The Modal and Epistemic Arguments Against the Invariance Criterion for Logical Terms*

Gil Sagi

There is a criticism of the isomorphism-invariance criterion for logical terms that is expressed in several variations in the literature on logical terms. The criticism in most cases was aimed against the criterion of invariance under isomorphism,[1] but it can be seen as applying to criteria of invariance under other transformations[2] just as well. The gist of the objection is that invariance criteria pertain only to the extension of logical terms, and neglect the *meaning*, or the way the extension is fixed. A term, so claim the critics, can be invariant under isomorphisms and yet involve a contingent or a posteriori component in its meaning, thus compromising the necessity or apriority of logical truth and logical consequence.

The criticism has been expressed in different variations by various authors. The first to present the criticism, to the extent of my knowledge, is McCarthy. Others include McGee

---

* Penultimate version (April 2015), to appear in the *Journal of Philosophy.*

[1] As in Gila Sher, *The Bounds of Logic* (Cambridge, MA: MIT Press 1991).

[2] Such as those proposed in Saul Fefereman, "Logic, Logics and Logicism," *Notre Dame Journal of Formal Logic*, XL, 1 (Winter 1999): 31-55, and Denis Bonnay, "Logicality and Invariance," *The Bulletin of Symbolic Logic*, XIV, 1 (March 2008): 29-68.

and Hanson.[3] I shall argue that the arguments underlying the criticism are flawed. I will divide the criticism into a *modal argument* and an *epistemic argument*, according to whether it is the necessity or the a priority of logic that is claimed to be undermined.

In the main part of the paper (§§1-2) I will concentrate on McCarthy's presentation of the arguments. Before that, we can look at some other formulations. So, for instance, Hanson gives the following putative counterexample to Sher's invariance criterion of logical terms: "Let *n* be the least number of whole seconds (that is, the least number of seconds, disregarding fractions of a second) in which, up through the end of the twenty-first century, a human being runs a mile. Now consider a quantifier that behaves exactly like the universal quantifier (over individuals) in models with domains of cardinality $\geq n$, but like the existential quantifier in models with domains of cardinality $<n$. Call this quantifier 'Q^*'. 'Q^*' is a logical term on Sher's account because it satisfies her semantic isomorphism conditions, although it seems bizarre to treat it as one. To see just how bizarre this is, consider the following argument:

(7)

$$(Q^*x)(Dog(x) \rightarrow Black(x))$$

$$(Q^*x)Dog(x)$$

$$\therefore\ (Q^*x)Black(x)$$

[3] Timothy McCarthy, "The Idea of a Logical Constant," *Journal of Philosophy*, LXXVIII, 9 (September 1981): 499-523, Vann McGee, "Logical Operations," *Journal of Philosophical Logic*, XXV, 6 (1996): 567-580, William H. Hanson, "The Concept of Logical Consequence," *Philosophical Review*, CVI, 3 (July 1997): 365-409.

As long as $n \geq 3$, argument (7) has counter models. So we know that (7) is invalid, since we know that no one will run a mile in less than three seconds before the end of the next century (or ever, for that matter). Yet we don't and can't know this a priori."[4]

As much as Hanson's argument shows a clash between the isomorphism-invariance criterion and apriority, it can also be used to show that the violation of necessity of logical consequence. That is since, arguably, the number of seconds a human being runs a mile is a contingent matter. Other critics have used other counterexamples to the same effect. In the next sections I shall confront the two types of argument.

Before we go on, however, there is an important distinction to be made. Note that the setting is that of model-theoretic semantics—this is dictated by the invariance criteria that

---

[4] Hanson, "The Concept of Logical Consequence," *op. cit.*, pp. 391f. Sher responds to Hanson in "The Formal-Structural View of Logical Consequence," *Philosophical Review*, CX, 2 (April 2001): 241-261, and is replied to by Hanson in "The Formal-Structural View of Logical Consequence: A Reply to Gila Sher," *Philosophical Review*, CIX, 2 (April 2002): 243-258. Sher's ultimate reply to the modal argument is reference to her condition (B) in her criterion for logical terms (Sher, "The Formal-Structural View of Logical Consequence," *op. cit.*, p. 249, Sher, *The Bounds of Logic*, *op. cit.*, pp. 56, 64). According to (B), a logical constant is defined by a single extensional function, and is identified with its extension. In other words, logical terms are rigid, in the sense that their definitions in the metatheory are rigid. Sher does not pose apriority as a requirement on logical consequence. Due to its problematic nature, Sher wishes to remain neutral on the subject of the a priori (though she contends that posing a further condition excluding so-called *empirical terms* does not conflict with her approach, only limits its generality). Hanson claims that given her assumptions, Sher cannot be neutral with respect to the apriority of logical consequence (Hanson, "The Concept of Logical Consequence", *op. cit.*, p. 392). Sher claims, however, that her account is in no way committed to the contentious concept of the *a priori* (Sher, "The Formal-Structural View of Logical

are here at stake. So let us assume we have a language L for which we give model-theoretic interpretations. We can assign the extensions of a term in models in two ways: (*i*) by a definition or abbreviation in the object language, and (*ii*) by rule in the metalanguage. Not all the relevant critics say explicitly whether the example of a recalcitrant term is given by way of (*i*) or of (*ii*). It seems like what Hanson is opting for in the above quote is (*ii*): he is explicitly referring to models and their domains. But in other cases this is less clear. McGee, for instance, defines *unicorn negation* as:

$$\mathcal{U}\phi =_{Def} (\neg\phi \wedge \text{there are no unicorns})$$

and claims that '$\mathcal{U}$' is intuitively nonlogical, though the operation described by '$\mathcal{U}$' is invariant: "Indeed, if "$\mathcal{U}$" were counted as a logical connective, "($\neg$0=0 $\vee \mathcal{U}$ $\neg$0=0)" would be counted, according to Tarski's definition, as a logical truth; and "($\neg$0=0 $\vee$ $\mathcal{U}$ $\neg$0=0)" surely oughtn't count as a logical truth, since it entails "There are no unicorns"."[5] McGee, who defends the invariance criterion for logical *operations*, contends that merely denoting a logical operation is insufficient for making a *connective* logical.

McGee does not say explicitly whether he ascends to the metalanguage to define '$\mathcal{U}$', that is whether he takes the way of (*i*) or of (*ii*). However, we should set the criticism employing (*i*) aside as an obvious *non sequitur*. If McGee's formulation is meant as an

---

Consequence," *op. cit.*, pp. 257-260). I will offer a further response to Hanson's concerns.

[5] McGee, "Logical Operations," *op. cit.*, p. 569f.

abbreviation in the object-language, the claim that the operation denoted by '$\mathcal{U}$' is invariant is problematic. In that case 'unicorn' is itself a term in the object-language. Indeed, '$\mathcal{U}$' would be invariant under isomorphisms (or any of the other transformation under consideration) only if 'unicorn' is, which McGee gives us no reason to suppose.

In the rest of the paper I focus on the criticism employing (*ii*), i.e. giving a rule in the metalanguage, and show that it is invalid. My main contention will be that the criticism misidentifies the role of the metalanguage in fixing the meaning of logical terms. I focus on McCarthy's formulation of the criticism, since it is formulated in the most general manner. The discussion will clearly apply to others' formulations as well.

1 The modal argument

The semantic framework McCarthy uses is similar to standard contemporary model-theoretic semantics. Given an object language, primitive predicates and individual constants are assigned extensions in structures over *ontologies*. Expressions with free variables are then assigned extensions over ontologies relative to a variable assignment over the same ontology. Expressions can thus be characterized by rules of satisfaction. Ontologies are not defined or constrained by McCarthy, but it seems that for our purposes we can speak of models and domains instead of structures and ontologies, as is currently more commonly accepted, and rephrase McCarthy's argument accordingly. The basic idea of McCarthy's argument is not bound to suffer from this change in framework.

Let us now reconstruct McCarthy's argument to the effect that some isomorphism-invariant terms suffer from some sort of contingency. Suppose that '*K*' is some contingently true statement of the metalanguage.[6] Consider a unary connective '*N*' with the following semantic clause:[7]

(N) $\forall s(s$ satisfies $\ulcorner N\phi \urcorner \leftrightarrow [(\sim s$ satisfies $\phi \wedge K) \vee (s$ satisfies $\phi \wedge \sim K)])$

Since '*K*' is true, '*N*' is coextensional with negation, and is invariant under isomorphism. If we allow '*N*' as a logical term, as the usual invariance criteria would do, we would have to fix its interpretation according to (1). And then, for any sentence φ:

(LT) '$N\phi \leftrightarrow \sim\phi$' is a logical truth.

But, McCarthy claims, in any counterfactual situation in which '*K*' is false, the sentence '$N\phi \leftrightarrow \sim\phi$' is false. This is based on the observation that,

(1) It could have been the case that '*N*' was not coextensive with negation.

And thus we have,

(2) '$N\phi \leftrightarrow \sim\phi$' could have been false,

---

[6] McCarthy also assumes that '*K'* is a posteriori, but that assumption is redundant in this version of the argument.

[7] McCarthy, "The Idea of a Logical Constant," *op. cit.*, p. 514.

in support for McCarthy's claim. But then, McCarthy contends that this is an unhappy result, since "it is a reasonable constraint on any theory of validity for a language L that a sentence of L that it construes as a logical truth is *not* possibly false in L".[8]

McCarthy is plainly alluding here to the condition of necessity of logical truths:

(NEC) If ϕ is a logical truth, then ϕ is necessarily true.

The problem with the modal argument is that it relies on an invalid inference from a modal statement about a linguistic entity, namely "The sentence '$N\phi \leftrightarrow \sim\phi$' may have been false" to the modal status of the content of that linguistic entity, "It may have been that $N\phi \leftrightarrow \sim\phi$ was not the case".[9] Now, if 'K' were not true, then '$N\phi \leftrightarrow \sim\phi$' would not have been true: merely because we would have attached to 'N' a different meaning.

Arguably, the array of extensions of a given term in standard first-order logic represents its meaning. This view especially congenial if we view models as representing possible worlds, as I will propose to do later on. In that case, we can view the function assigned to each term from models to the relevant extensions as representing *intensions*, understood as functions from possible worlds to relevant extensions in those worlds. Thus,

---

[8] *Ibid,* p. 515.

[9] This difference, between the necessary truth of a sentence and the necessity of its content is at the base of the *Church-Langford argument* and its successors, see Cooper H. Langford, "Beth EW. De significa van de pasigrafische systemen. Bijdrage tot de psychologie van het wiskurdig denkproces. (The signifies of pasigraphic systems. A contribution to the psychology of the mathematical thought process.) Euclides, vol. 13 (1936–1937), pp. 145–158," *The Journal of Symbolic Logic*, II, 1 (March 1937): 53-54, Alonzo Church, "A Formulation of the Logic of Sense and Denotation," *Structure, Method, and Meaning* (New York: Liberal Arts Press, 1951): 3-24, and also Casimir Lewy, "Truth and significance," *Analysis*, VIII, 2 (December 1947): 24-27. I thank Hannes Leitgeb for the references.

if we look at the *content* of '$N\phi\leftrightarrow\sim\phi$' given the meaning actually assigned to 'N', there is no possibility for it to fail to hold.

One may object to the presumption that the array of extensions of a term fully captures its meaning in each case. A possible defense would be that this is a limitation of the framework of extensional logic. Whatever can be represented by arrays of extensions is the most of the meaning of a term that can be captured in this framework. The metalanguage merely guides to this array. Some rules are more useful than others, but any difference between two metalinguistic rules giving the same array of extensions is irrelevant to the modal status of sentences in the object language.[10]

To make the point clearer, let us consider an analogous argument. It seems an obvious truth about the nature of language that,

(1') It could have been the case that '2' referred to 3, while '+', '=' and '4' referred to +, = and 4 respectively.[11]

---

[10] Thus, we encounter the well-known problem of viewing meanings as functions of possible worlds: any two expressions that are necessarily coextensive are rendered synonymous. So, for instance, the predicate *male-widow* is necessarily coextensive with *not-self-identical* (or *self-distinct*), and is thus synonymous with it by the current view (see Mario Gómez-Torrente, "The Problem of Logical Constants," *The Bulletin of Symbolic Logic*, VIII, 1 (March 2002): 1-37 for a criticism of invariance criteria based on such examples. See also McGee, "Logical Operations," *op. cit.*, p. 578 for a similar example).

[11] One might be missing a language parameter here, and claim that a term always refers to an object in some language. And then, if languages are individuated, *inter alia*, by the meanings they give to expressions, then it *couldn't* have been the case that '2' referred to 3 *in English*. But in the same vein, it couldn't have been the case that *N* was not coextensive with negation *in L*. Thus, in order for McCarthy's (1) to be true, if we parameterize with respect to a language, then the language parameter

So, we can conclude that

(2') '2+2=4' could have been false.

But still, we can say with certainty that necessarily, 2+2=4. The contingency of the sign-object relation does not infringe the necessity of arithmetical claims. The arithmetical case is analogous to McCarthy's example. As it happens, 'K' is true, so 'N' refers to the functor of negation. Indeed, what 'N' refers to is a contingent matter. But so is the reference of '2'.

McCarthy tells us that in some world, where 'K' is false, '$N\phi \leftrightarrow \sim\phi$' is false. Note that in that world '$N\phi \leftrightarrow \sim\phi$' is also *logically* false, while in ours it is logically true. They simply use a different language in that world. Logical truths should presumably be true in all possible worlds, but only so long as their meanings are fixed. As vexed an issue as meaning is, that much is assumed when it is assumed that logical truths are necessary truths. That is, talk of necessary truth that employs possible worlds assumes that sentences can be evaluated in possible worlds *given the meanings that they have*.

McCarthy's claim that the invariance criterion yields logical truths that defy (NEC) is thus falsified, for the intended meaning of that condition. But we can still ask, is '$N\phi \leftrightarrow \sim\phi$' a necessary truth? Rule (N) McCarthy gives us is presumably a model-theoretic rule, assigning an extension to '$N\phi$', for all 'ϕ', in each model. Assuming 'K' is true, '$N\phi \leftrightarrow \sim\phi$' is true in all models. But is it true in all possible worlds?

---

should be descriptive, such as "in the language defined here". Likewise, we can add to (1') "in the language we use here".

We receive a straightforward positive response to this question if we view models as representing possible worlds.[12] The logical truth of a sentence would guarantee its necessity. On the option of not relating models to possible worlds, we need an alternative method for determining the necessity of a statement, whether it is true in all possibilities. But even then, only the terms of the object-language will change their extension across possibilities. The metalanguage, just as in the model-theoretic interpretation, is used for telling us how to interpret the object-language, and is not re-interpreted or re-evaluated itself. Thus, observing rule (N), the only thing that can vary the truth-value of '$N\varphi$' at different possible worlds is the truth-value of 'ϕ'. Thus, given that 'K' is a true statement in the metalanguage, '$N\phi\leftrightarrow\sim\phi$' would be a necessary, logical truth.

2 The epistemic argument

The epistemic version of the argument for the insufficiency of invariance under isomorphisms is more delicate, due to the notion of the *a priori* employed. The epistemic version has the same structure as the modal one, where 'K' is an a posteriori truth, and the worry is that '$N\phi\leftrightarrow\sim\phi$' would be logically true yet a posteriori. If we were to deal with

[12] The idea is brought up in John Etchemendy, *The Concept of Logical Consequence*, (Cambridge, MA: Harvard University Press 1990), and has received criticism therein and e.g. in Sher, "Did Tarski commit 'Tarski's fallacy'?" *The Journal of Symbolic Logic*, LXI, 2 (June 1996): 653-686. However, a more plausible version of the approach that model represent possible worlds, which avoids a large part of the criticism, appears in Stewart Shapiro, "Logical Consequence: Models and Modality," in M. Schirn (Ed.), *The Philosophy of Mathematics Today* (Oxford: Oxford Univerity Press 1998): 131-156.

the apriority of '$N\phi \leftrightarrow \sim\phi$' as we did with its necessity, we would have to appeal to something like *epistemically possible worlds*. For anyone who equates a priori truth with truth at all epistemically possible worlds, our previous argument modified appropriately would suffice. However, this equation is more contentious than that of necessary truth with truth at all possible worlds. Arguably, the notion of the *a priori* does not pertain to some type of possibilities where a sentence is true, but to the way the truth of a sentence is determined, justified or known.

The question then is how an a posteriori statement of the metalanguage appearing in a semantic rule of interpretation for a term in the object-language would affect the apriority of sentences containing that term. My first claim is that the answer to this question is not obvious. Any shared intuitions or defining features of the a priori pertain to natural language sentences, which are not primarily understood by us through rules of interpretation in a metalanguage.

Further, suppose that 'K' is an a posteriori statement in the metalanguage, and the object-language term 'N' is defined again by the semantic clause:

$$\text{(N) } \forall s(s \text{ satisfies} \ulcorner N\phi \urcorner \leftrightarrow [(\sim s \text{ satisfies } \phi \wedge K) \vee (s \text{ satisfies } \phi \wedge \sim K)])$$

McCarthy contends that since our knowledge of the truth of the sentence '$N\varphi \leftrightarrow \sim\varphi$' depends on the correct application of rule (N), and thus on knowledge that K, it is therefore a posteriori. However, it's not obvious that the a posteriori status of a statement in the metalanguage confers its status to sentences in the object-language. As pointed out with regard to the modal version of the criticism of the invariance condition, 'K' is part of what determines the meaning of 'N'. It is not the truth of '$N\varphi\leftrightarrow\sim\varphi$' that depends on the a posteriori (or contingent) truth of 'K', it is its *meaning* that depends on 'K'.

Here, we can make an analogy between rules in a metalanguage to procedures by which natural language is learned. It is uncontentious that the lexicon of a natural language is learned in an a posteriori manner. A child might learn the meaning of "three" through the repeated use of the word in contexts where a group of three apples, balls, or fingers are presented. An adult might learn the meaning of "three" using a dictionary. These are just plausible examples, not psycho-linguistic theses. The point is that experience is used when learning a language. Any plausible account of the a priori should take this into account. "Three is the number of apples in the bowl" is an a posteriori sentence by which the meaning of three might be learned, yet this does not affect the apriority of arithmetic statements about the number three. Analogously, 'K' being a posteriori does not entail that '$N\varphi\leftrightarrow\sim\varphi$' is a posteriori.[13] The same analogy would hold as a counterargument to Hanson's use of 'Q^*' presented in the introduction.

[13] Another option is to make the analogy between rules in the metalanguage and procedures of reference fixing along the lines of Kripke's theory of reference. In those procedures, the way reference is fixed might well be contingent and surely a posteriori. But this does not entail that sentences with terms whose reference is fixed in such procedures are themselves contingent or a posteriori. I thank Gila Sher for suggesting this option.

At this point I do not wish to make the further positive claim that '$N\varphi \leftrightarrow \sim\varphi$' *is* a priori. This would require saying much more about the a priori than I have so far. But as I hope I have shown, the modal and epistemic arguments do not stand.

3 Conclusion

McCarthy's conclusion from his argument is that "the logical status of an expression is not settled by the functions it introduces, independently of how those functions are *specified*".[14] My objection in this paper was not aimed towards this conclusion, but rather to the way this conclusion was argued for. It is perfectly reasonable to impose restrictions on the way logical systems are formulated. Thus, it is consistent with invariance criteria to add a further condition, by which the rules of a logical system should not contain contingent or a posteriori statements.[15] But the arguments presented in the literature in support of such a condition do not achieve their intended aim. As we have seen, these arguments rely on an invalid inference from the modal or epistemic status of statements in the metalanguage to that of statements in the object-language.

Further, we may add that the philosophical intuitions and theorizing about the subjects of necessity and the a priori mainly pertain to natural language, which isn't normally conceived of as having set rules in a metalanguage. Thus, any reasoning based on formulations in a metalanguage should make a clear and well-argued connection between

[14] McCarthy, "The Idea of a Logical Constant," *op. cit.*, p. 516.
[15] *Cf.* Sher, "The Formal-Structural View of Logical Consequence," *op. cit.*, p. 256.

metalinguistic rules and the modal and epistemic status of sentences in the object-language, a task which the critics I discussed have failed to fulfill.