Chapter 11

Cognitive Significance Regained

11.1 Re-evaluating the problem of cognitive significance

The theory of meaning and theory of knowledge thrive on distinctions. We intuit them and motivate them. Then we formulate them and apply them. When we tire of that, or find our formulations defective, we turn to undermining them and discrediting them. Finally we discard them or (so we think) demolish them. Then we rediscover them, and start all over again.

In this chapter we take a fresh look at the problem of how one might formulate a criterion for cognitive significance: a criterion that will distinguish between, on the one hand, those sentences that depend, for their truth or falsity, on our understanding and experience; and, on the other hand, those sentences that would not so depend (even on the assumption that they had truth-values at all). I take heart in this endeavour from a recent concession by Quine. He writes:[1]

> Tennant writes: 'I believe there is still the prospect of fashioning a criterion that would correctly trace the lines of impregnation by empirical meaning from periphery inwards: a criterion that

[1]W. V. Quine, 'Comment', in W. Salmon and G. Wolters (eds.), *Logic, Language, and the Structure of Scientific Theories*, University of Pittsburgh Press, 1994, pp. 345–51; at pp. 349–50. The reference (Quine 1990) is *The Pursuit of Truth*, Harvard University Press. Quine's quotation is from 'Carnap and Quine', in W. Salmon and G. Wolters (eds.), *op. cit.*, pp. 305–44.

> will mark out "the best of these sentences." ' In Quine (1990) I remarked, 'It would be a Herculean labor, not to say Augean, to sort out all the premisses and logical strands of implication that ultimately link theory with observation if or insofar as linked they be' (p. 17). I agree that we should elicit these hidden lines of impregnation or strands of implication as best we can, and that a more discriminating analysis of cognitive significance might be hoped for than I have managed on the basis of 'critical semantic mass'.

The task, though, might turn out to be not too Augean. We shall be as hygienic as possible with logical notions in what follows.

Sentences that depend, for their truth or falsity, on our understanding and experience are those that are apt for assertion or denial. We allow for the possibility that some sentences counting as cognitively significant on our account will be prime targets for projectivist (or expressivist) construal, rather than descriptivist construal. That is, a modern irrealist may attempt to carve out some of his irrealist territory within the domain of what we shall be calling the cognitively significant. He may, for example, want to construe causal claims, or statements about theoretical entities, or hypotheses about a speaker's interpretation of an expression in his language, in an irrealist way. Yet these statements will be cognitively significant on the account to be offered here. This point is worth stressing, given that certain important areas of discourse that in the early days of Logical Empiricism were counted as not cognitively significant would be prime candidates for modern irrealist reconstrual: among them, most importantly, ethical discourse. We need to realize, however, that modern irrealists have ventured irrealist reconstruals even for areas of discourse whose status as cognitively significant was quite unproblematic for the early Logical Empiricists.

The distinction between sentences that are cognitively significant and those that are not cuts across, or is presupposed by, each of the following distinctions, should they be tenable:

- analytic/synthetic
- conceptual/empirical
- a priori/a posteriori
- (logically) necessary or impossible/contingent

- (metaphysically) necessary or impossible/contingent
- observational/theoretical
- decidable/undecidable
- reducible/irreducible (with respect to some favoured class of sentences)
- apt for descriptivist construal/apt for projectivist or expressivist (irrealist) construal

The reader will no doubt be familiar with the embarrassing failure of the most famous early attempt to explicate the notion of cognitive significance. It was Ayer's definition of 'indirectly verifiable' sentences, which he gave in the second edition of *Language, Truth and Logic*.[2] Church showed, in his *JSL* review,[3] that under very weak assumptions every sentence came out as indirectly verifiable on Ayer's definition. Later, Ullian sharpened and simplified Church's result.[4]

The inadequacy of Ayer's definition was about as distressing as it could be. It wasn't just that he located the distinction wrongly, with some sentences coming out on the wrong side of the dividing line, but with most of them correctly located. Rather, he got it spectacularly wrong. His criterion collapsed. It failed to drive a wedge between two non-empty classes of sentences; it just skidded right past the whole language. Every sentence was ruled in, and none ruled out.

Anyone who works now on the problem of cognitive significance does so with the spectre of such failure at their shoulder. It is with great trepidation, therefore, that a new criterion of cognitive significance is ventured here. We shall in due course be formulating some inductive definitions. One of our inductive definitions characterizes, for a sentence S and a theory Δ, and finite levels n, the notion 'S is in F_n with respect to Δ', where F_n will be the set of sentences falsifiable at level n with respect to a theory Δ.[5] Another

[2] A. J. Ayer, *Language, Truth and Logic*, Gollancz, London, 2nd edn., 1946. Isaiah Berlin had shown in pretty short measure that Ayer's simpler account in the first edition was inadequate. See his paper 'Verification', *Proceedings of the Aristotelian Society*, 39, 1938–9, pp. 225–48.

[3] A. Church, 'Review of Ayer's Language, Truth and Logic, 2nd Edition', *Journal of Symbolic Logic*, 14, pp. 52–3.

[4] J. Ullian, 'A Note on Scheffler on Nidditch', *Journal of Philosophy*, 62, 1965, pp. 274–75. We shall examine Ullian's result below, and shall show how to avoid it.

[5] In this chapter we shall depart from our earlier notational convention according to which the usual generic variable for a sentence is 'ϕ'. Here we shall use 'S' instead, since its sibilance reminds us of the word 'significant'.

characterizes, for a non-logical expression E and a theory Δ, and finite levels n, the notion 'E is in Λ_n w.r.t. Δ', where Λ_n will be the expressions legitimated as cognitively significant by level n with respect to Δ. Note that these notions are parametrized with respect to the theory Δ; although in our notation we suppress mention of Δ for easier reading.

We shall propose that:

> a sentence S is cognitively significant with respect to a given theory Θ
>
> if and only if
>
> if S is contingent then for some level n, S is a compound of expressions in Λ_n with respect to some subtheory of Θ.

It will follow that every non-contingent sentence will be cognitively significant with respect to every theory. But contingent sentences will be cognitively significant only with respect to theories that make them so. (They will do this by making those sentences beholden to basic sentences in a way to be spelled out in the formal theory below.) This new criterion, we submit, is faithful to the original intuitions of the logical empiricists, and reveals these intuitions to be coherent. It avoids collapse: the wrong sentences are ruled out. It avoids belittling: the right sentences are ruled in. It can be formulated simply and elegantly, as we shall see, with minimal logical materials, as a system of inductive definitions.

Cynicism about the prospects for a criterion of cognitive significance has become part of the contemporary philosophical tradition even among analytical philosophers. It is no exaggeration to say that the failure of past proposals, even in the absence of an impossibility proof, have led to a widely shared intellectual conviction that is now allowed to override strong intuitions (at least among the historically less informed) to the contrary. Now and again the central intuition will resurface even in the writings of a sophisticated philosopher of science and mathematics. An example is from Hartry Field's *Science Without Numbers* (at p. 14):

> there is a marked disanalogy between mathematical theories and physical theories about unobservable entities: physical theories about unobservables are certainly not conservative, they give rise to genuinely new conclusions about observables.

Precisely how physical theories do this, and, as a result, acquire their cognitive significance, is what we aim here (failed past attempts notwithstanding) to explicate. In his classic paper 'Empiricist Criteria of Cognitive Significance: Problems and Changes',[6] Hempel wrote

> I think that the general intent of the empiricist criterion of meaning is basically sound. . . . I feel less confident, however, about the possibility of restating the general idea in the form of precise and general criteria which establish sharp dividing lines . . . between those sentences which do have cognitive significance and those which do not.

Hempel's lack of confidence in such a possibility resulted from the demonstrable failure of the various criteria that had been put forward by the time of his writing. He did not claim to have any general impossibility proof. All he laid out was a record of piecemeal failure.[7]

Perhaps because Hempel gave such a pessimistic prognosis, the notion of cognitive significance fell from favour. No one appeared to be willing to devote any more intellectual effort to the task of clarifying it. Although Carnap himself never lost faith in the tenability of the notion, Quine's 'Two Dogmas of Empiricism' consummated the slide to pragmatic holism. Not only was the attempt to demarcate the cognitively significant sentences regarded as misguided and certainly fruitless; but also, the celebrated distinction within the class of cognitively significant sentences, between those that were analytic and those that were synthetic, was abandoned. One is tempted to wonder how the subsequent course of analytical philosophy of language and of science might have been altered had these two distinctions been rescued with adequate explications, and had their denial not become such widespread orthodoxy for decades. We shall not, however, enter into such counterfactual speculation here. Instead, we shall confine ourselves to picking up the pieces of the problem now that the dust has settled—all too heavily, unfortunately—on past attempts to formulate a criterion of cognitive significance. If the criterion that we are offering fails, it will, we imagine, fail miserably, just like its predecessors. Nevertheless, if it fails, but not miserably, it may be of some value. And subsequent improvements might lead one to embark on the counterfactual speculations that we must here defer.

[6]in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, Macmillan, New York, 1965, pp. 101–22. Note that Hempel's pessimistic conclusions were reached as early as 1950, in one of the original sources for the latter anthologized paper, 'Problems and Changes in the Empiricist Criterion of Meaning', *Revue Internationale de Philosophie*, 4, pp. 41–63.

[7]This assessment was confirmed by Professor Hempel in private discussion in Konstanz in May 1991.

11.2 Conditions of adequacy on a criterion of cognitive significance

11.2.1 Sentences are cognitively significant only within the context of a theory that makes them so

We owe to Hempel the insight that sentences are cognitively significant only within the context of a theory that makes them so. Certain sentences (usually: observation sentences) are obviously cognitively significant. These we shall call *basic*. The problem is to work out, in a principled way, which sentences involving non-basic expressions are cognitively significant, and which are not, with respect to a theory. (Non-basic expressions are usually called non-observational, or 'theoretical'.)

Consider modern physical theory with the theoretical terms 'electron' and 'quark'. The claim 'All electrons contain quarks' is cognitively significant because its constituent terms 'electron' and 'quark' have been 'legitimated' as the kind of expression that can be used to form a cognitively significant sentence. Each of these expressions has in turn been legitimated because it has been involved as a 'new' expression in some statement (within the theory) that has acquired the status of a cognitively significant sentence by some stage, by virtue of the special logical relationships it bears to sequents[8] that consist solely of sentences previously legitimated as cognitively significant.

The religious term 'angel' is not cognitively significant—at least, not by virtue of its having acquired that status within any current theory whose confirmation or disconfirmation derives from basic statements. It would be no counter to this remark about the term 'angel' to say that within current physical theory the term 'electron' could be uniformly replaced by the term 'angel' throughout, thereby conferring cognitive significance on talk about angels. This wouldn't be a case of cognitive significance being granted to talk about the kind of angels that current religious discourse is supposed by its participants to be about. The newly suggested talk about 'angels' (resulting from the proposed substitution of the term 'angel' for the term 'electron' in current physical theory) would serve only to make the term 'angel' render the same cognitively significant service that the term 'electron' currently does. If all our talk about angels had been the substituted talk all along, no one would have entertained any objection to the term 'angel' being cognitively significant. It would have been legitimated as such, and

[8]A sequent $X : Q$ is here thought of as a statement of an argument (not necessarily valid). It consists of a set X of premisses and a conclusion Q.

for precisely the same reason that our current term ‘electron’ is cognitively significant—namely, that it features in a physical theory, beholden to basic statements, in the way that it does.

It is when we can forge no sensible empirical connection between our current physical theory and some other stretch of discourse about ‘angels’ that we refuse to grant cognitive significance to the term ‘angel’ in that other stretch of discourse, or to contingent sentences within it that use that term. If, on the other hand, some stretch of discourse involving what might initially be regarded as the uninterpreted term ‘angel’ were to be brought into fruitful theoretical liaison with current physical theory, then we would grant cognitive significance to the term ‘angel’, and to contingent sentences containing it. But the term ‘angel’ would in that case, by virtue of the very grounds on which we grant it cognitive significance, be functioning very differently indeed from the term ‘angel’ of current religious discourse. It would not be the same kind of angel-talk at all.

These reflections point to the need for a definition of cognitive significance that allows one to grant cognitive significance to terms only because of the way they function within statements of a particular theory. So the predicate ‘. . . is cognitively significant’, when properly explicated, will be essentially relational, parametrized by theories. This point, as remarked above, we owe to Hempel.

11.2.2 Basic sentences

A point of terminological clarification is in order concerning the use we shall make of the words ‘basic’ and ‘basically’. These are theoretical labels in our analysis that would, in its most obvious application, be replaced by the words ‘observable’ and ‘observably’; or by the words ‘empirical’ and ‘empirically’. The distinction basic/non-basic is meant to mark the fact that some sorts of terms in our language—individual constants and predicates, or just plain propositional variables—have some sort of priority over the rest, and are accordingly called ‘basic’. The analysis or explication we have given rests only on the existence of such a distinction. Just what lies on each side of the distinction is irrelevant to the operation or application of the distinction itself. The logical virtue of our analysis is that it is neutral between competing choices of what is meant, substantively, by saying that certain items of vocabulary (or the things, properties or concepts for which they stand) are basic. The analysis serves only to give a clear sense to what it is for any given sentence of the language to be able to rest ultimately, for its own truth or falsity, on what is basically the case. It seeks to trace the lines of infection or impregnation, as it were, by one's preferred species of meaning. We want to analyse how it is that the truth-value of a sentence is grounded in a certain basic class of facts. Choice of basic class is another matter altogether. Thus the project should be of interest to any epistemologist or philosopher of language entertaining the prospect of any sort of evidential or conceptual foundationalism.

11.2.3 Metalogical neutrality

The treatment here enjoys a certain metalogical neutrality. We have no epistemological axe to grind in the choice of what sort of sentence counts as basic. All we are concerned to do is to show how, once one has chosen a class of basic sentences, the meanings of other sentences in the language can be appropriately conditioned by them, so that those sentences count as cognitively significant. Of course, the usual applications of a criterion of cognitive significance have in the past had observation sentences (or protocol sentences) as basic. But the emphasis on such perceptual primacy is unnecessary. All that matters is that the basic sentences should be interpreted, and that they should be accepted as cognitively significant.

It is not important to know which basic sentences a theory contains, in order to determine which sentences involving non-basic expressions will be cognitively significant within it. The role of cognitively significant sentences that involve non-basic expressions is, rather, to furnish inferential connections among basic sentences (in the first instance), regardless of which of these latter are held to be true, and which are held to be false. It is the provision of these inferential connections, prescinding from the question of what is basically the case, that allows higher-level theoretical laws to support counterfactuals. Indeed, high-level physical theories, even though evidentially constrained by observation sentences, avoid commitment to the truth or falsity of any particular observation sentences. The boundary conditions of experiments, and the observational findings they deliver, are of secondary importance in identifying the theory as a logical entity.

One need not be worried unduly by the fashionable contemporary insistence that we have no pre-theoretical grasp of what is to count as basic; and that certain 'theoretical' sentences are often used to make 'basic observational' reports within a developed scientific theory. First, this fashionable inclination to blur a commonsense distinction ignores the extent to which one really could retrench, if need be, to a region of pre-theoretic ordinary discourse in order to make theoretically unalloyed observation reports. Secondly, the invitation here is to entertain the basic/non-basic distinction as *tentative* and *revisable* in the context of seeking to reconstruct one's scientific theory so as to show that all of it is cognitively significant according to the criterion to be given below. Eventually *some* division of sentences into the basic and the non-basic ought to be found that will allow one to do this. Moreover, it would be reasonable to require that, once such a division has been tentatively effected, an ordinary English speaker who is not versed in the scientific theory in question should know, simply on the basis of his perceptual experience, how to apply or withhold the 'basic' terms on this tentative identification.

11.2.4 Inductive levels, new vocabulary and extension

One works out from the basic to the non-basic. The basic level can be embroidered upon. This metaphor is a useful one: it suggests colourful and sometimes intricate developments, using a network of threads, but threads that are always tightly rooted in the fabric of the basic. As we work out, we legitimate both constituent expressions and sentences involving them as significant, in so far as they are properly related to the basic. (Exactly how they ought to be so related will be defined in due course.) For the purposes of regimentation, we envisage this as involving finite levels (generically subscripted by n). With typical ambiguity, we shall call levels up to and including any given level n 'old', and levels from $n+1$ onwards 'new'. If the historical development of a theory closely matched the logical reconstruction in terms of levels, the 'old' theory would include just those expressions legitimated by level n; and the 'new' theoretical hypotheses would venture to introduce as yet unlegitimated expressions, making them legitimate at level $n+1$. They would do this by making possible the drawing of certain inferences among old sentences that could not be drawn before.[9]

These logical connections forged by new cognitively significant sentences involving new non-basic expressions eventually come to involve previously legitimated cognitively significant sentences other than the most basic ones, as these are recruited to a growing theory. The theory can grow not only by our adopting new statements within the current vocabulary. It can grow also because we adopt new hypotheses involving new vocabulary intended

[9] When talking thus of 'old' and 'new' theory, we of course have in mind theoretical *extension* rather than revolutionary theoretical *displacement*.

to refer to hidden mechanisms, unobservable entities, etc. This observation is what leads to our notion of theoretical extension defined below.

The main idea behind the notion of extension is that while a theory might be indifferent as to the truth-values of the sentences in various invalid sequents $X : Q$ whose invalidity is to be 'overcome', nevertheless the inferential gap between these Xs and Qs cannot be bridged using only the resources that are within the theory by that stage and that can be articulated within the language of the sequents concerned.[10] Certain (sets of) new sentences S will be needed to make sequents $S, X : Q$ valid, thereby overcoming the invalidity of $X : Q$. So it will follow not only that

(i) infinitely many sequents $X : Q$ have to be involved (as we shall see in Lemma 3 below); but also that

(ii) the extending sentence(s in) S will need to contain some 'new' term(s), not occurring in any of the sequents $X : Q$ (as we shall see in Lemma 2 below).

These extending sentences in S will feature in sequents $S, X : Q$ that are not only valid, but valid in a special way. It is the main burden of the investigations in this chapter to characterize the kind of validity involved here. We call it *constrained creative extension*; more on which below.

Both features (i) and (ii) are found in scientific theorizing. One introduces new theoretical terms in new hypotheses that are designed both to unify our understanding of disparate phenomena and to extend the range of predictions that can be made about phenomena yet to be observed. The unification and extension involves the postulation of inner or hidden mechanisms, or of micro-constituents, or of large-scale force fields, etc. These posits require new referential terms to be introduced in new hypotheses that 'extend', and thereby unify and widen our range of scientific explanations. Moreover these explanations concern types of phenomena whose possible instances are infinitely various. They can vary in respect to time and place; or in respect of the particular exemplars that might be chosen from any natural kind. It would run against the thrust for generality to limit those natural kinds so as to have only finitely many exemplars.

[10] Compare our quotation from Field above.

11.2.5 Verifiability and falsifiability

Since 'verifiable' derives from 'verify', which means establish as true, it is fair to say that Ayer's term 'criterion of verifiability' is a misnomer, even on the strength of its first part, which was concerned with direct verifiability. Recall that according to his definition a sentence S would be directly verifiable if there were observation sentences $O_1,\ldots, O_n$ that did not by themselves logically imply some observation sentence O, but would do so when combined with S as an extra premiss. Thus 'All swans are white' is directly verifiable on Ayer's account. For 'This is a swan' does not logically imply 'This is white'; but 'This is a swan', combined with the extra premiss 'All swans are white', does logically imply 'This is white'. Schematically:

$St \nvdash Wt$; but

$St, \forall x(Sx \supset Wx) \vdash Wt$

To say, however, that 'All swans are white' is directly verifiable seems to be an unhappy way of putting the matter. It would be better to say that the logical relations exhibited show rather that 'All swans are white' is directly *falsifiable*. For the observation of but a single non-white swan would establish it conclusively as false; while yet no number of observed swans, even be they all white, would establish it conclusively as true.

The term 'verifiability', then, in Ayer's own writings, should be taken to include falsifiability as part of its sense. As we tackle the problem, our interest in truth and falsity on the basis of observation could be made perfectly symmetric, by dealing with sequents of the form $X : Y$ (where X and Y are sets of sentences) and by having theories in the form $W|V$ (where one asserts all of W and denies all of V). We should be interested in identifying those sentences, very roughly, to the determination of whose truth-value the truth-values of observation sentences are relevant. The truth-value could be True, and it could equally be False. Positive answers are no more and no less informative than negative answers; observationally informed, however, both kinds of answer must be. Nevertheless, since such thoroughly symmetric metalogical treatments are unusual, we have decided, for ease of exposition, to limit ourselves to the more familiar, asymmetric approach. Thus we shall deal with sequents of the form $X : Q$ (where X is a set of sentences and Q is a sentence) and with theories as sets of assertions. Furthermore, all theoretical extensions will be effected 'on the left', by adding to the premisses. The result is an apparent pride of place for falsifiability of theoretical hypotheses. It should be borne in mind, however, that this asymmetry is only apparent, and a consequence only of our preference for a familiar mode of exposition.

Unlike Ayer's criterion, our criterion does not presuppose, in its formulation, any distinction between analytic and synthetic sentences. All it presupposes is the distinction between logically necessary (or impossible) and contingent sentences—surely congenial to a Quinean—and the distinction between so-called basic and non-basic primitive sentences. Although the interpretation of 'basic' is left open (so that the account can enjoy a certain generality across kinds of discourse), the motivating application is of course the one where the basic sentences are the observation sentences of the language. This too is surely congenial to the Quinean.

11.2.6 The principle of composition, or molecularity

Sentences consisting of only non-basic expressions will be cognitively significant if all their constituent expressions are significant.[11] These constituents will have been 'legitimated' as cognitively significant, and therefore be eligible to help form cognitively significant sentences. This is in keeping with the principle of compositionality, or molecularity. This 'legitimation' of new expressions, however, occurs relative to a theory as it is extended to include hypotheses involving those new expressions along with older, already legitimated expressions, in a legitimating way. The latter sentences that legitimate the new expressions are ones that have acquired their cognitive significance via the way they helped to forge new inferential links among statements already enjoying that status, and consisting of previously legitimated expressions.

11.2.7 Constrained extension: higher-level hypotheses

In the account given below, this forging of new inferential links whereby sentences involving new vocabulary become cognitively significant is called constrained creative extension of (families of) sequents consisting of sentences that are already cognitively significant. It is intended to explicate the manner in which cognitively significant sentences with 'new' expressions can owe their newly acquired cognitive significance to the way they function within a particular theory whose confirmation or disconfirmation derives, ultimately, from basic statements. Thus the whole account of how expressions and sentences acquire their cognitive significance is relativized to a (growing) theory throughout. Constrained creative extension is a refinement of the kind of

[11] The reader who asks whether one should strengthen this to 'if *and only if*' should consult the discussion below of Hempel's compositionality condition.

creative theoretical extension that Ayer appealed to in his ill-fated definition of indirect verifiability. Ayer was on to something important, but failed to shape it into the right metalogical form. He also stressed verifiability at the expense of falsifiability (even though Ayer's 'verifiability' was really a misnomer!—he should, as pointed out above, have spoken of falsifiability).

We start with the logical empiricist intuition that first found expression in Ayer's own account. This intuition is that cognitively significant sentences should effect extensions of arguments whose premisses and conclusions are already cognitively significant. This condition, however, is not enough—as the well-known collapses of Ayer's criterion make vivid. We therefore go one step further, by adding an extra constraint at the logical level so as to capture the logical empiricists' intuition more precisely. We require not just extension, but constrained creative extension, in the sense to be defined below. This extra condition embodies a plausible deepening of Ayer's motivating intuition. It also rules out the standard collapses of Ayer's revised criterion (for which, see section 5 below).

The original motivation behind the formulation of conditions for what Ayer called indirect verifiability (which, as we would also stress, was really indirect falsifiability) was to accommodate higher-level theoretical hypotheses. These were logically connected, within some theory, to observational and other directly verifiable (or directly falsifiable) sentences in a characteristic way. It is the way that we term extension ('on the left' for falsifiability). These high-level theoretical hypotheses made a difference to the class of theory-generated conclusions one could expect, given certain observational or otherwise significant sentences as premisses.

To take a simple example, consider various significant sentences $V_1, \ldots, V_n$ that do not, collectively, imply the significant sentence F. (Whence, by orthodox 'non-relevantist' logic, $V_1, \ldots, V_n$ are mutually consistent and F is not logically true.) But suppose that the theoretical hypothesis H is consistent with $V_1, \ldots, V_n$ and, when combined with $V_1, \ldots, V_n$ as premisses, yields F as a conclusion. (In other words, suppose $H, V_1, \ldots, V_n : F$ is perfectly valid.)[12] Then we have the prospect of bringing H into confrontation with observable facts, and having to appeal to the latter in falsifying H. For there will be some situation (world, model) making H and $V_1, \ldots, V_n$ all true, and the question will then be whether the actual world is such a world. Since $V_1, \ldots, V_n$ are significant, their truth in any world in which they are true will

[12] A sequent is perfectly valid just in case it is valid but has no valid proper sub-sequent. So with a perfectly valid sequent each of its sentences is needed for its validity.

rest on the observable (= basic) facts in that world. Since F is significant, it will, if false in a world, likewise be false by appeal perforce to some of the observable facts in that world. Thus H can in principle be tested by recourse to the observable facts. For $V_1, \ldots, V_n$ may well turn out to be true in the actual world, and F false; and this will be the case by virtue at least of some observable (= basic) facts. In such a case H will have been falsified, but only by making appeal to the observable facts. When we say '$V_1, \ldots, V_n$ may well turn out to be true in the actual world, and F false', we leave open the possibility that the epistemic situation may not be conclusive, especially with respect to $V_1, \ldots, V_n$. For, particularly in the first-order case, one may not have access to finitary truth makers for $V_1, \ldots, V_n$. Their truth-by-virtue-of-at-least-some-basic-facts may be beyond the reach of finitary proof. One may, however, entertain them confidently as lower-level, as-yet-unfalsified statements about the actual world, and test H modulo those lower-level statements in the way just described.

This was clearly the main idea behind Ayer's treatment, but he did not go far enough in explicating it logically. We shall impose this further condition on how an hypothesis like H makes its special contribution in allowing one to pass from $H, V_1, \ldots, V_n$ to F when one cannot pass from $V_1, \ldots, V_n$ alone to F:

> if H is helping to forge a logical connection between significant premisses and a significant conclusion that does not obtain between them on their own, it must be because of how the truth of H secures consequences at the level of basic facts in a way that the truth of $V_1, \ldots, V_n$ alone does not.

Now suppose that the actual world turns out (in a semantic, or logical sense, rather than necessarily an epistemic sense) to make $V_1, \ldots, V_n$ true but F false. Suppose, in other words, that the actual world @ is a counter-example to the inference $V_1, \ldots, V_n : F$. Then the way things are at the observable (= basic) level in @ must be involved in making H false. We explicate this extra requirement below by saying that

> every @-disproof of H must use at least some basic axioms.

That is, in every falsity-maker for H in @ at least some basic facts will be ultimately implicated. (The same holds for any other world in place of @. The notions of @-proof and @-disproof will be explained in due course.)

I have chosen to illustrate the leading idea behind our strengthening of Ayer's criterion by reference to a multiple premiss set and a single conclusion, with the hypothesis H reckoned to the premisses (on the left) for testing. But one could generalize our definitions to multiple conclusion sets and dualize to allow 'hypotheses' on the right. This would effect a thoroughgoing symmetry while remaining faithful to the leading idea. As remarked earlier, however, we prefer a conventional approach for ease and clarity of exposition.

As we shall see below, our newly and more amply characterized notion of cognitive significance appears to be robust and to answer also to intuitions about grades of theoreticity. Our account does justice to Hempel's view that cognitive significance depends both on linguistic framework and theoretical context; but it does so without going the way of Quinean holism. The new criterion belongs to that species of criterion, which we owe to Carnap, that exploits logical relations among sentences, as well as their compositional semantics.

Despite the emphasis on basic expressions in our discussion above of H, $V_1, \ldots, V_n$ and F, we do not have to insist in general that the cognitive significance of new terms be transmitted directly from the basic expressions themselves. Such insistence is appropriate only when $V_1, \ldots, V_n$ and F involve only basic expressions (hence are significant 'at level 0'). In general, however, when $V_1, \ldots, V_n$ and F (and their constituent expressions) are significant only by higher levels $k > 0$, the cognitive significance of new terms at level $(k + 1)$ could be transmitted from those expressions already legitimated as significant by level k. That is why we shall use the slightly more liberal notion of a *legitimate* axiom when describing the constraint on theory extenders. This is more faithful to actual scientific practice; for high-level theoretical terms are usually introduced in order to secure new inferential links among sentences involving only theoretical terms already legitimated. The latter sentences are not, in general, required also to contain basic terms; though of course they are not excluded from doing so.

11.2.8 A remark on significance via compounding

The simple example 'All swans are white' is instructive because it illustrates how a sentence might be thought to be cognitively significant for two sorts of reason: namely, by compounding and by theoretical extension. But in fact 'All swans are white' can be licensed as significant only by virtue of compounding. The provision for extension will anyway require, by Lemma 2 below, that the extender contain new vocabulary not yet legitimated within the language of the invalid sequents to be extended. 'All swans are white', however, is a logical compound of the basic atomic expressions ('... is a swan' and '... is white'), which are already legitimated at level 0. The sentence is therefore cognitively significant by compounding at level 1 (with respect to any theory—for they all legitimate its constituent terms—whether or not the theory contains the sentence itself).

Significance via compounding cannot, on the account we are proposing, be subsumed under significance via constrained extension. Constrained extension is required in order to legitimate new expressions in the sentences effecting such extension. Once those expressions have been legitimated, then any compound of them becomes significant. Hence, in particular, the sentences effecting the constrained extension are significant.

11.2.9 Hempel's compositionality condition

Our criterion meets Hempel's own necessary condition of adequacy on any proposed criterion of cognitive significance. Hempel's condition, however, is not quite correct in its original formulation. As the reader will recall, Hempel's condition was that the criterion should ensure that any sentence[13] is cognitively significant only if all its constituent expressions are cognitively significant. The converse of Hempel's condition certainly holds: any sentence is cognitively significant if all its constituent expressions are cognitively significant. But the original condition cannot hold, and cannot be shown to hold, until we have qualified it.

Hempel gave the following definition.

> Suppose S is a subsentence of S'. Then S occurs non-vacuously in S' if and only if there are distinct truth-value assignments, differing at most in what they assign to atoms in S, which assign different values to S and assign different values to S'.

Note that S can occur non-vacuously in S' only if S' is contingent. Hempel then stated his necessary condition of adequacy as follows:

> (A) If under a given criterion of cognitive significance, a sentence N is non-significant, then so must be all truth-functional compound sentences in which N occurs non-vacuously as a component. For if N cannot be significantly assigned a truth-value, then

[13]Here, we assume, the sentences are well-formed in such a way that licit combination of significant words will not produce gibberish. Thus, for example, we would have to say that Chomsky's example 'Green ideas sleep furiously', if regarded as gibberish, is ill-formed. (Perhaps it violates sortal restriction conditions in the underlying grammar of English.)

> it is impossible to assign truth-values to the compound sentences containing N [non-vacuously]; hence, they should be qualified as non-significant as well.

My insertion above of '[non-vacuously]' supplies a qualification that Hempel obviously intended. Otherwise he would be committed to saying, for example, that a logical truth like $\neg(N \wedge \neg N)$ cannot be assigned a truth-value. But this would be wrong, since the proof of $\neg(N \wedge \neg N)$ shows it to be true regardless of the truth-value of N. For N to occur non-vacuously in a compound S it must be the case that a difference in the truth-value of N could make a difference in the truth-value of S; that is, in order to work out the truth-value of S (by any means whatever) we would have to work out the truth-value of N.

These reflections show that Hempel is in error when drawing the following 'corollary' of his requirement (A):

> (A2) If under a given criterion of cognitive significance, a sentence N is non-significant, then so must be any conjunction $[N \wedge S]$ and any disjunction $N \vee S$, no matter whether S is significant under the given criterion or not.

This 'corollary' does not follow. Even if N is non-significant, if S is logically false, then so is $N \wedge S$; so $N \wedge S$ is cognitively significant. Likewise, even if N is non-significant, if S is logically true then so is $N \vee S$; so $N \vee S$ is cognitively significant.

It follows that Hempel's necessary condition of adequacy should be reformulated as follows:

> (A*) Any *contingent* sentence is cognitively significant only if all its constituent expressions are cognitively significant

Our compositionality theorem below secures this result.[14]

[14] Suppose that S is cognitively significant but that N is not. Then the compound sentence $S \wedge (S \vee N)$, it might be objected, should still count as significant, since N occurs vacuously within it. But (A*) would disallow $S \wedge (S \vee N)$ as significant, for contingent S.

This is not a telling objection. We do not see much being lost if $S \wedge (S \vee N)$ and its ilk are disallowed as significant. But for one who insists on their being significant, (A*) could be modified to

> (A**) Any contingent sentence is cognitively significant only if all its constituent expressions enjoying non-vacuous occurrences are cognitively significant.

Similar minor modifications in the definition below of significance via compounding would secure (A**) as a theorem.

Conventionally the logical empiricists counted all logical truths and logical falsehoods as cognitively significant. In this we follow them. For the truth-value of a logical truth or logical falsehood can in principle be determined 'am Symbol allein'. It is a matter of the form, not the content, of the sentence concerned. We therefore allow as cognitively significant even sentences containing some non-basic terms (indeed: even ones containing only non-basic terms), provided that they are logically true or logically false. Thus the sentence 'It is not the case that (the Absolute is perfect and it is not the case that the Absolute is perfect)' is cognitively significant. Because it has the form $\neg(P \wedge \neg P)$, we can determine it as true without concerning ourselves with what the sentence P itself means. Even if P has defective meaning, the defects do not hamper our determination of the truth-value of $\neg(P \wedge \neg P)$ on the basis of its logical form alone.

All that is important in the definition of cognitive significance is that it should exclude those *contingent* sentences whose defective meanings do obstruct our determination of their truth-values on the basis, ultimately, of the observable (or basic) facts and sentence meaning. It is precisely this task which our definition below is designed to accomplish.

11.2.10 The first-order case

Any workable account of cognitive significance must deal with the first-order case, and not be restricted merely to propositional languages. Historically, however, the collapses of proposed accounts have been effected at the propositional level. That might lead one to think that the level of propositional logic would be the most appropriate one for the initial exposition of the central ideas behind a new criterion. Even at the propositional level, it would seem, there is challenge enough to get the account right. We have resisted this temptation so to restrict the exposition of the main ideas behind the criterion we propose, for the following compelling reasons.

First, Lemma 4 below shows that after getting all contingent compounds of basic propositional atoms to be significant at level 1 with respect to any theory, we get no more significant sentences (in a propositional language) at all. For all $n > 1$, a sentence is significant at level n only if it is already significant at level 1. This should be no surprise. A finitary classical propositional language is hardly one in which to get a grip on regularities in the world that can manifest themselves in infinitely various ways. Whatever 'theoretical' sentence of a classical propositional language we may introduce into our theory (a sentence, that is, containing non-basic propositional atoms), the deductive links it effects among (compounds of) basic atomic sentences can be just as well effected by adopting as an hypothesis a compound of basic atomic sentences. But this is not so at first order!

Secondly, just as a criterion can fail through collapse, so too it can fail through belittling. A criterion of cognitive significance would be belittled if there were some obviously cognitively significant sentence (within some theory) that, according to the criterion, could not be seen to count as such. The danger, in restricting oneself to the propositional level for an exposition of a new criterion, is that critics of the new criterion will be entitled to conclude that it can be belittled. The serious theorist about scientific theories will want to work right away with examples of quantified sentences from the natural sciences, and (quite rightly) will be at a loss to see how they could be accommodated as cognitively significant by any proposed criterion that is limited to the propositional level.

11.2.11 The invariance of non-significance under reformulation

The cognitive non-significance of a theory should be invariant. 'Once nonsense, always nonsense' is what the logical empiricists wanted to be able to say about meaningless metaphysical discourse. It was important to Carnap to be able to show that bad theory could not insinuate or inveigle itself into a better status via any finagled mix of the good with the bad. A good, terminologically clean theory in the physical sciences, affording no respectability to the terms of Heideggerian metaphysics, cannot be reformulated in conjunction with Heideggerian metaphysics in such a way as to force one to concede cognitive legitimation for the terms of Heideggerian metaphysics. No matter how much the dodgy and caddish terms may be made to rub shoulders with the gentle and honest terms, they will remain, essentially, worlds apart. Holism provides no refuge for the shifty, footloose and phenomena-free language of the Heideggerian. We need to be able to show this. What we want is a theorem about linguistic class differences. Our Quarantine Theorem below improves on Carnap's result to this effect.[15] It shows that our criterion cannot be collapsed. That is, there can be no metatheorem to the effect that every sentence is (on the definition given) cognitively significant with

[15] Carnap's result was at p. 55 of 'The Methodological Character of Theoretical Concepts', in H. Feigl and M. Scriven (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minnesota Studies in the Philosophy of Science, Vol. I, University of Minnesota Press, 1956, pp. 38–76.

respect to any given theory.

11.3 The formal theory

11.3.1 Extension

Let X be a (possibly empty) set of sentences and let Q be a sentence. We include the absurdity sign $\perp$ as a special case of Q. $X : Q$ is a sequent. Instead of $X : \perp$ we can also write $X : \emptyset$ where $\emptyset$ is the empty set. This enables us to think of $X : \perp$ as a proper sub-sequent of $X : Q$ when $Q \neq \perp$. Also, if Y is a proper subset of X then $Y : Q$ is a proper sub-sequent of $X : Q$.

DEFINITIONS

A *countermodel to* $X : Q$ is a model that makes every sentence in X true, and Q false.

The sequent $X : Q$ is *valid* if and only if $X : Q$ has no countermodel.

$X : Q$ is *perfectly valid* if and only if it is valid and has no valid proper sub-sequent: that is, it ceases to be valid if any sentence is removed on the left or on the right.

$X : Q$ is *skeletally valid* if and only if it is perfectly valid and is not a proper substitution instance of a perfectly valid sequent: that is, not only does every sentence matter, but also every aspect of each sentence's logical structure.

We shall write $X \vdash Q$ when $X : Q$ is valid. Instead of $\emptyset \vdash Q$ we shall write simply $\vdash Q$. $\vdash Q$ means that Q is logically true. $X \vdash \perp$ means that X is not satisfiable. We shall write $X \nvdash Q$ when $X : Q$ is invalid (that is, has a countermodel). With this notation likewise we suppress mention of $\emptyset$ if it occurs on the left or on the right. $X \nvdash$ means that X is satisfiable; $\nvdash Q$ means that Q is falsifiable. '$X \nvdash Q$' can be made vivid as follows: the language can be partitioned (by some countermodel) so that on the left are all the true sentences, including all those in X, and on the right are all the false sentences, including Q.

Suppose $X : Q$ is non-empty, that is, either X is non-empty or Q is not absurdity. Suppose further that $P, X : Q$ is perfectly valid. Then the proper sub-sequent $X : Q$ is invalid, whence some model makes every sentence in X true, and Q false, and therefore makes P false (on pain of counter-exemplifying $P, X : Q$). Likewise, removing some member of X or replacing Q by absurdity leaves an invalid proper sub-sequent when P is on the left, whence there is a model that makes P true. Thus if $P, X : Q$ is perfectly valid, P can be made true and P can be made false; that is, P is contingent.

LEMMA 0: Suppose $X \nvdash Q$. Then for every sentence S, exactly one of the following holds:

(i) $\neg S, X \vdash Q$ and every countermodel to $X : Q$ makes $\neg S$ false
(ii) $S, X \vdash Q$ and every countermodel to $X : Q$ makes S false
(iii) $\neg S, X \nvdash Q$ and $S, X \nvdash Q$.

> *Proof*: Consider the set of countermodels to $X : Q$. Then exactly one of the following holds:
>
> (i′) every one of them makes S true
>
> (ii′) every one of them makes S false
>
> (iii′) some of them make S false and some of them make S true
>
> These are respectively equivalent to (i), (ii) and (iii) above. So we know also that if $X \nvdash Q$ then
>
> (iv) if $\neg S, X \vdash Q$ then every countermodel to $X : Q$ makes S true
>
> (v) if $S, X \vdash Q$ then every countermodel to $X : Q$ makes S false. QED

We now need some definitions that will enable us to talk about extensions of scientific theories. These arise from a given theory by adding to it new hypotheses—in particular, hypotheses involving new theoretical vocabulary. We need to provide in general for infinite theories, which may or may not be logically closed.

DEFINITIONS (regarding theories)

We shall treat *theories* as sets of sentences. A *model for* a theory makes all its sentences true. A *satisfiable* theory is one that has a model. We shall always assume our theories to be satisfiable.

W is *logically closed* if and only if W contains every sentence (in the language of W) that is true in every model for W. We do not assume logical closure for theories. (In this respect we do not follow the conventions observed by most mathematical logicians. But our own usage is arguably more in keeping with how philosophers of science actually talk about scientific theories—as sets of sentences not necessarily containing all their logical consequences.) $[W]$ will denote the logical closure of the theory W.

W is a *subtheory* of X if and only if $[W]$ is a subset of $[X]$. (Note that it might be the case that W is a subtheory of X even though it not be the case that W is a subset of X! This is because X might not be logically closed.) When W is a subtheory of X we shall also say that X is an *extension* of W.

DEFINITIONS (regarding sequents)

$X : Q \succeq Z : R$ if and only if every countermodel to $Z : R$ is a countermodel to $X : Q$.

$X : Q \otimes Z : R$ if and only if not both $X : Q \succeq Z : R$ and $Z : R \succeq X : Q$

$\succeq$ could be read 'is logically as strong as', or 'is as easy to counter-exemplify as'. Clearly $\succeq$ is reflexive and transitive. $\otimes$ could be read as 'is not logically equivalent to'.

We can relativize these notions to a given theory. Thus:

$X : Q \succeq Z : R$ modulo W if and only if every model of the theory W that is a countermodel to $Z : R$ is a countermodel to $X : Q$

$X : Q \otimes Z : R$ modulo W if and only if not both $X : Q \succeq Z : R$ modulo W and $Z : R \succeq X : Q$ modulo W

If a sequent $X : Q$ is thought of as the claim $(\wedge X \supset Q)$, it is easier to appreciate why '$\succeq$' can be read as 'is logically as strong as'. For if $X : Q \succeq Z : R$ modulo W, then $(\wedge X \supset Q)$ logically implies $(\wedge Z \supset R)$ modulo W. The relation $\succeq$ among sequents modulo a given theory is a partial ordering of those sequents.

Note that if $X : Q \succeq Z : R$ modulo W and W^* is an extension of W, then $X : Q \succeq Z : R$ modulo W^*.

$\otimes$ is the relation of logical non-equivalence (modulo W) by means of which we shall distinguish sequents.

Validity of sequents can also be relativized to a given theory:

$X : Q$ is *valid modulo* W if and only if no model of W is a countermodel to $X : Q$—that is, $X, W \vdash Q$. Hence

$X : Q$ is *invalid modulo* W if and only if some model of W is a countermodel to $X : Q$—that is, $X, W : Q$ has a countermodel.

Note that if $X : Q$ is valid modulo W and W^* is an extension of W, then $X : Q$ is valid modulo W^*. Ordinary validity is validity modulo the empty theory.

Note also that if a sequent $X : Q$ is invalid modulo W then so is any sub-sequent of $X : Q$.

DEFINITION
Let $\mathcal{F}$ be a set of sequents. W *clinches* (each member of) $\mathcal{F}$ if and only if every sequent in $\mathcal{F}$ is valid modulo W—that is, for every $X : Q$ in $\mathcal{F}$ we have $X, W \vdash Q$.

DEFINITION
A *family* for W is a non-empty set of sequents in the language of W that are invalid modulo W.

The sequents in a family for a theory W are to be thought of as those inferences (in the language L of one's theory W) that one might wish to be clinched by some theoretical extension of W.[16] In general these will include both those sequents expressing law-like regularities that we wish to be able to explain, and other sequents expressing regularities of which we might not yet be apprised. By introducing new theoretical vocabulary, and new theoretical hypotheses involving both the old and the new vocabulary, we generate the former sequents as explanations; whereas the latter sequents may be generated as new predictions by means of which the extended theory can be tested. We should therefore bear in mind that when we undertake theoretical extensions we may not be motivated at the time by the full range of invalid sequents (modulo the unextended theory) that will form the family

[16] The clinching, however, will have to be effected in a very special way; see below.

involved in the extension. Indeed, it is usually required of theoretical extensions motivated by the desire to explain a given range of regularities that it should also generate predictions concerning other regularities of which the theory-builder may not yet be aware. The sequents expressing these latter regularities can then be put to the test; hence also the extended theory that generates them.

DEFINITION
Suppose $\mathcal{F}$ is a family for W and S is a finite set of sentences. *S extends W for $\mathcal{F}$*

if and only if

S, W is satisfiable and for each of the sequents $X : Q$ in $\mathcal{F}$, there is some subset W_X of W and some (perforce non-empty) subset S^* of S such that the extended sequent $X, S^*, W_X : Q$ is skeletally valid.

When S extends W for $\mathcal{F}$ we shall call S an *extender of W for $\mathcal{F}$*. Note that an extender S is always finite. Note also that if S extends W for $\mathcal{F}$ then S is not a subtheory of W. For, if S were a subtheory of W, we would have each sequent in $\mathcal{F}$ valid modulo W, contrary to the requirement that $\mathcal{F}$ be a family for W.

LEMMA 1. If S extends W for $\mathcal{F}$, then every sequent in the family $\mathcal{F}$ is finite.

> *Proof*: Since for each such sequent $X : Q$, we have a skeletally valid sequent of the form $X, S^*, W_X : Q$, and validity is compact, X must be finite. QED

Even if W already clinches $\mathcal{F}$, some extender S might be theoretically desirable, on grounds of simplicity and unifying explanatory power. If W does not already clinch $\mathcal{F}$—and especially if $\mathcal{F}$ is a family for W—then an extender S is called for in order to clinch $\mathcal{F}$. Note that S might be a theory in the same language as W, even though it is not a subtheory of W. In that case, if S extends W for a family $\mathcal{F}$ then it does so in a rather uninteresting way. S would, in effect, merely state (whatever extra is needed, modulo W, to clinch) the regularities expressed by the sequents in $\mathcal{F}$, but would not need to resort to linguistic expressions other than those already used in (sequents in) $\mathcal{F}$. More interesting is the case where a finite extender

S extends W for a family $\mathcal{F}$, and does so by resorting to new vocabulary. And most interesting of all is the case where a finite extender S extends W for a family $\mathcal{F}$, and does so *by having to resort to new vocabulary.* In that case the theory W would have to have been $\mathcal{F}$-less in the following sense:

DEFINITION
Let $\mathcal{F}$ be a set of sequents in the language of W. The theory W *is $\mathcal{F}$-less* if and only if no satisfiable extension W, S obtained by adding only finitely many sentences (i.e. those in S) in the language of W, clinches $\mathcal{F}$

LEMMA 2. If S extends W for $\mathcal{F}$, and W is $\mathcal{F}$-less, then S must contain at least one new term not in the language of W.

Proof. Suppose that S extends W for $\mathcal{F}$, but that every term in S is in the language of W. Consider the extended theory $W^* = W, S$. *Ex hypothesi* W^* is satisfiable and is in the language of W. Moreover, for each sequent $X : Q$ in $\mathcal{F}$, since there is a skeletally valid sequent of the form $X, S^*, W_X : Q$, it follows by dilution that $X, W, S \vdash Q$, hence $X, W^* \vdash Q$, contrary to the assumption that W be $\mathcal{F}$-less. QED

LEMMA 3. Extension of an $\mathcal{F}$-less theory for a family $\mathcal{F}$ requires $\mathcal{F}$ to be infinite.

Proof. Suppose that S extends W for $\mathcal{F}$, and that W is $\mathcal{F}$-less. Suppose there are only finitely many sequents in $\mathcal{F}$. Call them $X_i : Q_i$ $(i = 1, \ldots, n)$. Note that W, S is satisfiable, and for each sequent $X_i : Q_i$ $(i = 1, \ldots, n)$ there is a (finite) subset W_i of W and a (non-empty, finite) subset S^* of S such that $X_i, S^*, W_i : Q_i$ is (skeletally) valid. Hence every model of S fails to be a countermodel to $X_i, W_i : Q_i$. Thus every model of W that is also a model of S fails to be a countermodel to $X_i : Q_i$ $(i = 1, \ldots, n)$. Hence every model of W that is also a model of S fails to make $(\wedge X_i W_i) \supset Q_i$ false; hence, makes it true.[17] Consider now the sentence σ:

[17]The acute reader will have noticed that this step is non-constructive, like certain others in what is here perforce a more classical exposition of ideas than we would prefer. The reader may rest assured, however, that we are taking this more relaxed, 'classical semantical'-looking route only for ease of exposition. We do not want to compound difficulties of comprehension that might arise from a reader's unfamiliarity with constructivism with difficulties of comprehension arising from this new attempt to characterize cognitive significance. The main ideas that we are capturing here in this more classical vein—ideas

(σ) $(\wedge X_1 W_1 \supset Q_1) \wedge \ldots \wedge (\wedge X_n W_n \supset Q_n)$

Note that every model of W that is also a model of S makes σ true. Hence, since W, S is satisfiable, it follows that W, σ is satisfiable also. σ is in the language of W. Each sequent $X_i, W_i, \sigma : Q_i$, i.e.

$$X_i, W_i, (\wedge X_1 W_1 \supset Q_1) \wedge \ldots \wedge (\wedge X_n W_n \supset Q_n) : Q_i \ (i = 1, \ldots, n)$$

is valid. By dilution

$$X_i, W, \sigma \vdash Q_i$$

But consider the extended theory $W^* = W, \sigma$. This, as we have seen, is satisfiable. Moreover we now have, for each sequent $X_i : Q_i$ $(i = 1, \ldots, n)$ in $\mathcal{F}$, that

$$X_i, W^* \vdash Q_i;$$

which is contrary to the requirement that W be $\mathcal{F}$-less. QED

LEMMA 4. For extension of an $\mathcal{F}$-less theory for a family $\mathcal{F}$ to be possible, the language used, if classical, has to be more powerful than a propositional language that has only finitely many propositional constants.

> *Proof*: Suppose we are dealing with a propositional language L that has only finitely many propositional constants. Suppose S extends the $\mathcal{F}$-less theory W for the family $\mathcal{F}$. By Lemma 3, $\mathcal{F}$ is infinite. Suppose its pairwise logically non-equivalent members modulo W are $X_i : Q_i$ $(i = 1, 2, \ldots)$. For each such sequent $X_i : Q_i$ in $\mathcal{F}$ there is a (finite) subset W_i of W and a (non-empty, finite) subset S^* of S such that $X_i, S^*, W_i : Q_i$ is skeletally valid (hence also finite). So the conjunction of all the members of S logically implies (modulo W) each of the infinitely many sentences $(\wedge X_i) \supset Q_i$ $(i = 1, 2, \ldots)$ which are pairwise logically non-equivalent modulo W. But this is impossible within a

about constrained creative extension, and of subformula occurrences having to 'do some work' within a proof—can be captured also within the more austere context of *IR*.

classical propositional language. QED

It is when S extends W for $\mathcal{F}$, and W is $\mathcal{F}$-less (hence $\mathcal{F}$ is an infinite family for W) that the extension deserves to be called *creative*. The *finite* extender S helps to clinch the *infinitely many* regularities (sequents) in $\mathcal{F}$ that *no* satisfiable extension of W, within the language of W, can clinch. And the finite extender S does this only by virtue of involving *new* theoretical vocabulary, and theoretical assertions involving that new vocabulary.

What we need to ensure, now, is that every member of S is involved in the clinching of at least infinitely many of the sequents in the family $\mathcal{F}$. We want to prevent the family $\mathcal{F}$ from containing finitely many 'rogue' sequents which alone are responsible for S's containing some of the sentences that it does. Every member of S has to be rendering an infinite amount of extending service, so to speak. In effect, therefore, in creative extension, there is a certain sort of *homogeneity* to the problematic sequents in $\mathcal{F}$, and a homogeneity of purpose in S's having all the (finitely many) member sentences that it does, in order to clinch these sequents.

Let us call a subset $\mathcal{F}^*$ of $\mathcal{F}$ a *cofinite* subset of $\mathcal{F}$ just in case it contains all but finitely many members of $\mathcal{F}$ (so $(\mathcal{F} \setminus \mathcal{F}^*)$ is finite). Note that if W is $\mathcal{F}$-less and $\mathcal{F}^*$ is a cofinite subset of $\mathcal{F}$ then W is $\mathcal{F}^*$-less. This is because, if some satisfiable extension W, S^* obtained by adding only finitely many sentences (i.e. those in S^*) in the language of W, were to clinch $\mathcal{F}^*$, then the adoption of only finitely many more sentences in the language of W would suffice to clinch in addition the finitely many remaining sentences in $(\mathcal{F} \setminus \mathcal{F}^*)$. For each of the finitely many sequents $X : Q$ in $(\mathcal{F} \setminus \mathcal{F}^*)$, the sentence $(\wedge X) \supset Q$ would do.

DEFINITION

S creatively extends W for $\mathcal{F}$

if and only if

(i) S extends W for $\mathcal{F}$, and W is $\mathcal{F}$-less; and, moreover,

(ii) no proper subset of S extends W for any cofinite subset of $\mathcal{F}$.

Since $\mathcal{F}$ is a cofinite subset of itself, it follows from (ii) that if S creatively extends W for $\mathcal{F}$ then no proper subset of S extends W for $\mathcal{F}$. So each member of a creative extender S is needed just for the plain extending that S accomplishes. Indeed, (ii) secures more: each member of S is needed for an infinite amount of the plain extending that S accomplishes.

This notion of creative extension can be illustrated with a simple mathematical example.[18] Let E_n be a sentence of first-order logic with identity that says 'There are at least n individuals'. Consider the family of sequents $\{\emptyset : E_n | n > 2\}$. This is a family for the empty theory in the language of identity, since E_n is logically contingent for $n > 2$. Moreover, a corollary of the compactness theorem for first-order logic tells us that no finite satisfiable theory (hence, no finite satisfiable extension of the empty theory) in the language of identity can clinch this family. So the empty theory is $\{\emptyset : E_n | n > 2\}$-less. Now consider a finite 'theory of infinity' S—that is, a finite satisfiable theory S that has only infinite models. S could extend the empty theory for $\{\emptyset : E_n | n > 2\}$, since for each E_n there could be a subset S^* of S such that each sequent $S^* : E_n$ is skeletally valid. Examples of such finite theories of infinity are

{0 is not a successor, Everything has a successor, No two successors are identical}

and

'R is a dense linear ordering of at least two elements', i.e.
$\{ \exists x \exists y Rxy, \forall x \forall y (Rxy \supset \neg Ryx), \forall x \forall y (Rxy \supset \forall z (Ryz \supset Rxz)),$
$\forall x \forall y (Rxy \supset \exists z (Rxz \wedge Rzy)), \forall x \forall y (y = x \vee Rxy \vee Ryx) \}$

The first of these involves the successor function symbol $s(\ )$; the second involves the two-place predicate R.

This example nicely confirms our general expectations about creative extension. We have seen (Lemma 2) that any extender has to involve at least one new term not in the language of the unextended theory (such as the successor function symbol s or the two-place predicate R). We have also seen (Lemma 3) that extension of an $\mathcal{F}$-less theory for a family $\mathcal{F}$ requires $\mathcal{F}$ to be infinite (such as $\{\emptyset : E_n | n > 2\}$).

11.3.2 How sentences depend on the atomic facts within a model for their truth or falsity

Partition the primitive extra-logical expressions of L into two classes. Call one the class of *basic* primitive expressions. Call the other the class of *non-*

[18] I am indebted here to Randy Dougherty.

basic primitive expressions. We are concerned to define a sense in which the truth-value of a given sentence rests on the basic facts rather than on the non-basic facts. (Note that 'non-basic' does not imply 'non-atomic'.)

Let M be a model that deals with both basic and non-basic primitive expressions, and suppose from now on that it deals with all the primitive expressions occurring in any sentence we mention. The identity predicate, if it occurs in the language, counts as basic. A *basic atomic sentence* is one made up entirely of basic primitive expressions. The adjectives 'basic' and 'non-basic' will transmit in the obvious way to assumptions, rule assumptions and axioms.

Consider a propositional language, based on $\neg$, $\vee$, $\wedge$ and $\supset$. Consider any truth-value assignment M dealing with the propositional variables. To any propositional variable, M assigns the value T or the value F. Let the *atomic diagram* corresponding to M be the set that results by taking each propositional variable that is assigned T by M, and taking the negation of each propositional variable that is assigned F by M. Let us call the atomic diagram obtained in this way $[M]$. Thus if $\{A, B\}$ is the set of propositional variables, and if M is the assignment

$$A \rightarrow T$$
$$B \rightarrow F$$

then the atomic diagram $[M]$ is the set $\{A, \neg B\}$. We can go one step further and form the inferential diagram $|M|$ corresponding to M by replacing each unnegated atom A in $[M]$ by the atomic axiom

$$\frac{}{A}$$

and replacing each negated atom $\neg A$ in $[M]$ by the atomic rule of inference

$$\frac{A}{\bot}$$

Thus in our previous example the inferential diagram $|M|$ would be the set

$$\{\ \frac{}{A}\ ,\ \frac{B}{\bot}\ \}$$

of atomic rules. For ease of layout we shall also write $\{|A\ ,\ B|\bot\}$. The inferential diagram represents 'the way the world is' according to M. Given a truth-value assignment M, one uses truth tables in the familiar way to work out the truth-value of any sentence formed from atoms to which M assigns a truth-value. Now corresponding to any such evaluation of any sentence S as true under M, we can provide a proof of S from (i.e. by using axioms and rules in) $|M|$ (called an *M-proof* of S); and for any sentence S false under M we can provide a proof of $\perp$ from $|M|, S$ (which we may call an *M-disproof* of S).

Each M-proof (and M-disproof) is built up by means of logical rules and also, possibly, the atomic rules in the inferential diagram $|M|$ described above. The logical rules consist of the familiar introduction and elimination rules for each logical connective.

KALMÁR'S THEOREM[19] (inferential version)

For any sentence S and any truth-value assignment M dealing with the atoms of S,
(i) if S is true under M then there is a proof of S whose only undischarged assumptions are rule assumptions drawn from $|M|$; and
(ii) if S is false under M then there is a proof of $\perp$ whose only undischarged sentential assumption is S and whose undischarged rule assumptions are drawn from $|M|$.

The proof is by induction on S, and is easy enough to leave to the reader.

We now have a precise sense for the claim that a sentence's truth-value is based on certain atomic facts according to a given evaluation. Each way of 'working out' that a sentence is true under a given truth-value assignment corresponds to a distinct proof of that sentence using rules in the inferential diagram corresponding to the assignment. To take a very simple example, if M makes both A and B true, then there will be two ways of showing that the sentence $(A \vee B)$ is true:

$$\dfrac{\dfrac{}{A}}{A \vee B}$$

[19]The reader will recognize this as the core of a well-known completeness proof for classical propositional logic. Cf. L. Kalmár, 'Über die Axiomatisierbarkeit des Aussagenkalküls', *Acta Sci. Math.* (Szeged) 7, 1934–5, pp. 222–43.

$$\frac{\overline{B}}{A \vee B}$$

Likewise for disproofs of false sentences.

'Evaluation' proofs with respect to a truth-value assignment M have one of the two forms

$$\begin{array}{l} \Sigma\ (\subseteq |M|) \\ \vdots \\ S \end{array}$$

$$\begin{array}{c} \underbrace{S, \Sigma}\ (\subseteq |M|) \\ \vdots \\ \bot \end{array}$$

where Σ, the set of undischarged rule assumptions, is drawn from the inferential diagram $|M|$. Proofs of the first form are M-proofs, and those of the second form (ending with $\bot$) are M-disproofs. For brevity of subsequent definitions, we shall say that Σ is the set of axioms of a proof of either form. In the case of proofs of the first form, Σ summarizes the atomic 'facts' (according to M) on which the *truth* of S rests. In the case of proofs of the second form, Σ summarizes the atomic 'facts' (according to M) on which the *falsity* of S rests.

Note that the truth or falsity of S rests on Σ according, respectively, to the proof or disproof in question. To stress this point, consider a complex sentence S true under M. There will in general be more than one M-proof of the form

$$\begin{array}{l} \Sigma\ (\subseteq |M|) \\ \vdots \\ S \end{array}$$

and in general the various Σ could be distinct. Such cases represent over-determination of truth, as we saw in the simple case of $(A \vee B)$ when both A and B are true. Likewise with falsity.

So there are in general different ways of seeing that a given sentence is true in M (if it is true) and different ways of seeing that a sentence is false in M (if it is false). The existence of exactly one way of arriving at a truth-value for a given sentence will be the exception rather than the rule.

Thus different selections of 'atomic facts' in M can be responsible—that is, sufficient—for the determination of a sentence's truth-value in M.

Thus far we have dealt only with the propositional case, where M is a truth-value assignment to propositional variables. We have spoken of M-proofs and M-disproofs. It is time now to generalize to the first-order case.

Here instead of a truth-value assignment M we shall have a *model* M consisting of a domain of individuals and with extensions as usual for primitive extralogical terms such as predicates. We can still talk of M-proof and M-disproof. We just have to bear in mind that these proofs will have (at the quantifier steps ∃E and ∀I) a distinct subproof for each individual in the domain of M, in a way to be explained presently. Just as an M-disproof Π of the conjunction $A \wedge B$ might proceed:

$$\begin{array}{c} \dfrac{A \wedge B}{A}{\scriptstyle \wedge E} \\ \Theta \\ \bot \end{array}$$

where Θ is an M-disproof of A, so too an M-disproof Π of $\forall x F(x)$ could proceed:

$$\begin{array}{c} \dfrac{\forall x F(x)}{F(t)}{\scriptstyle \forall E} \\ \Theta \\ \bot \end{array}$$

where Θ is an M-disproof of $F(t)$, and t is (the name for) a member of the domain of M. Matters would be correspondingly simple for M-proofs of existential sentences of the form $\exists x F(x)$.

But what about an M-proof of $\forall x F(x)$? And an M-disproof of $\exists x F(x)$? An M-proof of $\forall x F(x)$ is going to be as wide, at its final step, as M is large. That is, it will have as many branches (subordinate M-proofs) at the point of 'universal introduction' as there are members of the domain of M. Indeed, for each member a of the domain of M there will have to be an M-proof Θ_a of $F(a)$ as a subordinate proof for 'universal introduction' in the final step of an evaluation-as-true-in-M of the universal sentence $\forall x F(x)$:

$$\frac{\begin{array}{c}\ldots\ \ \Theta_a\ \ \ldots\\ \ldots F(a)\ldots\end{array}}{\forall x F(x)}$$

where the final step is 'M-relative universal introduction'.

Likewise an M-disproof of $\exists x F(x)$ is going to be as wide as M is large. That is, it will have at least as many branches (subordinate M-disproofs) at the point of 'existential elimination' as there are members of the domain of M. For each member a of the domain of M there will have to be an M-disproof Θ_a of $F(a)$ as a subordinate proof for existential elimination:

$$\frac{\exists x F(x) \qquad \begin{array}{c}\ldots \overline{F(a)}^{\,(i)}\ldots\\ \ldots\ \ \Theta_a\ \ \ldots\\ \bot\end{array}}{\bot}\,(i)$$

where the final step is 'M-relative existential elimination'.

If the domain of M is finite, then these M-relative proofs and disproofs will of course be finitary objects. But if the domain of M is infinite, these proofs and disproofs will in general be infinitary. Nevertheless, they will be well-defined mathematical objects. (Compare Carnap's use of the ω-rule in arithmetic to ensure that every arithmetical sentence came out as L-determinate.)

It is clear then that we have a notion of M-relative proof and disproof for any model M. Intuitively, such proofs correspond to evaluations of their conclusions as true-in-M; and such disproofs correspond to evaluations of a sentence as false-in-M. The assumptions in the former case, and the side assumptions in the latter case, must come from the 'atomic diagram' of the model M. That is, they will be singular (atomic) sentences or negations of such sentences. (We shall allow individuals in M to occur in such sentences as their own names, if necessary.)

We are now in a position to formulate the notion of constrained creative extension. Let Λ be a set of terms. A Λ-*axiom* is an atomic axiom that involves at least one Λ-term.

Definition

S effects a Λ-constrained extension of the theory W for the family $\mathcal{F}$
if and only if

S creatively extends W for $\mathcal{F}$

and (the Λ-constraint:)

for every sequent $X : Q$ in $\mathcal{F}$, for every countermodel M to $X, W : Q$ every M-disproof of any M-false member of S uses at least one Λ-axiom in the vocabulary of W.

When S effects a Λ-constrained extension we shall also speak of 'constrained extension via Λ' for grammatical ease in contexts below.

11.3.3 Some inductive definitions

We are now in a position to define, inductively, a relation that holds between sentences and theories, at finite levels, and a relation that holds between expressions and theories, at finite levels. The relation involving sentences is:

'S is basically falsifiable at level n with respect to theory Θ'.

The relation involving expressions is:

'E is legitimated at level n with respect to theory Θ'.

The set of all sentences that are basically falsifiable at level n with respect to theory Θ will be called 'F_n w.r.t. Θ'. We stipulate that if a sentence is basically falsifiable at a given level w.r.t. a theory, then it is basically falsifiable at that level w.r.t. any extension of that theory. We are also taking the levels to be cumulative. That is, once a sentence finds its way into F_n, with respect to a given theory, then it is automatically in F_{n+1} with respect to any extension of that theory.

The set of all expressions that are legitimated at level n with respect to theory Θ will be called '$\Lambda_n[\Theta]$'. We stipulate that if an expression is legitimated at a given level w.r.t. a theory, then it is legitimated at that level w.r.t. any extension of that theory. We are also taking the levels to be cumulative. That is, once an expression finds its way into Λ_n with respect to a given theory, then it is automatically in Λ_{n+1} with respect to any extension of that theory.

The set of cognitively significant sentences at level n w.r.t. Θ contains:

1. all logical truths;
2. all logical falsehoods; and
3. contingent sentences that are compounds of expressions in $\Lambda_n[\Theta]$.

We stipulate that if a sentence is cognitively significant at a given level w.r.t. a theory, then it is cognitively significant at that level w.r.t. any extension of that theory. And once a sentence is cognitively significant at level n w.r.t. a given theory, then it is cognitively significant at level $n+1$ w.r.t. any extension of that theory.

Basis step
Every basic sentence is basically falsifiable at level 0 w.r.t. the empty theory.
Every basic expression is legitimated at level 0 w.r.t. the empty theory.

Corollary of basis step:
All sentences compounded from basic expressions are cognitively significant at any level w.r.t. any theory.

Recall that a basic axiom was of the form 'infer A' or 'from A infer absurdity', where A was a basic atom (in the propositional case). A more liberal notion that can now be put in its place is that of a *legitimated* axiom. A legitimated axiom is one *all* of whose vocabulary has been legitimated within the theory with respect to which a constrained creative extension by S is to be made, as described by the following inductive subclause. Note how this subclause also keeps track of the set of terms that have been legitimated with respect to the growing theory. The subclause applies to (finite sets of) sentences S that contain arbitrarily but finitely many non-basic expressions not yet legitimated by the theory W at level n. These 'new' terms will be legitimated at level $n+1$, with respect to the theory that results by adopting S.

Inductive subclause 1 (for compounds containing expressions that are not yet legitimated, and for those expressions)

Suppose that S effects a constrained creative extension of W for the family $\mathcal{F}$ via $\Lambda_n[W]$ and that each sequent $X : Q$ in $\mathcal{F}$ involves only sentences that are significant at level n w.r.t. W. Then:

> S is included in F_{n+1} w.r.t. the extended theory W, S;
> and
> the result of adding the new terms of S to the already legitimated vocabulary $\Lambda_n[W]$ is the extended legitimated vocabulary $\Lambda_{n+1}[W, S]$.

Note here how the sentences in S get added to the theory in order to be in F_{n+1} with respect to the theory resulting from that addition. This respects the intuition that it is only within the context of the theory to which it belongs that a theoretical hypothesis (hence also: each of its terms) acquires its cognitive significance. As remarked earlier, these 'new' terms in S will be legitimated at level $n+1$, with respect to the extended theory that results by adopting S.

Inductive subclause 2 (for compounds of already significant expressions)

Suppose E_i is in $\Lambda_n[W_i]$ $(i = 1, \ldots, k)$. Then any contingent compound of $E_1, \ldots, E_k$ is significant at level $n+1$ w.r.t. the 'union of theories' $W_1, \ldots, W_k$.

There are the usual closure clauses for these inductive definitions.

11.3.4 Main results

COMPOSITIONALITY THEOREM

For any contingent sentence P, for any theory Θ,
for some n, P is significant at level n w.r.t. Θ
if and only if
there is some m such that each constituent
expression of P is in $\Lambda_m[\Theta]$

> *Proof*: The 'if' part follows from inductive subclause 2 above, which provides for the compounding of legitimate expressions. For the 'only if' part, note that a sentence can come to be cognitively significant at level n w.r.t. Θ only by being a compound of expressions in $\Lambda_n[\Theta]$. QED

PERSISTENCE THEOREM

If P is in F_n w.r.t. Θ then P is cognitively significant at level $n+1$ w.r.t. any extension of Θ.

> *Proof*: To get into F_n w.r.t. Θ, P (if not a compound of basic expressions) had to be in some finite S that effected a constrained creative extension of some subtheory of Θ. All the non-logical

> expressions of P thereby got into $\Lambda_n[\Theta]$. Thus P would be cognitively significant at level $n+1$ w.r.t. any extension of Θ. QED

The following observation may be thought of as a principle of relevance. For the propositional case it is obvious. In the first-order case it is a version of Robinson's consistency theorem:[20]

> If the sequents $W_1 : V_1$ and $W_2 : V_2$ have no terms in common, then $W_1, W_2 \vdash V_1, V_2$ only if either $W_1 \vdash V_1$ or $W_2 \vdash V_2$.

It is plain that successive applications of the inductive clause above involving extenders generates a sequence of sets of legitimated terms with respect to the respective satisfiable theories thereby generated. These theories will form a chain by inclusion. Each of the sets of terms in the sequence contains the new terms of the extender at that point in the sequence.

QUARANTINE THEOREM

Let Θ_1 be a satisfiable theory legitimating all its terms. Let Θ_2 be a satisfiable theory all of whose terms are non-legitimated w.r.t. Θ_1. Let Θ_0 be any theory logically equivalent to Θ_1, Θ_2, assuming the latter to be satisfiable. Then no sequence of sets of legitimated terms w.r.t. any chain of subtheories of Θ_0 can ever reach a point at which, in order to achieve the next extension accomplished 'within Θ_0', it has no alternative but to involve a set of terms containing a term of Θ_2.

> *Proof*: Θ_1 and Θ_2 have no terms in common. Suppose for *reductio* that there is a sequence σ, say, of sets of legitimated terms w.r.t. respective members of a chain of subtheories of Θ_0, that reaches a point at which, in order to achieve the next extension accomplished 'within Θ_0', it has no alternative but to involve a set containing a term of Θ_2, and a constrained creative extension using that term. Suppose that $\{E_1, \ldots, E_k\}$ is the first set in σ containing a term of Θ_2. σ reaches $\{E_1, \ldots, E_k\}$ via the constrained creative extender S, say, involving $E_1, \ldots, E_k$ as 'new' terms. S is a (finite) subtheory of Θ_0. Let Δ_0 be the corresponding subtheory, in the chain, with respect to which S effects the constrained creative extension for the family $\mathcal{F}$, say. So we are supposing S extends Δ_0 in this way for $\mathcal{F}$.

[20] See *Natural Logic*, p. 120.

We shall now construct an alternative extender to S, not involving any terms of Θ_2, but still extending Δ_0 for the family $\mathcal{F}$. This will establish our result.

Δ_0 is a subtheory of Θ_0. All the expressions in S, apart from $E_1,\ldots,E_k$, are legitimate w.r.t. Δ_0—hence, by our assumption concerning $\{E_1,\ldots,E_k\}$ within σ, w.r.t. Θ_1. We shall define *cleanliness* of X to be the following property:

> X contains only such terms as precede $\{E_1,\ldots,E_k\}$ in the sequence σ, and so are legitimate w.r.t. Δ_0—hence, by our assumption concerning $\{E_1,\ldots,E_k\}$ within σ, w.r.t. Θ_1

Cleanliness of a sentence or sequent, or of a set of sentences or sequents, will be indicated from now on by prefixing it with the superscript $^\circ$. We are assuming that S creatively extends Δ_0 for the family $\mathcal{F}$ (which we now call $^\circ\mathcal{F}$) and hence that

> (α) Δ_0 is $^\circ\mathcal{F}$-less.

It is clear from inductive subclause 1 above that the sequents in $^\circ\mathcal{F}$ contain only such terms as precede $\{E_1,\ldots,E_k\}$ in the sequence σ, and so are legitimate w.r.t. Θ_1. For every sequent $^\circ X : {}^\circ Q$ in $^\circ\mathcal{F}$ we have that for some subset Δ_{0X} of Δ_0 and for some non-empty subset S^* of S, the sequent $^\circ X, S^*, \Delta_{0X} : {}^\circ Q$ is skeletally valid. We have also that:

> (α) for no satisfiable extension $\Delta_0{}^*$ of Δ_0, obtained by adding only finitely many sentences in the language of Δ_0, do we have for each sequent $^\circ X : {}^\circ Q$ in $^\circ\mathcal{F}$ that $^\circ X, \Delta_0{}^* \vdash {}^\circ Q$

Since S is a subtheory of Θ_0, and Θ_0 is logically equivalent to Θ_1, Θ_2, it follows that there are finite sets Z_1, Z_2 of sentences drawn from Θ_1, Θ_2 respectively, such that

> $Z_1, Z_2 : \wedge S$ is perfectly valid $\ldots(1)$

where $\wedge S$ is the conjunction of all the members of S. This conjunction can be formed since S is finite. Remember that for each sequent $^\circ X : {}^\circ Q$ in $^\circ\mathcal{F}$, there is some subset Δ_{0X} of Δ_0 and some subset S^* of S such that the extended sequent

$^\circ X, S^*, \Delta_{0X} : {}^\circ Q$ is skeletally valid;

whence

$^\circ X, \wedge S, \Delta_{0X} : {}^\circ Q$ is valid . . . (2)

whence by cut on $\wedge S$ in (1) and (2) we have

$^\circ X, \Delta_{0X}, Z_1, Z_2 \vdash {}^\circ Q$

whence, since $^\circ X, \Delta_{0X}, Z_1 : {}^\circ Q$ and $Z_2 : \emptyset$ are vocabulary disjoint,

either $^\circ X, \Delta_{0X}, Z_1 \vdash {}^\circ Q$
or $Z_2 \vdash \emptyset$.

But $Z_2 \vdash \emptyset$ is contrary to (1).

So we have

for each sequent $^\circ X : {}^\circ Q$ in $^\circ\mathcal{F}$, there is some subset Δ_{0X} of Δ_0 such that $^\circ X, \Delta_{0X}, Z_1 \vdash {}^\circ Q$

hence

for each sequent $^\circ X : {}^\circ Q$ in $^\circ\mathcal{F}$, $^\circ X, \Delta_0, Z_1 \vdash {}^\circ Q$

Note that Z_1 is finite. But now, if Z_1 is clean, i.e. in the language of Δ_0, this will be contrary to (α), and we will have shown that the involvement of a term from Θ_2 was impossible. On the other hand, if Z_1 is not clean, it will at least not involve any term from Θ_2. And then the theory Z_1 will be a finite entender of the theory Δ_0 for the family $^\circ\mathcal{F}$. Since Z_1 is a subtheory of Θ_1 it is an alternative to S as the extender at that stage within Θ_0. One is not obliged to go the way of S into the territory of Θ_2-terms in order to forge the inferential connections one wants within Θ_0. Within Θ_0, no matter how it may be formulated, one may eschew all Θ_2-terms as illegitimate. QED

11.4 Comparison with Carnap's account

The best-known attempts to formulate a criterion of cognitive significance were all made in the shadow of the Carnapian doctrine, in *The Logical Syntax of Language*, according to which all interesting notions in the philosophy of science could be explicated by using only the notion of logical consequence—and a syntactically characterized one at that. It was only at the end of the heyday period of theorizing about cognitive significance that Carnap produced his most liberal formulation, in 'The Methodological Character of Theoretical Concepts'. Carnap's account is not without its problems, as we have seen; but it appears also to have been neglected by theorists who might have been able to improve upon it. In the exchange with Hempel in the Schilpp volume,[21] Carnap's proposal received scant attention. Hempel did not pursue its novel merits, and Carnap did not emphasize them as perhaps he should. The debate seemed to have lost steam, or become unfashionable. Admittedly, Carnap had not made his reader's work easy in his final foray on the topic.

Our new criterion is a much-mutated descendant of Carnap's. We have gone beyond Carnap, however, by incorporating extra safeguards against collapse (via the notion of constrained creative extensions), as well as some important liberalizations and refinements in the interests of faithful explication of scientists' theoretical practice. Despite these liberalizations, it is still possible to prove the theorem above, which we call the Quarantine Theorem. This theorem is crucial for showing that terms and sentences ruled non-significant by the criterion cannot defy that ruling by appearing in reformulations of the union of a significant theory with a non-significant one, and thereby forcing us to recognize them as legitimate.

It would be useful in conclusion to have a summary of points of contrast and points of affinity between Carnap's much neglected 1956 account and that put forward above. The following are the important similarities:

1. We both parametrize with respect to a theory.

2. We both subscribe to a compositionality principle: Carnap's D3 has the same effect as our inductive subclause 2.

The following are the important points of contrast between Carnap's approach and ours:

1. Carnap parametrizes with respect to just one theory, taken as fixed for the purposes of defining cognitive significance for terms and for

[21] C. G. Hempel, 'Implications of Carnap's work for the philosophy of science', in P. A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*, Library of Living Philosophers, Vol. XI, Open Court, La Salle, 1963, pp. 685–709; R. Carnap, 'Carl G. Hempel on Scientific Theories', in P. A. Schilpp (ed.), *op. cit.*, pp. 958–66.

sentences; whereas we build up the theory in the process of inductive definition.

2. Carnap deals with only finitely axiomatized theories; whereas we deal with arbitrary ones. (We both, however, countenance only finite extenders.)

3. Carnap has only the following features of perfect validity: the premisses are consistent (and one of them is the whole theory); the implication fails if the new sentence (with the new term) is missing. Hence the conclusion is contingent. We insist by contrast on skeletal validity. Thus our account too envisages only some, not necessarily all, the sentences of the theory featuring as premisses of the extended sequents. The insistence on skeletal validity also blocks both Kaplan's 'definitional extension' problem and his 'deoccamization' problem, which arise for any account based on perfect, but not skeletal, validity.[22]

4. Carnap distinguishes the correspondence rules from the theoretical postulates. For our account the distinction is unnecessary.

5. Carnap has the basic sentences only on the right of extended sequents, appealing to the deduction theorem to justify this. We have basic sentences both on the left and on the right, which is perhaps a good idea in case the deduction theorem fails for the logic concerned.

6. Carnap requires basic sentences to be in the invalid sequent being extended; whereas we require only sentences whose terms have already been legitimated, even if these terms are not basic.

[22] See D. Kaplan, 'Significance and Analyticity: A Comment on Some Recent Proposals of Carnap', in J. Hintikka (ed.), *Rudolf Carnap, Logical Empiricist: Materials and Perspectives*, Reidel, Dordrecht, 1975, pp. 87–94; and R. Creath, 'On Kaplan on Carnap on significance', *Philosophical Studies*, 30, 1976, pp. 393–400. It is interesting to note that Creath resorted instead to legitimation of sequences of sets of terms, rather than just sequences of terms, in response to Kaplan's problem. But he need not have!—the notion of skeletal validity would have done the trick. Ironically, Creath's introduction of sequences of sets of terms is needed for an altogether different reason—namely, to capture holistically interdependent theoretical concepts. Nor did Creath observe that his liberalization to allow sets of terms in legitimation sequences would render Carnap's own proof of what we have called his 'Quarantine Theorem' incorrect. Note that Creath's elaborate requirement on each set in the sequence is furthermore rendered otiose by our requirement of constrained creative extensions.

7. Carnap's extender always has its new non-basic term as its sole descriptive term. In our extenders, by contrast, several new non-basic terms may be brought into play simultaneously. Hence we can account for the introduction of holistically interdependent theoretical concepts.

8. Carnap has no analogue to our constraint condition on extension, which is what renders his account vulnerable to Ullian's collapse proof. We shall see presently how our account parries the general thrust of such collapse proofs.

There are two strategies embodied in our response to Ullian-style collapses of the original criteria of Ayer and Carnap.

One is the *Hempelian strategy*: to insist that the theory must feature as an extra relational term in the analysis of cognitive significance. We must appreciate that the 'new' expressions are initially uninterpreted, but acquire their interpretations by being incorporated into theoretical conjectures (in the theory) that turn out to be extenders. On this strategy, the extension can be required to take place in a stricter or in a more liberal form. The stricter form insists on forging new inferential links among basic sentences. The more liberal form requires only that new inferential links be forged among already legitimated sentences (some or all of which might not be basic). The constraint we are imposing on these new inferential links, however, makes up for the liberalism just mentioned.

The Hempelian strategy was of course first explicitly adopted by Carnap. It allowed him at the same time to focus on theoretical terms instead of whole sentences, and to exploit compositionality to form significant sentences from legitimated terms. As Schlesinger has emphasized,[23] this was

> the special power of Carnap's new meaning criterion

and it

> truly represents a turning point in the history of the efforts to formalize empirical significance. For ... the earlier efforts of verificationalists were concentrated upon the devising of a criterion to test the meaning of whole sentences and not of terms. All the criteria suggested in the past however could be circumvented by devising meaningless sentences which complied with those criteria.

[23] G. Schlesinger, 'The formalization of empirical significance', *Philosophy of Science*, 31, 1964, pp. 65–7; at p. 66.

In the light of this remark it is worth noting that the recent attempt by Wright[24] to improve upon the Ayerian account did not follow Carnap in adopting the Hempelian strategy.

The other strategy in response to Ullian-type collapses of Ayerian accounts is what we shall call the *constraint strategy*. It insists on more stringent conditions on the kind of logical entailment that is involved when the new 'extender' sentences in S forge inferential connections not previously available among already significant sentences. This was, in effect, the only strategy employed by Wright,[25] with his requirement of what he called 'S-compact entailment'. On our account the strategic thought here in question shows up as the requirement of constrained creative extension.[26] The constraint strategy also comes in both a stricter and a more liberal form. The stricter form requires that one appeal, in the evaluation proof or disproof within any countermodel to the unextended sequent, to basic axioms. The weaker form requires appeal merely to already legitimated axioms.

There is an important, independent element of intuitive appeal about the constraint strategy, even though the Hempelian strategy—indeed, the Hempelian strategy in its more liberal form—suffices for the proof of what we have called the Quarantine Theorem. We incorporate the constraint strategy as well, even though only in its more liberal form. A beneficial and non-*ad hoc* result is that we acquire a prophylactic to Ullian-style collapse proofs. But—more importantly—we thereby acquire also a much-needed immunity to Kaplan's problem of definitional extension and his problem of deoccamization, which Kaplan raised for Carnap's account.

Carnap seems, in the account in 'The Methodological Character of Theoretical Concepts', to neglect to discuss the original criterion put forward by Ayer, and Church's collapse of it. He does not make enough expository appeal to the consideration leading to the Hempelian strategy, with its insistence on the theory as an extra relational term. He was also writing in

[24] C. Wright, 'Scientific Realism, Observation and the Verification Principle', in G. Macdonald and C. Wright (eds.), *Fact, Science and Morality*, Blackwell, Oxford, 1986, pp. 247–74. See D. Lewis, 'Statements Partly About Observation', *Philosophical Papers*, 1988, pp. 1–31 for criticism of Wright's proposal, and C. Wright, 'The Verification Principle: Another Puncture - Another Patch', *Mind*, 98, 1989, pp. 611–22 for an emendation still within the self-imposed limitation of the second strategy discussed next in the text.

[25] loc. cit.

[26] Constrained creative extension by S is distinct from Wright's S-compact entailment. We require (for extension) that a whole (infinite) family of logically distinct invalid sequents be involved in theory extension by S, whereas Wright requires only that some invalid sequent be so involved.

the wake of Quine's celebrated essay 'Two Dogmas of Empiricism', in which the slide to pragmatic holism was consummated, with not even jaded regard for the fortunes of the technically demanding approach that Carnap was still pursuing. Perhaps Carnap's difficult prose, at the end of a long and wearying tradition of previous failed attempts to formulate a technically adequate criterion, lost the day for logical empiricism over pragmatic holism.[27] Quine's striking metaphors had greater appeal.

Carnap's account is at a disadvantage with its insistence on the whole theory occupying a uniform argument place in the analysis. This is in contrast to the approach adopted above, in which the theory itself is built up as new 'extenders' are adopted as theoretical conjectures. We need to show that we can deal with non-finitely axiomatizable theories; but we must avoid giving an analysis that works for these only by having such theories as wholes in the relational analysis of significance.

For this would be to concede that the context principle had taken a striking turn for the worse. If non-finitely axiomatizable theories had to appear as wholes in the relational analysis of significance, then the context principle would have at best the following invidious formulation:

> Only in the context of a whole theory does a theoretical term get its meaning

On our analysis, by contrast, we are still able to maintain that

> Only in the context of a theoretical sentence does a theoretical term get its meaning

This important result is secured, on our analysis, by the compactness of consequence.

What we have called the Quarantine Theorem has been proved here, rather surprisingly, on a very liberal formulation of the criterion for cognitive significance, even though that liberalization has entailed more work in the proof than was involved for the corresponding but cruder result of Carnap's. Carnap's analysis insisted on the new expression being the sole descriptive

[27] The secondary literature on Carnap's proposal is rather limited and indecisive. The problems that were raised for it seem to have been dealt with reasonably adequately by Carnap's defenders. Thus W. Salmon, 'Barker's Theory of the Absolute', *Philosophical Studies*, 10, 1959, pp. 50–3 against S. Barker, *Induction and Hypothesis*, Cornell University Press, Ithaca, 1957, pp. 136–42; and Creath, loc. cit. 1976 against Kaplan, loc. cit. 1975. None of these critics, however, raised the problem of otiose conjuncts: the problem, namely, that in the extended sequent $S \wedge (A \supset B), A : B$ the conjunct S is 'doing no work'.

term within the extending theoretical sentence. Inspection reveals that his own proof of his related result makes essential use of this condition. On our analysis, by contrast, we allow for arbitrarily finitely many new terms all at once within the extending theoretical sentence. This surely improves the prospects for legitimating concepts that are holistically interdependent in a scheme of explanation, in the way, for example, that Peacocke has tried to explicate.[28] The other liberal feature of our account, which has not jeopardized proof of the Quarantine Theorem, is that extension is required only among legitimated sentences, not basic ones.

11.5 Blocking Church–Ullian collapses

Ayer's original criterion can be expressed as a set of rules that serve to define the notions involved. The reader will easily check that the rules below are faithful to Ayer's own formulation. We shall use 'O', with or without subscripts or superscripts, for observation sentences. Note that Ayer, notoriously, omitted reference to any theory with respect to which significance was to be conferred on sentences. The rules capturing Ayer's criterion are:

$$\frac{}{O \textit{ is directly verifiable}}$$

$$\frac{\begin{array}{l} O_1, \ldots, O_n \nvdash O \\ S, O_1, \ldots, O_n \vdash O \end{array}}{S \textit{ is directly verifiable}}$$

$$\frac{\begin{array}{l} P_1, \ldots, P_n \nvdash P \\ S, P_1, \ldots, P_n \vdash P \\ P \textit{ is directly verifiable} \\ \textit{Each } P_i \textit{ is analytic or directly verifiable or can be shown} \\ \quad \textit{to be indirectly verifiable without showing that } S \textit{ is} \end{array}}{S \textit{ is indirectly verifiable}}$$

Instead of direct or indirect verifiability, let us talk instead of significance (at or by a certain level). Without changing the scope of what they license as

[28] See his *Holistic Explanation: Action, Space and Interpretation*, Clarendon Press, Oxford, 1979.

significant, the rules just given can be turned into a more clearly inductive form of definition as follows:

$$\frac{}{O\ \textit{is significant at level}\ 0}$$

$$\frac{S, O_1, \ldots, O_n : O\ \textit{is perfectly valid}}{S\ \textit{is significant at level}\ 1}$$

$$\frac{\begin{array}{l} S, P_1, \ldots, P_n : P\ \textit{is perfectly valid} \\ P_1, \ldots, P_n\ \textit{are significant at level}\ m \\ P\ \textit{is significant at level}\ 1 \end{array}}{S\ \textit{is significant at level}\ m+1}$$

These rules of Ayer's succumb to the following collapse proof originally due in a different form to Ullian, and presented here in the notation developed above. The steps in the proof are justified either by Ayer's rules above or by obvious facts about logical deducibility:

$$\dfrac{\dfrac{\dfrac{}{S, O \vdash S\wedge(O\vee\neg O_1)} \quad \dfrac{O \nvdash S}{O \nvdash S\wedge(O\vee\neg O_1)}}{S, O : S\wedge(O\vee\neg O_1)\ \textit{is perfectly valid}} \qquad \dfrac{\dfrac{\dfrac{}{S\wedge(O\vee\neg O_1), O_1 \vdash O} \quad S \nvdash \quad O_1 \nvdash O}{S\wedge(O\vee\neg O_1), O_1 : O\ \textit{is perfectly valid}}}{S\wedge(O\vee\neg O_1)\ \textit{is significant at level}\ 1}(\dagger)}{S\ \textit{is significant at level}\ 2}$$

The proof establishes that, if the observation sentence O_1 does not imply the observation sentence O, which in turn does not imply the contingent sentence S, then S is significant. These are extraordinarily weak conditions on S. We might as well say that any contingent sentence S comes out as significant (indeed, already at level 2) on Ayer's definition. This is a spectacular collapse indeed.

It is the step marked (†), in our view, that has to be blocked at all costs. Note that even Carnap's 1956 account,[29] if it is taken to apply at the propositional level, allows the step marked (†) to go through unchallenged, albeit only with the conclusion '$S \wedge (O \vee \neg O_1)$ is significant at level 1 with respect to any theory containing it'. But this is reason to be worried, even with such relativization with respect to a theory. For the intuition is that the embedded sentence S is not doing any work in securing the passage from the two premisses $S\wedge(O\vee\neg O_1)$ and O_1 to the conclusion O. S is theoretically idle. It should not be regarded as a significant sentence, despite the fact that it

[29] 'The Methodological Character of Theoretical Concepts', loc. cit.

would be significant—on Carnap's account—by virtue of being a constituent expression of the now supposedly significant compound $S \wedge (O \vee \neg O_1)$; and despite the fact that it would be significant—on Ayer's account—by virtue of the subsequent step in the proof above.

How, then, can we block the step (†)? We have suggested, in effect, a two-pronged defence. First, we are requiring the significance-conferring extensions to have to involve the new vocabulary in order to effect the extensions in question. In the Church–Ullian cases this requirement is violated. A simple truth-functional combination $O \vee \neg O_1$ of constituents already in the language will effect the sought extension of the sequent $O_1 : O$. The theory being extended is not, in our terminology, $\{O_1 : O\}$-less. We can clinch the family $\{O_1 : O\}$ by theoretical extension within the language already at hand. This first prong alone is enough to block the step (†).

The second prong of defence against the step (†) is to take seriously the idea that S is 'not doing any work' despite the perfect (indeed: skeletal) validity of the sequent $S \wedge (O \vee \neg O_1),\quad O_1 : O$. This second line of defence will be needed even when dealing with genuine creative extension for some family $\mathcal{F}$, where the unextended theory is indeed $\mathcal{F}$-less. For one has to avoid otiose conjuncts from piggy-backing on the theoretical assertions that effect the creative extension, and from thereby acquiring the status of cognitive significance without doing any proper work. In the Ullian example above, the otiose conjunct S is not doing any work because it fails to 'join forces with' O_1 and O in making the sequent $S \wedge (O \vee \neg O_1), O_1 : O$ valid. In other words, $S \wedge (O \vee \neg O_1)$ is not a *constrained* extender of the invalid sequent $O_1 : O$. There is a counter-example M to $O_1 : O$—namely, the one in which S is false—in which the evaluation, as false, of the compound sentence $S \wedge (O \vee \neg O_1)$ need not exploit the truth of O_1 or the falsity of O. The M-disproof

$$\cfrac{\cfrac{S \wedge (O \vee \neg O_1)}{S}}{\perp}$$

uses only the non-basic axiom (rule) $S|\perp$. It makes no use of the basic axioms $|O_1$ or $O|\perp$. This is what explicates the intuition that S 'does no work' in making the sequent $S \wedge (O \vee \neg O_1), O_1 : O$ skeletally valid.

The requirement of constrained creative extension also serves to dispose of Foster's problem,[30] whereby $Pa \wedge \neg Pb$ serves to secure $\neg a = b$, no matter

[30] J. Foster, *A. J. Ayer*, Routledge & Kegan Paul, 1985, pp. 19–20.

how 'metaphysical' the predicate P may be. First, the example cannot even get off the ground since no theory is $\{a = b : \emptyset\}$-less. So we would never have to resort to $Pa \wedge \neg Pb$ as a creative extender for the family $\{a = b : \emptyset\}$. But even if we did, the extended sequent $Pa \wedge \neg Pb : \neg a = b$ violates our constraint. Any countermodel M to the unextended sequent $\emptyset : \neg a = b$ makes $Pa \wedge \neg Pb$ false, and any M-disproof of $Pa \wedge \neg Pb$ uses only non-legitimated axioms. Once again, we have a two-pronged defence against an attempted counter-example, with each prong sufficient.

It is in order to avoid a simple but decisive problem for both Ayer's and Carnap's accounts that we have made use of the notion of constrained creative extension. All we have done is give formal expression to an intuition which, we are convinced, was guiding their thinking without coming explicitly to the surface. Ayer thought only in terms of sentences, not terms; he stayed at the propositional level; and he neglected the Hempelian 'theory parameter'. Carnap advanced the thinking on this topic at the time so as to cover terms as well as sentences; he ventured to first-order, and adopted a uniform theory parameter. We have followed Carnap in treating terms as well as sentences. But we have tried to show also that we not only may, but must venture to first-order. Furthermore, we have allowed the theory-parameter to vary (by 'growing monotonically') in a way that is faithful to the development of scientific theories as they extend their explanatory and predictive reaches in between incidents of refutation and revolutionary theory change. And we have tightened the logical linkages exploited to confer significance, by insisting on constrained creative extensions.

The result is an account of cognitive significance that, so far as we are able to determine, avoids extant problems for previous accounts; and provably avoids collapse. It is also formulated liberally enough, we believe, to find application in suitably regimented reconstructions of scientific theories. It is entirely in harmony, also, with the anti-realist outlook in semantics, epistemology and ontology. Our account of cognitive significance roots humanly graspable scientific contents in the humanly sensible. It pictures those contents as compositional. With its emphasis on the truth as in principle knowable (or at least, in the empirical case: on falsity as in principle knowable!) it motivates the search for a criterion to rule out metaphysical gobbledegook of the more pernicious varieties. Modern anti-realism, by insisting on compositionality, and insisting that the cognitively meaningful must be systematically graspable, averts the slide to pragmatic holism and emerges, albeit belatedly, as the proper heir to the tradition of Logical Empiricism.