Reflections on Consequence

John Etchemendy
Department of Philosophy
Stanford University

November 1, 1999
(Draft: Comments welcome, but please don't quote.)

In *The Concept of Logical Consequence* (CLC) [13], I presented an extended argument that the standard, Tarskian analysis of logical consequence and logical truth is wrong. In the years since its publication, over a score of authors have written reviews, articles, or portions of books criticizing various arguments I gave in the book.[1] Nearly all have presented what the authors considered devastating replies to some or all of my arguments. Many of the replies are very thoughtful and contain much with which I entirely agree. Other authors misunderstood crucial parts of my argument, no doubt because I expressed them poorly. But all in all, the attention the book has gotten is gratifying. My only regret is that due to an onerous administrative appointment at my university, I was unable to reply to individual articles as they came out.

This paper is not meant to be a "reply to my critics." Such a reply would be of very little interest to any one reader, inasmuch as the critics themselves disagree so sharply on fundamental points, and so the lines of criticism are often at odds with one another. Instead, the paper is meant to be a rethinking of my overall argument in light of what I have learned from the various critiques, in particular what I have learned about ways in which CLC was confusing, incomplete, or otherwise misleading. Where appropriate, I indicate in footnotes how points made in this paper relate to specific criticism that has appeared in print.

Let me say at the outset that I still believe that all of the significant points made in the book are essentially correct. Indeed, I am confident that most readers will not need my help answering some of the criticism that has been offered in the literature. A fair amount of that criticism has centered on historical questions about Tarski's 1936 paper "On the Concept of Logical Consequence" [37], which I took as the philosophical locus of the standard analysis. Tarski's short paper is remarkably puzzling in many ways, but as I said in the book, and as most commentators agree, the important issue is not what Tarski was

[1] A partial list is contained in the references at the end of this article. Some of these were reactions to my earlier articles [11] and [12] (plus private correspondence) which covered some of the points discussed at greater length in CLC. I apologize to any authors whose work I have inadvertantly overlooked.

thinking when he wrote the paper, but whether the account he proposed is correct. Though I still hold to the reading of Tarski's paper described in the book, the historical debate is more a distraction than a useful guide to the important issues surrounding the account.

The articles and reviews have, however, convinced me that I made at least two serious mistakes in writing CLC: one a sin of commission, the other a sin of omission. The sin of commission was that I simply included too much, trying to anticipate objections, rationales, and modifications that might be raised in defense of Tarski's account. This is all to the good in one sense, but in another made it hard to see the forest for the trees. My overall argument, the "forest," can be summed up quite succinctly. Tarski's analysis involves a simple, conceptual mistake: confusing the symptoms of logical consequence with their cause. Once we see this conceptual mistake, the extensional adequacy of the account is not only brought into question—itself a serious problem given the role the semantic definition of consequence is meant to play—but turns out on examination to be at least as problematic as the conceptual adequacy of the analysis. To put it bluntly, the account fails both conceptually and, in most applications—in fact in virtually all applications—extensionally as well. That is the forest. In this paper I'll try to fill in just enough of the trees to make the justification of these claims clear.

Fixing the sin of omission, unfortunately, pulls in the opposite direction. In the book I intentionally avoided discussing my own views on consequence and model-theoretic semantics. At the time, I thought it better not to muddy the discussion with both a negative argument directed at Tarski's analysis and a positive argument for an alternative view. Since the defect in the analysis is entirely independent of my positive views, I did not want readers to imagine that the positive views, with which they might disagree, were somehow part and parcel of the critique.

I see now that this strategy was a mistake. First, several authors have criticized the book by proposing views of model-theoretic semantics remarkably similar to my own. The fact that they consider their proposed accounts of model theory criticism of my arguments suggests that they have seriously misconstrued the target of my critique. In particular, the critique is not aimed at model-theoretic techniques, properly understood, nor at the view that logical consequence is a fundamentally semantic, not syntactic, notion. Second, I now realize that without the positive side of the story, readers were legitimately puzzled about the overall significance of my arguments. If we acknowledge that the Tarskian analysis is wrong, does it mean that large tracts of accepted logical practice must be abandoned? Do many of the main technical results in the field suddenly lose their intuitive or philosophical significance? In fact, I think the significance is quite the opposite. Recognizing the flaw in the standard analysis has a liberating effect: it allows us to give a sensible account of much work that does not fit neatly into the picture that results from the flawed analysis, and opens up new areas of legitimate study that seem precluded by the analysis. Although I suggested as much at the end of my book, it is clear that the assertion alone was not enough to give readers a clear understanding of why this might

be so.

My second goal in this paper, then, is to provide the missing, positive account that I should have included in CLC. By this I don't mean I will propose a competing analysis of logical consequence: this project will have to wait until later. But I will sketch what I consider the proper understanding of model-theoretic semantics and its relation to the pretheoretic notions of logical consequence and logical truth. I claim that model theory, properly understood, does not yield an analysis of the logical properties, but presupposes them. This is not to say that a model-theoretic semantics for a particular language does not illuminate the logical properties of the target language. It does, but the illumination results not from an analysis of the basic logical notions, but rather from an entirely different characteristic of the model-theoretic technique, a characteristic which I will isolate and explain.

In order to add the positive account without obscuring the main argument even more—adding more trees, so to speak—I devote the first two sections of this paper to my criticism of Tarski's analysis, and postpone discussing my positive views on logic, model theory, and the consequence relation to sections 3 and 4. I am convinced that the point of view described in the latter two sections is in fact widely held, though not widely discussed, and so make no claims for the originality or novelty of that view. Still, I emphasize once again that my criticism of Tarski's account does not rely on the acceptance of that point of view.

1. Conceptual adequacy of Tarski's account

Let me begin by briefly recounting Tarski's analysis and sketching the main objections I raise in the book. For the reasons explained, I won't try to repeat the detailed arguments presented there, but will simply suggest their flavor and encourage the reader to go back to the original if the current summary is unsatisfying or if it seems I've overlooked an obvious point.

Tarski proposes a reductive analysis of the logical properties. The analysis purports to reduce logical truth to ordinary truth (or satisfaction) of an associated generalization (or open formula). Similarly, it reduces logical consequence to material consequence plus generalization: an argument is logically valid, according to the analysis, if every argument in an associated class of arguments preserves truth, where by "preserves truth" we mean simply that it has one or more false premises or a true conclusion. The analysis ensures that a logically true sentence is true because it is an instance of the associated generalization, and so could not be false without falsifying the generalization. It ensures that a logically valid argument preserves truth—has a false premise or true conclusion—because the argument is a member of its associated class.

I'll say more in a moment about how we get from a sentence to its associated generalization, or from an argument to its associated class of arguments. But for now let me note the remarkable appeal of such a reductive analysis. Surrounding the intuitive concepts of logical consequence and logical truth are a host of vague and philosophically difficult notions—notions like necessity, certitude, *a*

prioricity, and so forth. Among the characteristics claimed for logically valid arguments are the following: If an argument is logically valid, then the truth of its conclusion follows necessarily from the truth of the premises. From our knowledge of the premises we can establish, without further investigation, that the conclusion is true as well. The information expressed by the premises justifies the claim made by the conclusion. And so forth. These may be vague and ill-understood features of valid inference, but they are the characteristics that give logic its *raison d'être*. They are why logicians have studied the consequence relation for over two thousand years.

In spite of the importance of these characteristics, we needn't be happy about the vagueness. And that is why the reductive analysis of consequence is so attractive. Tarski shows us, if he is right, how the logical properties can be reduced to the well-understood notion of truth, plus whatever is involved in specifying the associated generalization or class. Like magic, the vague and obscure notions that sit at the core of our discipline simply disappear. No wonder we find Tarski's account so appealing: if it works, it allows us to set aside a breathtaking number of philosophical issues. Who could not want the account to succeed?

So how does Tarski characterize the associated generalization and the associated class of arguments? Let's consider arguments first. According to Tarski's account, the associated class consists (roughly speaking) of all arguments displaying a similar "logical form," where logical form is defined by the appearance of the so-called "logical constants," plus the pattern of the remaining expressions. So, for example, assuming *if... then* is the only logical constant in the following argument:

> *If Tarski was right, then Etchemendy is wrong.*
> *Tarski was right.*
> *So, Etchemendy is wrong.*

the associated class consists of all arguments displaying the following form:

> *If P then Q*
> *P*
> *So, Q*

What Tarski means by "all arguments" of the displayed form is not just arguments actually expressible in the language, but all arguments expressible in sufficiently similar languages. Exactly what is meant by this isn't crucial for our purposes. But a more modern way of expressing the analysis is to stick with the original argument and ask whether it preserves truth (in the actual world) regardless of how we interpret the constituent sentences *Tarski was right* and *Etchemendy is wrong*, or alternatively, regardless of how we interpret the names *Tarski* and *Etchemendy*, and the predicates *was right* and *is wrong*. If the argument preserves truth on all interpretations of the non-logical constants, then it is said to be logically valid.

The definition of logical truth is similar. We hold fixed the logical constants in the sentence, and quantify away the contribution of the remaining expressions. For example, the sentence:

If Tarski was right, then Tarski was right

is logically true, according to the analysis, because every instance of the form *If P then P* (or *If R(a) then R(a)*) is simply true. Or in modern parlance, the displayed sentence is logically true because it remains true however we interpret the expressions *Tarski* and *was right*.

I have already mentioned the great attraction of the reductive analysis: the fact that it replaces a host of obscure modal or epistemic notions with the vastly clearer notion of truth. But we can now add to that attractiveness the observation that the account is also quite plausible, at least at first glance. Consider our sample argument above. This is a logically valid argument, and so of course preserves truth—that is, has a false premise or true conclusion. What's more, it is an instance of an argument form, *modus ponens*, all of whose instances preserve truth. Indeed, instances of any of the well-known rules of inference—the Aristotelean syllogisms, universal instantiation, even the omega rule—display these same features. Each such argument is an instance of an argument form all of whose instances preserve truth. The reductive analysis simply takes the natural step of proposing these obvious features of valid rules of inference as *definitive* of the logical consequence relation. This is surely an attractive and plausible proposal.

In CLC, I argued that the reductive analysis fails. In what sense? Roughly speaking in any sense that would give it philosophical, logical, or foundational interest. In this section, I begin with the easy part: isolating the conceptual flaw in the account. Suppose we ask whether the account captures, directly or indirectly, any of the intuitive characteristics of the consequence relation mentioned above. Does it guarantee, or for that matter give us any reason to expect, that inferences that qualify as valid according to the analysis will have any of the modal, epistemic, semantic, or informational characteristics ascribed to logically valid arguments? My concern here is not with the extension of the account, which I will discuss at length in the next section, but with its conceptual adequacy: with the question of whether there is any *conceptual* assurance that arguments declared valid by the account will in fact be genuinely so.

Perhaps not surprisingly, the answer is *no*. This is simply a generalization of an observation first made by Wittgenstein when Russell entertained a similar, reductive account of logical truth. I will make the point in my own words, rather than Wittgenstein's, and about logical consequence rather than logical truth, since that is the important concept. The crucial point is this. The property of being logically valid cannot simply consist in membership in a class of truth preserving arguments, however that class may be specified. For if membership in such a class were all there were to logical consequence, valid arguments would have none of the characteristics described above. They would, for example, be epistemically impotent when it comes to justifying a conclusion. Any uncertainty about the conclusion of an argument whose premises we know

to be true would translate directly into uncertainty about whether the argument is valid. All we could ever conclude upon encountering an argument with true premises would be that *either* the conclusion is true *or* the argument is invalid. For if its conclusion turned out to be false, the associated class would have a non-truth-preserving instance, and so the argument would not be logically valid. Logical validity cannot guarantee the truth of a conclusion if validity itself depends on that self-same truth.

It might help to look at an analogous, but obviously faulty definition of consequence where the same problem arises. Suppose we defined a logically valid argument as an argument that simply preserves truth, that is, has either a false premise or a true conclusion. Then of course any "valid" argument with true premises would have to have a true conclusion, since otherwise it would be invalid. Nevertheless this definition of validity misses the crucial feature of genuine consequence: *the fact that we can draw conclusions about sentences whose truth values we do not antecedently know, based on our knowledge of other sentences that logically imply them.* If logical validity were nothing more than truth preservation, then our knowledge that the premises of an argument are true would only tell us that either the conclusion of the argument is true (and hence the argument "valid") or the conclusion is false (and hence the argument invalid). Whatever enables us actually to infer consequences from our knowledge of other claims has simply dropped out of the account.

I claim that Tarski's reductive analysis of consequence, though certainly more involved, suffers from precisely the same conceptual omission. The fact that validity is tied to a larger class of arguments does not help, nor does the appeal to logical constants as a means of specifying that class. Indeed the problem remains no matter how narrowly one construes the term "logical constant," no matter how general the resulting argument forms. Indeed, consider again the instance of *modus ponens* shown above. Surely, if anything is a logical constant, *if... then* must be. But even here, if the logical validity of the argument came down to nothing more than the fact that every instance of the illustrated form preserves truth, then the truth of the premises could never be used to establish the truth of the conclusion. The conclusion might be true and the argument valid, or the conclusion false and the argument invalid. The consequence relation, as characterized by the reductive account, involves nothing that would incline us toward one of these possibilities rather than the other.

Of course, this problem does not infect genuinely valid arguments, like the instance of *modus ponens* illustrated above. Indeed, that's the point. The crucial feature of *modus ponens* is that we can recognize that all of its instances preserve truth without knowing the specific truth values of the sundry instances. My own view is that we recognize this by virtue of the meaning of the expression *if... then* and our knowledge of how the remaining constituents can contribute to the truth values of the premises and conclusion. This gives us an independent guarantee—independent, that is, of the actual truth values of premises and conclusions—that all the instances of the argument form preserve truth. This independent guarantee, and only this independent guarantee, is what enables us to infer that a conclusion is true on the basis of the truth of the premises.

It is obvious that in the absence of such a guarantee, we would not have a logically valid argument, regardless of which expressions we considered logical constants and regardless of the truth preservation of its instances. Suppose we have an argument form all of whose instances preserve truth, just as the reductive account requires, but suppose that the only way to recognize this is, so to speak, serially—by individually ascertaining the truth values of the premises and conclusions of its instances. In other words, suppose there is no independently recognizable guarantee of truth preservation, as there is with *modus ponens*, only the brute fact that the instances preserve truth. Would an instance of this argument form be logically valid? Clearly not. For example, we could never come to know the conclusion of such an argument in virtue of our knowledge of its premises. Indeed, the premises would provide no justification whatsoever for a belief in the conclusion. For, by hypothesis, knowing the specific truth value of the conclusion in question would be a prerequisite to recognizing the "validity" of the argument.[2]

We can and do recognize that all instances of *modus ponens* preserve truth, and we do this without having the foggiest idea of the actual truth values of most of the sentences that make up those instances. The characteristic that enables us to do this, at least in the case of simple valid forms, is clearly the essential feature of logical consequence. For an argument *with* this characteristic can be used to extend our knowledge: we can know antecedently that the argument preserves truth, subsequently discover that its premises are true, and thereby infer that its conclusion is true as well. By contrast, an argument form *without* this characteristic could never be so deployed.

I said a moment ago that the conceptual problem with the reductive account applies regardless of how tightly constrained our notion of a logical constant. Let me expand on this a bit. When presented with the omission just described, people often reply that the reductive account is not defective, but rather incomplete. What is required is a careful characterization of the logical constants. This characterization, it is thought, will explain why when these and only these expressions are held fixed, the only arguments that can possibly satisfy the reductive definition are those that display the required guarantee of truth preservation, that is, those that are genuinely valid. In other words the crucial guarantee, it is thought, flows jointly from the truth preservation of the associated arguments plus certain special features of the logical constants. Of course, until we see what those features are, this is little more than an article

[2]Graham Priest [25] accuses me of confusing the "epistemic order" with the "definitional order," drawing an analogy with the notion of computability and the Church/Turing analyses of that notion. As Priest correctly points out, we might know that a function is effectively computable without knowing that it is Turing computable. But Priest has misunderstood my argument, the point of which is that it is possible for an argument that is not in fact logically valid—one that has none of the epistemic or other characteristics of a valid argument—to satisfy Tarski's definition. The point has nothing to do with knowing whether Tarski's definition applies, but rather with the characteristics (or lack of characteristics) of the arguments to which the definition could in fact apply. The right analogy would be if we could show that it is possible for functions that are not effectively computable to be Turing computable (which of course we can't).

of faith, though an article of faith that has sustained many a supporter of the reductive analysis.

But this article of faith is simply false. It is not hard to prove that there are no features of the logical constants capable of providing this assurance, at least on the assumption that the truth-functional connectives are logical constants. To see this, consider the following argument form:

$P(a) \wedge Q(a)$
$\neg P(b)$
So, $P(c) \rightarrow Q(c)$

This is obviously not logically valid: from premises of the indicated forms we are in no way justified in inferring the corresponding conclusion. Now it happens that there are non-truth-preserving instances of this argument form. But notice that this is not guaranteed by any features, global or local, of the truth-functional connectives appearing in the argument. For there would *not* be any non-truth-preserving instances if the world contained only two objects, or if all the objects in the world fell into two indistinguishable types. And yet the truth-functional connectives would presumably still have whatever features we thought definitive of the logical constants. This shows that the conceptual flaw in the reductive analysis will never be corrected by specifying characteristics of the logical constants, at least on the assumption that these characteristics are enduring features of the truth-functional connectives. *The source of the guarantee observed in genuinely valid arguments is not the truth preservation of their instances plus special features of the logical constants.*

No selection of logical constants rules out the possibility discussed four paragraphs back: arguments that satisfy the reductive definition due to the "brute fact" that their instances preserve truth, but which do not display the guarantee of truth preservation that makes an argument genuinely valid. No matter how we characterize the logical constants, Tarski's definition provides no assurance that every argument that satisfies the definition will be logically valid.

Let me summarize. Genuinely valid arguments carry with them an independent guarantee of truth preservation, a guarantee that can be recognized antecedent to our knowledge of the actual truth values of their premises and conclusion. Now what is important for present purposes is not how we diagnose this crucial guarantee. I've indicated that I think it emerges from semantic characteristics of the language, but others may disagree. Kant, if I remember correctly, attributed it to the *a priori* structure of the understanding; others seem content to appeal to a primitive notion of logical necessity. But in any event, what is important for now is only that we recognize the following two points. The first is that without such an independent guarantee of truth preservation, logical consequence would be a completely flaccid relation. It would be impossible to use logically valid arguments to extend our knowledge, to justify the truth of a conclusion, or to prove that a given theorem follows from accepted axioms. The second point is that the reductive analysis just omits the guarantee, attempting to replace it with that which the guarantee is a guarantee *of*. We ignore whatever it is that assures us that every instance of a logically valid

form preserves truth, and say that logical validity simply *consists* in every instance of the given form preserving truth. It is like confusing the symptoms with the disease, effects with their cause: understandable confusions, but confusions nonetheless.

I take both of these points to be undeniable, the first about the consequence relation, the second about the reductive analysis. When you think carefully about these two points, you will see that they show that none of the central characteristics of the consequence relation—whether modal, epistemic, semantic, or informational—are captured by Tarski's analysis. Since much of my discussion in CLC focused on the standard modal characterization, many commentators have not quite understood this point. For example, Timothy Smiley interprets my book as an attack on the "non-modal aspect" of Tarski's definition. Smiley goes on to say:

> A debate is called for, but it will be more fruitful if it asks for what purposes necessity is an essential ingredient of consequence. For example, someone who does not endorse Aristotle's doctrine of proof and episteme may well be content with proofs that establish the bare truth of theorems, and it is not obvious that this requires a modal relation of consequence. [36, p. ??]

My reply should be obvious: Fine, jettison all talk of modalities.[3] Concentrate on nothing more than the fact that the consequence relation allows us to establish the truth of sentences based on the truth of others. But that is precisely the problem, precisely the characteristic that Tarski's definition ignores. If the consequence relation involved nothing more than what the reductive definition maintains, then the relation could never be used to "establish the bare truth of theorems." When we encounter a new inference of any specified form, our sole guarantee would be that it either preserves truth or constitutes a counterexample in virtue of which the argument form is invalid. This guarantee, the only one Tarski's definition offers, can never establish the truth of anything.

This was, as I said, the easy part. It is clear that Tarski's definition tries to reduce a "cause"—the logical consequence relation—to its "symptoms," the truth preservation that the consequence relation guarantees. And it is equally clear that this guarantee of truth preservation is the essential feature of logical consequence, the feature that makes it possible to infer the conclusion of a valid argument from its premises. In short, the reductive analysis omits the single most important characteristic of the consequence relation. Let's consider what follows from this fact.

[3]Note, by the way, that my own explanation of the validity of *modus ponens* is semantic, not modal. I am no particular fan of modal characterizations of consequence, and thought I had made that relatively clear in CLC. In spite of that, some commentators seem to have concluded that I identify logical truth with necessary truth and logical consequence with necessary truth preservation. Nothing could be further from the truth.

2. Extensional adequacy of Tarski's account

We might summarize the observations of the preceding section with a simple slogan: All of the instances of *modus ponens* preserve truth because it is a logically valid argument form. This is true. What is false is that *modus ponens* is logically valid *simply because* its instances preserve truth. What follows from the fact that Tarski's definition is based on the latter, faulty assumption?

First and most obvious, it follows that we have no assurance that the reductive account will yield the correct extension. When we apply the definition to an arbitrary language, choosing some subset of its expressions as logical constants, we have no blanket assurance that the arguments declared logically valid in fact *are* logically valid. Let me call this the question of *overgeneration*. There is an equally problematic question of *undergeneration*—are there any logically valid arguments in the language which, on the given selection of logical constants, are not declared valid—but let me set this second issue aside for the moment.

In CLC, I discuss at length how, when, and where the reductive account overgenerates. Obviously, applications of the account will overgenerate if there are argument forms all of whose instances in fact preserve truth, yet which do not provide the guarantee of truth preservation required of logically valid arguments. Intuitively, this can happen if the truth preservation is an upshot of facts that have nothing to do with logic, the consequence relation, or anything plausibly related to it.[4] As I explain in the book, this happens whenever the language, stripped of the meanings of the non-logical constants, remains relatively expressive, or if the world is relatively homogeneous, or both. This is not hard to see. If the expressions we've chosen as logical constants are sufficiently expressive, then it will always be possible to come up with argument forms whose instances uniformly preserve truth in spite of the fact that they are not logically valid: we need only find a non-logical generalization that is expressible in these terms and cast it into an argument form. The more homogeneous the world, the easier this task becomes, since fewer expressive resources are required.

So when does Tarski's account work, in the sense of not overgenerating? Well, we can say with confidence that the account works when applied to the language of propositional logic, treating the truth-functional connectives ($\neg$, $\wedge$, $\vee$, $\rightarrow$, etc.) as the sole logical constants. This is an exceedingly weak language, and so long as the world provides us with an adequate supply of true and false propositions, the only argument forms whose instances universally preserve truth are those whose truth preservation is guaranteed by the meanings of the chosen connectives. They are all, in other words, genuinely valid arguments. Mind you, it is easy to introduce a new "logical constant" for which this is not the case, for example the operator $\odot$ discussed in Chapter 9 of CLC. But if we

[4] I am being intentionally vague here, but only to maintain an ecumenical stance about the source of the consequence relation. Those who share my view that the relation emerges from the semantic characteristics of the language can replace the vague phrase "facts that have nothing to do with logic" with "facts that have nothing to do with the semantics of the language." Others can make corresponding replacements, depending on their views of consequence. In what follows, I will continue to speak of "non-logical" or "substantive" facts in this way, leaving it to the reader to make appropriate substitutions.

treat the truth functions as the only logical constants, this application of the account successfully avoids the problem of overgeneration. This is not because, even here, the reductive account is testing for the right thing, but only because the truth-functional operators are expressively weak and the world sufficiently heterogeneous.

What happens when we move to the language of first-order logic? To put it mildly, things get complicated. If we apply Tarski's unmodified definition to such a language, adding the first-order quantifiers and identity predicate to our list of logical constants, then the account overgenerates right and left. For one thing, the quantifiers and identity predicate allow us to express many numerical truths that are substantive, non-logical claims about the world. For example, we can express the fact that there are more than three billion objects using a sentence that contains only the quantifiers, the identity predicate, and the truth-functional operators. Call this sentence β. According to the reductive account, β is a logical consequence of any premises whatsoever, since any argument with this conclusion preserves truth. What is happening here is exactly what we predicted two paragraphs back: the chosen logical constants are now sufficiently expressive that the basic conceptual flaw in the account manifests itself in concrete, extensional errors. Every instance of the argument form:

P
So, β

preserves truth, in spite of the fact that these arguments do not display the required guarantee that would justify an inference from premise to conclusion. The fact that the argument form contains only expressions traditionally considered logical constants is no protection against the fundamental mistake made by the reductive account: taking the symptoms of the genuine consequence relation as definitive of the cause. The above argument form is just one of many examples of the misdiagnoses that result; I refer the reader to CLC for additional examples and a more extensive discussion of this phenomenon.

This sort of example was used by Wittgenstein to convince Russell of the flaw in the reductive account. Modern applications of the analysis avoid this embarrassment by adding, without explanation or rationale, a new twist. Rather than say that any argument form all of whose instances preserve truth is logically valid, we require additionally that the instances preserve truth *in every (actual but possibly restricted) domain of quantification*. Since there are domains of quantification containing fewer than three billion objects, we thereby dodge these particularly blatant instances of overgeneration.

In CLC, I argue that this new twist is at best unmotivated and at worst inconsistent with the original, reductive account.[5] Be that as it may, it is not

[5]Giving this argument in detail would be impossible in an article, but let me rule out one motivation that may spring to mind. We might reason that although the universe is the size it is, it could have been larger or smaller, and varying the domain is meant to take account of these possibilities. But considerations of the possible size of the universe is completely irrelevant to the reductive account, which is based on the assumption that we can reduce logical consequence to facts about how the world actually is, not how it could have been. Surprisingly enough, many commentators have missed this basic point.

worth repeating those arguments here, for the revised account is subject to the same fundamental flaw as the original reduction. It is still possible for an argument form to have only truth preserving instances—instances that preserve truth in every existing domain—without being logically valid, without having the guarantee of truth preservation needed to support an inference from premises to conclusion. In the book I emphasize this in various ways, including pointing out the peculiar position of the finitist who, if the reductive account were correct, would be forced to accept as logically valid many first-order arguments that obviously are not. For example, if there is a largest domain, then any inference whose conclusion asserts that there are no more objects than the cardinality of this domain will be incorrectly declared logically valid. Similarly, the inference from the claim that a relation is transitive and irreflexive to the claim that the relation has a "least" or "greatest" element will be declared logically valid if all domains are finite, even if there is no largest domain.

Many people have misunderstood the point of this argument, which is not directed at the finitist, nor meant to show that Tarski's analysis presupposes the axiom of infinity—perhaps an interesting point, but not an objection I would consider significant. Rather, it shows that even the modified account, incorporating varying domains of quantification, suffers from the same conceptual flaw described in Section 1. It still provides no conceptual assurance—whether due to the truth preservation of instances, characteristics of the logical constants, or the variation in the domain—that all arguments which satisfy the definition are actually valid. The symptoms of consequence on which Tarski's account is based are not a reliable indicator of genuine consequence, even when we vary the domain of quantification.

This is an important realization, even if you believe, as do I, that finitism is false; indeed, even if you believe that finitism is necessarily false.[6] For although the availability of infinite domains may assure us that the *specific* examples mentioned two paragraphs back are not mistakenly declared valid, *it does nothing to assure us that there aren't other arguments that are.* It is clear that the output of even the modified account depends on facts, such as the size of the universe, that have no bearing on the logical consequence relation. What is not clear is whether any such facts expressible in the first-order language cause this application of the account to overgenerate. This uncertainty is a direct conse-

[6] Vann McGee [21, 23] has replied to my argument by claiming that finitism is not simply false but necessarily false, since mathematical objects like pure sets exist necessarily. I am not sure how to assess the truth of this claim, though I suspect I agree with it. Still, it does not weaken my argument. When we apply Tarski's account to a particular language we make a host of decisions about the kinds of objects we will take as legitimate interpretations of the non-logical constants, for example whether predicates are interpreted by properties or arbitrary sets, and whether individual constants may refer to abstract objects as well as concrete ones. When we make the standard decisions on these matters—using sets to interpret predicates, and so forth—it follows from McGee's (entirely reasonable) assumptions about mathematical truth that an application of the account will indeed *have the extension it has necessarily.* When the extension is right, it will be necessarily right; when it is wrong, it will be necessarily wrong. If we accept McGee's assumption, then, the problem is that Tarski's account provides no general assurance that such an application of the account will be necessarily *right* rather than necessarily *wrong.* As we'll see, it is sometimes the former, sometimes the latter.

quence of the conceptual flaw discussed in Section 1, and the flaw applies with full force to the modified definition.

Assuming there are no finitists among us, and assuming as well a reasonably powerful set theory, it turns out that the modified reductive account does not overgenerate when applied to the first-order language. Now if you understand what I've said so far, a question should immediately come to mind: How do we know that this application of the account does not overgenerate? This is not an idle question. If we apply even the modified account to first-order languages, and limit ourselves to considerations internal to the Tarskian definition of consequence, *there is absolutely no way to determine whether all of the arguments declared logically valid are in fact logically valid.* There is no way to rule out the possibility that general, extralogical facts expressible using the first-order quantifiers, identity and the truth-functional connectives give rise to truth preserving argument forms that are not logically valid, that display the symptoms of validity but not the underlying cause. Perhaps these facts are more complex cardinality claims similar to β, but whose truth is not blocked by the trick of varying the domain of quantification. Or perhaps there are obscure algebraic or set-theoretic facts that are true in every domain, but not because they are logically true. If we think we are assured that this application of the account does not overgenerate—whether on general philosophical grounds or because of any characteristics of the expressions we've chosen as logical constants—we are simply fooling ourselves.

Though it is not an idle question, it does have an answer. In fact we can prove that this particular application of the account does not overgenerate by appealing to an entirely different tool for studying the consequence relation: a system of deduction. Now it is generally accepted that deductive techniques do not provide an analysis of the logical consequence relation. Nevertheless, it is possible to set down a simple collection of deductive rules whose repeated application will never permit us to prove a sentence that is not a genuine consequence of the assumed premises. How is this possible? First, we set out a handful of argument forms whose instances are all logically valid, that is, whose instances all display the requisite guarantee of truth preservation. Obviously, we choose forms like *modus ponens*, while avoiding those like "from any sentence, infer β," since the former do, but the latter do not, display the requisite guarantee. Second, we observe that the logical consequence relation is transitive, and hence repeated application of the primitive valid rules can never lead to a conclusion that is not a genuine consequence of the original premises.

What these two points show is that the careful application of deductive techniques allows us to design systems that are recognizably sound, systems we can be sure do not "overgenerate." And with first-order logic, it happens that we can use such a system to prove that the (modified) Tarskian account does not overgenerate, either. This follows from the so-called "completeness" theorem for first-order logic. The theorem assures us that any argument declared valid by the (modified) Tarskian account is provable in the deductive system, and hence is sure to be logically valid, thanks to the intuitive soundness of that

system.[7] Seen in this light, the theorem is actually misnamed, for its import is to transfer our assurance of the soundness of one characterization of consequence, the deductive system, to another characterization of consequence, the Tarskian definition, whose "soundness" we can never independently ascertain.[8]

What I have just argued is that the application of the (modified) reductive account to first-order languages can be proven correct. Or rather, I've argued that we can prove this application does not overgenerate, since I've set aside for the moment the issue of undergeneration. But if you understand the argument, you will begin to see why I claim that the faulty analysis has little philosophical, logical, or foundational interest. The common mythology is that the Tarskian definition is important because we have an independent, conceptual assurance of its extensional adequacy, and this allows us, among other things, to prove the extensional adequacy of other characterizations of consequence, such as our system of deduction. But once we recognize the conceptual flaw in Tarski's account, we see that it is not, contrary to mythology, in better shape than our deductive characterization of consequence. Quite to the contrary, the deductive techniques are actually in better shape: as we have seen, the careful application of these techniques can at least give us a characterization of consequence that is recognizably sound. This is more than we can say for the Tarskian definition of consequence, where our assurance of "soundness" is entirely derivative from the deductive system.

We can emphasize this point by asking what we can conclude in cases where we have an intuitively sound deductive system and a Tarskian definition of consequence, but the completeness theorem fails. In these cases, the Tarskian definition asserts the logical validity of arguments that go beyond what the deductive system can prove. The standard mythology would have it that in such cases the deductive system is incomplete. But this presupposes that the Tarskian definition is guaranteed not to overgenerate. Not only do we have no such guarantee, once we appreciate the flaw in the reductive analysis, we see that it predictably will overgenerate. So in the absence of a completeness theorem, our only legitimate conclusion is that *either* the deductive system is incomplete, *or* the Tarskian definition has overgenerated, *or possibly both*.

We have already seen a simple example of this. If we apply the unmodified Tarskian definition to a first-order language, we will not be able to prove completeness, since the standard deductive system cannot prove sentences like

[7] A more cautious and correct statement would be that the completeness theorem shows that *if* there are infinite domains, and *if* the presupposed axioms of set theory hold, *then* the modified reductive account does not overgenerate. The theorem obviously provides no assurance for the finitist, for if the finitist is right, this application of the account demonstrably overgenerates.

[8] This is closely related to Kreisel's construal of the completeness theorem in [20]. The difference is that Kreisel accepts the reductive account of logical consequence, but is worried about the fact that standard applications of it survey only domains that are sets, while we often use the first-order language to talk about proper classes of objects. But even if we included proper classes among the domains surveyed, we would still have to worry about the possibility of overgeneration. The completeness theorem assuages both worries, Kreisel's and mine.

β from arbitrary premises. But this does not show that the deductive system is incomplete, but rather that this application of the reductive definition overgenerates: β is not a logical truth, and in fact should not be provable from random premises.

A more interesting example is second-order logic, where the problem appears even in the modified account. If we apply Tarski's account of consequence to this language in the most natural way, treating both first- and second-order quantifiers as logical constants, then the resulting consequence relation extends well beyond any intuitively sound deductive system for the language. But what can we conclude from this? Can we infer that there is no complete deductive system for second-order languages, that any candidate system leaves some *genuinely* valid arguments unprovable? Or is the problem that the Tarskian definition of consequence overgenerates when applied to these languages, declaring sentences logically true and arguments logically valid that in fact are not? Or perhaps both?

The answer is that we can't really tell, at least not based on Tarski's reductive definition of consequence. It is well known that when the Tarskian definition is applied to second-order languages, certain highly abstract set-theoretic claims are declared logically true. For example, we can easily formulate sentences containing only identity, truth functions, and first- and second-order quantifiers that are true in all domains if and only if the Continuum Hypothesis is true, and other sentences that are true in all domains if and only if the Continuum Hypothesis is false. Let CH and $\neg CH$ be representative sentences of this sort.[9] If the Continuum Hypothesis is true, then CH will be declared, by the reductive account, a logical consequence of any sentence whatsoever. If the Continuum Hypothesis is false, the latter sentence will have a similar fate. But from an intuitive standpoint it seems that neither the argument form:

P
So, CH

nor the argument form:

P
So, $\neg CH$

displays the guarantee of truth preservation required of logically valid arguments.[10] It does not seem that we are logically justified in concluding CH from a random premise, even if the Continuum Hypothesis happens to be true

[9]For example, we can take CH to be the second-order sentence that says there are no subsets of the domain larger than $\mathbf{N}$ and smaller than $\mathbf{R}$. Since "larger than $\mathbf{N}$" and "smaller than $\mathbf{R}$" are definable in (full) second-order logic, this sentence will be true in all domains just in case the Continuum Hypothesis is true. Similarly, we can take $\neg CH$ to be the sentence that asserts that if the domain is at least the size of $\mathbf{R}$, then there *are* such intermediate-sized subsets. One of these is true in all structures, though we don't know which.

[10]I should mention, in case it is not obvious, that the fact that I've used argument forms with trivial premises is not significant. In both this and the first-order case I could give examples in which the forms of the premises are significant. The examples I've chosen are just easier to describe.

(perhaps unbeknownst to us). A more reasonable hypothesis is that the identification of the symptoms of consequence with genuine consequence here fails, thanks to the expressive power of this language. All of the instances of one of these argument forms will indeed preserve truth, but not because it is logically valid.[11]

As I said, the wayward behavior of the reductive analysis when applied to second-order languages is well known. Some philosophers, including Quine, have concluded from examples of this sort that *second-order logic is not logic.* This conclusion has to count as one of the more surprising and implausible conclusions of recent philosophy. After all, second-order languages, like all languages, have a logical consequence relation. Some inferences employing the expressive devices of these languages are logically valid, and others are not. True, the consequence relation for these languages may be vague or underspecified, depending on the vagueness or underspecification of the expressions that make up the language, and perhaps also because of the vagueness of our understanding of the consequence relation itself. But the idea that studying the logic of these languages is somehow not the business of logic is hardly a supportable conclusion. If we are convinced that the above argument forms are not logically valid—certainly a reasonable position—then we should simply conclude that this is a case where Tarski's analysis overgenerates, something that we know, for conceptual reasons, is bound to happen. The symptoms of consequence—truth preserving instances—are not reliable indicators of the sought-after cause: genuine logical consequence.

So far, I have only discussed the problem of overgeneration, but the reductive account can undergenerate as well. Most obviously, it will fail to detect logically valid arguments if the validity of those arguments depends on expressions not in the chosen collection of logical constants. To take a simple example, suppose we apply the account to an interpreted propositional language, treating only the truth-functional connectives as logical constants. Although all the arguments identified as valid in this case are genuinely so, there may well be valid arguments that are overlooked. For example, suppose the following argument is expressible in the language (where $Triangle(a)$ asserts that a is a triangle):

$Triangle(a)$
$a = b$
So, $Triangle(b)$

[11] Much more could be said about second-order logic than I can say here. For example, there are various ways to modify the interpretation of the second-order quantifiers to decrease the expressive power of the language. We can, for instance, construe them in the manner of so-called "weak" second-order logic, or perhaps as plural quantifiers (appropriately generalized to handle relation variables). And as the expressive power of the chosen logical constants decreases, so too will the instances of overgeneration, for reasons I have already explained. But this does not affect the point made in the text. When the quantifiers are interpreted as quantifying over all subsets of the domain—surely a possible interpretation, and probably the most natural—the problems discussed here unavoidably arise. All that matters is that this is a possible interpretation of second-order quantification; whether there are also weaker interpretations is irrelevant.

This argument would not be declared valid for obvious reasons: its validity depends on the meaning of the identity predicate, which is not among the expressions we've treated as logical constants, plus the fact that $Triangle$ is an extensional predicate. Similarly, if the language contains quantifiers, but we do not treat them as logical constants, most of the interesting arguments of first-order logic will be judged invalid. None of this, of course, is the least bit surprising.

The real problem with undergeneration arises when a language contains expressions that figure into valid argument forms, but which we cannot treat as logical constants in Tarski's account for fear of the opposite problem: overgeneration. For recall that the more expressive the list of logical constants, the more likely it is that the reductive account will overgenerate. But how often does this occur? In the book, I give a very simple, but artificial example of a language in which no selection of logical constants characterizes what intuitively seems the right set of logically valid arguments. But let's avoid artificial examples and jump headlong into a controversial one.

Suppose the language in question contains a binary predicate, say $\simeq$, that asserts that two objects are identical in shape. Then it seems at least arguably the case that the conclusion of the following argument follows logically from its two premises, much like our previous example:

$Triangle(a)$
$a \simeq b$
So, $Triangle(b)$

Surely, this conclusion must be true provided the premises are true: indeed, its truth preservation is guaranteed by the meanings of the constituent expressions. One could even argue that it is formally valid, since it holds for any a and b, and even holds when we replace one shape predicate with another. Unfortunately, if we treat enough predicates as logical constants to validate this (and similar) arguments, Tarski's account is sure to overgenerate. Contingent facts about the shapes of objects in the universe will result in arguments that are declared logically valid, but which do not display the guarantee of truth preservation that seems evident in our chosen example.[12]

As I said, this is a controversial example, but it is worth mentioning for a couple of reasons. For one thing, it is not obvious that this example is all that different from the previous argument, which is universally acknowledged to be logically valid. Numerical identity justifies substitution of individual constants, so long as the predicates involved are extensional. Identity of shape would seem to justify similar substitutions, albeit within a more narrow class of predicates.

[12]For instance, suppose $P(x)$ is a shape predicate satisfied by only finitely many objects, say n. Then $\neg P(a)$ will be declared a logical consequence of any collection of premises that imply that n objects not equal to a satisfy $P(x)$. Just as we noted earlier that in the first-order case we must assume an infinite universe or the (modified) Tarskian account will overgenerate, here we would, for a start, have to hope that every shape predicate is actually satisfied by infinitely many objects. Again, note that appeals to "possible objects" and "possible satisfaction" are completely irrelevant to the reductive account.

Given the similarity of these inferences, it is hard to see why they should be treated differently. Of course, most philosophers have been raised, under the influence of Quine, to say that the former inference is an instance of logical consequence, while the latter is something quite different: "analytic consequence," or something of the sort. But the idea that the justification underlying the first inference is different in kind from the second is supported by nothing more than the fact that the reductive account of consequence can be made to work in the first case but not in the second. Given the flaw in the reductive account, this is hardly a persuasive consideration.

Still, philosophers are extremely wary of any mention of the meaning of predicates—with the exception of identity, which receives special dispensation. So are there other cases where this problem occurs, where any selection of logical constants either overgenerates or undergenerates? Well, how about first-order logic? Before modifying the reductive account, this was precisely the situation we were in. If we include the standard collection of logical constants, then sentences like β turn up as logical truths. But if we delete any of these from the list, many obviously valid arguments are not so declared. And how about second-order logic? Once again we have the identical problem, only this time varying the domain doesn't come to the rescue. If we include the second-order quantifiers among the logical constants, then claims like CH (or $\neg CH$) are declared logically true. But if we exclude them, many intuitively valid arguments are judged invalid.

Many logicians and philosophers react to my conceptual critique of Tarski's account by retreating to an extensional stance, saying the only thing that really matters is that the analysis be extensionally correct. I have no doubt that, when push comes to shove, Tarski would have said the same thing—as, in fact, would I. So let's try to assess the "material adequacy" of the definition, as Tarski would have put it. What can we say about the account from a purely extensional standpoint? We can say that it is an unqualified success in one case: propositional languages in which the atomic sentences are logically independent. *But that is about all we can say without adding significant caveats.* With a first-order language, the analysis fails unless we add an important modification whose consistency with the original analysis has yet to be explained.[13] Adding this

[13]Greg Ray [27] argues that Tarski originally intended to employ varying domains, presumably to prove that the feature is consistent with the original account. But Ray's interpretation is inconsistent with the motivations Tarski gives for the reductive account, with Tarski's explicit description of the account, and with the consequences that he expressly draws from it. Yet even if Tarski himself were assuming varying domains without telling his readers, which he clearly was not, an explanation would be needed for what is in fact a radical departure from the core idea of the reductive analysis.

In what is surely one of the more interesting examples of defensive zeal, Ray claims that Tarski is wrong about one of the simplest consequences of the account (that logical consequence reduces to material consequence when all expressions are treated as logical constants), since this does not accord with the account Ray tries to impose on the article. Ray then goes on to accuse me of presenting an "invalid argument," which (we find out only in a footnote) is invalid because it takes Tarski at his word about this obvious consequence of the account [27, p. 648].

Why, according to Ray, would Tarski actually intend an account that is at such variance with

modification avoids some obvious extensional errors in first-order languages, but does not help when we move, for example, to second-order languages.

The issue of extensional adequacy is even more troubling than this tiny survey indicates. So far, I have focused on three rather similar languages: propositional, first- and second-order logic. When we widen the scope of our survey, the Tarskian analysis becomes increasingly implausible. For example, much of the most interesting work in logic during the past thirty years has grown out of so-called "index" or "possible world" semantics, pioneered by Saul Kripke, Stig Kanger, and others. This work includes modal logic, epistemic logic, temporal logic, deontic logic, the logic of indexicals, and so forth. Yet in none of these cases does the consequence relation studied admit of a plausible Tarskian characterization.

This deserves emphasis, since it is on the one hand so obvious, yet on the other, so consistently overlooked. If we were to follow Tarski's lead, the way to study the logic of, say, knowledge and belief, would be to treat these operators as logical constants and consider the argument forms whose instances, purely as a matter of fact, preserve truth. But this way madness lies. Contingent, but perfectly general facts about knowledge and belief, perhaps of the depressing sort studied by Kahneman and Tversky, would be enshrined as logically valid arguments of epistemic logic. For example, suppose that the inference from φ to ψ is a particularly subtle fallacy of first-order logic. Then to decide whether $Bel_a(\psi)$ is a logical consequence of $Bel_a(\varphi)$ we would have to find out if anyone—any actually existing believer—saw through the fallacy, that is, believed an instance of φ but not the corresponding instance of ψ. If so, it would not be a logical consequence; if not, it would. Similarly, to decide whether $Bel_a(\varphi)$ logically implies $\exists x(x \neq a \wedge Bel_x(\varphi))$, we would have to find out if any propositions are, as a matter of fact, believed by one and only one person. My guess is not, but that would not make this a logically valid inference, as Tarski's analysis implies. And of course we'd also have to settle the question of whether any propositions are believed by a thousand people but not a thousand and one, or a thousand and one but not a thousand and two. Since there are only finitely many believers, the answer would eventually be dubbed a logically valid inference, according to the reductive account. To take yet another example, we'd also have to determine whether anyone believes that there are more than three but fewer than seven things in the universe, for if not (and I suspect not), the negation of this claim would be a logical truth. And so forth. As I said,

his explicit description, with his express motivation, and with the consequences he draws from it? Because, Ray says (following Wilfrid Hodges [18] and Gila Sher [34]), he was addressing the paper to philosophers. It is interesting that Tarski should have been concerned that his philosophical audience would not understand the clause "and you must vary the domain of quantification," though he assumes they will follow his discussions of omega incompleteness, Gödel's theorems, satisfaction, and so forth. Is this concept really so difficult? It is even more interesting that he would provide a motivation at odds with his "real" account, and go on to draw consequences that follow from the stated account but not from the "real" account. One wonders if Tarski could have said anything to convince these commentators that he actually said what he meant. I urge readers interested in this exegetical issue to read Tarski's article; nothing I could add would provide a more convincing refutation.

this way madness lies.

Similar issues arise when we try to apply the reductive analysis to any of the other languages mentioned, from modal logic to the logic of indexicals. These applications immediately involve us in a host of empirical or quasi-empirical questions similar to those mentioned for belief. No one knows, or has ever tried to find out, what the actual extension of the Tarskian consequence relation would be in any of these cases. I leave it as an exercise for the reader to try out any of these applications to see why.[14]

This is not, of course, how these logics are investigated. Kripke semantics, in its many variations, bears no relation to the reductive account of logical consequence given by Tarski. To be sure, in Kripke semantics we use semantic techniques pioneered by Tarski to define the relation of truth in a structure, and we define logically true sentences to be those that come out true in every structure. But this vague similarity is as far as the resemblance goes. We conduct no investigations of which sentences involving knowledge and belief (or necessity and possibility, or "I," "here," and "now") are actually true—that is, true of actual knowers and believers out there in the actual world. Yet such issues would be an essential part of these investigations if Tarski's reductive analysis were correct. The answer, of course, is that in characterizing the logic of these languages we are doing something quite different. I will return to what that is in section 4. For now, what is important to recognize is that the reductive analysis is not used in studying these languages, and would not work if it were.

Tarski's reductive definition of consequence works for propositional languages with logically independent atomic sentences. It can be made to work, with some significant tinkering, for first-order languages (again, with logically independent predicates and functions), and certain close relatives of these.[15] But it fails as soon as we add any logically interesting expressions that go significantly beyond the truth-functional connectives, first-order quantifiers, and identity. It fails if those expressions are predicates and relations; it fails if they are non-truth-functional sentential operators; it fails if they are higher-order quantifiers. In all cases, it fails for exactly the reason explained in the first section: having uniformly truth preserving instances is no guarantee of logical validity. I think it is clear—even from a narrowly extensional standpoint—that the reductive account of consequence is a failure.

[14] One might think that of all these applications, modal logic would be the one most likely to succeed. But even here, we are immediately embroiled in substantive issues expressible using the modal operators that one does not ordinarily consider part of modal logic. For example, we would have to decide (or discover) whether there are any properties which one object has necessarily, but which another object has contingently. If not, then $\forall x(\Box Px \vee \Box\neg Px)$ will be a (Tarskian) consequence of $\exists y\, \Box Py$. This is not a logical consequence in any modal logic I am familiar with. Again, as with epistemic logic, there are a host of similar examples.

[15] The "close relatives" I am referring to are first-order languages supplemented with various numerical quantifiers. This is obvious in the case of quantifiers already definable in first-order logic, but interestingly, the account produces plausible results when we add the quantifier *there exist uncountably many* and even, I believe, the quantifier *there exist infinitely many*. Of course, in all of these cases we have to employ the modified reductive account, in which we vary the domain of quantification; the unmodified account gets the extension radically wrong.

3. Consequences of Tarski's account

The conclusions of the last paragraph may seem a damning indictment of Tarski's account of logical consequence. But in an odd quirk of intellectual history, these conclusions are actually embraced by the most ardent defenders of the reductive analysis. When you accept Tarski's analysis as capturing the essence of logical consequence, its haphazard behavior on most choices of logical constants gives rise to a seemingly important issue. That issue is sometimes raised with the question "What are the *genuine* logical constants?"; sometimes under the rubric "What is logic?"

What is really being asked here—though the defenders of the reductive account would never phrase it this way—is simply this: when does the Tarskian analysis get the extension of the consequence relation right? And not surprisingly, most of the answers we find in the literature are roughly the same as mine. What are the genuine logical constants? The truth-functional connectives, first-order quantifiers and identity, plus or minus epsilon. What is logic? First-order logic, give or take a bit.

I have already intimated what I consider the correct—indeed the obvious—answer to the question "What is logic?" Any language, regardless of its expressive devices, gives rise to a consequence relation, a relation that supports inferences from sentences in the language to other sentences in the language. The study of this relation is the study of the logic of that language. When Carnap, Kanger, and Kripke studied languages with modal operators, they were doing logic. When Hintikka applied similar techniques to epistemic notions, he was doing logic. When Kaplan investigated indexicals, he was doing logic. Second-order logic is logic (though the Continuum Hypothesis may not be). And these are logic not in a derivative or secondary or lesser sense: they are studying precisely the same thing we study in first-order logic, though in languages with additional expressive resources. Logic is not limited, *de jure*, by the expressive power of the devices in the language, as the reductive account unavoidably implies, though it may be limited, *de facto*, by the clarity of those devices and the availability of techniques to study them.

This is why the issue of the adequacy of the reductive analysis is important. It is not simply an abstract question about a piece of philosophical analysis. Accepting the faulty account leads to an extraordinarily limiting view of the appropriate subject matter of logic. Consider an analogy. Suppose in the early days of chemistry a technique had been developed that worked reasonably well in classifying inorganic compounds. Suppose further that this technique had been taken as definitive of the subject matter of chemistry: chemistry was just the study of those compounds that could be classified using the technique. But suppose the technique simply failed when applied to organic compounds, and as a consequence organic chemistry was declared "not chemistry." No doubt this would have impeded the development of organic chemistry, though I'm sure it would not have stopped it completely. Organic chemists would have pushed ahead, recognizing the importance of their work regardless of what it was called.

In many ways, this is similar to what has happened in logic. Lip service

is given to the Tarskian analysis of logical consequence, and to the extremely narrow view of logic that it implies. But much of the interesting work in logic is done outside the confines of that view. I have pointed to some important examples that are well known among philosophers and logicians, but these are only the tip of the iceberg. Let me gesture toward two additional examples of a very different sort. Both fall, by my lights, squarely within the legitimate boundaries of logic; neither admits of a Tarskian analysis.

I said above that any language gives rise to a consequence relation, and that this relation is a legitimate subject of logical investigation. I actually believe this is true of any well-defined system of representation, whether it takes the form of a traditional language or not. A good example is a database. A database stores information in a systematic format, and it is often an extremely important question whether a given piece of information is a consequence—yes, a logical consequence—of the information the database contains. A good deal of work has gone into the study of such questions, and recently into issues that arise when dealing with information contained in heterogeneous databases, where the same information may be represented in very different forms. For instance, one database may contain a field recording an individual's date of birth; another may record the person's age in years at the time of entry, along with a record of when the entry was made; a third might simply indicate whether the individual was a minor when the record was created. The information stored in any one of these databases bears a host of logical relations to the others. They can be inconsistent with one another; one may logically imply information that allows us to update another; and so forth.[16]

When characterizing the logic of a database, or of a collection of heterogeneous databases, the one thing we cannot ignore is the specific structure and interpretations of the various fields. Indeed, it is rarely the case that anything like the traditional logical constants are found as components of a database. The logical constants are often used in query languages designed for accessing information in a database, but almost never in the database itself. The important logical issues that arise here are not amenable to a Tarskian analysis. In fact the issues bear more relation to what Quine would disparage as "analyticity" or "analytic consequence." But this is just Quine's way of marking the artificial boundary that results from the reductive account of consequence.

The second example is closer to home—my home, at any rate. For the past ten years, Jon Barwise and I, along with many students and colleagues, have been studying the logic of various forms of graphical and diagrammatic representation.[17] Barwise and I are particularly interested in what we call het-

[16]Logicians who are loath to give serious consideration to representational systems other than traditional languages are inclined to say that databases are models. This is simply a confusion. Models are abstract, set-theoretic entities which we use to characterize the semantics of a system of representation. Databases, in contrast, are full-fledged representations. They have a semantics; they can be true or false, accurate or inaccurate; they bear logical relations both to other databases and to sentences in more traditional languages. It happens that they have what I have elsewhere [5] called a homomorphic semantics, but this does not make them models.

[17]See for example Barwise and Etchemendy [3, 4, 5, 6] and the papers collected in Allwein

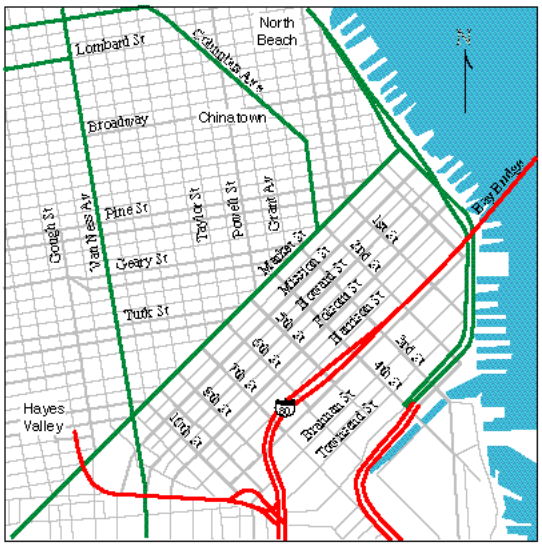

Figure 1: A map of downtown San Francisco.

erogeneous reasoning, reasoning that involves information provided in multiple forms. A simple example of such reasoning is the following. Suppose you are given two pieces of information: the map of San Francisco shown in Figure 1 and the assertion "The Old San Francisco Mint is at the corner of Mission and Fifth Streets." Here, now, is a quiz. Which of the following sentences follows logically from the information you've been given: "The Old Mint is south of Chinatown," or "The Old Mint is east of Golden Gate Park." If you know San Francisco, you may realize that both of these assertions are true. But whether you know San Francisco or not, you can see that only the first is a consequence of the information provided.

It takes only a moment's thought to appreciate how hopeless the Tarskian account of consequence is when applied to this sort of inference. What features of the map would be our candidate logical constants? At least when we discussed inferences involving predicates, non-truth-functional operators, or second-order quantifiers, our only problem was that the reductive account gets the wrong extension. With the present example, there is no clear way even to begin applying the analysis. Yet I dare say that inferences of this sort, and the logic that underlies them, are far more common in everyday life than those studied in first-order logic.

Much more could be said about this example, particularly about the implausibility of replacing such inferences with inferences characterizable in first-order logic, but that would take us away from the basic point. That point is this. Tarski's analysis of consequence is based on a simple mistake: the identifica-

and Barwise [1].

tion of the symptoms of consequence with their cause. When we accept this identification, based on the fact that the symptoms and cause happen to be co-extensive in a tiny collection of languages with very limited expressive resources, we risk missing the greater part of logic. Taken seriously, the analysis would rule out any reasonable treatment of modal, epistemic, deontic, or temporal logic. Taken seriously, it precludes the systematic study of the logic of predicates and relations, the logic of noun phrases, and the logic of at least some quantifiers. Finally, taken seriously it rules out the logical investigation of representational systems that take forms other than that of a traditional language, both those that have been around since before recorded history, like maps and diagrams, and those of more recent origin, like computer databases.

4. Model theory and the modeling perspective

In the years since Tarski published his article on logical consequence, model theory has become one of the dominant disciplines in mathematical logic. The history of model theory is complex, and includes much work that predates both Tarski's article on consequence and his seminal monograph on truth. Still, there is little question that model theory in its present form owes more to Tarski's work than to the work of any other single individual.

It is important to understand that my rejection of the reductive analysis of logical consequence is not an attack on model theory or model-theoretic semantics *per se*, but rather on a particular view of these techniques. In this section, I will try to make clear what I consider the proper understanding of model-theoretic techniques for studying the logic of a language. My explanation will, of necessity, be fairly dense, but I hope it is sufficient for those already familiar with standard applications of model theory. Readers not interested in model-theoretic semantics should feel free to skip to the final section.

When we give a model-theoretic semantics for a language, we characterize a class of set-theoretic objects alternatively called *structures*, *interpretations* or *models* for the language. Once these are described, we use Tarski's semantic techniques to define a relation between these structures and the sentences of the language, a relation known as *truth in a structure* and usually written $\mathfrak{A} \models \varphi$ for structures $\mathfrak{A}$ and sentences φ. A sentence is said to be logically true if it is true in all structures; a sentence φ is said to be a logical consequence of a set Σ of sentences if φ is true in every structure in which the members of Σ are all true, that is, if it preserves truth in every structure.

In CLC I described how the model theory for propositional and first-order languages can be seen as a more or less direct outgrowth of Tarski's reductive account of logical consequence. I devote a chapter of the book to this explanation, but I can describe the gist of it in a paragraph or two. The structures for these languages are quite simple. For the most part, they simply assign objects of a semantically appropriate type to certain expressions of the language. The expressions are the atomic sentences of propositional languages and the predicates, functions, and individual constants of first-order languages. These are, of course, exactly the expressions that in these languages are traditionally

considered non-logical constants. Now if we think of the non-logical constants as a special kind of variable and structures as assignments of values to these variables, the relation of truth in a structure is nothing more than the ordinary satisfaction relation: not truth *in* a structure, but satisfaction of an open formula by actual objects in the actual world.[18] Thus, on this view, a sentence φ is logically true just in case the universal closure $\forall v_1 \ldots \forall v_n \, \varphi$ that explicitly quantifies the special variables is simply true, just as Tarski's reductive analysis would have it.

When model-theory is seen through this lens, structures are often called *interpretations* of the language, since the assignment of a value to a non-logical constant can equally well be thought of as assigning an interpretation to the expression. If the expressions of the language have antecedent interpretations, the structure that assigns each non-logical constant its actual semantic value is called the *intended* interpretation of the language. The model-theoretic characterization of logical truth would, using this nomenclature, go like this: A sentence is logically true if it is true (in the actual world) no matter how the non-logical constants are interpreted. Again, the parallel with Tarski's reductive account should be apparent.

We have seen that the reductive analysis of the logical properties is mistaken. The same mistake is inherited by the interpretation of model-theoretic semantics just described. But this is a problem with the described interpretation, not with model theory itself. For there is an alternative view that makes perfectly good sense of model-theoretic practice—much better sense, in fact, than the Tarskian view. In CLC I called this alternative *representational* semantics and briefly described it in order to distinguish it from the Tarskian perspective. But I clearly did not say enough to forestall confusion, so let me try to rectify that here.

The guiding idea of the representational view of model theory is simple, and in fact widely held, though not widely articulated. The idea is this. The set-theoretic structures that we construct in giving a model-theoretic semantics are meant to be mathematical models of logically possible ways the world, or relevant portions of the world, might be or might have been. They are mathematical models in a sense quite similar to the mathematical models used to study, say,

[18] Dale Jacquette [19], Graham Priest [25], Gerhard Schurz [30, 31] and Gila Sher [34] all think that Tarski was assuming, or should have been assuming, that structures contain both existing and non-existing (merely possible, or perhaps even impossible) objects. I'm not sure which is more difficult to accept, the idea that we can build structures out of non-existent objects or the idea that Tarski had this in mind. Structures are built from actual objects, whether concrete or abstract, and the truth values had by sentences in those structures are determined by the actual properties and relations of those objects. I'm not sure how to make sense of the envisioned alternative. Are structures containing only non-existent objects actual, or are they too non-existent? If the former, this is a truly remarkable set-theoretic feat; if the latter, then do we have to revise Tarski's definition to quantify over all existing and non-existing structures? In any event, once we decide to appeal to non-existent objects, we forsake the principle benefit of the reductive account, the elucidation of a philosophically difficult notion by means of concepts that are significantly clearer and easier to understand. I do not deny, by the way, that we can use actual objects to *represent* alternative possibilities: I will say more about this in a moment.

the possible effects of carbon dioxide in the atmosphere, only they are used to study semantic phenomena, not atmospheric, and specifically to characterize how variations in the world affect the truth values of sentences in the language under investigation. The main difference is that in model theory we generally use discrete mathematics rather than the continuous mathematics used in physical modeling, though one can easily devise languages where discrete tools do not suffice. I called this view of model theory "representational" because the set-theoretic structures are seen as full-fledged representations: models of the world.

I will say more in a moment to add texture to the representational perspective, but for now let me finish this simple, initial sketch. According to the representational view, our goal in constructing a semantics is to devise a class of models that represents all logically possible ways the world might be that are relevant to the truth or falsity of sentences in the language, and to define a relation of truth in a model that satisfies the following constraint: a sentence φ should be true in model $\mathfrak{A}$ if and only if φ would be true if the world were as depicted by $\mathfrak{A}$, that is, if $\mathfrak{A}$ were an *accurate* model. The models are designed to represent the world in a particularly straightforward way, and this is important. Any individual model represents a logically possible configuration of the world and any two (non-isomorphic[19]) models are logically incompatible: at most one can be accurate. But jointly, they are meant to represent all of the possibilities relevant to the truth values of sentences in the language. In other words, if we've designed our semantics right, the models impose an exhaustive partition on the possible circumstances that could influence the truth of our sentences. Because of this, it is a trivial consequence that sentences which are true in every model are logically true, and arguments that preserve truth in every model are logically valid, at least if the representational criteria are genuinely satisfied.[20] Note, however, that this does not give us an analysis of the logical properties, since the logical notions are presupposed from the very start, in the criteria by which we assess our class of models. I will return to elaborate on this point later.

This is only a very rough sketch of the representational perspective. Let me add some texture to the sketch before discussing it in detail. When we study the logic of a language, we are generally interested in the logic of only some of

[19]Throughout this section, when I speak of "two" models, I presuppose the modifier "non-isomorphic." If I had more room to discuss representational semantics, I would explain how in certain semantics, non-isomorphic models can be representationally equivalent, that is, represent exactly the same possible circumstances. I will ignore this fact here, since it is irrelevant to the present issues. I also set aside issues that arise when we allow partial models in the semantics, or when we compare models from different semantics.

[20]I will set aside the important question of how we know our models actually depict every relevant possibility. Merely intending our semantics in this way is not sufficient, since limitations of our modeling techniques may rule out the depiction of certain possibilities, despite the best of intentions. This is arguably the case in the standard semantics for first-order logic, for example, where no models have proper classes for domains. Similarly, if we built our domains out of hereditarily finite sets we would have no model depicting an infinite universe. These are not problems with representational semantics *per se,* but with our choice of modeling techniques. Analogous problems arise in mathematical models of physical phenomena.

its expressions. In propositional logic we are interested in the truth-functional connectives; in first-order logic we additionally focus on the quantifiers and identity; in modal logic we add necessity and possibility; in epistemic logic, knowledge and belief; and so forth. Because we focus on only some of the expressions in the language, the semantics of the remaining expressions can be treated differently from those whose logic we aim to explicate. I will say we treat them "categorematically" for lack of a better term.

When we are not focusing on the logic of a particular expression or category of expressions, we need not model the specific semantic behavior of that expression. It is enough to characterize the minimal semantic behavior common to expressions of the same semantic category. This is what I mean by the categorematic treatment of an expression. In propositional logic, any sentence with no truth-functional structure is simply treated as providing a truth value, true or false, to the larger sentences of which it is a constituent. This is not to say there is no interesting or even logically relevant semantic behavior among these sentences, but only that we are not attending to it at present. Similarly, in first-order logic, where we are not concerned with the logic of predicates, a model can simply assign an arbitrary set to represent the semantic contribution of a monadic predicate, since all monadic predicates, whatever the details of their semantic behavior, will have some extension or other.

Note that we needn't treat all members of a given category categorematically just because we treat some that way. For example, in studying the logic of indexicals, Kaplan focuses on a handful of singular terms, "I," "here," and "now," while treating all other singular terms categorematically. In propositional logic, we treat sentences with no truth-functional structure categorematically, while giving sentences built using the truth-functional connectives a more detailed semantic treatment. In first-order logic, we characterize the specific semantics of the identity predicate, but treat the remaining binary predicates categorematically. Note also that an expression treated categorematically in one semantics may be given a detailed semantic treatment in another. Nothing special hangs on the choice of whether to treat an expression categorematically or not. It simply depends on which expressions we wish to focus on. This too is an important point that I will come back to later.

Now the fact that we treat expressions categorematically does not change the fundamentally representational perspective of the semantics we construct, though it can give rise to some confusion. What is confusing is that the categorematic treatment of certain expressions ignores any specific meanings these expressions may have, if in fact they were drawn from an antecedently interpreted language. For example, if we start with the language of elementary arithmetic or a first-order language containing the predicates $\simeq$ and $Triangle$ from our earlier example, there will be models that do not represent genuine possibilities—that is, possibilities consistent with the antecedent meanings of these expressions. For instance, there will be models in which an object a is in the extension assigned to $Triangle$, an object b is not in that extension, and yet the pair $\langle a, b\rangle$ is in the extension assigned to $\simeq$. But this is hardly surprising or problematic, given that we have only characterized the minimal semantic

behavior shared by all expressions of the respective categories. These models represent relevant possibilities for some expressions of those categories, though perhaps not all. This is analogous to what happens when, in a propositional semantics, we assign the truth value *false* to the atomic sentence $a = a$. This truth value assignment represents a genuine possibility for *some* atomic sentences, and since we are treating such sentences categorematically, the specific meaning of $a = a$ is irrelevant to the semantics.

How does the categorematic treatment of certain expressions affect the representational view of model theory? Well, not much at all. Where earlier we said that every model is meant to represent a logically possible configuration of the world, we now need to add a qualification to handle expressions treated categorematically. Each model is meant to represent a semantically relevant circumstance for at least *some* expressions in those categories—or, if you will, for *some* interpretations of the expressions so treated. This is a minor change, but it is potentially confusing because there are now two dimensions of variation. Structures still represent possible circumstances—this is the important dimension of variation—but our decision to treat some expressions categorematically introduces a second dimension, since these expressions receive uniform treatment in spite of potentially significant variations in their meanings. Thus in propositional logic, we afford $a = a$ the same treatment as $Triangle(a)$, despite the fact that the first may express a logical truth, while the second may express a contingent claim about the world.

From the representational standpoint, there is only one significant effect of the categorematic treatment of expressions. If we do not treat any expressions categorematically and our semantics meets the representational guidelines—that is, if every logically possible circumstance is represented by some model and our definition of truth in a model is correct—then we can be sure that *all and only* logically valid arguments of the language will be declared valid by the semantics. But suppose we are dealing with an antecedently interpreted language, and yet treat some expressions categorematically. Then it is always possible that our semantics will not declare some genuinely valid arguments valid, namely those whose logical validity depends on specific meanings our semantics ignores. Again, there are no surprises here. In propositional logic, the following argument is not revealed to be logically valid:

$Triangle(a)$
$a = b$
So, $Triangle(b)$

But the validity of this argument emerges as soon as we give a more detailed semantics, one that does not simply treat the sentences categorematically.[21]

[21] In CLC I discussed the representational view of model-theoretic semantics in some detail, and described how it differs from the Tarskian perspective. But I did not elaborate on the common practice of treating certain expressions categorematically, since I assumed it would be clear how the practice fits into the representational perspective. But this clearly confused some commentators. For example, Gila Sher [34] claims that no one interprets model theory as a representational semantics, but her evidence is simply the categorematic treatment of

Now it is important to see that representational semantics is not simply a minor redescription of Tarski's reductive analysis of consequence, or of "interpretational semantics," as I called the view of model theory that emerges from that analysis. There are many ways in which they differ, but the most important is that they impose different and conflicting criteria of adequacy on the semantics. What this means is that a semantics that is acceptable from the representational stance may be completely unacceptable from the reductive stance, and conversely, one that satisfies all criteria from the reductive perspective may be entirely inadequate from the representational. It is like the difference between billiards and pool: there are obvious similarities, to be sure, but pretty quickly the difference in rules (and the presence or absence of pockets) can no longer be ignored.

Let's look at a few examples, since in fact the criteria diverge almost immediately. For example, I earlier alluded to the fact that on the reductive analysis of consequence, it is hard to understand why in the semantics for first-order languages we vary the domain of quantification. Certainly, there are languages with restricted quantifiers, and even languages (such as English) in which the restrictions may be determined contextually from one use to the next. But suppose we are interested in the consequence relation for a language in which $\forall$ really means "*for all*," a language in which this expression quantifies over everything that happens to exist. From the reductive perspective, if we treat this expression as a logical constant, we should fix its meaning and survey the argument forms whose instances uniformly preserve truth, for example any argument with β as its conclusion. Replacing the unrestricted quantifier with various restricted quantifiers, quantifiers that do not quantify over everything, seems clearly inconsistent with the stated goal of determining the unrestricted quantifier's logic. When we view the same issue from the representational perspective, however, the inconsistency immediately goes away. First-order structures, viewed as representations of the world, should of course have different domains: this is simply our way of representing the fact that, although the world is the size it is, this could have been different. The same feature that seems a straightforward violation of the reductive account is in fact demanded by the representational perspective.

A more illuminating example, or collection of examples, are the various Kripke semantics mentioned earlier. Let me describe these in representational terms, since I know of no other way to view them. In a Kripke semantics, a structure consists of a set of indices, I, plus a relation R on I that specifies whether one index is "accessible" from another. Each index is associated with

names and predicates in the standard first-order semantics. I think, on the contrary, that it is patently obvious that many logicians and most philosophers (including Sher) adopt a fundamentally representational stance, though they may be unclear how different this is from the Tarskian analysis of consequence. My reasons for saying this will become clear later, since most model-theoretic semantics can only be understood representationally. Another commentator, Manuel García-Carpintero, gives an excellent and thoughtful analysis of the intuitions underlying the representational semantics of first-order languages in [14]. My main disagreement with García-Carpintero is his assumption that this was what Tarski had in mind in his analysis.

what is in effect a first-order structure, specifying a domain, extensions of the predicates, and so forth. The members of I are for heuristic reasons called *possible worlds* (though they are in fact simply set theoretic objects of some sort), and one of them, sometimes denoted @, is singled out as the *actual world*. According to the heuristic, the first-order structure associated with a given index represents the non-modal facts of the possible world corresponding to that index. Thus the first-order structure assigned to the index @ represents the non-modal facts in the actual world. The remaining members of I, along with the accessibility relation R, are simply an ingenious way of representing modal (or epistemic, deontic, temporal, etc.) facts about the world. For example, if a particular state of affairs holds at an index i accessible from @, this represents that this is a possible (though perhaps not actual) state of affairs. In other words, an entire Kripke structure represents a world—a single world—replete with both modal and non-modal facts.[22]

The crucial feature of a Kripke semantics is that for any logically possible configuration of the world, including both modal and non-modal facts (or epistemic and non-epistemic facts, etc.), there will be a Kripke structure representing that configuration. To hark back to our discussion of epistemic logic, there will be structures representing worlds in which $Bel_a(\psi)$ is true whenever $Bel_a(\varphi)$ is true, but also worlds in which this is not the case, however subtle the fallacious inference from φ to ψ may be. There will be worlds in which no one believes there are more than three but fewer than seven objects, but also worlds in which some people do. And so forth. What this means is that any sentence that comes out true in every Kripke structure must be true regardless of how the semantically relevant circumstances—in this case, epistemic and non-epistemic facts—happen to shake out.[23]

Kripke semantics obviously satisfies the guidelines for a representational semantics. Indeed, the great contribution of the semantics is that it gives us a remarkably flexible way to represent facts that play determining roles in the truth or falsity of sentences in a wide range of languages. But there is no sensible way to construe it as an application of Tarski's reductive account of consequence, as should be clear from our earlier discussion of these languages. To apply the reductive account, we would have to hold fixed the meanings of the operators in question and determine which arguments preserve truth—in the actual world—under various interpretations of the non-logical constants. A Kripke semantics does nothing even vaguely resembling this, and so by these criteria would have to be judged an out and out failure.

[22]Many people mistakenly believe that Kripke semantics commits us to the existence of possible worlds of some sort or other. This is just a confusion resulting from a simplistic view of the technique used in the semantics to represent modal (or epistemic, etc.) facts. The semantics is neutral about the issue of whether there are possible worlds in any ontologically significant sense; it simply uses the heuristic as a technique for representing various alternative modal facts.

[23]Similarly, returning to our example from footnote 14, there are Kripke structures in which $\exists y \Box Py$ is true but $\forall x(\Box Px \vee \Box\neg Px)$ is false, and others in which this is not the case, showing that the latter is logically independent of the former, in contrast to what the Tarskian account will say about such cases.

Once we appreciate how far removed representational semantics is from the reductive account of consequence, it becomes clear that there is no language, indeed no system of representation, whose logic cannot in principle be studied using model-theoretic techniques. For example, suppose we are interested in studying the logic of color predicates, perhaps in the context of a first-order language. Clearly the traditional first-order semantics, in which all predicates except identity are treated categorematically, would have to be supplemented. But it is not hard to see how the supplement might go. One simple option would be to assign to color predicates appropriate regions in a color space, and then have models map (some or all) objects in their domain to random points in color space. This would give us a more detailed representation of the range of logically possible circumstances relevant to the truth values of sentences involving these predicates. Naturally, important questions would arise in constructing such a semantics, but it is clear enough how it would be done. Again, the resulting semantics, like Kripke semantics, would bear no relation to Tarski's reductive account of consequence.

To take another example, I mentioned in the last section that it is unclear how we would even begin to apply the Tarskian analysis of consequence to diagrammatic forms of representation, since these are so different, both syntactically and semantically, from traditional languages. But this does not mean that a *representational* semantics is difficult to construct for these forms of representation. As long as we can devise model-theoretic techniques for representing circumstances relevant to the truth or falsity, accuracy or inaccuracy, of these types of representation, nothing prevents us from studying their logic as well. Such semantic accounts are no more difficult to provide than a model-theoretic semantics for traditional languages.[24]

Let me conclude this discussion by returning to a couple of issues touched upon earlier. The first is the difference between the categorematic treatment of expressions in a representational semantics and the distinction between logical and nonlogical constants central to Tarski's reductive analysis. I have already mentioned one respect in which these are very different. When we accept Tarski's analysis, it becomes an extremely important question which expressions are the legitimate logical constants, for choosing the wrong ones will yield a radically incorrect consequence relation. In representational semantics, in contrast, the decision to treat certain expressions categorematically is entirely arbitrary, depending only on whether we are interested in the logic of those expressions. Of course, we will sometimes treat expressions categorematically for important practical reasons—for example, we may not have a clue how to give a detailed treatment of their semantics—but there is nothing logically or philosophically significant about this choice. A second difference is that in Tarski's analysis, we genuinely hold fixed the meaning, both intension and extension, of the chosen logical constants. To treat "believes" as a logical constant, we must survey actual believers and actual beliefs. In representational semantics, even

[24]See for example Barwise and Etchemendy [5], Shin [35], and the papers collected in Allwein and Barwise [1].

the behavior of expressions not treated categorematically enjoys more flexibility. The fact that models are simply representations of semantically relevant circumstances allows us to survey alternative extensions of these expressions—not different meanings, but different ways the world might be. This is as it should be: we are studying the logic of these expressions, not general facts that may be expressed using them. Epistemic logic is not psychology, modal logic is not metaphysics, and second-order logic is not set theory.

Several commentators on CLC have argued that Tarski had in mind something like representational semantics when proposing his analysis, but this question should be finally laid to rest by the observations of the preceding paragraph. The fact that Tarski saw the choice of logical constants to be a crucial step in applying his analysis, the fact that he explicitly points out that the choice is not arbitrary, and finally his acknowledgment that logical consequence reduces to material consequence when all expressions are treated as logical constants, show that he could not have had in mind representational semantics. If we are engaged in representational semantics, none of this is even remotely the case.[25]

[25]Gila Sher, though she claims no one views model-theoretic semantics representationally, goes on to propose an interpretation of Tarski's analysis that looks suspiciously like a representational semantics with the categorematic treatment of names and predicates. But Sher continues to claim that the choice of logical constants is crucial, for reasons that are obscure. She begins by emphasizing the importance of the notion of formality: "Necessity. . . is by itself a problematic notion, but formality can be viewed as a modifier of necessity: not all necessary consequences are logical, only *formal*-and-necessary (or *formally* necessary) consequences are. The key to understanding logical consequence is, thus, formality." [34, p. 672.] This may seem a promising start, despite the rather abrupt "thus." But Sher goes on to describe a notion of formality which, among other things, implies that the formal rules of modal and epistemic logic are not, contrary to appearances, formal in the required sense. Predictably, it turns out that Sher's formality requirement is only satisfied by first-order logic and minor variants.

What is the relationship between formality and necessity that justifies Sher's "thus"? Sher explains it this way: "[The concern about non-logical generalizations] does not apply to my conception, where logical consequence is reducible not to just any kind of generality, but to a special kind of generality, namely, *formal and necessary* generality. Speaking in terms of models: Suppose there is an accidental property H, of all models for a given language. The notion of model is defined within some background theory, T, based on its notion of 'formal structure.' If T is an adequate theory of formal structure, then T includes the theorem 'Some formal structure A does not possess the property H' and, in accordance with this theorem, the apparatus of models defined in T will include a model representing a formal structure in which H does not hold." [34, p. 681.]

Sher is saying here that if there is an accidental or non-logical feature that holds of all models (say the Continuum Hypothesis or the absence of proper classes among the models' domains), then you simply need a background theory of formality that says the feature does not really hold of all models. But this is simply nonsense. Every class of models has such features—indeed infinitely many such features—including the models used in first-order logic. For example, if the Continuum Hypothesis is true (or false), that will have an effect even on what first-order models there are; similarly, in standard first-order model theory there are no proper classes among the domains; and so forth. Having a "theory of formal structure" that says these features are not there doesn't help, it simply means your theory is false. Or, to put the point another way, if an "adequate theory of formal structure" must be able to prove that no such features hold of the class of models, then there can be no such (true) theory, any more than there can be a true theory proving that two plus two is five.

Any class of structures will have features that are not logically necessary. The crucial question is not whether there *are* such features, but whether the features are expressible using the chosen logical constants, and hence whether they have an impact on the extension of the

The second issue is the issue of analysis itself. If Tarski's analysis worked, it would be a genuine analysis, in the sense that it characterizes the logical properties in terms that do not presuppose those very same properties. As I mentioned earlier, representational semantics gives us no such analysis, since the logical notions are used to assess the class of models devised for the semantics. Each model is meant to depict a logical possibility; no two are logically consistent; and the "sum" of the models is logically necessary—that is, every semantically relevant possibility is represented. This has a consequence that some may find disappointing, though it should hardly be surprising. We cannot look to model-theoretic semantics to answer the most basic foundational issues in logic. For example, if we have serious doubts about whether the principle of excluded middle is a logical truth, the classical semantics for propositional languages will not provide an answer. For the same intuitions that suggest that it *is* a logical truth are used in defining the class of structures—truth value assignments—that are employed by this semantics.

Does this mean that model-theoretic semantics, construed representationally, provides no illumination about the logical properties of the languages studied? Not by any means. We can see how it illuminates these properties both concretely and abstractly. Concretely, a well-designed semantics shows us how the truth values of sentences in the target language vary as the non-linguistic facts represented by the structures vary, and accordingly explains persistent patterns of truth values that emerge due to the semantics of these sentences. Logical truth and logical consequence are just two such persistent patterns. Naturally, the explanation is only as clear as the semantics, and in particular relies on a clear understanding of how the structures used in the semantics are meant to represent possible circumstances. If, so to speak, the "semantics" of our models is obscure, this will detract from or even negate the explanatory power of the model-theoretic semantics. But assuming a clear understanding of the states of affairs depicted by our models, the semantics shows precisely how the logic of the language arises from the meanings of its constituent expressions, modulo any basic logical assumptions incorporated into the models themselves. For example, the classical semantics for propositional logic may not provide a fully grounded explanation of the principle of excluded middle, but it does explain why, given this basic assumption, a complex sentence like $\neg(P \wedge (\neg P \vee (Q \wedge R))) \vee Q$ is necessarily true. This provides illumination of a very real sort.

There is another, more abstract way to describe this illumination. In a model-theoretic semantics, although the class of models is itself a representational system, it is a system with a particularly simple logic—in fact the simplest logic possible: no model is logically true or logically false, no model follows logically from another, and so forth. I will say that the system of models is logically "transparent," since there are no non-trivial consequence relations between representations in the system. Thus in a representational semantics we describe the logical properties of a logically *complex* system of representation in terms

reductive definition. The only way to prove that they don't have such an impact is by means of a completeness theorem, as explained earlier.

of the logical properties of a transparent system of representation. We show why, for example, $Triangle(b)$ is a consequence of the premises $Triangle(a)$ and $a = b$ by characterizing the semantics of these sentences in terms of a class of representations in which there are no non-trivial consequence relations of this sort. Since we are presenting the semantics of our target language in terms of another representational system, this is the best we can possibly do.[26]

5. Concluding philosophical postscript

The reductive analysis of consequence is by no means a silly or trivially mistaken account. On the contrary, it is both attractive and plausible: attractive, because it promises to eliminate a host of obscure notions at the core of logic in favor of the vastly clearer notion of truth; plausible, because the definition is based on features that are indeed important characteristics of logically valid arguments. This is why the account has been put forward repeatedly, not only by Tarski, but in slightly different forms by Bolzano, Russell, Quine, and others. Yet in spite of its plausibility and attractiveness, the account is wrong: the identified features are not what underlie logical consequence, but merely symptomatic of the genuine relation. Sometimes these symptoms are coextensive with the cause, but more often they are not.

The main problem with the account, however, has nothing to do with subtle philosophical issues, but rather with the wide-ranging consequences of accepting the faulty analysis. In Section 3, I discussed the consequences for logic if we take the reductive account seriously. I also noted that the account is only given lip service among many working logicians, who of necessity abandon the analysis in order to study the logic of languages with expressive resources that go much beyond propositional or first-order logic. But the analysis has also had a significant impact in philosophy proper, perhaps even more so than in logic itself. Let me conclude this paper with some very brief remarks about the influence of the account on work in the philosophies of logic, mathematics, and language.

The influence of the reductive account has been most direct in the philosophy of logic, where the analysis provides the field with one of its principal problems. A great deal of effort has been devoted to the question of which expressions are "genuine" logical constants, and precisely what features make them so. This of course is an extremely important question if the reductive analysis is correct. After all, if the expression *if. . . then* turns out not to be a logical constant, then according to the reductive account *modus ponens* is not a logically valid argument form. On the other hand, if *believes* or *same shape* or *is red* turn out to be logical constants, then logic becomes, in effect, an empirical discipline. Sentences like:

> *No one believes there are more than three but fewer than seven objects.*

will then qualify as truths of logic. Once we accept the reductive account, the problem of the logical constants appears to hold the key to the difference be-

[26] For a more extensive discussion of ways in which model-theoretic semantics illuminates the semantics of a language, see Barwise and Etchemendy [2] and Etchemendy [11].

tween genuinely valid inference and inference that obviously is not. With stakes this high, this becomes a philosophical issue that demands attention.

Wittgenstein is well known for his claim that the problems of philosophy arise out of fundamental confusions, and that their proper solutions lie in clarifying those antecedent confusions. Personally, I think that this is not at all true of most philosophical problems. But the problem of the logical constants, and the closely related question of what is logic, are clear examples of Wittgenstein's claim. The problem arises for no other reason than our acceptance of a incorrect account of logical consequence. When we fail to distinguish the symptoms of consequence from genuine consequence, we are bound to get faulty results. The idea that these results are due to the correctness or incorrectness of our selection of logical constants is simply a misdiagnosis of what went wrong. Any expressive device—predicates, adverbs, indexicals, quantifiers—can in principle affect the logical properties of a language, can give rise to arguments that are guaranteed to preserve truth in virtue of the way those devices work. The expressions traditionally singled out in the argument forms studied by Aristotle or Boole or Frege or Gentzen are simply expressions whose logic is particularly clear, interesting, and widely applicable. These traditional constants will no doubt share many properties, as will any finite collection of expressions, but the idea that the properties they share are somehow *definitive* or *determinative* of logic is based on a confusion.

The reductive account of consequence has had an equally extensive influence in the philosophy of mathematics, though the influence is more diffuse. Most of the influence comes via the claim that logic is identical to first-order logic. We have seen why this view seems inevitable given the reductive account of consequence: as soon as we venture very far from the expressive resources of first-order logic, the resulting "logical consequence" relation bears little or no relation to logic—not due to the real logic of these expressions, but due to the faulty analysis of consequence. The identification (or misidentification) of logic with first-order logic has important consequences for how we understand the nature of mathematics and mathematical truth. To take the most obvious example, the logicist claim that arithmetical truth is reducible to logical truth is clearly false if we accept this identification. But it is arguably true when we consider the logic of languages containing more powerful expressive devices. Of course the conclusion that the reduction is possible in a more powerful language may not carry with it some of the epistemological benefits envisioned by the early logicists—there is no getting around Gödel's incompleteness theorems—but it may nonetheless provide illumination about the nature of mathematical truth. I do not pretend to have solutions to the longstanding debate inspired by the logicist's claim. But it is obvious that sorting these issues out requires a reasonably clear understanding of logical truth and logical consequence, not one based on an analysis incapable of dealing with more powerful logics.

A very different example is our understanding of geometrical reasoning, which has been hampered by the absence of any account of valid reasoning that involves diagrammatic or other non-linguistic forms of representation. As long as we adhere to the reductive account of logical consequence, we will never

make progress understanding this sort of reasoning, for reasons discussed in Section 3. This point in fact applies much more widely than the philosophy of mathematics. Logic is in part a service discipline, providing precise, idealized models of valid and invalid reasoning, models which in turn help us describe and understand the process of rational investigation, whether in mathematics, the sciences, or everyday life. To the extent that the models we develop fail to address important types of deductive reasoning, we make the task of philosophers investigating those domains correspondingly difficult. Since the most highly developed model of deductive reasoning is that provided by first-order logic, philosophers naturally try to model, say, scientific reasoning by applying the notions derived from this theory. But it is likely that, as in the case of geometrical reasoning, the first-order model is inadequate to capture significant portions of the deductive reasoning that takes place in these disciplines.

I have already alluded to the impact of the reductive account in the philosophy of language. Here, the influence flows largely from the work of Quine. Quine and his followers have long disparaged the notions of analytic truth and analytic consequence, arguing that it is impossible to sensibly distinguish analytic truths from deeply held empirical beliefs, that the distinction is simply an unfounded "dogma" of empiricism. But most followers of Quine pull their punches when it comes to logical truth and logical consequence: these notions, unlike analyticity, can be clearly and definitively characterized by means of the reductive analysis, or so the Quinean would like to believe. This allows them to assume the legitimacy of logic—or at any rate, first-order logic—while denying the legitimacy of any appeal to the analytic/synthetic distinction, or to a full-bodied conception of meaning.

Needless to say, a detailed analysis of Quinean epistemology and philosophy of language is far beyond the scope of the present article. But it should be clear that the distinction between analyticity and first-order logic, or between first-order logic and more powerful logics, is brought into question once we recognize the defect in the reductive analysis. Quine's attack on analyticity applies equally to the notions of logical truth and logical consequence (as Quine himself sometimes acknowledges). My own view is that the attack fails in both cases, but a Quinean can consistently maintain that it succeeds in both, that the concept of logical consequence, like that of analyticity, is an unfounded dogma of empiricism. Whether he would be willing to accept the model of rationality that emerges—wherein the web of belief comes to resemble a disconnected pile of sand—is an open question. But these are questions that must be postponed until later.

References

[1] Allwein, Gerard, and Jon Barwise, eds. *Logical Reasoning with Diagrams.* Oxford University Press, 1996.

[2] Barwise, Jon, and John Etchemendy. "Model-theoretic Semantics." In *Foundations of Cognitive Science*, Michael Posner, ed., MIT Press, 1989, 207-243.

[3] Barwise, Jon, and John Etchemendy. "Visual Information and Valid Reasoning." In *Visualization in Mathematics*, Walter Zimmermann and Stephen Cunningham, eds., Mathematical Association of America, 1991, 9-24. Reprinted in [1].

[4] Barwise, Jon, and John Etchemendy. "Hyperproof: Logical Reasoning with Diagrams." In *Proceedings of the 1992 AAAI Spring Symposium on Diagrammatic Reasoning*, AAAI, 1992, 80-84. Reprinted in *Reasoning with Diagrammatic Representations*, AAAI Press, 1994.

[5] Barwise, Jon, and John Etchemendy. "Heterogeneous Logic." In *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, Janice Glasgow, N. Hari Narayanan and B. Chandrasekaran, eds., MIT Press, 1995, 211-234.

[6] Barwise, Jon, and John Etchemendy. "Computers, Visualization, and the Nature of Reasoning." In *The Digital Phoenix: How Computers are Changing Philosophy*, T. W. Bynum and James H. Moor, eds., Blackwell, 1998, 93-116.

[7] Blanchette, Patricia. "Models and Modality." Forthcoming.

[8] Chihara, Charles. *The Worlds of Possibility.* Oxford University Press, 1998.

[9] Chihara, Charles. "Tarski's Thesis and the Ontology of Mathematics." In *The Philosophy of Mathematics Today*, Matthias Schirn, ed., Clarendon Press, 1998, xxx-xxx.

[10] Curtis, Gary. "Review of *The Concept of Logical Consequence*, by John Etchemendy." *Nous* 28 (1994): 132-135.

[11] Etchemendy, John. "Models, Semantics and Logical Truth." *Linguistics and Philosophy* 11 (1988): 91-106.

[12] Etchemendy, John. "Tarski on Truth and Logical Consequence." *Journal of Symbolic Logic* 53 (1988): 51-79.

[13] Etchemendy, John. *The Concept of Logical Consequence.* Harvard University Press, 1990. Reissued by CSLI Publications and Cambridge University Press, 1999.

[14] García-Carpintero Sánchez-Miguel, Manuel. "The Grounds of the Model-Theoretic Account of the Logical Properties." *Notre Dame Journal of Formal Logic* 34 (1993): 107-131.

[15] Gómez-Torrente, Mario. "Tarski on Logical Consequence." *Notre Dame Journal of Formal Logic* 37 (1996): 125-151.

[16] Hansen, William. "The Concept of Logical Consequence." *Philosophical Review* 106 (1997): 365-409.

[17] Hart, W. D. "Review of *The Concept of Logical Consequence*, by John Etchemendy." *Philosophical Quarterly* 41 (1991): 488-493.

[18] Hodges, Wilfrid. "Truth in a Structure." *Proceedings of the Aristotelean Society* 86 (1986): 135-151.

[19] Jacquette, Dale. "Tarski's Quantificational Semantics and Meinongian Object Theory Domains." *Pacific Philosophical Quarterly* 75 (1994): 88-107.

[20] Kreisel, Georg. "Informal Rigour and Completeness Proofs." Reprinted in *Problems in the Philosophy of Mathematics*, Imre Lakatos, ed., North Holland, 1969, 138-171.

[21] McGee, Vann. "Review of Etchemendy, *The Concept of Logical Consequence*." *Journal of Symbolic Logic* 57 (1992): 254-255.

[22] McGee, Vann. "Two Problems with Tarski's Theory of Consequence." *Proceedings of the Aristotelean Society* 92 (1992): 273-292.

[23] McGee, Vann. "*The Concept of Logical Consequence* and the Concept of Logical Consequence." San Francisco: American Philosophical Association, Pacific Division Meeting, March 1993.

[24] O'Hair, Greg. "Logical Consequence and Model-Theoretic Consequence." *Logique et Analyse* 35 (1992): 239-249.

[25] Priest, Graham. "Etchemendy and Logical Consequence." *Canadian Journal of Philosophy* 25 (1995): 283-292.

[26] Ray, Greg. "On the Possibility of a Privileged Class of Logical Terms" *Philosophical Studies* 81 (1996): 303-313.

[27] Ray, Greg. "Logical Consequence: A Defense of Tarski." *Journal of Philosophical Logic* 25 (1996): 617-677.

[28] Read, Stephen. "Formal and Material Consequence." *Journal of Philosophical Logic* 23 (1994): 247-265.

[29] Sagüillo, José. "Logical Consequence Revisited." *Bulletin of Symbolic Logic* 3 (1997): 216-241.

[30] Schurz, Gerhard. "Logical Truth: Comments on Etchemendy's Critique of Tarski." In *Sixty Years of Tarski's Definition of Truth*, B. Twardowski and J. Wolenski, eds., Philed, Kraków, 1994, 78-95.

[31] Schurz, Gerhard. "Tarski and Carnap on Logical Truth—or: What is Genuine Logic?" In *Alfred Tarski and the Vienna Circle*, J. Wolenski and E. Köhler, eds., Kluwer, 1998.

[32] Shapiro, Stewart. "Logical Consequence: Models and Modality." In *The Philosophy of Mathematics Today*, Matthias Schirn, ed., Clarendon Press, 1998, 131-156.

[33] Sher, Gila. *The Bounds of Logic: A Generalized Viewpoint.* MIT Press, 1991.

[34] Sher, Gila. "Did Tarski Commit 'Tarski's Fallacy'?" *Journal of Symbolic Logic* 61 (1996): 653-686.

[35] Shin, Sun-Joo. *The Logical Status of Diagrams.* Cambridge University Press, 1994.

[36] Smiley, Timothy. "Consequence, Conceptions Of." In *The Routledge Encyclopedia of Philosophy*, Edward Craig, ed., Routledge, 1998, ???-???.

[37] Tarski, Alfred. "O pojciu wynikania logicznego." *Prezglad Filozoficzny* 39 (1936): 97-112. Translated as "On the Concept of Logical Consequence" in [38], 409-420.

[38] Tarski, Alfred. *Logic, Semantics, Metamathematics.* Oxford University Press, 1956.