GILBERT HARMAN

IF AND MODUS PONENS

*A Study of the Relations between Grammar and Logical Form**

Modus ponens is the principle that *P* and *if P, then Q* imply *Q*. I want to argue that modus ponens is not a principle of logic.

I agree that *P* and *if P, then Q* do imply *Q*. This is so, at any rate, if *P* and *Q* are indicative statements, with certain exceptions that I will mention. Nevertheless, I do not agree that *P* and *if P, then Q* logically imply *Q*.

I want to say that an implication is a logical implication if it holds solely by virtue of logical form. When I say that *P* and *if P, then Q* do not logically imply *Q*, it is because of my view about the logical form of the conditional *if P, then Q*.

Compare *if P, then Q* with *that P implies Q*. *P* and *that P implies that Q* imply *Q*. But I do not want to say that this is a logical implication, since I hold that it depends not only on logical form but also on the meaning of the word *implies*. If, for example, the word *suggests* replaces the word *implies*, the implication does not always hold. *P* and *that P suggests that Q* can both be true even though *Q* is not true. Therefore, even though *P* and *that P implies that Q* always imply *Q*, I want to say that the implication does not hold solely by virtue of logical form and is, consequently, not a logical implication.

In saying this, I am assuming that the word *implies* is not a logical particle. For consider this. *P and Q* logically implies *Q*. This implication depends only on logical form. But it obviously depends on the meaning of the word *and*. If, for example, the word *or* were to replace the word *and*, the corresponding implication would not always hold. *P or Q* can be true even though *Q* is not true. But I do not take this to show that the implication from *P and Q* to *Q* is not a logical implication. For I assume that *and* is a logical particle. If I say that an implication holds solely by virtue of logical form, I mean that it does not depend on the meanings of any terms other than logical particles.

I want to say that the difference between logical particles and nonlogical terms is this: logical particles are members of small closed classes whereas

Theory and Decision **11** (1979) 41–53. 0040–5833/79/0111–0041$01.30

nonlogical terms are members of large open classes.[1] However, I do not want to say that a word is a logical particle if and only if it belongs to a small closed grammatical class of terms. For, the class of English prepositions is a relatively small closed class and I do not want to say that the preposition *between* is a logical particle. I want to say that *between* represents a nonlogical relational predicate. Similarly, the class of English modal auxiliary verbs is small and closed. Its members are *may*, *might*, *can*, *could*, *will*, *would*, *shall*, *should*, and *must*. But I do not want to say that any of these words are logical particles.

There is a connection between grammar and logic, but it is important not to oversimplify it. Logical categories are related to grammatical categories, but the relation is indirect and grammar makes many distinctions that are irrelevant to logic. For example, logic sees predicates where grammar distinguishes nouns, verbs, adjectives, and prepositions.

Some linguists have suggested that this difference between grammar and logic is superficial, a matter of "surface structure." These linguists say that a good transformational grammar will show how nouns, verbs, and adjectives derive from underlying predicates. But this suggestion itself derives from a misunderstanding of the relation between grammar and logic.

One way to appreciate the indirectness of the relation is to note that the logician's category "predicate" covers complex as well as simple predicates. A complex predicate will ordinarily fail to correspond to a linguistic expression of a single grammatical category. Consider the following principle of logic: $(\exists x)(y)Rxy$ logically implies $(y)(\exists x)Rxy$. Here R is to be any relational predicate at all. Such a predicate might correspond to a linguistic expression of one or another grammatical category, say, the verb *loves*. *Someone loves everyone* logically implies *everyone is loved by someone*. But R can also be a complex predicate. Instead of *x loves y* we might have *x steals y's money*. *Someone steals everyone's money* logically implies *everyone's money is stolen by someone*. But there is no phrase of a single grammatical category that corresponds to the relational predicate *x steals y's money* in the way that the verb *loves* corresponds to the relational predicate *x loves y*.

In other words, the category "predicate" is a logical category for which there is in general no corresponding grammatical category. Similarly, although verbs correspond to simple or atomic predicates, the category "verb" is a grammatical rather than a logical category.

A predicate can be thought of as a sentence with a hole or holes in it. The verb *loves* is not, strictly speaking, a relational predicate. It corresponds to a relational predicate which we can represent as *x loves y* or better as . . . *loves* - - - or even better as 1 *loves* 2 (I number the holes since I want clearly to distinguish the relational predicate 1 *loves* 2 from the converse relational predicate 2 *loves* 1, which is equivalent to 1 *is loved by* 2).[2]

A predicate can also be thought of as a function which maps a sequence of names into the sentence that results when the holes in the predicate are plugged with the respective names. For example, the predicate 2 *loves* 1 is a function that maps the sequence consisting of the name *John* followed by the name *Mabel* into the sentence *Mabel loves John*. The predicate 1 *steals* 2*'s money* is a function that maps the same sequence into *John steals Mabel's money*.

One connection between grammatical categories and the logical category "predicate" is that grammatical categories are relevant to the construction of grammatical sentences. Grammar refers to these categories in telling us what the grammatical sentences are and what names are. Given that we know what sentences and names are, we can then define predicates, in the logician's sense, as sentences with holes where names could go, or, equivalently, as the related functions from sequences of names to sentences containing those names.

Another connection between grammatical categories and the logical category "predicate" is that nouns, verbs, adjectives, and certain prepositions correspond to atomic predicates. The verb *loves* corresponds to the atomic predicate 1 *loves* 2. The noun *friend* corresponds to the atomic predicate 1 *is a friend of* 2. The adjective *fond* corresponds to the atomic predicate 1 *is fond of* 2. The proposition *with* corresponds to the atomic predicate 1 *is with* 2. (I am ignoring tense and time. Let me continue to do so.)

What I have been saying about predicates also applies to other logical categories. For example, sentential connectives are not, strictly speaking, words like *and* and *or*. They are, rather, compound sentences with holes to be filled with sentences. Logical conjunction might be represented as . . . *and* - - - or better as A *and* B; logical disjunction as . . . *or* - - - or better as A *or* B. There are many other connectives, for example, *that* A *implies that* B.

We can also think of sentential connectives as functions from ordered sequences of sentences to the sentences that result when the holes in the connectives are filled with the relevant sentences in the sequence. For example, the connective *A and B* is a function which can be applied to the sequence consisting of the sentence *John loves Mabel* followed by the sentence *Mabel loves John*, yielding as a result *John loves Mabel and Mabel loves John*. If we apply to this sequence the different function, *that A implies that B*, the result is the sentence *that John loves Mabel implies that Mabel loves John*.

Just as predicates do not in general correspond to expressions of some single grammatical category, similarly, sentential connectives do not in general correspond to expressions of some single grammatical category. On the other hand, just as nouns, verbs, adjectives, and certain prepositions correspond to atomic predicates, similarly the words *and* and *or* correspond to atomic sentential connectives.

The atomic operator *that* ◯ is a complex name with a hole to be plugged by a sentence. In other words, it is a function that applies to a sentence to form a complex name. For example, it applies to the sentence *John loves Mabel* to yield the name *that John loves Mabel*. This might be the name of something that is said to imply something else, as in the sentence *that John loves Mabel implies that Mabel loves John*. Or it might be the name of something said to be believed, as in *Sue believes that John loves Mabel*.

The operator *that* ◯ is atomic. An example of a complex non-atomic operator of this sort is *the belief that* ◯ *and snow is white*. The latter operator converts the sentence *grass is green* into the name *the belief that grass is green and snow is white*.

Other atomic name forming operators of roughly this sort are *for* ◯, *-ing* ◯, and *WH-*◯. The operator *for* ◯ applies to a sentence in the infinitive. (In other words it applies to a sentence whose main verb does not contain tense.) It converts such a sentence into a name of the sort of thing toward which one can have a positive or negative attitude. For example, it converts *George come to the party* into *for George to come to the party* which can name what someone might want, as in *I want very much for George to come to the party*.

Often the words *that* and *for* are omitted. We say *Sue believes John loves Mabel* or *I want George to come to the party*. I assume that this is just a matter of "surface structure" and that the complementizers *that* and *for* are present at some stage of the grammatical derivation of these sentences.[3]

The operator *-ing* ◯ also applies to tenseless or infinitive sentences to yield a name. Applied to *John love Mabel* it yields the name *John's loving Mabel*, as in *John's loving Mabel saddened Sue*.

It is less clear what the operator *WH-*$\bigcirc$ applies to. The result of applying the operator is an "indirect question" in surface structure. *Albert does not know who loves John. Sue wonders whether John loves Mabel. Tom asked Albert how much Sue loved Herbert*. One might suppose that in some cases this operator applies to a predicate, so that one might apply *WH-*$\bigcirc$ to $\bigcirc$ *loves John* to get *who loves John*, a complex name of something that Albert is said not to know.

However, that idea is not easily applied to *whether* clauses. *Whether* is sometimes said to be *WH* plus *either*. Thus we say *whether or not*. One might suggest that *whether or not John loves Mabel* is the result of applying *WH-*$\bigcirc$ to the disjunctive sentence *either John loves Mabel or not*. But that is rather different from what was suggested for other kinds of indirect questions. So, in what follows, I will simply leave open exactly what sort of operator *WH-*$\bigcirc$ is.

The logical categories are "name", "sentence", and various categories of schemata representing functions that map items of one logical category into terms of another (possibly the same) category. Predicates map names into sentences. Certain name forming operators convert sentences into names. Functional operators, which I have not discussed, map names into names. For example, *father of* $\bigcirc$ maps the name *Oscar* into the name *father of Oscar*, and so on. For example quantifiers map predicates into sentences. There are in this way an infinite number of logical categories.

Now, except for the categories "name" and "sentence", the logical categories are not the same as any grammatical categories. However, it is often possible to associate terms of various grammatical categories with atomic members of certain logical categories. Verbs, nouns, adjectives, and prepositions represent atomic predicates. Complementizers like *that*, *for*, and *-ing* represent atomic name forming operators that apply to sentences. Coordinating conjunctions, like *and* and *or*, represent sentential connectives, and so on, and so forth. In this way, many words can be associated with logical categories. Notice that words of different grammatical categories are associated with the same logical categories. For example, nouns and verbs are both associated with predicates.[4]

For convenience in what follows, I will speak loosely as if nouns and verbs *are* atomic predicates, although, strictly speaking, they are, as I have said, only *associated with* atomic predicates.

I can now say more precisely what I take the difference between logical particles and nonlogical terms to be. Recall that I want to say that logical particles are words that are members of small closed classes and that nonlogical terms are members of large open classes. The relevant classes are not syntactic classes; they are logical classes. More exactly, they are classes of atomic predicates, the class of atomic sentential connectives, the class of atomic name forming operators that apply to sentences, and so forth. The word *and* is a logical particle because it is an atomic sentential connective and the class of atomic sentential connectives in English is small and closed in the sense that it significantly changes the language to add a new atomic sentential connective in a way in which it does not significantly change the language to add a new atomic predicate. The word *implies* is a nonlogical term because it is an atomic predicate and the class of atomic predicates in English is large and open.

I have had to go into all this because I want to deny that *modus ponens* is a principle of logic. I agree that *P* and *if P, then Q* imply *Q* but I deny that they logically imply *Q*. I want to say that an implication is logical only if it does not depend on the meanings of nonlogical terms. What I have been explaining is how I distinguish logical particles from nonlogical terms.

There is an element of circularity here, since my reasons for distinguishing logical particles from nonlogical terms in this way is that this way of marking the distinction yields results that correspond with my antecedent intuitions. I was prepared ahead of time to say that *and* is a logical particle and that *loves* and *implies* are nonlogical terms. My way of distinguishing logical particles from nonlogical terms was designed to yield such results. (I do not know of any other plausible way to get such results.)

Supposing that this is the way in which logical particles are to be distinguished from nonlogical terms, we might speculate as to why it is the way to do so. However, I will not speculate on this occasion.[5]

Returning, then, to my principal thesis, I want to argue that, just as *P* and *that P implies Q* do not (according to me) logically imply *Q*, similarly *P* and *if P, then Q* do not logically imply *Q*. A natural objection, given what I have been saying, is that the latter implication depends only on the meaning of *if-then* and that *if-then* unlike *implies* is a logical particle rather than a non-logical term. If that were true, the implication from *P* and *if P, then Q* to *Q* would hold solely as a matter of logical form and would, on my account, be a logical implication.

In order to meet this objection, I must say what the logical status of *if* is. (In what follows I will not try to distinguish *if* from *if-then*. I believe that the word *then* has mainly a stylistic function.) I will begin by mentioning a couple of theories about *if* that I do not want to accept and then will say what account I do accept.

A first, rather implausible theory would be that *if* or *if-then* logically represents a predicate, like *implies*, the predicate *if* 1 *then* 2. The theory would be that, when this predicate is applied to the names *that P* and *that Q*, say, the result is *if that P, then that Q*, which is grammatically realized as *if P, then Q*. The theory would also have to invoke a special principle of grammar to explain why the alleged predicate *if* 1 *then* 2 cannot be applied to the demonstratives *this* and *that* to yield the nonsentence *if this, then that*.

What makes the theory implausible is just this ad hoc appeal to grammar. Such a theory would complicate any account of the relation between grammar and logic.

If the theory were right, though, I would be right about modus ponens. Modus ponens would not be a logical principle, since it would depend not only on logical form but also on the meaning of the nonlogical predicate *if* 1 *then* 2. But I do not want to defend my thesis about modus ponens in this way, since I do not want to accept this first, rather implausible theory about *if-then*.

However, I do want to say that something like this first theory is true. In particular, I want to say that the logical form of *if P, then Q* is like the logical form of *that P implies that Q*. But I do not want to say that *if* or *if-then* functions as the word *implies* does, as a predicate. Instead I want to say that *if* functions as the word *that* does in *that P implies that Q*. Furthermore, I want to say that there is a hidden relational predicate in *if P, then Q*. It is this hidden unexpressed predicate that, I want to say, functions as *implies* functions in *that P implies that Q*.

In other words, I agree that the logical form of *if P, then Q* is just like the logical form of *that P implies that Q*. And I agree that an element of this logical form is unexpressed at the "surface" in *if P, then Q*. But I disagree with the first theory concerning what the unexpressed element is. According to the first, implausible theory, the hidden unexpressed element in *if P, then Q* is a word like *that*, representing an atomic operator that converts a sentence into a name. My view is that the word *if* represents just such an operator and that the hidden unexpressed element therefore represents the main predicate of the sentence, a predicate similar to that represented by *implies*.

Before elaborating my own view, I must mention an obvious alternative. One might suppose that there are no hidden unexpressed elements in *if P, then Q*. In other words, one might suppose that *if* or *if-then* represents an atomic sentential connective *if A, then B*, so that the logical form of *if P, then Q* was similar to the logical form of *P and Q*. In that case, one would have to conclude that modus ponens is a principle of logic. For one would have to say that the implication from *P* and *if P, then Q* to *Q* depended only on logical form, since one would have to say that it depended only on the meaning of *if-then*, a logical constant. For, given that the class of atomic sentential connectives is small and relatively closed (as I have been assuming) and given that my criterion for distinguishing logical particles from nonlogical terms is accepted, it would follow that *if-then* is a logical particle and, therefore, that the implication indicated by the principle of modus ponens holds solely by virtue of its logical form.

However, there are sentences in which *if* clearly does not function as an atomic sentential connective. For example, *Albert wondered if Mabel loved John. Mabel asked if John was going to the party. John does not know if Mabel loves him.* Here *if* has a meaning somewhere between the meaning of *whether* and the meaning of *that*. Logically it resembles the complementizer *that*, since it represents a name forming operator *if* ◯ which combines with a sentence to form a name of something like a proposition – but more "iffy."

A grammarian counts such *if* clauses as indirect questions. Observe that we have not only *Mabel asked if John was going to the party* but also *Mabel asked who John was going to the party with*, *Mabel asked where the party was*, *Mabel asked how John was getting to the party*, *Mabel asked which party John was going to*, and so forth.

One might suppose that in this use of *if* it is simply a variant of *whether*. But that is not quite right. We say *Mabel asked whether or not John was going to the party* but not *Mabel asked if or not John was going to the party*. We say *I doubt if John is going* but not *I doubt whether John is going*.

Still, one might suppose that the *if* that can appear in indirect questions cannot be the same as the *if* in *if P, then Q*. So it is important to notice that variants of indirect questions often appear as the antecedents of conditionals. *Whether or not John goes to the party, Mabel will go. Whoever John would have taken to the party, Mabel would have been unhappy. Whichever party John had gone to, Mabel would have come to ours.* So there is reason to believe that the *if* in the conditional *if P, then Q* is the same *if* that appears in *Mabel asked if P.* Since that *if* is not a sentential connective, there is reason to believe that *if-then* is not at atomic sentential connective.

Chomsky and Lasnik "regard *if* as a variant of *that* before subjunctives" because of the contrast between "I was surprised that they came" and "I would be surprised if they came", etc. and because this assumption simplifies certain principles of grammar.[6] *If* also seems to be a variant of *that* in statements of conditional probability. *If John is ugly, Mabel probably won't like him.* Modus ponens does not seem to hold for such statements. We can have *if John is ugly, Mabel probably won't like him* but also *if John is ugly and rich, Mabel probably will like him.* It can be true that, if John is ugly, Mabel probably won't like him, and true that John is ugly, without its being true that Mabel probably won't like him.

I want to say that the statement *if John is ugly, Mabel probably won't like him* means that Mabel's not liking John is probable in relation to the consideration that John is ugly. Furthermore, I want to say that *Mabel probably won't like John* means that Mabel's not liking John is probable in relation to considerations that are supposed to be indicated by context; usually these are "all things considered".

Furthermore, I want to say that *if John is ugly, Mabel probably won't like him* is derived, perhaps by rules of transformational grammar, from *that Mabel won't like John is probable if John is ugly*. This indicates what I take to be its logical form: *that P is probable if Q*. As I see it, *probable* or *probably* represents an atomic relational predicate *1 is probable 2*, which relates a proposition to an "iffy" proposition (a condition). *That Mabel won't like John* is the name of a proposition formed by applying *that* ◯ to the sentence *Mabel won't like John*. I say that *if John is ugly* is the name of an "iffy" proposition or condition formed by applying *if* ◯ to the sentence *John is ugly*.

So, I want to say that the *if* clause in *if John is ugly, Mabel probably will* *not like him* is of the same sort as the *if* clause in an indirect question like *Mabel asked if John was going to the party*.

These considerations do not absolutely refute the idea that *if* is a sentential connective in *if John is ugly, Mabel probably will not like him*. For a defender of the sentential connective theory might argue that such a statement is really about the probability of a conditional, perhaps deriving from something like *it is probable that if P, then Q*, i.e. *that (if P, then Q) is probable*. Modus ponens would not apply to statements of conditional probability statements since these would not be conditional statements at all but rather statements about conditionals.

This defense would seem to require the assumption that the probability of a conditional is the same as the corresponding conditional probability, an assumption David Lewis has shown to be untenable.[7] Furthermore, as I have indicated, a more unified account of *if* is also possible which treats *if* always as a name forming operator.

Returning to my main argument, I want to analyze statements of conditional possibility, conditional necessity, conditional certainty, and so forth, in exactly the same way as I have analyzed statements of conditional probability. For example, consider the conditional possibility statement *if John is rich, possibly Mabel will like him*. Modus ponens fails, since this statement is compatible with *if John is rich and also ugly, it is not possible that Mabel will like him*. I want to say that the first statement derives from *that Mabel will like John is possible if John is rich*. I say that *possible* represents an atomic relational predicate *1 is possible 2*, which relates a possibility named by a *that* clause to a condition named by the *if* clause. Similarly for sentences whose superficial form is *if P, necessarily Q; if P, certainly Q;* and so on.

Next, consider the statement *if John is rich, Mabel may like him*. This can be interpreted to mean roughly the same thing as the statement *if John is rich, possibly Mabel will like him*. I want to say that, on this interpretation, it derives from something like *that Mabel likes him may be, if John is rich*. Notice, for instance, that we can say, perhaps, *it may be that Mabel likes him, if John is rich*. I want to say that *may* here represents an atomic relational predicate *1 may be 2*, which relates a possibility named by a *that* clause to a condition named by an *if* clause.

I am speaking inexactly when I say that the possibility here is named by a *that* clause. What I am really supposing is *that Mabel may like John* is related to *it may be that Mabel likes John* in the way in which *Mabel seems to like John* is related to *it seems that Mabel likes John*. Both *Mabel may like John* and *Mabel seems to like John* probably derive from sources in which *Mabel likes John* is in the infinitive. However, I will ignore this complication in what follows.

Next, consider *if John is rich, Mabel should like him*. This can be interpreted to mean roughly the same thing as *if John is rich, Mabel will probably like him*. I want to say that it derives from something like *that Mabel likes him should be, if John is rich*. I want to say that *should* represents an atomic relational predicate *1 should be 2* which relates a proposition to a condition named by an *if* clause.

Other interpretations of conditional *should* and conditional *may* statements are possible, e.g. interpretations according to which these statements say what morally should or may be done. These interpretations raise problems of logical analysis that I cannot go into here. However, I would argue that in every case, *should* or *may* represents an atomic relational predicate and that one of the things it relates is a condition named by an *if* clause.[8]

I want to give exactly the same sort of analysis to subjunctive conditions like *if John were sick, Mabel would like him*. I want to say that this derives from something like *it would be that Mabel liked him, if John were sick*. I want to say that *would* represents a relational predicate *1 would be 2* which relates a proposition to a condition named by the *if* clause.

Finally, consider simple indicative conditionals containing no overt relational predicates of the sort I have been discussing, for example, *if John went to the party, Mabel stayed home*. I want to suppose that such a statement contains a hidden unexpressed relational predicate which I will call *IMP*. In other words, I want to suppose that this derives from something like *that Mabel stayed home is IMP if John went to the party*. I assume that *IMP* represents a relational predicate that relates a proposition to a condition named by the *if* clause.

This account of ordinary simple indicative conditionals is to some extent forced and artificial, but it also has certain advantages. In particular it allows me to suppose that the *if* in conditionals is the same as the *if* in indirect questions.

The alternative is to suppose that there are two *if's*, a name forming operator in indirect questions and a sentential connective in conditionals. If this alternative were accepted, one would have to suppose that it is a coincidence that indirect questions can appear in conditionals as in *whether or not John goes to the party, Mabel will go* and *whoever John would have taken to the party, Mabel would have been unhappy*. On the other hand, this is just what one would expect given my analysis of conditionals.

When I weigh the artificiality involved in the one theory, because of its assumption that there is a hidden unexpressed predicate in certain indicative conditionals, against the facts that the alternative does not give a unified account of *if* and does not account for this connection between indirect question clauses and conditionals, I conclude that, on balance, the unified theory is better.

So, I conclude that there is no sentential connective *if* or *if-then* and, therefore, that there is no principle of logic like modus ponens. To the extent that this principle holds for indicative conditionals, it holds by virtue of the meaning of an unexpressed predicate *IMP*. *IMP* is not a logical particle, since it represents an atomic predicate and is therefore a nonlogical term. Modus ponens, where it holds, does not hold solely by virtue of logical form.

NOTES

* I am indebted to Wayne Davis for helpful comments.

[1] Cf. W. V. Quine, *Philosophy of Logic* (Englewood Cliffs, New Jersey: Prentice Hall, 1970), pp. 28–29.

[2] See W. V. Quine, *Elementary Logic*, revised edition (New York: Harper, 1965), pp. 99–101, and *Methods of Logic*, third edition (New York: Holt, 1972), p. 145.

[3] See Joan Bresnan, 'On Complementizers: Toward a Syntactic Theory of Complement Types', *Foundations of Language* 6 (1970): 297–321 and Noam Chomsky and Howard Lasnik, 'Filters and Control', *Linguistic Inquiry* 8 (1977): 425–504.

[4] We might tentatively define 'atomic' and 'corresponds' as follows: x is an atomic element of logical category C if and only if (1) x is an element of category C and (2) there are not elements f and y of any logical categories such that $x = f(y)$; and expression e corresponds to atomic element x of logical category C if and only if (1) x is an atomic element of C, (2) e is of some single grammatical category, (3) e is part of x, and (4) e is not part of any other atomic element of any logical category. (I assume here that in deep structure different senses of the same expression are distinguished, e.g. by subscripts.)

[5] I have speculated about this, e.g., in 'The Logic of Ordinary Language', in Parviz Morewedge (ed.), *Philosophies of Existence* (to appear).

[6] Chomsky and Lasnik, *op. cit*., p. 53.
[7] David Lewis, 'Probabilities of Conditionals', *Philosophical Review* **85** (1976), pp. 297–315.
[8] See Gilbert Harman, *The Nature of Morality* (New York: Oxford, 1977), pp. 115–124.