An interpretationalist and structuralist account of logical consequence

Philipp Blum

Very much work in progress

Representational and interpretational semantics

In his seminal book on Tarski's account of logical consequence, John Etchemendy (1990) distinguishes between what he calls representational and interpretational semantics. They differ in the way they assign counterfactual truth-values to the schematic letters: interpretational semantics attributes values for *truth-in-*L, where L ranges over some class of different languages; representational semantics attributes values for *truth-in-*w, where w ranges over some class of worlds different from the actual one. Logical truth, for interpretational semantics, is truth under all (admissible) reinterpretations of the non-logical constants, and it is truth in all possible worlds for representational semantics, truth no matter what the extensions of the non-logical constants are. Absolute, non-relational, truth is truth-in-English for interpretationalists, while it is truth-in-the-actual-world for representationalists.

Representational and interpretational semantics also differ with respect to the account they give of the logical constants: while the first implicitly characterises them as invariants among the assignments of extensions ("worlds") deemed admissible, the latter explicitly holds them constant while reinterpreting the non-logical parts of the language.

The main question of interest to Etchemendy (1990), "how do we know that our semantic definition of consequence is extensionally correct?" (1990: 4) receives different treatments in the two cases. While the adequacy problem in representational semantics is relegated to the prior and non-semantic specification of the relevant realm of "possible worlds" (1990: 60,77), it comes to the fore with interpretationalism:

> "A sentence is true in a given model if, so to speak, what it would have said about the world on the suggested interpretation is, in fact, the case. Thus, sentences that come out true in all models are true regardless of how we interpret a subset of their component expressions. Here, too, the "regardless" must be qualified: the result holds only modulo our circumscription of the class of semantically well-behaved reinterpretations of the variable terms." (1990: 61)

In the following, I will first sketch Etchemendy's interpretation and critique of Tarski's account of logical consequence, then give an independent argument for interpretational semantics and try to meet Etchemendy's criticisms and finally sketch my own theory of logical consequence which, or so I hope, keeps what is good in Tarski's.

Etchemendy's critique of Tarski

Etchemendy (1990: 23) claims that Tarski (1936b), while usually taken to be a representationalist, has given an interpretationalist account of logical consequence: a sentential function S is a logical consequence of some set of sentential functions K with respect to some expressions $\mathcal{F}$ iff any pair $\langle K', S, \rangle$ differing from $\langle K, S\rangle$ by uniform replacement of all atomic expressions not in $\mathcal{F}$ with expressions of the same grammatical type is such that every sequence f either satisfies S' or does not satisfy some member of K' (1990: 49).

His critique of Tarski is three-fold: he notes, first, that Tarski's account depends on a prior account of semantic categories, defining the domains for the satisfaction function; second, that it relies on cross-term

restrictions that cry out for independent motivation; third, and most importantly, that it is unable to account for the distinctive modal element of our intuitive notion of consequence: instead, by "Tarski's fallacy", he equates logical truth with simple truth of an appropriate universal generalisation. I will take up this three criticisms in turn and then try to answer them in the next section.

Etchemendy poses a dilemma for the interpretationalist: in order to satisfy two important desiderata for any account of logical consequence or logical truth, persistence of non-logicality through contractions – "if a sentence S is *not* a logical truth of a given language, then neither should it *become* a logical truth simply by virtue of the deletion of expressions *not occurring in* S" (1990: 31) – and, equivalently, persistence of logicality through expansions of the language, Tarski has to generalise Bolzano's purely syntactical notion of substitution to the semantic notion of satisfaction as a relation between entities (as opposed to expressions) and sentential functions. Satisfaction, however, cannot be satisfactorily defined without an account of *semantic categories*, where a semantic category is a class of terms whose contributions to the truth-value of sentences in which they occur are of the same type. This already severely limits the insight we might gain in the "logicality" of the logical constants.[1]

The need for cross-term restrictions arises as soon as we want to define logical consequence for a quantificational language. In order for "something was president" to be a logical consequence of "George Washington was president", an interpretationalist has to impose cross-term restrictions on his acceptable reinterpretations, by only allowing reinterpretations of "George Washington" that assign to it an entity that falls within the set of objects assigned to the existential quantifier (1990: 68). If, on the other hand, we treat "something" as constant, then any cardinality statements come out logically true (1990: 74). By preserving persistence, cross-term restrictions invalidate the original substitution criterion: by ruling out some reinterpretations, they declare arguments logically valid *despite* their having false substitution instances.

What Etchemendy takes to be "Tarski's fallacy" is the erroneous belief that S being a Tarskian consequence of K entails that the argument from K to S is *necessarily* truth-preserving (1990: 87). All that follows from Tarski's account is that the step from true premisses to a false conclusion cannot be a Tarskian consequence, but not that being a Tarskian consequence implies anything stronger than actual truth-preservation. It is therefore impossible to know both that S is a consequence of K and that all sentences in K are true without *antecedently* knowing that S is true: the premisses of Tarskian consequences do nothing to justify their conclusions (1990: 93). In a later paper, he formulates this "conceptual flaw" of the Tarskian analysis thus:

> "The fact that validity is tied to a larger class of arguments does not help, nor does the appeal to logical constants as a means of specifying that class. Indeed the problem remains no matter how narrowly one construes the term "logical constant," no matter how general the resulting argument forms. [...] if the logical validity of the argument came down to nothing more than the fact that every instance of the illustrated form preserves truth, then the truth of the premises could never be used to establish the truth of the conclusion." (2008: 268)

The reason for this inadequacy, Etchemendy (1990: 96) claims, is that Tarski equates logical truth of a sentence S of L with the truth of the universal closure of the sentence S' obtained from S by uniformly replacing the variable terms with variables in an expanded language L^+. Because the latter depends on contingent facts, any guarantee for the extensional adequacy of the Tarskian account is lost (Etchemendy 1990: 108).

Even if what we held constant is stipulated to be just the expressions "of a distinctively logical sort" occurring in the original sentence (1990: 110), the account both overgenerates and undergenerates. It will count as logically true some cardinality statements about the size of the universe (1990: 113), count – if and only if the universe is finite – as logically true the claim that every irreflexive and transitive relation has a maximal element and depend on the pair-set axiom to guarantee that "$\forall x, y \forall F(Fx \rightarrow Fy)$" is not a logical truth:

[1] Etchemendy (1990: 50) takes such an account to fall within the domain of an adequate theory of satisfaction: "...the goal of achieving a persistent account of the logical properties makes no sense except in the context of a theory of (or assumptions about) how existing members of a category contribute, and how potential members could contribute, to the truth values of sentences in which they occur. The required account of satisfaction must provide such a theory, both to give precise (and plausible) sense to the demand for persistence, and of course to give us resources with which to meet that demand."

> "What makes for the correct extension are such things as the existence of an infinite number of objects, the assumed distinguishability of those objects, the existence of transitive, irreflexive relations with and without minimal elements, and so forth." (1990: 127)

Because the only way of guaranteeing that there are no logical truths which depend on the meanings which we vary is to hold *every* expression constant and we thereby inevitably overgenerate, we cannot avoid the danger of undergeneration either (1990: 132–133). The question whether there can be an independent, i.e. logical guarantee that Tarski's account is extensionally correct thus receives a negative final answer (Etchemendy 1990: 130). We are trapped between two complementary dangers without any recipe to avoid them both:

> "To avoid undergeneration, we cannot exclude from $\mathcal{F}$ [the set of expressions we hold constant] any expression whose meaning plays an essential role in any valid arguments expressible in the language. [...] But this general characteristic of expressions – figuring essentially into valid arguments – has nothing to do with the question of whether substantive claims can be made employing just those expressions plus an initial string of quantifiers. And it has even less to do with whether those substantive claims turn out to be true – in other words, with whether the account will, on that selection of $\mathcal{F}$, overgenerate." (1990: 133–134)

Interpretationalism defended

In this section, I would like to point to some independent advantages of interpretationalist semantics which will make it appear worthwile to try to save it from Etchemendy's criticisms.

The major drawback of representationalist semantics, built right into its motivation, is its inability to distinguish logical truth from necessary truth. To avoid classifying "water is H_2O" as a logical truth, further resources are necessary and it is not clear where they can be found if not in a partial return to an interpretationalist stance. Furthermore, as Kaplan (1989) has argued, there are good reasons for accepting contingent logical truths like "I am here now".

As Etchemendy clearly recognises, representationalist semantics is of not much help when it comes to variations in the meaning of logical constants:

> "If we try to view such models [where "or" is considered variable] representationally, we must somehow imagine that "or" receives its ordinary inferpretation and that our assignment of various truth functions to this expression is just a technique for representing possible configurations of the "nonlinguistic world". But there is no plausible way of understanding, representationally, models in which "or" is assigned, say, the truth function ordinarily expressed by "and"." (1990: 62)

Another drawback, not mentioned by Etchemendy, comes with our obligation to interpret every truth-value assignment as the representation of a possible world. It seems clear to me that all of the following sentences are logically true:

(1) Either Holmes is a detective or he is not.

(2) Either some unicorns are black or all of them are not.

(3) If $2 + 1 = 3$ then $3 - 1 = 2$.

That these sentences should be counted as logically true entails that they should be counted true *for the right reasons*. Representationalist semantics is in no position to afford these. In order to classify (1) as true in all models, for instance, it has to classify models according to which Holmes does not exist as models of "Holmes is not a detective". In particular, this sentence would be true in the model representing the actual world.

With respect to (2), we do not want to say that it is necessarily true just because worlds without unicorns are such that all their unicorns are not black, for we do not want to count "all square circles are red" as

necessarily true. The problem with (3) is related: how are we to count it as logically true *as opposed to* the quite different "If $2 + 1 = 3$ then Fermat's theorem is true"?

Interpretationalism easily deals with all of these problems: whatever our account of the referents of fictional names, "Holmes" clearly is a name for a person; whatever our policy regarding empty predicates, "unicorn" is a name for a species of animals; whatever our theory of mathematical truth, "1", "2" and "3" are terms denoting natural numbers.

Another concern is ontology. Despite its ideological drawback of depending on a prior account of semantic categories and the heavy theoretical burden generated by it, there is an obvious ontological advantage to interpretational semantics, even though Etchemendy does not make much out of it. Representational semantics relies heavily on our intuitions concerning counterfactuals and the whole apparatus of "ways the actual world might be" connected to them: a sentence is declared true in a given model if it *would be* true were the world actually as depicted by that model. Interpretational semantics, on the other hand, at most involves "counterlinguisticals", hypotheses about what sentences would be true had their component words different meanings than they actually have. Languages, however, as opposed to possible worlds, are ontologically cheap: the language L' we need to reinterpret L *actually* exists and is an abstract object of exactly the same kind than L, differing from it only in that it is not a language we speak.

Etchemendy's argument for the need of cross-term restrictions in the definition of logical consequence for quantificational languages gets the logic of the existential quantifier wrong. When we are reinterpreting "something", we are not interpreting it as expressing some different kind of restricted quantifier, but as denoting some particular individual in the domain. We are treating it as a proper name for an *arbitrary object*, a name that denotes different individuals in different interpretations.

As Etchemendy argues, the interpretational account of necessary truths – "...sentences which are true in all models [...] are just those whose truth would survive any semantically well-behaved reinterpretation of the atomic sentences of the language" (1990: 59) – relies heavily on a prior account of when a reinterpretation of the language is semantically well-behaved. It may well be possible, as he insinuates (Etchemendy 1990: 76), that no non-circular such account is possible.

It is by no means mandatory, however, to characterise logical truth as a certain species of necessary truth and it seems worthwile to try a different route and characterise it as a species of analytic truth. As Etchemendy (1990: 101–103) observes, it seems right to classify the following three sentences:

(4) Either Lincoln was president or Lincoln was not president.

(5) Either everyone is happy or someone is not happy.

(6) Either Leslie is a man or Leslie is not a bachelor.

as true in virtue of the meanings of "or" and "not" (4), "or", "not", "everyone" and "someone" (5) and "or", "not", "man" and "bachelor" (6) respectively. Etchemendy (1990: 104) is right that the truth of the universal generalisation of S', in which all expressions occurring in S except $e_1, \ldots, e_n$ are replaced by variable expressions, gives us only a necessary, but not a sufficient condition for S's being true solely in virtue of the meanings of $e_1, \ldots, e_n$.

Even apart from the fact that the generalisation may be true for whatever reasons whatsoever, the generalisability account would force us to say that "Leslie is a man or Socrates is a philosopher or is not a philosopher" is analytic (also) in "Leslie" and "man", while it seems clear that it is true *solely* in virtue of "or".

References

Etchemendy, John, 1990. *The Concept of Logical Consequence*. Cambridge, Massachusetts: Harvard University Press.

Etchemendy, John, 2008. Reflections on Consequence. In Patterson, Douglas Eden (editor) *New Essays on Tarski and Philosophy*, pp. 262–299. Oxford: Oxford University Press.

Kaplan, David, 1989. Demonstratives. In Almog, Joseph, John Perry & Howard K. Wettstein (editors) *Themes from Kaplan*, pp. 481–563. Oxford: Oxford University Press. Widely circulated from 1977 on.

Tarski, Alfred, 1936a. O pojciu wynikania logicznego. *Przegląd Filozoficzny* 39: 58–68. Translated into German (Tarski 1936b) and English (Tarski 1956).

Tarski, Alfred, 1936b. Über den Begriff der logischen Folgerung. In Institut International de Collaboration Philosophique (editor) *Actes du Congrès International de Philosophie Scientifique, Paris, 1935. Volume 7: Logique*, pp. 1–11. Number 394 in Actualités scientifiques et industrielles, Paris: Hermann & cie. URL `http://gallica.bnf.fr/`. German translation of Tarski (1936a).

Tarski, Alfred, 1956. The Concept of Logical Consequence. In *Logic, Semantics, Metamathematics*, pp. 409–420. Indianapolis, Indiana: Hackett Publishing Co. English translation based on Tarski (1936a) and Tarski (1936b).