Is Logic about Validity?

Timothy Williamson

[draft of chapter to appear in Elke Brendel, Massimiliano Carrara, Filippo Ferrari, Ole Hjortland, Gil Sagi, Gila Sher, and Florian Steinberger (eds.), *The Oxford Handbook of the Philosophy of Logic*. Oxford: Oxford University Press, slightly corrected on 21 November 2022]

A common understanding of logic is that, centrally, it is about *validity*. A recent proponent of this view is Graham Priest (2016: 8). Ole Thomassen Hjortland defends a similar view: 'a logical theory should be about validity, consistency, formality, truth preservation, provability, among other things' (2017: 641), which sounds somewhat broader, but in practice his discussion too gives pride of place to validity. Indeed, when one opens an elementary textbook of logic, one expects it to start with a rough explanation of the difference between valid and invalid arguments. Occasionally, a textbook starts with consistency rather than validity, but since each is definable in terms of the other, that makes little difference (Hodges 1977: 13).

When logicians study arguments, they normally treat them as linguistic constructs. Their premises and conclusions are sentences of a natural or formal language. Thus, the study of validity, consistency, and provability is a metalinguistic pursuit. If that is what logic is, all logic is metalogic. Conversely, since metalogic is part of logic, all metalogic is logic. On this view, logic and metalogic are identical; the 'meta' is redundant.

Such a view seems simplistic once one looks at logic as a discipline. It has its own academic journals and conferences, though very few of its own university departments: one might describe the discipline as semi-institutionalized. There are still some departments labelled 'Logic and Philosophy of Science', a legacy of logical positivism, though contemporary logic and contemporary philosophy of science are only tenuously linked. The articles in logic journals are mostly by authors in departments of mathematics or computer science or—less often—philosophy. Unsurprisingly, logic as a discipline is too diverse in its concerns for any monistic account of what it is about to be adequate.

For example, much of logic is *mathematical logic*, of which the four most salient branches are model theory, proof theory, set theory, and computability theory. Obviously, theorems are proved in all four of these branches, and the proofs need to be valid arguments, but that no more makes those branches *about* validity than it does any other branch of mathematics. When one studies recursive functions in computability theory, one is not primarily studying validity, or even provability. Model theory can reasonably be described as about validity, given the dominant model-theoretic account of validity (logical or semantic consequence) as truth-preservation in all models, but proof theory is not about validity in that sense. Some results in set theory are model-theoretic: for instance, the independence of the Continuum Hypothesis from first-order Zermelo-Fraenkel set theory with the Axiom of Choice. But set-theorists trying to prove or refute a conjecture within their favoured set theory may be primarily interested in the structure of the hierarchy of sets, not in validity. Neither set theory nor computability theory is happily described as a branch of metalogic. To insist that such activities are not 'genuine logic' would be a pointlessly tendentious use of the term 'logic'.

Someone might concede that not all mathematical logic is metalogic, but still hold that the most philosophically interesting part of logic is metalogic, and perhaps specifically the study of validity. This chapter concerns a philosophically more fundamental kind of logical inquiry, which is not metalogic. It is best understood as abstract metaphysics.

Section 1 explains how the idea of logic as metaphysics is a natural upshot of Tarski's famous account of logical consequence and logical truth, when we respect the mundane distinction between mentioning symbols and simply using them. Section 2 explains why the natural methodology for logic as metaphysics is an *abductive* one, like that familiar from theorizing in natural science. It shows how confusion between logic and metalogic has caused misunderstanding of methodological issues, particularly about the role of logical *strength* in evaluating logics.

1. *Logic, metalogic, and logical truth*

A simple logical truth is $\forall x\ x = x$, a formula of first-order logic with identity. The claim '$\forall x\ x = x$ is a logical truth' is a metalogical truth, but the formula itself is not a distinctively *metalogical* truth, since it belongs to the object-language and contains no distinctively metalinguistic vocabulary. By contrast, if we prefix it with the metalogical symbol $\vDash$ for logical truth, we get the formula $\vDash \forall x\ x = x$, which belongs to the formal metalanguage, not to the object-language.

In English, we can roughly paraphrase $\forall x\ x = x$ as 'Everything is self-identical'. To interpret it fully, we must clarify how widely the quantifier $\forall x$ or 'everything' is ranging. For present purposes, we may stipulate that, by default, quantifiers are to be read as completely unrestricted, and so as ranging over absolutely everything (see Williamson 2003 for a defence of absolute generality, and Rayo and Uzquiano 2006 for an introduction to the debate on it). Thus the formula says, without restriction: everything is self-identical.

Neither the formula $\forall x\ x = x$ nor the English sentence 'Everything is self-identical' is in any way meta-linguistic. The constituents of each are non-metalinguistic and occur in many entirely non-metalinguistic sentences, nor do those familiar ways of composing them import any metalinguistic element. That much is just a routine application of the standard distinction between use and mention. Of course, by a standard disquotational principle for truth, the non-metalinguistic statement 'Everything is self-identical' is equivalent (given the semantics of English) to the metalinguistic statement '"Everything is self-identical" is true', just as the non-metalinguistic statement 'There are mountains in Scotland' is equivalent (given the semantics of English) to the metalinguistic statement '"There are mountains in Scotland" is true'.

To say '"Everything is self-identical" is true' is not yet to say '"Everything is self-identical" is logically true'. After all, someone who accepts the truth of 'Everything is self-identical' might still wonder whether it is a *logical* truth, perhaps because they wonder whether identity is a matter of pure logic. However, identity has all the structural hallmarks typical of the standard logical constants, such as invariance under permutations of individuals, so for present purposes we can continue to treat it as one, and to assume that 'Everything is self-identical' or $\forall x\ x = x$ is indeed a logical truth.

Logic tells us that everything is self-identical. Someone might object that we don't need logic to tell us that, since it is obvious anyway. But that just reflects my choice of a very simple example; there is no end of ever more complex logical truths. Anyway, no logical truth is too simple to be denied by a sufficiently perverse metaphysician. An analytic philosopher once illustrated his difficulties in engaging fruitfully with his non-analytic colleagues, the majority in his department, by telling me that they all took it for granted that *nothing* is self-identical. At a guess, they thought that everything changes, and that no changing thing is self-identical, in effect because they committed the usual fallacy in applying Leibniz's law (the indiscernibility of identicals) to change over time and so reached the conclusion that it is inconsistent with numerical identity, although they would probably not have put it that way (the fallacy is diagnosed in Wiggins 1980). They did not mean something different from the rest of us by 'identical'. They were just not very good at logic. The logical message 'Everything is self-identical' really is news to some. It matters in metaphysics, even though getting it right does not take us far, for getting it *wrong* takes us so far astray.

To generalize the point, Tarski's seminal account of logical consequence provides a convenient starting-point (Tarski 1936). It is often treated as the origin of the contemporary model-theoretic account, on which a conclusion α is a logical consequence of a set of premises Γ (in other words, the argument from Γ to α is valid) if and only if every model of Γ is a model of α. The premises and conclusion are formulas of a given language. A model in the modern sense is a set-theoretic structure with a set domain over which the non-logical constants (atomic predicates and names) of the language are assigned extensions, and over which the quantifiers are interpreted as ranging. By recursive compositional clauses, the model determines a truth-value (truth or falsity) for each formula of the language (relative to an assignment of values over the domain to all variables). A model is a model *of* a formula just in case the formula is true in the model, and it is a model of a set formulas just in case it is a model of each formula in the set. Thus logical consequence is truth-preservation across all models.

Tarski was one of the founders of model theory as a branch of modern logic, and more specifically as a branch of modern metalogic. However, his original 1936 account of logical consequence differs significantly from the contemporary model-theoretic account. His 1936 models are not set-theoretic structures but simply assignments of values of appropriate types to all variables. There is no domain. Instead, the quantifiers are mandatorily interpreted as ranging unrestrictedly over the appropriate type, in the spirit of a simplified version of Russell and Whitehead's type theory in *Principia Mathematica*. In particular, the first-order quantifiers range over absolutely every individual. The omission of domains was no oversight or popularizing short-cut of Tarski's; he took careful account of it in his mathematical practice for some years subsequently (Mancosu 2006).

To determine whether a conclusion is a logical consequence of a set of premises, Tarski proceeded in 1936 like this. First, all non-logical atomic constants are replaced in all formulas of the language by new variables, replacing occurrences of the same constant by the same variable and occurrences of different constants by different variables. Let α' and Γ' result from applying this substitution to a formula α and a set Γ of formulas respectively. Assignments of values to variables assign values to the new variables too. Thus, each assignment determines a truth-value for α' and each member of Γ', since the only constants in them are logical ones, which are interpreted standardly. An assignment $\underline{a}$ is a *1936-model* of a formula α if and only if α' is true under $\underline{a}$; likewise, $\underline{a}$ is a 1936-model of a set of formulas Γ if and only if every member of Γ' is true under $\underline{a}$. One then stipulates that α is a logical consequence of Γ if and only if every 1936-model of Γ is a 1936-model of α. Logical consequence is truth-preservation across all 1936-models, as on the modern model-theoretic account, but with domain-free models. As a special case, one stipulates that α is a logical truth if and only if it is a logical consequence of the empty set of formulas, in other words, if and only if α' is true under every assignment (since every assignment is vacuously a model of the empty set).

Tarski's 1936 account, like the modern model-theoretic one, makes logical consequence and logical truth independent of the intended interpretations of the non-logical constants, since those constants are replaced by variables; only their logical type is retained. It makes logical consequence and logical truth purely *formal* matters, where 'form' abstracts from all non-logical aspects of meaning.

In 1936, Tarski explicitly refrains from providing a criterion for distinguishing between logical and non-logical constants. He leaves it open whether there is an objectively correct criterion, or whether instead the line has to be drawn *ad hoc*, perhaps in different places on different occasions, for pragmatic reasons. In much later work, he proposed a unique criterion: an expression is a logical constant just in case its extension remains constant under all permutations of individuals (Tarski 1986). The identity sign '=' is a logical constant by that criterion, as are the usual truth-functors and quantifiers.

A revealing special case for Tarski's account is where the argument from Γ to α contains no non-logical constants and no free variables. Thus the replacement operation makes no difference; Γ' and α' are just Γ and α. Since they contain no free variables, varying the assignment leaves the truth-values of α and the sentences in Γ fixed, so the generality over 1936-models in characterizing logical consequence becomes redundant. The upshot, as Tarski noted in 1936, is that in this special case α is a logical consequence of Γ unless α is false and every member of Γ true. In particular, when Γ is empty, α is logically true if and only if α is true. For example, since $\forall x\ x = x$ is a closed sentence with no non-logical constants, it is logically true because it is plain true.

A more striking example is $\exists x \exists y\ x \neq y$, another closed sentence with no non-logical constants, which says in effect that there are at least two things. It too is plain true, for there are indeed at least two things—for instance, Warsaw and Berkeley—so it too is logically true, on Tarski's 1936 account. In this respect, it contrasts sharply with a contemporary model-theoretic account, on which $\exists x \exists y\ x \neq y$ is not logically true, since it is false in any model with a one-membered domain.

Many philosophers of logic regard the contrast as a major advantage of the contemporary model-theoretic account over Tarski's 1936 account. Some would argue that logical truths are metaphysically necessary, whereas $\exists x \exists y\ x \neq y$ is metaphysically contingent. Others would argue that logical truths that knowable *a priori*, whereas $\exists x \exists y\ x \neq y$ is not knowable *a priori*. Still others would argue that logical truths are non-substantive, whereas $\exists x \exists y\ x \neq y$ is substantive.

Those objections are problematic for specific reasons: if elementary set theory is metaphysically necessary, it is metaphysically necessary that there are at least two individuals ({} and {{}}); if elementary set theory is knowable *a priori*, it is knowable *a priori* that there are at least two individuals; what counts as a substantive truth is utterly obscure. But the objections are also problematic for a more general reason: where are the alleged constraints on logical truth supposed to come from?

Philosophers of logic sometimes judge accounts of logical consequence as if by their fit with a pre-theoretic folk conception of logical consequence. But why think the folk have any such pre-theoretic conception? They may indeed have some pre-theoretic conception of *good reasoning* or *inconsistency*; that conception will presumably count arguing from 'This is scarlet' to 'This is red' as good reasoning, and accepting 'This is scarlet' while denying 'This is red' as inconsistent. But, it is generally agreed, the argument from 'This is scarlet' to 'This is red' is not *logically* valid, although it may be valid in some looser, non-logical sense, and accepting 'This is scarlet' while denying 'This is red' is not *logically* inconsistent, although again it may be inconsistent in some looser, non-logical sense. For logical connections are supposed to be *formal*, while the connection between 'This is scarlet' and 'This is red' is not formal; it depends on the specific meanings of the non-logical words 'scarlet' and 'red'. The folk had no pre-theoretic need of any standard of consequence higher than those looser, non-formal ones. Thus there is unlikely to be any pre-theoretic folk conception of specifically *logical* consequence.

Admittedly, Tarski himself opens his discussion thus (1936: 409):

> The concept of *logical consequence* is one of those whose introduction into the field of strict formal investigation was not a matter of arbitrary decision on the part of this or that investigator; in defining this concept, efforts were made to adhere to the common usage of the language of everyday life.

But he then notes the unclarity of ordinary language and warns (ibid.):

> Any attempt to bring into harmony all possible vague, sometimes contradictory, tendencies which are connected with the use of this concept, is certainly doomed to failure. We must reconcile ourselves from the start to the fact that every precise definition of this concept will show arbitrary features to a greater or less degree.

In the paper, Tarski argues for his preferred account against then-prevalent alternatives on technical, structural grounds, such as mathematical cases where logical consequence is clearly non-compact (a conclusion is a logical consequence of an infinite set of premises but of no finite subset of it), and the need for logical consequence to be robust with respect to accidental limitations in the expressive power of the language (it should conservative with respect to extensions of the language, so an argument in the old vocabulary is valid in the old language if and only if it is valid in the new language). He invokes no theoretically loaded metaphysical or epistemological constraint on logical consequence, and would surely have disapproved of doing so.

Probably, philosophers' expectations of logical consequence derive from traditional paradigms of logical reasoning, and so reflect older images of logic. But what authority have those older images? Logic has made more progress in the past century than in all of previous history. Like any other science, it is entitled to draw and study whatever distinctions it finds theoretically fruitful.

Of course, the flourishing contemporary sub-discipline of model theory is mainly about models with domains, so the later model-theoretic account of logical consequence has itself proved to be extremely theoretically fruitful—not least through Tarski's own contributions—and is clearly legitimate in its own right. That is not in question. Its emergence too had far more to do with mathematical naturalness than with any metaphysical or epistemological constraints. From a Tarskian perspective, the absence of theoretically motivated philosophical accretions from both the 1936 account and the later one is a positive advantage, yielding clean, spare, precise definitions of just the sort conducive to rigorous proofs of metalogical results. Which of them better deserves the title of 'logical consequence' is a question not worth disputing.

Tarski's 1936 account has its own distinctive interest. Indeed, its domain-free approach makes it even more attractively austere and simple than the later version. We can bring out its significance by focusing on the special case of logical truth.

Recall that, on the 1936 account, a formula α is a logical truth, true in all 1936-models, if and only if α' (the result of substituting variables for non-logical constants in α) is true on all assignments of appropriately typed values to variables. But that generality over assignments can be captured in the object-language, if it has universal quantifiers for all types for which the language has variables. We can add universal quantifiers for those types if the original language does not already have them. Let $\forall(\alpha')$ be the closed sentence which results from prefixing α' with unrestricted universal quantifiers for all its free variables (in some specified order). Then $\forall(\alpha')$ is true if and only if α' is true on all assignments. In other words, α is *logically* true if and only if $\forall(\alpha')$ is plain true: roughly, the logical truth of a formula is the truth of its universal generalization. But since the truth predicate behaves disquotationally, an ascription of truth to $\forall(\alpha')$ is equivalent to $\forall(\alpha')$ itself. Thus an ascription of logical truth to α is also equivalent to $\forall(\alpha')$, which is just a universal generalization in the (possibly expanded) object-language.

Typically, the formula $\forall(\alpha')$ contains no metalinguistic expressions, except in the unusual case where the original formula α contained some metalinguistic device (such as quotation marks or a truth-predicate) which is being treated as a logical constant. Normally, $\forall(\alpha')$ will not contain the kind of vocabulary capable of generating semantic paradoxes such as the Liar, and so will not be vulnerable to associated doubts about the disquotational equivalence of the sentence '$\forall(\alpha')$ is true' to $\forall(\alpha')$ itself.

To appreciate what is going on, an example will help. Let α be an instance of excluded middle, $A \vee \neg A$, where A is a sentence letter, a non-logical constant. Disjunction ($\vee$) and negation ($\neg$) are treated as logical constants. Then α' is $P \vee \neg P$, where 'P' is a sentential variable, and $\forall(\alpha')$ is $\forall P\ (P \vee \neg P)$, a fully general statement of the *law* of excluded middle, in the original object-language or an expansion thereof. We can paraphrase it as ‘Every state of affairs either obtains or does not obtain’. In one key respect, that paraphrase is misleading, for it suggests first-order quantification—quantification into *name* position—over special objects, states of affairs, a singular term for which combines with the monadic predicate ‘obtains’ to form a sentence—whereas the quantifier $\forall P$ binds a variable already in *sentence* position. English and other natural languages may lack the resources for a more accurate paraphrase, but that does not make the formula unintelligible on its intended reading, for we can get the hang of such locutions by the direct method of language learning (Williamson 2013: 235-40). More generally, such comments apply to attempts to paraphrase type-theoretic semantics in a natural language, and in particular to the required gloss that the quantification is over all entities of the relevant type, not just to those that happen to be expressible in the language. Quantification into sentence position or any other non-nominal position is not substitutional quantification; it is no more substitutional or somehow metalinguistic than quantification into name position. In particular, the law of excluded middle, $\forall P\ (P \vee \neg P)$, is just a very general structural law about the mostly non-linguistic world—a basic law of both logic and metaphysics, whatever Tarski would have said.

Unifying excluded middle as one universal generalization, rather than distributing it through all the instances of a schema, also grants it a negation, $\neg\forall P\ (P \vee \neg P)$. By contrast, negating the schema $\alpha \vee \neg\alpha$ just yields the negative schema $\neg(\alpha \vee \neg\alpha)$, which has as instances the negations of *all* instances of excluded middle, whereas anti-classical logicians who reject the law only argue that it is not always right, not that it is always wrong. This is especially significant for intuitionistic logicians, who deny the law of excluded middle but cannot deny any instance of it on pain of contradiction, since the schema $\neg\neg(\alpha \vee \neg\alpha)$ is intuitionistically valid. Their position is in effect to assert $\neg\forall P\ (P \vee \neg P)$, with the intuitionistically compelling argument that we could never have a proof of $\forall P\ (P \vee \neg P)$, since such a proof would provide a decision procedure for all yes/no questions, which is absurd. That is consistent in intuitionistic logic, even though $\exists P\ \neg(P \vee \neg P)$ is inconsistent, for $\neg\forall P\ (P \vee \neg P)$ does not imply $\exists P\ \neg(P \vee \neg P)$. The intuitionistic failure of excluded middle consists in a failure of generality, not a counterexample. As a classical logician, I reject that approach root and branch, but I do not find intuitionists’ denial of excluded middle *unintelligible*. They should be allowed the linguistic resources to express their erroneous view. Of course, some logicians regard the dispute between classical and intuitionistic logicians as merely verbal, on the grounds that the two schools use the logical constants with different senses, but many self-described intuitionists have meant their critique of classical logic more seriously than that, and have genuinely denied the law of excluded middle on the very reading on which classical logicians assert it; those are the intuitionists I have in mind (for more discussion see Williamson 2023).

To return to the main thread of argument, here is an example from first-order logic. Let α be the classical logical truth $\exists x\ (Fx \rightarrow \forall y\ Fy)$, where ‘$F$’ is a non-logical monadic predicate constant. Then α' is the formula $\exists x\ (Xy \rightarrow \forall y\ Xy)$, where ‘$X$’ is a monadic predicate variable, and $\forall(\alpha')$ is $\forall X\ \exists x\ (Xx \rightarrow \forall y\ Xy)$, a fully general statement of the law, in the original object-language or an expansion thereof. We can paraphrase it as ‘For every property, there is something which has it only if everything has it’. In one key respect, that paraphrase too is misleading, for it suggests first-order quantification—that is, quantification into *name* position—over special objects, properties, a singular term for which combines with another singular term and the dyadic predicate 'has' to form a sentence—whereas the quantifier $\forall X$ binds a variable already in *monadic predicate* position. Exactly analogous comments to those already made about the understanding of quantification into sentence position apply to the understanding of quantification into predicate position; $\forall X\, \exists x\, (Xx \rightarrow \forall y\, Xy)$ is another very general structural law about the mostly non-linguistic world—another law of both logic and metaphysics.

These examples also illustrate how the need to make the generality of our results explicit leads us naturally beyond first-order logic to higher-order logic. Wherever the object-language has non-logical constants of a given type, in following Tarski we replace them by variables of the same type, and quantifying those variables in the object-language captures the very same generality he captures in the metalanguage when he defines logical truth and logical consequence. From this perspective, first-order logic looks theoretically unstable, because it cannot capture the generality implicit in its treatment of non-logical predicate constants.

What does this whole approach imply about cases of logical consequence with a *non*-empty set of premises. In classical logic, α is a logical consequence of the set Γ if and only if $(\wedge \Gamma) \rightarrow \alpha$ is logically true, where $\wedge \Gamma$ is the conjunction of all members of Γ (in a specified order), which holds if and only if $\forall((\wedge \Gamma') \rightarrow \alpha')$ is plain true, so in that way the account generalizes. If the premise set Γ is infinite, we must add an infinitary conjunction operator to the language. Since an infinite conjunction may contain infinitely many variables, we must also allow a universal quantifier to bind infinitely many variables (if such devices are not already available in the language). This role for infinitary formulas is unsurprising, since it arises only for arguments that already have infinitely many premises.

Admittedly, not all logics provide a conditional that bridges between logical consequence and logical truth. Moreover, we also care about logical consequence in its own right, irrespective of its connection with logical truth. After all, when we *apply* logic—classical or non-classical—to other fields, our primary concern is arguments whose premises and conclusion are not logical truths. For example, syllogistic logic studies only arguments with multiple premises. In logic's auxiliary role of drawing out deductive consequences of hypotheses and theories, what matters is the relation of logical consequence. But logic also has another role, in codifying very general structural truths about the world; for that purpose what matters are the true universal generalizations corresponding to logical truths. Their primary significance is in what they tell us about the world, not in what they tell us about logical truth or validity; our interest in them is no more primarily metalinguistic than is our interest in the true sentences of the language of physics.

Even when our primary interest is in logical truths, we may take a strong derivative interest in other cases of logical consequence, since we use them to establish logical truths. This is clear in a standard system of natural deduction, where the proof of any logical truth involves making hypothetical assumptions which are later discharged, since that is how the basic introduction and elimination rules are designed. The simplest case is the proof of $\alpha \rightarrow \alpha$, which moves from the trivial sequent $\alpha \vDash \alpha$ with the assumption α to the sequent $\vDash \alpha \rightarrow \alpha$ by a step of →-introduction (conditional proof). Proof-theoretically, logical truth is just a by-product of logical consequence, even though, metaphysically, logical consequence is just a means to the end of logical truth.

Where does this approach to logic as metaphysics leave the distinction between logical and non-logical constants, presupposed by Tarski's account? We could simply follow his later line of demarcation, equating the logicality of a constant with its invariance under all permutations of individuals. That seems as good a criterion as any, with the virtues of clarity, naturalness, and simplicity. Of course, the symbol $\vDash$ for logical consequence is neither a logical nor a non-logical constant of the object-language, because it is not a symbol of that language at all.

However, once we increase the expressive capacity of the language, it becomes much less obvious how to apply Tarski's criterion. For example, we may add the monadic sentence operators $\Box$ and $\Diamond$ to the object-language. If we leave them uninterpreted, we cannot treat them as logical constants, since the truth-value of a formula involving them under an assignment is undefined. Once we interpret them, say as operators for metaphysical necessity and metaphysical possibility respectively, we can treat them as logical constants, although we are not obliged to do so. If we do treat them as logical constants, the necessitist formula $\Box\forall x\Box\exists y\ x = y$ ('Necessarily everything is necessarily something') and its contingentist negation, equivalent to $\Diamond\exists x\Diamond\forall y\ x \neq y$ ('Possibly something is possibly nothing'), are both free of non-logical constants, so whichever of them is true is logically true, on Tarski's account (see Williamson 2013 for more discussion). They make the metaphysical potential of logical truth vivid. But we could also interpret $\Box$ and $\Diamond$ as deontic operators such as 'obligatorily' and 'permissibly', or as temporal operators such as 'always' and 'sometimes'. Once we consider the range of possible interpretations from a metaphysical perspective, the need for a once-and-for-all criterion of logical constanthood evaporates. Which interpretations we hold fixed in evaluating logical consequence and logical truth, and which we abstract away from, just depends on what we want to study. If we want to study the structure of time, we hold the interpretation of temporal operators fixed, thereby treating them as logical constants. If instead we want to abstract away from the structure of time, we allow the interpretation of temporal operators to vary, thereby treating them as non-logical constants. The same goes for deontic and alethic modal operators. Thus we can leave the line between logical and non-logical constants as a parameter in a Tarskian account of logical consequence and logical truth to be filled in according to the needs of a given inquiry.

Of course, some interpreted expressions will *reward* being treated as logical constants more than others do, by leading to a more fruitful inquiry, for example by yielding more informative general laws. Treating the word 'British' as a logical constant would not yield interesting results. But effects on the fruitfulness of inquiry are a messy, pragmatic matter; nothing is gained by trying to build them into the definition of 'logical'.

The assessment of such proposed object-language universal laws as excluded middle and necessitism is properly one branch of logic not only by historical tradition and the intimate connection with logical truth. It is fundamental work that has to be done, by *someone*, and philosophical logicians are in practice the people with by far the most relevant training, skills, and interests to do it.

2. *Abductive methodology and logical strength*

Once we have settled on a specific object-language and set of logical constants in fixing our inquiry, our target theory comprises every universal generalization $\forall(\alpha')$ corresponding to a logical truth α in the object-language. We try to discover which those universal generalizations are. We cannot expect them to be self-evident. But we do not start from a position of total ignorance. We already have some patchy knowledge of our chosen field, else our attempts to theorize it would be hopelessly premature. The aim is to systematize and extend, to generalize and deepen, what we already know, and to correct the mistakes, whatever they are, in what we currently believe. That schematic picture fits logic as well as natural science.

The law of excluded middle is an example. With reference to any point at issue, one can feel the crude pre-theoretic force of 'It is or it isn't'. But that hardly settles the question whether the law, $\forall P\ (P \vee \neg P)$, is true—plain true, let alone logically true. All sorts of counterexample to the universal generalization have been proposed, concerning the open future, vagueness, semantic paradoxes, infinite sequences, and so on. In my view, all the alleged counterexamples to the law are mistaken, but they are not all *trivially* or even *blatantly* mistaken. Reasonable people can doubt any proposed law of logic. However compelling it sounds on first hearing, they may suspect that, although it holds in everyday situations, it fails in less familiar but still actual or at least possible ones—or even that its appearance of holding in everyday situations is the artefact of some glitch in the human cognitive system. If we take such doubts seriously, they call for more systematic, rigorous inquiry, to explore and compare the results of accepting the law with those of rejecting it in various ways.

Even for logicians who accept all of classical propositional logic, assessing quantified claims poses tricky new challenges. Consider this formula $\varphi(R)$, which we can paraphrase as 'R totally orders everything', in other words, 'R is a total, antisymmetric, transitive relation' (where 'R' is a non-logical predicate constant):

$$\forall x\forall y(Rxy \vee Ryx) \wedge \forall x\forall y((Rxy\ \wedge Ryx) \rightarrow x = y) \wedge \forall x\forall y\forall z\ ((Rxy \wedge Ryz) \rightarrow Rxz))$$

If we think of a finite or countably infinite domain, we can find, or at least imagine constructing, such a total ordering of it, by enumerating its members, which verifies $\varphi(R)$ on one interpretation of 'R'. We may therefore be tempted to conclude that its negation $\neg\varphi(R)$ is no logical truth, and correspondingly that its universal generalization $\forall X\ \neg\varphi(X)$ is plain false. But that would be too quick. On their present reading, the first-order quantifiers in $\varphi(R)$ range over absolutely everything. Without some version of an axiom of global choice, one cannot show that there is a total ordering of absolutely everything (see Friedman 1999 for discussion). Thus the attempt to settle the logical status of the first-order formula $\neg\varphi(R)$ on its unrestricted interpretation quickly takes one into abstract, speculative regions.

The speculative, metaphysical character of the inquiry becomes even more marked once we consider modal, deontic, and temporal logic and other extensions of the expressive power of the language of predicate logic. We already saw how the metaphysical dispute between necessitism and contingentism boils down to a comparison between two mutually contradictory sentences, each of which is logically true if true at all. At the level of propositional modal logic, one currently disputed issue is whether the S5 axiom $\lozenge A \rightarrow \Box\lozenge A$ is universally true on the metaphysical reading of the modal operators, that is, whether the universal generalization $\forall P(\lozenge P \rightarrow \Box\lozenge P)$ is plain true on that reading: is every metaphysical possibility metaphysically necessarily metaphysically possible? (For related discussion see Bacon 2018.)

Just as in the natural sciences, *ad hoc* answers to such theoretical questions are unsatisfying. As far as we can, we want to answer them on the basis of general principles. We prefer logical theories that are not only consistent with our evidence—with what we already know—but answer many of our questions, so the more informative they are the better, and we prefer them to do so in ways which are not *ad hoc*, so the less gerrymandered they are the better. To use a standard formulation, we want the optimal combination of *strength* and *simplicity*, where strength corresponds to informativeness, and simplicity to avoidance of gerrymandering.

The criterion of simplicity raises notoriously tricky issues in the philosophy of science, and the same problems are no easier to solve in the philosophy of logic. A robust definition of ‘simpler’ is elusive, but without some such criterion one is hard pressed to explain exactly what is wrong with even the most outrageously *ad hoc* hypotheses. By contrast, the criterion of strength looks more tractable, especially as applied to precisely characterized logical theories, since logicians already describe some logics as stronger than others in a precise sense. However, the relevance of strength to logical theorizing has been called into questions. Disentangling the issues reveals some of the methodological harm done by equating logical theories with metalogical theories of validity. That will be the main task for the rest of this chapter.

Logics are routinely compared for ‘strength’. Textbooks and articles often include diagrams to map the logics in play by comparative strength. This is straightforward in areas such as modal logic, where a logic is usually identified with the set of its theorems—its logical truths. Then a logic L_1 is said to be *at least as strong as* a logic L_2 if and only every theorem of L_2 is a theorem of L_1 ($L_2 \subseteq L_1$); ‘at least as strong as’ expresses a reflexive partial ordering of logics. L_1 is said to be *stronger than* L_2 if and only if L_1 is at least as strong as L_2 and L_2 is not at least as strong as L_1 ($L_2 \subset L_1$); ‘stronger than’ expresses a strict partial ordering of logics.

So defined, logical strength is intended as a way of ordering logics in the *same* object-language. To compare logical theories in different object-languages, without prejudging any semantic relations between them, one must fall back on a standard of *relative interpretability* in some sense. One might also use such a standard to compare logical theories in the same language if one wanted not to prejudge semantic relations between syntactically the same symbol as used in one theory and as used in another. Interpretability strength is a much coarser standard than logical strength, since the interpretations (despite often being called ‘translations’) are not required to preserve meaning. In the present setting, logical strength is more appropriate, because we are treating the object-language as already interpreted—we theorize about the world in it.

If a logic is equated with the extension of 'logical consequence' rather than that of 'logical truth', the subset relations defining 'at least as strong as' and 'stronger than' can be read accordingly. This makes no difference to the ordering for logics with the bridging equivalence between $\Gamma \vDash \alpha$ and $\vDash (\wedge \Gamma) \rightarrow \alpha$ (for compact logics, a finite subset of Γ will do), since the extension of logical truth already encodes the extension of logical consequence, so ordering in terms of logical consequence has the same effect as ordering in terms of logical truth. But for logics with no such bridging equivalence, ordering in terms of logical consequence can make a difference. For instance, some non-classical logics have no theorems at all, but still differ from each other in their consequence relation: one may be stronger than another in logical consequences even though they coincide in logical truths.

Logical strength does not capture all aspects of informativeness. For example, recall the comparison between standard classical and intuitionistic propositional logic (considered as rival theories in the same language). Every intuitionistic theorem is a classical theorem but not *vice versa*, so classical propositional logic is logically stronger than intuitionistic propositional logic. The result is the same whether we consider logical truth or logical consequence, since both logics have the bridging equivalence. Now expand the object-language by adding quantification into sentence position, governed by appropriate principles. Then the classical logic will have the theorem $\forall P\ (P \vee \neg P)$, while the intuitionistic logic will have the theorem $\neg \forall P\ (P \vee \neg P)$, as explained in section 1. Since both logics are consistent, each has a theorem the other lacks, so neither is logically at least as strong as the other; they are incommensurable in logical strength. Nevertheless, adding quantification into sentence position does very little to diminish the difference in strength between the two logics, informally understood. The law of excluded middle as a universal generalization is far more informative deductively than its negation. As an instrument for mathematical reasoning, for instance, classical logic remains much more powerful than intuitionistic logic.

Analogous examples arise in natural science. Let SR be Einstein's theory of special relativity, formulated as a big universal generalization. An anti-Einsteinian bluntly asserts ¬SR ('Einstein is wrong'). Pretheoretically, SR is much more informative than ¬SR. SR gives a very full, exact account of motion, applicable to all times and places. Its negation merely says that SR goes wrong *somewhere*, without specifying where, or when, or how. SR, is a paradigm of a theory of physics; ¬SR hardly deserves the title of a 'theory' at all. Yet SR is logically no stronger than ¬SR, indeed it does not even count as logically at least as strong as ¬SR, for ¬SR is a logical consequence of itself but not of SR. One can summarize the comparison by saying that SR is *scientifically stronger* (more informative) than ¬SR, but not *logically stronger*. In general, if T_1 is logically stronger than T_2, then T_1 is also scientifically stronger than T_2 (since T_1 already says everything that T_2 says), but the converse fails.

Other things being equal, we prefer scientifically stronger theories to scientifically weaker ones—though buying a little more strength at the cost of a lot more complexity may be a bad bargain. The reason for preferring scientifically stronger theories is of course *not* that they are more probable. On the contrary, increasing scientific strength tends to decrease probability. In particular, by the standard Kolmogorov axioms of probability, if T_1 is logically (and so scientifically) at least as strong as T_2, then T_1 is no more probable than T_2. A better reason for preferring scientifically stronger theories is that they answer more of our questions, which is something we want our scientific theories *for*. Moreover, greater scientific strength carries additional epistemic benefits: it tends to increase explanatory and predictive power, and thereby the scope for abductive confirmation or inference to the best explanation. But having gained such confirmation for T_1, we could always increase probability again by watering T_1 down to some logically and scientifically weaker disjunction $T_1 \vee T_2$. We avoid such retreats because they deprive us of answers to some of our questions. Theory comparison is typically between rival theories of similar scientific strength.

These considerations apply to comparisons between rival logical theories too. In their own right, scientifically stronger logics answer more of our questions in logic. In their auxiliary role as background logics for theories in other fields, they increase the scientific strength of those non-logical theories, and enhance their explanatory and predictive power, by extracting more relevant consequences from them. But stronger logics can also play an explanatory role in logic itself, obviously not by providing causal explanations, but by subsuming isolated logical observations under illuminating generalizations—for example, in the twentieth-century streamlining of modal logic.

One special case deserves mention. In classical logic (and many others), an inconsistent theory has all sentences of the language as theorems, and so is logically stronger—and therefore scientifically stronger too—than any consistent theory. But an inconsistent theory is not *better* than a consistent theory, not even by giving (mutually inconsistent) answers to all our questions. One can have too much of a good thing; an inconsistent theory is *too* strong. But what is so bad about an inconsistent theory is not its maximal scientific strength as such. Rather, what is so bad about it is its inconsistency with all our evidence, an inevitable consequence of its maximal strength.

We can test this rough sketch of the place of logical and scientific strength in an abductive methodology for theory choice in logic against a recent challenge by Gillian Russell (Russell 2019). She is inclined to agree with me that 'scientific strength is a virtue' but argues that 'logical strength does not entail scientific strength' (2019: 557), and more generally that logical strength is not a virtue in a logical theory. As one of the main sources for the view she is attacking, she repeatedly cites an article of mine (Williamson 2017). Identifying her misunderstandings may help to clarify my abductive account. That such a perceptive and open-minded philosopher should get the view so wrong may indicate how far it runs contrary to well-entrenched expectations.

Russell opens her article by noting 'renewed interest in a broadly abductive approach to the epistemology of logic' (Russell 2019: 548, citing Priest 2006, Russell 2014, Hjortland 2017, Beall 2017, and Williamson 2017). She writes (ibid.):

> The details vary but the central idea is that rival logics are different theories of the relation of logical consequence, and the best theory is the one which is adequate to the data and possesses the most theoretical virtues—perhaps simplicity, strength, elegance, unity, symmetry, or ontological parsimony—and least theoretical vices—such as ad hockery, inelegance, or ontological profligacy.

But central to my idea (and explicit in Williamson 2017) is that rival logics are *not* 'different theories of the relation of logical consequence', for that would make them theories of a metalogical subject matter, which is exactly what I deny. The symbol for logical consequence, $\models$, belongs to the metalanguage, not to the object-language, whereas the logical theories I have been discussing are formulated in the object-language itself; their theorems are its logical truths, in which the symbol $\vDash$ does not occur. They are theories not of the somewhat esoteric relation of logical consequence but of the referents of the logical constants—negation, conjunction, disjunction, identity, universality and other quantifiers, various modalities, and so on, many of which are involved in virtually all serious theorizing (a more accurate but less concise statement would respect type distinctions). When universally generalized, the theorems state very abstract general patterns in the mostly non-linguistic world, not relations between sentences.

When logicians apply the terminology of ‘stronger’ and ‘weaker’ to logics conceived as sets of theorems, they understand those theorems in the usual way, as formulas of the object-language, not as formulas of the metalanguage with $\vDash$. As a slight generalization, when they conceive a logic as the extension of a consequence relation, they in effect treat it as the set of ordered pairs $<\Gamma, \alpha>$ such that the relation holds between the set Γ of object-language formulas and the object-language formula α. In both cases, a weaker logic is simply a proper subset of a stronger one, in the usual way. By contrast, when Russell defines ‘logical strength’ for logics, she gets the intended extension only by the Procrustean device of ‘thinking of logics as sets of sentences of the form $\Gamma \vDash A$’ (2019: 554). That hardly fits her official view of logics as ‘theories of the relation of logical consequence’, for such theories should be allowed sometimes to say that a given formula is *not* a logical consequence of a given set of formulas, and to state the introduction and elimination rules for a system of natural deduction, which involve conditional statements with several different constituent sentences of the form $\Gamma \vDash \alpha$. Russell’s need for such an artificial restriction indicates that her approach does not really fit the way logicians think of stronger and weaker logics.

Russell’s restriction of logics to sentences of the form $\Gamma \vDash A$ is also inconsistent with what she says about them elsewhere in the paper, for instance (2019: 557):

> Modern logic is mathematical, and logics are formulated so that they are determinate, in the sense that for any set of premises, Γ, and conclusion, A, in the language on which the logic is defined, they say whether or not $\Gamma \vDash A$.

For if the logic were a theory consisting only of metalinguistic sentences of the form $\Gamma \vDash A$, it would be consistent with the hypothesis that the relation $\vDash$ is universal, holding between any set of object-language formulas and any object-language formula, in which case the logic would *not* say whether or not $\Gamma \vDash A$ in cases where $\vDash$ did not in fact hold (if $\vDash$ is not universal, in other words, there is no inconsistency at the level of the object-language).

Incidentally, if a metalogic for standard first-order logic were ‘determinate’ in Russell’s sense, it would not even be recursively enumerable, and so could not be formally axiomatized. For otherwise it would provide a decision procedure for first-order logic. To decide whether a formula α of the first-order object-language was a theorem of first-order logic, we could simply let the metalogic start enumerating its (metalogical) theorems, and wait; since it was determinate, it would sooner or later output either $\vDash \alpha$ or $\nvDash \alpha$. But first-order logic is undecidable; there is no such decision procedure.

The special case of inconsistency brings out another difficulty for Russell’s treatment of logics as metalogical. A logic is inconsistent if and only if it has every formula of the object-language as a theorem. In the metalanguage, we can easily give a complete and consistent description of the inconsistent logic. But if we identify the logic with its complete metalinguistic description in the way Russell does, in calling the logic inconsistent we are calling its complete metalinguistic description inconsistent, which it is not! One might wriggle out of the problem by multiplying senses of 'inconsistent', but the need for such artificial manoeuvres is another indication that something has gone wrong.

The difference between logics and their metalogical descriptions is vivid for the weakest logics as well as the strongest ones. Consider the minimal logic Min, treated as a consequence relation, in which any set of formulas entails only its own members: for all Γ and α, $\Gamma \vDash \alpha$ if and only if $\alpha \in \Gamma$. Min obeys all the standard structural rules for a consequence relation: reflexivity ($\{\alpha\} \vDash \alpha$), monotonicity (if $\Gamma \vDash \alpha$ and $\Gamma \subseteq \Delta$ then $\Delta \vDash \alpha$), cut (if $\Gamma \vDash \alpha$ and $\Delta \cup \{\alpha\} \vDash \beta$ then $\Gamma \cup \Delta \vDash \alpha$), and closure under uniform substitution (if $\Gamma \vDash \alpha$ then $s\Gamma \vDash s\alpha$, for any uniform substitution s). Any consequence relation that obeys all those rules includes Min (by contrast, the empty relation violates reflexivity). As a logic, Min is pitifully weak. It has no theorems (since no formula belongs to the empty set), and it prevents one from ever deriving a new conclusion from premises. But this metalogical description of Min is just as informative as an (accurate) description of any other logic, since it enables one to work out exactly what does or doesn't follow from what in Min. Min is scientifically weak as well as logically weak: since it has no theorems, it says nothing in its own right about the structure of the world, and as an auxiliary logic for other sciences it would be a disaster, since it would not enable one to draw any further consequences from given hypotheses or theories in those sciences. It is exactly this scientific weakness which disqualifies Min as an auxiliary logic for science. Many other logics, not as weak as Min but still much weaker than classical logic, are similarly far too weak scientifically to serve as the background logic for science.

Russell claims that a whole range of propositional logics—including the empty logic, various many-valued logics, classical logic, and the inconsistent logic—'are all equal in scientific strength' (2019: 557). Her basis for this claim is in effect that their metalogical descriptions are all equally informative, because she conflates the logics with their metalogical descriptions. That basis is clearly quite irrelevant to the abductive assessment of those logics either as abstract, structural theories of the world or as background logics for science. Nor does it support her rejection of my elementary point that greater logical strength entails greater scientific strength, because she is using the terms in senses quite alien to mine.

To her credit, Russell is aware that simply to treat all logics as equal in scientific strength would be to miss something. She therefore labours to explain a new sense in which logics are *not* all equal in scientific strength. However, she denies that this new sense will vindicate the role of logical strength in the abductive comparison of logics, for on it 'although logically weaker logics are often less scientifically strong than logically stronger logics, *any* of the logics in our set can be strengthened—without damage to the logic—to make them as scientifically strong as our logically strongest logic' (2019: 557, her italics).

Russell's idea is that when a logic classifies an argument form as invalid, it does not thereby tell us *which* instances of the argument fail; by contrast, when a logic classifies an argument form as valid, it thereby tells us that *no* instances of the argument fail. Since weaker logics classify more argument forms as invalid, they tend to be less informative. But the weaker logics can be supplemented with the missing information as to which instances fail, so equality in scientific strength is restored.

The details of Russell's account are confusing, because consequence relations as usually understood involve not argument *forms* but *arguments* from a specific set of object-language sentences to a specific object-language sentence. It is correspondingly unclear what she means by 'instances'. They might be interpreted substitution instances of the original argument in a given language, but an argument can be invalid even though all its substitution instances are truth-preserving, through expressive limitations of the language (Tarski 1936). To return to an earlier example, there may be complete orderings of everything, even though no predicate of any actual language expresses any of them; in that case, the formula $\neg\varphi(R)$ is not a logical truth, even though every interpreted substitution instance of $\neg\varphi(R)$ is true.

We need not press these subtle difficulties, for Russell's account faces more blatant problems. She explains that, in Tarski's 1936 account of logical consequence, an argument form fails in a given instance with true premises unless the conclusion is also true (2019: 561n30). Her account is supposed to apply both to that framework and to a broad range of others. In any case, Tarski's 1936 framework is the one relevant to her critique of my abductive account.

Consider the fallacious argument form of *affirming the consequent* for the simple case of the material conditional (→). To accept affirming the consequent as valid is to accept the metalogical schema or argument form $\{\alpha \rightarrow \beta, \beta\} \vDash \alpha$. In a given instance, affirming the consequent fails if and only if $\alpha \rightarrow \beta$ and β are both true and α is false. Since the truth of β implies the truth of $\alpha \rightarrow \beta$, that means that affirming the consequent fails in any instance where α is false and β true. The same applies to the even more crassly invalid argument form $\beta \vDash \alpha$. Thus to identify all the instances in a given language where invalid argument forms fail, one must be able to identify all the true sentences of the language as true, and all the false sentences as false: in effect, one must be more or less omniscient. The result generalizes to many non-classical logics and many non-standard semantic frameworks, with minor complications, but for present purposes Tarski's 1936 framework is the most relevant. Thus the strengthenings of logics Russell invokes to restore equality in scientific strength are available only to a quasi-omniscient being.

Compare the difference in scientific strength between Einstein's theory of special relativity, SR, and its mere negation ¬SR. If opponents of SR propose ¬SR as their alternative theory, they face the complaint that ¬SR is too weak scientifically to be a serious competitor theory to SR. Following Russell's example, they might respond that once ¬SR is supplemented by a full account of every instance in which SR holds and every instance in which it fails, their theory will be just as scientifically strong as SR. But that response is hopeless. For scientific purposes, the relevant abductive comparison is *not* between notional strengthenings of SR and ¬SR, of which their proponents can only dream, but between the theories currently on the table in the envisaged case, SR and ¬SR. What counts is the pitiful scientific weakness of ¬SR, not the putative scientific strength of its imagined extension. Exactly the same applies in the methodology of logic. For scientific purposes, the relevant abductive comparison is *not* between notional strengthenings of (say) classical logic and a sub-classical alternative logic, of which their proponents can only dream, but between the logics currently on the table, classical logic and the alternative logic. What counts is the scientific weakness of the alternative logic itself, not the putative scientific strength of its imagined extension. Thus Russell's attempt to restore equality in scientific strength between logics is irrelevant to the application of the abductive methodology.

A further methodological remark is worth making. The kind of case-by-case supplementation of a logic with classifications of individual instances as failing or not failing various invalid rules would involve a massive increase in the complexity of the resultant theory. It would be like supplementing a physical theory with a vast array of miscellaneous observations consistent with it but not explained by it. Trading off simplicity for strength in that way would normally be considered a bad bargain.

In the case of mathematics, proponents of weak non-classical logics often propose to recover the full strength of classical mathematics by supplementing their logic with versions of classical schemas such as excluded middle restricted to instances in the language of mathematics (Hjortland 2017). That strategy runs into trouble with *applications* of classical mathematics in the natural and social sciences, because accepting the relevant instances of the classical schemas often conflicts with the proposed rationale for going non-classical in the first place (Williamson 2018).

The strange contortions of Russell's account all result from her initial conflation of logic with metalogic, of logical theories with their metalogical descriptions. It is an object lesson in what goes wrong once one assumes that logic is about validity. In particular, it shows how that confusion can distort the application of an abductive methodology to logic, by neutralizing one of the key criteria for abductive theory comparison, scientific strength.

That moral also bears on Russell's comment on how different authors, all professing to apply a broadly abductive methodology to logic, have in practice arrived at such divergent conclusions (2019: 548-9, following Hjortland 2017). The divergence is unsurprising; after all, there are many deep disagreements in physics and other natural sciences where all parties employ a broadly abductive methodology. But a more specific point is also relevant. The conception of logic as the study of validity, and resultant conflation of logic with metalogic, are widespread amongst philosophers of logic, including some of those who profess to apply a broadly abductive methodology to logic: for example, Russell herself, Graham Priest (2006, 2016), and Ole Thomassen Hjortland (2017). Given how badly that conflation can distort the application of an abductive methodology to logic, it may well have contributed to the divergence in results. In particular, the way it confuses the issue of scientific strength may help explain some professed abductivists' preference for non-classical logics whose striking weakness might have been expected to disqualify them by normal abductive standards.

Acknowledgment

Thanks to participants in an Oxford graduate class and Daniel Kodsi for comments on earlier versions of this material.

References

Bacon, Andrew. 2018: 'The broadest necessity', *Journal of Philosophical Logic*, 47: 733 -783.

Beall, J.C. 2017: 'There is no logical negation: True, false, both, and neither', *The Australasian Journal of Logic*, 14, 1.

Friedman, Harvey. 1999: 'A complete theory of everything: satisfiability in the universal domain'. https://cpb-us-w2.wpmucdn.com/u.osu.edu/dist/1/1952/files/2014/01/ACompThyEver101099-1mkg42b.pdf

Hjortland, Ole Thomassen. 2017: 'Anti-exceptionalism about logic', *Philosophical Studies,* 174: 631-658.

Hodges, Wilfrid. 1977: *Logic*. London: Penguin.

Mancosu, Paolo. 2006: 'Tarski on models and logical consequence', in Jeremy Gray and José Ferreiros (eds.), *The Architecture of Modern Mathematics*, 209-237. Oxford: Oxford University Press.

Priest, Graham. 2006: *Doubt Truth to be a Liar*. Oxford: Oxford University Press.

Priest, Graham. 2016: 'Logical disputes and the *a priori*', *Logique et Analyse*, 236: 347-366.

Rayo, Agustín, and Uzquiano, Gabriel (eds.). 2006: *Absolute Generality*. Oxford: Clarendon Press.

Russell, Gillian. 2014: 'Metaphysical analyticity and the epistemology of logic', *Philosophical Studies*, 171: 161-175.

Russell, Gillian. 2019: 'Deviance and vice: strength as a theoretical virtue in the epistemology of logic', *Philosophy and Phenomenological Research*, 99: 548-563.

Tarski, Alfred. 1936: 'O pojciu wynikania logicznego', *Przegląd Filozoficzny*, 39: 58-68. Page references are to English translation by J.H. Woodger, 'On the concept of logical consequence', in Alfred Tarski (ed. John Corcoran), *Logic, Semantics, Metamathematics*, 409-420. Indianapolis: Hackett, 1983.

Tarski, Alfred. 1986: 'What are logical notions?' (ed. John Corcoran), *History and Philosophy of Logic*, 7: 143-154. Based on a 1966 lecture.

Wiggins, David. 1980: *Sameness and Substance*. Oxford: Blackwell.

Williamson, Timothy. 2003: 'Everything', *Philosophical Perspectives*, 17, 415-465.

Williamson, Timothy. 2013: *Modal Logic as Metaphysics*. Oxford: Oxford University Press.

Williamson, Timothy. 2017: 'Semantic paradoxes and abductive methodology', in Brad Armour-Garb (ed.), *Reflections on the Liar*, 325-346. Oxford: Oxford University Press.

Williamson, Timothy. 2018: 'Alternative logics and applied mathematics', *Philosophical Issues*, 28: 399-424.

Williamson, Timothy. 2023: 'Accepting a logic, accepting a theory', in Romina Padró and Yale Weiss (eds.), *Saul Kripke on Modal Logic*. New York: Springer.