respect to the philosophical issues which our question raises, given that the question itself *has* arisen.

On scope: there are many ontological or metaphysical questions which might be raised about meanings, propositions, operations, truth-functions, formulae, and other things. For the most part, I do not consider such questions.

I use the term 'wf' for well-formed formulae, and 'atom' for atomic wfs. I sometimes call the formula yielded by putting a '~' at the front of a formula F 'the *tilde* of F', in order to avoid the semantic connotations of the term 'negation'.

1. The denotational approach

The denotational approach is of a quite different character from the other three. In its classic manifestations, it seems relatively innocent or naive; not a theoretical response to problems so much as something which just came naturally to early mathematical logicians. In discussions about notation, symbols were often said to 'symbolize' and 'stand for' things. Such phrases leave it somewhat indefinite whether it is a referential or some kind of abbreviational or proxy relation being discussed, and both disambiguations were common. The approach was later examined critically (most notably by Quine), and has sometimes been adopted as a reaction to problems with the currently more mainstream abbreviational and truth-conditional approaches.

The characteristic thing about the denotational approach is that formulae are regarded

[2] cf. Hacker (1975).

as denoting (designating, referring). The objects of reference may be propositions, sentences, truth-values, thoughts, and perhaps other things. For definiteness, let us take this description from Quine (1934) as our starting point:

> The usual sort of system treats of some manner of elements, say cardinal numbers or geometrical points, which are denoted ambiguously by variables; operative upon these elements are certain operations or relations, appropriately expressed within the language of the system. The theory of deduction, when construed as a calculus of propositions, is a system of this kind; its elements are propositions denoted by the variables '*p*', '*q*', etc., and its operations are the propositional operations of denial, alternation, material implication etc., denoted by prefixture or interfixture of the signs '$\sim$', '$\vee$', '$\supset$' etc.

So, atoms denote propositions and connectives denote operations. We shall take it that complex formulae are also to denote propositions (namely the results of operations denoted by the connectives). Note that Quine does not say exactly that connectives denote operations, but rather that operations are denoted by *the prefixture or interfixture* of connectives. I just take this as meaning that connectives appear as prefixtures and interfixtures, and that they denote operations. (I do not mean to make it seem as though it should be totally clear what a propositional operation *is*, or whether such things even exist. This matter will not be discussed.)

1.1 Grammatical issues

So far, all we have is a kind of naming system. On an alternative construal of this naming system, formulae denote truth-values, and connectives denote functions on truth-values. On another, formulae denote the numbers 0 or 1, connectives functions on these numbers. Or formulae may denote Plato or Socrates, connectives functions on those philosophers, etc.

But are formulae also to be regarded as sentences? The question is pertinent, because many logicians who adopt some kind of denotational conception *also* regard formulae as sentence-like (classic examples being Frege and the pre-Wittgenstein Russell). Here we shall explore the prospects for doing both simultaneously.

For such an approach to work, it seems like it will need to be the case that sentences - at least, sentences of PC - denote (although in 1.3. we consider an alternative). Historically, the main candidate objects for denotation are propositions (variously construed) and truth-values. But what reason is there to think that sentences denote at all? People do not ordinarily speak that way; in ordinary language, sentences are said to 'refer to' such things as their *constituents* designate. The doctrine that sentences denote could be argued for purely on the basis of consideration of PC (though today this would seem somewhat myopic given the availability of other approaches), or on a broader theoretical basis. The classic example of the latter course is Frege's semantical theory, as put forth in *On Sense and Reference*. It has sentences refer to truth-values.

It is worth noting that for the purpose of interpreting PC, the "sentential-denotationalist" need not maintain that sentences in *natural language* refer - it may be that sentences of logical languages are peculiar in getting their meaning in a way which involves their referring.

But: even if we assume that some theoretical justification for the doctrine that (some) sentences denote can be provided, it does not follow that the formulae of PC *are* sentences; an expression can denote without being a sentence. Atoms could perhaps be taken as abbreviations of pre-existing denoting-sentences, but if connectives are to denote as well (as they do on Quine's description above), a complex formula such as '$p \vee q$' will then just be a string of denoting expressions with no verb. This is not a sentence in any ordinary grammatical sense.[3]

What, then, are the options available to the would-be sentential-denotationalist?

1.1.1. Positing a special grammatical category

Theoretically, one option would be to argue that PC formulae are of a special verb-less grammatical category. Sentences of this form, the story would go, only contain noun-phrases, but are sentences nonetheless. If '~' refers to a function/operation called negation, and '*p*' refers to the proposition that snow is white, then the PC formula '~*p*' could be rendered without symbols as 'Negation The proposition that snow is white'. If, on the other hand, atoms refer to truth-values, we would get 'Negation The True'. This latter course would apparently restrict PC to talking about truth-values (and functions/operations). However, whatever theory was previously used to motivate the view that sentences refer will presumably have implications for this matter. For example, on a Fregean theory, the translation of '~*p*' as 'Negation The True' is inappropriate, since, while '*p*' (i.e. 'snow is white') and 'The True' might have the same reference, they do not have the same sense.

We have just been considering the view that a string of nouns (perhaps of a certain kind) can form a sentence of a special category, in the wider context of our consideration of the sentential-denotationalist's options. This view is strictly independent of the view that sentences denote. However, adoption of the former without the latter prohibits iteration (i.e. formulae with multiple connectives): on the denotational approach, connectives must be flanked by denoting phrases, but all connective-*containing* phrases will be sentences, and if sentences don't denote, they won't be able to flank further connectives. Also, since atoms are to denote, then they won't be sentences: formulae with *less* than one connective will not get a sentential interpretation either. Such a 'one-connective' view can be had *without* the theoretical cost of defending a special 'string of nouns' category of sentences, as we will see in a moment.

1.1.2. Verbs

A less highly theoretical option for the sentential-denotationalist would be to introduce verb-like expressions. For example, one might use verbs to 'connect' connectives with formulae (e.g. 'snow is white' *is related by* Conjunction *to* 'grass is green', or something of the kind). Both economy and terminology, however, suggest simply dropping the idea that connectives denote, and making *them* verbs. '∨' could be read as something like 'disjoins with', '⊃' as 'materially implies'. '~' could be read as a grammatically reversed 'is false' or 'is not true'. (English-speakers may here use archaic grammar as a heuristic, i.e. 'False is *p*'.) Just like the 'special grammatical category' strategy, this move is strictly independent of the thesis that sentences

[3] This grammatical issue is reminiscent of the more metaphysical issue of 'the unity of the proposition', which occupied leading philosophers in the late 19th and early 20th centuries. cf. Frege (1892), Bradley (1893, pp. 27-28), Russell (1903, §52), and Wittgenstein (1918, 2.03 - 2.034, 3.14 - 3.142).

denote, but without that doctrine formulae will be limited to containing one connective. *This* version of one-connective-ism is less obscure than the previous, however, since it does not rest on an unusual grammatical doctrine. One plainly can make sentences of *this* sort, as logicians and philosophers do when for example they claim of two sentences/propositions that one materially implies the other.

Connectives-as-verbs can straightforwardly enough be read as applying to sentences or propositions, but the case of truth-values calls for comment, with respect to negation especially. 'False is' ('is false' reversed), as applied to sentences or propositions, is a predicate which ascribes a truth-value, but truth-values themselves do not have truth-values. A quick fix would be to regard the phrase as not a predicate, but part of a statement where 'false' refers to a truth-value, and 'is' is the 'is' of identity, not predication. (Likewise, if 'is not true' is preferred, we read 'is not' as the 'is not' of distinctness, not dispredication, and we regard 'Not true is' as a re-arrangement of the same material.) Finally, just as with the truth-value version of the previous strategy, we get the issue of formulae apparently being *about* truth-values, in lieu of any special theory such as that of Fregean senses.

We have now examined two ways of getting a sentential version of the denotational approach - that is, two ways of making PC more than just a naming system. The first involves recognizing a special grammatical category of sentences, while the second involves reckoning connectives as verbs. On both conceptions, to get formulae with more than one connective (and thus to retain the 'naming system' aspect) requires that sentences denote.

1.3 The Quinean strategy

In his (1934), Quine suggested a different strategy for getting a sentential interpretation of denotational PC. His motivations were two. The first motivation was essentially the avoidance of the issues we have just been dealing with; Quine regarded the doctrine that sentences denote as *ad hoc*. Using the word 'proposition' to mean the kind of object, if such there be, which sentences denote (and thus not ruling out truth-values), he complained that 'while we are apprised of a wide array of logical properties of propositions, concerning which there is little essential disagreement, on the other hand as to the residual character of propositions we have that full latitude of choice which attends the development of gratuitous fictions'.

Quine's second motivation was quite subtle, and has to do with the possibility of reinterpreting the calculus, or considering it in abstraction from any interpretation:

> It was suggested above that in the ordinary calculus of propositions the theorems are expressions denoting certain of the elements of the system. This is an anomaly upon which mathematicians have looked askance. It is customary to consider systems in abstraction from the nature of their elements; the theorems of a system, thus viewed, become sentences telling us various properties of unidentified elements. But to abstract from the fact that the elements of the propositional calculus are propositions is to deprive the theorems themselves of their character as sentences, since in that calculus the theorems are symbols of elements of the system. The student of systems in the abstract thus comes to an *impasse* when he takes up the calculus of propositions.

The point is: if the elements of the propositional calculus are denoted by formulae-as-sentences, and if sentences by their nature denote propositions and not other things, then it would seem that the elements *must* be propositions.

Quine's foremost recommendation regarding these problems is to abandon the denotational approach in favour of the abbreviational (which we discuss in section 2):

> Without altering the theory of deduction internally, we can so reconstrue it as to sweep away such fictive considerations; we have merely to interpret the theory as a formal grammar for the manipulation of sentences, and to abandon the view that sentences are names.

He also offers a more conservative strategy which retains denotationalism:

> But there is a way of gaining these advantages without persisting in the exclusion of the theory of deduction from the orthodox realm of systems. The theory can be reinterpreted in such a way that the signs '*p*,' '*q*,' etc., resume their status of variables denoting elements of the system, without return to the fiction of propositions as denotations of sentences. We can reconstrue the theory of deduction as a branch of semantic, a system whose *elements* are shapes, signs, specifically sentences.

On this view, formulae are *not* sentences; they merely *denote* sentences. To turn a wf into a sentence, the symbol '⊢' ('turnstile') is introduced, 'which may be read as a predicate to the effect that the element denoted in its wake is a true (i.e. truthful, truth-telling) sentence'. *This* system, turnstile included, can be reinterpreted and considered in the abstract. So we no longer have any anomaly in 'the theory of systems', and nor do we need to maintain that sentences denote.

In proposing this strategy, Quine simultaneously recommends that we treat the elements as sentences, rather than, say, propositions. But the main idea could have considerable appeal to someone who wants to speak of propositions, while regarding them as being *expressed*, but not *denoted*, by sentences.

1.4 The Russellian strategy

We conclude this section with a look at an intriguing interpretation strategy employed by Russell in *The Principles of Mathematics*. It is likely to be of interest as a curiosity, and as further evidence of the great variety of possible strategies, but it could also be used as a reply to Quine's charge that the typical denotative approach makes for an anomalous, un-reinterpretable system. It is an ingenious and "outside the square" idea, characteristic of the early Russell.

Sentences are taken to denote propositions. Atoms are treated as variables; not "propositional" variables, but ordinary term variables. So expressions like '$p \supset q$' become open sentences which can be universally quantified. These quantifiers are *completely* unrestricted (which is what makes this a reply to Quine).

Theorems, e.g. '$p \vee {\sim}p$', as they stand, still do not yield truths when quantified; it is not the case that, for absolutely all objects p, p disjoins with (false is p). (Remember: sentences are taken to denote.) Russell adds, to each theorem, a prefix containing, for each of its atoms p, '$(p \supset p)$', followed by a '$\supset$'. Thus '$p \vee {\sim}p$' becomes '$(p \supset p) \supset (p \vee$

$\sim p$)'. *Then* quantifying, we get '$\forall p[(p \supset p) \supset (p \vee \sim p)]$', or:

> For all *p*, (*p* materially implies *p*) materially implies (*p* disjoins with (false is *p*)).

Now this, according to the early Russell, is a completely general truth. The idea is that for values of *p* which are not propositions, e.g. Russell himself, the quantified sentence comes out vacuously true. So, the following instance is true according to Russell:

(Russell materially implies Russell) materially implies (Russell disjoins with (false is Russell)).

Russell, not being a proposition, does not materially imply himself, so the first parenthesis is false, i.e. denotes a false proposition. Thus the whole sentence, being a material implication with a false antecedent, comes out true.

This is a notable kind of interpretation in its generality and "fixedness": at no point is anything assigned to, or conventionally associated with, any formula. Wittgenstein explicitly criticized it in the *Tractatus* (5.5351):

> (It is nonsense to place the hypothesis '$p \supset p$' in front of a proposition, in order to ensure that its arguments shall have the right form, if only because with a non-proposition as argument the hypothesis becomes not false but nonsensical, and because arguments of the wrong kind make the proposition itself nonsensical, so that it preserves itself from wrong arguments just as well, or as badly, as the hypothesis without sense that was appended for that purpose.)

2. The abbreviational approach

On this approach, atoms abbreviate sentences and connectives abbreviate expressions of English (or another natural language). Brackets don't abbreviate - the bracket conventions can be thought of as a regimentation of English. Thus formulae come into a representational relation to the world by abbreviating expressions which naturally stand in such a relation.

The two main questions which arise on this approach are:

(1) Which connectives are to be taken as basic? (In principle, any functionally complete set will suffice.)

(2) Which expressions of English, and which occurrences thereof, are to be abbreviated by the basic connectives?

We are not free to answer (2) in just any grammatically acceptable way: we must answer it such that standard proof theory does not lead us from truth to falsity, and such that the truth-conditions of compound sentences agree with the "truth-tables".

As regards (1), and the fact that in principle any functionally complete set of connectives will suffice, I shall not take advantage of this by confining my attention to some small set such as {The Sheffer Stroke}, since much of the interest of the calculus has to do with its applicability or otherwise to natural language conjunctions, disjunctions, negations and conditionals. Accordingly, we shall consider in turn the prospects of taking '∧', '∨' , '~' and '⊃' as basic.[4]

2.1. '∧'

'∧' is taken to abbreviate 'and' - but not just any occurrences. For example, the occurrence in 'John and Mary had a picnic' is not fair game. This gives rise to a simple syntactic constraint: '∧' is to abbreviate such occurrences of 'and' as appear between grammatically complete sentences. But purely syntactic criteria are not enough. For a start, in some cases the very same string of characters can, in English, express both a conjunction of two sentences, as well as a different sentence, e.g. 'Paper folds and tears well'. This could be read as a conjunction of a general statement about paper and a general statement about the objects produced by tear-ducts, and what they do in eyes. It could also be read as saying that there are two things which paper does well: fold and tear. Another source of failure of syntactic criteria is the use of conjunctions to imply temporal or causal relations, for example: 'John announced his resignation and his colleagues were shocked'. This doesn't *just* affirm both conjuncts: it is not equivalent to 'John's colleagues were shocked and he announced his resignation.' So we need a semantic criterion for abbreviability by '∧'. The obvious criterion, commonly employed in logic books, is truth-functionality: '∧' is to abbreviate such occurrences of 'and' as form conjunctions which are true iff both conjuncts are true. On some views, this may not narrow down the abbreviable occurrences to one meaning: some may be "dry", while others may harbour subtle *semantic* implicatures which do not give rise to a difference in truth-conditions. But none of that need be seen as a problem: PC will never lead us from truth to falsity, whether or not we regard some of the compounds as carrying implicatures which affect meaning.

Another question is whether there *are* any truth-functional occurrences of 'and'. I think the answer is obviously 'yes'.[5] It is not so clear in other cases, however - 'if' being the most vexed.

Without this restriction to truth-functionality, PC might look like a substantial, revisable theory about the truth-values of certain sentences. With the restriction in place, it does not have that character, but remains a codification, depiction, and calculating device.

It could be asked whether a "weaker" semantic criterion might be possible - one which does not speak of truth-functionality. This is not an entirely clear question, but one thing is clear: if we are not to be led into falsity by the calculus, any such criterion will have to be co-extensional with, or less extensive than, the truth-functionality criterion. And so it is hard to see how it could be "weaker" in any attractive way.

We shall then settle on using '∧' as an abbreviation of *truth-functional occurrences* of 'and', where the truth-function is the conventionally associated one. We could call this

---

4 For reasons of space, '≡', '≢', and their respective relations to 'if and only if' and an exclusive sense of 'or', will not be discussed. These, rather than others, were omitted on the following grounds: (1) The issues associated with '' and 'iff' are in large part (though, it is true, not entirely) present in the case of '⊃' and 'if'. (2) Exclusive uses of 'or' at the sentential level are notoriously hard to come by in natural language, and this issue is not fundamental to our inquiry.

5 By 'truth-functional occurrences', nothing more is meant than that the truth-value of the compound is a function of the truth-values of its components. No presumption is made that these compounds refer to, or concern, functions.

sort of abbreviation 'semantic', as opposed to purely syntactic abbreviation, to mark the fact it that can yield expressions which do not in general have the same meaning as the abbreviated terms.

2.2. '∨'

Having allowed restriction to truth-functionality, we can afford to be quite cursory in our consideration of '∨'. The only thing we need settle is that there *are* truth-functional occurrences of 'or' (where the function is that associated with '∨'). The answer is not quite as obvious as for 'and', due to the distinction between inclusive and exclusive disjunction. We must assure ourselves that not all truth-functional occurrences of 'or' go by the exclusive function. This is easily done: consider the sentence 'I saw a cat, or I saw some animal'. It is clearly true if I saw a cat *and* I saw some animal (namely the cat), so this is not an exclusive 'or'. It is clearly true if I saw a cat, or if I saw some animal. And it is clearly *not* true if I neither saw a cat nor any other animal.

2.3. '~'

'~' is commonly pronounced by logicians as 'not'. So perhaps we could have it abbreviate truth-functional occurrences of 'not'. The trouble is: it's not clear that there are any, at least in ordinary speech.[6] We typically use 'not' *in* sentences, often for dispredication ('John did not go to the concert'), and for what might be called 'dis-subjection' ('John, not Mary, went to the concert').

One might wonder what all the fuss is about; surely we can all understand a negation sign, and are free to use 'not' - or any word - for this purpose! The fuss is about the following: we do not want to regard this as the denial that a predicate (relation, identity, etc.) holds of some subject, since we want to be able to deny, in a literal way, statements which presuppose the existence of subjects we do not believe in - subjects which we ourselves do not want to attempt to literally refer to. On the other hand, we do not exactly want to talk *about* a proposition or sentence and deny that it possesses truth, for that would be to change the subject, in a grammatical sense: we should have to *quote* or otherwise name the negatum, whereas we want '~' to attach to *sentences*. We want something, as it were, grammatically *above and between* these two kinds of "negation": the denial *of a whole sentence*, in contrast to denial with respect to some subject that such-and-such holds of it, whether that subject be the subject of the negatum (the first kind), or the negatum itself (the second kind).

With this in mind, 'It is not the case that' seems like a good candidate English expression for '~' to abbreviate. It attaches to sentences. Furthermore, it doesn't appear to involve ontological commitment to the objects mentioned in the negatum; let '*p*' abbreviate 'Santa came this Christmas'. '~*p*' then abbreviates 'It is not the case that Santa came this Christmas'. This sentence, unlike 'Santa did not come this Christmas', does not seem to commit one to the existence of Santa Claus. Furthermore, unlike '"Santa came this Christmas" is not true', it retains grammatical mention of Santa, but not ontological commitment, thanks to the natural semantic properties of the phrase 'It is not the case that'. Finally, 'It is not the case that' seems mildly preferable to 'It is not true that', since the occurrence of the word 'true' arguably gives a more "metalinguistic flavour" to the negation.

---

[6] The speech fad of "asserting" a sentence, followed by '...not!', might seem like an exception. (Thanks to N.J.J. Smith for this suggestion.) But perhaps such negative assertions often carry existential import, which (as we shall see below) is not desirable for negations in PC.

The proposed reading of '~', unlike our readings of '∧' and '∨', appears not to require any semantic restrictions; 'It is not the case that' seems quite unequivocal. It seems no coincidence that this reading is more prolix than the readings 'and' and 'or'. Perhaps one could fashion less ambiguous but more prolix expressions to go in their place.

2.4. '⊃'

And now we come to the notorious question of '⊃' and 'if...then'. Or rather, a certain version of it. *Our* question is not whether natural language conditionals in general, or even all indicatives, can be validly abbreviated in PC by '⊃'-formulae, but: can '⊃'-formulae in PC be regarded as abbreviations of *any* kind of natural language conditional?

The "if/⊃"-question has an interesting history. It had evidently been considered (in essentials) by the Stoics, and by some mediaevals (Abelard especially). By the 19th century, many logicians endorsed the view that '⊃' (or whatever symbol was used) could be read as 'if...then'. This continued through the early years of the 20th century, but conscientious objectors came into view. This is socially and historically interesting, in that (as we shall see) the essential matter of the controversy had lain dormant in logic books for years beforehand, without being much discussed. It is as though logic had started to come to life again: gradually, more people were moved to think *critically* (but without complete dismissal) about what they read in logic books. In the English-speaking world, MacColl was one of the earlier dissenters, though his criticisms were partly obscured by his own unpopular doctrines and procedures.[7] More successful criticisms came later from Strawson. By the time of Quine's (1953) review of Strawson's *Introduction to Logical Theory*, the former was able to treat the semantic divergence between '⊃' and 'if...then' as rather old news:

> The well-known failure of the ordinary statement operators 'or', 'if-then', 'and', and 'not' to confirm in all cases to the precept of truth-functional logic is well expounded by Mr. Strawson. Because 'and' and 'not' deviate less radically than the others, I have found it pedagogically helpful (in *Elementary Logic*) to treat the translation of ordinary language into logical form, at the truth-functional level, as funnelled through 'and' and 'not'; and Mr. Strawson follows suit.

And later:

> Mr. Strawson is good on '⊃' and 'if-then'. He rightly observes the divergences between the two, and stresses that '$p \supset q$' is more accurately read as 'not (p and not q)' than 'if p then q'.

This state of affairs did not last. A series of post-1960 events has changed things

---

[7] In 1908, in a short polemic against Russell, MacColl wrote: 'For nearly thirty years I have been vainly trying to convince [logicians] that this assumed invariable equivalence between a conditional (or an implication) and a disjunctive is an error'. (This is a reference to the Or-to-If Argument, which we consider in 2.4.3.) Russell's reply was made easy by the fact that MacColl had, in his objection, overlooked the former's distinction between propositions and propositional functions. After correcting this, Russell addressed the main issue swiftly, writing 'I say that p implies q if either p is false or q is true. This is not to be regarded as a proposition, but as a definition', and admitting happily that this definition does not give 'implies' its usual meaning. As we shall see in 2.4.3., this does not square well with the *justification* of the 'Definition of Implication' given in *Principia*.

irrevocably, so that Quine's comments above seem to come from a bygone era when things were much simpler. In my own view, the Quine-Strawson view was basically right, but one cannot make a respectable case for that today without discussing the post-1960 events. Therefore I shall now give a summary of the events, followed by a series of critical comments. (Following that, one important thing will remain to discuss: the Or-to-If Argument. This is a simple, initially-compelling deductive argument-form which, if valid, would suggest that '⊃' can be read as 'if...then'.)

2.4.1. The resurgence of '⊃': a potted history

Phase 1: In his William James Lectures at Harvard in 1967, Grice makes public his theory of implicature and conversational maxims. People are impressed by this idea: 'John is poor *but* honest' has the same *truth*-conditions as 'John is poor *and* honest', but the former (in some contexts) strikes people as objectionable and unassertable, even when the latter may be both true and *a*ssertable, the difference being that the former can carry an *implicature* that poor people aren't honest. Secondly, the maxim of 'Assert the Stronger' is developed; if someone asks where John is, and I know he's at the library, it's not proper to respond that he is either in the library or at the pub. Similarly, Grice argues, sentences like 'if snow is green then I am king' are *true* (just because snow isn't green), but unassertable, since we should assert the stronger: that snow isn't green. (The work is published in Grice (1975).)

Phase 2: Meanwhile, other philosophers had been continuing to develop more sophisticated accounts of the truth-conditions of conditionals. Among these is the possible worlds account of Stalnaker (1968), who, following Adams (1965) (who himself wasn't interested in the question of truth-conditions), conjectured that the probability (in some sense) of a conditional 'If A then C' is the probability of 'C' given 'A'. That is: P(If A then C) = P(C/A) = P(C & A)/P(A) (where P(A) is positive).

Phase 3: David Lewis proves his triviality results in Lewis (1976), to the effect that 'there is no way to interpret a conditional connective so that, with sufficient generality, the probabilities [of truth] of conditionals will equal the appropriate conditional probabilities'. He considers the possibility of accommodating this with a theory on which conditionals do not have truth-values (i.e. are not truth-apt): 'Why not? We are surely free to institute a new sentence form, without truth conditions, to be used for making it known that certain of one's conditional subjective probabilities are close to 1. But then it should be no surprise if we turn out to have such a device already.' He writes: 'I have no conclusive objection to the hypothesis ... . I have an inconclusive objection, however: the hypothesis requires too much of a fresh start. ... [W]hat about compound sentences that have ... conditionals as constituents? We think we know how the truth conditions of compound sentences of various kinds are determined by the truth conditions of constituent sentences, but this knowledge would be useless if any of those subsentences lacked truth conditions.' This boosts Grice's proposal, which Lewis has come to endorse: 'It turns out that a quantitative hypothesis based on Grice's ideas gives us just what we want: the rule that assertability goes by conditional subjective probability.' And so the truth-conditions of indicative conditionals are identified with those of '⊃'-statements. And for sophisticated reasons.

(To complete the story, though this is less important for what follows: in a postscript to his (1973) in his *Philosophical Papers, Volume II*, Lewis admitted that in 'special cases', assertability and conditional probability diverge. Secondly, he abandoned the 'Assert the Stronger' explanation of apparent counterexamples to the '⊃'-analysis, due to apparent counterexamples to the 'Assert the Stronger' maxim itself, in favour of an ingenious alternative theory devised by Frank Jackson: one may assert 'if A then C', even when one is in a position to assert the stronger 'C', if one wants to give information which is *robust* with respect to 'A' (which could have low probability): information which, even if 'A' turned out true, would still hold. For more details on how this theory works, see Lewis's postscript and Jackson (1979).)

Thus the Grice-inspired Lewis-Jackson version of the '⊃' analysis is today regarded as a serious proposal, even if it is not widely accepted. Some other major accounts on the market deny truth-aptness, either completely (cf. Edgington 1991, 1995) or in certain cases, such as when the antecedent is false (cf. McDermott 1996). All these accounts have in common that they are error-theoretic with respect to many or most competent speakers: the '⊃' analysis implies that competent speakers often get a conditional's truth-value wrong, while accounts which partially or totally deny truth-aptness have it that competent speakers often mistakenly ascribe truth-values to sentences which have none.[8]

2.4.2. Commentary on the resurgence

Comment on Phase 1: Note a fundamental difference between the cases of 'but' and 'or' on the one hand, and the case of 'if' on the other: people do not generally judge it *false* to say that a poor and honest person is poor *but* honest, but rather *wrong* in some other sense. This is even more pronounced in the case of 'or'. In that case, we can see perfectly well that the misleading statement about John is *true*. By contrast, competent speakers will confidently classify a sentence like 'If grass is blue, it isn't blue' as *not true*. Thus it seems any view which says that for every '⊃' sentence, there is a corresponding 'if' sentence with the same truth-conditions, will inescapably be an error theory with respect to competent speakers.

Comment on Phase 2: The notion that assertability or probability of conditionals goes by conditional probability may seem initially appealing, but apparent counterexamples abound: sentences such as 'If 6 is greater than 5, then 7 is greater than 6' and 'If Gödel's proof really was valid, the sun will thankfully rise again' do not seem at all assertable or probable. They seem like bits of nonsense. Furthermore, the idea that assertability can be quantified, and that it equals any sort of probability, seems odd; if I attach a probability of only .5 to some proposition P, why would I *assert* it? Such a proposition seems not assertable *at all* in a normative sense - and therefore not 'half assertable' either, whatever that means. A common proposal in response to this is that assertability remains low until probability gets high, at which point it shoots up. This has been criticized by Dudman (1992), using lottery cases: someone who has a ticket in a lottery will usually *not* be prepared to assert that they won't win, even though they may realize that not winning is very highly probable indeed.

Comment on Phase 3: Lewis, wanting to maintain that assertability of conditionals goes by conditional probability ('A = CP' for short), uses his triviality results to argue

[8] This claim might be resisted on the grounds that I am imposing a surreptitious philosophical interpretation on the "data", but I do not think this is so. With respect to the '⊃' analysis: people won't *only* say that there is 'something wrong' with conditionals such as 'if my house burns down tonight, it will stand tomorrow' - they will say they are *not true*. Regarding Edgington-style and McDermott-style accounts: people will say that 'if my house burns down tonight, it will *not* stand tomorrow' *is* true (whether or not my house does burn down). It could be objected that these judgements do not concern the *pure*, philosophical conception of truth, but: whatever *that* is, I do not mean to invoke any such thing. (In this connection, cf. Ness (1938) - a Norwegian work of experimental/empirical philosophy, cited by Tarski (1944)).

in effect that, since we can't give any truth-conditional analysis of conditionals such that probability of truth will equal conditional probability, *any* truth-conditional account will (by A = CP) have to explain divergences between assertability and probability of truth, so why not at least start with something simple like the '⊃' analysis? The quite different course of denying truth-aptness remains open, but - says Lewis - that requires too much of a fresh start.

The first thing to note about this line of argument is that, for reasons given in the previous comment, A = CP is really not independently attractive, once you consider certain examples. So perhaps no 'divergences' need explaining at all, and philosophers can go on looking for a non-gappy truth-conditional account of conditionals which is more plausible than the '⊃' analysis.

The second thing to note is that the logical space between giving a truth-conditional *analysis* of conditionals and denying truth-aptness remains largely unexplored. Consider the case of subject-predicate statements about explanatorily basic things possessing explanatorily basic properties: unless you're a pretty far out nominalist, you're likely to think that this is a class of truth-apt statements for which no non-circular truth-conditional analysis can be given - what we might call an 'analytically basic' class of statements.[9] (And if you *are* a pretty far out nominalist, maybe there is some other class of statements which you take as primitive.) A view on which conditionals are analytically basic - an antitheory of conditionals - can happily avoid the error-theoretic consequences of prevailing views, although it could be retorted that such a view is error-theoretic with respect to analytic philosophers. Surely the response to that is: when faced with a choice between a set of accounts which are error theoretic with respect to (almost) all competent speakers, and an error theory with respect to some philosophers, one of whom also believed in other universes inhabited by donkeys which speak[10], the latter should at least be examined properly. (This, of course, would go beyond the scope of the present inquiry.)[11]

There is a different family of accounts, known as "support" theories, which are *not* strikingly error-theoretic. Such accounts are for the most part out of favour today, but a highly sophisticated one has been developed by my teacher Adrian Heathcote, in unpublished work. In my view, all such accounts - if they purport to be reductive - will face circularity problems. (A defence of this view is beyond our scope here.) However, even if they don't succeed as reductive analyses, the key ideas behind "support" theories seem important for understanding the logic and context-sensitivity of conditionals.

2.4.3. The or-to-if argument

Before we look at this argument, it needs to be made clear exactly how, if valid, it

[9] This is not so straightforward as perhaps I make it sound. The notion of non-circularity here is problematic, and lies behind 'the paradox of analysis'. In any case, the notion of 'analytically basic' does not seem devoid of content, since its application to conditionals would clearly rule out extant analyses.
[10] I refer of course to David Lewis's thesis of modal realism, as defended in his diabolical (1986b). (There seems to be an analogy here: both modal realism and, in light of present knowledge, a truth-functional theory of conditionals, trade on a theoretical feeling of 'Wouldn't it be nice?'.)
[11] If one were to make this examination, a good first question would be: Why has this idea - that conditionals are truth-apt but analytically basic - been resisted at all? The resistance seems to have to do with a feeling that conditionals do not "match the furniture" (in some ontological sense), in the way 'the cat is on the mat' matches the furniture. (Something similar appears to be operative in the philosophies of mathematics and modality.) But the notion that analytically basic propositions always ought to match the furniture is one which must be questioned, and perhaps abandoned. (I read the opening pages of Wittgenstein's *Investigations* as being concerned with just this issue.)

would support the view that '⊃' can be taken as an abbreviation of 'if...then'. Canonically, the argument proceeds from a truth-functional disjunction of the form '$\sim p \vee q$' to a conditional of the form 'if p then q'. And the truth-function associated with '$p \supset q$' in PC is equivalent (or identical, extensionally speaking) with the truth-function associated with '$\sim p \vee q$'. Secondly, it is widely accepted that 'if p then q' implies '$\sim p \vee q$'. (This has been questioned on subtle grammatical/syntactic grounds, but we will not discuss that here.) Thus it seems that if the Or-to-If argument is valid, '⊃'-statements can be taken as logically equivalent to indicative conditionals.

2.4.3.1. Historical preliminary

The origin of the idea that one can infer a conditional from a disjunction appears to be unknown. There has been speculation that it originated with Stalnaker. Priest (2001, p. 17) says the Or-to-If Argument was 'given by' Faris (1968) - and it was, but not for the first time. While one of those authors might have made the first use of the inference form as an explicit argument for a truth-functional reading of 'if' *after* the issue had become controversial in our era, the form itself has a long and venerable history. We find it on p.64 of Cohen and Nagel (1934):

> *Equivalence of Compound Propositions*
> . . .
> Consider next the alternative proposition *Either a triangle is not isosceles or its base angles are equal*. To assert it means to assert that *at least one* of the alternants is true. If, therefore, one of the alternants were false, the other would have to be true. Hence we may infer from the alternative above the following hypothetical *If a triangle is isosceles, its base angles are equal*.

This textbook, which was popular in its day, also contains quite extensive discussion of the relation between 'formal' and 'material' implication - including a resolution of the 'paradox' attending to the latter (there is no paradox, since the term 'implication' is just given a special technical use in PC). Curiously, this 'paradox' is not related to hypotheticals (conditionals). In fact, hypotheticals are not discussed in the chapter on 'the calculus of propositions' at all, but in two more old-fashioned chapters near the beginning called 'The Analysis of Propositions' and 'Relations between Propositions'. (No doubt this has partly to do with the dominance of the denotational approach to PC at that time.) It is in the latter that the Or-to-If argument and its conclusion appear as a bland lesson.

Even C.I. Lewis, who famously raised the 'paradoxes of material implication' in his 1918 *Survey of Symbolic Logic* (and articles written earlier), had no problem with '⊃' being read as 'if...then'. He appeared to regard the latter as ambiguous between an "extensional" and an "intensional" reading. A curious passage on p.225 reads:[12]

> we can *now prove* that we have a right to interchange the joint assertion of p and q with $p \times q$, "If p, then q", with $p \subset q$, etc. We can demonstrate that if p and q are members of the class K, then $p \subset q$ is member of K, and that "If p, then q", is equivalent to $p \subset q$. And we can demonstrate that this is true not merely as a matter of interpretation but by the necessary laws of the system itself. We can thus prove that writing the logical relations involved in the theorems—"Either ... or

[12] Lewis was using the notation of the algebraic tradition.

...," "Both ... and ...," "If ... , then ..."—in terms of +, ×, ⊂, etc., is a valid procedure.

In this case, the "proof" does not proceed from Or to If , but by the previously "established" theorem '(1 ⊂ *a*) is equivalent to (*a* = 1)', together with the rather Tarskian postulate 'For any proposition *p*, *p* = (*p* = 1)', and a tacit use of something like Conditional Proof (which, we shall see, is crucial in the Or-to-If Argument). Today this reasoning would be regarded as metalinguistic, not 'by the necessary laws of the system itself'.

Earlier, we find the Or-to-If Argument given in support of the very first definition in *Principia Mathematica*, 'Definition of Implication':

> *1·01. $p \supset q . = . \sim p \vee q$ Df.
> . . .
> According to the above definition, when '$p \supset q$' holds, then either *p* is false or *q* is true; hence if *p* is true, *q* must be true. Thus the above definition preserves the essential characteristic of implication . . .

This was then taken to be authoritative in Hankin (1924), a widely-cited legal article on 'Alternative and Hypothetical Pleadings'[13], with the groan-inducing remark:

> "If A, then B" is equivalent to the statement "either A is false or B is true". To persons not engaged in the study of logic this may at first appear absurd; yet it can be proved.

In Boole (1847) p.54, the supposed equivalence - except with the negation in the conditional instead of the disjunction - is baldly stated:

> To express the conditional Proposition, If X be true, Y is not true.
> The equation is obviously
>
> xy=0, (37);
>
> this is equivalent to (33), and in fact the disjunctive Proposition, Either X is not true, or Y is not true, and the conditional Proposition, If X is true, Y is not true, are equivalent.

Earlier still, according to Ashworth (1968), 'The Spanish scholastic, Petrus Fonseca ... [wrote] that the name 'hypothetical' most properly applies to conditional propositions, but can also be used of disjunctions, because they imply a conditional.' Ashworth tells us that Abelard discussed the point in his *Dialectica*. It is known that Abelard learnt about the theory of hypothetical syllogisms from Boethius, whose *De Hypotheticis Syllogismis*, written during the years 516–22, appears to be the earliest extant source of the idea:

> [1.3.2] Fiunt uero propositiones hypotheticae etiam per disiunctionem ita:
>
> Aut hoc aut illud est.
>
> Nec eadem uideri debet haec propositio quae superior, quae sic enuntiatur:
>
> Si hoc est, illud non est.
>
> haec enim non est per disiunctionem sed per negationem.[14]

---

[13] Hankin (1924).
[14] This may be translated as:
[1.3.2] But propositions become hypothetical also through disjunction, thus:

2.4.3.2. Criticism of the argument

Here is the Or-to-If argument in schematic form, with connectives employed abbreviationally, in accord with what was said in previous subsections:

1. $\sim A \vee B$.	(Premise)
2. A.	(Hyp)
3. B.	(1, 2, Disj. Elim.)
4. If A then B.	(2 - 3, Cond. Proof)

Consider the following instance:

1. ~grass is green ∨ grass isn't green.	(Premise)
2. Grass is green.	(Hyp)
3. Grass isn't green.	(1, 2, Disj. Elim.)
4. If grass is green then grass isn't green.	(2 - 3, Cond. Proof)

I think there is something wrong with this argument, and I suspect most unindoctrinated people who comprehend it would agree. If a demon somehow *convinced* me of the truth of '(~grass is green ∨ grass isn't green)', and if I were rational, I would conclude that grass isn't green. In that situation, it would *not* appear rational (valid, truth-preserving) to conclude further that if grass is green, then grass isn't green. Of course, a defender of '⊃' as 'if' will argue that I have been deceived by appearances on this point. I have tried to undermine the motivation for this in subsection 2.4.2. However, the question remains: *what* should we say is wrong with the argument?

The fallacy occurs, I think, in the step of discharging the hypothesis and deriving a conditional. That is, in the application of the rule of Conditional Proof (roughly speaking, the natural language analogue of the Deduction Theorem for PC - I say 'roughly' because DT is strictly a metatheorem, not a proof-rule). Notice that, together with (2) (whose scope it appears in), (3) is an absurdity; it *can't* be that grass is and isn't green. Accordingly, I propose that CP becomes unavailable once an absurdity has been derived within the scope of the supposition. (Here I count as an 'absurdity' anything which, when conjoined with the supposition, yields an absurdity in an ordinary sense.) That CP is unavailable in such circumstances should not be surprising; if it were not so, all sound *reductio* arguments could be used to establish bizarre conditionals.[15]

Either this is, or that is.

Neither should the proposition pronounced as follows:

If this is, then that is not.

seem the same as the one above. For this one is not through disjunction but through negation. (Thanks to P.V. Spade for this translation.) Boethius intends exclusive disjunction. To help corroborate this interpretation: Lagerlund (2010) makes much the same suggestion (without specific reference to the text).

[15] This constraint is arguably insufficient to make Conditional Proof valid. There will remain the problem of Strengthening the Antecedent, and perhaps others. For a more thorough treatment of this matter, see Thomason (1970) (thanks to Adrian Heathcote for the reference). According to King (2004), Abelard rejected something like Conditional Proof. Given his interest in the semantics of conditionals, it is conceivable that his reasons were closely related to ours.

Essentially the same point can be seen from another side, if we change the premise to something we actually believe, such as: ~grass is blue ∨ grass isn't blue. Coming to step (2), in this case the hypothesis that grass is blue, if we really want to assume this hypothesis for the sake of argument, then we can hardly use the above disjunction in the ensuing reasoning, *unless* we are trying for a simple reductio of the proposition that grass is blue. And that would be epistemically queer, since it is hard to see how we could rationally be more sure of the disjunction than the "conclusion" that grass is not blue.

What I think all this shows is that the Or-to-If Argument form is not generally valid, as it would have to be if '⊃' could be read as 'if...then'. Therefore '⊃' cannot be read as 'if...then'. There is, of course, much more to say, in particular concerning the wide range of cases in which one seemingly *can* argue from Or to If; it seems that while '⊃'-sentences aren't conditionals, assurance of the truth of a '⊃'-sentence can in many cases serve as a *basis* for a conditional. The common talk about ordinary conditionals differing from '⊃'-sentences in asserting some kind of natural "connection" between antecedent and consequent is, for this reason, highly suspect.

For a differently orientated discussion of the Or-to-If Argument which culminates in the same verdict - that it is not valid - see Bennett (2003).

3. The truth-conditional approach

The crucial thing about this approach is that the meanings of compounds are supposed to be given by means of stipulation of their truth-conditions. It is perhaps the presently dominant approach to our main question. The meaning of atoms may be supposed to be determined either by means of abbreviation, or by truth-conditions. Our discussion will be taken up with two related sorts of issue: (1) issues to do with the semantic level of the definitions, i.e. use-mention issues, and (2) uniqueness issues: on this approach, meaning is in some sense meant to be determinable by truth-conditions; the worry is that a mere set of truth-conditions will not determine a *single* meaning, but will rather leave open a class of possible meanings.[16] Issues of semantic level could, at least in part, be subsumed under the heading of uniqueness issues, but they seem particularly important to our discussion, so I consider them separately.

Talk of truth-conditions is common in semantics and philosophy. We must not let this familiarity prevent us from engaging in a proper examination. So, before talking about the practise of stipulating truth-conditions for PC formulae, I shall first attempt to get a bit clearer, with some distinctions, about truth-condition talk in general, before working up to PC. I am not satisfied that I have attained a clear view on the matter. The topic seems more difficult than is commonly supposed.

3.1. On giving truth-conditions in general

[16] I learnt of the idea of uniqueness problems with respect to compounds (i.e. the definition of the connectives) from unpublished work of Adrian Heathcote, where it is suggested that the idea was partially anticipated in Church (1956). Both Church and Heathcote meet these challenges with sophisticated versions of the denotational approach. In the present discussion, my aims are quite different - to explore the problem, rather than to circumvent it. And my investigation has suggested to me that the philosophical problem raised in this context (compounds of PC) is actually a general problem attaching to truth-conditional stipulation. (A reference pertaining to the general problem is given in the next footnote.)

3.1.1. Some distinctions

One relatively clear distinction is between *stipulative* and *truth-apt* talk about truth-conditions: In stipulative cases, we are laying down a convention, so there is no question of truth or falsity (but there are questions of meaning, coherence, utility, etc.). In the truth-apt case, we are making an assertion about something which already has meaning. It is stipulative talk we are interested in here.

We shall now look at some cases and define some terms with which to carve up the field. (Some of these terms could be carried over to truth-apt talk.)

Consider this case:

(StipGladknow) Let an agent S gladknow a proposition P iff S knows P and S is glad to know P.

(I modelled this stipulative sentence on sentences used truth-aptly by would-be conceptual analysts of knowledge.)

On the LHS of the 'iff', the object of definition is not *mentioned*, but *used* in a special way. Let us call cases with that feature *object-use* cases - to be contrasted with *object-mention* cases. Furthermore, the object of definition is not a sentence, but a part of one - a predicate. (Or, we might say, we are partially defining all sentences with such a part.) Let us call such cases *subsentential object* cases - to be contrasted with *sentential object* cases, which we will later discuss and distinguish further among. Note further that components *used* in the LHS ('S' and 'P') are also *used* in the RHS: call cases with this feature *use-then-use*. We can similarly speak of cases which are *use-then-mention*, *mention-then-use*, *mention-then-mention*, and shall call cases were nothing used or mentioned on the LHS is used or mentioned on the RHS *then-else* cases. Note that these features are not all mutually exclusive: for example, a sentence might, on its LHS, use some expression E1, then mention E2, and then the RHS might *mention both* E1 and E2: such a case would be both *use-then-mention* and *mention-then-mention*. I do not use all of these terms in what follows, but it is helpful to have such vocabulary available in advance.

3.1.2. The uniqueness problem for predicates

As we said, (StipGladknow) is a subsentential-object case - more specifically, a predicate is being defined. The case of predicates raises a particular kind of uniqueness issue: the need for unique intension. Let us discuss this as practise for when we go on to consider PC.

It is widely recognized that the meaning of predicates has to do with intension, and that intension is something over and above extension. A classic example concerns the predicates 'renate' and 'cordate', which mean 'kidney-possessing' and 'heart-possessing' respectively. It is said (and we shall assume) that, as a matter of fact, everything which has a heart also has a kidney: thus 'renate' and 'cordate' have the same extension. But they clearly differ in meaning, and hence to settle meaning one needs to settle more than just extension. So if (StipGladknow) is a proper definition, it must settle intension also. It seems as though, in practice, it does indeed. But it does not do this in an *explicit* way, in the sense that no phrase is used to *pick out* or *identify* an intension for the object of definition, or even to pick out a synonymous object (i.e. an object with the same intension). Our question is thus: how is intension settled, here? In other words, how is it that

(StipRenate) Let a creature C be R iff C has a kidney

gives 'R' the intension of 'renate' rather than that of 'cordate'? After all, both of the following truth-apt sentences will be true once (StipRenate) is made:

(AptRenate) A creature C is R iff C has a kidney.
(AptCordate) A creature C is R iff C has a heart.

I think the answer is, roughly: (StipRenate) works by *ensuring* (AptRenate) - that is, making it analytic. (StipRenate) brings it about *by itself* that (AptRenate) is true, whereas (AptCordate)'s truth is a matter of further fact. We might try to capture the point as follows: (StipRenate) isn't simply saying 'give "R" an intension such that it gets an extension such that (AptRenate) comes out true', but rather 'give "R" an intension such that (AptRenate) comes out analytic'.

That's still not quite right. Consider:

(AptConj) A creature C is R iff (C has a kidney and C is either dead or alive).

Given that it is analytic that creatures are either dead or alive, it would seem that (AptRenate) and (AptConj) *therefore* stand and fall together with respect to analyticity. So we still have a uniqueness problem: there are at least two possible intensions of 'R' such that (AptRenate) comes out analytic: 'kidney-possessing' and 'kidney-possessing and either dead or alive'.

Suppose we accordingly amend our last characterization of (StipRenate) to: 'give "R" an intension such that (AptRenate) comes out analytic *purely in virtue of this stipulation plus the meanings of auxiliary terms* such as 'let', 'true', 'iff', etc.'. This, I think, rules out the intension 'kidney-possessing and either dead or alive', since the resulting analyticity of (AptConj) also rests on facts about the meanings of other words. (One worry about this fix is that perhaps the meanings of the definition's auxiliary terms will be enough to generate unintended candidate intensions. The intuitive idea behind it seems sound, however: definitions like (StipRenate) only explicitly *talk about* extension, but they do it in *such a way* as to implicitly provide the intension; we understand that we are to take the path of least resistance - that we have now got what we need, and should not introduce extraneous aspects of meaning.)

Now to PC.

3.2 On giving truth-conditions for PC

3.2.1. The definition of atoms

Atoms may be taken to abbreviate sentences which already have meaning, in which case their definition is quite unproblematic. Now we consider the case where they are given truth-conditions.

Presumably, we do not want definitions of atoms to be object-use. That is, they should not take the form, where '*p*' is an atom:

Let *p* iff ...

As a matter of fact this is never done, and it seems odd, for reasons which should become apparent in a moment. So we shan't consider this option any further. We want an object-mention definition. That is, something of the form:

Let '*p*' be true iff ...

Note that this form of definition seems quite different from what we were considering before (e.g. 'Let a creature C be R iff ...'). It is a sentential-object definition. Now, this is not the only kind of sentential-object definition. One may give a kind of general truth-conditional definition of a logical form, for example this "definition" of the subject-predicate form: 'Let “P*s*” be true iff “P” expresses a property which is possessed by an object which is denoted by “*s*”'. There is also the kind of case where one defines specific but *multiple* parts of a sentence in one pass. Our present case is unlike all those, in that on the LHS, we have been given an expression which is *simple*; no logical form has yet been indicated. So, in this case, unlike the predicate definitions we considered earlier, the logical form of the object of definition must be supplied on the RHS.

Unless '*p*' is to be self-referential, its definition will be what we called 'then-else': no expression used or mentioned on the LHS will be used or mentioned on the RHS. A question now arises about semantic level; suppose we want to define '*p*' to say something about snow, namely that it is white. Should we say:

Let '*p*' be true iff snow is white

or

Let '*p*' be true iff 'snow is white' is true?

Both can seem right (in virtue of the T-schema, effectively): the first definition can seem right because 'snow is white' is about snow, and *it* is true iff snow is white, just like '*p*' on this stipulation. The second can seem right insofar as it gives one the idea that, since '*p*' and 'snow is white' will be true under the same conditions, they must come to the same thing. But we cannot afford to be indiscriminate about this. We need a way of clearly defining '*p*' so that it says something about the English sentence 'snow is white', namely that it is true - *as distinct from* saying something about snow.

This difficulty can be partly removed by simply insisting that, whichever train of thought is followed, one then remains consistent. But what is it to remain consistent, here? One feels like saying: in the first case, one has to give, on the RHS, a sentence which has *the meaning* one intends, and in the second case, one must name such a sentence and attach a truth-predicate. Otherwise, one would still be haunted by the T-schema. But then both conventions are, pragmatically speaking, just ways of providing a sentence which *has the intended meaning* of the symbol to be defined - so the formulation in terms of truth-conditions starts to seem quite idle and misleading.

The problem above is compounded by (other) uniqueness issues. If we stipulate only that '*p*' is to be true iff snow is white, and if 'snow is white and snow is black or not black' is also true iff snow is white, what is to stop '*p*' meaning *that*? And what if we actually *wanted* '*p*' to say something tautologous? How could we differentiate it from any other tautology by means of *truth-conditions* if all tautologies are true in all conditions?

Here one cannot convincingly reply: 'it doesn't matter, since it is idle to interpret atoms as tautologies'. Firstly, this is just a pragmatic evasion, but secondly, it overlooks the fact that some tautologies are not *obviously* tautologous. And the sorts of things conventionally recognised as tautologies are just the tip of the iceberg: what about other kinds of statements which, in some sense, seem to be true in all conditions? '2 + 2 = 4', for example - or more complicated mathematical truths?[17]

And now the thought arises - what about *impossible* "conditions"? - perhaps '2 + 2 = 4' *isn't* true in all conditions after all, since it would be false if 2 + 2 equalled 5. But what sort of condition is that? There is an ambiguity in the word 'condition': one may talk of the real condition of things, and also about conditions as proposition-like objects. And if we let in impossible conditions such as '2 + 2 = 5', it starts to look like we are dealing with the second meaning, in which case it is unclear what advantage there is in the "truth-conditional" form of definition, as opposed to just *giving the meaning* - producing an expression with the intended meaning, or referring to the proposition one intends.

We are not doing very well so far. I think we ought to consider seriously the possibility that the whole idea that one can *stipulate* truth-conditions for expressions at all, and furthermore that one *thereby* determines their meaning, cannot stand up to scrutiny. One thing which makes this hard to accept is that, since the definitions appear to often work in practise, it might seem like you really *can* get meaning from truth-conditions alone, if you just get into the right spirit about it, so to speak. But perhaps we ought to regard this as an illusion. (I.e., perhaps the things we call 'truth-conditional definitions' have a "misleading surface-structure".)

Let us try to look objectively at the idea that one can, by stipulating truth-conditions, give a propositional meaning to a simple symbol (which may be a rock for all it matters), as though we had never heard of the idea in logic and philosophy (where it has the appearance of actually doing work). In fact, let us imagine that our first exposure to the idea is this: someone, quite out of the blue, pulls a stone out of their pocket, and says to us 'Now, let this stone be true if, and only if, snow is white.' We might say 'What do you mean "let this stone be true"? A stone can't be true! A stone doesn't *mean* anything.' And one could have a similar thought about a symbol on a piece of paper. It can't be true in *any* situation unless it *already* has a meaning. It is as though the stipulation of an atom's truth-conditions is elliptical for something like: 'let '*p*' *mean something* such that it is true iff ...'. But then, of course, uniqueness problems arise. One could perhaps start to add more and more constraints, until just one possible meaning is left.[18] But what would be the point of such a practise? Why not just come out and say what '*p*' is to mean, letting truth-condition talk fall by the wayside?

Perhaps philosophers favour superficially truth-conditional talk because it enables them to get their hands semantically dirty without raising Quinean alarm-bells about the definiteness of meaning-talk. Quine himself - after championing the abbreviational approach in the early-mid 'thirties - was one of the early supporters of

---

[17] For general discussion of some of these issues, as well as a source for further references, a good starting-point is Heck (2002).

[18] I say 'perhaps', because something strange happens when we try to think of constraints: it is natural to bring in considerations of grammar, reference and use-mention, however these notions usually apply to *expressions*, not *meanings* thereof. (This raises difficult issues to do with thought.)

the truth-conditional approach. (Cf. his (1940).)

3.2.2. The definition of compounds

We are now working under a considerable shadow, but we must press on, since the definition of compounds is the most crucial part of the truth-conditional approach, and there are significant differences from the atomic case.

The truth-conditional definition of compounds typically proceeds something like this:

For all formulae 'A' and 'B',
'~A' is true iff 'A' is false.
'A ∧ B' is true iff 'A' is true and 'B' is true.

and so on for other connectives. For '⊃', some authors have used 'if' on the RHS, and made some comment to the effect that truth-functionality is being assumed. The very notion that 'if' *has* a truth-functional meaning being controversial, most modern authors proceed with a less controversial RHS such as '"A" is false or "B" is true'.

One can also regard the "truth-tables" as encoding these definitions.

Note that these are mention-then-mention definitions. As with atoms, we must be careful about semantic level. In this connection, we must distinguish this approach from one where we semantically descend, to the level of use, for the RHS. On this second approach, which we classify as mention-then-use, the above definitions must be replaced with schemata:

'~A' is true iff it is not the case that A
'A ∧ B' is true iff A and B.
etc.

As with the definition of atoms, we get the problem of uniqueness in the definition of compounds. However, there is an important difference from the atomic case, which may affect how we think about the problem: with compounds, on the LHS we are given not a simple sign, but something with a grammatical form. In terms of our previous discussion of the lack of explicitness of truth-conditional stipulations, this may give rise to a difference in the functioning of these compound definitions: perhaps in this sort of case, the syntax of the LHS guides its semantics. For example, the fact that there are no quoted expressions *within* '$p \wedge q$' suggests that the compound is to be on the same semantic level as the components, referring to the same things as they do. Furthermore, the absence of anything else besides '∧' suggests that nothing *else* is to come within the terms of reference.[19] This calls to mind the notion of 'taking the path of least resistance', as mentioned in 3.1.2. when discussing the case of predicates.

What is the moral of all this discussion? I am not sure, since there appear to be two quite different morals we could draw, but I can't see any good reason to prefer one over the other. Here's one possible moral we could draw: truth-conditional stipulative definitions to not literally *say* enough to pin down a single meaning for the

[19] Here again we get the strange situation mentioned in the previous footnote, although in this case the problem is less vicious because we are not talking about official definitions, but merely trying to understand the pragmatics of the situation we are in when we work with truth-conditional definitions of compounds.

expressions being defined - they say something like 'Give this expression a meaning such that it is true iff ...' - and so, if and when they are used successfully as definitions, this is in virtue of 'pragmatics': an understanding between communicators that goes beyond knowledge of literal meaning. Alternatively, we could say: given that truth-conditional stipulative definitions *always* require the kind of interpretation we just called 'pragmatic', and given that there appears to be a considerable amount of uniformity in this - the path of least resistance, etc. - we should regard all this as part of their semantics, admitting that they may have a misleading (because elliptical, in a sense) surface structure. How to choose? What is at stake? This is essentially the problem of the semantics/pragmatics distinction and its use in philosophy. Since I presently have nothing special to say about it, I close this section with what I find to be a helpful analogy, from Wittgenstein: use flows from meaning in the sense that behaviour flows from character. If someone behaves quite differently from before, for a short time, we might call it 'out of character'. But the more lasting and integrated the change, the more we will be inclined to say that the person's character has changed.

4. The modelling approach

Since I know of no previous philosophical discussion of this approach to PC, this section will be largely expository, in contrast to the more investigative previous sections.

The previous three approaches have in common that they are supposed to make formulae represent *by themselves* - more specifically, as linguistic objects which refer or have truth-values.

By contrast, on this approach, formulae don't denote, abbreviate or have truth-conditions. Rather, they are reckoned as elements of an abstract structure which can be stipulated into a representational relation to the world. This can be done with both valuation- and proof-theory. I will first explain this using valuation-theory, and will then show how the same thing can be done using proof-theory.

Here is an example of a similar kind of non-linguistic representation: near where I live, there is a fire danger sign consisting of a half-dial and an arrow to indicate the level of danger. (The arrow is manually re-positioned as perceived conditions change.) The arrow by itself does not mean anything, and neither does the dial. But *that* the arrow is in such-and-such a position *does* mean something. Similarly, on this approach to PC, a formula by itself does not mean anything, but *that* it has such-and-such a value (1 or 0, in the classical case) can mean something. But there is an important difference between the fire-danger dial and PC here: the fire-danger dial is a model which we have to configure manually and directly. If we say 'Let the fire-danger dial correctly represent the current level of fire danger', this incantation will, of course, achieve nothing. We *are* able to make an analogous kind of stipulation with respect to PC conceived as an abstract model, however. This is a sort of modelling whereby the model (if successfully configured to represent at all) will always be correct, although we may not know what state it is in, and indeed may be wrong about what state it is in. This should become clearer with some illustrations.

4.1. Valuation-theoretic interpretation

In order to see how the "machinery" works, let us first consider a non-representational case where we simply stipulate values for atoms directly:

Let $v(p) = 1$ and $v(q) = 0$.

This can be read as 'Let the value of "*p*" be 1 and the value of "*q*" be 0'. (In the symbolic definition, formulae are used autonymously, i.e. to name themselves. The values, incidentally, don't have to be 1 and 0 - any two objects will do.) Given this stipulation, we can talk truth-aptly about the values of '*p*' and '*q*'. Values for all compounds made up of those atoms are then determined via the value-functions or value-tables. Thus, in virtue of the above stipulation, $v(\sim p) = 0$, $v(q) \neq 1$, $v(p \vee q) = 1$, $v(p \wedge q) = 0$, etc. So, if we don't know which values were stipulated, but are told for example that $v(\sim p) = 0$, we can "work backwards" via the value-tables and conclude that $v(p) = 1$. Or if we are told that $v(p \vee q) = 1$, we can work out that either $v(p) = 1$, or $v(q) = 1$, or both.

In order to bring this apparatus into a representational relation to the world, we stipulate the values of the atoms *conditionally*. For example:

Let $v(p) = 1$ if snow is white, 0 otherwise.
Let $v(q) = 1$ if grass is green, 0 otherwise.

With these stipulations in force, one can then assert things about the values of formulae involving '*p*' and '*q*', and these assertions will have implications concerning the colours of snow and grass. For example, if someone asserted that $v(p) = 0$, most of us would disagree, since that assertion now carries, via the stipulation, the implication that it is not the case that snow is white. Likewise if someone asserted that $v(p \wedge q) = 0$; this, via the value-tables, implies that it is not the case that $v(p)$ and $v(q) = 1$, which implies that it is not the case that snow is white and grass is green. In a setting where people trust each other to give correct statements, statements about the values of formulae could in this way be used for the purpose of conveying information.

4.2. Proof-theoretic interpretation

Now let us see how an analogous thing can be done with proof-theory. We do not proceed by associating formulae with one of two objects. Corresponding to that stipulative basis of the valuation-theoretic interpretation, we have in this case the population of a 'start-set' S according to the following convention: for each atom '*p*' to be used in the interpretation, exactly one of either '*p*' or '~*p*' is to be a member of S. As before, we shall illustrate the "machinery" before using it representatively. As a proof-theoretic equivalent to our stipulation that $v(p) = 1$ and $v(q) = 0$, we have:

Let $p \in S$ and let $\sim q \in S$.

Given this stipulation, it becomes true to say that '*p*' and '~*q*' follow from S (trivially) via proof-theory. In symbols:

$S \vdash p$ and $S \vdash \sim q$.

It also becomes true to say that, e.g., $S \vdash (p \vee q)$, $S \nvdash (p \wedge q)$, $S \vdash \sim(p \wedge q)$, etc.[20] As before, for a representational configuration, we stipulate conditionally:

[20] It may be wondered why we need to include tilded atoms (e.g. '~*q*') in S - that is, why it doesn't suffice that '*q*' is *not* in S. This is because we need compounds such as '$\sim(p \wedge q)$' to follow from S via ordinary proof-theory, and in ordinary proof-theory, nothing follows from the *absence* of a formula.

Let $p \in$ S if snow is white, and let $\sim p \in$ S otherwise.
Let $q \in$ S if grass is green, and let $\sim q \in$ S otherwise.

With these stipulations in force, we can assert that certain formulae involving '*p*' and '*q*' follow from the start-set, and these assertions will have implications concerning the colours of grass and snow. For example, the statement that S *p* implies, via the stipulation, that snow is white. Furthermore, if we come to believe that, e.g., S $\vdash (p \vee q)$, we can work out via proof-theory that either S $\vdash p$ or S $\vdash q$ or both, and thus that snow is white or grass is green (or both). This working out will typically be easiest in tree systems, followed by natural-deduction systems (in virtue of their having separate rules for each connective). In an axiom system which allows assumptions, the working out will be possible in principle but often very involved.[21] Soundness and completeness make value-tabular methods indirectly available also.

4.2.1. Summary

To recap: on this approach to PC, formulae have no linguistic meaning, but are reckoned as forming, together with either valuation- or proof-theory, elements of an abstract structure which can be stipulated into a representational relation. I find it natural to picture the valuation-theoretic structure as a mechanical object, with a set of levers or switches at the base (for atoms), which assume positions when we make stipulations about them, below a vast structure of connected 'compound' levers. The proof-theoretic structure could be pictured as a box into which we "insert" (by stipulation) "seeds" (i.e. atoms or tilded atoms), and out of which grows a vast, rigid tree-like structure, with nodes for compounds (or rather, compounds other than tilded atoms - in mathematical terminology, 'non-literals'). An alternative sort of conceptualization would be to envisage a space of all the possible structures (mechanisms, trees, etc.) in all possible configurations. The stipulations could then be thought of as narrowing the space down to a smaller set of structures, or just one.

The stipulations above were conditional on such things as grass being green, but of course "sentential" or "propositional" interpretations are possible as special cases of this technique.

This way of thinking about the interpretation of PC recommended itself to me as a way of answering our main question without getting involved in grammatical or logico-semantic difficulties. It keeps all that at arm's length, so to speak, from the logical machinery, which *itself* forms the means of representation. I have found this approach to be very helpful in thinking about nature of logic[22], and more specifically, valuation-theory, proof-theory and the close relationship between the two.

4.3. Sketch of an application to inferentialism

'Inferentialism' is a typically loose "ism" in philosophy. Inferentialism *about* a kind of

[21] In an axiom system which does not allow assumptions, an amended interpretation strategy would be required, exploiting the correspondence between the consequence relation and '$\supset$' formulae. The start-set would in this case be thought of as a set of axioms.

[22] Note that, on this approach, we can alter or mangle our value tables, rules or axioms in any way we please, but this will never lead us into *falsity*. We might lose "expressive power", and the worst that can happen is that the system becomes completely trivial and can no longer represent anything. This is reminiscent of the remark in the *Tractatus*, 'In a certain sense, we cannot make mistakes in logic.' (5.473).

expression X is canonically described as the view that the meanings of expressions of kind X are determined by the 'inferential role' these expressions play in the language, or that their meaning can be given in terms of inferential rules. 'Inferentialism' by itself is sometimes used for a global view about the nature of meaning. That is far from our concern here.[23] Our concern is inferentialism about the meanings of the connectives of PC (hereafter 'inferentialism'), or rather, a debate which has been going on under that heading.

I do not want to give a false sense of sense here. The doctrine is vague and amorphous. The debate can look like a battle of precisifications; one party says: 'if *this* is what it means, it's wrong', and another will respond with: 'yes, but in *this* sense, it could be right'. No *definitive* sense ever seems to get attached to the form of words 'the meanings of the logical connectives are determined by the rules of inference'.

In a sense, what I want to say here involves just another precisification. However, my motivation is not to argue for or against - since I'm not sure there's anything *in particular* to argue about. Rather, I want to use the ideas of section 4 to try to clarify one of the most interesting and confusing threads of the inferentialism debate.

Before that, a word about a more well-known thread. In his (1961), Arthur Prior showed that not every set of natural deduction in/out (introduction/elimination) rules yields a "well-behaved" connective. His example, which he called 'tonk', is governed by the in-rule of '$\vee$' and the out-rule of '$\wedge$'. This enables anything to be inferred from anything: from 'A', 'A tonk B' can be derived by the in-rule, and from that, 'B' follows by the out-rule. This has been treated as some kind of argument against inferentialism, but it is not clear why; as noted by Peregrin (2009) and others, tonk would directly threaten the view that *any* set of in/out rules determines a meaningful expression, but all the inferentialist seems to maintain is that the rules belonging to the actual connectives we use - the "well-behaved" ones - determine *their* meaning. Be that as it may, the phenomenon brought to our attention by Prior has been of great technical interest to people, since it raises the question: what properties characterise the sets of rules which yield "good behaviour"? Research on this has been carried out, often under the heading 'proof-theoretic harmony'.[24]

Our present concern is a more newly active, less well-understood thread of the inferentialism debate, most recently initiated in Raatikainen (2008). Its substance, however, goes back to a book published by Carnap in 1943, *Formalization of Logic*, which was largely forgotten until Raatikainen printed his article.[25] The results obtained by Carnap were not framed in terms of the inferentialism debate, so in this sketch I will not discuss his book (which deserves our attention) directly, but rather will attend to one result in particular, as employed by Raatikainen. (Rich technical discussions of Carnap's result have appeared since Raatikainen's article. See especially Hjortland's (2009) PhD thesis, and Peregrin (2010). For a reply to Raatihainen, see Murzi and Hjortland (2009).)

4.3.1. Carnap's result

[23] Incidentally, global inferentialism, as a theoretical claim about the nature of meaning, seems to be in danger of undermining itself. A claim about the *nature* of meaning may have implications for the *meaning* of 'meaning' which fail to square with the claim itself - i.e., if the word 'meaning' turns out *not* to play an inferential role which fits the hypothesis that meaning has crucially to do with inferential roles.

[24] Examples: Belnap (1962), Prawitz (1965, 1971), Tennant (1978, 1987, 1997), Weir (1986), Milne (1994).

[25] Two notable exceptions are Smiley (1996) and Rumfitt (2000).

In valuation-theory, we assign values to atoms, and the value-tables determine values for compounds. The resulting complete valuation can be thought of as a total function from formulae to {1, 0} - a mapping on which every formula is assigned exactly one value. Consider the set of all such total valuation functions *V*. Now, there are obviously other total valuation functions which do not conform to the value-tables - for example, a function on which, for some formula 'A', both 'A' and '~A' are sent to 1. Let us call such functions 'non-normal'. We know by soundness and completeness that a formula will be a theorem of proof-theory (in our framework, derivable from the empty set - and thus, by monotonicity, derivable from all start-sets) iff it is assigned 1 by all functions in *V* (that is, all normal functions). What Carnap showed was that we can add certain non-normal functions to *V* without destroying soundness and completeness. The most straightforward example is the function *t* which assigns *all* formulae to 1. A formula F is a theorem of proof-theory iff it is assigned 1 by all functions in the set $V \cup \{t\}$. This holds, because in the case where F is not a theorem, there will be some function in $V \cup \{t\}$ which sends F to 0, and in the case where F is a theorem, all functions in $V \cup \{t\}$ will send F to 1. In neither case does *t* get in the way.

Raatikainen's gloss of this result is:

> It can be shown that no ordinary formalization of logic, and not the standard rules of inference (of the natural deduction) [sic] in particular, is sufficient to 'fully formalize' all the essential logical properties of logical constants. That is, they do not exclude the possibility of interpreting logical constants in any other than the ordinary way.

I do not wish to say this is wrong - only that it does not give the full picture. Note first what Raatikainen means by 'interpretation': he means a total valuation function conceived as a set of truth-value assignments. This is confirmed when he writes of proof-theory not ruling out 'non-normal interpretations' such as *t.* From the point of view of the modelling approach to the interpretation of PC, this way of speaking rather begs the question against proof-theory.

4.3.2. A converse result

My object in this sketch is to show that, *from the present point of view* (on which valuation- and proof-theory are given "equal rights"), there is no asymmetry of the kind suggested - no underdetermination on the part of proof-theory. I will show that Carnap's result has an exact converse - that there are non-normal *proof-theoretic* "interpretations" which do not destroy soundness and completeness with respect to normal valuation theory.

To be more specific about what I mean by "equal rights": on the modelling approach, both valuation- and proof-theory are conceived as abstract structures; neither of them is, in itself, defined or described in terms of "the notion of truth" or any such thing. One can then stipulate both kinds of apparatus into a representational relation, and *perhaps* the notion of truth could be said to be implicit in this - but then it is equally available in both cases.

In both the valuation- and proof- theoretic interpretation procedures explained earlier in this section, we are making what could be called a *duality assumption*. More specifically, we have:

The Valuation-theoretic Duality Assumption: In an interpretation, every atom is to get exactly one value. Then *via the value-tables*, for every formula 'A', one of 'A' and '~A' will get the value 1, the other 0.

The Proof-theoretic Duality Assumption: In an interpretation, the start-set S is to contain, for every atom, either that atom or its tilde, but not both (and via proof-theory, for every formula 'A', either 'A' or '~A' will be derivable from S, but not both).

The non-normal valuation *t* which we added to *V* breaches the Valuation-theoretic Duality Assumption. This can be seen from two angles. One can say that the value-tables, such as that for '~', are contravened. Or, one can say that the value-tables are still in effect, and so formulae will now have more than one value: by definition of *t*, for any formulae 'A' and '~A', both will have the value 1, *but*, via the table for '~', both with thereby also get the value 0.

Now: consider the set of all start-sets *AS* populated in accordance with the Proof-theoretic Duality Principle (i.e. the 'normal' start-sets). As we observed before, a formula can be derived from all sets in *AS* (including the empty set) iff it is a valuation-theoretic tautology (i.e. gets 1 on all normal valuations). Let us consider the set $AS \cup \{N\}$, where *N* contains all formulae whatsoever. Our converse to Carnap's result can now be stated: a formula A can be derived from all sets belonging to $AS \cup \{N\}$ iff A is a valuation-theoretic tautology. This holds, because if A is *not* a tautology, then there will be some member of *AS* (and thus of $AS \cup \{N\}$) from which A cannot be derived. If it is a tautology, then of course it will be derivable from *AS*, and *N* does not get in the way, since everything is derivable from *N*. $AS \cup \{N\}$, of course, breaches the Proof-theoretic Duality Principle, just as $V \cup \{t\}$ breached the Valuation-theoretic.

We could now turn Raatikainen's gloss on its head, and say that the normal valuation-theoretic apparatus does not rule out non-normal "interpretations" of the proof-theoretic apparatus.

To stress, this is not meant to be an argument *against* Raatikainen, nor a vindication of inferentialism. The point is to show how the modelling approach can shed light on Carnap's result, which has recently become an important factor in the inferentialism debate.

4.3.3. The tilde and meaning-determination

A final word is in order about the involvement of the symbol '~' in the stipulative base of the proof-theoretic apparatus, with just double-negation elimination and introduction in the proof-theory, in contrast to the valuation-theoretic case, where we stipulate one of two values for each atom, and have '~' entirely taken care of by its value-table. In *this* sense, it might be argued that the value-tables in some sense contain *more* meaning with respect to '~' than the proof-theoretic rules/axioms. Be that as it may, we can make sense of the difference: roughly speaking, proof-theoretic interpretation uses tilded atoms (along with atoms) to do the job done in valuation-theory by {0,1}.

4.4. Sketch of an extension to quantification theory

For PC, the proof-theoretic notion of 'following from a start-set' in 4.2 corresponds to the valuation-theoretic notion of 'getting 1 on a valuation'. Here, the object is to show that proof-theory can play an analogous role with respect to quantification theory (the quantificational calculus, QC).

The task is to define a *proof*-theoretic, model-theoretic notion which is equivalent to the valuation-theoretic, model-theoretic notion commonly called 'truth on a model' (or 'truth in a model').

A clarification is in order. 'Valuation-theory' in this context (QC) *still* means the theory of assigning values 1 or 0 (truth or falsity) to *formulae* and nothing else - the notion of a semantic value, or mapping, for a term or predicate of QC is *not* considered as part of valuation-theory. Thus the notion commonly called 'truth on a model' is both valuation-theoretic (as indicated by the word 'truth') and model-theoretic ('model'). These two things are commonly put under one heading, 'semantics', as against proof-theory, which is filed under 'syntax'. In the present setting, however, we must keep them clearly distinguished and must beware of such traditional categories as 'semantic' and 'syntactic'.

So, for our task, we shall use proof-theory *in conjunction with model-theory*, in much the same way as valuation-theory is used in the standard concept of 'truth in a model'. We shall show how, when given a model M, to populate a start-set S such that a formula will follow from S iff it is 'true on that model' according to the standard definition. After this, we shall distinguish four conceptually different but co-extensive notions of tautologousness for QC. In keeping with our theme, the fact that all this can be done should shed light on the close relationship between proof- and valuation-theory.

4.4.1. Proof-theoretic 'Truth'-definition for QC:

I take the universal quantifier as basic.

Model structure: <D, *c*, *p*>
D: domain set
c: total function from constants to elements of D
p: total function from n-placed predicates to sets of n-tuples of elements of D[26]

All atomic formulae A contain a predicate P which is mapped by *p* to a set of n-tuples. Let us call this set *the instance set* of the atomic formula A.

Formulae and expressions are often used autonymously below.

A formula F which contains a single free variable *x* (which may occur multiple times) will be referred to as F(*x*). The result of replacing all occurrences of *x* in F(*x*) with some term *t* will be referred to as F(*t*/*x*).

4.4.1.1. Definition of start-set S of a model M

- If A is an *n*-place atomic formula containing *n* constants (i.e. no variables), let A $\in$ S iff for some n-tuple P in A's instance set, and for all places in *m* in the formula A, *c* maps A's *m*-th constant to the object occupying the *m*th place in the n-tuple P. Otherwise, let ~A $\in$ S.

[26] For ease of exposition, monadic predicates will be mapped to sets of 1-tuples.

- If A is an (n > 1)-place valuation-functional compound, let A ∈ Sv iff there exists a set D ⊂ Sy = {x: x is a component or tilded component of A} such that D ⊢ A, otherwise let ~A ∈ Sv.

- If A is of the form (∀x)B(x) where B is a wf and x is the only free variable occurring in it, let A ∈ S iff, on all models MT which differ from M only in containing an extra constant t which is mapped to an object of the domain, A(t/x) is in MT's start-set. Otherwise, let ~A ∈ S.

- Let nothing else be in S.

4.4.1.2. Correspondence thesis

For all formulae A, given a model structure, S ⊢ A iff A is "true" in that structure in valuational model theory, and S ⊢ ~A iff A is "false".

4.4.2. Four notions of QC tautologousness/validity

The most well-known notions of tautologousness for QC, taught in introductory logic courses, are the valuation-theoretic model-theoretic notion of 'truth in all models', and the proof-theoretic notion of theoremhood (which makes no reference to models, of course). But there are at least two other kinds available.

A valuation-theoretic notion which makes no reference to models was given in LeBlanc (1968), following a remark in Beth (1959, §89). The result was stated as follows: 'a formula A of QC may be declared valid if satisfied by every assignment of truth-values to its atomic subformulas'. ('Satisfy' and 'subformula' are special terms defined in LeBlanc's article.)

Secondly, using our notion of a start-set for a model, we can easily define a proof-theoretic model-theoretic notion of tautologousness: A formula A of QC is tautologous iff it is derivable from the start-set for every model.

It is useful to have the four in a table:

Valuation-theoretic model-theoretic 'Truth in all models'	**Proof-theoretic model-theoretic** 'Derivability from all models' start-sets'
Minimal valuation-theoretic 'Satisfiability by all valuations of atomic subformulae'	**Minimal proof-theoretic** 'Theoremhood'

Bibliography

Adams, Ernest W. 1965. 'The Logic of Conditionals', *Inquiry* 8, pp. 167-197.

Adams, Ernest W. 1975. *The Logic of Conditionals*, Dordrecht, Reidel.

Ashworth, E.J. 1968. 'Propositional logic in the sixteenth and early seventeenth centuries' in *Notre Dame Journal of Formal Logic*, Vol. 9, No. 2, 179-192.

Belnap, N. D. 1962. Tonk, plonk and plink. *Analysis*, 22(6), 130-134.

Bennett, Jonathan Francis. 2003. *A Philosophical Guide to Conditionals*. Clarendon Press, Oxford University Press.

Beth, Evert W. 1959. *The Foundations of Mathematics.* Studies in Logic and the Foundations of Mathematics. Edited by L. E. J. Brouwer, E. W. Beth, A. Heyting. 1959. North Holland Publishing Company, Amsterdam.

Boethius, Anicius Manlius Severinus. 516–22. *De Hypotheticis Syllogismis*.
Original Latin available at Peter King's website:
<http://individual.utoronto.ca/pking/resources.html> (Last accessed: 18/10/2010.)
Published in Italian:
(Istituto di Filosoofia dell'Università di Parma, Logicalia 1).
ed. Obertello, L. Brescia, 1968.

Boole, George. 1847. *The Mathematical Analysis of Logic, Being an Essay Towards a Calculus of Deductive Reasoning*, Cambridge: MacMillan, Barclay and MacMillan. London: George Bell.

Bradley, F. H. 1893. *Appearance and Reality*, 2nd ed. 1897, ed. W. J. Mander, Bristol, 1999.

Carnap, Rudolf. 1943. *Formalization of Logic.* Cambridge, MA: Harvard University Press.

Church, Alonzo. 1956. *Introduction to Mathematical Logic Vol. 1*, Princeton University Press, Princeton.

Cohen, Morris R. and Nagel, Ernest. 1934. *An Introduction to Logic and Scientific Method*, London: Routledge & Kegan Paul Ltd.

Dudman, V.H. 1992. 'Probability and Assertion', *Analysis*, 52:204-11.

Edgington, Dorothy. 1991. 'Do Conditionals Have Truth-Conditions?' in Jackson. ed. (1991, pp. 176-201).

Edgington, Dorothy. 1995. 'On Conditionals', *Mind* 104.414., (Apr. 1995), pp. 235-329.

Faris, J.A. 1969. 'Interderivability of "⊃" and "if"' in *Logic and Philosophy: Selected Readings*, ch. 7, Iseminger, G., ed., Appleton-Century-Crofts, New York.

Frege, Gottlob. 1893. 'On Sense and Reference', reprinted in *Translations from the Philosophical Writings of Gottlob Frege*, Geach, P.T. and Black, M. (eds.), Oxford: Blackwell, 1952.

Frege, Gottlob. 1892. 'On Concept and Object', reprinted in *Mind*, New Series, Vol. 60, No. 238 (Apr., 1951), pp. 168-180. Translated by Geach, P.T. (revised by Black, M.).

Grice, Herbert Paul. 1975. 'Logic and Conversation', in *The Logic of Grammar*, D. Davidson and G. Harman (eds.), Encino, California, Dickenson, pp. 64-75. Reprinted in Grice (1989).

Grice, Herbert Paul. 1989. *Studies in the Way of Words*, Cambridge MA, Harvard University Press.

Hacker, Peter M.S. 'Frege and Wittgenstein on Elucidations' in *Mind*, New Series, Vol. 84, No. 336 (Oct., 1975), pp. 601-609.

Hankin, Gregory. 1924. 'Alternative and Hypothetical Pleadings' in *The Yale Law Journal*, Vol. 33, No. 4 (Feb., 1924), pp. 365-382.

Heck, Richard G. Jr. 2002. 'Meaning and Truth-Conditions: A Reply to Kemp' in *The Philosophical Quarterly*, Vol. 52, No. 206 (Jan., 2002), pp. 82-87.

Hjortland, Ole T. 2010. *The structure of logical consequence: proof-theoretic conceptions*. PhD thesis at The University of St. Andrews.
Available in PDF at <http://hdl.handle.net/10023/892>

Jackson, Frank. 1979. 'On assertion and indicative conditionals.' in *The Philosophical Review* 88, 565-589. Reprinted in Jackson, ed. (1991, pp. 111-135).

Jackson, Frank, ed. 1991. *Conditionals*. Oxford: Oxford University Press.

King, Peter, "Peter Abelard", *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/abelard/>.

Korhonen, Anssi. 2007. *Logic as the Universal Science: Bertrand Russell's Early Conception of Logic and Its Philosophical Context*, Philosophical Studies from the University of Helsinki 18 (Nov. 2007).
Available in PDF at <http://ethesis.helsinki.fi>.

Lagerlund, Henrik, "Medieval Theories of the Syllogism", *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2010/entries/medieval-syllogism/>.

LeBlanc, Hugues. 1968. 'A Simplified Account of Validity and Implication for Quantificational Logic' in *The Journal of Symbolic Logic*, Vol. 33, No. 2, (Jun., 1968), pp. 231-235.

Lewis, Clarence Irving. 1918. *A Survey of Symbolic Logic*, University of California Press, Berkeley, 1918.

Lewis, David. 1976. 'Probabilities of conditionals and conditional probabilities.' in *Philosophical Review*, 85(3):297–315. Reprinted with Postscript in *Philosophical Papers, Volume II*, pp. 133-152.

Lewis, David. 1986a. *Philosophical Papers, Volume II*. Oxford University Press, Oxford.

Lewis, David. 1986b. *On the Plurality of Worlds*. Blackwell, Oxford.

MacColl, Hugh. 1908. '"If" and "Imply"' in *Mind*, New Series, Vol. 17, No. 65 (Jan., 1908), pp. 151-152.

McDermott, Michael. 1996. 'On the truth conditions of certain "If"-sentences' in *The Philosophical Review*, Vol. 105, No. 1 (Jan., 1996), pp. 1-37.

Milne, Peter. 1994. 'Classical Harmony: Rules of Inference and the Meaning of the Logical Constants' in *Synthese* 100, pp. 49-94.

Murzi, Julien and Hjortland, Ole T. 'Inferentialism and the categoricity problem: reply to Raatikainen' in *Analysis* 69(3), pp. 480-488.

Ness, Arne. 1938. "Truth" as Conceived by Those Who Are Not Professional Philosophers (*Skrifter utgitt av Det Norske Videns- kaps-Akademi i Oslo, II. Hist.-Filos. Klasse.* Vol. 4.) Oslo: Jacob Dybwad.
(Ness also went by and published under his original name 'Naess'.)

Peregrin, Jaroslav. 2009. 'Inferentialism and the Compositionality of Meaning' in

International Review of Pragmatics 1, pp. 154–181. (Koninklijke Brill NV, Leiden.)

Peregrin, Jaroslav. 2010. 'Inferentializing Semantics' in *Journal of Philosophical Logic* Vol. 39 No. 3, pp. 255 - 274

Prawitz, D. 1965. *Natural deduction: A proof-theoretical study*. Stockholm: Almqvist and Wiksell.

Prawitz, D. 1971. 'Ideas and results in proof theory' in *Proceedings of the 2nd scandinavian logic symposium* pp.235-308. J. E. Fenstad, ed. Amsterdam: North Holland.

Priest, Graham. 2001. *An Introduction to Non-Classical Logic*, Cambridge University Press.

Prior, Arthur N. 1961, 'The Runabout Inference Ticket' in *Analysis* 21, pp. 38-39.

Quine, W.V.O. 1934. 'Ontological Remarks on the Propositional Calculus' in *Mind*, New Series, Vol. 43, No. 172. (Oct., 1934), pp. 472-476.

Quine, W.V.O. 1940. *Mathematical Logic*. Harvard University Press, Cambridge, Mass., revised edition, 1951.

Quine, W.V.O. 1953. 'Mr. Strawson on Logical Theory' in *Mind*, New Series, Vol. 62, No. 248 (Oct., 1953), pp. 433-451.

Raatikainen, Panu. 2008. 'On rules of inference and the meanings of logical constants' in *Analysis* 68.4, pp. 282-287.

Rumfitt, Ian. 2000. '"Yes" and "No" ', *Mind*, New Series, Vol. 109, No. 436, pp. 781-823.

Russell, Bertrand. 1903. *The Principles of Mathematics*.
Published with a new Introduction:
London: George Allen and Unwin. 1937.

Russell, Bertrand. 1906. 'The Theory of Implication' in the *American Journal of Mathematics*, 28, 159-202.

Russell, Bertrand. 1908. '"If" and "Imply", A Reply to Mr. MacColl' in *Mind*, New Series, Vol. 17, No. 66 (Apr., 1908).

(For *Principia Mathematica* see 'Whitehead'.)

Smiley, Timothy. 1996. 'Rejection', *Analysis*, Vol. 56, No. 1, pp. 1-9.

Stalnaker, Robert: 'A Theory of Conditionals', *Studies in Logical Theory: American philosophical quarterly monograph*, Oxford, Blackwell 1968, pp. 98-112.

Tennant, N. 1978. *Natural logic*. Edinburgh: Edinburgh University Press.

Tennant, N. 1987. *Anti-realism and logic*. Oxford: Clarendon Press.

Tennant, N. 1997. *The taming of the true*. Oxford: OUP.

Tarski, Alfred. 1944. 'The Semantic Conception of Truth and the Foundations of Semantics', in *Philosophy and Phenomenological Research*, 4, pp. 13–47.

Thomason, Richmond H. 1970. 'A Fitch-style formulation of conditional logic' in *Logique et Analyse*, 52:397–412.

Weir, Alan. 1986. 'Classical Harmony' in *Notre Dame Journal of Formal Logic* 27, pp. 459-

482.

Whitehead, Alfred North and Russell, Bertrand. 1910. *Principia Mathematica*, Vol. 1. Cambridge: Cambridge University Press. Second edition 1925.

Wittgenstein, Ludwig. 1918. *Tractatus Logico-Philosophicus* (Originally *Logisch-Philosophische Abhandlung*)
Published in parallel German-English:
trans. C.K. Ogden, assisted by Frank Ramsey and G.E. Moore, Routledge, 1922.

Wittgenstein, Ludwig. 1953. *Philosophical Investications*.
Published in parallel German-English:
trans. Anscombe, G.E.M. eds. Anscombe, G.E.M. and Rees, Rush.
Blackwell.