

Logical Form and Radical Interpretation

TIM McCARTHY*

Abstract This paper concerns the empirical constraints on a characterization of logical relations in a natural language. Syntactic characterizations are distinguished from model-theoretic ones. It is shown that the structure of syntactic characterizations is largely underdetermined by the empirical constraints that naturally suggest themselves. However, an explanation of the notion of a logical constant is suggested that renders the model-theoretic characterization of logical relations in an extensional language determinate, relative to an idealized intentional psychology for its speakers.

In the simplest cases a semantic theory for a natural language assigns semantic data to the syntactically primitive expressions of the language, and specifies how the semantic properties of complex expressions are determined by the semantic properties of their parts. A *logical* theory for such a language identifies certain semantically primitive expressions whose interpretations are to be held fixed, in some sense, in characterizing a consequence relation for the language. These are the *logical constants* of the characterization. A sentence *B* is said to be a consequence of a sentence *A* if, very roughly, the interpretations assigned to the logical constants alone guarantee that *B* is true when *A* is.¹ The *manifestation problem* for semantics is that of saying what sort of empirical content a semantic theory has; or what constitutes evidence in semantics; or what sort of empirically accessible facts a semantic theory predicts or explains and how it explains them.

This paper is about the manifestation problem for *logic*. Although the manifestation problem for semantics has been widely discussed, the manifestation problem for logic—the question of what constitutes evidence for a character-

*This paper has grown out of a paper [8] I prepared for a symposium on William Lycan's book *Logical Form in Natural Language* at the American Philosophical Association Central Division meetings in Cincinnati, April 1988. I would like to thank Lycan for valuable comments.

Received April 11, 1989

ization of logical relations in a natural language—is less familiar. The problems are, of course, distinct. A semantic interpretation of a language does not, by itself, settle the issue of which expressions of the language should be counted as logical constants. Nor does it fix the precise role to be assigned to those constants in the characterization of logical relations for the language. Thus our question is in part one of what empirical reasons can be given for a choice of logical constants and in part what empirical reasons can be given for a particular interpretation of them. In what follows I will distinguish syntactic from model-theoretic characterizations of logical relations for an extensional language. The body of the paper explores the empirical consequences of characterizations of both types.

I Let L be a natural language or a fragment thereof. We shall assume both a syntactic and semantic description of L to be given, and for the purposes of this discussion we shall assume that these descriptions are of simple but familiar types. In particular, we assume a syntax for L that generates the well-formed formulas of L from formulas of a categorial language L_0 . L_0 contains two categories of *formulas* and *terms*, and two subcategories of syntactically primitive terms, *variables* and *individual constants*. The *descriptive constants* of L_0 comprise its formulas and terms, and are built up from the primitive terms by operators that apply to cartesian products of the four categories. Thus, n -place predicates map n -tuples of terms to formulas, quantifiers map variable-formula pairs to formulas, and n -place function symbols map n -tuples of terms to terms.

Secondly, we shall assume L_0 to have an extensional semantics of a rather simple sort.² If D is a nonempty set and Z a descriptive constant of L , the *extensions of appropriate type for Z over D* are functions from D^ω to D if Z is a term and subsets of D^ω if Z is a formula. If v is the n -th variable in a standard enumeration, the unique extension of appropriate type for v over D is the projection function from D^ω to D taking any sequence onto its n -th component, and if c is a constant the extensions of appropriate type for c over D are the constant functions from D^ω to D . If K is an operator that maps expressions from a product $T_1 \times \dots \times T_n$ of categories to a category T , then the extensions of appropriate type for K over D are functions mapping sequences of extensions of appropriate type for expressions of T_1, \dots, T_n over D to extensions of appropriate type for expressions in T over D .

An *extensional interpretation I* of L consists of the following data:

- (i) A nonempty set $|I|$, the *domain* of I
- (ii) A function F_I that maps each constant of L onto an object from $|I|$
- (iii) A function G_I that maps each operator of L onto an extension of appropriate type for it over $|I|$.

Given such an interpretation I of L , the extension of any descriptive constant of L in I is determined inductively in the obvious way. The extension of a constant c is the constant function mapping each sequence over $|I|$ onto $F_I(c)$. The extension of any variable is the (unique) extension of appropriate type for it over $|I|$. If K is an operator of L and Z_1, \dots, Z_n are expressions to which K applies, then the extension of $K(Z_1, \dots, Z_n)$ in I is obtained by applying $G_I(K)$

to the sequence consisting of the extensions of Z_1, \dots, Z_n in I . A sequence from $|I|$ that appears in the extension of a formula in I is said to *satisfy* that formula on I . A formula is *true on I* if it is satisfied by all sequences on I .

A *logic* for L will be a collection of operators from L , comprising a possible demarcation of the logical constants. These are the expressions of L whose interpretations are to be held fixed in characterizing the consequence relation for L . But there are different sorts of characterization; most familiar are the model-theoretic ones. On a model-theoretic characterization a formula A is a consequence of a set S of formulas in L roughly if A is satisfied by any semantic interpretation of L that satisfies each member of S and which respects the interpretations of the logical constants. For this to make sense, the "interpretation" of a logical constant must assign it an extension over each nonempty domain. Accordingly, a model-theoretic interpretation of a logic for L will be a function mapping each operator K of the logic onto a function (more precisely, a functional class) which in turn maps any nonempty set D onto an extension of appropriate type for K over D . If J is an interpretation of a logic for L , an interpretation I of L is said to be *admissible for J* iff

$$I(K) = J(K)(|I|)$$

holds for each operator K in the domain of J . A formula A is said to be a *consequence* of a set S of formulas in L with respect to J iff, for any interpretation I of L such that I is admissible for J , and any sequence s from $|I|^\omega$, if s satisfies each formula in S on I then s satisfies A on I .

Opposed to (or perhaps beside) model-theoretic characterizations of the consequence relation stand various syntactic ones. As I will use the phrase "syntactic characterization" *both* syntactic *and* model-theoretic characterizations of logic construe the consequence relation as a *semantic* one: a consequence of a set S is a sentence whose *truth* is ensured, in some as-yet-unspecified sense, by the truth of each sentence in S . The difference is that a syntactic characterization associates a chosen logic with a formal theory from which the relevant semantic relations of sentences are *deducible*. For example, a sequent calculus for the logical constants of a language L may be interpreted as a syntactic characterization of the consequence relation for L : a sequent of the form $X \vdash Y$ is naturally interpreted (and was interpreted by Gentzen [2]) as saying that if each of the formulas in the set X is valued as true in L then at least one of the formulas in Y must be so also. The rules of the sequent calculus connect semantic relations of this type. Another example would be a Tarskian truth theory for L or a fragment thereof. In either case, a formula A may be said to be a consequence of a formula B with respect to the characterization if (an equivalent of) the sentence

(T) B is true in $L \rightarrow A$ is true in L

is deducible in it. Call a sentence of the form (T) a *T-conditional* for L . An expression is a logical constant for a syntactic characterization if a rule describing the semantic behavior of the expression appears in the characterization.

The manifestation problem for characterizations of either sort is to say what counts as evidence for a characterization of that sort and why. It is to be hoped that a conception of evidence for logic can be found which is sufficiently

robust to enable us to choose between *rival* characterizations; but it is not yet clear when two characterizations are rivals. The primary constraint on the compatibility of characterizations is that compatible theories cannot generate divergent consequence relations. But this is only a necessary condition, for two characterizations may determine the same consequence relation in incompatible ways. Thus, for example, two model-theoretic characterizations of logical relations in a language *L* might determine the same consequence relation for *L* and agree on what counts as a logical constant, but assign different extensions to the constants over certain domains. It seems natural to say that two model-theoretic characterizations for *L* are compatible when, and only when, they identify the same expressions of *L* as logical and assign the same extensions to them over each domain. It follows from this that they determine the same consequence relation for *L*.

What of syntactic characterizations? A syntactic characterization of the notion of logical consequence for *L* is a formal theory *T* that gives a partial semantic description of *L*. What makes *T* into a characterization of the *logic* of *L* is our decision to count as a *logical* consequence of a set of sentences in *L* any sentence whose truth is deducible from *T* in conjunction with the hypothesis that the members of the set are true. If two such theories are phrased in a common first-order language, it seems natural to say that they are compatible when they are jointly consistent and determine, in the present sense, the same relation of logical consequence for *L*.³

2 William Lycan has suggested a perspective on syntactic characterizations which may be described briefly as follows:

- (1) A syntactic characterization of logical relations in a language *L* is generated by a fragment of a formal theory of truth for *L* (roughly, the part dealing with the logical constants). Such a theory satisfies the constraints described by Davidson.

Lycan also has an explicit view about how such a characterization is confirmed:

- (2) A characterization of logical consequence for *L* is confirmed by reference to its predictions of intuited or felt implications in *L*. Such a prediction is provided by a deductive explanation of a T-conditional in terms of the semantical rules associated with the logical constants.⁴

The present suggestion does not, as stated, throw any light on the question of how to *identify* the logical constants. One might naturally suppose that a logical expression is distinguished by the fact that a semantic rule for it plays the right sort of role in explaining implications, though this is left unspecified. I shall return to this matter below. In any case, even without an explicit demarcation of the logical constants, Lycan's proposal is still of interest, for it purports to explain the sort of data that are associated with the constants in a characterization of the logic of a natural language, and the sort of empirical content such a characterization has.

One question facing us is that of *how* the ability of a truth theory to generate explanations of intuited implications (assuming for the moment that it *does* have this ability) confers *empirical content* on the theory. It is important to dis-

tinguish *intuited implications* from *intuitions of implication*.⁵ Intuitions of implication are empirically manifested in a variety of ways; for example, an agent that instantiates an intuition that p implies q will normally attach a high conditional epistemic probability to q relative to p , and this fact will normally be reflected in his linguistic (and other) behavior. But it is not at all clear what the empirical consequences of the intuited *implications* are.

It might be suggested that logical theory is, in this respect, no worse off than physical theory. A theory in physics may be empirically confirmed if it affords a good explanation of facts from which other beliefs are empirically (typically, perceptually) derived, e.g., meter readings; it need not explain the beliefs themselves. It might be argued that intuitions of implication are analogous to beliefs derived from perception, and that a characterization of logical relations in a natural language has empirical content in virtue of explaining the *objects* of such intuitions: the implications themselves.

This response is suitable only if implications, like meter readings, are in an appropriate sense "empirically accessible"; that is to say, only if speakers can have "empirical knowledge" of implications, whatever that may mean. The mere fact that speakers regularly have intuitions about implication does not suffice to establish the empirical accessibility of implication (the fact that individuals in medieval society regularly had intuitions about the presence of demonic forces does not establish the empirical accessibility of demonic forces). The intuitions must be generated by an epistemic mechanism of the appropriate type.

There is, however, another component of Lycan's view which might seem to offer at least the beginnings of an account of such a mechanism. Lycan holds that the semantic competence of a speaker of a natural language L is explained by the fact that the speaker instantiates a theory of truth for L of the sort mentioned under (1) above. The theory is postulated to be phrased in an internal code whose interpretation is given independently of that of L . The speaker is postulated to understand a sentence of L (the public, natural language) by tacitly deriving a representation of its truth-condition from the internally coded truth theory.⁶ This bold empirical hypothesis raises many interesting questions which I cannot discuss here. If it is correct, however, it immediately suggests a possible explanation of intuitions of implication: a speaker acquires intuitions of implication by tacitly deriving a sentence representing the implication from an internally coded truth theory. Such a representation might be given by a T-conditional.

Now the present suggestion does not, by itself, yield quite what we want. What is wanted is an explanation of the possibility of *empirical knowledge* of implications; what we have so far is a sketch of an explanation of intuitions of implication. But no account has yet been given of the epistemic status of these intuitions. Let us suppose, for example, that we discovered that certain individuals derive intuitions about demonic forces from unconsciously represented demonologies ("internally represented demonic theories"). The intuitions in question have as their objects states of affairs whose existence is asserted by those very theories. But the intuitions are (presumably) quite without epistemic force.

What is needed, then, is an account of how the intuitions of implication derived from an internally coded truth theory yield knowledge (or, at least, an epistemic attitude analogous to knowledge) of those implications. Although it

is rather unclear what form such an account ought to take, the following two constraints will perhaps be assented to readily:

- (3) The truth theory for *L* instantiated by a speaker of *L* must be true.
- (4) There must be an appropriate connection between the facts reported by that truth theory and the fact that the speaker instantiates the theory.

Both (3) and (4) are suggested by the idea that, if a speaker's internal representation of a truth theory is to ensure the epistemic status of intuitions derived from it, the representation itself must satisfy conditions which are at least analogous to those which an explicit belief must satisfy in order to constitute knowledge. One such condition is the truth of the belief; another is that the right sort of connection exists between the belief and the state of affairs it represents. (3) is analogous to the first condition, and (4) to the second.

In fact, it seems to me that Lycan's perspective affords some reasons for thinking (3) and (4) to be true. Lycan's theory seems to offer something like the following reduction of the semantic properties of a public natural language *L*: *L* has its semantic properties (in our present context, this would be to say that a certain extensional interpretation is true of *L*) in virtue of the fact that speakers of *L* generally instantiate truth theories that *ascribe* these properties to *L*. A bit more explicitly:

- (5) A speaker's internally coded truth theory for *L* is true if and only if it stands in an appropriate relation of equivalence to the theories instantiated by *other* speakers of *L*.
- (6) The fact stated by a semantical axiom for an expression of *L* in a speaker's theory (that the expression in question has a certain extension in *L*) consists in or reduces to the fact that the theories instantiated by speakers of *L* generally ascribe that extension to that expression.

It seems further to be Lycan's view that if a speaker's semantic representations are caused in the right sort of way (by appropriate linguistic/environmental interactions with other speakers), then they will stand in the relevant equivalence relation to the theories instantiated by other speakers. By (5), then, the speaker's theory will be true, whence (3) will hold. Moreover, in virtue of (6) the facts stated by the theory do seem closely connected with the fact that the theory is instantiated by the speaker. In particular, if the theories instantiated by other speakers of *L* generally associated an expression with an extension *other* than its actual one, the speaker would not instantiate the semantic theory that he does. It is not implausible that this sort of connection between the facts stated by the speaker's semantic theory and the fact that that theory is instantiated by the speaker is sufficient to underwrite (4).

I cannot argue that in detail here. However, if (4) could be motivated in this way, we would have at least the beginnings of an account of the epistemic status of semantical intuitions. In particular, intuitions of implication derived in the appropriate way from a speaker's semantic theory can yield knowledge of those implications, or at least an epistemic attitude analogous to knowledge. And if a logical theory for *L* can *explain* items that are objects of such attitudes, the underlying intuitions may confirm the theory.

3 The above, I wish to stress, is multiply speculative: the empirical support for Lycan's hypothesis (that speakers of a natural language instantiate truth theories, and understand sentences of the language by deriving truth-conditions for them in those theories) is, to say the very least, rather thin. Further, more would have to be said about the connection between semantic facts and the instantiation of semantic theories to show that assumption (4) is satisfied. Finally, (3) and (4) were claimed to constitute only *necessary* conditions for the existence of the required epistemic attitude.

However, even if these matters can be taken care of, the status of our original question is still unclear; for, as I noted above, no explicit account is given in Lycan's proposal of how the logical constants of the language L are to be circumscribed. What is worse, from the standpoint of our problem, is that *it is not clear that drawing the line in one place rather than another will make an empirically definable difference*. A natural suggestion would be to count as a logical constant of L any expression of L whose semantic characterization is relevant to explaining implications on the above model. A bit more explicitly:

- (C) An expression E of L is a logical constant relative to a truth theory T for L iff there exist explanations in T of implications in L in which the semantical rule for E occurs essentially.

Unfortunately, however, certain intuitively nonlogical expressions are counted as logical on the present proposal. The difficulty is that a semantical rule for a nonlogical expression may occur essentially in an explanation of a *nonlogical* implication in T ; for example, let R be a predicate modifier of L associated with the following semantical rule in T :

(SR) $\forall a \forall F (a \text{ satisfies } RF \leftrightarrow a \text{ satisfies } F \ \& \ a \text{ is red})$.⁷

Assuming first-order logic in the metalanguage, we may then derive the T-conditional

$RF(t) \text{ is true} \rightarrow F(t) \text{ is true}$

for any predicate F and term t , and thus explain, on the above model, why $RF(t)$ implies $F(t)$. Since R is, intuitively, a nonlogical predicate modifier, (C) results in a too liberal demarcation of logic. However, if we contract the set of logical constants in the light of this intuition the resulting demarcation of logical from nonlogical expressions has no obvious explanatory relevance. We can still, after all, account for the implication between $RF(t)$ and $F(t)$: only now we shall call it a *nonlogical* implication. The difficulty is that the form of explanation we are now considering makes no explicit *use* of a distinction between logical and nonlogical expressions; the availability of such explanations will be unaffected by our demarcation of the logical constants, assuming that we are willing to countenance a sufficiently broad class of nonlogical entailments. What is required, then, is not simply a demarcation of the logical constants, but an account of the explanatory relevance of the demarcation, an account that will reveal the empirical differences between one demarcation and another.

This problem arises for other syntactic theories against the background of the assumption that the empirical content of the theory resides in its ability to explain the objects of judgments of implication that speakers of the language

in question make *in virtue of bearing a (tacit or explicit) epistemic attitude to the theory*. Another example is provided by the idea of taking the syntactic characterization to be generated by a sequent calculus of the sort described by Gentzen (see [2]). Hacking has proposed in [3] that the logical/nonlogical boundary for an extensional language be drawn in terms of the characterizability of the logical constants by a theory of this sort. On this conception, a logical constant is roughly an expression for which a set of Gentzen-type operational rules can be given that makes a finitistic proof of cut-elimination possible for derivations of sequents. The notion of logical consequence may then be characterized in terms of sequential derivability.

It might be suggested that such a theory constitutes a representation of what a speaker knows by knowing the meanings of the logical constants: the operational rules may be regarded as describing the cognitive transitions of an idealized speaker, and as *approximately* describing, in some appropriate sense, the cognitive operations of actual speakers. The present suggestion thus fits nicely into the framework described above: the empirical content of a sequent calculus for the constants of a language may be identified with its ability to explicate the objects of judgments of entailment that speakers actually make in virtue of bearing the relevant epistemic relation to the calculus. However, the problem is again that a speaker might bear the *same* epistemic relation to rules that characterize *non*logical expressions. Consider, for example, the nonlogical modifier *R* above. *R* may be described by the following operational rules:

$$(R1) \quad \frac{J, F(t) \vdash K}{J, RF(t) \vdash K}$$

$$(R2) \quad \frac{J, \text{RED}(t) \vdash K}{J, RF(t) \vdash K}$$

$$(R3) \quad \frac{J \vdash F(t), K \quad J \vdash \text{RED}(t), K}{J \vdash RF(t), K}$$

for any parameter *t*, where RED is a predicate of *L* with the obvious meaning. If we suppose, then, that a speaker of *L* instantiates these rules (in the relevant sense), we can give just the same sort of explanation of the nonlogical relation between *RF(t)* and *F(t)* as one would give of logical relations in terms of the operational rules characterizing the logical constants.

It would appear that modifiers such as *R* constitute an obstacle to the proposal that the logical constants be circumscribed in terms of their characterizability by a sequent calculus. I do not claim that it is an insurmountable obstacle. In fact, there are natural restrictions on the form of operational rules that would exclude modifiers of this sort; for example, the subformula constraint.⁸ The difficulty lies in seeing how, on the suggested account of how a syntactic characterization is confirmed, there could be any empirical reason for invoking such a constraint. On that account, the following two hypotheses seem to sit equally well with the evidence:

- (a) The logical constants of *L* are those characterized by Gentzen-type operational rules satisfying the subformula constraint, and the entailments generated by the rules *R1*–*R3* are nonlogical.

- (b) The logical constants of L are characterized by operational rules without the subformula constraint, and the entailments generated by R1–R3 are logical.

Although we are inclined to say that modifiers such as R are nonlogical, this judgment seems to have no empirical basis given the conception of evidence for syntactic characterizations of logic that we have been examining.

4 I shall now discuss the manifestation problem as it arises for model-theoretic descriptions of the logic of an extensional language. Our starting point is a minimal structural constraint on a definition of logical consequence: the logical consequences of a sentence should be implied by that sentence. This is vacuous in the absence of something substantive to say about the general notion of implication; thus our first task will be to characterize the notion of implication in such a way that requiring the model-theoretic entailment relation to be a species of implication places an empirical constraint on the model-theoretic relation. I will then give a characterization of the logical constants that makes available an explanation of *why* instances of model-theoretic entailment in an extensional language are instances of implication therein. This explanation will provide an empirical reason for identifying the property of being a logical constant in an extensional language with the characterizing property mentioned. The effect of the proposed identification will be to reduce the manifestation problem for logic to that of an idealized intentional psychology for speakers of the language in question.

Let L be the language of an idealized cognizer whom we shall call 'Karl'. We assume that an extensional semantics for L has been given. We also assume that a description of Karl's *attitudes* (in particular, his beliefs and desires) in the familiar form of a demarcation of the possible situations *compatible* with his attitudes has been given. We shall say, for example, that Karl holds a sentence of his language to be true if it is true in each possible situation compatible with his beliefs; and analogously for his desires. By a 'possible situation' or 'world' in this context I will mean an *epistemically* possible situation, described by a set of propositions which incorporate the propositions that Karl holds true in some possible state of evidence.⁹

For our present purposes it will suffice to take a nominalistic view of these situations, identifying them roughly with certain sets of sentences in Karl's language or a syntactic extension thereof. For the purpose of describing epistemically possible situations, the parts of these languages that outrun Karl's actual language may be viewed as uninterpreted; the formulas of the uninterpreted fragment function as placeholders for epistemically possible properties, and the constants as placeholders for possible objects. The extension of a formula $A(x_1 \dots x_n)$ at a world w is represented by the sequences $\langle c_i \rangle_{i \in \omega}$ of individual constants such that the sentence $A(c_1/x_1 \dots c_n/x_n)$ holds at w . Such a syntactic ersatz world, then, fully describes the *structure* of an epistemically possible situation, but only partially describes its content.

Suppose we are given a description of Karl's attitudes in terms of such a set of syntactically characterized epistemic possibilities. Call this set the *inten-*

tional space of the description or, where the description is fixed, simply *Karl's intentional space*. The elements of such a set will be called *epistemically possible situations*, or simply *worlds*. Within this framework there is a natural characterization of implication in Karl's language L : if A and B are sentences of L , then A implies B iff B occurs in each epistemically possible situation in which A occurs. Roughly, this is to say that A implies B iff it is not *epistemically possible for Karl* that A is true and B is false in L . If Karl's intentional space is fixed, then the relation between the notions of consequence and implication in Karl's language constrains a characterization of the consequence relation by forcing the consequences of a sentence to appear in each world in which the sentence is realized.

It is easy to see in terms of an example how a model-theoretic characterization can run afoul of this constraint. Let F and G be two predicates and c a constant such that Fc does not imply Gc in L . That is to say, Fc is true at some point in the relevant intentional space at which Gc is false. Suppose that F and G are both counted as logical constants in a model-theoretic characterization of the consequence relation for L , and let us assume that F and G have the same extension E over the ontology of L , and that this interpretation is extended to an arbitrary domain D by taking their extensions over D to coincide with $D \cap E$.¹⁰ Then Gc is true in any *model* (i.e., any extensional interpretation of L respecting the logical constants) that satisfies Fc , and so counts as a logical consequence of Fc in L . Thus we cannot simultaneously count the predicates F and G as logical constants and assign to these predicates the indicated extensions over an arbitrary domain.

A system of attitudes defined in the above-indicated way on a set of epistemic alternatives for Karl has an obvious empirical content. It is the basis of decision-theoretic predictions about Karl's behavior, and can be tested in terms of these predictions.¹¹ A characterization of the notion of consequence for Karl's language is constrained by a map of his intentional attitudes, and thus acquires some empirical content: it may be empirically undermined, disconfirming the systems of attitudes with which it is compatible. Of course, at this point the possibility exists that our intentional description of Karl may radically underdetermine the model-theoretic characterization of Karl's logic. Below I shall argue that this is not the case, that in fact the model-theoretic characterization of the notion of consequence for Karl's language L is *determined by the structure of Karl's intentional space*. On a model-theoretic account, logical relations in L are determined by a choice of logical constants together with a choice of extension for them over each domain. If we can see how the structure of Karl's intentional space determines both of these choices, the manifestation problem for model-theoretic characterizations of Karl's logic will have been reduced to the manifestation problem for descriptions of Karl as an intentional system.

5 In [10] Mostowski suggested a constraint on generalized quantifiers that I have elsewhere argued to be relevant to a demarcation of the logical constants.¹² One succinct statement of Mostowski's condition is that the extension of a generalized quantifier should be invariant under isomorphism. For this to make sense, of course, it is assumed that the extension of such a quantifier sym-

bol is defined for an arbitrary domain: we saw that this assumption was in any case required if such an expression is to play the role of a logical constant in a model-theoretic characterization. Thus, for example, suppose that Q is a quantifier symbol that combines with a variable v and a formula F to yield a formula QvF . The extension of Q over a domain D is then a function $\{Q^*\}$ mapping each pair (f, E) , where $f: D^\omega \rightarrow D$ is a projection function and $E \subseteq D^\omega$, onto a set $\{Q^*\}(f, E) \subseteq D^\omega$. Then the requirement is that any isomorphism of the structures (D, f, E) and (D', f', E') induce an isomorphism of the structures $(D, \{Q^*\}(f, E))$ and $(D', \{Q^*\}(f', E'))$. The invariance condition for expressions of other types is characterized analogously. I argued that it is reasonable to require the logical constants to be extensionally invariant in just this sense, on roughly the ground that we should expect the semantic role of a logical constant in a sentence to be determined by the *structure* of the subject matter of the expressions to which it is applied.

However, the present constraint does not, as it stands, ensure that a quantifier symbol satisfying it is a logical constant. The problem is that there can be accidental invariance. Consider, for example, a quantifier symbol Q of L translated into English by the phrase “for some . . . , if P , and for all . . . if $\sim P$ ”, where ‘ P ’ represents a contingent truth of English. Q extensionally coincides with \exists on any (actual) domain. If Q were treated as a logical constant, then the formula $(Qx)A(x)$ would be a logical consequence of $(\exists x)A(x)$ in L . If, as we have assumed, the model-theoretic entailment relation for L is imbedded in the implication relation for L , for any choice of $A(x)$, $(\exists x)A(x)$ must then imply $(Qx)A(x)$ in L . But it does not do this, assuming only that there exists a possible situation in which P fails and in which some but not all individuals fall under $A(x)$.

But notice that Q would be excluded as a logical constant if the suggested invariance constraint were applied to the extension of Q in an arbitrary pair of *possible* situations. That is to say, we would require of Q that, for any pair of worlds, $w, w^\#$ and any pair of predicates A, B , any isomorphism mapping the extension of A at w onto the extension of B at $w^\#$ also maps the extension of $(Qx)A$ at w onto the extension of $(Qx)B$ at $w^\#$. This condition will not be satisfied by a pair of worlds $w, w^\#$ if P holds at w , P fails at $w^\#$, and A and B are monadic predicates in x such that the extension of A in w is nonempty, nonuniversal, and isomorphic to the extension of B in $w^\#$. If Q fulfills the present generalization of the Mostowski property with respect to a collection M of possible situations, I will say that Q is *rigidly invariant over M* . I now suggest that the logical constants of Karl’s language be characterized as those operators of Karl’s language that are rigidly invariant over Karl’s intentional space.

Let us characterize this property a bit more fully. To do this, we shall have to be somewhat more explicit about the representation of the epistemically possible situations in Karl’s intentional space. I have assumed a syntactic characterization of these situations, one that identifies a possible situation with a state description phrased in some extension of Karl’s language. Let such a description take the form of a triple $w = (T_1, T_2, K)$, where T_1 and T_2 are theories in an extension L_w of Karl’s language and K is a collection of constants in L_w . The constants of K are to be regarded as correlated in a one-to-one way with the individuals in the possible situation described by w . T_1 and T_2 comprise respectively

the sentences of L true (false) at that situation, so that $T_1 \cap T_2$ is empty and $T_1 \cup T_2$ is the collection of all sentences in L_w . We suppose that the constants of elementary logic can be identified in L by reference to their syntactic behavior in structures of this sort.¹³

The *extension* of a formula $A(x_1 \dots x_n)$ of L_w at w is the set of sequences $\langle c_i \rangle_{i \in \omega}$ of constants from K such that the formula $A(c_1/x_1 \dots c_n/x_n)$ is in T_1 . The extension of a term $t(x_1 \dots x_n)$ is the function f mapping each sequence $s = \langle c_i \rangle_{i \in \omega}$ of constants from K to K such that $f(s) = c$ iff $\text{Id}(t(c_1/x_1 \dots c_n/x_n), c)$ appears in T_1 , where Id is the identity predicate of L_w . In view of the constraints placed on Id in note 13, f is well-defined. If Z is a finite sequence (Z_1, \dots, Z_n) of expressions from L and $w = (T_1, T_2, K)$ is a world, by the extension of Z at w we mean the structure (K, A_1, \dots, A_n) , where A_i is the extension of Z_i at w for each i , $1 \leq i \leq n$. Let Q be an operator that applies to sequences of expressions from a product $S = T_1 \times \dots \times T_n$ of syntactic types in L to yield an expression of type T . Then Q is *rigidly invariant* in L iff, for any pair of worlds w, w' from Karl's intentional space and any pair of sequences Z, Z' from S , any function that maps the extension of Z at w isomorphically onto the extension of Z' at w' also maps the extension of $\langle Q(Z) \rangle$ at w isomorphically onto the extension of $\langle Q(Z') \rangle$ at w' . The present proposal, then, is that the property of being a logical constant of Karl's language be identified with the property of being rigidly invariant therein.

6 What is the evidence for such an identification? If, as I suggested above, a basic empirical constraint on the relation of logical consequence is that its converse be a species of implication, a fundamental fact that wants explaining on any construal of what a logical constant *is* is why B is a logical consequence of A only if A implies B . Beyond that, we should want recognized logical constants classified as logical by the construal, and recognized nonlogical constants classified as nonlogical. We should want it to imply the Mostowski constraint and other entrenched model-theoretic generalizations about logic. Finally, if the suggested description of the logical constants is to generate a model-theoretic characterization of logical relations, it must determine the extension of a constant over an arbitrary actual domain. I believe that the identification of logical constanthood in Karl's language with rigid invariance accomplishes these aims, at least in the context of three natural assumptions about the structure of Karl's intentional space.

To state the first assumption, we require a definition. If $w = (T_1, T_2, K)$ is a world, two formulas of L_w will be said to be (*extensionally*) *equivalent* in w when they contain the same instances and their instances co-occur in T_1 and T_2 . Terms s and t are *equivalent* in w when the formulas $\text{Id}(s, x)$ and $\text{Id}(t, x)$ are equivalent in w , for a fixed variable x occurring neither in s nor in t . Finite sequences of expressions are *equivalent* in w when they are of the same length and are componentwise equivalent in w . Now, our first assumption is that operators behave extensionally within each world. A bit more explicitly, if w is any world and Q is an operator that is defined for a product T of syntactic types in L_w , then the expressions that result from applying Q to sequences from T which are extensionally equivalent in w are also extensionally equivalent in w .

The second assumption we shall require concerning the possibilities described in Karl's intentional space is, roughly stated, that *the actual is possible*. Within the present framework, this assumption may be formulated as follows. By an *extensional structure* I will mean a finite sequence (D, E_1, \dots, E_n) where D is a nonempty set of actual objects (the domain of the structure), and E_1, \dots, E_n are extensions over D appropriate for expressions of an extensional language. Let $A = (D, E_1, \dots, E_n)$ be an extensional structure, $w = (T_1, T_2, K)$ a world, and I an interpretation of L_w . I will be said to *satisfy* w iff each sentence of T_1 is true in I and each sentence of T_2 is false in I ; and I will be said to *realize* A in w iff

- (i) $|I| = D$ (the domain of A)
- (ii) I maps K one-to-one and onto D
- (iii) I satisfies w
- (iv) For each i , $1 \leq i \leq n$, there is an expression u_i of L_w such that $I(u_i) = E_i$.

We say that a structure A is *realizable* in Karl's intentional space if there is a world w and an interpretation I of L_w such that I realizes A in w . The second required assumption, then, is:

(A) All extensional structures are realizable in Karl's intentional space.

This is roughly to say that any actual structure can be characterized up to isomorphism by some state description which is epistemically possible for Karl.

One immediate consequence of (A) is that rigidly invariant expressions of a language satisfy the Mostowski constraint: if all actual structures are realizable in Karl's intentional space, any operator invariant under any isomorphism of the possibilities described therein is *ipso facto* invariant under isomorphisms of actual structures. Another slightly less obvious consequence of the assumption is that the extensions of the logical constants over each actual domain are determined uniquely. The example of a rigidly invariant quantifier symbol Q of Karl's language L that maps predicate-variable combinations Ax onto formulas $(Qx)Ax$ is typical. To determine the action of Q at a structure (D, E) ,¹⁴ find an interpretation I that realizes (D, E) in a world w , and suppose that $I(F) = E$ for a formula F in L_w . Then Q maps the structure (D, E) onto the extension that I assigns to $(Qx)F$. It is immediate from the invariance condition that this procedure yields the *same* extension for Q at (D, E) , regardless of which realization I of (D, E) is considered.¹⁵ However, it is not yet clear that it determines the *correct* extension for Q over the *ontology* of L , i.e., the domain of the given extensional interpretation of L . For this we require a further constraint, which is also quite natural:

(B) Any extensional structure whose domain coincides with the ontology of L is realized in some world by an interpretation that coincides with the intended interpretation of L on L .

Of course, such a structure may contain extensions that are not definable in the intended interpretation of L , and in this case I will assign these extensions to additional expressions of L_w . It is clear that, given (B), the indicated procedure allows us to recover the correct extension for Q over the ontology of L .

The provision of extensions over each actual domain for the logical constants fixes a model-theoretic characterization of the consequence relation for L in the manner described in Section 1 above. The assumptions I have stated now allow us to fulfill the main empirical constraint placed on the characterization, namely that it lead to an explanation of why model-theoretic entailments are instances of implication. The explanation flows from the following result:

Completeness Theorem *Any world is satisfied by an interpretation J that assigns to each logical constant of L its extension over $|J|$.*

*Sketch of Proof:*¹⁶ Fix a world $w = (T_1, T_2, K)$, and let p_1, \dots, p_n be the operators of L_w . Define a structure $A = (D, A_1, \dots, A_n)$ as follows: $D = K$. If u is a formula or term of L_w in the variables x_1, \dots, x_n , we denote by $\{u\}_w$ the extension of u at w . Suppose that p_i is defined for a product $T_1 \times \dots \times T_k$ of types from L_w ($1 \leq i \leq n$). If p_i is a logical expression, A_i is the intended interpretation of p_i on D ; if p_i is a nonlogical expression, A_i is any function on the corresponding product of types over D such that, if $p_i(u_1, \dots, u_k)$ is defined and equal to an expression v , then

$$A_i(\{u_1\}_w, \dots, \{u_k\}_w) = \{v\}_w.$$

Let I be a realization of the structure (D, A_1, \dots, A_n) in a world $w^* = (T_1^*, T_2^*, K^*)$. Define a translation F from L_w to the language of w^* as follows. For any constant c , $F(c)$ is the constant from K^* which in I designates the unique constant b from K such that $b = c$ appears in T_1 . $F(v) = v$ for each variable v , and for any operator p_i , $F(p_i)$ is an expression of L_{w^*} designating A_i in I . F is now extended to arbitrary formulas and terms in L_w inductively in the obvious way. Then, using the invariance property of the logical constants as between w and w^* , it may be shown that F maps T_1 into T_1^* and T_2 into T_2^* . The world w is now satisfied by the interpretation J of L_w on D defined by $J(e) = I(F(e))$ for each semantically valuable expression e of L_w . It is clear that J assigns to each logical expression its extension on D .

The desired property of the model-theoretically characterized consequence relation for L is an obvious consequence of the completeness theorem: if S is a set of sentences of L and A is a consequence of S , then S implies A . Otherwise, for some world $w = (T_1, T_2, K)$, we must have that $S \subseteq T_1$ and $A \notin T_2$. However, by the completeness result there exists an interpretation that satisfies w and assigns to the logical constants their intended interpretations. In this interpretation S is true and A false, which contradicts the fact that A is a consequence of S . Similarly, a sentence A is model-theoretically valid only if A is epistemically necessary, i.e., A is realized in each world.

7 The suggested demarcation of the logical constants is, I believe, extensionally reasonable: it includes intuitively clear cases and excludes clear noncases. Perhaps the only completely uncontroversial cases are the logical notions of elementary quantification theory. These, it may be verified, are invariant over any collection of worlds that satisfies the constraints described above. Uncontroversial noncases include (but are not limited to) those that violate the Mostowski condition; these are excluded by the suggested criterion, since, as I noted above, it implies the Mostowski condition. The criterion also excludes such oddities as

the quantifier Q above, which, though extensionally invariant, are not rigidly so.¹⁷

However, the present account of logical constanhood is not intended simply as a demarcation. It does more than sort expressions into two classes, logical and nonlogical, in a way consonant with intuition. A demarcation however consonant with intuition need give us no reason for taking the intuitions seriously. The suggested criterion is intended to embody a reductive explanation of what a logical constant *is*. I have suggested that the property of being a logical constant in Karl's language L be identified with a structural property of the extension of the constant over the collection of epistemic alternatives for Karl. One empirical reason for accepting the reduction lies in its ability to explain an entrenched generalization that is the main source of our empirical control over the characterization of the logic of L . It enables us to say that \models constitutes a *kind* of implication (and validity a kind of epistemic necessity), in much the same way as a physical explanation of a law of the form $(\forall x)(Ax \rightarrow Bx)$ in terms of a structural characterization of the A 's enables us to say that the A 's are a kind of B .

It may be objected that, construed in this way, the empirical basis of a model-theoretic characterization of logical relations in a natural language is rather thin. Our empirical control on \models has been said to consist mainly in its being a species of implication, where a map of implication in a natural language is confirmed in the decision-theoretic way alluded to above. There will typically be many *non*logical instances of implication; however, in this respect the situation of model theory is by no means peculiar. For example, several substances may share a "thin" qualitative stereotype. That stereotype may represent all of the empirical information we have about these substances at a certain time. There exist typical instances of each substance, but the qualitative properties that we ascribe to the instances of one are common also to the instances of the others. The structural explanations of the stereotype will differ for the several substances, but if there exists a uniform structural explanation for each substance the stereotype is an adequate empirical basis for the demarcation of several natural kinds, one for each relevant structural characterization. Of course, we do not expect this paucity of information to persist; we expect that the underlying structural characterizations of the several substances will allow us to explain additional properties of their typical instances, properties that distinguish instances of one substance from those of the others. But the original stereotype is enough to get us started.

Quite similarly, there may be several types of implication in Karl's language L , distinguished by several sorts of explanation of why their instances fall into the implication relation. An explanation of why \models is a species of implication made possible by a reductive characterization of the logical constants in terms of the structure of Karl's intentional space will license our saying that \models is a type in an intentional psychology of Karl. The typical instances of this type exemplify additional properties which are explicable in terms of the characterization. Model-theoretic properties that flow from the Mostowski condition are examples, as are certain epistemological traits of the logical constants that have been discussed elsewhere.¹⁸

Above I faulted several syntactic characterizations of logical relations, on

the ground that they seemed incapable of generating a demarcation of the logical constants that is simultaneously empirically motivated and extensionally reasonable. The suggested basis for model-theoretic characterizations fares better. The problem with the syntactic characterizations was that, although they are capable of explaining logical implications, in each case explanations of the same type can be given of *nonlogical* implications. Thus a logical constant cannot simply be said to be an expression for which a semantic rule of the relevant sort exists that plays the required explanatory role; for *too many* expressions would then count as logical. By placing further suitable restrictions on the form of the rules, a demarcation of the constants might be secured that agrees better with intuition, but the restrictions lack empirical motivation. I believe that the suggested framework for model-theoretic characterizations avoids this problem rather nicely. As noted, the identification of logicity with rigid invariance leads to an extensionally reasonable demarcation of the logical constants; and it is precisely the characterizing property of the constants that is needed to explain the fact that instances of model-theoretic entailment are instances of implication. Since making this connection was our leading empirical constraint, the suggested demarcation of the logical constants has an obvious empirical motivation.

Let us review. I have suggested that the logical expressions of an extensional language L be characterized in terms of a structural property of their extensions in all epistemic alternatives for its speaker (or population). The suggested characterization was seen to fix the extension of such an expression over each actual domain, and thus to determine the model-theoretic characterization of the consequence relation for L . I adduced a result which shows that the consequence relation thus generated is a subspecies of the implication relation for L . The characterization thus serves to explicate our main empirical constraint on the notions of logical truth and consequence, and to reduce the manifestation problem for the logic of a natural language to the manifestation problem for a theory of the intentional states of its (idealized) speakers. All of this suggests that *logical* necessity, our generic name for whatever sort of necessity is exemplified by the logical truths, does not, as is sometimes supposed, constitute a fully autonomous necessity concept. Rather, the logical necessity of a sentence consists in the fact that its epistemic necessity has a special explanation.

NOTES

1. There are, of course, different ways of making this more precise; but it is difficult to go beyond this rather vague formulation without begging substantive questions. Below we shall distinguish several explications of the relevant guarantee.
2. These present assumptions are made for simplicity's sake: but the results of the paper extend to more sophisticated settings, for example, to intensional languages characterized along the lines of Montague's [9].
3. It should also be noted that a syntactic characterization can be compatible with a model-theoretic one. What seems essential to the compatibility relation in this case is that the assignment of extensions to the logical constants given by the model-theoretic characterization satisfy the syntactic one, and that the two theories determine the same consequence relation for the language in question.

4. See [5], pp. 21, 36. Davidson's constraints are by now familiar, but were first explicitly formulated in [1].
5. This distinction is not explicit in [5], but was clearly drawn by Lycan in our Cincinnati exchange. What follows is much indebted to Lycan's comments on my [8].
6. See [5], Chapter 10, for further discussion.
7. Here ' a ' ranges over objects in the ontology of L and ' F ' over 1-place predicates of L . Strictly speaking, our assumptions require a rule in which ' F ' ranges over arbitrary formulas and ' a ' over sequences. In that case we could say, somewhat less naturally, that a sequence satisfies a formula RF when it satisfies F and the objects in the sequence that correspond to the free variables in F are red.
8. Which says that the formulas mentioned in an operational rule for a constant should be instances of subformulas of the *principal* formula, the formula in the rule involving the constant. The suggested rules for R run afoul of this constraint, involving, as they do, ineliminable reference to the formula $RED(x)$. However, it is worth noting that adding the subformula constraint will not suffice to avoid all counterexamples of this type. Let c be a fixed individual constant of L . Consider the following rules for the quantifier E_c :

$$\frac{J, F(c) \vdash K}{J, E_c x F(x) \vdash K} \quad \frac{J \vdash F(c), K}{J \vdash E_c x F(x), K}.$$

These rules satisfy the subformula constraint and characterize E_c in the same sense as the usual operational rules characterize \exists . But E_c is not plausibly a logical constant; for it functions simply as the restriction of the existential quantifier to the individual concept of the denotation of c .

9. The terminology 'epistemically possible situation' is due to Kripke [4], who sharply distinguishes it from the notion of a *metaphysically possible world*. I shall do likewise. For example, 'water is H_2O ' fails in some *epistemically* possible situations. Perhaps the continuum hypothesis does also, even if it is true.
10. For the purposes of the example, we need only suppose that F and G have the *same* extension over each actual domain.
11. By a decision-theoretic explanation in this context I mean an explanation of a piece of Karl's behavior that renders that behavior intelligible in terms of Karl's beliefs and desires. Of course, the notions of belief and desire which arise in this way are idealized, and an actual agent can be expected to instantiate them only approximately. In a more refined treatment, I would take Karl's beliefs to be described by an assignment of epistemic probabilities to sentences generated by an idealized inductive method (which again would be instantiated only approximately by an actual agent), and his desires to be described by a coherent assignment of utilities. Karl's intentional space would then be described in terms of the relevant notion of epistemic probability; roughly, a set would qualify as epistemically possible if it is assigned nonzero probability relative to some evidential situation.
12. I defended an analogue of Mostowski's condition as a constraint on the logical constants in Section III of [6], and suggested that a sort of amalgam of that constraint with a criterion proposed by Peacocke in [11] might serve as a demarcation of logic. The present proposal derives from these observations in conjunction with a technical result of [7]. One form of the present proposal was sketched very briefly in Section 6 of [7].

13. A bit more explicitly, we assume that there is an operator *neg* (negation) such that any sentence of the form *neg*(*A*) appears in T_1 iff *A* appears in T_2 , and appears in T_2 iff *A* appears in T_1 , an operator *dj* (disjunction) such that a sentence of the form *dj*(*A*, *B*) appears in T_1 iff either *A* or *B* appear in T_1 and appears in T_2 iff both *A* and *B* appear in T_2 , and so on. For the universal quantifier we require a variable binding operator *Uquant* that applies to a one-variable formula *A*(*x*) and variable *x* to yield a sentence that appears in T_1 iff *A*(*c*/*x*) appears in T_1 for each constant $c \in K$, and in T_2 iff *A*(*c*/*x*) appears in T_2 for some $c \in K$, and analogously for existential quantification. For identity we require a two-place predicate *Id* such that (i) *Id*(*a*, *a*) is in T_1 for each $a \in K$ and *Id*(*a*, *b*) appears in T_2 for each pair of distinct constants *a*, *b* from K , (ii) if *A* is a predicate and *b*, *c* constants such that *A*(*b*) and *Id*(*b*, *c*) appear in $T_1(T_2)$, then *A*(*c*) appears in $T_1(T_2)$, and (iii) for each closed term *t*, *Id*(*t*, *c*) appears in T_1 for a (unique) constant *c* from K .
14. Here *D* is an arbitrary nonempty set and *E* a set of sequences over *D*, an extension of the sort appropriate for formulas over *D*. I suppress reference to the extensions of variables, which are determined uniformly from *D*.
15. To see that this procedure need not yield a well-defined extension for *Q* at (*D*, *E*) in the absence of the invariance constraint, let *Qx* represent the nonlogical quantifier phrase "for some red *x* . . .". In this case one could have two interpretations, I_1 and I_2 , which realize (*D*, *E*) in worlds w_1 and w_2 and which assign *E* to a one-variable condition *F* in *x* such that (*Qx*)*F* is true at w_1 but false at w_2 . This will occur if w_1 is a world in which there exist red satisfiers of *F*, and w_2 a world in which the same objects satisfy *F*, but in w_2 none of these objects are red.
16. For a more detailed proof of a related but somewhat more general result, see [7].
17. On the other hand, it is worth noting that if the possible situations populating Karl's intentional space satisfy some rather weak closure conditions, the quantifier "for infinitely many *x*" is rigidly invariant (see [7], p. 440). Thus the suggested criterion does not generally classify all nonelementary quantifiers as nonlogical.
18. For example, the criterion of logicity suggested by Peacocke in [11] can be naturally formulated within the present framework. Thus formulated, it is a consequence of the suggested characterization of the logical constants. In [6] I argued that Peacocke's criterion does not provide a sufficient condition for an expression to be a logical constant, but it is plausibly a necessary one.

REFERENCES

- [1] Davidson, D., "Truth and meaning," *Synthese*, vol. 17 (1967), pp. 304–323.
- [2] Gentzen, G., "Investigations into logical deduction," translated by M. E. Szabo, pp. 68–131 in *The Collected Papers of Gerhard Gentzen*, North Holland, Amsterdam, 1969.
- [3] Hacking, I., "What is logic?," *Journal of Philosophy*, vol. 76 (1979), pp. 285–318.
- [4] Kripke, S., *Naming and Necessity*, Harvard University Press, Cambridge, 1980.
- [5] Lycan, W., *Logical Form in Natural Language*, Bradford, Cambridge, 1984.
- [6] McCarthy, T., "The idea of a logical constant," *Journal of Philosophy*, vol. 78 (1981), pp. 499–523.

- [7] McCarthy, T., "Modality, invariance, and logical truth," *Journal of Philosophical Logic*, vol. 16 (1987), pp. 423-443.
- [8] McCarthy, T., "Lycan's neo-Davidsonianism and logical theory," Symposium Paper, American Philosophical Association, Central Division, April 1988.
- [9] Montague, R., "Pragmatics and intensional logic," *Synthese*, vol. 20 (1970), pp. 68-94.
- [10] Mostowski, A., "On a generalization of quantifiers," *Fundamenta Mathematicae*, vol. 44 (1957), pp. 12-36.
- [11] Peacocke, C., "What is a logical constant?," *Journal of Philosophy*, vol. 73 (1976), pp. 221-240.

University of Illinois
Department of Philosophy
Urbana, Illinois 61801