University of Puget Sound
Sound Ideas

Summer Research

Summer 2016

What Do We Mean by Logical Consequence?

Jesse Endo Jenks
University of Puget Sound, jjenks@pugetsound.edu

What Do We Mean by Logical Consequence?

Jesse E. Jenks *

September 2016

Abstract

In the beginning of the 20th century, many prominent logicians and mathematicians, such as Frege, Russell, Hilbert, and many others, felt that mathematics needed a very rigorous foundation in logic. The standard approach in the early part of the 20th century was to use a syntactic or proof-theoretic definition of logical consequence. This says that "for one sentence to be a logical consequence of [a set of premises] is simply for that sentence to be derivable from [them] by means of some standard system of deduction"[3, pg. 65]. However, many famous results of the time, especially Gödel's incompleteness theorems led to logicians such as Tarski to define logical consequence with what was eventually developed into the standard "model-theoretic" definition. This way of defining logical consequence says that a argument of a certain form is a logically valid argument if it is *impossible* for the premises to be true and the conclusion false[9, 1, pg. 5, pg. 12]. Many philosophers have written about the effectiveness of this definition, but in 1990 John Etchemendy offered a fundamental criticism of Tarski's definition both as to whether it is conceptually correct, and whether it captures the right set of arguments, or interpretations. This paper explores Etchemendy's argument and various responses from prominent philosophers.

1 Introduction and Background

In the past century, the landscape of modern logic has grown into an extremely wide and varied field[see 2, ch. 1]. In fact, William H. Hanson begins his paper *The Concept of Logical Consequence* by noting that if one were to try to decide if experts in the field of logic were generally in agreement, "one is tempted to conclude that the foundations of the subject are in disarray."[6, pg. 366] But what exactly are logicians trying to do? Generally speaking, they are trying to define precisely what it is for a sentence (a conclusion) to be a *logical consequence* of some other sentences (the premises), and the various implications of that definition. There are, I think, two natural questions which come up. First and foremost, what is logical consequence? The second one is, essentially, why should we care? A majority of this paper (as the title suggests) will focus on the first question, but I think the second one is important to consider as well. What exactly is the point of all of this and what are the goals we hope to achieve? I believe there are two main goals in the field of logic. The first is to try to characterize how we as humans reason, and how

*I would like to thank my advisor Professor Douglas Cannon for his generosity and support.

to reason (well). The second goal is to capture the apparent a priori nature of this reasoning. As Benson Mates puts it, "Is logic about the way people think, the way they ought to think, or neither of these?"[9, pg. 3] The idea that logic is a priori is very widespread in philosophy. As Hanson says, "Logic, as with mathematics, has long been considered a science in which problems can in principle be solved just by thinking about them. And many logicians would agree that the relation of logical consequence holds between premises and conclusion of an argument only if a rational agent can come to know, independently of experience, that it is impossible for the premises to be true and the conclusion false."[6, pg. 377] If there is an objective nature to logical consequence, we would hope that our reasoning coincides with it. That is, our a priori or intuitive understanding of logic coincides with 'actual' logic, whatever that may be.

One of the reasons I believe this second question is important is because many logicians, including Tarski emphasize the importance of the 'pretheoretic' (intuitive) notion of logical consequence. However, if logic is meant *only* to characterize reasoning, it seems quite circular to require any reasoning in the characterization. That is to say, if a determination of logical validity[1] cannot be done without reasoning, logically validity cannot (itself) be a characterization of reasoning. In that case, only proof-theoretic methods could possibly satisfy this. This is essentially what the philosophers and mathematicians like Hilbert were trying to do in the early 20th century [2] Perhaps this circularity of reasoning is not a problem after all. Dictionaries, for example, define the meaning of words in terms of other words. This seems highly circular, but this seems to work because we think of the meaning of words as being independent of the word itself[3]. Similarly, sentences in two different languages may have a different grammar and syntax, but they can share the same meaning. In the same way, we can allow this circularity of reasoning since, just as with the dictionary, we have no other way of communicating the underlying ideas. However, it seems that logical consequence can be understood (the 'rational agent' really can arrive at a conclusion) only if logic is indeed an (objective) metaphysical property of the universe and, coincidentally, human reason (say, for evolutionary reasons). Essentially my point is this: if logic truly is independent of subject matter, and is an objective, fundamental part of reality, it must also be independent of our ability to understand it. Simply put, realism of logical consequence does not imply a priority of logical consequence. If anything, it gives us reason to believe the opposite. If, on the other hand, one is not a realist about logical consequence, the whole subject reduces to the study of various kinds of necessary consequence. Both of these views could explain the widespread disagreement about the nature of logic.

As for the answer to the first question, so far, nobody can seem to agree on anything. More specifically, while there are widely accepted technical systems of logic, there are many disagreements are about whether these systems capture the pretheoretic ideas involved. Logical consequence is generally thought to be a relation between a set of premises and a conclusion. But exactly what conditions need to be met in order for a conclusion to be a logical consequence of the premises is at the heart of the debate in modern logic, and the source of most, if not all disagreements. However, there are some common themes, though not everyone agrees with these either. One of the most common ideas dates back to Aristotle and his theory of the syllogism. The idea is that it is not the actual content of an argument, but how the argument is presented, or the form of an argument

[1] An argument is logically valid if and only if (iff) the conclusion is a logical consequence of the premises.

[2] Do Turing's and Church's proofs mean that logic is not actually about reasoning?

[3] Whether words like 'triangle', 'three', or 'soul' are meaningful may be problematic due to the questionable nature of their existence. But if they mean anything at all, it should be independent of the words in their definition.

that determines its validity. For example,

All books are blue.
All blue things are cold.
Therefore, all books are cold.

Clearly the first two sentences are false. I only need to look at a single red book or a hot blue object to see that. Yet most people would agree that *if* the first two sentences were true, the third sentence would *have to be* true. On the other hand

Some books are blue.
Some blue things are cold.
Therefore some books are cold.

Most people would agree that all three of these sentences are true, yet the third sentence is not a *logical consequence* of the first two sentences. You may also notice that the fact that we were talking about books seems irrelevant to the first argument. In fact, the argument remains valid even if we replace 'book' with 'chairs' or 'blue' with 'heavy'. In some sense, it is the *same* argument. On the other hand, if we replace 'books' with 'flames' in the second argument, our premises remain true, while our conclusion becomes false. This is the basic intuition behind what is called 'The Hypothesis of Formal Logic', which states that "Whether or not a conclusion follows logically from some premises depends solely upon the form of the argument they compromise."[1, pg. 8].

2 Proof theory

Before we explore Tarski's definition of logical truth and logical consequence, we should get a little historical context. For thousands of years mathematicians have been primarily concerned with proofs. A proof in mathematics sits within an axiomatic theory, and the idea is that results are deduced through repeatedly applying simple rules, called "rules of inference" starting from what seem like obviously true premises (the 'axioms' or 'postulates'). When we combine them using the simple rules, we get a new statement which, if we agree with the axioms and the rules we used to combine them, it seems, must be true. The most famous example is in book I of Euclid's *Elements*. In it, he states 5 postulates for geometry from which all results follow. One potential drawback of this approach is that the truth of the derived results rests entirely on the truth of the original axioms. This became especially clear in the 19th century when it was discovered that Euclidean geometry would remain completely consistent when we changed the 5-th axiom (the so called 'parallels postulate'). This brought into question the a priori nature that the axioms were meant to have. Euclidean geometry was supposed to *be* geometry[4], what else could there be? While this had a significant effect on the philosophy of physics, we could still feel safe in the nature of logical deduction. After all, the consistency of non-euclidean geometry was directly due to the consistency of euclidean geometry. In fact, is was in response to these discoveries of 'non-euclidean' geometries that inspired David Hilbert to try to axiomatize geometry. Unlike Euclid, however, he had 20 axioms!

[4]There were other kinds of geometry before the 19th century such as projective geometry, but for the most part, this was mainly developed as a tool for artists to accurately depict perspective, not a fundamentally different kind of geometry.

While mathematics has been the basis for the popularity of proof theory, it was eventually applied to logic. Naturally the most familiar rules of inference are mathematical ones, though we typically don't think of them this way. For example the sum of two integers is another integer. This idea that we intuitively know that this simple rule will always hold is the basic motivation for proof theory. If we have an intuitive understanding of the way the integers "really" behave, then this rule formalizes our intuition. A similar idea can be found in sentential or propositional logic. For example, one of the simplest ways to combine two true statements it with the word 'and'. Again, we don't typically think of 'and' as referring to a rule of some sort, but just as with addition of integers, we have the intuition that if 'p' is a true sentence and 'q' is a true sentence, then 'p and q' is also a true sentence. This gives us the rule "Given p and q as premises, we can derive $p \ \wedge \ q$" (where $\wedge$ is the logical 'and' symbol. Sometimes the more familiar ampersand, & is used.). When we combine sentences with the standard connectives and operators $\{\wedge, \vee, \rightarrow, \leftrightarrow, \neg\}$, we get, what are called formulas. However, we cannot combine them in any arbitrary way, such as $\wedge pq$, but if a formula is constructed by combining sentences in the appropriate ways[5], we get what is called a 'well-formed formula' or wff for short.

Propositional sentences however, are only the simplest building blocks for sentential logic. If we combine two wffs with the standard connectives, we get another wff. This means that wffs can become arbitrarily complex. Now we can look at proofs. Proofs are essentially instruction manuals on how to derive the wff we are trying to prove (the conclusion) from the wff we have as premises by repeatedly applying the rules of inference. These rules allow us to manipulate the wffs by, say by joining two sentences with $\wedge$. For example we could derive $p \wedge q$ from the set of premises $\{p, q\}$, but not from the set $\{p\}$. We often write this as $p, q \vdash p \wedge q$ meaning $p \wedge q$ is derivable from the premises p and q. Following all these rules, we are guaranteed that, given true premises, any conclusion we derive through application of these rules of inference is also true. Thus any argument with a conclusion that can be derived from the premises is valid. This idea that if we are given true premises, we will only derive true conclusions is called "truth preservation" and is an important concept in logic. More generally, truth preservation is the property of an argument in which conclusion is never false when the premises are true. This leads us to model theory.

3 Model Theory

Model theory, on the other hand, can determine if an argument is *not* valid. This is done by finding an instance in which all the premises are true and the conclusion false. In other words, it violates truth preservation. If we can find a case which violates truth preservation, we call it a counterexample. The second argument given at the beginning of the paper is an example of an argument which is not valid, precisely because we could find an argument of the same form which had true premises and a false conclusion. How we specify the form is intuitively clear in these simple cases, but how we generalize this idea is still contested. The simplest examples come, once again from sentential logic. In proof theory, we equate validity with derivability. That is, an argument is valid if there is a proof of the conclusion from the premises. This implicitly guarantees truth preservation. However, with model theory, we directly check that arguments are truth preserving by seeing if the conclusion is true whenever the premises are true. We do this by assigning truth values to each atomic

[5] Outlined in [1, ch. 3-5] and most introductory texts.

sentence, and defining the connectives as binary truth functions. We can encode this information in a more concise table, called a 'truth table'. For example, with the connective '$\wedge$'(and):

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

We can see that $\wedge$ is truth preserving since the top row, which represents the case when p and q are both true, the resulting sentence $p \wedge q$ is also true. Thus, as we would expect, under this simple model theory $p \wedge q$ is a logical consequence of p, q. To distinguish this model-theoretic notion of logical consequence from proof-theoretic consequence, we write $p, q \vDash p \wedge q$.

What I did not mention in the previous section is that formal languages can become much more complex than sentential logic. The motivation for this is that we can formalize much more complex sentences but requires a bit more mathematics. The standard model theory sits in a broader set theory, such as ZFC[11]. This set theoretic foundation of model theory has become extremely standard and was first developed for what are called first-order theories. These are a large class of different logical systems capable of defining validity for a large variety of languages. First, we specify a non-empty set called a "universe of discourse" D. A model or interpretation[6] of a first-order language then assigns elements of D to variables, subsets of D to predicates (called "extensions" of the predicates) and quantifiers[7], while subsets of D^n (ordered lists of n elements from D) are assigned to n-ary relations. For example, we could have the binary relation "taller than" which is assigned pairs of elements which satisfy this relation, e.g., the pair <Dan, Dave>, meaning Dan is taller than Dave. 'Truth' of sentences like "Neil Armstrong walked on the moon" in this context, reduces to simply whether the element in D assigned to the name "Neil Armstrong" is in the extension of the predicate "walked on the moon" in a particular model. A sentence is logically true if it is true under all interpretations, i.e. true in all models.

One natural question that arises when comparing these model theoretic and proof theoretic techniques is how can we be sure that they will give rise to the same results? In other words, how can we be sure that if we have a proof of a formula we will not be able to find a model that serves as a counterexample? For first-order theories, we have what are called 'soundness' and 'completeness' theorems. These theorems tell us that for every proof we will not have a counterexample, and if we don't have a counterexample, then there is a proof. Specifically, it is fairly straightforward to establish that any first-order logic with a model theory has a proof theory, i.e. will not produce a contradiction. However, the question of whether a proof theory is enough to guarantee a model was much more complex. A full proof of completeness for first-order theories was first established by Kurt Gödel in 1929[8]. However, he later showed that if we have a more powerful, higher order logic, specifically one that allows us to make arithmetical claims, Gödel's famous first incompleteness theorem tells is that, while we can have soundness, we cannot have completeness. This result was no doubt one of the most significant results of the 20th century and has had a lasting impact in

[6]Some people make a distinction between a model and an interpretation, but that will not be discussed here.

[7]Quantifiers are expressions like 'everyone' or 'something'. Set theoretically they are interpreted as saying "for all elements in the set of people, _" or "There exists at least one object in the set of things such that _". These are called "universal" and "existential" quantifiers respectively and will play a big role in Etchemendy's criticism.

[8]Letting Γ represent a set of premises and σ the conclusion, we can symbolize Gödel's result as "$\Gamma \vDash \sigma$ iff $\Gamma \vdash \sigma$".

philosophy. This result, as well as results from Turing and Church led many to try to re-exam how we understand logical consequence. This result, in some ways "freed" model theory from proof theory.

In this context, it is not clear what particular models are supposed to depict. For example, if <Dan, Dave> is a an element of the extension of the "taller than" relation, <Dave, Dan> should not be assigned as it is arguably impossible for Dan to be (strictly) taller than Dave, and for Dave to be taller than Dan. But there is no such restriction in this set-theoretic model theory. This led, according to Etchemendy, to two distinct ways of thinking about model theory . To illustrate the difference, consider the sentence "Neil Armstrong walked on the moon". We could view a model which assigns the person Neil Armstrong to the name "Neil Armstrong", but does not assign him to the set of people who walked on the moon, i.e. assign him to the extension of the predicate "walked on the moon" as representing a possible world in which Neil Armstrong did not walk on the moon. On the other hand, if the composer John Cage is assigned to the name "Neil Armstrong", we could view this as reinterpreting the name "Neil Armstrong". In either case, the sentence "Neil Armstrong walked on the moon" will be false, but for very different reasons. Etchemendy calls these two fundamentally different ways of understanding model theory *representational* and *interpretational* semantics.

Although interpretational semantics can be introduced through the standard set-theoretic model theory, it is important that we distinguish them. But before we do, we need to look at substitutional semantics. Substitutional semantics is based on the observation of Aristotle given at the beginning of the paper. The arguments about books and blue objects were valid or invalid, not because they were about books, but in spite of being about books. That is to say, the argument was truth preserving even after substituting the word "flames" for "books". After Aristotle, this idea of trying to capture the form of an argument by analyzing the generalized form of the argument was further developed by monks in the middle ages[1], but was developed with the modern formalization by Bolzano.

Tarski was not aware of Bolzano's work, but their work shares some important intuition. If we see an argument and we believe that it is a logically valid argument, we can see if the argument is truth preserving when we substitute terms. One significant difference between the standard model theory and both Bolzano and Tarski's accounts is that they relativize their account to an arbitrary selection of "fixed" terms, or "logical constants". Sher calls this the "second fallacy of material consequence". In the standard model theory, these would be the connectives and other logical terms we had before. One of the reasons for this is because it is difficult to define exactly what makes some terms 'logical' and others not. It is easy to see that we need to have some selection of fixed terms, otherwise every argument would be deemed invalid. Similarly, we need some variable terms, otherwise every sentence remains the same under any interpretation, and we would simply be asking if a sentence is (materially) true.

To avoid confusion, Tarski introduced the idea of a "metalanguage" and an "object language". The metalanguage allows us to talk about the object language in question without confusing descriptions *about* a language with statements *in* the language. This allows us not only to talk about an object language $\mathfrak{L}$, but also *expansions* of the language. One of the main reasons for this is due to what Sher calls "the first fallacy of material consequence" for substitutional semantics. Hanson mentions that Quine also disagreed with the substitutional account of logical consequence since it could produce "false positives" in a limited language. Essentially the problem is this: In substitutional semantics, we look at all possible substitutions of variable terms. This means that if we are looking at a particular language $\mathfrak{L}$ with some finite collection of names. If we use Etchemendy's example, the set of all U.S. presidents, a sentence like "If Abe Lincoln was a man then Abe Lincoln was president", will be deemed logically true by the substitutional account. That is to say, if we replace both instances of "Abe Lincoln" with any other name in our language, we get a new sentences which is true as a matter of fact (materially true). Clearly this is not a logically valid argument, despite the fact that (pending the upcoming election) there are no counterexamples. Of course, this problem would not have come about if we had not restricted ourselves to U.S. presidents. However, there some some instances in which our language could not possibly have enough names. For example, if we had a statement about real numbers, but only had countably many names, we could never establish the validity of even the simplest algebraic statements. This peculiarity of an argument which is deemed logically valid in a substitutional semantics but *would have been false* had we had enough names is what Sher calls the "first fallacy of material consequence", and Etchemendy calls a 'violation of persistence'. This is because logical validity should 'persist' through expansions of (say) the class of names in a language. But this is also due to the fact that we designated 'Abe Lincoln' as a variable term, while we held some (traditionally) non-logical predicates like "was president" and "was a man" as fixed terms. But, even if we only fix traditionally logical expressions, we still get problems.

Tarski begins his article *On the Concept of Logical Consequence* by saying "The concept of *logical consequence* is one of those whose introduction into the field of strict formal investigation was not a matter of arbitrary decision on the part of this or that investigator; in defining this concept, efforts were made to adhere to the common usage of the language of everyday life."[14, pg. 409] Clearly Tarski believed that the proof-theoretic systems that had been developed did not agree with "the common concept of consequence". In another article, Tarski talks about a concept introduced in Gödel's paper on his incompleteness theorem, known as ω-inconsistency. Tarski develops an analogous system which is ω-incomplete and he argues that the existence of systems like this are a reason to reject proof-theoretic systems[9].

This led Tarski to develop one of the most important concepts introduced in this paper he called "convention T". Tarski argues that "[no] definition of true sentence which is in agreement with the ordinary usage of language should have any consequences which contradicts the principle of the excluded middle[10]. This principle, however, is not valid in the domain of provable sentences. [For example, the disjunction of a sentence and its negation, neither of which is provable.] The extension of the two concepts is thus not identical. From the intuitive standpoint all provable sentences are without a doubt true sentences ... [t]hus the definition of true sentence which we are seeking must

[9] The system Tarski developed is a deductive theory in which we can prove each of the sentences '0 has property P, '1 has property P', ... and so on for each natural number. It seems reasonable that we could conclude
A : every natural number possesses the property P
That is to say, A is *clearly* a logical consequence of the premises A_0, A_1, and so on. But without an explicit axiom known as the *rule of infinite induction* telling us that we can derive A given that we have derived $A_0, A_1, \ldots$, we can't. Even if we define a new sentence B as "It is possible to prove $A_0, A_1, \ldots$" and so on for every natural number, this would require an infinite sentence which Tarski tries avoid. However, supplementing the system with this axiom, in light of Gödel, is not satisfactory for Tarski, nor does it avoid the incompleteness results.

[10] The law of excluded middle is the idea that all propositions are either true or false, and there is no "middle", or alternative to true and false. One of the reasons for this is the existence of undecidable problems. The most famous example in set theory is known as the continuum hypothesis. In 1940, Gödel showed that the continuum hypothesis was consistent with the axioms of ZFC, but in 1963, Paul Cohen showed that its negation was also consistent. This means the axioms of ZFC are not powerful enough to prove either statement.

also cover the sentences which are not provable. ... Using the symbol *Tr* to denote the class of all true sentences ...

> Convention T. *A formally correct definition of the symbol 'Tr', formulated in the metalanguage, will be called an* adequate definition of truth *if it has the following consequences:*
>
> (α) *all sentences which are obtained from the expression '$x \in$ Tr if and only if p' by substituting for the symbol 'x' a structural-descriptive name* ["names which describe the words which compose the expression denoted by the name"[13, pg. 156-7]] *of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage;*
>
> (β) *the sentence 'for any x, if $x \in$ Tr then $x \in S$' (in other words '$Tr \subseteq S$').*

[13, pg. 186-8] where S is the set of sentences in the object language. This marks a significant departure from the usual syntactic definitions promoted in the mid to late 19th century. Tarski uses this as a guideline for defining a 'truth predicate', or a way of saying x is true. In defining truth this way, the consistency of the metatheory/metalanguage and a guarantee of (every instance of) convention T is a guarantee of consistency in the object language, one of Tarski's main concerns.

It is important, I think, to note that what Etchemendy is arguing against is what has become the standard model theory based on the work of Tarski on logical consequence (what Priest calls the "Tarskian Orthodoxy"[10]). But in a 1988 article[3], Etchemendy emphasizes an aspect of Tarski's work which has been confused in the modern formulation of model theory. Tarski's real goal in defining convention T was to provide an analysis of truth, not to study what is now called formal semantics. convention T is used only as a way to understand what we mean by "truth". And, although the recursive way in which Tarski defines consequence has become standard[see 9] "[A] definition of truth need not give us any effective procedure for deciding or enumerating the set of truths"[3, pg. 55]. In fact Tarski gives an explicit example of a finite language in which a list-like definition of truth, which gives us no semantic information whatsoever, is able to satisfy convention T[13, pg. 188]. (A similar point is made in footnote 4 of Kreisel[8, pg. 87] in reference to defining inclusion in the set of integers.) This, Etchemendy argues, clearly shows that in defining his 'adequate definition of truth' Tarski was distinguishing between a definition of truth and the semantic theory which came from his analysis of logical consequence. It is easy to see why this confusion can come about. For example, the sentence

> 'Snow is white' is true if and only if snow is white

can be understood in two ways. In defining convention T, Tarski was trying to use sentences like this (what he called "partial definition of truth") to give us a better understanding of what the word "true" means. But intuitively, we already know what true means, especially with these simple examples. Under this second view, convention T is an almost trivial observation. Instead, this would give us a way of understanding what the phrase 'Snow is white' means, instead of telling us what "true" means. For example, Tarski says p is the translation of x into the metalanguage. So if our object language contained Japanese, but our metalanguage contained only English, we would have

> '雪は白い' is true if and only if snow is white

which is even more misleading. This makes it look like we are trying to understand the meaning of the Japanese phrase in terms of the English one, when really this is a (partial) definition of truth with a translation required only due to the choice of languages. So "[g]etting from a Tarskian definition of of truth to a substantive account of the semantic properties of of the object language may involve as little as the reintroduction of a primitive notion of truth."[3, pg.59] In fact, Etchemendy claims that Tarski was trying to remove semantics from defining truth altogether. This, Etchemendy argues, has been the cause of much confusion surrounding Tarski's work in modern model theory. Both Hanson and Sher talk about the "semantic tradition begun by Tarski"[7, pg.393] which seems to confirm Etchemendy's argument at the beginning of his article.

The motivation for Tarski in defining Convention T was straightforward. We can't even talk about truth preservation unless we know what 'true' means. Etchemendy seems to take this to mean that the definition of logical consequence Tarski was outlining was a material one, and Tarski himself said that logical consequence had to be materially adequate. But notice that the definition given does not explicitly require that the sentences in the metalanguage be materially true. That is, the sentences need not correspond to the actual world.

4 Etchemendy's argument

There were many responses to Tarski's account. But one of the most influential criticisms of Tarski came from John Etchemendy over fifty years later with his book, *The Concept of Logical Consequence*. Etchemendy begins by outlining the difference between "truth in a language" and "truth in a world". These are the interpretational and representational semantics discussed earlier. Instead of defining them in terms of set theory, however, we construct them as a new kind of model theory. Etchemendy does not go into detail about representational semantics since this leads to some metaphysical problems. As Sher says," [T]he model theory of metaphysical [representational] semantics is couched in a background theory based on general metaphysics... The strategy of vindicating Tarski's conjecture by developing a metaphysical [representational] semantics is, given the daunting problems facing general metaphysics, not feasible."[11, pg. 659-61]. Etchemendy calls this approach representational semantics, because it represents various ways the world could have been. Specifically, "[T]he class of models should contain representatives of *all* and *only* intuitively possible configurations of the world."[4, pg. 23] Sher points out that this is not a very clear or well defined notion. For example, she asks whether it is possible for her dog Cerberus to be New York City. Is it possible for Cerberus to be rational? to be moral?[11] Hanson also brings up this question when citing Kripke's examples of the inferences "If x is red then x is colored" or "If x is water, x is H_2O." This question of what, if any bounds there are on metaphysical possibility is very important if we want to pursue the representational approach.

Etchemendy claims, however, that Tarski's real goal (or what is closer to his real goal) was what he calls 'interpretational semantics'. This is the "truth in a language". Here, we hold the world fixed and vary the meaning of words. He is careful to point out that these two ways of understanding model theory are not simply two different perspectives on the same thing (i.e. they are not "extensionally equivalent", meaning they don't always produce the same results). For example, if we consider the pair of sentences "Roses are red" and "roses are yellow", under an interpretational

[11] This seems to be more of a question of identity. Is it possible for me not to be me?

semantics, if we interpret the words “roses”,“red”, and “yellow”, to mean, say cherry pie, warm, and delicious respectively, both sentences can be true (or false if you don’t like cherry pie). On the other hand, it would be (arguably) impossible for (all) roses to be both red and yellow simultaneously in any possible world. That is to say, this is a metaphysically impossible situation.

The way Tarski avoids the first fallacy of material consequence is through the metalanguage. Instead of checking whether a sentence is true after substitution, we introduce the concept of satisfaction. Satisfaction is a relation that holds between elements in the object language and what are called sentential functions. Sentential functions are now usually called “open formulas”, and are essentially sentences where every variable term (such as ‘Abe Lincoln’) is replaced by variables. Saying a name, or a *sequence* of terms satisfies a sentential function is analogous to saying “The golden ratio’ satisfies the formula $x^2 - x - 1 = 0$’, where x is a variable, and ‘the golden ratio’ names a specific number ($\frac{1+\sqrt{5}}{2}$), while English words and symbols like 2 and $=$ are fixed. The metalanguage allows us to describe exactly how this relation holds, and unlike the substitutional method, our metalanguage allows us to describe what would have been expansions of our object language in the substitutional account. Instead we define ‘satisfaction domains’ for each grammatical category. For example, the ‘name domain’ for the satisfaction relation is made up of “*Individuals*, things that *could* have been picked out by names”[4, pg. 41], or the predicate domain as a collection of objects which assert possession of a property. As you can see, it would be easy to describe this in terms of a set or class. In fact Sher, in recounting Etchemendy talks about these satisfaction domains as “maximal substitution classes”. However, these satisfaction domains cannot be defined just as objects which obey the same grammatical rules, since, for example ‘Pegasus’ can act as a name grammatically, but not semantically. For this reason, Etchemendy requires that these domains be “semantically well-behaved”. Just as the expansions of the language in substitutional semantics needed to be “semantically well-behaved”, the members of each satisfaction domain must behave semantically similar. That is, they “contribute to the truth value of the sentence in a similar way”. Unfortunately, this requires us to delve into philosophical problems which were avoided by proof theory, since it relied on syntactic definitions, rather than semantic ones. However, we can now think about functions which take elements from these satisfaction domains, and assign them to the corresponding variables in our sentential function. That is to say, if we take the sentence we had before,“If Abe Lincoln was president, then Abe Lincoln was a man.”, fixing everything except ‘Abe Lincoln’, our corresponding sentential function is “If n was president then n is a man”. We can then define *models* for our interpretational semantics. A model is a particular assignment of elements from our satisfaction domain onto the variables in our sentential function that satisfies the sentential function[12]. We can now use this concept of an interpretational model to define logical consequence! Here is Tarski’s original definition:

> *The sentence X* follows logically *from the sentences of the class K if and only if every model of the* [sentences in the] *class K is also a model of the sentence X*.[14, pg. 417]

One of the key arguments Etchemendy makes against Tarski’s analysis is what he calls *Tarski’s fallacy*. Since the Tarskian, interpretational account of consequence is meant to capture our intuitive understanding of consequence, we can state the following:

[12]Etchemendy actually defines models slightly differently so that we can talk about sentences being consequences of other sentences rather than sentential functions.

(L') If S is a consequence (in the ordinary sense) of K, then S is a Tarskian consequence of K

where S is a sentence, and K is a set of premises. Tarski's analysis would then agree with our intuitive understanding of consequence if we could also show the converse of (L'). In trying to do so, he argues that what Tarskian analysis can show is that the following three statements are jointly incompatible:

(1) S is a Tarskian consequence of K (for some $\mathfrak{F}$).
(2) All the members of K are true.
(3) S is false.

where $\mathfrak{F}$ is a set of 'fixed terms' we treat as logical constants. Traditionally these are terms such as 'and' or 'not'. Tarski's fallacy is to then draw the conclusion from (1), that is, given that S is a Tarskian consequence of K, we can say that the following *two* statements are incompatible:

(2) All the members of K are true.
(3) S is false.

Gila Sher gives a nice recapitulation in her paper in terms of modal operators. "Etchemendy takes Tarski's proof to have the following structure: Assume

$$\text{(i)}\ \ \Gamma \vDash \sigma$$
$$\text{(ii)}\ \ \neg(\text{All the sentences of } \Gamma \text{ are true } \rightarrow \sigma \text{ is true})$$

... This yields a contradiction since (i) says that the argument is true in all models, while (ii) says that our actual world serves as an example of a model in which the argument fails [to preserve truth]"[11, pg. 654]. Letting $\Box$ be a "modal operator" meaning "necessarily", this would show that

$$\Box\,(\Gamma \vDash \sigma \rightarrow \text{All the sentences of } \Gamma \text{ are [actually] true } \rightarrow \sigma \text{ is [actually] true})$$

"Tarski' fallacy" would then be to go from the previous formula to the following:

$$\Gamma \vDash \sigma \rightarrow \Box(\text{All the sentences of } \Gamma \text{ are true } \rightarrow \sigma \text{ is true})$$

Sher argues that since Tarski never provided a full proof of the adequacy of his analysis, the idea that he committed this fallacy is pure speculation. (Sher actually provides a short proof of the adequacy of Tarski's analysis for first order logic which avoids Tarski's fallacy.) Instead, Sher believes that "Tarski's pretheoretic notion of logical consequence involves two ideas: the idea that *logical consequence is necessary* and the idea that *logical consequence is formal*."[11, pg. 654][13] Sher agrees that Etchemendy is correct in concluding that the interpretational account he gives does not agree with our understanding of validity. However, she does not agree that an interpretational semantics cannot be made into account that does not avoid the problems presented. One of the problems raised by Etchemendy is the possibility of necessary, but not logical truths. The examples he gives are related to the size of the (actual) universe, though there are other universal closures

[13] I agree with Sher in that a priority is not a property we should require of logical consequence. This is opposed to Hanson's view who specifically requires a priority.

which depend on extralogical facts (Sher's third fallacy). These examples are deemed logically true because they are composed entirely of logical constants, and thus are true under all interpretations if they are materially true. This is essentially Etchemendy's argument against, what he calls the "reduction principle". Sher, on the other hand, argues against this material view, as does Priest[10, pg. 286]. While she does not deny that logical consequence should be materially adequate, as Tarski says, she instead says it must be "intuitively formal and necessary"[11, pg. 667]. Etchemendy, like Tarski and Hanson, does not believe there is any single set of properties that distinguishes a logical constant from any other term. Of course the standard logical connectives seem to work most of the time, they don't seem to work all the time. Sher, on the other hand, believes that logical consequence should be formal. And in order for this formality to hold, there has to be a, well, form. Thus she develops a (somewhat technical[14]) definition for *formal operators*, and in turn, logical constants.

One interesting thing to note is that, as I mentioned in footnote (12), Etchemendy actually changes his description of Tarski's account slightly to fit the modern model-theoretic semantics. In a 1988 article *Tarski on Truth and Logical Consequence*, he argues that Tarski's real goal was not formal semantics (at least, not directly). Instead, Tarski chose to use semantics, and as a result, a recursive definition in order to avoid the problem of infinitely many possible sentences in the object language, while having finite sentences in the metalanguage. However, there are other possible systems which will still satisfy convention T. An example he gives is the "list-like" definition of truth[3, pg. 55]. This is where he first introduces the idea of a recursive definition, right below convention T.). This seems to clearly indicate that Tarski was interested only in an analysis of truth, as can be seen by his convention T, and only afterwards does he decide to use a recursive, semantic definition of logical truth to accommodate the infinity of words in the object language.

One of the main problems Etchemendy has with interpretational semantics is due to quantifiers. Quantifiers are terms like "everyone" or "something". These clearly have a logical aspect to them, but traditionally they are treated like predicates, and their meaning varies over models. This means, while quantifiers are logical in some sense, they cannot be in the selection of fixed terms. However, this would deem sentences like "If Abe Lincoln was president, then someone was president", which is considered a very standard argument to be invalid if, for example the set of dogs is assigned to the existential quantifier "someone". But, both ways Etchemendy outlines of interpreting the existential quantifier leads to a restriction on the class of models. This, he argues, is a significant departure from Tarski's original analysis. The restrictions are that, under a particular interpretation, either Abe Lincoln, (or if Abe Lincoln is a variable term, whatever Abe Lincoln has been reinterpreted as) must be a member of the set assigned to the quantifier "someone". "If Fido was president, then some dog was president" or "If Fido was president, something in the set of dogs was president". In general, "if Pn then $\exists x(Px)$ is a valid inference iff $n \in f(`\exists')$, meaning n is in the set assigned to the existential quantifier. This restricts the class of models in a complex way (these are actually quite similar to the restrictions Sher places on representational models). Clearly the inference "If Abe Lincoln was president then some dog was president" is not a true statement. On the other hand "Abe Lincoln was president" and "Someone who fought in the revolutionary war was president" are both true. He offers various solutions to this problem, such as using "meaning postulate" and allowing cross-term restrictions to "tighten the 'semantic similarity' relation".

[14]The basic idea is an extension of work done by Mostowski on generalized quantifiers. If there is a bijection or isomorphsim between the universes, including onto itself (e.g. a permutation of a sequence) for different models, a logical operator is one which "does not distinguish between isomorphic arguments"[11, pg.677].

However, all of these seem to be circular, in that they are essentially equivalent to the two types of cross-term restriction he outlines. Perhaps the most obvious solution is, instead of assigning a subset of the universe to the existential quantifier, we take it to mean that there simply is something in the universe at all which satisfies the formula. In other words, we always assign the entire universe to the existential quantifier. This would not restrict the class of models, nor would it require any implicit predicate, so have we solved the problem? With this view of existential quantification, we still have a problem.

Etchemendy argues that formulas involving the cardinality of the universe (how many things there are), like

$$\begin{aligned}
\sigma_2 &: \ \exists x \exists y \, (x \neq y) \\
\sigma_3 &: \ \exists x \exists y \exists z \, ((x \neq y) \wedge (x \neq z) \wedge (y \neq z)) \\
&\vdots
\end{aligned}$$

or their negations cannot be a logical truth, because they depend on extralogical facts about the size of the universe, despite the fact that they are made up of only logical constants. Etchemendy calls the property of generating "false-positives" like this, "overgeneration", since the theory "generates" a logical truth when there shouldn't be one.

This overgeneration is caused by the material nature of interpretational semantics and leads Etchemendy to define what he calls the reduction principle. The material nature of interpretational semantics means that, universally quantified sentences which are materially true, that is they are true in every model, are deemed logically true. The basic idea is quite similar to the problem we had with substitutional semantics. For example, the sentence "Every president is a man" is universally quantified sentence in which every instance is materially true, but not logically true. Again this is due to a non-standard selection of fixed terms, but the problem persists with the standard selection.

In a model theory based in a set theory like ZFC[15], the falsehood of each $\neg\sigma_n$ is avoided by making use of infinite models. But this requires the use of the axiom of infinity, which Etchemendy says is not a matter of logic. At the very least, a material theory of consequence should not have any infinite models. Furthermore, Tarski was explicitly trying to avoid undecidable problems, which ZFC has been shown to have (see footnote 10). To have a model theory couched in a set theory with an axiom which we were intentionally trying to avoid seems to defeat the point altogether. This is because each σ_n is made up entirely of traditionally logical terms, so as long as σ_n is *materially* true, that is, there are actually n distinct objects, it is true under all interpretations, and is thus judged to be a logical truth. This, Sher argues, is one of the weaknesses of Etchemendy's argument. The modern mathematical logic/semantics is *not* an interpretational one. This is why she agrees that Etchemendy is correct in his argument against interpretational semantics, but does not agree that interpretational semantics is modern semantics. Instead Sher thinks that "ZFC ... is the right kind of theory for [Tarskian logic]"[11, pg. 676]. This view of a logic based in set theory has several immediate objections. First and foremost, if mathematics requires a rigorous foundation in logic, how could our theory of logical consequence be based on a mathematical theory?. But this view, argues Sher, is a foundationalist approach to logic, which she does not subscribe to. "[In t]he foundationalist view, ... logic is viewed either as something that does not require justification ... or as something that cannot be justified (i.e. we accept or discard, logic as arbitrary, accidental,

[15]This stands for 'Zermelo-Fraenkel set theory with the axiom of choice.

conventional). Either way, the foundationalist does not ask, and does not explain, why logic is the way it is"[11, pg. 680][16]. While mathematics and logic are interrelated, if logic is not required as a foundation for mathematics "logic requires a powerful background theory of formal structure as much as set theory requires a powerful logical machinery. ... [But w]e cannot reduce logic to mathematics and then reduce mathematics to logic."[11, pg. 680]]. Is it possible that there could have been no objects in the universe? Most people would argue that the existence of multiple objects in the universe is a contingent fact. Although, arguably even if no people are around to think about the, abstract objects exist and thus each of Etchemendy's σ_n would be, at the very least true of the natural numbers as abstract objects. Especially since Etchemendy believes that it is metaphysically impossible for sentences like "2+2=4" to be false. If he is not a realist about the existence of the natural numbers, I am not sure what his reasoning is, then to believe that "2+2=4" is a necessary truth, since, presumably he is arguing that in a representational semantics, each σ_n is false in some possible world. Thus if "2+2=4" is true in all possible worlds, at the very least the world should contain the natural numbers. That is, if arithmetical statements are necessarily true, they necessarily must exist. Does this explain the σ_n and $\neg\sigma_n$? This brings up the question of what 'materially true' means. If we take it to mean empirically or physically true, then interpretational semantics could not possibly make any mathematical claims. Many logicians make a distinction between logical and necessary consequence.

5 Logical vs. Necessary Consequence

The intuition that has guided much of the work done in logic is that if the conclusion is a logical consequence of the premises, then the conclusion is a necessary consequence of the premises. Now how do we talk about necessary consequence? The most common idea is called truth preservation. Other than formulas, there are two main schools of though on truth preservation. These are the interpretational and representational semantics. Truth preservation in proof-theoretic formulas and interpretational semantics is fairly similar. This says that a (propositional) sentence is made up of logical, or fixed terms, and variable terms. Thus a sentence can be turned into a kind of 'generalized' sentence, called a sentential function. Truth preservation in this case would be, in every possible substitution (interpretation, sequence, model) of the variable terms, whenever the premises are (materially) true, the conclusion is also (materially) true. Hence people (Sher) consider(s) this a characterization of material consequence. Or it is materially adequate, as Tarski says. Truth preservation in representational semantics on the other hand, simply requires that it is (metaphysically) impossible for the conclusion to be false while the premises are true (this is usually done with a 'modal' logic). Clearly this is much harder to deal with, as Sher talks about. There does not seem to be any theory of metaphysics which would allow us to definitively say one way or the other if an argument is metaphysically truth preserving. Is it metaphysically possible for "I am sitting on a bench and I am not sitting on a bench", or perhaps, "2+2≠4" to be true? Etchemendy argues that the latter sentence is a mathematical claim, and is necessarily false, since it is metaphysically impossible for 2+2=4 to be false. On the other hand, in interpretational semantics, since the set of fixed terms is arbitrary, 2+2=4 is false under various interpretations and selection of fixed terms. Of course, if we hold 2,4,+,=, to all be fixed as we traditionally do, it is also a logically true since it is (trivially) true under all interpretations.

[16]This is discussed in section 7.

One of the key distinctions which Etchemendy uses to argue against Tarski is the difference between necessary and logical consequence. As Gila Sher says, "[D]rawing a sharp distinction between [logical and necessary consequence] is not simple, and the attempts to account for the distinctive features of logical consequence have often taken too much for granted."[11, pg. 658] Clearly, if σ is a logical consequence of Γ, then σ must be a necessary consequence. However, simply being a necessary consequence is not sufficient to deem it a logical one. Hartry Field, on the other hand, argues that necessary truth preservation is not a good way to define validity at all. "The claim at issue ... is that genuine logical disagreement is disagreement about which inferences preserve truth by logical necessity ... one can make the account slightly more informative by explaining logical necessity in terms of some more general notion of necessity together with some notion of logical form, yielding that an argument is valid iff (necessarily?) every argument that shares its logical form necessarily preserves truth."[5, pg. 36, his question mark]. In his footnote 5 he mentions that of course, logical form is not a well defined notion and that, in light of Tarski, who says that 'there is no privileged notion of logical constant' and that what is considered a logical constant is really just a fixed term so logical form is relative to our set of fixed terms or logical constants.

Hartry Field suggests a different way of understanding validity. Instead of defining validity directly, he argues that the role of validity is in fact not about necessary truth preservation. One of the examples he gives concerns the Curry Paradox. Letting K be equivalent to the sentence "if K is true, then $0 = 1$", the paradox is that, given the equivalence, we can prove that $0 = 1$. However, the proof requires the use of conditional proof, and the common solution to this paradox is to reject this move. However, he then says that if one accepts this solution, but defines validity in a way that requires truth-preservation, we can reason from K to $0 = 1$ (as our solution to the paradox has no problem with this), but to say this is valid is to accept True($\langle K \rangle$)$\rightarrow$True($\langle K \rangle$), implying we accept the reasoning True($\langle K \rangle$)$\rightarrow$ $0 = 1$. So in defining validity in a way that requires truth preservation, we prevent ourselves from accepting our solution to the paradox. Thus Field concludes "[o]n the [necessary truth preservation] definition of validity, not only can good logical reasoning come out invalid, but fallacious reasoning can come out valid"[5, pg. 41]]. Instead he argues that the "conceptual role of validity" is, to accept that an argument is valid is to put constraints on your belief. He outlines a very general way to understand validity in terms of our degree of belief in each of the premises and conclusion(s). Letting $\Gamma \Rightarrow B$ mean "B is a genuinely valid consequence of Γ", where Γ is a set of sentences (premises), and letting $Cr(A)$ be a number between 0 and 1 representing our degree of belief in a sentence A (0 means we do not belive A at all, while 1 means we fully believe in A). Letting A_i be premises in Γ, we get the following[17]:

$$\text{If } A_1, \ldots, A_n \Rightarrow B \text{ then } Cr(B) \geq \sum_{i \leq n} Cr(A_i) - n + 1$$

I agree with Field that it is important to consider the 'conceptual role' of validity. This relates to the second question in the introduction. Unfortunately, Field doesn't mention any problems relating to monotonicity. Monotonicity is the idea that, if we have an argument in which B is a logical consequence of Γ, then B must also be a logical consequence of $\Gamma \cup \{A\}$, where A is any sentence. Intuitively this is straightforward, since the new assumption would simply not have any

[17]Equivalently, if we let *disbelief* in A, or $Dis(A) = 1 - Cr(A)$, we get
If $A_1, ..., A_n \Rightarrow B$ then $Dis(B) \leq \sum_{i \leq n} Dis(A_i)$

effect on our derivation of B. In other words new information should not change our belief in the conclusion. Field does talk about the role of evidence in our degrees of belief by making an analogy with conditional probability. What 'should' (in a "non-subjective" sense) our degree of belief be? However on Field's account, if we add a premise A_{n+1}, and we don't fully believe it, we would end up believing in B even less, even though the addition of A_{n+1} should not affect our belief in B. If we fully believe in all of the A_i, so $Cr(A_i) = 1$ for each premise, then Field's inequality tells us we must also fully believe in B. However, if we add the premise A_{n+1}, and we only believe it say to degree $\epsilon < 1$, then

$$\begin{aligned} Cr(B) &\geq Cr(A_{n+1}) + \sum_{i \leq n} Cr(A_i) - (n+1) + 1 \\ &\geq \epsilon + n - n - 1 + 1 \\ &\geq \epsilon. \end{aligned}$$

In general, if we fully believe each of the premises $A_1, \ldots, A_n$, while $A_{n+1}, \ldots, A_{n+k}$ are our 'extra' premises, then we get a lower bound that could potentially be any number less than or equal to 1. It seems odd that we are now not obligated to fully believe B. Of course, this can be avoided with a relevance logic in which the addition of unnecessary premises is not allowed. However, Field was trying to apply his idea a range of logics, so we need a simple modification: Let $\Gamma' \subseteq \Gamma$ be the smallest subset of a set of premises such that $\Gamma' \Rightarrow B$. This should provide the "true" minimum degree of belief in B, or at least, maximize it. If $\Gamma \setminus \{A_i\} \not\Rightarrow B$ for any $A_i \in \Gamma$, then Γ will give us the best lower bound. In other words, if removing a premise means that B is no longer a logical consequence, then we have exactly the best lower bound[18]. This is essentially what a relevance logic would do since, intuitively, the minimum set of premises should be only those that could (or should) affect the truth value of the conclusion. But Field's method avoids making reference to any kind of semantic notion. This is actually good news, because this means that even with extra premises, we don't fool ourselves into thinking we believe B more than we 'should'. Of course, this assumes that we can determine whether B is a 'genuine' logical consequence of a set of premises which is precisely what is in question.

Unfortunately, we cannot turn this around and define the minimum set of premises as the set that maximizes our belief. If we remove too many premises, say to leave us with only the premises we are most confident in, the greater our bound will be, but it will not give us the 'true' bound. In other words, adding extra premises can only decrease our bound, but removing premises can make us overconfident. This would not be a problem if, for example the extra premise brought some new evidence to light and made us less confident in the validity of the argument. But this has two problems. A higher degree of belief in the new premise should result in a lower degree of belief in the conclusion. The second problem is that this is precisely what Field is arguing against! So, although I agree with Field that to say an argument is logically valid is to accept conditions on ones own degrees of belief, I am not confident that this is a fruitful method of investigation.

[18]Even if we don't fully believe in the premises, this will maximize our lower bound. A short proof with one 'extra' premise: Suppose the sum of the $Cr(A_i)$ is μn, with $\mu \leq 1$. Then in order for $Cr(A_{n+1}) = \epsilon \leq 1$ to actually increase the lower bound, after cancelling we get $\epsilon + n(\mu - 1) > n(\mu - 1) + 1$ which immediately leads to $\epsilon > 1$, a contradiction. If we see what conditions we need on μ for the extra premise to get us full confidence in B, we get that $\mu = \frac{1-\epsilon}{n} + 1 \geq 1$, with equality only if $Cr(A_{n+1}) = 1$. So the extra premise cannot increase the bound.

6 The Role of Intuition

What exactly is the relation between intuition and logic? Can we learn something about logical consequence from only intuitive properties of the relation? In his paper *Informal Rigour and Completeness Theorems*, Georg Kreisel defends the idea that rigorous proofs outside of any formal system are possible. This is a very attractive idea as it allows us to get a "big picture" of logical validity and its relation to other concepts. Letting α be a sentence in some first-order language, we define $D(\alpha)$ to mean α is derivable "by means of of some fixed (accepted) set of formal rules"[8, pg. 89], $V(\alpha)$ means α is valid in all set-theoretic models, and finally, $Val(\alpha)$ means that α is intuitively valid. In the next paragraph, Kreisel makes this clearer by saying $Val(\alpha)$ "asserts that α is true in *all* [models]". This allows us to use intuitive properties of each predicate to relate one to the other, giving us a theorem. For first-order languages, we can use Gödel's completeness theorem to say

$$\forall\alpha(V(\alpha) \leftrightarrow D(\alpha)).$$

Since our formal rules are intuitively valid,

$$\forall\alpha(D(\alpha) \rightarrow Val(\alpha)),$$

and if our sentence is true in all models, i.e. logically valid, it is true in all set-theoretic models, so

$$\forall\alpha(Val(\alpha) \rightarrow V(\alpha)).$$

From these formulas, we can show that

$$\forall\alpha(Val(\alpha) \leftrightarrow V(\alpha)) \text{ and } \forall\alpha(Val\alpha) \leftrightarrow D(\alpha)),$$

which Kreisel takes as a proof that for first order languages, set-theoretic, proof-theoretic, and Tarskian model-theoretic methods will give us exactly the same extension. However, Etchemendy's problem with this proof is that, while this proof works for the definition of $Val(\alpha)$ given, it does not mean anything about our intuitive understanding of logical validity, since his whole argument is that in an interpretational model theory, truth in all models does *not* coincide with our intuitive understanding of validity. But Kreisel's proof strategy can be generalized. Letting $\{L_M\}_{M\in\mathcal{M}}$ be a collection of languages with the same set of sentences, but with different interpretations, since interpretational semantics overgenerates truths, the collection of logical truths must be a subset of the true sentences. In fact, the set of common logical truths must be a subset of the common truths of these languages. Etchemendy expresses this as

$$\bigcap_{M\in\mathcal{M}} LTr(L_M) \subseteq \bigcap_{M\in\mathcal{M}} Tr(L_M).$$

One condition Etchemendy places on this class of models $\mathcal{M}$ is that it needs to be "rich". He defines this as a collection of models which can serve as a "basis for the completeness theorem", in other words $\mathcal{M}$ is rich if for any first-order sentence S which is true in all models in $\mathcal{M}$, S is derivable. Finally, we can define $Val_{\mathcal{M}}(\alpha)$ as analogous to $V(\alpha)$, that is, α is true in all models in $\mathcal{M}$, not just all models, and $CLTr(\alpha)$ as being in the collection of common logical truths, defined earlier, which is analogous to $Val(\alpha)$. From these new definitions, he makes an Kreisel-like proof that these predicates are "coextensive".

Do these theorems really mean anything? If intuition provides a basis for our understanding of logic, these results are meaningful, and they tell us about the relation between these different views of logic. But intuition is, after all inherently subjective, so we might instead say someone is rational (a rational agent) if their intuition coincides with logical consequence. That is, to be rational is to think logically. So if someone is rational, their intuition would, by definition agree with logic. But we would not be able to establish that this person is rational without first understanding logical consequence.

7 The Meaning of Logical Consequence

Although the study of logic can be quite abstract, it is important to remember how fundamental the subject is. When Hanson says mathematics and logic are sciences in which the problems can be solved just by thinking about them, this really applies to any area of philosophy. Philosophy is predicated on the idea that it is possible to derive meaningful results by thinking about something. But this is only possible if we can eliminate possibilities, and the most basic boundaries for possibility are logical ones. That is to say, no argument could possibly hold up if either side of the argument were equally valid. Even if they seem to be equally valid at the moment, the reason people continue to study these subjects is the belief that ultimately one side of the argument really is correct, or one explanation really is better than the other. Since logic is essentially the study of arguments, this belief implies a realist view of logical consequence. But this is the basic problem of talking about logic. We need to use logic to argue about the properties of logic.

So what do we mean by logical consequence? We can interpret this question as asking "what is required to decide that a sentence is a logical consequence of some premises?", which is what logicians are primarily concerned with. But we can also interpret the question as asking "what exactly *is* logic?". While I disagree about the applicability of Field's approach, his question of the conceptual role of validity is very important. It seems that there are a few choices: i) Logic provides bounds on human reasoning (and, according to Field, on our degrees of belief) ii) it captures some metaphysical characteristics of our universe iii) or i) and ii) coincide. Intuition has always been a major part of the study of logic and many logicians do not make any distinction between i) and ii)[6, 11, 14], so it would then be useful to try to understand why we believe in iii). The first reason is that it simply *feels* that way. It seems intuitively obvious. But it seems very circular to use our intuition to argue for the adequacy of our intuition. As I mention in the introduction, this circularity may not be a problem, but is not a very satisfying explanation. But if such an explanation were possible, what would a "satisfying" explanation of logical consequence look like? Sher acknowledges and embraces this circularity by basing a logical system on an axiomatic set theory, and not insisting that logic be a basis for mathematics, or vice-versa. She uses set-theory only as a tool to explain properties of logic. This view is consistent with the dictionary analogy.

Many authors try to establish various intuitively necessary properties that the logical consequence relation must have, but for these reasons, it is difficult to say these properties are sufficient in characterizing the relation. Kreisel sums up the problem nicely.

> The 'old fashioned' idea is that one obtains rules and definitions by analyzing intuitive notions and putting down their properties. This is certainly what the mathematicians thought they were doing when defining length or area or, for that matter, logicians

> when finding rules of inference or axioms (properties) of mathematical structures ... The general idea applies equally to the so-called realist conception of mathematics which supposes that these intuitive notions are related to the external world ... and to the idealist conception which denies this or, at least, considers considers this relation as inessential to mathematics. What the 'old fashioned' idea assumes is quite simply that intuitive notions are *significant*, be it to the external world or in thought (and a *precise* formulation of what is significant in a subject is the result, not a starting point of research into that subject).

[8, pg.78] Although I am inclined to the realist view of mathematics, I can easily see how one can be an intuitionist. But I cannot see any justification for an intuitionist view of consequence. Perhaps mathematics does not need any relation to the external world, but *only* because we take for granted the reality of logical consequence. Perhaps one could view logic as simply a complex set of rules, and see what kinds of results it produces, but then we remove a seemingly fundamental aspect of logic. Why do we study some systems of logic and not others? If the rules are arbitrary, how can we decide what 'good reasoning' is? If they aren't arbitrary, why not? Some authors seem to have a more utilitarian view of logic, e.g. Field. But then the same questions come up. Why would logical principles apply so well to the external world?

The property that the premises are true whenever the conclusion is true seems to be a necessary property of the logical consequence relation, but it seems that the argument is still intuitively obviously valid regardless of the truth of the premises or conclusion. That is to say, even if I don't agree with, or don't know whether the premises are true, I would still say the argument is valid. So necessary truth preservation doesn't seem to fully capture our intuition of validity. The first argument from the beginning of the paper, for example, not only seems reasonable in a practical sense, but also in a necessary sense. Something about the very meaning of the words requires the conclusion to be a logical consequence of the premises. But what evidence do we have to say that it is an a priori fact that the conclusion follows from the premises? On the other hand, if an argument is obviously valid, does it need any other justification? Logic may be independent of our intuitive understanding of it, but what is logic to us without it?

Here is where we could form a combination of the realist and intuitionist views of logic. Perhaps logical consequence is a real, absolute relation, but we are inherently limited by the fact that our intuition is the only tool we have available to study the relation. So whatever properties logical consequence may have which lies outside of our intuition should not be considered a part of the study of logic. But if we adopt this view of consequence, Kreisel's theorem, or Etchemendy's version of the theorem relates our intuition about validity to model theory and proof theory, but it doesn't necessarily mean anything about 'absolute' validity. As far as empirical evidence, I am, of course not denying that simple logical consequence is observed in all sciences, nor that mathematical proofs (with no contested premises) are incorrect. Instead I am asking whether logical validity is knowable in all instances. In the *Questions and Objections* section of Sher's 1996 paper[11], she describes the foundationalist view that logic cannot be justified as meaning "we accept, or discard, logic as arbitrary, accidental, [or] conventional"[11, pg. 680]. However, in this realist-intuitionist view, it is not arbitrary, nor is it a convention, but still possibly unjustifiable. But is it accidental? Is it possible for there to be a logically valid argument for which there can only be counter-intuitive justifications? To many logicians, this is absurd. A logically valid argument which is counter-intuitive?

These observations seem to suggest that (regardless of whether we can develop a theory to determine validity in all instances) logical consequence is an absolute relation between the meaning expressed by the sentences in the premises, and the meaning expressed by the conclusion. The form of an argument is, then simply telling us precisely which parts of the sentence are relevant in the relation. In this case, the selection of fixed terms may vary from argument to argument, but are the same for arguments of the same form. The question then, is whether logically valid, but counter-intuitive arguments can exist, and if so, what problems does this create?

The importance of logic in philosophy should prompt us to take a careful look at these foundational questions. Do we need a justification for the correctness of our intuitive ideas of logic? Is this possible? Is 'obviousness' a meaningful justification for the validity of an argument or does it need some other justification? The role logic plays in our thoughts both motivates further study, while simultaneously making it very difficult (if not impossible) to talk about. For this reason, philosophers often place restrictions on logic. For example, Aristotle's 256 forms, or first-order logic. But the underlying idea is that, whatever logic *is*, it can be understood through our intuition. At the same time, logical consequence seems to be independent of this intuition. Understanding this connection between intuition and logical consequence is key to our understanding of logic as a whole.

Bibliography

[1] Cannon, D. (2016). Logic for Philosophers. (not yet published).

[2] Caret, C. R. and Hjortland, O. T., editors (2015). *Foundations of Logical Consequence*. Oxford University Press.

[3] Etchemendy, J. (1988). Tarski on Truth and Logical Consequence. *The Journal of Symbolic Logic*, 53:51–79.

[4] Etchemendy, J. (1990). *The Concept of Logical Consequence*. Harvard University Press.

[5] Field, H. (2009 2013). What is Logical Validity? In *Foundations of Logical Consequence*. Oxford University Press.

[6] Hanson, W. H. (1997). The Concept of Logical Consequence. *Philosophical Review*, pages 241–261.

[7] Hanson, W. H. (2002). The Formal-structural View of Logical Consequence: A Reply to Gila Sher. *Philosophical Review*, pages 243–258.

[8] Kreisel, G. (1969). Informal Rigour and Completeness Proofs. In Hintikka, J., editor, *The Philosophy of Mathematics*. Oxford University Press.

[9] Mates, B. (1972). *Elementary Logic*. Oxford University Press.

[10] Priest, G. (1995). Etchemendy and Logical Consequence. *Canadian Journal of Philosophy*, 25:283–292.

[11] Sher, G. (1996). Did Tarski commit 'Tarski's Fallacy'? *The Journal of Symbolic Logic*, 61:653–686.

[12] Sher, G. (2001). The Formal-structural View of Logical Consequence. *Philosophical Review*, pages 241–261.

[13] Tarski, A. (1931-35a). *Logic, Semantics, Meta-Mathematics*. Oxford University Press.

[14] Tarski, A. (1931-35b). On The Concept of Logical Consequence. In *Logic, Semantics, Meta-Mathematics*. Oxford University Press.