Part V

CONCEPTS OF LOGICAL CONSEQUENCE

14

Necessity, Meaning, and Rationality: The Notion of Logical Consequence

STEWART SHAPIRO

There are many ways of saying that a given proposition, or sentence, Φ is a *logical consequence* of a set Γ of propositions, or sentences: Γ *entails* Φ, Γ *implies* Φ, Φ *follows from* Γ, Φ is a *consequence of* Γ, and the pair $\langle\Gamma,\Phi\rangle$ is *valid*. If a given Φ is a logical consequence of the empty set, we say that Φ is *logically true*, Φ is a *tautology*, or Φ is *valid*.

The notion of logical consequence has always been an important item on the agenda of philosophy. What is it? How do we determine that a given Φ is a consequence of a given Γ? Can we know this infallibly? A priori? What role does consequence play in our efforts to obtain knowledge? If Φ is a logical consequence of Γ, then what is the epistemic status of Φ vis-à-vis the members of Γ?

Logic is the study of correct reasoning, and has something to do with justification. Logical consequence is an important ingredient in *proof*. Thus, we broach issues concerning what *reasoning* is, and questions about what it is to reason *correctly*. The very notion of rationality is tied in here. Other slogans are that logic is topic neutral, and completely general, and that logical consequence is a matter of *form*. What do these slogans mean?

There are actually several different notions that go by the name of 'logical consequence,' 'implication,' etc. Some of them are controversial and some are, or may be, related to others in interesting and important ways. The purpose of this article is to sort some of this out.

1 Modality

As far as we know, the first systematic treatment of logic is found in Aristotle's *Prior Analytics*. In chapter 2 of book 1, we find:

> A deduction is a discourse in which, certain things having been supposed, something different from the things supposed results of necessity because these things are so. By "because these things are so," I mean "resulting through them" and by "resulting through them," I mean "needing no further term from outside in order for the necessity to come about".

I will not attempt to recapitulate the efforts of scholars to probe the subtleties in this text. To attempt a paraphrase, Aristotle's thesis is that a given proposition Φ is a consequence of a set Γ of propositions if (1) Φ is different from any of the propositions in Γ, (2) Φ *necessarily* follows from the propositions in Γ ('because these things are so'), and (3) propositions not in Γ are not needed for this necessity 'to come about.'

Contemporary practice is to drop clause (1) and allow that Φ follows from Γ when Φ is a member of Γ, as a trivial instance of logical consequence. Aristotle's phrase "because these things are so" seems to imply that in order to have a consequence, or 'deduction,' the premises in Γ must all be *true*. With one notable exception (Gottlob Frege), most modern conceptions of consequence do not follow this, and allow instances of logical consequence in which the premises are false. For example, 'Socrates is a puppy' follows from 'All men are puppies' and 'Socrates is a man.'

What of Aristotle's gloss of "because these things are so" as "resulting through them," and that as "needing no further term from outside in order for the necessity to come about." These clauses might indicate that the premises *alone* guarantee the conclusion, or that the premises are sufficient for the conclusion. Our first conception of logical consequence is modeled on this reading of Aristotle's definition:

(M) Φ is a logical consequence of Γ if it is not possible for the members of Γ to be true and Φ false.

It is common nowadays to think of modal notions in terms of possible worlds. For what that is worth, our thesis (M) becomes:

(PW) Φ is a logical consequence of Γ if Φ is true in every possible world in which every member of Γ is true.

According to (M) and (PW), 'Al is taller than Bill' seems to follow from 'Bill is shorter than Al,' since it is impossible both for Bill to be shorter than Al and for Al to fail to be taller than Bill. Surely, 'Al is taller than Bill' holds in every possible world in which 'Bill is shorter than Al.' Or so one would think. For another example, according to (M) and (PW), 'Hilary is wealthier than Barbara' and 'Barbara is wealthier than Nancy' seems to entail that 'Hilary is wealthier than Nancy.' Again, it is simply not possible for Hilary to be wealthier than Barbara, Barbara to be wealthier than Nancy, and for Hilary to fail to be wealthier than Nancy. To adapt an example from Bernard Bolzano, a religious person who accepts (M) or (PW) might say that 'Caius has an immortal soul' follows from 'Caius is a human' since (according to the person's theology), the premise *cannot* be true and the conclusion false.

On most contemporary accounts of logic, none of these conclusions is a logical consequence of the corresponding premise(s). It is a routine exercise to formalize these arguments and show that the conclusions do not follow (see below). Perhaps we can bring (M) and (PW) closer to the contemporary notions by articulating the involved modality, invoking a special notion of *logical* possibility and necessity. One tactic would be to invoke Aristotle's final clause, that a true logical consequence needs "no further term from outside in order for the necessity to come about." In the example about Bill and Al, we need to invoke some 'outside' fact about the relationship between shortness and tallness in order for the 'necessity to come about.' In the second example, we need the fact that relative wealth is transitive, and in the third example, our theologian needs to invoke some theology for the conclusion to follow. The idea is that 'Caius has an immortal soul' follows from 'Caius is a human' *together with* the relevant theology, but 'Caius has an immortal soul' does not follow from 'Caius is human' alone.

On the other hand, I would think that there just is no possible world in which Bill is shorter than Al without Al being taller than Bill. Al being taller than Bill is part of what it is for Bill to be shorter than Al. And I presume that most theologians would insist that there are no possible worlds in which the relevant theology is false. Having an immortal soul is part of what it is to *be a human*.

2 Semantics

According to Alberto Coffa (1991), a major concern of philosophers throughout the nineteenth century was to account for the necessity of mathematics and logic without invoking Kantian intuition. Coffa proposed that the most successful line came from the 'semantic tradition,' running through the work of Bolzano, Frege, and Ludwig Wittgenstein, culminating in the Vienna Circle. The idea is that the relevant necessity lies in the use of language, or *meaning*. This suggests the following proposal:

(S) Φ is a logical consequence of Γ if the truth of the members of Γ guarantees the truth of Φ in virtue of the meanings of the terms in those sentences.

Thesis (S) rules out our theological example. I presume that even the most religious linguist or philosopher of language does not take it to be part of the *meaning* of the word 'human' that humans have immortal souls. One can perfectly grasp the relevant meaning and not know the relevant theology. So according to (S), 'Caius has an immortal soul' does not follow from 'Caius is human.' We are, however, still left with our other two examples. According to (S), 'Al is taller than Bill' does indeed follow from 'Bill is shorter than Al,' since the meanings of 'taller' and 'shorter' indicate that these relations are converses to each other. Similarly, meaning alone determines that 'Hilary is wealthier than Barbara' and 'Barbara is wealthier than Nancy' together guarantee that 'Hilary is wealthier than Nancy.' The meaning of 'wealthier' indicates that it is a transitive relation.

The thesis (S) captures what is sometimes called 'analytic consequence,' which is often *distinguished* from logical consequence, due to examples like those considered here. We now turn to a refinement of the semantic idea.

3 Form

As noted above, there is a longstanding view that logical consequence is a matter of *form*. As far as I know, Aristotle does not explicitly endorse this, but his work in logic is surely consonant with it. He sometimes presents 'deductions' by just giving the forms of the propositions in them. Moreover, to show that a given conclusion does *not* follow from a given pair of premises, Aristotle typically gives an argument in the same form with true premises and false conclusion. It is most straightforward to interpret these passages as presupposing that if an argument is valid, then every argument in the same form is valid.

Consider a paradigm case of a valid argument:

All men are mortal; Socrates is a man; therefore, Socrates is mortal.

The validity of this argument does not turn on anything special about mortality and Socrates. Any argument in the form

All *A* are *B*; *s* is an *A*; therefore *s* is a *B*

is valid. That is, if one fills in the schematic letters *A*, *B* with any predicates or common nouns and *s* with any name or definite description, the result is a valid argument.

We might say similar things about the examples we used to illustrate the theses (M), (PW), and (S). Consider the following 'forms':

s is human; therefore *s* has an immortal soul
s is shorter than *t*; therefore *t* is taller than *s*
s is wealthier then *t*; *t* is wealthier than *u*; therefore *s* is wealthier than *u*

An argument in one of these forms has the same status, vis-à-vis (M), (PW), or (S), as the argument it was taken from. So one might think of these arguments as valid in virtue of form.

Nevertheless, the prevailing view is that the examples illustrating the above theses do not have a valid form. Although the one about shortness and tallness, for example, does not turn on anything special about the denotations of 'Bill' and 'Al,' it does turn on specific facts about (the meaning of) 'shorter' and 'taller.' On the prevailing view, the requisite logical forms of the above arguments are

s is *A*; therefore *s* is *B*
s is *S* than *t*; therefore *t* is *T* than *s*
s is *S* than *t*; *t* is *S* than *u*; therefore *s* is *S* than *u*

It is, of course, straightforward to find arguments in each of these forms that have true premises and false conclusion. So the original arguments are not valid in virtue of *these* forms.

At this point, a stubborn opponent might complain that even though the paradigm argument does not turn on anything special about Socrates, humanity, or mortality, it does turn on the specific meaning of 'all,' 'are,' and 'is.' We can give the following 'form' to our paradigm valid argument:

Π *A* is *B*; *s* is *A*; therefore, *S* is *B*,

and then give a counter-argument in that same 'form':

Some men are British; Clinton is a man; therefore Clinton is British

This has true premises and a false conclusion. So even the paradigm argument is not valid in virtue of the last-displayed 'form.'

The standard response would be to claim that the last-displayed 'form' is not a *logical* form of the paradigm argument. How, then, are we to characterize logical form? One might say that a form is logical if the only terms it contains (besides the schematic letters) are *logical terms*. Typically, these consist of truth-functional connectives ('not,' 'and,' 'or,' 'if . . . then'), quantifiers ('some,' 'all'), variables, and perhaps the sign for identity.

We now face the task of characterizing the logical terms. How do we go about designating a term as logical? The logician, or philosopher of logic, has three options. One is to attempt a principled definition of 'logical term,' perhaps by focusing on some of the traditional goals and purposes of logic (see, for example, Peacocke 1976; Hacking 1979; McCarthy 1981; Tarski 1986; Sher 1991). The proposals and theories cover a wide range of criteria and desiderata, such as a priori knowledge, analyticity, formality, justification, and topic-neutrality. It would take us too far afield to examine the proposals here. A second tactic, implicitly followed in most logic textbooks, is to merely provide a list of the logical terms, and to leave our task with this act of fiat. This is perhaps a safe route, since it avoids some sticky philosophical questions, but it might leave the readers wondering what is going on, and of course it provides no insight into the choice of logical terms. A third option (following Bolzano) is to make the notions of logical form and logical consequence relative. That is, one defines an argument to have a certain logical form *relative to* a given choice of logical terms. The same argument might be valid relative to one set of logical terms and invalid relative to another.

Some medieval logicians combined the notion of form with the *modal* conception of consequence. They defined a conclusion to be a *formal consequence* of some premises if (1) the argument is a consequence in the sense much like our (M) above, and (2) the result of any uniform substitution of the terms results in an argument that is also a consequence in that sense. In other words, suppose that Φ is a formal consequence of Γ. Then if Γ', Φ' are the result of a uniform substitution of the terms, then it is not possible for the members of Γ' to be true and Φ' false.

Our next notion of consequence combines the notion of logical form with a *semantic* conception like (S) above:

(FS) Φ is a logical consequence of Γ if the truth of the members of Γ guarantees the truth of Φ in virtue of the meanings of the logical terminology.

Another popular option is to avoid explicit mention of semantic notions like meaning altogether:

(Sub) Φ is a logical consequence of Γ if there is no uniform substitution of the nonlogical terminology that renders every member of Γ true and Φ false.

According to (Sub), logical consequence is defined solely in terms of logical form, substitution, and ordinary truth and falsehood. *Prima facie*, no metaphysically troublesome modal or semantic notions are involved.

4 Epistemic Matters

We still have not directly addressed the role of logical consequence in organizing and extending knowledge. As noted above, a common slogan is that logic is the study of correct *reasoning*. In particular, we reason from premises to conclusion via valid arguments. If we believe the premises, we must believe the conclusion, on pain of contradiction – whatever that means.

Let us propose another definition of consequence:

(R) Φ is a logical consequence of Γ if it is irrational to maintain that every member of Γ is true and that Φ is false. The premises Γ alone *justify* the conclusion Φ.

A theologian might admit that it is not irrational to hold that Caius is a human being without an immortal soul. The theologian should concede that someone can know that Caius is a human without knowing that he has an immortal soul. Some poor folks are ignorant of the relevant theology. So prima facie, the conception of consequence underlying (R) differs from the one underlying (M) above. On the other hand, there does seem to be something irrational in maintaining that Bill is shorter than Al while denying that Al is taller than Bill – unless of course one does not know the meaning of 'shorter' or 'taller.' But perhaps one can also rationally deny that Socrates is mortal while affirming that all men are mortal and Socrates is a man – if one pleads ignorance of the meaning of 'all.'

What is the penalty for being irrational? What exactly is the 'pain' of contradiction? The idea is that one who affirms the premises and denies the conclusion of a valid argument has thereby said things which *cannot* all be true. This broaches modal notions, as in (M) and (PW) from Section 1 above, but the pain of contradiction goes further than this. The charge is that our subject *could have known better*, and indeed should have known better. In this sense, logical consequence is a *normative* notion.

The most common way to articulate the modality and normativity here is in terms of *deduction*. If Φ is a consequence of Γ in this sense, then there should be a process of inference taking one from members of Γ to Φ. One purpose of such a deduction is to provide a convincing, final case that someone who accepts the members of Γ is thereby committed to Φ. So we have:

(Ded) Φ is a logical consequence of Γ if there is a deduction of Φ from Γ by a chain of legitimate, gap-free (self-evident) rules of inference.

Arguably, this notion also has its pedigree with Aristotle. He presents a class of syllogisms as 'perfectly' valid, and shows how to reduce other syllogisms to the perfectly valid ones by inference (see Corcoran 1974).

5 Recapitulation

I do not claim that the foregoing survey includes every notion of logical consequence that has been seriously proposed and maintained. For example, there is a tradition, going back to antiquity and very much alive today, that maintains that Φ is not a logical consequence of Γ unless Γ is *relevant* to Φ. But to keep the treatment from getting any more out of hand, we will stick with the above notions. Here they are:

(M) Φ is a logical consequence of Γ if it is not possible for the members of Γ to be true and Φ false.

(PW) Φ is a logical consequence of Γ if Φ is true in every possible world in which every member of Γ is true.

(S) Φ is a logical consequence of Γ if the truth of the members of Γ guarantees the truth of Φ in virtue of the meanings of the terms in those sentences.

(FS) Φ is a logical consequence of Γ if the truth of the members of Γ guarantees the truth of Φ in virtue of the meanings of the logical terminology.

(Sub) Φ is a logical consequence of Γ if there is no uniform substitution of the non-logical terminology that renders every member of Γ true and Φ false.

(R) Φ is a logical consequence of Γ if it is irrational to maintain that every member of Γ is true and that Φ is false. The premises Γ alone justify the conclusion Φ.

(Ded) Φ is a logical consequence of Γ if there is a deduction of Φ from Γ by a chain of legitimate, gap-free (self-evident) rules of inference.

Our next question concerns what to make of all these notions. They do not *seem* to be pointing in the same direction. Nevertheless, one might hold that there is but a single underlying notion of logical consequence. On this view, if there is a divergence between two of the above notions, then we must conclude that (at least) one of them is incorrect. It fails to capture the true notion of logical consequence. On the other hand, the logician might be more eclectic, proposing that there are different notions of consequence, some of which are captured by the above notions. In this case, of course, the various notions are not necessarily rivals, even if they differ from each other.

In any case, there are connections between the above notions. Trivially, if an argument is valid in the sense (FS) then it is valid in the sense (S). If the premises guarantee the conclusion in virtue of the meaning of the logical terminology, then the premises guarantee the conclusion in virtue of meaning. As we have seen, the converse of this fails in cases where premises guarantee a conclusion in virtue of the meanings of the *non-logical* terminology. If an argument is valid in the semantic sense (S) then presumably it is valid in the modal sense (M) (and perhaps (PW)). That is, if the meaning of the terms guarantees that the premises cannot all be true and the conclusion false, then surely it is not possible for the premises to be true and the conclusion to be false. The converse, from (M) to (S), depends on whether there are necessary truths that do not turn on the meanings of terms. Our foregoing theologian thinks that there are such truths (e.g. about immortal souls).

The relationship between (Sub) and (S) turns on the boundary between logical and non-logical terms and the expressive resources available in the base language. For example, suppose that we are dealing with a 'language' in which the only predicates are 'was US President sometime before January 1, 2000' and 'is male,' and the only singular terms are 'Bill Clinton' and 'Hilary Clinton.' Then the argument:

Bill Clinton is (or was) US President; therefore Bill Clinton is male

comes out valid according to (Sub). Any uniform substitution of the (available) non-logical terminology that makes the premise true also makes the conclusion true. But, of course, this argument is not valid on any of the other conceptions. Its coming out as a (Sub)-consequence turns on the fact that the 'language' in question is amazingly impoverished.

Suppose that we follow standard practice and assume (or stipulate) that the logical terminology consists of truth-functional connectives, quantifiers, variables, and the sign for identity. Consider the following argument:

(weird) for every x there is a z such that $x \neq z$; therefore for every x and every y there is a z such that $x \neq z$ and $y \neq z$.

The premise 'says' that there are at least two things and the conclusion 'says' that there are at least three things. Both are true. Notice that neither of these propositions contains any non-logical terminology. So there are no substitutions to make, and so the argument is valid according to (Sub). This is not a comfortable result. Clearly, it is not part of the *meaning* of the logical terminology that if there are at least two things then there are at least three things. So (weird) is not valid according to (S) or (FS). Whether (weird) is valid according to (M) depends on whether it is necessary that if there are two things then there are three things. Whether (weird) is valid according to (R) depends on whether one can rationally maintain that there are two things while denying that there are three things in the universe. I do not venture an opinion on these matters of metaphysics and epistemology.

Suppose that an argument is valid according to the modal conception (M), so that it is not possible for its premises to be true and its conclusion false. Does it follow that it is *irrational* to believe the premises and deny the conclusion? Can an argument be valid in the sense of (M) even if no one knows, or can know, that the argument is valid? Conversely, suppose that it is irrational to believe some premises and still deny a conclusion. Does it follow that it is *impossible* for the premises to be true and the conclusion false? The philosophical literature reveals no consensus on these matters, and I propose to stay out of the debates here.

Suppose that an argument $\langle\Gamma,\Phi\rangle$ is valid in the sense of (S) (or (FS)). Then if someone knows that each member of Γ is true, then she can determine that Φ is true just by reflecting on the meanings of the words. In other words, anyone who knows the language and also knows every member of Γ thereby has the wherewithal to know that Φ is true. To adapt Aristotle's phrase, nothing 'from outside' the premises is needed to determine the truth of the conclusion. Presumably, the meaning of the premises is not outside of them. Thus, it is *prima facie* irrational to believe that the premises are true and the conclusion false. So the argument is valid according to (R). Turning to the converse, suppose that it is irrational to believe some premises while denying a conclusion. Does it follow that the premises guarantee the conclusion in virtue of meaning? Once again, it depends on the nature of the underlying notions. Are there any beliefs whose irrationality does not turn on meaning?

In addition to the nature of rationality, any connections between (Sub) and (R) turn on issues concerning the logical/non-logical boundary and the expressive resources of the language. I leave this as an exercise.

Turning to the deductive notion (Ded), we encounter the notion of a legitimate, gap-free, self-evident rule. Another slogan of logic is that rules of inference are truth-preserving. This seems to entail that if a legitimate, gap-free (self-evident) rule of inference takes one from some premises to a conclusion then it not possible for the premises to be true and the conclusion false. Thus, if an argument is valid in the sense of (Ded) then it is valid in the sense (M).

In articulating (Ded), we can maintain the theses that consequence turns on meaning (S) and that consequence is a matter of form (FS) by insisting that the *only* legitimate, gap-free rules of inference are those that flow from the meaning of the logical terminology (see Hacking 1979; Tennant 1987).

So it is plausible that if an argument is valid in the sense (Ded) then it is valid in the senses (M), (PW), (S), and (FS). W. V. O. Quine argues that if the logical/non-logical boundary is chosen judiciously and the language has sufficient expressive resources (as above), then an argument is valid in the sense (Ded) only if it is valid in the sense (Sub).

The converses of these implications are more problematic. Are there necessary truths that are not knowable via a derivation using only legitimate, gap-free self-evident rules? If so, then there are arguments that are valid in the sense (M) (and (PW)) but invalid in the sense (Ded).

What if the necessity in question turns on meaning alone (as in (S)), or what if the necessity turns on the meaning of the logical terminology? In that case, can we conclude that there is a chain of legitimate, gap-free, self-evident rules that go from premises to conclusion? This depends on whether all truths concerning meaning can be negotiated via the requisite type of derivation. I reiterate the emerging policy of not taking sides on debates like this.

The notions (Ded) and (FS) are equivalent (at least in extension) if the meaning of every logical term is exhausted by legitimate, gap-free (self-evident) rules of inference involving the term. This is also a matter of controversy. Some philosophers claim that a term is logical only if its meaning is determined completely by matching introduction and elimination rules (Hacking 1979; Tennant 1987). This view rules out the sort of non-effective consequence relation advocated by other philosophers and logicians (see Shapiro 1991: chapter 2).

6 Mathematical Notions

The foregoing notions, from (M) to (Ded), are intuitive conceptions of logical consequence, dealing with either sentences in natural languages or propositions expressed by such sentences. Most textbooks in logic give scant treatment to these intuitive notions, and quickly move to developing a formal language, which is a rigorously defined set of strings on a fixed alphabet. The books then focus exclusively on this 'language,' and at least seem to leave the intuitive notions behind. The resulting mathematics is, of course, interesting and important, but we can query its philosophical ramifications.

Typically, parts of a formal language correspond, roughly, to certain parts of a natural language. Characters like '&,' '$\vee$,' '$\rightarrow$,' '$\neg$,' '$\forall$,' and '$\exists$' approximately correspond to the English expressions 'and,' 'or', 'if . . . then,' 'it is not the case that,' 'for every,' and "there is," respectively. As above, these are logical terms. Some formal languages include specific non-logical terms, such as the sign for the less-than relation over the natural numbers, but it is more common to include a stock of schematic letters which stand for arbitrary, but unnamed, non-logical names, predicates, and functions. So one can think of a formula of a formal language as corresponding to a logical *form* in a natural language (or in the realm of propositions). The correspondence thus engages the slogan that logic is a matter of form (as in (FS)).

Let γ be a set of formulas and ϕ a single formula of a formal language. A typical logic text formulates two rigorous notions of consequence, two senses in which ϕ follows from γ. For one of the notions of consequence, the author presents a *deductive system* S, which might consist of a list of axioms and rules of inference. An argument $\langle\gamma,\phi\rangle$ in the formal language is *deductively valid* (via S) if there is a sequence of formulas in the formal language ending with ϕ, such that each member of the sequence is either a member of γ, an axiom of S, or follows from previous formulas in the sequence by one of the rules of inference of S. If $\langle\gamma,\phi\rangle$ is deductively valid via S, we write $\gamma \vdash_S \phi$, or simply $\gamma \vdash \phi$ if it is safe to suppress mention of the deductive system.

The other rigorous notion of consequence invokes a realm of *models* or *interpretations* of the formal language. Typically, a model is a structure $M = \langle d,I\rangle$, where d is a set, the *domain* of M, and I is a function that assigns extensions to the non-logical terminology. For example, if c is a constant, then Ic is a member of the domain d, and if R is a binary predicate, then IR is a set of ordered pairs on d. Then one defines a relation of *satisfaction* between interpretations M and formulas ϕ. To say that M satisfies ϕ, written $M \models \phi$, is to say that ϕ is true under the interpretation M.

Finally, one defines ϕ to be a *model-theoretic* consequence of γ if every interpretation that satisfies every member of γ also satisfies ϕ. In other words, ϕ is a model-theoretic consequence of γ if there is no interpretation that satisfies every member of γ and fails to satisfy ϕ. In this case, we write that the argument $\langle\gamma,\phi\rangle$ is model-theoretically valid, or $\gamma \models \phi$.

Model-theoretic consequence and deductive validity (via S) are both sharply defined notions on the formal language. So relations between them are mathematical matters. The system is *sound* if every deductively valid argument is also model-theoretically valid, and the system is *complete* if every model-theoretically valid argument is also deductively valid.

Typically, soundness is easily established, by checking each axiom and rule of inference. Completeness is a usually a deep and interesting mathematical result. Virtually every system presented in a logic text is sound. Gödel's (1930) completeness theorem entails that first-order logic (with or without identity) is complete. Second-order logic is not complete (see Shapiro 1991: chapter 4).

To begin an assessment of the philosophical import of the technical work, one must explore the relation between the rigorous notions (of deductive and model-theoretic consequence) and the intuitive notions broached above ((M) to (Ded)). Probably the closest conceptual connection is that between the deductive notion of consequence (Ded) and deductive validity via a standard deductive system. In so-called "natural deduction" systems each rule of inference corresponds to a legitimate, gap-free (self-evident) inference in ordinary reasoning. So if an argument $\langle\gamma,\phi\rangle$ in the formal language is valid via such a system, and if a propositional or natural language argument $\langle\Gamma,\Phi\rangle$ corresponds to $\langle\gamma,\phi\rangle$, then Φ is a consequence of Γ in the sense (Ded). Although it is not quite as straightforward, something similar holds for other deductive systems. One typically indicates how each rule of inference corresponds to a chain of legitimate, gap-free inferences concerning ordinary reasoning, and that each axiom can be established by such rules.

Let *S* be a fixed, standard deductive system. One would like to establish a converse to the above conditional linking (Ded) to deductive validity via *S*. Call the following biconditional Hilbert's thesis:

> There is a deduction of a proposition (or natural language sentence) Φ from a set Γ of propositions (or natural language sentences) by a chain of legitimate, gap-free, self-evident rules of inference if and only if there is a corresponding argument $\langle\gamma,\phi\rangle$ in the formal language such that $\langle\gamma,\phi\rangle$ is deductively valid via *S*.

Perhaps one might restrict Hilbert's thesis to cases where the 'chain of gap-free, self-evident rules of inference' flow from the meaning of terminology that corresponds to the logical terminology of the formal language. This would focus attention on arguments that are valid in virtue of their logical form.

The philosophical interest of formal deductive systems depends on something like Hilbert's thesis. If there is no interesting connection between (Ded) (or (R)) and formal deductive validity, then the technical work is a mere academic exercise. Hilbert's thesis is the same kind of thing as Church's thesis, in that it identifies an intuitive, pre-theoretic notion with a precisely defined mathematical one. The exact nature of the identification depends on the relationship between formulas in the formal language and propositions or natural language sentences (or whatever it is that (Ded) applies to). At the very least, deductive validity via *S* is meant as a good mathematical model of (Ded).

Let us turn to model-theoretic consequence. The technical notion of satisfaction is a relation of truth-under-an-interpretation. Roughly, the relation $M \models \phi$ says that if the domain of *M* were the whole universe and if the non-logical terms are understood according to *M*, then ϕ is true. So model-theoretic consequence recapitulates the slogan that logical consequence is truth-preserving.

One might think of an interpretation as a possible world, which would link model-theoretic consequence to the modal notion (PW) and thus to (M). However, the complete freedom one has to 'interpret' the non-logical terminology (in the realm of model-theoretic interpretations) does not sit well with the modal notions. Consider one of our standby arguments: 'Hilary is wealthier than Barbara; Barbara is wealthier than Nancy; therefore Hilary is wealthier than Nancy.' A straightforward formalization would be *Whb*; *Wbn*; therefore *Whn*. To see that this formal argument is not model-theoretically valid, consider an interpretation whose domain is the natural numbers, and where *W* is 'within 3' (so that *IW* is $\{\langle x,y\rangle : |x-y| \geq 3\}$); *Ih* is 0; *Ib* is 2; and *In* is 4. This interpretation satisfies (i.e. makes true) the premises but not the conclusion. However, this interpretation has nothing to do with the *modal* status of the original argument about the relative wealth. It does not represent a genuine possibility concerning the relative wealth of those women. In terms of (PW), the given interpretation does not correspond to a genuine possible world.

For much the same reason, model-theoretic consequence systematically diverges from the semantic notion (S), according to which Φ is a logical consequence of Γ if the truth of the members of Γ guarantees the truth of Φ in virtue of the meanings of the terms in those sentences. Again, it is part of the meaning of 'wealthier' that the relation is transitive. This feature of the meaning is lost in the given interpretation of the formal argument over the natural numbers.

Model-theoretic consequence does better with (FS): Φ is a logical consequence of Γ if the truth of the members of Γ guarantees the truth of Φ in virtue of the meanings of the *logical* terminology. Within the framework, the extension of the nonlogical terminology varies from interpretation to interpretation. So one can claim that the model-theoretic validity of a given formal argument $\langle\gamma,\phi\rangle$ is independent of the meaning of the nonlogical terminology. So, to the extent that model-theoretic validity depends on meaning, it depends only on the 'meaning' of the logical terminology in the formulas.

But does model-theoretic consequence depend *only* on meaning (as required for (FS))? Recall that the different interpretations have different domains. This feature of model-theoretic semantics does not seem to correspond to anything in (FS). Why should we vary the domain, in order to determine what follows from what *in virtue of the meaning* of the logical terminology? What do the varying domains have to do with meaning at all? One might show that the variation in the domains keeps arbitrary and nonlogical features of the universe (such as its size) from affecting logical consequence, ruling out arguments like the above (weird). But we need an argument to establish a direct link between the semantic notion of meaning and the variation of domains from interpretation to interpretation.

The fact that each interpretation has a domain, and that different interpretations can have different domains, does fit in nicely with the *modal* notion (M). The domain corresponds to what the totality of the universe might be. If we think of an interpretation as a possible world (perhaps invoking the notion (PW)), then the domain would be the universe of that world.

So perhaps model-theoretic consequence corresponds to a blending of a modal notion like (M) or (PW) with the notion (FS) that revolves around logical form. We say that Φ is a logical consequence of Γ in this blended sense if it is not possible for every member of Γ to be true and Φ false, and this impossibility holds in virtue of the meaning of the logical terms. In the terminology of possible worlds, Φ is a logical consequence of Γ in this blended sense if Φ is true in every possible world under every reinterpretation of the non-logical terminology in which every member of Γ is true.

In sum, perhaps we have several intuitive notions of consequence corresponding, at least roughly, to the formal notions of deducibility in a deductive system and model-theoretic validity. From the soundness and completeness of first-order logic (with or without identity), we have the two rigorous notions corresponding to each other exactly. To stretch things a bit, the situation is analogous to that of Church's thesis, where it was shown that a number of different mathematical notions (recursiveness, λ-definability, Turing computability, Markov computability, etc.) each corresponding to a different pre-theoretic idea of computability, are all coextensive with each other. In the case of Church's thesis, this is sometimes taken to be evidence that all of the notions are correct

– that they do accurately capture the underlying notion of computability. Given the wide range of notions of consequence noted above (not to mention those not noted above), and the tenuous connections between them, we should not make too strong a conclusion here in light of completeness. Perhaps we can tentatively suggest that validity in a standard deductive system and model-theoretic validity correspond to something like a natural kind. The exact nature of this natural kind, and its relationship to notions like necessity, possibility, meaning, form, deduction, and rationality requires further study.

References

Coffa, A. (1991) *The Semantic Tradition from Kant to Carnap*. Cambridge: Cambridge University Press.

Corcoran, J. (1974) Aristotle's natural deduction system. In J. Corcoran (ed.), *Ancient Logic and its Modern Interpretations* (pp. 85–131). Dordrecht: Reidel, 85–131.

Gödel, K. (1930) Die Vollständigkeit der Axiome des logischen Funktionenkalkuls. *Montatshefte für Mathematik und Physik*, 37, 349–60; translated as "The completeness of the axioms of the functional calculus of logic," in Jean van Heijenoort (ed.), *From Frege to Gödel*, Cambridge, MA: Harvard University Press, 1967, 582–91.

Hacking, I. (1979) What is logic? *Journal of Philosophy*, 76, 285–319.

McCarthy, T. (1981) The idea of a logical constant. *Journal of Philosophy*, 78, 499–523.

Peacocke, C. (1976) What is a logical constant? *Journal of Philosophy*, 78, 221–40.

Shapiro, S. (1991) *Foundations without Foundationalism: A Case for Second-order Logic*. Oxford: Oxford University Press.

Sher, G. (1991) *The Bounds of Logic*. Cambridge, MA: MIT Press.

Tarski, A. (1986) What are logical notions. In John Corcoran (ed.), *History and Philosophy of Logic*, 7, 143–54.

Tennant, N. (1987) *Anti-Realism and Logic*. Oxford: Oxford University Press.

Further Reading

The following is only a sampling of the many articles and books on our subject. I apologize to neglected authors.

Anderson, A. and Belnap, N. (1975) *Entailment: The Logic of Relevance and Necessity I*. Princeton, NJ: Princeton University Press. (An extensive study of logical consequence, arguing that the premises of a logical consequence must be relevant to its conclusion.)

Anderson, A., Belnap, N. and Dunn, M. (1992) *Entailment: The Logic of Relevance and Necessity II*. Princeton, NJ: Princeton University Press. (A sequel to the above.)

Bolzano, B. (1837) *Theory of Science*, trans. R. George. Berkeley, University of California Press, 1972. (A presentation of a substitutional account of consequence, along the lines of (Sub).)

Corcoran, J. (1973) Meanings of implication. *Dialogos*, 25, 59–76. (Presents a lucid account of several different notions of consequence, focusing on the role of formal deductive systems and a relation like (Ded).)

Etchemendy, J. (1990) *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press. (A critical study of the influential account in Tarski (1935) and of contemporary model-theoretic consequence. This book generated many responses, including Sánchez-Miguel (1993), Sher (1996), and Shapiro (1998).

Quine, W. V. O. (1986) *Philosophy of Logic*, 2nd edn., Englewood Cliffs, NJ: Prentice-Hall. (An influential substitutional account, along the lines of (Sub), relating that conception to the formal notions of model-theoretic consequence and deducibility in a standard deductive system.)

Sánchez-Miguel, M. (1993) The grounds of the model-theoretic account of the logical properties. *Notre Dame Journal of Formal Logic*, 34, 107–31. (A nice study of the role of model-theoretic consequence, responding to Etchemendy (1990).)

Shapiro, S. (1998) Logical consequence: models and modality. In M. Schirn (ed.), *The Philosophy of Mathematics Today*. Oxford: Oxford University Press, 131–56. (A defense of model-theoretic consequence, relating it to the various intuitive notions.)

Sher, G. (1996) Did Tarski commit "Tarski's fallacy"? *Journal of Symbolic Logic*, 61, 653–86. (A reaction to Etchemendy (1990), relating Tarski's account to the model-theoretic one.)

Tarski, A. (1935) On the concept of logical consequence. In A. Tarski, *Logic, Semantics and Metamathematics*. Oxford: Clarendon Press, 1956; 2nd edn., edited and introduced by John Corcoran (Indianapolis: Hackett, 1983), 417–29. (A very influential article, developing a notion of consequence evolving from (Sub) but avoiding the tie to the expressive resources of the base language. The notion of satisfaction is introduced. The connection between this Tarskian notion and the contemporary model-theoretic one is a matter of controversy.)