Semantic Paradoxes and Abductive Methodology

(draft of paper to appear in Brad Armour-Garb, ed., *The Relevance of the Liar*, Oxford: Oxford University Press)

Timothy Williamson

Understandably absorbed in technical details, discussion of the semantic paradoxes risks losing sight of broad methodological principles. This essay sketches a general approach to the comparison of rival logics, and applies it to argue that revision of classical propositional logic has much higher costs than its proponents typically recognize.

1. *Logical truths and universal generalizations*

As a first step, we rehearse a version of Tarski's account of logical consequence in his famous early paper (Tarski 1936).

For present purposes, we are *not* trying to analyse a pre-theoretically given concept of logical consequence. Although the folk reason, and sometimes even reflect on differences between specific instances of good and bad reasoning, they have no need to distinguish between *logically* valid reasoning and reasoning valid in some much broader sense. Nor should logic as a developing theoretical discipline tailor its basic theoretical terms to fit whatever pre-theoretic prejudices and stereotypes may happen to be associated with the word 'logic', any more than physics should tailor its basic theoretical terms to fit whatever pre-theoretic prejudices and stereotypes may happen to be associated with the word 'physics'. Like every other form of systematic inquiry, logic has the right to identify and employ whatever fundamental distinctions help it formulate the most fruitful questions and their answers. Such distinctions cut through pointless accretions and complications in folk logic.

We start from an interpreted object-language L. It could be a natural language, although in practice we shall find it more convenient to deal with an interpreted formal language, such as a mathematical notation, since we have a clearer overview of the totality of its sentences and their syntax. We do not assume that the speakers of L have agreed an explicit semantic theory of L, any more than the speakers of a natural language have agreed an explicit semantic theory of it. Rather, as with a natural language, the speakers of L share ordinary linguistic competence in L, using it as a public language, relying on the publicly available, informally explained meanings of its sentences. They acquire such competence in the first place by the direct method of immersion in the language, supplemented with occasional *ad hoc* explanations of specific points, usually given by already competent speakers. As always, this shared competence permits, and indeed enables, deep disagreements between speakers to be expressed in the common language.

We are going to define a relation of *logical consequence* between sets of sentences of L and sentences of L, but first we need to sketch in some more background.

For simplicity, we assume that all sentences of L are declarative. If L has context-sensitive expressions, we provisionally fix a preferred context — ideally, one that leaves quantifiers unrestricted. Taking non-declaratives and indexicality into account would complicate the arguments below without seriously impeding them. For concreteness, we further assume that L has at least the expressive power of a first-order language with identity: L has negation, conjunction, disjunction, universal and existential quantifiers, an identity predicate, and an appropriate variety of other predicates. We assume that open or closed formulas of L can be evaluated as true or otherwise on an assignment of values to variables; after all, L is an interpreted language, not a mere formalism.

The *constants* of L are the atomic non-variable expressions of L, of whatever type. From amongst them we select a set of *logical constants*. In Tarski's original paper, he left it open where and how the distinction between logical and non-logical constants is to be drawn, and indeed whether it can be drawn non-arbitrarily. In a later paper, he proposed that the logical constants are those invariant under permutations of individuals (Tarski 1986). If a once-and-for-all criterion of logicality is wanted, something along those lines may be the best we can do. But for present purposes a once-and-for-all criterion is not wanted. Rather, the choice of logical constants is pragmatic. In effect, one is investigating which general structural principles the logical constants satisfy. Varying the extension of 'logical constant' amounts to varying what one is investigating the general structural features of. For example, in modal logic, if one is interested in the specific structural features of metaphysical necessity and metaphysical possibility, one may make the operators □ and ◊ logical constants (typically, in addition to the standard ones) with the respective intended interpretations (as in Williamson 2013). Alternatively, one may be interested in the structural features general to all operators, and make □ and ◊ non-logical. Both choices are legitimate. Indeed, both investigations need to be carried out sooner or later. Of course, some choices of logical constant will send the investigation outside the pragmatically appropriate domain of logic. For instance, if one counts 'mass' and 'energy' as 'logical' constants, one will be doing physics rather than logic in any distinctive sense. Experience and good judgment are needed to determine which choices determine promising inquiries.

In what follows, we assume that a set of logical constants has been given, and that it includes at least the usual suspects: negation, conjunction, disjunction, the universal and existential quantifiers, and the identity predicate. Others will be added as required.

For each non-logical constant *e* of L we extend the language with a new variable v_e of the same semantic type as *e*, where $v_e = v_f$ only if $e = f$. For each sentence α of L, let α^v be the result of substituting v_e for *e* throughout α for each non-logical constant *e* of L. For each set Γ of sentences of L let $\Gamma^v = \{\alpha^v: \alpha \in \Gamma\}$. Some of the new variables will be of higher type, for instance those replacing atomic predicates and sentences, and will need to be assigned values appropriate to their type. If the context restricted the values of the variables to a contextually relevant domain, the assignments would need to meet those restrictions, but for present purposes we can work on the simplifying assumption that the quantifiers are unrestricted. Although in this paper we will continue speaking in terms of 'values' and 'assignments', a more rigorous treatment would be articulated in part by quantifiers of suitably higher types in the meta-language, in order to avoid Russellian paradoxes with unrestricted quantification.[1] These complications do not affect the broad methodological issues which concern this paper.

We can now define logical consequence. A sentence α of L is a *logical consequence* of a set Γ of sentences of L ($\Gamma \models \alpha$) if and only if α^v is true on every assignment on which every member of Γ^v is true. As a special case, α is *logically true* ($\models \alpha$) if and only if α^v is a logical consequence of the empty set, in other words, α^v is true on every assignment. We can think of α^v as representing the logical form of α, and the pair $<\Gamma^v,\alpha^v>$ as representing the logical form of the argument from Γ to α, subject to the stipulation that logical form is invariant under permutations of the variables of a given type.

Logical consequence in the sense of $\models$ obeys the standard structural rules for a consequence relation. That is, the following hold for all sentences α and β of L and all sets Γ and Δ of sentences of L:

Assumption $\{\alpha\} \models \alpha$

Monotonicity (Thinning) If $\Gamma \models \alpha$ then $\Gamma \cup \Delta \models \alpha$

Cut If $\Gamma \models \alpha$ and $\Delta \cup \{\alpha\} \models \beta$ then $\Gamma \cup \Delta \models \beta$

Logical consequence also obeys a rule of closure under uniform substitution. To be more precise, let a *uniform substitution* be a function s that maps each expression *e* of L to an expression s*e* of L of the same type as *e*, maps every logical constant to itself, and commutes with all grammatical constructions.[2] For example, since ¬ is a logical constant, $s(\neg\alpha) = s(\neg)s(\alpha) = \neg s(\alpha)$. A uniform substitution may substitute variables for non-logical constants and constants for logical or non-logical constants, within the same type; thus in L^+ a uniform substitution may map α to α^v or vice versa. As usual, $s\Gamma = \{s\alpha: \alpha \in \Gamma\}$. Then:

Uniform Substitution If $\Gamma \models \alpha$ then $s\Gamma \models s\alpha$

Someone might ask what logical consequence in the sense of $\models$ has to do with *logic*. Notoriously, Tarski's account does not impose any modal or epistemic constraints, yet many philosophers assume that principles of logic should be necessary or *a priori* or analytic. They often present themselves as speaking on behalf of some intuitive pre-theoretic conception of logic. But it has already been emphasized that we are not trying to be faithful to any such conception. Rather, we are trying to *construct* a definition that will best serve the purposes of theoretical inquiry in this general area. We want to *define* logical consequence rather than treating it as primitive, because any pre-theoretic conception associated with the phrase 'logical consequence' will be too inchoate, too primitive in the adverse sense, to be adequately constrained. Tarski's austerely clean and clear definition is perfectly suited to a discipline as fundamental as logic. In abstracting from the specific meanings of the non-logical constants, it enables us to recognize the patterns formed by the logical constants, which pick out the field of our interest. Once we have that dimension of generality, adding a second dimension of necessity, a priority, or analyticity needlessly complicates the picture, mixing together questions that our fundamental terminology should hold carefully apart so that it can represent their interrelations perspicuously.

The wisdom of not adding such accretions to logical consequence is confirmed when we come to the semantic paradoxes. Suppose that we are confronted with a Liar-like derivation of an absurd conclusion. We want a diagnosis which tells us *where* the derivation goes wrong. Roughly speaking, we want to know *which* step goes from true to false, or at least to untrue (this is only rough, in part because 'true' is itself one of the contested terms in the semantic paradoxes). If we are told that the conclusion of some step is not a logical consequence of its immediate premises, because the connection is contingent, or a posteriori, or synthetic, but the step turns out still to be materially truth-preserving, then we have not yet been given the needed diagnosis, because we have not been told where to stop going along with derivation. Of course, we want to *know* whether the step is materially good, but that is just an instance of the banal general need to know the results of a science in order to apply them; it does not stem from any special epistemic condition on logical consequence. When we are trying to solve the semantic paradoxes, any special epistemic or metaphysical constraint inserted into the definition of logical consequence would be a pointless distraction. In answer to the corresponding worry about the generality condition itself in Tarski's account, that generality is the minimal requirement for a theoretically useful relation in the vicinity.

One manifestation of the austerity of Tarski's account, already noted in his 1936 paper, is that if a closed sentence of L contains no non-logical constants, then it is logically true if and only if it is (simply) true: the quantification over assignments is redundant here because α^v contains no free variables. For instance, since $\exists x \exists y \neg x=y$ contains no non-logical constants, it is logically true, because it is true: there are indeed at least two things. Although logical consequence is a linguistic relation, because it holds only between sets of sentences and sentences, and logical truth is a linguistic property, because it holds only of sentences, they are nevertheless closely connected to how things are in the mostly non-linguistic world.

The equivalence between ascriptions of logical truth and non-metalinguistic sentences can be generalized in an extension of L. Let L^+ be the result of adding to L both the new variables v_e and universal quantifiers for all the corresponding types. A universal quantifier for a given type is interpreted as ranging unrestrictedly over all members of that type. For any sentence α of L, let UG(α) be the result of prefixing α^v with a sequence of universal quantifiers for the relevant types on all its free variables (in some fixed order). For instance, if α is *a*=*a*, then α^v is $v_a = v_a$, and UG(α) is $\forall v_a\ v_a = v_a$. If if α is $\neg\forall x\ Fx$, where *F* is a non-logical predicate constant, then α^v is $\neg\forall x\ V_F x$, and UG(α) is $\forall V_F \neg\forall x\ V_F x$. These extra universal quantifiers simply articulate in the extended object-language the effect of the universal quantification over assignments in the definition of logical truth. Thus any sentence α of L is logically true ($\models \alpha$) if and only if UG(α) is (simply) true. But UG(α) is simply a non-metalinguistic generalization (unless α itself contains a metalinguistic logical constant, such as a truth predicate): there is no more reason to regard higher-order quantification as metalinguistic than there is to regard first-order quantification as metalinguistic.

The question whether UG(α) is true still has a metalinguistic aspect, at least superficially. More generally, one is asking which *sentences* of L^+ of a given universally generalized form are true. However, that is an artefact of presentational convenience. After all, one could present physics as asking which generalizations in an appropriate language for physics are true, but that would not make physics a metalinguistic inquiry in any deep sense. Rather, we may simply be using the truth predicate here in its familiar role as a convenient device for generalization ('Everything the policeman said is true'). Our underlying interest is typically not in the sentences themselves.

We can go further if we provisionally assume that a disquotational principle for truth applies to sentences of L^+ (the assumption will be reconsidered shortly). Thus an ascription of truth to UG(α) is equivalent to UG(α) itself. But, as we have just seen, an ascription of logical truth to α is equivalent to an ascription of truth to UG(α). Hence an ascription of logical truth to α is equivalent to UG(α) itself. Investigating which sentences of L are logically true is tantamount to trying to decide universal generalizations of L^+ not containing non-logical constants. Such an investigation is not semantic or epistemological in any distinctive sense. It is more like an investigation in mathematics or physics, an attempt to determine which relevant principles hold. Its subject matter is of obvious scientific interest, in a sense of 'science' that includes mathematics as well as the natural and social sciences.

One might feel that a universal generalization like $\forall x\, x=x$ is too trivial to be of scientific interest, but of course $a=a$ was chosen for heuristic reasons as an example of a logical truth for its extreme simplicity: in general, logical truths can have all the complexity and difficulty of mathematical theorems. Indeed, not even $\forall x\, x=x$ is wholly uncontroversial. Some non-analytic philosophers deny that anything is really self-identical, probably because (like many analytic philosophers) they misapply Leibniz's Law. However confused their reasons, the reflexivity of identity is what they reprove their analytic colleagues for naively accepting. I have argued at length elsewhere that disputes even over elementary logical principles can be non-verbal, involving genuine disagreement of a non-metalinguistic sort (Williamson 2007, 2013). I will not repeat those arguments here.

Of course, the semantic paradoxes throw doubt on the assumed disquotational principle about truth, at least at the margins, since restricting that principle is one of the main strategies for resolving the paradoxes. Nevertheless, even if we cannot assume that an ascription of truth to UG(α) is always exactly equivalent to UG(α) itself, we might find the latter more interesting or fundamental than the former. After all, such non-metalinguistic generalizations constitute in effect a large part of mathematics. Moreover, UG(α) itself should precede ascriptions of truth to UG(α) in the order of explanation. Just as we should ask whether grass is green before asking whether the sentence 'Grass is green' is true, once we separate the questions, so we should ask whether everything is self-identical before asking whether the sentence 'Everything is self-identical' is true, once we separate the questions. Allowing that the questions may have different answers does not reverse that order of priorities. In such cases, the metalinguistic question of the logical truth of α may be just a convenient but approximate device for raising the non-metalinguistic question of UG(α) itself. To continue the analogy with physics: even if we are forced to restrict the disquotational principle, that does not make investigating which physical principles are true a fundamentally metalinguistic inquiry. We are still talking *with* words such as ‘mass’ and ‘energy’, not *about* them. The same goes for the logical constants in the sort of logical inquiry just sketched.

In asking the non-metalinguistic question ‘UG(α)?’, we are outside the realm of *metalogic*, but that does not mean that we are outside the realm of *logic*. Indeed, we may be able to answer the question by deducing UG(α) in the object-language simply by using standard rules of logic. However, we cannot always rely on such rules. In some cases, there are no standard rules: we are breaking fresh ground. In other cases, there are standard rules, but paradoxes — such as the semantic paradoxes — call them into question. We are then forced to adopt a more speculative mode of inquiry, to determine which rules *should* be standard. In non-normative terms, we ask which universal generalizations really hold. Such an inquiry resembles an early stage of mathematics when there is still widespread disagreement over which first principles we may rely on in proving mathematical results. The investigation has a more philosophical flavour. Even here, for suitable choices of the logical constants, no discipline is better fitted than logic to evaluate the universal generalizations at issue, given their extreme generality.

Of course, the term ‘logic’ covers a wide variety of legitimate inquiries. The diverse branches of model theory, proof theory, set theory, and recursion theory all count as logic, and these days there are more logicians in departments of mathematics or computer science than in departments of philosophy, as a glance at a logic journal will indicate. Those more technical branches of logic are pursued by purely mathematical methods of an established sort. Moreover, insofar as they investigate logical consequence at all, they do so from a metalinguistic standpoint: their results about the object-language hold independently of its intended interpretation. By contrast, although the sort of non-metalinguistic inquiry just sketched is logic, it is not primarily metalogic (see also Williamson 2014). While it in no way supplants those more technical parts of logic, it has its own fundamental significance. It is, in one good sense of the term, *philosophical logic*.

2. *Logical truths and logical consequence*

The focus so far has been on logical truth, the special case of logical consequence where the set of premises is empty. One might fear that it is a very misleading special case: when it comes to practical applications, what typically matters is deducing conclusions that are not logical truths from premises that are not logical truths. Indeed, from a proof-theoretic perspective in particular, logical truth has no privileged status. The introduction and elimination rules for standard systems of natural deduction make that obvious. Praising Gentzen's work in proof theory, Michael Dummett wrote: 'The generation of logical truths is thus reduced to its proper, subsidiary, role, as a by-product, not the core, of logic' (1981, p. 434).

Alternative logics standardly provide a candidate relation of logical consequence, not just a candidate set of logical truths. This is crucial for their treatment of the semantic paradoxes, many of which start from a premise that is not a logical truth, such as the observed identity λ = '$\neg$True(λ)' in one version of the Liar. How well does the present approach extend from logical truth to logical consequence?

The extension is not trivial. We have taken the target of the logical inquiry to be the typically non-metalinguistic general principles involving only the selected logical constants, just as the target of a physical inquiry may be general principles involving only terms from a language for physics. In the logical case, those general principles correspond to logical truths. It is not obvious how logical consequence might be supposed to enlarge such a target.

To define a framework for discussion, let a *consequence relation* for L be any relation between sets of sentences of L and sentences of L that, in place of |=, obeys the standard structural rules above of Assumption, Monotonicity, Cut, and (with respect to the selected set of logical constants) Uniform Substitution. L will of course have many different consequence relations in this sense. The narrowest consequence relation for L holds between any set of sentences of L and any of its members, and in no other case. The widest consequence relation for L holds between any set of sentences of L and any sentence of L. The *theorems* of a consequence relation $\vdash$ are the sentences α for which $\{\} \vdash \alpha$. Thus the narrowest consequence relation for L has no theorems, whereas every sentence of L is a theorem of the widest consequence relation for L. We can treat different consequence relations as rival attempts to theorize the typically non-metalinguistic subject matter indicated in section 1. For simplicity, we may assume that the language L has already been chosen to be expressive enough for present purposes, so we need not extend it further to L^+.

In building the four rules into the definition of 'consequence relation', the intent is not at all to put them above question, but simply to make the picture more definite for the sake of clarity. The arguments below can be adapted to wider classes of relations.

Of course, we could simply say that the relevant standard of success for consequence relations is how closely they approximate logical consequence as defined in section 1. More specifically, the consequence relation $\vdash$ should if possible be both *sound* ($\Gamma \models \alpha$ whenever $\Gamma \vdash \alpha$) and *complete* ($\Gamma \vdash \alpha$ whenever $\Gamma \models \alpha$). But that looks like a fundamentally metalinguistic inquiry. The question is how, if at all, the fundamentally non-metalinguistic conception in section 1 of the goal of the inquiry can be extended from logical truth to logical consequence.

One strategy is to take a consequence relation into account by trying to encode it in its set of theorems. Suppose that L contains a two-place sentence operator $\rightarrow$ which obeys conditional proof (the deduction theorem) and modus ponens, the standard introduction and elimination rules for a conditional, with respect to the consequence relation $\vdash$. In other words, for any formulas α and β and set of formulas Γ:

($\rightarrow$E) If $\Gamma \vdash \alpha \rightarrow \beta$ and $\Delta \vdash \alpha$ then $\Gamma \cup \Delta \vdash \beta$

(→I) If $\Gamma \cup \{\alpha\} \vdash \beta$ then $\Gamma \vdash \alpha \rightarrow \beta$

Then, using the standard structural rules above, one can easily show that $\{\alpha_1, \ldots, \alpha_n\} \vdash \beta$ if and only if $\vdash \alpha_1 \rightarrow (\alpha_2 \rightarrow (\ldots (\alpha_n \rightarrow \beta)\ldots)$. Thus, at least for finite sets of premises, each consequence is encoded in a corresponding theorem, so in a way one loses nothing by concentrating on logical truth. Even for infinite sets of premises, $\vdash$ may be *compact*, like the standard consequence relation for first-order logic, in the sense that $\Gamma \vdash \alpha$ if and only if $\Gamma^* \vdash \alpha$ for some finite subset Γ^* of Γ, in which case the consequence relation is still reducible to its theorems. Of course, $\vdash$ may be non-compact, like the standard consequence relation for second-order logic. In that case, we may still be able to reduce logical consequence to logical truth in an extension of the language with an infinitary conjunction operator $\wedge$. For suppose that every set of sentences Γ has a conjunction $\wedge\Gamma$, a single sentence, where for every sentence α, $\Gamma \vdash \alpha$ if and only if $\wedge\Gamma \vdash \alpha$. Then $\Gamma \vdash \alpha$ if and only if $\vdash \wedge\Gamma \rightarrow \alpha$, as required.

However, for some non-classical logics such reductions are unavailable. Cases in point are the weak and strong Kleene logics. They are based on three-valued tables; we may label the values 'True', 'False', and 'Neutral'. True is the only designated value: $\Gamma \vdash \alpha$ if and only if whenever every member of Γ is assigned True, so is α. For the strong Kleene tables, if # is an *n*-place connective then $\#(\alpha_1, \ldots, \alpha_n)$ is assigned Neutral if every one of $\alpha_1, \ldots, \alpha_n$ is assigned Neutral. For the weak Kleene tables, $\#(\alpha_1, \ldots, \alpha_n)$ is assigned Neutral if at least one of $\alpha_1, \ldots, \alpha_n$ is assigned Neutral. Either way, any sentence built up by connectives out of atomic sentences is assigned Neutral on the line of the three-valued table on which every atomic sentence is assigned Neutral.[3] Thus no such formula is assigned True on every line: the logic has no theorems at all. In particular, no conditional → is definable in it for which (→I) holds, since that would require $p \rightarrow p$ to be a theorem, because $\{p\} \vdash p$ by Assumption. Nevertheless, neither the strong nor even the weak Kleene logic trivializes logical consequence. For example, the usual introduction and elimination rules for conjunction still hold: $\{\alpha \,\&\, \beta\} \vdash \alpha$, $\{\alpha \,\&\, \beta\} \vdash \beta$, $\{\alpha, \beta\} \vdash \alpha \,\&\, \beta$. For such logics, we cannot afford to focus exclusively on logical truth. We especially cannot afford to do so when comparing non-classical logics proposed in response to the semantic paradoxes, since one of the most salient is strong Kleene logic. If we just look at its theorems, it seems powerless; if we look at its consequence relation, it seems moderately powerful.

Even for logics with the full set of classical theorems, theoremhood may be a poor guide to the consequence relation when →E or →I fails. Here is an extreme example of a kind opposite to the Kleene logics. For a propositional language, let $\Gamma \vdash \alpha$ if and only if α is either a classical tautology or a member of Γ. Thus $\vdash$ yields the same set of theorems as classical logic. One can easily check that $\vdash$ obeys the structural rules for consequence relations. Nevertheless, it is radically impoverished. In particular, it lacks both →E and →I. For example, where p and q are distinct atomic sentences, we have $\{p \rightarrow q\} \vdash p \rightarrow q$ and $\{p\} \vdash p$ but not $\{p \rightarrow q, p\} \vdash q$, so →E fails, and we have $\{p, q\} \vdash q$ but not $\{q\} \vdash p \rightarrow q$, so →I fails. If we just look at its theorems, the logic seems very powerful; if we look at its consequence relation, it seems almost powerless.

We need our methodology for comparing logics to be capable of taking seriously logics that lack a connective → obeying →I and →E. Clearly, then, it must take account of the full consequence relation directly, not just of its precipitate in theorems.

There is a natural proposal. For a consequence relation $\vdash$ and a set Γ of sentences, let $Cn_{\vdash}(\Gamma)$ be $\{\alpha: \Gamma \vdash \alpha\}$, the set of consequences of Γ with respect to $\vdash$, in other words, the theory generated by $\vdash$ from Γ. Suppose that we are comparing the consequence relations $\vdash$ and $\vdash^*$. Then we should not simply compare the theorems of $\vdash$ with the theorems of $\vdash^*$. Rather, we should compare the theories they generate from independently well-confirmed sentences, such as well-established principles of physics. That is, we should compare $Cn_{\vdash}(\Gamma)$ with $Cn_{\vdash^*}(\Gamma)$ as theories for various independently well-confirmed sets Γ of sentences of L. We require Γ to be highly confirmed because the best of logics will draw some bad conclusions from bad premises, and for reasons of methodological fairness we require the confirmation to be independent in the sense that it is not too sensitive to the choice of logic.[4] Comparing the theorems of the consequence relations is just the limiting case where $\Gamma = \{\}$, for the empty set is vacuously well-confirmed.

The proposal vindicates the idea that the comparison of consequence relations is primarily non-metalinguistic, since the comparison of the theories $Cn_{\vdash}(\Gamma)$ and $Cn_{\vdash^*}(\Gamma)$ will not in general be a primarily metalinguistic inquiry, unless the subject matter of Γ itself happens to be metalinguistic. The proposal gives full weight to the service role of logic in drawing out the consequences of assumptions or beliefs that are not themselves logical truths, but it also takes the candidate logical truths into account in their own right.

The proposal need not displace the ideal for our consequence relation of soundness and completeness with respect to logical consequence ($\models$). Rather, it may be a means to achieving that ideal when it is unclear what rules of logic we should reason by. For we can use normal scientific standards of theory comparison in comparing the theories generated by rival consequence relations. Thus the evaluation of logics is continuous with the evaluation of scientific theories, just as Quine suggested (1951). We must now say something about what those standards are.

3. *Abductive methodology in philosophical logic*

We make the standard assumption that scientific theory choice follows a broadly abductive methodology. Scientific theories are compared with respect to how well they fit the evidence, of course, but also with respect to virtues such as strength, simplicity, elegance, and unifying power. We may speak loosely of inference to the best explanation, although in the case of logical theorems we do not mean specifically *causal* explanation, but rather a wider process of bringing our miscellaneous information under generalizations that unify it in illuminating ways. We do not fully understand *why* this methodology works so well. In particular, much remains to be done in clarifying the relevance of aesthetic or pragmatic criteria like simplicity and elegance to questions of truth and falsity. Nevertheless, it is clear that if we do not prefer them in theories to complication and ugliness, we face a hopeless proliferation of *ad hoc* projections from our data. The abductive methodology is the best science provides, and we should use it.[5] In particular, we should use it when comparing the theories generated from a given set of premises by rival consequence relations.

The idea of applying an abductive methodology to logic and mathematics is far from new. According to Bertrand Russell, what he calls 'mathematical philosophy'

> proceeds, by analysing, to greater and greater abstractness and logical simplicity; instead of asking what can be defined and deduced from what is assumed to begin with, we ask instead what more general ideas and principles can be found, in terms of which what was our starting-point can be defined or deduced.

The role of such an inquiry is 'to take us backward to the logical foundations of the things that we are inclined to take for granted in mathematics' (Russell 1919, pp. 1-2).

We cannot attempt a general discussion of abductive methodology here. However, some specific comments on the criteria may be useful, concerning their application to logic.

First comes fit with the evidence. At a minimum, one might think, it requires *consistency* with the evidence. But that raises the worry that the operative standard of consistency is set by some transcendental background logic, thereby undermining the ideal of fair process in comparing alternative logics. Another worry is that a standard of consistency is unfair to dialetheist treatments of the semantic paradoxes, which embrace the contradictions and restrict the logic to block their trivializing consequences.[6] We can answer the first worry by using the logics under test to set their own standard of consistency. Moreover, we can answer the second worry too by treating consistency as avoidance of trivialization rather than avoidance of contradiction. To be more precise, let E be the relevant evidence, expressed in a set of sentences of L. Then a consequence relation ⊢ is *consistent* with E if and only if for not every sentence α of L does it hold that E ⊢ α. That criterion should be acceptable even to dialetheists, who are almost as keen as everyone else to avoid trivialization, as well as to classical logicians and others who accept that everything follows from a contradiction.

Such consistency with the evidence does not exhaust fit with the evidence. A theory to which the evidence is simply irrelevant is consistent with the evidence. A theory fits the evidence better if the evidence verifies some of its predictions as well as falsifying none of them. Evidence here is not confined to observations. We may use anything we know as evidence (Williamson 2000). For example, in the case of propositional modal logic, we may know that the coin could have come up heads, and could have not come up heads, but could not have both come up heads and not done so, and on that basis eliminate this proposed law:

$$(\Diamond p \,\&\, \Diamond q) \rightarrow \Diamond(p \,\&\, q).$$

By contrast, this law identifies a useful pattern in the modal data:

$$(\Diamond p \vee \Diamond q) \leftrightarrow \Diamond(p \vee q)$$

In that sense, we can verify some predictions of the law by using our pre-theoretic ability to evaluate particular modal claims. The law even goes some way towards unifying and explaining its instances, by bringing them under an illuminating generalization.

The criterion of strength also requires clarification in the context of logic. In one standard logical sense, a theory T is *stronger* than a theory T* if and only if T entails T* but T* does not entail T: every theorem of T* is a theorem of T, but not every theorem of T is a theorem of T*. Similarly, one might call a consequence relation ⊢ stronger than a consequence relation ⊢* if and only if whenever ⊢* holds, so does ⊢, but ⊢ sometimes holds when ⊢* does not. Since we are concerned with theories in a given interpreted language L, and consequence relations for L, we do not consider ‘translations’ between languages, or non-homophonic ‘translations’ of L into itself, in making comparisons of strength.[7] However, relative strength of the logical sort just explained is rarely at issue in the abductive comparison of scientific theories. Typically, we are comparing internally consistent theories that are inconsistent with each other (on any reasonable standard of consistency). But if T is stronger in the sense above than T*, so T entails T*, yet T is also inconsistent with T*, then T is internally inconsistent. For abductive purposes, we need a looser sense in which one internally consistent theory may be stronger than another with which it is inconsistent, because the former is more specific or informative than the latter. For instance, ‘The time is between 3.14 and 3.16’ is more specific than ‘The time is between 4.00 and 12.00’, even though they are inconsistent with each other. To take an extreme case, let T be a consistent scientific theory axiomatized by a conjunction of universal generalizations with many interesting consequences, and let T* be the consistent theory axiomatized by the negation of that conjunction. T* is inconsistent with T and concerns the same subject matter; if the probability of T on our evidence is less than 0.5, then T* is more probable than T on our evidence. Nevertheless, T* would typically not even be treated as a rival theory to T, because it is too uninformative: it says that there is a counterexample somewhere or other to one of the universal generalizations in T, but nothing more. One role for the informal scientific standard of strength is to provide a minimal threshold of informativeness below which theories do not even come up for serious abductive evaluation. We want scientific theories to inform us about their subject matters; weak theories do too little of that to give us what we want. Furthermore, strength contributes to explanatory power in the broad sense sketched above, the capacity to bring our miscellaneous information under generalizations that unify it in illuminating ways.[8]

If T is stronger than T* in the strict logical sense, then T is also stronger than T* in the looser scientific sense, but the converse fails. Both senses are applicable to logical theories. For instance, let PC be standard classical propositional logic, and IC be intuitionist propositional logic. Then every theorem of IC is a theorem of PC but not conversely, since $p \vee \neg p$ is a theorem of PC but not of IC;

likewise for the corresponding consequence relations. Thus PC is stronger than IC in the strict logical sense, and so also in the looser scientific sense. Now extend the language to include quantification into sentence position (propositional quantification). Thus the natural extension PC^+ of PC to the extended language has the universal generalization $\forall p\ (p \vee \neg p)$ as a theorem. Arguably, the appropriate extension IC^+ of IC to the extended language should have its negation $\neg \forall p\ (p \vee \neg p)$ as a theorem: for intuitionists, it is absurd to suppose $\forall p\ (p \vee \neg p)$ assertible, since that would require a decision procedure for all sentences of the language, which is impossible, and intuitionistically that absurdity makes $\neg \forall p\ (p \vee \neg p)$ assertible. They deny the universal generalization even though they cannot deny any instance of it (since $\neg(p \vee \neg p)$ is intuitionistically as well as classically inconsistent). Thus neither PC^+ nor IC^+ is stronger than the other in the strict logical sense. Nevertheless, PC^+ is stronger than IC^+ in the looser scientific sense, just as the scientific theory T was stronger than T* above, in the way that a universal generalization is typically more informative than its negation. Remember, we are *not* concerned with 'translations' between classical and intuitionistic languages because we are considering the corresponding logics as formulated in a single already interpreted language.

In discussion of alternative logics, it is not always recognized that strength is a strength, in logical theories as in others. One often encounters various forms of exceptionalism about logic, according to which weakness is a strength in logic, because weak logics leave open more possibilities, prejudge fewer issues, and achieve higher levels of neutrality. However, such tendencies have no natural stopping-off point short of an empty consequence relation, since any logical principle whatsoever is in principle open to challenge. Indeed, virtually every salient logical principle has actually been challenged by some philosopher or other. Attempts to argue that the challenges are just verbal typically fail to do justice to the role of an interpreted public language in which all parties are competent for the formulation of such challenges: one is reminded of a bland spokesman for a totalitarian regime, assuring us that its critics do not mean what they say. Since I have discussed this issue at length elsewhere, I will not labour it further here (Williamson 2007, 20013, 2014). Henceforth, I will assume the abductive methodology.

Once we assess logics abductively, it is obvious that classical logic has a head start on its rivals, none of which can match its combination of simplicity and strength. Its strength is particularly clear in propositional logic, since PC is Post-complete, in the sense that the only consequence relation properly extending the classical one is trivial (everything follows from anything). First-order classical logic is not Post-complete, but is still significantly stronger than its rivals, at least in the looser scientific sense, as well as being simpler than they are; likewise for natural extensions of it to more expressive languages. In many cases, it is unclear what abductive gains are supposed to compensate us for the loss of strength involved in the proposed restriction of classical logic.

None of this is yet to say that no non-classical logic can overcome the initial advantages of classical logic once we move to a wider setting, by considering fit with evidence or with other scientific theories, or by treating more expressions as logical constants. Quantum logic is an obvious test case. Quine (1951) invoked proposals to revise the law of excluded middle in order to simplify quantum mechanics, and Putnam (1969) more pertinently proposed rejecting one of the distributive laws of PC as a precondition for understanding what is physically occurring in two-slit experiments. In both cases, the idea was that the intrinsic abductive advantages of classical logic over non-classical alternatives are trumped by the abductive advantages of quantum mechanics plus the preferred non-classical logic over quantum mechanics plus classical logic. The general methodology proposed in this paper does not preclude such a challenge to classical logic. In the terminology of section 2, comparing the consequence relation $\vdash$ of classical logic with the consequence relation $\vdash^*$ of the non-distributive logic at issue involves comparing the theory $Cn_{\vdash}(\Gamma)$ with the theory $Cn_{\vdash^*}(\Gamma)$, where Γ comprises principles of quantum mechanics. However, both Quine (1970, pp. 85-6) and Putnam (2012) later came to a negative verdict on quantum logic. Most significantly, in practice rejecting classical logic just does not seem to help us understand the nature of quantum reality. Thus quantum logic is not in practice an encouraging precedent for critics of classical logic. Nevertheless, the challenge had to be argued through in immanent detail; it could not merely be dismissed on transcendental grounds.

The strong *prima facie* abductive case for classical logic just noted does *not* depend on a principle of conservativism. It does not rely on the position of classical logic as the status quo, the logic we more or less currently accept, nor does it appeal to the benefits of familiarity or the costs of change. It concerns intrinsic features of classical logic, such as simplicity and strength, which it would have even if we currently accepted some non-classical logic. The case may indeed be strengthened by reference to the track record of classical logic: it has been tested *far* more severely than any other logic in the history of science, most notably in the history of mathematics, and has withstood the tests remarkably well. Nevertheless, the initial abductive case for classical logic would be quite powerful, even if we had only stumbled across that logic a few weeks ago.

4. *Application to the semantic paradoxes*

We can now consider the semantic paradoxes as a challenge to classical logic, and apply the abductive methodology just sketched to the challenge.

At first sight, the semantic paradoxes constitute unusually promising ground for an abductive critique of classical logic. They seem to rely on a combination of classical logic with identity and a disquotational principle for truth to derive absurd consequences from easily verified premises, such as that λ = '$\neg$True(λ)', where the sentence 'λ is not true' has indeed been labelled 'λ'; such a sentence simply articulates part of our evidence. Thus we seem to be forced to restrict either classical logic or disquotation. But if we restrict classical logic in suitable ways, then we can hold on to an unrestricted disquotation principle which allows us to treat any sentence α as freely intersubstitutable with a sentence True('α') that applies the truth predicate to a quotation name 'α' of α. Thus although the restriction of classical logic involves a loss of both simplicity and strength, it compensates us by saving the simplicity and strength of unrestricted disquotation. Saving the simplicity and strength of unrestricted classical logic forces us to sacrifice the simplicity and strength of unrestricted disquotation. Which is the better deal?

In this respect, the case against classical logic from the semantic paradoxes is better than most cases against classical logic, such as that from the sorites paradoxes. For the latter typically involve no principles as simple and strong as disquotation to compensate for the lost simplicity and strength of classical logic (for the case of the sorites paradoxes see Williamson 1994).

To apply our abductive methodology to the semantic paradoxes, we should add both the truth predicate and the quotation device to the usual list of logical constants. The reason is of the pragmatic kind explained in section 1: in this context part of our interest is in what general principles the truth predicate and the quotation device obey, so we should hold them fixed, and not worry whether they are *really* logical devices in some mysterious deep sense. Thus we count as revising logic whether we revise excluded middle or disquotation. However, the term 'classical logic' will as usual be confined to the logic of the more standard logical constants: negation, disjunction, conjunction, the quantifiers, and the identity predicate. Thus, given our evidence, the choice is indeed between revising classical logic and revising disquotation.

Let us return to the choice between restricting classical logic and restricting disquotation. On second thoughts, one might doubt the apparent symmetry between the options. For the constants of classical logic seem to express absolutely fundamental structure. By contrast, the constants at issue in the disquotational principle — the truth predicate, quotation marks — seem to express much less fundamental matters, specific to the phenomenon of language. Thus the comparison between classical logic and disquotation looks analogous to the contrast between a successful theory in fundamental physics and a successful theory in one of the special sciences, such as economics. Suppose that the economic theory is found to be inconsistent with the fundamental physical theory. Faced with the choice as to which theory to restrict in order to preserve the other unrestricted, which would you choose? Perhaps one can imagine unusual circumstances in which it would be better to restrict the fundamental physical theory in order to preserve the economic theory unrestricted. Nevertheless, on general methodological grounds, that would usually be a perverse choice. It would normally be better to make the opposite choice, and restrict the economic theory in order to preserve the fundamental physical theory unrestricted. By analogy, then, on general methodological grounds it would normally be better to restrict disquotation in order to preserve classical logic unrestricted, and perverse to do the opposite.

Friends of disquotation may dispute the analogy. In particular, they may argue that the concept of truth is itself implicitly fundamental to our usual understanding of classical logic, through both the standard truth-conditional account of the meanings of the classical logical constants and the standard Tarskian account of logical consequence as generalized truth-preservation. The truth predicate and quotation marks are then just the linguistic devices required to articulate that fundamental conception. On this view, the disquotational principle for truth concerns as fundamental a level as does classical logic.

Our general abductive methodology helps us resolve the apparent impasse. As we saw, it recommends us not to compare two logics only in isolation, but also to compare the results of combining each of them with well-confirmed results from outside logic, such as principles of natural science. But now a crucial asymmetry becomes visible. For any complex scientific theory, especially one that involves some mathematics, will make heavy use of negation, conjunction, disjunction, the quantifiers, and identity. Thus restricting classical logic will tend to impose widespread restrictions on its explanatory power, by blocking the derivation of its classical consequences in particular applications. By contrast, most scientific theories make no use whatsoever of a truth predicate and quotation marks, because they are not metalinguistic theories. Hence restricting disquotation makes no difference to their explanatory power. Once we take such extra-logical applications into account, restricting classical logic must involve a vastly greater abductive loss than does restricting disquotation.

Quine already made a similar point, in relation to both the semantic and the set-theoretic paradoxes:

The classical logic of truth functions and quantification is free of paradox, and incidentally it is a paragon of clarity, elegance, and efficiency. The paradoxes emerge only with set theory and semantics. Let us then try to resolve them within set theory and semantics, and not lay fairer fields to waste. (1970, p. 85)

However, two differences may be noted. First, Quine bases his argument on 'the maxim of minimum mutilation'. By contrast, I argued above that abductive comparisons of the relevant kind need not make such appeals to a principle of conservativism. Second, the quotation comes from *Philosophy of Logic*, in which Quine notoriously claims of the deviant logician: 'when he tries to deny the doctrine he only changes the subject' (1970, p. 81). Here Quine takes disputes over alternative logics to be verbal. By contrast, I have argued that they typically involve genuine non-verbal disagreement (a view perhaps closer to that of Quine 1951).[9]

But wait: the friends of disquotation are not finished yet. Like other opponents of classical logic, they argue that it fails only in exceptional cases, and can be recovered in non-exceptional ones. When we need it, we can have it. For example, many proponents of a non-classical approach to the semantic paradoxes opt for the strong Kleene logic K_3. True is the only designated value. Liar-like sentences are supposed to be neither True nor False. The law of excluded middle fails in K_3, because both $\neg p$ and $p \vee \neg p$ are Neutral when p is Neutral. However, adding $p \vee \neg p$ to K_3 for each atomic sentence p restores classical logic, since it is True when and only when p is True or False, and the Kleene tables yield the standard bivalent outputs whenever all the inputs are bivalent. Thus the Kleenean recovers classical logic for non-paradoxical sentences by adding appropriate instances of excluded middle. This goes for Hartry Field's approach (2008), and for many other non-classical logics.

Of course, it can be far from obvious which sentences *are* paradoxical. Semantic paradoxes can be contingent on all sorts of circumstance (Kripke 1975). Even if we are working in a language for physics, with no overtly semantic vocabulary, a reductionist about semantics may worry that the property of Truth is expressed by some complex predicate or other of the language for physics, so that confining our sentences to that language is no guarantee of non-paradoxicality.

There is a more general concern. The retreat to K_3 invalidates vast swathes of ordinary mathematical reasoning, since mathematicians freely reason in ways that depend on the law of excluded middle. Such restrictions on mathematics in turn restrict its applications to natural science. The natural scientists might overcome the restrictions by postulating instances of excluded middle as needed.[10] But then their explanations invoke those auxiliary assumptions, which reduces their explanatory value; elegant explanations get as much as possible out of as little as possible. The point is not that the auxiliary assumptions exceed the classical logician's commitments: they do not, because the classical logician is anyway committed to the unrestricted law of excluded middle. Rather, the point is that the auxiliary assumptions are *ad hoc* for the Kleenean in a way they are not for the classical logician, who derives them all from the simple, elegant, general principles of classical logic. The Kleenean can give them no such general explanation. The best the Kleenean can do is derive them from a metalinguistic principle such as that all instances of excluded middle in the language of physics are true; but it would be bizarre to claim that ordinary physical explanations which happen to need excluded middle actually involve such metalinguistic considerations. Thus the

Kleenean strategy pays a heavy abductive cost across a vast range of ordinary science, by remodelling ordinary scientific explanations in ways that introduce numerous *ad hoc* assumptions. By standard abductive criteria, the classical strategy does significantly better, because its abductive costs are restricted to metalinguistic discourse.

Similar considerations apply to other non-classical treatments of the semantic paradoxes. For instance, a dialetheist may argue for truth-value gluts rather than gaps: if p is paradoxical, then p is both true and false, so the contradiction $p \wedge \neg p$ is both true and false too. The dialetheist may classify the rule of disjunctive syllogism as invalid, on the grounds that when p is both true and false but q is simply false, the disjunction $\neg p \vee q$ is both true and false, so the argument from $\neg p \vee q$ and p to q has true premises (even though one of them is also false) and a simply false conclusion. That would invalidate a vast array of arguments in mathematics and the rest of science. The dialetheist may respond by permitting instances of disjunctive syllogism in non-paradoxical cases, whichever they are. But, just as before, that still involves a heavy abductive cost across a vast range of ordinary science, by remodelling ordinary scientific explanations in ways that introduce numerous *ad hoc* elements.

The piecemeal reintroduction of instances of missing classical principles involves heavy abductive costs through loss of simplicity and elegance. The appeal to the greater complexity of non-classical logics here involves more than the usual impressionistic claim that classical logic is simpler than its non-classical rivals. In the context of the semantic paradoxes, technical results are available about the extreme complexity of some non-classical proposals.[11] By standard abductive criteria, it is far better to keep classical logic unrestricted and restrict disquotation than to keep disquotation unrestricted and restrict classical logic.

To sharpen our sense of the abductive loss involved in the non-classical proposals, we may consider the opposite process, of gain through theoretical unification. Suppose that we have explained many different physical phenomena, using specific auxiliary hypotheses on a case-by-case basis. Now someone notices that vast numbers of those auxiliary hypotheses can all be subsumed as instances of a single simple universal generalization, which we then postulate as a law. That would normally be regarded as progress, a very significant abductive gain. The non-classical proposals for the semantic paradoxes amount to reversing just such a step. That involves an abductive loss equal to the abductive gain in taking the original step forward.

5. *Complications*

The foregoing argument over-simplifies in various way. Although they complicate the overall picture, they change its overall outline little. Still, they deserve mention.

The competing theories of truth were assumed to be all formulated within the same interpreted language L. But many theories of truth essentially involve extensions of the original language specific to their approach. Tarski's account is an obvious example, since it requires a hierarchy of languages.[12] Even non-hierarchical theories of truth may require distinctive new vocabulary, for instance to classify pathological sentences. This proliferation of new technical terms is of course normal for scientific theories, and poses no insuperable obstacle to abductive comparisons between them. It may, however, initially obscure their logical relations, since they may use homophonic expressions with different meanings, or non-homophonic expressions with the same meaning. Serious issues of translation between the rival technical vocabularies may arise, as in any other science. And, as in any other science, such problems do not warrant overblown claims of wholesale incommensurability between rival theories, such as Kuhn and Feyerabend once made fashionable on the basis of a mistaken philosophy of language. We must be sensitive to semantic issues and flexible in handling them when we compare theories of truth abductively, but we can still make the comparisons.

The assumption of constancy in reference across contexts in the semantics of the given language L also needs to be lifted, since some important explanations of the paradoxes postulate variation in the reference of semantic terms such as 'true' across contexts (Parsons 1974, Burge 1979). This too complicates abductive comparisons between theories without rendering them impossible. Criteria such as strength, simplicity, and evidential fit can be applied even to theories formulated partly in context-sensitive terms.

A related idea is that in paradoxical contexts semantic terms like 'true' may *crash*, becoming locally semantic defective. Then meaningful sentences involving them may fail to say anything. Such malfunctions in the working of the linguistic mechanism may occur unsystematically, as malfunctions in general tend to do.[13] This idea may even suggest that we were too quick in assuming that the semantic paradoxes make classical logic and disquotation incompatible.[14] For the bad results of substituting semantically defective terms into principles of classical logic and disquotation no more show something wrong with those principles than the bad results of substituting 'this little green man' or '0/0' into a law of physics show something wrong with that law. We may simply regard such semantically defective instances as not genuine instances. Can we then hold on to *both* classical logic *and* disquotation after all?

If crashes depend on context, then to theorize explicitly about such context-dependent phenomena we must relativize 'true' and other semantic expressions to contexts. We cannot expect to generalize disquotational principles across contexts. Although "It is cold here" is true as uttered in the present context if and only if it is cold here, and that whole biconditional sentence is true as uttered in other contexts too, it is not cold here (in Oxford) even though "It is cold here" is true as uttered in various other contexts (such as the South Pole). In that banal sense, disquotation must be restricted anyway. But that still leaves disquotation with respect to the theorist's own context. The danger is that the semantic paradoxes force the theorist of crashes into locating some of those crashes for the terms of the crash theory itself in the very context of the theorizing, which looks like some sort of defeat for the crash theory itself. Of course, once we are willing to theorize about L in a metalanguage for L that contains a predicate 'true' for L not in L itself, we can have a disquotational principle for that predicate over all sentences of L, but such a potentially hierarchical approach falls short of the originally envisaged form of disquotation. And presumably the term 'crash' will have to get the hierarchical treatment too. But we can still hope for a simple, strong version of the hierarchical approach that fits our evidence.

Many ways of handling the semantic paradoxes are consistent with classical logic. There is no need to try to rank them here. But when they are ranked, the usual abductive methodology will of course apply.

The methodological case made here for maintaining classical logic even in the face of the semantic paradoxes is a very natural one. Why have such obvious considerations been so widely neglected? Presumably, the explanation lies in the tendency to discuss the paradoxes in isolation. A narrow focus has many advantages, but from time to time it is worth looking at the bigger picture.

Notes

1 Although Tarski 1936 is often cited as the origin of the model-theoretic conception of logical consequence, its assignments are not accompanied by a specification of a domain in the model-theoretic sense. He seems to have envisaged the quantifiers as ranging unrestrictedly over the relevant type in a type-theoretic hierarchy. Thus first-order quantifiers range over absolutely all individuals. See Williamson 2003 and 2013, pp. 221-261, for more discussion.

2 We spare the reader the further constraints needed to avoid clashes of variables. For instance, $\forall Rxx$ should not count as a substitution instance of $\forall Rxy$.

3 Since L is an already interpreted language, the status of the three-valued semantics is that of a *theory* about the semantics of L, comparable to that of a theory about the semantics of English or some other natural language. It is no mere stipulation.

4 We require the sentences in Γ to be jointly well-confirmed, not just individually so, to avoid problems with lottery and preface paradoxes.

5 For a general discussion and assessment of abductive methodology see Lipton 2004.

6 A seminal work here is of course Priest 1987.

7 The reason for the inverted commas is that such mappings typically fail to preserve meaning.

8 Proof-theoretic criteria of strength are also too coarse-grained for present purposes.

9 See Cozzo 1998, pp. 271-9, for an interesting development of Quine's view in a similar spirit to that of this paper (although I regard the commitment to evidence-transcendent truth he discusses as an interesting outcome rather than a *cost* of classical logic with known auxiliary assumptions).

10 Halbach (2011, p. 293) points out that the use of K_3 for treating the semantic paradoxes involves an essential loss of the principle of transfinite induction up to the ordinal ε_0, and a corresponding loss of combinatorial principles of arithmetic, although that would presumably not be crippling for most natural science.

11 See the appeal to complexity considerations in the critique of Field (2008) by Welch (2011).

12 See Tarski 1935. Williamson 1998 postulates a less systematic sort of linguistic variation provoked by the semantic paradoxes.

13 See Smiley 1993 for a sketch of such an approach.

14 Of course, as well as classical logic and disquotation, we also need some elementary syntax or the equivalent to derive the paradoxes by diagonalization or the like. Above, the equation λ = '$\neg$True(λ)' was treated as part of our evidence. Alternatively, one could treat it as up for abductive assessment too. That would only reinforce the main moral of the paper, that classical logic is the very last thing to revise.

15 Previous versions of this material were presented at a conference in honour of Michael Dummett at the Institute of History and Philosophy of Sciences and Techniques in Paris and a seminar at Oxford University, and earlier in some of the 2013 Hägerstrom Lectures at Uppsala University. It has benefited greatly from questions and comments on those occasions.

References

Burge, Tyler 1979: ‘Semantical paradox’, *Journal of Philosophy*, 76, pp. 169-198. Reprinted with postscript in Martin 1984, pp. 83-117.

Cozzo, Cesare 1998: ‘Epistemic truth and excluded middle’, *Theoria*, 64, pp. 243-282.

Dummett, Michael 1981: *Frege: Philosophy of Language*. London: Duckworth.

Field, Hartry 2008: *Saving Truth from Paradox*. Oxford: Oxford University Press.

Halbach, Volker 2011: *Axiomatic Theories of Truth*. Cambridge: Cambridge University Press.

Kripke, Saul 1975: ‘Outline of a theory of truth’, *Journal of Philosophy*, 72, pp. 690-716.

Lipton, Peter 2004: *Inference to the Best Explanation*, 2nd ed. London: Routledge.

Martin, Robert (ed.) 1984: *Recent Essays on Truth and the Liar Paradox*. Oxford: Clarendon Press.

Parsons, Charles 1974: ‘The Liar Paradox’, *Journal of Philosophical Logic*, 3, pp. 381-412. Reprinted with postscript in Martin 1984, pp. 9-45.

Priest, Graham 1987: *In Contradiction: A Study of the Transconsistent*. The Hague: Martinus Nijhoff. 2nd (extended) edition, Oxford: Oxford University Press, 2006.

Putnam, Hilary 1969: ‘Is logic empirical?’, in R. Cohen and M. Wartofsky (eds.), *Boston Studies in the Philosophy of Science*, 5, pp. 216-241. Dordrecht: Reidel.

Putnam, Hilary 2012: ‘The curious story of quantum logic’, in his *Philosophy in an Age of Science: Physics, Mathematics, and Skepticism*, M. De Caro and D. Macarthur (eds.) pp. 162-177. Cambridge, Mass.: Harvard University Press.

Quine, Willard van Orman 1951: ‘Two dogmas of empiricism’, *Philosophical Review*, 60, pp. 20-43.

Quine, Willard van Orman 1970: *Philosophy of Logic*. Englewood Cliffs: Prentice Hall.

Russell, Bertrand 1919: *Introduction to Mathematical Philosophy*. London: George Allen & Unwin.

Smiley, Timothy 1993: ‘Can contradictions be true?’, *Proceedings of the Aristotelian Society*, sup. 67, pp. 17-33.

Tarski, Alfred 1935: ‘The concept of truth in formalized languages’, in Tarski 1983, pp. 152-278. First published as ‘Der Wahrheitsbegriff in den formalisierten Sprachen’, *Studia Philosophica Commentarii Societatis Philosophicae Polonorum*, 1, 261-405.

Tarski, Alfred 1936: ‘On the concept of logical consequence’, in Tarski 1983, pp. 409-20. First published as ‘O pojciu wynikania logicznego’, *Przegląd Filozoficzny*, 39, pp. 58-68.

Tarski, Alfred 1983: *Logic, Semantics, Metamathematics*, 2nd edn. Trans. J.H. Woodger, ed. J. Corcoran. Indianapolis: Hackett.

Tarski, Alfred 1986: ‘What are logical notions?’, *History and Philosophy of Logic*, 7, pp. 145-154.

Welch, Philip 2011: ‘Truth, logical validity, and determinateness: a commentary on Field’s *Saving Truth from Paradox*’, *Review of Symbolic Logic*’, 4, pp. 348-359.

Williamson, Timothy 1994: *Vagueness*. London Routledge.

Williamson, Timothy 1998: ‘Indefinite extensibility’, *Grazer Philosophische Studien*, 55, pp. 1-24.

Williamson, Timothy 2000: *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, Timothy 2003: ‘Everything’, *Philosophical Perspectives*, 17, pp. 415-465.

Williamson, Timothy 2007: *The Philosophy of Philosophy*. Oxford: Wiley-Blackwell.

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

Williamson, Timothy 2014: 'Logic, metalogic, and neutrality', *Erkenntnis*, 79, pp. 211-231.