

Logical Consequence and Logical Expressions*

Mario GÓMEZ-TORRENTE

ABSTRACT: The pretheoretical notions of logical consequence and of a logical expression are linked in vague and complex ways to modal and pragmatic intuitions. I offer an introduction to the difficulties that these intuitions create when one attempts to give precise characterizations of those notions. Special attention is given to Tarski's theories of logical consequence and logical constancy. I note that the Tarskian theory of logical consequence has fared better in the face of the difficulties than the Tarskian theory of logical constancy. Other theories of these notions are explained and criticized.

Keywords: logical consequence, logical constant, Tarski, validity, necessity

1. Modality and form

The intuitive relations between the notion of logical consequence and those of form and modality are the basic guide when one reflects philosophically on the former notion¹. There are two traits of the notion of logical consequence based on these intuitive relations which often allow us to distinguish the logically correct arguments from the incorrect or the merely correct. One of these traits, based on the relation with modality, is this: if a conclusion follows logically from some premises then *that conclusion follows by logical necessity from those premises*. It's not quite clear what it is that makes a conclusion follow by logical necessity from some premises. A good deal of the unclarity surfaces in the vagueness of the notion of consequence by logical necessity: some arguments are clear instances of it, some arguments are clear non-instances, but many arguments which are correct in some sense appear to lie in a dimly illuminated borderline zone. Much of this vagueness appears to be due to the fact that that notion is itself associated (in ways which are themselves unclear and vague) to unclear and vague notions like analytical implication, *a priori* implication and implication by necessity (*tout court*). These notions transfer their vaguenesses (and their obscurities in general) to the notion of consequence by logical necessity, and conspire to make very fuzzy the line dividing the instances from the non-instances.

As in all cases of vagueness, some examples lie far from the line, at opposite extremes. It is reasonable to think that the argument with premise 'The ravens I have seen till today are all black' and conclusion 'The ravens I will see tomorrow will be black' is in some sense correct, but intuitively its conclusion doesn't follow by logical necessity from its premise (and so it's not logically correct). On the other hand, for example, there is a fairly extended intuition that from the premise 'Some widows are

* This article responds to a request from Andoni Ibarra, editor of *Theoria*, who had originally thought it would be useful to publish a summary in English of my book *Forma y Modalidad* (Gómez-Torrente (2000a)). Later the idea arose of making my piece more useful, by turning it into an item to which some distinguished specialists on the general topic of logical consequence and logical expressions could react in a forum provided by *Theoria*. I thank Andoni Ibarra for his interest and encouragement, and the other participants in this forum for their collaboration. I also wish to express my gratitude to Josep Macià for his help.

¹ And they give its title to my book Gómez-Torrente (2000a).

merry' follows by logical necessity the conclusion 'Some women are merry'. Note that these examples also lie far from the lines that divide the instances from the non-instances of analytical implication, *a priori* implication and necessary implication (and at opposite extremes).

But the argument with premise 'Some widows are merry' and conclusion 'Some women are merry' is not logically correct, as is indicated by the second trait I referred to, which mentions the notion of logical form: if an argument is logically correct, then *every argument with the same logical form is logically correct*. It's common to suppose that logical form is a certain schematic form that results approximately from replacing the *non-logical expressions* of the argument by letters without a specific meaning, in a revealing and uniform way (with different expressions replaced by different letters but with the same expressions replaced always by the same letters)². In the merry widows example, the logical form would look somewhat like this:

Some *F* are *G*

Some *H* are *G*

Notice that the argument with premise 'Some widows are merry' and conclusion 'Some sofas are merry' has this same logical form but is intuitively logically incorrect. Notice too that the fundamental notion in the formulation of the second trait of the notion of logical consequence is the notion of a logical expression. This notion is again considerably vague, and this vagueness conspires again with the other vaguenesses to make the notion of logical consequence a very vague one indeed.

The two intuitive traits of the notion of logical consequence are the touchstone for the attempts at philosophical clarification of this notion. Some important attempts are based on the search and proposal of characterizations of logical consequence for arguments of formal languages which mirror fragments of natural language, characterizations given in terms of notions which are better understood (or with which one can at least do fruitful intellectual work). Much of the recent philosophical work on the topic (including my own) has concentrated on examining the adequacy and clarifying power of some classical characterizations. There are two main kinds of classical characterizations: one based on the notion of derivability and one based on the notion of validity.

2. *Derivability and validity*

With his use of the notion of derivability, Frege approached logical consequence in a way analogous, in essence, to the way employed by Aristotle and the Stoics and Megarians. Frege invented a formal language (that we today often decompose in a series of languages), designed especially for the formalization of mathematical argu-

² I don't make an attempt to be more precise regarding the replacing method. By 'the logical form' of an argument I understand intuitively the schema which reveals as much as possible about its logical structure (in the sense in which a quantificational schema reveals more than a propositional schema of the same argument).

ments, and conjectured that at least every logically correct mathematical argument could be formalized by means of an argument in his language which would be logically correct in a certain technical sense³. The language invented by Frege was what we would now call a higher-order quantificational language, containing a first-order fragment. Frege also increased tremendously the rigor of the presentation of an inferential system (with respect to previous attempts), to the point that he can be considered the creator of formal systems.

Using the notion of derivability in a system like Frege's, it is possible to propose a very precise characterization of the set of logically correct formalized arguments of the system's language. One can propose that a conclusion C logically follows from some premises P_1, P_2, P_3 , etc. exactly when there is a series of applications of the inference rules which, starting from P_1, P_2, P_3 , etc., and possibly from axioms, ends in C (when such a series exists one says that C is *derivable* from P_1, P_2, P_3 , etc.). As I will stress later, if the formal system is built with care, when one is done one will be able to convince oneself that the derivability of a conclusion from some premises is a sufficient condition for the corresponding argument to be logically correct. The question whether one can convince oneself that it is also a necessary condition will also be examined later.

In logic there has been also, from Aristotle himself, an alternative (and complementary) kind of approach to the characterization of logical consequence. This kind of approach is based on the two intuitive traits of the notion of logical consequence. Recall that the second trait is that every argument with the same form as a logically correct argument is itself logically correct. So this gives a necessary condition of logically correct arguments, though in terms of the notion of logically correct argument. But it also suggests a necessary condition in terms of the notion of truth. Note that if an argument is logically correct then it's not the case that its premises are true and its conclusion false; if this were the case the premises would not entail the conclusion by logical necessity, and then, uncontroversially, by the first trait of logical consequence, the argument would not be logically correct after all. So, by the second trait, an argument is logically correct only if *no argument with the same logical form has true premises and a false conclusion*; call this necessary condition ' (Φ) '.

The alternative kind of approach to characterizing logical consequence uses always a variant of condition (Φ) , proposing it in each case as both necessary and sufficient for logical consequence. Without doubt Tarski's characterization is the paradigmatic representative of this alternative kind of approach. Tarski (1936) offered his characterization for the Fregean formal languages, accepting the notion of logical form for arguments of these languages implicit in Frege. It is good to point out, however, that Tarski's abstract method can be used, and is used, to give similar characterizations of logical consequence even for languages which extend Frege's languages.

³ No attention will be paid here to the important question whether or how formalized arguments faithfully mirror natural language arguments. Often the claims in this article are about formalized arguments, but also often there are abstract versions of these claims that hold for natural language arguments.

Tarski's proposal is to strengthen the requisite expressed by the condition (Φ) ⁴ so as to incorporate, at least partially, the idea that the non-logical expressions of a logically correct argument cannot be *reinterpreted* so as to make true the premises and false the conclusion (and not merely the idea embedded in (Φ) , that a logically correct argument cannot be turned into one with true premises and false conclusion by *replacing* non-logical expressions in a fixed language). In other words, the idea that a sentence X is a logical consequence of a set of sentences K when *every interpretation of the non-logical expressions in which all sentences of K are true is an interpretation in which X is true*⁵. Or, as is sometimes said, when every interpretation *preserves the truth* of the premises in the conclusion. When every interpretation preserves truth it is also said that the argument is *valid*. If an argument is valid then, even if it's not logically correct, the conclusion can be inferred from the premises without worry that it will be false if the premises are true. So all valid arguments are correct at least in this sense.

Tarski proposes to build a mathematically manageable version of the notion of validity by employing the apparatus developed by him for giving mathematical definitions of semantic concepts, such as satisfaction, definability, and truth. Tarski's method cannot be described here, but its current presentation is well-known⁶. It's based on defining, in a way analogous to the way Tarski defines truth in a language in Tarski (1935), the notion of truth in a set-theoretic structure. For a Fregean language, a structure is a set-theoretic assignment of denotations to the non-logical expressions (including a set from which those denotations are drawn and which gives the range of the first-order variables and induces ranges of the higher-order variables). The condition by means of which one characterizes the notion of a sentence X being a logical consequence of a set of sentences K (of that language) is then the following:

- (IV) Every structure in which all the sentences of the set K are true is also a structure in which the sentence X is true. (We'll abbreviate this by means of the notation ' $\text{Val}_T(K,X)$ '.)

'TV' stands for "Tarskian validity". The subindex 'T' is used to stress the fact that ' $\text{Val}_T(K,X)$ ' denotes Tarskian validity and that this is different from the validity relation. As I'm using here the notion of an interpretation which appears in the characterization of validity, this is a very imprecise and intuitive notion, while the notion of a structure appearing in a characterization of Tarskian validity is a fairly precise and technical one. Every Fregean formal language can be endowed with a condition of Tarskian validity using the Tarskian method. The same is true of many languages which extend the Fregean languages, and for which minimally reasonable notions of a structure have been defined. A notion of structure is minimally reasonable when it is

⁴ Tarski (1936) formulates a requisite closely related, if not identical, to condition (Φ) , which he calls '(F)'.

⁵ This condition is meant to take into account the fact that Fregean languages are interpreted by interpreting not merely their non-logical constants but also their first-order variables (which can be taken to be non-logical expressions); in interpreting them, a domain is given over which they range.

⁶ A detailed exposition of Tarski's 1936 method, and of some differences between it and the way in which Tarski's idea is presented nowadays, will be found in Gómez-Torrente (2000a); for related issues see also my (1996) and (2000b).

clear that every structure models the power of one or several interpretations to make true the premises and false the conclusion of some argument⁷.

3. *Soundness and completeness*

As pointed out earlier, if one builds a formal system with care, one will be able to convince oneself that all arguments whose conclusion is derivable from their premises in the system are logically correct. The reason is that one can have used one's intuition very systematically to obtain that conviction: one can have included in one's system only axioms of which one is convinced that they are logical consequences of every set of premises; and one can have included as rules of inference rules of which one is convinced that they produce sentences that follow logically from the sentences to which they are applied. Using another terminology, this means that, if one builds one's system with care, one will be convinced that the derivability characterization of logical consequence for arguments of the formal language in question will be *sound* with respect to logical consequence.

It is equally obvious that if one has at hand a notion of Tarskian validity for a formal language which is based on a minimally reasonable notion of structure, then all logically correct arguments (of that language) will be Tarski-valid. The reason is simple: if an argument is not Tarski-valid then there is a structure which makes true its premises and false its conclusion; but this structure ought then to model an interpretation (or interpretations) in which the premises are true and the conclusion false; so it would in principle be possible to construct an argument with the same logical form, whose non-logical expressions would have, by stipulation, the particular interpretations drawn from that collective interpretation, and which would have true premises and a false conclusion. But then the two intuitive traits of logical consequence uncontroversially imply that the original argument is not logically correct. Using another terminology, we can conclude that the Tarskian validity characterization is *complete* with respect to logical consequence.

Let's abbreviate 'X is derivable from K in system S' by 'Der_S(K,X)' and 'X follows logically (in the intuitive sense) from K' by 'LC(K,X)'. Then, if S is a system built with care, the situation can be summarized thus:

$$(1) \text{ Der}_S(K,X) \Rightarrow \text{LC}(K,X) \Rightarrow \text{Val}_T(K,X).$$

The first implication is the soundness of derivability; the second is the completeness of Tarskian validity. But in order to convince ourselves that the characterizations of logical consequence in terms of Der_S(K,X) and Val_T(K,X) are appropriate we should convince ourselves that the converse implications hold too:

$$(2) \text{ Val}_T(K,X) \Rightarrow \text{LC}(K,X) \Rightarrow \text{Der}_S(K,X).$$

⁷ As a later note mentions, the converse property, that each interpretation's validity-refuting power is modeled by some structure, is also a natural but more demanding requirement of adequacy on a notion of structure.

But obtaining this conviction, or the conviction that these implications don't in fact hold, turns out to be difficult for a great many languages, even Fregean languages for which Tarski originally intended his characterization. I'll examine some difficulties raised by a particular kind of Fregean languages shortly. But before that, I'll mention a remark of Kreisel (1967) that establishes that the implications in (2) hold in some cases.

Among the formal languages for which (1) holds there are many for which it is possible to give a mathematical proof that derivability (in some specified systems) is complete with respect to Tarskian validity, i.e. a proof of

$$(3) \text{Val}_T(K, X) \Rightarrow \text{Der}_S(K, X).$$

Kreisel called attention to the fact that (3) together with (1) implies that Tarskian validity is sound with respect to logical consequence, i.e., that the first implication of (2) holds. (Strictly speaking, this is a strong generalization of Kreisel's remark⁸.) This means that when (3) holds the notion of Tarskian validity offers an extensionally correct characterization of logical consequence⁹. Also, (3), together with (1), implies that the notion of derivability is complete with respect to logical consequence and hence offers an extensionally correct characterization of this notion.

An especially significant case in which (3) and (1) hold and so (2) holds, is that of first-order quantificational languages (for certain typical formal systems S)¹⁰. This means that one can convince oneself that both derivability and Tarskian validity (defined in appropriate ways for those languages) are extensionally correct characterizations of the intuitive notion of logical consequence for those languages. The situation is not so clear in other languages of special importance for the Fregean tradition, the higher-order quantificational languages.

4. Higher-order languages

It can be proved that already for a second-order language there is no formal system¹¹ S in which derivability is sound with respect to Tarskian validity and which makes true (3)—for the notion of Tarskian validity as usually defined for such a language. We

⁸ Something closer to Kreisel's remark is this: a version of (1) holds for Fregean languages with 'LC(K, X)' replaced by 'X is true in all the *class* structures (with a class as domain of the individual variables) in which all the sentences of K are true' (or, 'the argument (K, X) is class-valid'); then (3), together with this version of (1), implies that Tarskian validity is sound with respect to this notion of class-validity.

⁹ An interesting application of this pattern of reasoning is Hanson's (1997) argument that his intuitive characterization of first-order logical consequence is coextensional with first-order Tarskian validity.

¹⁰ Perhaps a qualification is necessary in view of the fact that sentences like $\exists x(x=x)$ are Tarski-valid consequences of the empty set of premises. (Certainly, even if some interpreted sentences of that form may be logically necessary, intuitively some others are not.) That fact is due to the usual convention of not contemplating empty universe structures when defining Tarskian validity. But Tarski's method for defining validity can be used and the corresponding (3)-type theorem proved for more inclusive conventions.

¹¹ In the technical sense of 'formal system' arising from any intuitively adequate characterization of the notion of effective enumerability, e.g., in terms of recursive functions.

may call this result *the incompleteness of second-order formal systems with respect to Tarskian validity*. In fact a stronger result holds: there is no set of second-order sentences for which a formal system sound with respect to Tarskian validity allows the derivation of all the Tarskian consequences of the set; said another way: for every set of sentences J and every system S sound with respect to Tarskian validity there will be a sentence X such that $\text{Val}_T(J, X)$ but it is not the case that $\text{Der}_S(J, X)$. Let's call this result *the strong incompleteness of second-order formal systems with respect to Tarskian validity*.

In this situation it's not possible to apply Kreisel's argument for (2). In fact, the incompleteness of second-order formal systems shows that, given any formal system S satisfying (1), one of the implications of (2) is false (or *both* are): *either* derivability in S is incomplete with respect to logical consequence *or* Tarskian validity is unsound with respect to logical consequence. A similar conclusion holds for cases of incompleteness in other formalized languages.

Different authors have extracted opposed lessons from incompleteness. A common reaction is to think that Tarskian validity must be unsound with respect to logical consequence. An intuitive idea occasionally appealed to in this connection is the idea that logically correct arguments, or at least those among them with a finite set of premises, must be arguments in which the conclusion is able to be drawn by humans from the premises through *a priori* reasoning. To this idea it is often added that every *a priori* reasoning must be reproducible in an effective way, if it must be capable of being performed by human beings. From all this, together with the strong incompleteness of second-order systems, the conclusion is reached that there must be arguments with a finite number of premises which are Tarski-valid but are not justifiable by *a priori* reasoning.

This way of arguing is incorrect, at least if understood as follows. Suppose that (i) every (human) *a priori* reasoning must be reproducible by a formal system. We accept also, of course, that (ii) for every finite set of premises and *for every* formal system S sound with respect to Tarskian validity *there is* a Tarski-valid argument with those premises and whose conclusion is not derivable from them in S . But from all this it doesn't follow that (iii) for every finite set of premises *there is* a Tarski-valid argument with those premises and such that *for every* formal system S sound for Tarski-validity its conclusion is not derivable from the premises in S . From (iii) and (i) it follows of course that there are Tarski-valid arguments that are not justifiable by (human) *a priori* reasoning. But the step from (ii) to (iii) is a typical quantificational fallacy. From (i) and (ii) it doesn't follow that there is any Tarski-valid argument which is not justifiable by (human) *a priori* reasoning. The only thing that follows (from (ii) alone under the assumptions that Tarskian validity is sound and that logically correct arguments with a finite set of premises are (humanly) *a priori* knowable implications) is that no formal system by itself can model all the *a priori* reasonings that people are able to make to extract Tarski-valid conclusions from (finite) sets of premises; but it's not sufficiently clear that this should be intrinsically problematic¹².

¹² The issue is related to the question whether some particular Turing machine can model human reasoning in general or mathematical human reasoning in particular. I dare to say that the common argu-

Etchemendy (1990) proposed this other argument for the unsoundness of Tarskian validity. (What follows is a reconstruction.) The first premise is that there are familiar second-order sentences \mathcal{A} and B such that

- (a) $\text{Val}_T(\emptyset, \mathcal{A})$ iff CH;
- (b) $\text{Val}_T(\emptyset, B)$ iff $\neg\text{CH}$.

Here ‘CH’ stands for some natural formulation of the continuum hypothesis. The equivalences (a) and (b) are true and provable in set theory. The next premises are:

- (c) it is not analytically true that CH;
- (d) it is not analytically true that $\neg\text{CH}$.

(c) and (d) are perhaps not as obvious as one might want, but might be conceded for the sake of argument. The final premise is

- (e) for every sentence X , if it is not analytically true that $\text{Val}_T(\emptyset, X)$ then X is not true by virtue of the meanings of its logical expressions.

From (a) and (c) on the one hand and (b) and (d) on the other Etchemendy derives

- (f) it is not analytically true that $\text{Val}_T(\emptyset, \mathcal{A})$;
- (g) it is not analytically true that $\text{Val}_T(\emptyset, B)$.

In fact, (f) and (g) clearly don’t follow from (a), (b), (c) and (d), since there is no reason to think that (a) and (b) state analytical equivalences. But (f) and (g) may again be conceded for the sake of argument. From (e), (f) and (g) it follows that neither \mathcal{A} nor B are true by virtue of the meanings of their logical expressions. However, either $\text{Val}_T(\emptyset, \mathcal{A})$ or $\text{Val}_T(\emptyset, B)$, since either CH is true or $\neg\text{CH}$ is. Therefore, concludes the argument, there are Tarskian consequences of the empty set of premises which are not true by virtue of the meanings of their logical expressions. Given the reasonable assumption that a sentence is true by virtue of the meanings of its logical expressions iff it is a consequence of the empty set of premises by virtue of the meanings of its logical expressions, from this conclusion it follows that there are Tarski-valid arguments whose conclusion doesn’t follow from their premises by virtue of the meanings of their logical expressions. And this suggests that these Tarski-valid arguments are not cases of logical consequence.

The main problem with this argument is perhaps that thesis (e) is not fully transparent and, as I understand it, I doubt that its denial is incompatible with the vague content of the general notion of analyticity and of the special notion of truth by virtue of the meanings of the logical expressions¹³. It may even be shown that some second-

ments for a negative reply to this question are inconclusive, but this is not the same as an argument for a positive reply, or even for the presumption that a positive reply is even likely to be correct.

¹³ (e) may be fallaciously mistaken for a more transparent thesis which Etchemendy has forcefully defended:

- (h) for every sentence X , it is not analytically true that [if $\text{Val}_T(\emptyset, X)$ then X is true by virtue of the meanings of its logical expressions].
- (h) is certainly plausible, but to go from (h) to (e) is to commit a modal fallacy.

order sentences X pre-theoretically and spontaneously called logically true by logicians (and presumably true by virtue of the meanings of their logical expressions) are such that the claim that $\text{Val}_T(\emptyset, X)$ is presumably not analytically true¹⁴.

In general, there are no fully satisfactory philosophical arguments for the thesis that Tarskian validity is unsound with respect to logical consequence in higher-order languages¹⁵. Are there then any good reasons to think that derivability (in any system sound for Tarskian validity) must be incomplete with respect to logical consequence? I don't think there are any absolutely convincing reasons either. Let's examine an argument in defense of this incompleteness, with which we will illustrate a fundamental difficulty—a difficulty which, by the way, could also be adduced to cast doubts on the ultimate value of the preceding arguments for the unsoundness of Tarskian validity.

The argument for incompleteness appears in Tarski (1936)¹⁶. Tarski notes that in some higher-order languages, e.g. in the language of a simple theory of finite types, it is possible to formalize arguments of the following form, without using numerals as primitives and without taking the predicate 'natural number' as primitive, but using only expressions of higher-order languages usually taken as logical (higher-order quantifiers in particular), plus variables and any predicate P :

$P(0),$
 $P(1),$
 $P(2),$
 ...

For every natural number n , $P(n)$

(instead of '...' one would have the rest of sentences of the form $P(n)$, for every natural number $n > 2$). We may call these arguments ω -arguments. Now, by Gödel's first incompleteness theorem, for every formal system S of the language of the simple theory of finite types which is sound with respect to Tarskian validity there will be an ω -argument such that its conclusion is not derivable in S from its premises. We may call this the ω -incompleteness of the formal systems of the language of finite types.

¹⁴ For more on the problems of Etchemendy's argument see Gómez-Torrente (1998/9) and Soames (1999).

¹⁵ There is another type of arguments for the unsoundness of Tarskian validity in the second-order case. In McGee (1992) there is a good example. The idea in these arguments is to find sentences X such that $\text{Val}_T(\emptyset, X)$ but which are intuitively false in a class structure. These arguments thus question the claim that each interpretation's validity-refuting power is modeled by some set-theoretic structure, a claim which is surely a corollary of the first implication in (2). (A critical analysis of McGee's argument can be found in Gómez-Torrente (1998/9).) But these arguments offer a challenge only to the idea that validity is adequately modeled by set-theoretic validity, not to the soundness of a characterization of logical consequence in terms of validity itself, or in terms of a species of validity based on some notion of structure different from the usual set-theoretic one. (The arguments we have examined in the text would have deeper implications if correct, for they offer a challenge to a characterization in terms of validity as well).

¹⁶ For a more detailed historical exposition see Gómez-Torrente (1996).

Tarski stresses the fact that ω -arguments are valid, in the sense that there is no interpretation of P (and the variables) that makes the premises true and the conclusion false. Certainly the conclusion can be concluded without worry from the premises. Perhaps it can even be said that ω -arguments are cases of implication by logical necessity, and of analytical implication¹⁷—though perhaps not of humanly *a priori* knowable entailment, since they have an infinite number of premises. A detailed evaluation of Tarski's claims ought to examine these matters more deeply¹⁸.

From ω -incompleteness Tarski concludes that derivability (in any formal system) must be incomplete with respect to logical consequence—and this leads him to propose his own characterization in terms of Tarskian validity. But a fundamental problem is that this conclusion is based on the assumption that the expressions typically cataloged as logical in higher-order languages, and in particular the quantifiers in quantifications of the form $\forall X$ (where X is a higher-order variable), are in fact logical expressions. If this assumption is made, then Tarski's observation may well convince us that every higher-order formal system sound for Tarskian validity, hence every system satisfying (1), will leave out some arguments which are cases of implication by logical necessity, and such that all arguments with the same form are cases of implication by logical necessity. And this makes it certainly very reasonable to think that if higher-order quantifiers are logical expressions, then derivability, in any formal system satisfying (1), must be incomplete with respect to logical consequence. But in the absence of additional considerations, a critic may question the assumption on which the argument is based, and deny relevance to the argument. (The assumption that the higher-order quantifiers are logical actually underlies any conviction one may have that (1) holds for any one particular higher-order system.) This same critic could for the same reason deny that the arguments presented above against the soundness of Tarskian validity are relevant at all.

This brings us directly to the last topic we'll examine in this article. What expressions are logical? (Or, what comes to the same: what forms are logical?)

5. *Logical expressions*

The principles guiding logicians (both implicitly and explicitly) in cataloging certain expressions as logical have been essentially pragmatic principles with a considerably vague content. Typically, logic has been seen as a discipline that deals with general reasoning, reasoning applicable in all argumentative fields. It is therefore reasonable to think that one principle underlying choices of expressions as logical is the principle

¹⁷ This is not to say that Tarski relies or would have relied on intuitions about logical necessity or analytical implication to make his point: he doesn't and he wouldn't. (See Gómez-Torrente (2000b) and (1996) for arguments that Tarski lacked a theoretical use for modal notions.)

¹⁸ It is interesting to note that Etchemendy, who probably would draw the borderline of the vague notions of logical consequence and analytical implication quite close to the zone of the clear instances, takes ω -arguments to be clearly logically correct and analytical implications (at least in first-order formalizations with '0' and 'successor' as primitive, but presumably also in higher-order formalizations without arithmetical primitives).

that logic must deal with arguments which are correct in virtue of the properties of expressions employed in general reasoning, expressions not specific of any argumentative field but common to all.

This principle suggests only a necessary condition of logical expressions¹⁹. Doubtless there are other implicit principles at stake. Logicians probably require implicitly that logical expressions be *very relevant* in general reasoning, or that their study be *useful* in the resolution of particularly *significant* problems in reasoning, to mention only two possibilities. If this is the right conception, then there is a great deal of vagueness and complexity in the intuitive concept of a logical expression. The pragmatic principles underlying the use of this concept leave much (perhaps too much) space for divergent conceptions, and for incompatible ideas on *what expressions* are logical. Nevertheless, this doesn't mean that the question of what expressions are logical is arbitrary, since the mentioned principles are not compatible with just any idea about what expressions are logical. Vagueness does not imply arbitrariness.

Logicians and philosophers of logic have often tried to offer philosophically richer characterizations of the notion of a logical expression than the one(s) suggested by the vague pragmatic principles just mentioned. These attempts usually characterize the notion of a logical expression (or of the logical expressions to be found in a restricted set of expressions) in terms of alleged semantic, epistemic or mathematical peculiarities of the logical expressions. I conjecture that these attempts will not succeed, since it must be nearly impossible to model closely the vague and pragmatic notion of a logical expression in those terms. I'll briefly describe a couple of attempted characterizations, which are representative of the two main kinds of approaches in the literature²⁰, and I'll indicate that they look deficient, which confirms the conjecture.

The first characterization is also due to Tarski, and is based on the idea that a permutation of a universe of discourse induces permutations throughout the hierarchy of types of objects determined by that universe. Thus, a permutation P of a universe of individuals U induces a permutation of the class of n -adic relations of elements of U , a permutation of the class of m -adic relations of n -adic relations of relations of elements of U , etc. Let's say that an object o of a certain type t is invariant under all permutations of the universe of discourse if, for every permutation P of this universe, the permutations \tilde{P} induced by P in the class of notions of type t are all such that $\tilde{P}(o)=o$. Let's say that $den(C, U)$ is the denotation of a expression C in a universe U . Then the property by means of which Tarski characterizes the logical expressions (more exactly, the logical expressions which get denotations in a hierarchy of types of objects) is this:

¹⁹ The fact that contemporary professionals of logic often study expressions which are not common to all argumentative fields does not show *per se* that the principle is conceptually incorrect nor even that it misdescribes logical practice. It shows at most that logical techniques are *applied* not merely to the study of logically correct arguments but also to the study of correct arguments *tout court*.

²⁰ A more detailed analysis of a number of questions having to do with the notion of a logical expression will be found in Gómez-Torrente (2002). (See also Sher (1991) for a good discussion of the notion and a defense of a characterization close to (ILC) below.)

(TLC) For every universe U and every permutation P of U , $\tilde{P}(den(C, U)) = den(C, U)$. (See Tarski and Givant (1987), p. 57.) (Let's abbreviate this by 'C is a Tarskian logical expression'.)

Thus, C is a Tarskian logical expression when it denotes, in every universe, a notion invariant under permutations. Tarski gave a plausible convention on how to assign denotations in the typical hierarchies to all the classical quantifiers and the truth-functional connectives, and under this convention they turn out to be Tarskian logical expressions.

The fact that higher-order quantifiers are cataloged as logical expressions by Tarski's characterization supports his argument for the incompleteness of derivability that we explained at the end of section 4. But consider the predicate 'MW', which we could add to a quantificational language, with the intuitive meaning "is a male widow". This predicate denotes the empty set in all actual and possible universes, since there are no male widows, nor can there be. So 'MW' is a Tarskian logical expression. But in none of the typical uses of 'logical expression' compatible with its vague meaning would 'MW' count as logical. There are of course many other similar though more complicated examples with expressions having denotations in higher types of the hierarchy of objects of a universe. These examples seem to show conclusively that the Tarskian characterization of the concept of a logical expression doesn't satisfy the requirement of determining a set of expressions compatible with the use (vague as it is) of 'logical expression'.

The second characterization I'll mention is due to Hacking (1979). It employs the following condition:

(H) The denotation of C is determined by a set of rules of a Gentzenian sequent calculus which are used in the inferences involving C ²¹. (Let's abbreviate this by 'C is a Hackingian logical expression'.)

Hacking's characterization is given together with a result which claims that the Gentzenian rules of the intuitively logical expressions of first-order quantificational languages "determine" the conventional Tarskian denotations of these expressions. Hacking also affirms that it is possible to show that the denotations of higher-order quantifiers are not "determined" by their Gentzenian rules (hence they are not Hackingian logical expressions).

If it is true that the Gentzenian rules of the intuitively logical expressions of first-order quantificational languages "determine" their denotations, then (H) is not a sufficient condition for an expression to be logical in the intuitive sense. The reason is that it is possible to imagine, or simply to stipulate, the existence of expressions governed

²¹ A different version of (H) has been proposed by other authors, Kneale (1956) being perhaps the first:

(K) The *sense* of C is determined by a set of rules of a Gentzenian sequent calculus which are used in the inferences involving C .

But (K) doesn't seem to be a property of many uncontroversial logical expressions, since the senses of these appear to be richer than any senses that can be "determined" by their typical Gentzenian rules (on this see Gómez-Torrente (2002)).

in part by the same Gentzenian rules as an intuitively logical expression, which are, nevertheless, intuitively non-logical. Consider the first-order quantification ‘for some x , if there are no male widows, or for every x , if there are male widows’, which we can stipulate to be synonymous with the expression $\exists x$, with \exists a primitive with the mentioned sense²². \exists has the same denotation as the usual first-order existential quantifier (in every universe). Unlike normal existential quantifiers, \exists is not intuitively a logical expression. But if Hacking’s result is true, then \exists satisfies condition (H), since the typical Gentzenian rules for the existential quantifier govern the use of \exists , and then its denotation is “determined” by a set of Gentzenian rules which are used in the inferences involving \exists . \exists is a Hackingian logical expression, but not a logical expression. On the other hand, the thesis that (H) gives a necessary condition of logical expressions is not so clearly false as the thesis that it gives a sufficient condition, but it is far from compelling.

In my view, then, characterizations like Tarski’s and Hacking’s don’t provide a good basis for accepting or rejecting higher-order quantifiers as logical expressions. The vague principles with which we started this section don’t have clear implications about this, either. It is tempting to think, though, that the vagueness of these latter principles is compatible both with taking higher-order quantifiers as logical and with denying them this status. But if both options are acceptable, we should probably work under the hypothesis that they are logical when we evaluate arguments such as those we saw in section 4. The Tarskian argument for the incompleteness of derivability looks then a bit better than the examined arguments for the unsoundness of Tarskian validity. A little more can be said about all these arguments, and probably more will be said in the future. But I will stop here.

BIBLIOGRAPHY

- Etchemendy, J. (1990), *The Concept of Logical Consequence*, Cambridge (Mass.): Harvard University Press.
- Gómez-Torrente, M. (1996), “Tarski on Logical Consequence”, *Notre Dame Journal of Formal Logic*, vol. 37, 125-151.
- (1998/9), “Logical Truth and Tarskian Logical Truth”, *Synthese*, vol. 117, 375-408.
- (2000a), *Forma y Modalidad*, Buenos Aires: Eudeba.
- (2000b), “A Note on Formality and Logical Consequence”, *Journal of Philosophical Logic*, vol. 29, 529-539.
- (2002), “The Problem of Logical Constants”, *Bulletin of Symbolic Logic*, vol. 8, 1-37.
- Hacking, I. (1979), “What is Logic?”, *Journal of Philosophy*, vol. 76, 285-319.
- Hanson, W. H. (1997), “The Concept of Logical Consequence”, *Philosophical Review*, vol. 106, 365-409.
- Kneale, W. (1956), “The Province of Logic”, in H. D. Lewis (ed.), *Contemporary British Philosophy, 3rd Series*, London: Allen & Unwin, 237-26.
- Kreisel, G. (1967), “Informal Rigour and Completeness Proofs”, in I. Lakatos (ed.), *Problems in the Philosophy of Mathematics*, Amsterdam: North-Holland, 138-171.
- McGee, V. (1992), “Two Problems with Tarski’s Theory of Consequence”, *Proceedings of the Aristotelian Society*, n.s., vol. 92, 273-292.

²² More clearly: $\exists xF$ is true iff [(there are no male widows and $\exists xF$) or (there are male widows and $\forall xF$)].

- Sher, G. (1991) *The Bounds of Logic. A Generalized Viewpoint*, Cambridge (Mass.): M.I.T. Press.
- Soames, S. (1999), *Understanding Truth*, New York: Oxford University Press.
- Tarski, A. (1935), "The Concept of Truth in Formalized Languages", in Tarski (1983), 152-278.
- (1936), "On the Concept of Logical Consequence", in Tarski (1983), 409-420.
- (1983), *Logic, Semantics, Metamathematics* (2nd edn.), Indianapolis: Hackett.
- Tarski, A. and S. Givant (1987), *A Formalization of Set Theory without Variables*, Providence (R.I.): American Mathematical Society.

Mario GÓMEZ-TORRENTE is a researcher at ICREA and teaches at the University of Barcelona. He is also a researcher at the Instituto de Investigaciones Filosóficas, UNAM. His main interests are in the philosophies of logic and language. He is the author of several publications in these and other areas of philosophy.

Addresses: ICREA & Univ. de Barcelona, Depto. de Lógica, Baldiri Reixac, s/n, 08028 Barcelona, Spain, and Instituto de Investigaciones Filosóficas, UNAM; México DF 04510, Mexico. E-mail: *mariogt@servidor.unam.mx*