The Place of Quine
In
Analytic Philosophy

Scott Soames
USC
School of Philosophy

To Appear in

A Companion to W.V.O. Quine

Edited by Ernie Lepore and Gilbert Harman
John Wilely and Sons Publisher

The Place of W.V.O Quine in Analytic Philosophy
Scott Soames

Quine was born on June 25, 1908 in Akron Ohio. From 1926 to 1930 he attended Oberlin College, from which he graduated with a B.A. in mathematics that included reading in mathematical philosophy. He received his PhD from Harvard in 1932 with a dissertation on *Principia Mathematica* advised by Whitehead. The next year traveling on fellowship in Europe, where he interacted with Carnap, Tarski, Lesniewski, Lukasiewicz, Schlick, Hahn, Reichenbach, Gödel, and Ayer. He was back in Cambridge between 1933 and 1936 as a Junior Fellow at the Harvard Society. In 1936, he joined the Harvard faculty, where he remained for 42 years, except for 3 years in the Navy in World War II. Returning after the war, he was promoted to Professor in 1948. Although he retired in 1978, he retained his office and remained active through much of the 1990s. Quine died on Christmas Day 2000.

The Harvard faculty he entered in 1936 included Whitehead, Ralph Barton Perry (who edited the papers of William James), Henry Sheffer (of the "Sheffer stroke"), and C.I. Lewis. An eclectic thinker, Lewis combined the perceptual realism of Perry with the Kantianism of Josiah Royce, and the pragmatism of Peirce. Like Perry, Lewis believed that perception and knowledge require an independent reality given in experience. Like Royce, he believed experience to be structured by concepts added by the mind. Like Peirce he held that these concepts are revisable in light of experience and that the "meanings" of concepts and thoughts lie in their success in predicting new experience and grounding successful action.[1] Despite having some commonalities with logical empiricists, Lewis was never one of them. While sharing their scientific naturalism, their emphasis on analysis, and their view of testable consequences as the basis of empirical significance, he opposed their

---

[1] These ideas are developed in Lewis (1929) and (1946), which were widely read – the former being the subject of a seminar at Oxford led by J. L. Austin and Isaiah Berlin in 1936-7.

noncognitivism about value, their physicalism, and their "linguistic turn." For him, the primary bearers of meaning and truth were thoughts, a point on which he differed from Quine. Lewis also differed from Quine in embracing analyticity and modal logic, to which he contributed the axiomatic S-systems.

This was the milieu into which Quine stepped as a young professor. It wasn't until after the war that his impact was widely felt. Even in 1948 his influence on his own department was too weak to secure the proposed appointment of his friend Rudolf Carnap. By 1953, when Lewis retired, Quine was ready to lead the first great analytic department in America. With the publication of Quine (1948, 1951a,b, 1953a,b,c, 1956), he was recognized as a world leader in philosophy. The first American to achieve this status, he changed analytic philosophy by transcending the limitations of logical empiricism, and replacing it with a more thoroughly empiricist view.

The Logical Empiricist Background

The logical empiricism of Quine's predecessors was built on (i) the Frege-Russell rejection of the Kantian synthetic apriori in favor of a notion of analyticity encompassing logic, arithmetic, and mathematics (except for geometry about which Frege followed Kant, but which the logical empiricists took to be an empirical theory), (ii) the Russellian version of Humean empiricism in which the material objects were said to be *logical constructions* out of perceptual experience, and (iii) the tractarian idea that a test for meaning, or intelligibility, is central to philosophy. By 1934 Carnap had synthesized and extended these ideas into a new version of empiricism developed in Carnap (1928a,b, 1930, 1931, 1932a,b, 1934).

To the logicist program of Frege and Russell, he added Wittgenstein's conception of logical truths as tautologies, guaranteed to be true by the meanings of their logical terms. Accepting logicist definitions as explicating the meanings of arithmetical terms, Carnap

extended this status to all of mathematics, thereby attributing the apriority and necessity of logic and mathematics to analyticity. Correct philosophical analyses were treated similarly. Having said, "Philosophy is to be replaced by the logic of science – that is to say by the logical analysis of the concepts and sentences of the sciences,"[2] Carnap maintained that analyticity was philosophy's stock and trade. To him, the point seemed obvious. If a truth is apriori, the reason it can be known without justifying empirical evidence *must be* that it places no constraint on the world, but rather is true in virtue of meaning. Since a necessary truth provides no information about which possible state the world is in, it too must be analytic, and empty of empirical content. Finally, experience can tell us only about the way the world actually is; so anything we *know to be necessary* must be something that doesn't constrain the world at all, again because it is true in virtue of meaning.

By contrast, the aposteriori, contingent, and synthetic were subject to Carnap's version of Wittgenstein's intelligibility test. In the *Tractatus,* empirical descriptions of the world are meaningful, but other (non-analytic, non-contradictory) claims are not, including those of traditional (metaphysical or moral) philosophy. Whereas Wittgenstein's test maintained that the truth/falsity of all empirical statements is determined by the totality of atomic facts (which correspond to atomic truths), Carnap dropped talk of correspondence, and characterized empirical meaningfulness in terms of verifiability or falsifiability. Unlike the *Tractatus,* which offered an ineffable metaphysical parallel between language and the world, the early Carnap rejected talk of relations between words and things (a position he revised after encountering Tarski's theory of truth).

This is where things stood on the eve of Quine's first major article, "Truth by Convention," which attacked Carnap's linguistic theory of the apriori. Although Quine's

---

[2] Page 277 of the 1937 translation of Carnap (1934).

critique was powerful, for many years it didn't attract much attention or change many minds – probably because its target seemed undeniably correct to its proponents. Nevertheless, central tenets of the target theory are clearly problematic. Most obviously, the bearers of analyticity are sentences, whereas the bearers of apriority and necessity seem not to be. When one says that *it is necessary, and knowable apriori, that all squares are rectangles,* what is said to be necessary and knowable apriori is not the sentence, 'All squares are rectangles,' or any other. How, in light of this, is one supposed to move from the claim that S is analytic to the truth of claim ⌈it is necessary / knowable apriori that S⌉?

Although this wasn't a worry to which Quine or the logical empiricists paid much attention, it's not obvious how a proponent of the linguistic theory of the apriori (and the necessary) should deal with it. Here is one line of thought, the failure of which may be instructive. Let S be an analytic truth expressing p.

(i)   Since S is analytic, one can know that S expresses a truth by learning what it means.

(ii)   One will thereby know the metalinguistic claim q -- *that S expresses a truth* -- on the basis of the evidence E provided by one's experience in learning the meaning of S.

(iii)   Since one has come to understand S, one will also know, on the basis of E, that S expresses p (and only p).

(iv)   Combining (ii) and (iii), one will thereby know, on the basis of E, that p is true. Since p is an apriori consequence of this claim, one will be in position to come to know p.

(v)   However, the claim that E *justifies* – by ruling out possibilities in which it is false -- is not p, but q.

(vi)   Since p can be known without justifying evidence ruling out possibilities in which it is false, there must be no such possibilities.

(vii)   So, if S is analytic, p must be necessary, and (by the present reasoning) capable of being known to be so; p is also apriori, since knowledge of p doesn't require evidence justifying it.

Though one might be fooled by this reasoning, if it were left implicit, the problems with it – apart from (i) which we here accept for the sake of argument – are evident.[3] The most

---

[3] See chapters 3 and 4 of Williamson (2007) for a catalog of well-taken worries about (i).

obvious difficulty concerns the knowledge of p reached at step (iv). Anyone who comes to know p by this route will know it *aposteriori* – whether or not p is *knowable* apriori. Worse, p will be knowable apriori only if there is a *different* route to such knowledge – which threatens to undermine the point of the linguistic theory. Further, the reasoning described, by which one comes to know both p and the necessity of p, requires one to employ apriori logical knowledge independent of the linguistic conventions in question. So, even if there were no other problems with it, the argument would presuppose much of what the linguistic theory purports to explain. Finally, (vi) falls afoul of the contingent apriori.

Quine's "Truth by Convention"

The theory under attack holds that logical truths are true by convention, and so are analytic, apriori, and necessary. Let us begin by taking the language L under discussion to be a first-order language with an infinite set LT of sentences true in all models of L. Quine observes that speakers cannot have adopted a separate convention for each member of LT. Rather, the proponent of the linguistic theory must maintain, L-speakers have adopted a finite set of conventions from which the truth of every member of LT *follows*. But this won't do. If the linguistic theory of logic, apriority, and necessity must presuppose logical, apriori, and necessary consequence, then it can't explain them. The point can be illustrated an example.

1. All sentences of the form 'S or ~S' are true.
2. The sentence 'Los Angeles is in California or it isn't' is of the form 'S or ~S'
3. So, 'Los Angeles is in California or it isn't' is true.

Suppose that (1) is the statement of a convention, and so is true by stipulation. Since (2) is obviously true, (3) must also be true, where the sentence mentioned is a logical truth. Imagine that every other logical truth is similarly treated. Although this is supposed to establish the analyticity/apriority/necessity of all such logical truths, it doesn't.

First notice a problem Quine doesn't mention. Whether or not (1) was used initially to stipulate a convention, for the argument to explain agents' knowledge of (3), agents must know *that all sentences of the form 'S or S' are true* by virtue of knowing *that linguistic convention stipulates that they are*. Surely, this is *aposteriori knowledge* of an empirical fact about the linguistic community. So we have a problem at the outset. The problem on which Quine does focus is that to derive (3) one must recognize the inference from (1) and (2) to (3) to be truth preserving, which requires knowing *that if all F's are G, and a is an F, then a is G*. Since this logical knowledge is required in order for one to come to know (3) on the basis of (1) and (2), *any appeal to knowledge of linguistic conventions to explain (apriori) logical knowledge will presuppose (apriori) logical knowledge that is not explained by knowledge of those conventions*. It is no good to object that one can derive (3) from (1) and (2) without thinking to oneself "*if all F's are G, and a is an F, then a is G.*" This makes no difference. If there was *nothing* behind A's moving from (1) and (2) to (3), other than a blind process *insufficient to credit A with knowledge of the rule,* we wouldn't credit A's accepting (3) as showing that A *knew it* by virtue of knowing (1) and (2). So, the critique stands.

Quine's second objection is that since logical words are needed to state the conventions, some logical words must get their meanings independently, and some logical truths must not be true by convention. To the objection that we can be guided by conventions that are never formulated, he replies that an explanation that appeals to *truth by unstated conventions* is empty without a compelling story of what such conventions amount to.

> "In dropping the attributes of deliberateness and explicitness from the notion of linguistic convention we risk depriving the latter of any explanatory force and reducing it to an idle label. We may wonder what one adds to the bare statement that the truths of logic and mathematics are *a priori*, or to the still barer behavioristic statement that they are firmly accepted, when he characterizes them as true by convention in such a sense."[4]

---

[4] Quine (1936) at pp. 99 of Quine (1966).

To this we add that *if* the imagined conventions by which words are introduced rely on prior beliefs and intentions to guide later linguistic behavior, then presumably those attitudes will have negative, disjunctive, and quantified contents. This confronts the defender of the linguistic theory of the apriori with a dilemma. To insist that *none* of the propositions toward which prelinguistic agents are capable of having attitudes are knowable apriori – and hence that none is identical with the (apriori) proposition that *either o is red or o isn't red* – would be to cast doubt on the idea that such agents have contentful attitudes at all (thereby undermining the account of how unstated conventions arise). To admit that such agents *do* grasp propositional contents that are knowable apriori is to invite questions to which the linguistic theory has no answers.

(i)   How, if the truth of a sentence depends on the truth of what it expresses, can introducing sentences expressing contents one can already entertain possibly enable one to know apriori things one could not know apriori before they were linguistically expressed?

(ii)   Why suppose that prelinguistic agents who already have the concepts of negation, disjunction, and quantification introduce words for them by stipulating that 'not', 'or', and 'all' will mean whatever the must in order to make certain sentences true, rather than by simply resolving to use them to express the concepts they already posses?

Quine's Battle Against Quantified Modal Logic

Although Quine was critical of the linguistic theory of the *apriori*, he shared two presuppositions of its proponents – that necessity is apriority and that both are defensible only if they are reducible to analyticity. Having rejected the reduction of apriority to analyticity, he concluded that there is no apriori knowledge and no necessary truths. But his attack on the later still hadn't been made explicit. When it came it proceeded in two stages. The first stage – in Quine (1943, 1947, 1953a,b) -- attacked quantified modal logic as developed in the forties by Ruth Marcus and Rudolf Carnap. The second, in Quine (1951b), was directed at analyticity/ necessity itself. In this section, I will discuss the first stage of the attack.

At this stage Quine was willing (for the sake of argument) to take the notion *analyticity* for granted -- defined as a sentence that can be turned into a logical truth by replacing synonyms with synonyms. Interpreting necessity as analyticity, taking S to be possible iff its negation isn't

7

analytic, and assuming logicism, he could make sense of claims like (4) and (5), which are instances of the first-grade of modal involvement.

4.   '9 is an odd number' is necessary.
5.   'The number of planets is even' is possible.

The second grade is illustrated by (6) and (7).

6.   It is necessary that 9 is an odd number.
7.   It is possible that the number of planets is even.

Quine's strategy was to reduce this grade of modal involvement to the first. When modal operators aren't iterated, the truth conditions of (6) and (7) are those of (4) and (5). When they are, he assigns sentences to a hierarchy, depending on the number of modal operators embedded under such operators. Each level is governed by a definition of logical truth and analyticity, with the truth conditions of ⌈It is necessary that S⌉ (which is of level n+1, when S is of level n) being given in terms of the definitions of logical truth and analyticity at level n.[5]

At the third level of modal involvement '□' and '◊' are operators (expressing necessity and possibility) that can be prefixed to open formulas, allowing quantifying in.

8.   ∃x □ x is an odd number.

9.   ∃x (x is the number of planets & ◊ x is even)

Even if we understand analyticity and objectual quantification, this doesn't guarantee that we can assign intelligible truth conditions to sentences like these. If necessity is analyticity, it is a property of *sentences*. To make sense of (8) and (9) we must decide whether *an open formula relative to an assignment of an object to a variable* is *a logically true sentence or one that can be turned into a logical truth by replacing synonyms with synonyms*. Since open formulas are *not* sentences and variables relative to assignments are *not* terms with meanings or definitions, it is puzzling what the truth conditions of (8) and (9) are supposed to be. Quine did *not* foist this puzzle on the modal logicians of his time; they brought it on themselves by sharing the

---

[5] Section 2 of Quine (1947).

identification of necessity with analyticity that generated it. Quine was right to insist that if quantified modal logic was to progress, it had to solve this puzzle or give up that identification.

From here, he developed two lines of argument that neither he nor his opponents consistently distinguished. One was that, if necessity is analyticity, there is no way of solving this puzzle. The other was that, the interpretation of necessity aside, quantified modal logic violates fundamental logical and semantic principles, and so must be rejected.

The second, more ambitious, attack depends on A1, which is true, A2 and A3, which are false, plus the definitions D1-D3.[6]

A1.   The modal operators '□' and '◊' are referentially opaque.

A2.   Occurrences of objectual variables in the scope of referentially opaque operators are not purely referential.

A3.   Bindable occurrences of objectual variables must be purely referential.

D1.   An occurrence of a term in a formula or sentence S is *purely referential* iff what it contributes to the truth or falsity of S (relative to an assignment) is simply what it designates or denotes (relative to the assignment).

D2.   A position in S is *referentially transparent* iff for any pair of terms t and t*, the results S(t) and S(t*) of substituting these terms into that position will have the same truth values (relative to an appropriate assignment) iff ⌈t = t*⌉ is true (relative to that assignment). A position is *referentially opaque* iff it is not referentially transparent.

D3.   A sentential operator is *referentially transparent* iff any referentially transparent position in a sentence remains so when the operator is prefixed to the sentence. A sentential operator is *referentially opaque* iff it is not referentially transparent.

The idea behind A2 is this: Let 'O' be a referentially opaque operator, let ⌈O F(x)⌉ be a formula in which 'x' occurs free (in position p), and let ⌈O F(t)⌉ and ⌈O F(t*)⌉ be sentences that differ in truth value, which arise from substituting distinct terms t and t* designating the same object o for 'x' (at p). (There must be such terms if 'O' is referentially opaque.) The truth value of ⌈O F(x)⌉ relative to an assignment A of o to 'x' differs from the truth value of *one* these two sentences even though ⌈t = t* = x⌉ is true relative A. Suppose ⌈O F(t*)⌉ differs in truth value

---

[6] The explication of this argument given below is of the reasoning, implicit and explicit, in Quine (1943, 1953b). The category of "terms" in D1, D2 includes definite descriptions (whether Fregean or Russellian).

from ⌈O F(x)⌉ (relative to A). Then, Quine concludes, occurrences of t in the former and 'x' in the latter both fail to be purely referential, verifying A2. As noted in section III of Kaplan (1986) and Kazmi (1987), this argument is fallacious. From the fact that ⌈O F(t*)⌉ differs in truth value from ⌈O F(x)⌉, we can conclude that either the occurrence of t* in the former or the occurrence of 'x' in the latter is not *purely referential*, but we cannot conclude that the occurrence of 'x' isn't. Also, one can construct opaque operators, as Kaplan does in sections IV, VII, and VIII-XIII, for which occurrences of variables in their scope *are* purely referential, and bindable from outside by objectual quantifiers. So A2 is false.[7]

Kaplan introduces the notion of the *valuated sentence* associated with F(x) relative to an assignment of o to 'x'. It is what one gets by substituting o for *'x'* in the syntactic structure F(x). Given this, one can define referentially opaque operators that allow quantifying in and have extensions that include both ordinary and valuated sentences. For example, we might define an operator O_1 that maps an ordinary sentence S onto truth iff Ralph utters S, while mapping a valuated sentence VS onto truth iff he utters any complete sentence that results from replacing a occurrence of an object o in VS with an occurrence of any proper name of o. So understood, occurrences of variables under O_1 are *purely referential*, and the standard "law" (10) of quantification theory is retained. ('F' is used as a schematic letter in (10).)

10. $\forall x,y [x = y \supset (O (Fx) \supset O (Fy))]$

There is, however, nothing in the nature of quantification that requires (10) to be true. Let a *finely valuated sentence* be just like a valuated sentence except that instead of replacing 'x' with o, we replace 'x' with <'x',o>. Now we stipulate that O_2 maps a finely valuated

---

[7] Quine (1947) uses different reasoning in attempting to establish A2. Quine notes that if O is a referentially opaque operator, there will be truths ⌈t = t* & O (S(t) & ~ O S(t*)⌉. If t and t* occupy positions open to objectual quantification, and if existential generalization is universally truth preserving, then ⌈∃x ∃y (x = y & Sx & ~Sy)⌉ must also be true. Since this violates "the law of the substitutivity of identity for variables" and requires some occurrences of variables to be non-purely referential, he thinks this is impossible. Below, I argue that Quine is wrong about this. The other flaw is his incorrect assumption that existential generalization is fundamental to objectual quantification. Although it is always truth preserving in certain contexts, it fails to be so in others.

sentence FVS onto truth iff Ralph utters any complete sentence that results from replacing all occurrences of each variable/object pair <v,o> in FVS with occurrences of a proper name of o, *provided that different occurrences of the same pair are replaced by occurrences of the same name*. Quantification into contexts governed by O_2 is as intelligible as quantification into contexts governed by O_1, even though (10) fails with O_2. So (10) *isn't* really a law of quantification, and bindable occurrences of variables need not be purely referential.

To understand this one must *not* confuse schema (11a) with the indiscernibility principle that may be formulated by (11b) or (11c).[8]

11a. $\forall x, y \ [x = y \supset (S(x) \supset S(y))]$
   b. $\forall x, y \ (x = y \supset$ every property of x is a property of y)
   c. $\forall x, y \ [x = y \supset \forall P(Px \supset Py)]$

The instance of (11a) that arises from replacing '$S(x) \supset S(y)$' with '$O_2(x \neq y) \supset O_2(y \neq y)$' is false, if Ralph has uttered 'Hesperus ≠ Phosphorus' but not ⌈n≠n⌉ for any name designating Venus. This is consistent with the truth of (11b) and (11c), since the property Venus must have iff '$O_2(x \neq y)$' is true (relative to an assignment A of Venus to 'x', 'y') is *being designated by some pair of names t_1 and t_2 such that Ralph utters ⌈$t_1 \neq t_2$⌉*, while the property Venus must have iff '$O_2(x \neq x)$' is true (relative to A) is *being designated by some name t such that Ralph utters ⌈$t \neq t$⌉* In short, the failure of Quine's principle A3 does not threaten the indiscernibility of identicals.[9]

However, the failure of this Quinean argument – that quantifying into referentially opaque constructions violates fundamental semantic and logical principles -- doesn't resolve his worries about the quantified modal logic of his day. To do that, one must make positive sense of quantifying into modal contexts *when necessity is identified with analyticity*. Quine argues, in (1947, 1953b), that this is impossible because the truth conditions of sentences of the third grade

---

[8] See Kazmi (1992).

[9] Two interesting analyses of propositional attitude verbs that lead to violations of (11a) are Mark Richard (1987) and Kit Fine (2007). These are critically discussed in Soames (1987, 2012) and chapter 7 of Soames (2002).

of modal involvement can't be specified in terms of the truth conditions of those of the second grade. As he notes, it is natural, to appeal to (i) and (ii) in attempting to do so.

(i) $\exists x$ ...x... is true only if ...a... is true for some term a.
(ii) $\exists x$ ...x... is true if ...a... is true for some term a.

Principle (i) is problematic because there will typically be no guarantee that unnamed, or even unnamable, objects might be the only ones underwriting the truth of an existence claim. Principle (ii) is also problematic. Suppose there are two names, a and b, such that (12a) and (13a) are both true. Then, by (ii), (12b) and (13b) must also be true.

12a.   a = b & □ aRa
   b.   ∃x [x = b & □ xRa]

13a.   b = b C & ~ □ bRa
   b.   ∃x [x = b & ~ □ xRa]

But, since (12b) and (13b) are contraries, they can't both be true. So, in order to prevent (ii) from being falsified, one must restrict the terms used to specify the truth conditions of quantified sentences to members of a class T of terms coreferential members of which are *analytically equivalent.*

(iii)   If a and b are members of T, then ⌈a = b⌉ is analytic if true, and substitution of one for the other in any analytic sentence preserves analyticity.

If this restriction is observed, (12a) and (13a) can't be jointly true, which in turn will block the erroneous characterization of (12b) and (13b) as jointly true.

However, to adopt (iii) as the means of specifying the truth conditions of third-grade modal sentences in terms of second-grade sentences requires one to drastically limit the domain of objects and the class of terms designating them. As cases involving 'Hesperus'/'Phosphorus' and 'Cicero'/'Tully' illustrate, ordinary proper names of empirically given objects must be excluded. That's not all. As Quine notes in (1947, 1953b), the severity of the needed restrictions would undercut any significant philosophical interest in quantified modal logic. Nor does there seem to be another way of specifying truth conditions of the third grade in terms of those of the

second. Quine was right: *if necessity is nothing more than analyticity,* then quantified modal logic is of little interest. His error was in taking it for granted, along with most of those against whom he argued, that if there is such a thing as necessity, it must be analyticity.[10]

Analyticity, Necessity, and Meaning

"Two Dogmas of Empiricism" (1951), which was among the most widely influential works in philosophy of its time, changed the self-conception of analytic philosophy in two ways. By undermining the analytic/synthetic distinction, as then understood, it decisively challenged the picture of philosophy as conceptual analysis; by embracing a holistic view of empirical confirmation, it drove the final nail in the coffin of the logical empiricists' verificationist criterion of meaning. Though in retrospect Quine's moves were simple, they were essential to freeing philosophy from the once liberating but by then confining "linguistic turn."

The first phase of the attack on analyticity is the circle argument in sections 1-4 of "Two Dogmas." A sentence is characterized as analytic iff it is a logical truth or can be turned into one by putting synonyms for synonyms. Synonymy is intersubstitutivity that always preserves truth value. In what environments? The answer, in contemporary terms, is that for A and B to be synonymous, substitution of one for the other must preserve truth value in *intensional* constructions (though not necessarily in hyperintensional ones). These constructions are identified with those in the scope of a modal operator. But this requires an independent conception of necessity that Quine's targets didn't have. Rather, they insisted, it is only by explicating necessity as analyticity that the former can be made defensible. Quine agreed, while insisting that the proposed explication was worthless, because we can't explain analyticity without presupposing necessity. Far from vindicating necessity, he argued, the logical empiricists' treasured reduction infused analyticity with necessity's fatal defects.

---

[10] The discussion in this section has profited greatly from the contributions of Ali Kazmi.

13

Although Quine's attack succeeded against its intended targets, today we realize that analyticity and necessity need not be yoked together. His failure to see what others of his time also failed to see was connected to his rejection in "Two Dogmas" of synonymy, and his later rejection, in Quine (1960) of hyperintensionality. As for analyticity, the jury is still out. Despite the trenchant criticism in Williamson (2007) of recent attempts to rehabilitate epistemic conceptions of analyticity, it is not obvious that no such conception can be sustained. What is clear is that no conception of analyticity with the broad philosophical significance accorded to it by Quine's opponents will ever be forthcoming. On this crucial matter, he was right.

After giving the circle argument, Quine devotes the last two sections of "Two Dogmas" to improving the logical empiricists' faulty conception of meaning. His second "dogma" was the view that every meaningful sentence S is associated with sets C and D of observational claims such that the truth of any member of C would add to the degree to which S is confirmed, while the truth of any member of D would add to the degree to which S is disconfirmed. This dogma is connected to the dogma that there is an analytic/synthetic distinction by a conception that identifies the meaning of a synthetic sentence with the sets of experiences that would confirm/disconfirm it. As Quine observes, if S1 and S2 are confirmed/disconfirmed by the same experiences, they are synonymous and their biconditional will be analytic. So one who rejects analyticity (and synonymy) must reject the second dogma too.

Quine's reason for so doing is rooted in the Duhemian idea that what counts as confirmation or disconfirmation of a hypothesis H depends on the background assumptions we hold fixed in testing H. Because he thinks we often have a wide range of choice in deciding which background assumptions to appeal to, and which to give up when they plus H entail a falsehood, he rejects the idea that H's meaning determines the evidence that would confirm or

disconfirm H. If that idea were correct, the meaning of H would underwrite analytic truths specifying which experiences would confirm, or disconfirm, H. But if understanding H were sufficient to determine when it was confirmed, or disconfirmed, we wouldn't have the range of theoretical choice of when to hold onto H, and when not, that we know we do have.

To this plausible criticism of Carnap, Quine adds two more dubious claims: (i) that only entire theories, rather than individual hypotheses, are confirmed or disconfirmed by empirical evidence, (ii) that the meaning of a theory is the totality of empirical evidence that would confirm or disconfirm it. In short, he agreed with logical empiricists that meaning (empirical significance) is verifying or falsifying experience, while insisting that "the unit [of meaning] accountable to an empiricist critique" is not the individual sentence but "the whole of science."[11]

In the final section of "Two Dogmas" Quine sketches the following theses of this holistic version of logical empiricism.

QT1. Holistic Verificationism
    a. The meaning of a theory = the class of possible observations it fits
    b. Two theories have the same meaning iff they fit the same class of possible observations.

QT2. The totality of our beliefs is a "man-made fabric which impinges on experience only along the edges."

QT3. Any statement can be held true come what may (by making adjustments elsewhere).

QT4. Any statement can be rejected, or held to be false (by making adjustments elsewhere). Thus, no statement is immune from revision.

QT5. Underdetermination
For any consistent theory T_1, and class of possible observations O that fit it, there is a theory T_2 incompatible with T_1 which also fits O.

In QT1 and QT5, we take the class of possible observations a theory fits to be the class of observational conditionals, $\lceil O_1 \supset O_2 \rceil$ it entails. (O_1 and O_2 specify observable events.) For the theory to be true, it is necessary that all these conditionals be true. For two theories to mean the same thing, it is necessary and sufficient that they entail the same observational conditionals.

One who accepts this view can identify mistakes made by defenders of the analytic/ synthetic distinction. For Quine, their mistakes were to have accepted (14a-c) instead of (15a-c).

14a.　Experience is relevant to the confirmation of individual synthetic, but not analytic, sentences.

   b.　Analytic sentences can, without error or change of meaning, be held true in the face of any experience. Synthetic sentences cannot be.

   c.　Analytic sentences cannot be rejected without error, unless we change what we mean by them. Synthetic sentences can be so rejected.

15a.　Experience is not relevant to the confirmation of individual nonobservation statements, taken in isolation. It is relevant to their confirmation taken in their role as contributing to our total theory of the world.

   b.　Any nonobservation sentence, can, without error, be held true in the face of any experience – by making compensatory changes elsewhere in one's theory.

   c.　Insofar as it makes sense to talk of the meanings of individual sentences at all, changes in one's total theory (which involve changes in which sentences one accepts and which one rejects) should be seen as implicitly changing the meanings of all one's sentences.

Although Quine takes no sentences to be immune from revision in light of experience, he does recognize that the degree to which we are ready to revise them varies from sentence to sentence, depending on how central they are to our conceptual framework. There is, he thinks, a continuum on which sentences that have traditionally been characterized as analytic typically occur at one end while those that have been characterized as synthetic often occur at the other.

One cannot, in evaluating Quine's holistic verificationism, avoid the role played by the underspecified wild card, observation, in his system. In Quine (1948, 1951b) he speaks of "sense experiences" as observational touchstones of theories. This is what underlies his perversely even-handed comparison of phenomenalistic vs. physicalistic ontologies, including his characterization of physical objects as mythic intermediaries, "comparable epistemologically to the gods of Homer,"[12] that are imported into theories as aids in predicting future experience in light of past experience. As I have argued elsewhere, to combine holistic verification with this

---

[11] Quine (1951b), p. 42 of the 1980 reprinting of Quine (1953a).

[12] P. 44 of the 1980 reprinting of Quine (1951b).

conception of observation is to reproduce some of the worst absurdities of the earlier empiricist systems.[13] Fortunately, his sensory conception of evidence was temporary, to be dropped in later years along with the pseudo-profundity of his earlier talk about myths and the gods of Homer.

Nevertheless, observation remained a problem. If, for the holistic verificationist, statements reporting the contents of ordinary, unaided observations can play the role of data statements that give empirical content to theories, how far should we go? Do observations using magnifying glasses count? How about binoculars, telescopes, microscopes, radar, electron microscopes, ratio telescopes, and cat scans? The more we include in the observational base of theories, the less radical, but also less interesting, holistic verificationism becomes. Is there a principled way to draw the line between the observational and the non-observational that would render even a weakened holistic verificationism plausible? It will, with some justice, be objected that Quine himself was no friend of a sharp distinction between the two. But this is less a defense of holistic verificationism than a recognition that no definite thesis about meaning and verification can be extracted from his discussion. Rather than attempting to extract such a view, one might do better by rejecting his gestures in this direction as too imprecise and global to have a chance of success. We simply have no clear idea what a comparison of two theories that differ radically in their ideologies might amount to.

Another reason to be suspicious of holistic verificationism is that it leads to paradox. Suppose, according to QT1 and QT5, that a certain consistent theory T_1 means the same as T_2 while being logically incompatible with T_2. Since two theories that mean the same thing must make the same claim about the world, they must agree in truth value.

(i)   If two theories mean the same thing, then they make the same claim about the world, in which case they cannot differ in truth value. Hence one is true if and only if the other is.

---

[13] Pp. 398-405 of Soames (2003a).

It follows that either T_1 and T_2 are both true or both false. Since they are logically incompatible, they can't both be true. So, they both must be false. But surely there are *some* true theories of *some subject matters*.

(ii). Some theories of some subject matters are true.

Given this, we may simply select some *true* theory T_1, and run the argument again. But now we get the result that T_1 and T_2 must *both* be true, because they have the same meaning, while also getting the result that they *can't* both be true, because they are logically incompatible. So either holistic verificationism or underdetermination, or both, must be rejected. Although this conclusion is unassailable, it is not clear whether it shows Quine's view to be irretrievably wrong or merely to need of some modification. Quine (1975) discusses this difficulty. But the best defense of his view is the modification proposed in Harman (1979), which is critically discussed in Soames (2003a).

On What There is: Quine, Carnap, and Ontology

The connection between Quine's ontology and his theory of meaning is more fully displayed in Quine (1948), the first part of which sets out his criterion of ontological commitment. One is not, he argues, committed, merely by using a name, to the existence of something named. Nor is one committed, in using any meaningful term, to their being some thing it means. It is a substantive theoretical position, which Quine rejects, that words are meaningful only if there exist entities they mean. In using the predicate 'is red' or the adjective 'seven', one is not thereby committed to the existence of colors or numbers, though one is committed when one says that there exist primary colors, or prime numbers between 6 and 12. One is committed to the existence of *so-and-so's* only when one says there exist *so-and-so's*.

That is the idea behind the slogan, "To be is to be the value of a bound variable." The point is *not* that to exist amounts to being the value of a variable, but that to *commit* oneself to the existence of something is *to say* that there exists such a thing. To commit oneself to the

existence of $\mathcal{F}$s is to say something the proper regimentation of which is, or entails, $\ulcorner \exists x\ Fx \urcorner$ --

the truth of which requires the existence of an object o that makes $\ulcorner Fx \urcorner$ true when o is assigned

as value of 'x'. The qualification about regimentation is a crucial to avoid unwanted

commitments. Quine has no problem saying that *there is a possibility that S*, without thereby

committing himself to the existence of possibilities.[14] The justification of his nonchalance is

that proper regimentation of the remark involves no quantification over possibilities, but simply

recognition that *it may be true that S*. Using the flexibility provided by such regimentation, he

holds that the *only* way to commit oneself to the existence of $\mathcal{F}$s is by asserting something the

proper regimentation of which entails the existentially quantified claim that there exist $\mathcal{F}$s.[15]

Quine puts this idea to use in discussing abstract objects.

"When we say that some zoological species are cross-fertile we are committing ourselves to recognizing as entities the several species themselves, *abstract though they are*. We remain so committed at least until we devise some way of so paraphrasing the statement as to show that seeming reference to species…was an avoidable manner of speaking…Classical mathematics …is up to its neck in commitments to an ontology of abstract entities. Thus it is that the great mediaeval controversy over universals has flared up anew in the modern philosophy of mathematics…The three main mediaeval points of view regarding universals are designated by the historians as *realism, conceptualism,* and *nominalism*. Essentially these same three doctrines reappear in twentieth-century surveys of the philosophy of mathematics under the new names *logicism, intuitionism,* and *formalism*. *Realism*…is the Platonic doctrine that universals or abstract entities have being independently of the mind; the mind may discover them but cannot create them. *Logicism*, represented by Frege, Russell, Whitehead, Church, and Carnap, condones the use of bound variables to refer to abstract entities known and unknown, specifiable, and unspecifiable."[16]

Quine suggests that Carnap's commitment to numbers is a form of Platonism. Though the label

may seem apt, Carnap, who had long repudiated metaphysics as meaningless nonsense, resented

having it applied to him. He devoted Carnap (1950) to explaining why the charge is unfair.

---

[14] In this example 'S' is used as a schematic letter.

[15] P. 12 of the 1980 reprinting of Quine (1948).

[16] Ibid., pp. 13-14.

His key thesis was that ontological questions are intelligible only within a framework for describing the world. Such a framework is a formalizable language with semantic rules interpreting its expressions and assigning truth conditions to its sentences. Ordinary English contains terms for observable physical objects and events. Carnap assumes that rules constituting their meanings specify possible observations that would confirm or disconfirm sentences containing them. So, he thinks, whether or not there are things of a given sort reduces to whether or not observable events occur that, as a matter of linguistic rule, confirm the relevant sentences. Since these *internal questions* can be answered by evidence, they aren't metaphysical. He contrasts *internal questions* with *external questions*, which can't be settled by evidence, but nevertheless purport to be about the world. Traditional metaphysical questions about the reality of the external world are of this sort. ⌈Are there Fs?⌉ is properly understood to be an internal question, resolvable by empirical evidence of the kind given by the semantic rule governing F. However, philosophers have traditionally misunderstood the question as not being settled by such evidence. Their mistake has been to divorce the application of F from the linguistic rules that constitute its meaning. In this way, they have been led to ask cognitively meaningless pseudo-questions that can't be answered.

This mistake is compounded by another one that disguises it. Philosophers are prone to run together the proper, though often trivial, internal *theoretical* question ⌈Are there Fs?⌉ with the non-trivial *practical* question of whether to adopt a theoretical framework incorporating F. Regarding physical objects, Carnap says:

> "Those who raise the question of the reality of the thing world itself have perhaps in mind not a theoretical question…but rather a practical question, a matter of a practical decision concerning the structure of our language…[W]e are free to choose to continue using the thing language or not; in the latter case we could restrict ourselves to a language of sense-data and other "phenomenal" entities…If someone decides to accept the thing language, there is no objection against saying that he has accepted the world of things. But this must

not be interpreted as if it meant his acceptance of a *belief* in the reality of the thing world; there is no such belief or assertion or assumption, because it is not a theoretical question.**"**[17]

We are asked to imagine a choice between our ordinary physical-object framework and a Berkeleyan alternative that speaks only of minds and "sense data." This, we are told, is *simply* a choice between two linguistic schemes for describing experience. There is, we are assured, no belief, assertion, or assumption *in the reality of the thing world* that one adopts when one opts for the physical, rather than the phenomenal, framework. If there were, what would it be? Not an unverifiable and unfalsifiable pseudo-statement, since they lack cognitive content. It would have to be an empirical statement of some sort. But then the assertion, belief, or assumption would require empirical justification -- in which case the choice between frameworks would be genuinely *theoretical*, rather than the purely *practical* decision Carnap takes it to be.

From here it is a short step to the conclusion that the cognitive contents of empirically equivalent theories stated in the two languages are the same. Since they have the same content, there is no fact on which they differ, and no claim about the world made by one of them that isn't made by the other. This is why Carnap insists that the choice between the two theories is "not cognitive in nature," but to be made solely on practical grounds.[18] We are justified in adopting the physicalistic theory because (i) we find it more efficient to use than the phenomenalistic one, and (ii) it doesn't make any contentious claims about the world beyond those made by the phenomenalistic theory.

Ontological questions about abstract objects are treated similarly. When F is a predicate applying to physical objects or events, Carnap takes its meaning to supply analytic truths specifying empirical evidence that would confirm or disconfirm statements containing F. The internal question ⌈Are there Fs?⌉ is answered by gathering this evidence, while the external

---

[17] Carnap (1950) at pp. 207-208 of its 1956 reprinting, my emphasis.

[18] Ibid. p. 208.

ontological question is dismissed as meaningless. When F is a predicate of abstract objects, empirical evidence is often irrelevant, and the meaning of F is given by rules specifying logical properties of sentences containing it. In these cases, the answer to the internal question ⌈Are there Fs?⌉ is analytic, while the external question is meaningless. Since, ⌈There are numbers⌉ is analytic, it makes no claim about the world, and so cannot be metaphysical.

That, in a nutshell, was Carnap's response to Quine. Quine's reply to Carnap was the rejection of analyticity in "Two Dogmas." Although this was fine as far as it went, it didn't go far enough. Since at this stage Quine was a holistic verificationist, he agreed with Carnap that it makes no difference to the empirical contents of whole theories, and hence to their truth or falsity, how they differ on *any non-observational statements*, so long as their observational consequences coincide. Hence, it makes no difference what their ontologies are. Theories that "posit" numbers, sets, physical objects, propositions, and properties do not differ in content from theories that don't, *as long as the theories are observationally equivalent*. This is the counterintuitive bedrock of agreement between Carnap and Quine.

The contrast between physicalist and phenomenalist ontologies is a case in point. For Carnap, physicalist and phenomenalist theories compatible with the same sense experience have the same content, and so make the same claims about the world. So the choice between them must be made on practical grounds. Quine agrees.

"As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience. Physical objects are conceptually imported into the situation as convenient intermediaries…as irreducible posits comparable, epistemologically, to the gods of Homer.…[I]n point of epistemological footing the physical objects and the gods differ only in degree and not in kind. The myth of physical objects is epistemologically superior to most in that it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience."[19]

---

[19] Quine (1951b), p. 44 of the 1980 reprinting.

Quine stresses that the "myth of physical objects," though useful, is not indispensable for making predictions about sense experience. The same predictions could, in principle, be made by a phenomenalistic theory. He makes this point with an analogy in which the phenomenalistic theory of nature is said to stand to the physicalistic theory as the algebra of the rational numbers stands to the algebra of the reals. He notes that in the algebra of the rationals, functions like square root sometimes go undefined, complicating the laws. "Then," Quine says:

> "it is discovered that the rules of our algebra can be much simplified by conceptually augmenting our ontology with some *mythical entities*, to be called irrational numbers. *All we continue to be really interested in, first and last, are rational numbers*; but we find that we can commonly get from one law about rational numbers to another much more quickly and simply by *pretending* that the irrational numbers are there too. Now I suggest that experience is analogous to the rational numbers and that the physical objects, in analogy to the irrational numbers, are posits which serve *merely* to simplify our treatment of experience…The salient differences between the positing of physical objects and the positing of irrational numbers are, I think, *just two*. First the factor of *simplification* is more overwhelming in the case of physical objects than in the numerical case. Second, the positing of physical objects is far more archaic, being indeed coeval, I expect, with language itself." (my emphasis)[20]

Quine's point is (i) that the phenomenalistic theory tells us the whole truth and nothing but the truth about nature, (ii) that what it talks about are "all that we are really interested in first, and last;" and (iii) that since the physical theory adds nothing new about the world, the *only* reason to prefer it is that it makes the needed predictions about sense experience more simply and conveniently. Carnap couldn't have said it better.

The Road to Word and Object

The view that emerged from "Two Dogmas" was unstable, not because of anything inherent in the rejection of analyticity, but because of the rejection of synonymy to which

---

[20] Ibid., 41-2 of the *original 1951 version* of "Two Dogmas" printed in *The Philosophical Review*. Quine tells us, in the section "On the Origins of These Essays" in Quine (1953a), that the passage was deleted from later reprintings because it overlapped with a passage from "On What There Is" that also appears in the collection. This is unfortunate, since though there is such an overlap, the original passage from "Two Dogmas" is more revealing. For further discussion, see pp. 401-403 of Soames (2003a).

Quine's argument committed him. Since synonymy is sameness of meaning, it is hard to reject it without rejecting meaning itself. And if meaning is rejected, can reference be far behind? Surely, it was objected -- in Grice-Strawson (1956), and in Carnap (1955) -- we can't do without meaning and reference. No matter that the notion of synonymy rejected by Quine in "Two Dogmas" was an artificial one, according to which for expressions to be synonymous was for them to be substitutable *salva veritate* in modal (rather than hyperintensional) constructions. The fact that this isn't the ordinary notion of sameness of meaning didn't prevent the parties from writing as if meaning in the ordinary sense, and perhaps reference too, would be vulnerable, if the notion of synonymy that Quine rejected were to go. Since, in the end, this road led to *Word and Object* and "Ontological Relativity" the objectors were vindicated.

The argument in Grice-Strawson (1956) is that it is absurd to reject synonymy, because doing so would require rejecting meaning altogether, which was taken to be obviously untenable. If expressions can be meaningful at all, it was argued, there must be true answers to the question *What does this, or that, expression mean?* But if there are such answers, we can identify synonymous expressions as those for which the answers are the same.

"To say that two expressions x and y are cognitively synonymous seems to correspond…to what we should ordinarily express by saying…that x means the same as y. If Quine is to be consistent…then it appears that he must maintain not only that the distinction we suppose ourselves to be marking by the use of the terms "analytic" and "synthetic" does not exist, but also that the distinction we suppose ourselves to be marking by the use of the expressions "means the same as," and "does not mean the same as" does not exist either...Yet the denial that the distinction…really exists, is extremely paradoxical...For we frequently talk of the presence or absence of relations of synonymy between kinds of expressions … where there does not appear to be any obvious substitute for the ordinary notion of synonymy… *Is all such talk meaningless? Is all talk of correct or incorrect translation of sentences of one language into sentences of another meaningless?* It is hard to believe that it is… *If talk of sentence-synonymy is meaningless, then it seems that talk of sentences having a meaning at all must be meaningless too. For if it made sense to talk of a sentence having a meaning, or meaning something, then presumably it would make sense to ask "What does it mean?" and if it made sense to ask "What does it mean?" of a sentence, then sentence-synonymy could be roughly defined as follows: Two sentences are*

synonymous if and only if any true answer to the question "What does it mean?" asked of one of them, is a true answer to the same question, asked of the other."(my emphasis)[21]

This powerful argument – that one can give up synonymy only if one gives up meaning and translation too – turned out to be prophetic. Fours years later, this is what Quine proposed to do.

Carnap also turned out to be prophetic. In Carnap (1955), he tried to show that the meaning of a term, over and above its reference, can play a role in empirical theorizing about language users. He argued that although there are empirical uncertainties in establishing the meaning and reference of a term, there are sound empirical methods for bringing evidence to bear on both questions. As a result, he concluded, meaning and reference are in the same boat. Thus, he thought, Quine was wrong to dismiss intensional notions like meaning and synonymy, while apparently retaining extensional ones like reference (and truth). For Carnap, the intensional and the extensional were scientifically on par – both required and both respectable. Quine's response came in *Word and Object*, which he dedicated to Carnap. There, he agreed that meaning and reference are on a par, but upped the ante by (in effect) rejecting both.

Meaning and Translation

In *Word and Object* Quine argues that *meaning*, as ordinarily understood, plays no empirical fact-stating role. He uses theories of translation between languages spoken by different communities to illustrate his point. Although translation is a useful practical activity, we cannot, he argues, take it's claims about meaning to be genuine truths. Assuming that no other empirical theory can provide such truths, he concludes that the ordinary notion of *meaning* has no role to play in serious descriptions of the world.[22]

His two main theses are:

---

[21] Grice-Strawson (1956), 141-158, at pp. 145-146.

[22] Quine (1960), p. 221.

<u>Underdetermination of Translation by Data</u>
Let L_1 and L_2 be arbitrary languages, and D be the set of all observational truths (known and unknown) relevant to translation from L_1 to L_2. For any theory of translation T for L_1, L_2, compatible with D, there is a theory T', incompatible with T, that is equally well supported by D.

<u>Indeterminacy of Translation</u>
Translation is not determined by the set N of all truths of nature, known and unknown. For any pair of languages and theory of translation T for those languages, there are alternative theories of translation, incompatible with T, that accord equally well with N. All such theories are equally true to the facts; there is no objective matter of fact on which they disagree, and no objective sense in which one is true and the other is not.

To reconstruct Quine's argument for these theses, we need some understanding of what theories of translation are, what evidence they are answerable to, and what empirical predictions they make. Quine understands them along roughly the following lines.

<u>Theories of Translation</u>
A theory of translation for two languages correlates individual words of each with words or phrases of the other; this correlation is then used to correlate the sentences of the two languages. Any system of establishing such correlations can be counted as a translation manual, or theory. We may take such a theory as yielding (infinitely many) theorems of the form:

Word or phrase w in L_1 means the same as word or phrase w* in L_2.
Sentence S in L_1 means the same as sentence S* in L_2.

<u>Empirical Data</u>
Empirical data relevant to theories of translation are statements about the behavior of language users. More specifically, they are statements about the *stimulus meanings* of certain classes of sentences.

<u>Stimulus Meaning</u>
The stimulus meaning of a sentence S (for a speaker at a time) is a pair of classes -- the class of situations in which the speaker would assent to S and the class of situations in which the speaker would dissent from S.

<u>Occasion Sentences</u>
S is an occasion sentence for a speaker x iff the x's assent to or dissent from S depends on what x is observing.

<u>Observation Sentences</u>
S is an observation sentence in a language L iff (i) S is an occasion sentence for speakers of L, and (ii) the stimulus meaning of S varies very little from one speaker of L to another. (Sameness of the stimulus meaning of such a sentence S across the members of a population is taken to be a reasonable approximation of the maximal degree to which assent to S, or dissent from S, is dependent on observation alone, free of contextually varying background assumptions.)

Empirical Predictions

(i) Translation of observation sentences must preserve stimulus meaning. If a translation theory states that an observation sentence S_1 in L_1 means the same as a sentence S_2 in L_2, then the theory predicts that S_1 and S_2 have the same stimulus meanings in their respective linguistic communities.

(ii) Translation must preserve the stimulus synonymy of pairs of occasion sentences. If S_1 and S_2 are occasion sentences of L_1, and if a translation theory states both that S_1 means the same in L_1 as P_1 in L_2 and that S_2 means the same in L_1 as P_2 in L_2, then the theory predicts that S_1 and S_2 have the same stimulus meaning for speakers of L_1 iff P_1 and P_2 have the same stimulus meaning for speakers of L_2.

(iii) Translations of truth functional operators -- *and* , *or, not* etc. -- have recognizable effects on stimulus meaning. For example, if a theory translates an expression n of a language L as meaning the same as *not* in English, then the theory predicts that adding n to sentences of L reverses stimulus meaning. Similar claims hold for other truth functional operators.

In addition, Quine sometimes talks as if these predictions might be augmented by a fourth kind defined in terms of his notions of *stimulus analyticity* and *stimulus contradiction*.

S is stimulus analytic in L iff virtually all speakers of L assent to S in all stimulus situations.

S is stimulus contradictory iff virtually all speakers of L dissent from S in all stimulus situations.

(iv) Translation must preserve stimulus analytic and stimulus contradictory sentences. If S_1 is a stimulus analytic (contradictory) sentence of L_1, and if a theory of translation states that S_1 means the same in L_1 as S_2 in L_2, then the theory predicts that S_1 is stimulus analytic (contradictory) iff S_2 is.

Because Quine conceives of (i)-(iv) as exhausting all possible evidence for theories of translation, he accepts the underdetermination thesis. In arguing for it, he imagines tying to translate some language L into English that has never previously been encountered by outsiders. (He calls this "radical translation.") Suppose the linguist can discover how questions in L, as well as assent and dissent to them, are expressed. The linguist might then discover that L-speakers will assent to the one-word sentence 'Gavagai?' in roughly the same situations in which an English speaker would assent to 'Rabbit?' This would support the hypothesis that the two sentences have the same stimulus meaning, and so are

intertranslatable. But, Quine thinks, once we go beyond observation sentences, the limitations on behavioral data will never provide an empirical basis capable of ruling out different incompatible theories of translation, each compatible with all the evidence.

His point is illustrated by the difficulty moving from translations of the *sentence* 'Gavagai!' by 'Rabbit!' to a translation of the *word* 'gavagai' by 'rabbit'. He observes:

> For, consider 'gavagai'. Who knows but what the objects to which this term applies are not rabbits after all, but mere stages, or brief temporal segments, of rabbits? In either event the stimulus situations that prompt assent to 'Gavagai' would be the same as for 'Rabbit'. Or perhaps the objects to which 'gavagai' applies are all and sundry undetached parts of rabbits; again the stimulus meanings would register no difference.[23]

> Does it seem that the imagined indecision between rabbits, stages of rabbits, integral parts of rabbits…must be due merely to some special fault in our formulation of stimulus meaning, and that it should be resoluble by a little supplementary pointing and questioning? Consider, then, how. Point to a rabbit and you have pointed to a stage of a rabbit, to an integral part of a rabbit...Nothing not distinguished in stimulus meaning itself is to be distinguished by pointing, unless the pointing is accompanied by questions of identity and diversity: 'Is this the same gavagai as that?', 'Do we have here one gavagai or two?'.[24]

For Quine, the question of whether 'gavagai' means the same as, and refers to (i.e., is true of) the same things as 'rabbit' -- as opposed to meaning the same as, and referring to the same things as, our expressions 'undetached spatial part of a rabbit' or 'temporal stage of a rabbit' -- depends on what expression of L, if any, means what 'is the same thing as' means in English. Evidence about stimulus meaning won't decide this question. Imagine 'squiggle' to be a word in L that is a candidate for expressing our notion of identity. Pointing at different spatial or temporal parts of the same rabbit we might ask "Gavagai (point) squiggle gavagai (point)?" with the thought that if the L-speaker assents, we will know that 'gavagai' *doesn't* mean what 'undetached spatial part of a rabbit' or 'temporal stage of a rabbit' mean in English, since the spatial or temporal parts pointed at are different. But, as Quine observes,

---

[23] Quine (1960), pp. 51-52.

[24] Ibid., pp. 52-53.

such a conclusion wouldn't be justified; for as far as the evidence goes, 'squiggle' might be translated 'is the same thing as', 'is an undetached spatial part of the same extended whole as', or 'is a temporal stage of the same enduring complex as'. The choice of a translation for 'gavagai' depends on a choice for 'squiggle', and vice versa. We can, if we like, translate 'gavagai' as 'rabbit' and 'squiggle' as 'is the same thing as', or we can translate 'gavagai' as 'undetached spatial rabbit part' and translate 'squiggle' as 'is an undetached spatial part of the same extended whole as' -- or we can make other choices, always adjusting our translations of 'gavagai' and 'squiggle' as a pair. Seeing these different translation theories as *incompatible* with each other, while being equally compatible with all possible evidence, Quine takes himself to have established the Underdetermination of Translation by Data.[25]

But how does one move from this to the Indeterminacy of Translation? Not being clear about this in *Word and Object*, Quine left himself open to an objection made in Chomsky (1969). As Chomsky observed, interesting empirical theories are always underdetermined by the data against which they are tested. If one is not already a behaviorist, one won't regard the mundane fact that theories of translation are underdetermined by Quine's behavioral data to be any more threatening to those theories, and the notion of meaning they employ, than similar underdetermination is threatening to theories in physics, chemistry, and biology. Since Quine *was* a behaviorist, this may not have worried him, but he did need a route to Indeterminacy that didn't presuppose behaviorism to convince the rest of us. In Quine (1969), he offered one. Recognizing that many might agree with him that facts about stimulus meanings won't resolve indeterminacies about meaning, he suggested that adding *any other physical* facts won't help. After all, if one can't deduce from an L-

---

[25] Although Quine doesn't clarify in what sense the theories are supposed to be incompatible, I suspect he intends logical incompatibility. If so, we must think of translation theories as being augmented by certain claims that go beyond those he specifies. This complication is addressed on pp. 326-328 of Soames (1999), and pp. 238-240 of Soames (2003b).

speaker's behavior that his word 'gavagai' means the same as 'rabbit', how would it help to add that certain neurons in his brain fire when he uses the word? We can no more read off the contents of a person's words from neurological claims than we can read off the contents of his words from statements about the sounds he emits in different environments. So, if non-intensional descriptions of linguistic behavior don't determine meaning, adding neurological – or any other physical facts -- won't help.

These thoughts led to Quine's most powerful route to the Indeterminacy of Translation, which was to derive the thesis from the following (roughly stated) doctrines.[26]

> Physicalism
> All genuine truths (facts) are determined by physical truths (facts).
>
> The Underdetermination of Translation by Physics
> Translation is not determined by the set of all physical truths (facts), known and unknown. For any pair of languages and theory of translation T for those languages, there are alternative theories of translation, incompatible with T, that accord equally well with all physical truths (facts).

If these doctrines are correct, then claims about what our words mean -- e.g. *that word$_1$ used by person$_1$ means the same as word$_2$ used by person$_2$* (for different persons and words) – never express genuine truths. Hence, the Indeterminacy of Translation. However, in order to properly assess this argument, the *determination relation* between sets of potential truth bearers must be carefully specified. When one does this, one encounters a serious problem; determination relations that make the Underdetermination of Translation by Physics most plausible make Physicalism implausible, and vice versa. To date, no clearly specified relation has been offered that makes both plausible. Although, this doesn't show that none ever will be, it raises a worry. Perhaps, the initial plausibility of Quine's argument for the Indeterminacy of Translation lies in equivocation about the determination relation.[27]

---

[26] Soames (1999) contains a more nuanced discussion.

[27] See Soames (1999) and chapter 10 of Soames (2003b).

The upshot, for Quine, of the indeterminacy thesis is that no claim that L-speakers use 'gavagai' to mean the same as what I mean by w (for any word in my language) expresses a truth. Similar reasoning leads to the result that no claim that they use 'gavagai' to refer to the same thing I use w to refer to is true. From here it is not far to the conclusion that they don't use it to refer to anything.[28] Although this sounds extreme, and Quine doesn't explicitly say it, what he does say in a crucial passage in "Ontological Relativity" is tantamount to it. Earlier in the article, he notes that the Indeterminacy of Translation, according to which it is indeterminate whether 'gavagai' means the same as 'rabbit', is inextricably linked to the Inscrutability of Reference, according to which it is indeterminate whether 'gavagai' refers to rabbits. He notes that if it is indeterminate whether the native's word 'gavagai' refers to rabbits, then it is indeterminate whether *my neighbor* uses to 'rabbit' to refer to rabbits. At this point he considers the possibility of making the same claim about himself.

> "I have urged…that the inscrutability of reference is not the inscrutability of a fact; there is no fact of the matter. But if there is really no fact of the matter, then the inscrutability of reference can be brought even closer to home than the neighbor's case; we can apply it to ourselves. If it is to make sense to say even of oneself that one is referring to rabbits… and not to rabbit stages…then it should make sense equally to say it of someone else…We seem to be maneuvering ourselves into the absurd position that there is no difference on any terms, interlinguistic or intralinguistic, objective or subjective, between referring to rabbits and referring to rabbit parts or stages…Surely this is absurd, for it would imply that there is no difference between the rabbit and each of its parts or stages…Reference would seem now to become nonsense not just in radical translation but at home."[29]

The implicit argument here can be reconstructed as follows:

R1.   It is indeterminate (hence there is no fact of the matter) regarding whether the native uses 'gavagai' to refer to rabbits, temporal stages of rabbits, etc. It is similarly indeterminate whether my neighbor uses 'rabbit' to refer to rabbits, rabbit stages, etc.

R2.   If (R1), then it is indeterminate (hence there is no fact of the matter) regarding whether I use 'rabbit' to refer to rabbits, temporal stages of rabbits, etc.

---

[28] See pp. 260-264 of Soames (2003b).

[29] Quine (1968) at pp. 47-48 of the version in *Ontological Relativity and Other Essays*.

R3.   So it is indeterminate whether I use 'rabbit' to refer to rabbits, temporal stages of rabbits, etc. More generally, it is never determinate that a word (used by anyone) refers to (i.e. is true of) all and only rabbits, as opposed to all and only temporal stages of rabbits, all and only undetached rabbit parts, etc.

R4.   But if that (R3) is so, then there is no difference between referring to all and only rabbits and referring to all and only temporal stages of rabbits, or to all and only undetached rabbit parts – i.e. *for any a word w, it is true that w refers to all and only rabbits only if it is equally true that w refers to all and only temporal stages of rabbits, and to all and only undetached rabbit parts*.

R5.   If the consequent of R4 is true, then there is no difference between rabbits, temporal stages of rabbits, undetached rabbit parts – i.e. something is a rabbit iff it is a temporal stage of a rabbit iff it is an undetached rabbit part.

R6.   So, there is no difference between rabbits, temporal stages of rabbits, undetached rabbit parts, and so on.

Since R6, is false, at least one of R1, R2, R4, or R5 must also be false.

In selecting the offending premise it is important to remember that the indeterminacy and inscrutability theses are ontological, not epistemic. They claim that all physical truths don't determine the truth of statements about meaning and reference. Since the physical truths don't distinguish between L-speakers, my neighbors, and me, they must determine reference for all of us if they determine truths for any. So, one can't avoid the absurd R6 by rejecting R3. Nor can Quine reject R1, which is his own doctrine. This leaves R4 and R5. What, from his point of view, is wrong with them? The passage quoted contains a comment on R4 suggesting that the previous steps require us to accept its "absurd" consequent. He concludes with the comment, "Reference would seem now to become nonsense not just in radical translation but at home." Two paragraphs later he adds that reference *really is nonsense*, if taken "absolutely," but it isn't nonsense if it is relative to a background theory or language.

Quine sees his indeterminacy theses as requiring us to reject our ordinary "absolute" notion of reference, which is employed in the reductio throughout. Although Quine himself isn't clear about this, properly understood, this rejection blocks R6 by allowing him to reject R5, the consequent of which is the absurd conclusion to be avoided. This is possible only if

the conditional to which the antecedent of R5 predicates truth is itself true by falsity of antecedent – i.e. only if no word refers to (is true of) all and only rabbits. Since the same argument could be repeated for any class of things, Quine must deny that any word ever refers to anything, in the ordinary sense of 'refer' ('is true of').

From here we move to Quine's reconstruction. Just as he doesn't simply reject ordinary meaning, but replaces it with 'stimulus meaning', so he doesn't simply reject ordinary reference, but replaces it with 'Tarski-reference' in his own present language, $L_{Quine\ Now}$. The "reference" of others, or of his own reference at other times, is then taken to be the product of translation of their words onto expressions of $L_{Quine\ Now}$, plus Tarski-reference for that language. On this view, when one says that x's use of word w at t refers to (is true of) $\mathcal{F}$s, one is saying that *relative to some underdetermined theory T of translation from x's words at t onto the words of one's own present language*, x's use of w at refers to $\mathcal{F}$s -- which is to say that there is some word or phrase P in one's own present language such that (i) according to T, w, as used by x at t, means the same as P, and (ii) P (as one now uses it) refers to (is true of) $\mathcal{F}$s. As for Tarski reference, it is strictly disquotational and can, in principle, be defined along the following lines.

Tarski-reference of my present words:
For all names n of my present language and objects o, n refers to o iff n = 'Alfred' and o is Alfred, or n = 'Willard' and o is Willard, or ...

For all predicates P of my present language and objects o, P refers (applies) to o iff P = 'rabbit' and o is a rabbit, or P = 'dog' and o is a dog, or P = 'white' and o is white, or ....

Just as Tarski-reference replaces ordinary reference, so Tarski-truth replaces ordinary truth for one's own present language. Thus, the neither the claims one makes about the Tarski-referents of the words of one's own present language, nor the claims one makes about the Tarski-truth or Tarski-falsity of its sentences are relativized to underdetermined theories

of translation. Underdetermination arises only when reporting the speech of others, or of oneself at other times.[30]

Quine's Self-Refuting Eliminativism

As I have explained, the Indeterminacy of Translation and the Inscrutability of Reference lead to eliminativism about: (i) our ordinary notion of meaning; including the notion of a sentence's meaning that P, (ii) our ordinary notion of reference or application – *being true of*. In addition, as Quine himself notes, this eliminativism requires him to reject standard propositional attitudes, such as believing and asserting that P. He says:

> "For, using the intentional words 'believe' and 'ascribe', one could say that a speaker's term is to be construed as 'rabbit' if and only if the speaker is disposed to ascribe it to all and only the objects that he believes to be rabbits. Evidently, then, the relativity to non-unique systems of analytical hypotheses [which give rise to indeterminacy] invests not only translational synonymy but intentional notions generally. Brentano's thesis of the irreducibility of intentional idioms is of a piece with the thesis of indeterminacy of translation. One may accept the Brentano thesis either as showing the indispensability of intentional idioms and the importance of an autonomous science of intention, or as showing the *baselessness* of intentional idioms and the *emptiness* of a science of intention. My attitude…is the second. To accept intentional usage at face value is, we saw, to postulate translation relations as somehow objectively valid though indeterminate in principle relative to the totality of speech dispositions. Such postulation promises little gain in scientific insight if there is no better ground for it than that the supposed translation relations are presupposed by the vernacular of semantics and intention. Not that I would forswear daily use of intentional idioms, or maintain that they are practically dispensable. But they call, I think, for bifurcation in canonical notation. Which turning to take depends on which of the various purposes of a canonical notation happens to be motivating us at the time. If we are *limning the true and ultimate structure of reality*, the canonical scheme for us is the austere scheme that knows no quotation but direct quotation and no propositional attitudes but only the physical constitution and behavior of organisms."[31]

The message can't be missed. If claims about propositional attitudes were determinate, translation would also be. Since translation is indeterminate, claims about propositional attitudes are too. Strictly speaking there are no facts about attitudes like belief

---

[30] Replacement of the rejected ordinary notions of reference and truth by language-relativized Tarski-reference and Tarski–truth is discussed on pp. 270-277 of Soames (2003b). Though not explicit in "Ontological Relativity," such replacement and relativization becomes so in section 20 of Quine (1992).

[31] Quine (1960) pp. 220-221, my emphasis.

and assertion, just as there are no facts about translation. Whatever the practical utility of such notions, they have no place in a properly scientific description of the world. This is icing on the cake, further underlining Quine's rejection of our ordinary notions *truth* and *being true of,* which are conceptually connected with the eliminated notions of meaning, reference, what is said, and what is believed. Our ordinary notion of truth, which applies to these things, as well as to sentences of arbitrary different languages, must be replaced with Tarski-truth, which is restricted to one's own present language.

With this, Quine's position has become not just implausible but self-defeating. Elimination of ordinary semantic notions undermines the Quinean's ability to formulate and argue for indeterminacy and inscrutability. Consider the claim that biology is determined by the physical truths, but semantics and non-behaviorist psychology are not. In speaking of truths, the Quinean has to be speaking of Tarski-truths-of-his present-language. So we have a thesis that a certain set of physical sentences of his present language determines a certain set of biological sentences of his present language, but does not determine a third set of semantic and psychological sentences of that language. But why should this matter? Suppose present physics lacks some concepts needed for an accurate and comprehensive scientific description of the world. *If the physics capable of being expressed in one's-present--language is incomplete in this way, then the fact that some set S of claims is not determined by one's Tarski-true physical sentences won't provide any compelling reason to doubt the genuine truth of members of S.* If someone three hundred years ago had maintained that the set of Tarski-true physical sentences of his then present language failed to determine the set of Tarski-true biological sentences of that language, nothing of philosophical interest would follow. So, the Quinean faces a dilemma. Either his theses of physicalism, underdetermination, indeterminacy, and inscrutability state something about the totality of

physical truths in the ordinary sense, in violation of his eliminativism, or they state something only about the totality of Tarski-truths of his present language, in which case physicalism and indeterminacy are indefensible, while underdetermination is uninteresting. This is the sad end of Quine's long flight from both intension and intention.

Analytic Philosophy After Quine

The failure of Quine's eliminativist project does not detract from the value of his earlier work pointing out the shortcomings, and correcting the errors, of his logical-empiricist predecessors. He was right to attack their linguistic theory of the apriori, to observe that quantified modal logic could not succeed as long as necessity was identified with analyticity, and to reject their use of analyticity as an all purpose philosophical tool. He was similarly effective in moving philosophy away from conceptual analysis and toward something continuous, though not identical, with science. Today's conception of the overlap of philosophy of language with linguistics, of philosophy of mind and epistemology with psychology, of philosophy of physics and biology with physics and biology, and of metaphysics with science more generally, owes much to his influence. To be sure, his behaviorism was a dead end, and his flights from intension and intention were mistaken. Fortunately, they were spectacular mistakes that couldn't be ignored. Just as his own era progressed by correcting the fundamental errors of the logical empiricists, so the era succeeding his -- led by his sometime students Saul Kripke and David Lewis -- progressed by correcting his errors about meaning and modality.

References

Ayer, A.J. (1959), *Logical Positivism*, Glencoe, Illinois: Free Press.

Carnap, Rudolf, (1928a), *Der Logische Aufbau Der Welt*, Berlin-Schlachtensee, Weltkreis-Verlag; trans. by Rolf. A. George, *The Logical Structure of the World*, Berkeley and Los Angeles: University of California Press, 1967.

_____, (1928b), *Scheinproleme in der Philosophie*, Berlin-Schlachtensee, Weltkreis-Verlag; trans. by Rolf. A. George, *Pseudoproblems in Philosophy*, Berkeley and Los Angeles: University of California Press, 1967.

_____, (1930), "Die alte und die neue Logik," *Erkenntnis,* 1:12-26; trans. by Issac Levi, "The Old and the New Logic," in Ayer (1959).

_____, (1931), "Die logizistische Grundlegung der Mathematik," in *Erkenntnis* 2: 91-105; trans. "On the Logicist Foundations of Mathematics," in Paul Benacerraf and Hilary Putnam, eds. *Readings in the Philosophy of Mathematics*, Englewood Cliffs, N.J.: Prentice Hall, 1964.

_____, (1932a), Uberwindung der Metaphysik durch logische Analyse der Sprache," *Erkenntnis*, 2: 219-241; "The Elimination of Metaphysics Through the Logical Analysis of Language," trans. by Arthur Pap in Ayer, (1959).

_____, (1932b), "Die physikalische Sprache al Universalsprache der Wissenschaft," *Erkenntnis*, 2: 432-465; trans. by Max Black, "The Unity of Science," in *Psyche Miniatures, General Series*, no. 63, London: Kegan Paul.

_____, (1934), *Logische Syntax der Sprache*, Vienna: Springer; trans. by A. Smeaton, *The Logical Syntax of Language*, London: Kegan Paul Trench, Trubner & Co, 1937.

_____, (1950), "Empiricism, Semantics, and Ontology," *Revue Intern. De Phil*. 4: 20-40; revised and reprinted in Carnap (1956).

_____, (1955), "Meaning and Synonymy in Natural Languages," *Philosophical Studies*, 6: 33-47; reprinted in Carnap (1956).

_____, (1956) *Meaning and Necessity*, 2nd edition, Chicago: University of Chicago Press.

Chomsky, Noam (1969), "Quine's Empirical Assumptions," in Davidson and Hintikka (1969).

D. Davidson and J. Hintikka (1969), *Words and Objections*, Dordrecht: Reidel.

Fine, Kit (2007), *Semantic Relationism*, Oxford: Blackwell.

Grice, Paul, and Peter Strawson (1956), "In Defense of a Dogma," *Philosophical Review*, 65: 141-158.

Harman, Gilbert (1979), "Meaning and Theory," *Southwestern Journal of Philosophy*, 9: 9-20.

Kaplan, David (1986), "Opacity," in L.E. Hahn and P.A. Schilpp, eds., *The Philosophy of W.V. Quine*, La Salle, Illinois: Open Court.

Kazmi, Ali (1987), "Quantification and Opacity," *Linguistics and Philosophy,* 10: 77-100.

_____, (1992), "Some Remarks on Indiscernibility," *Canadian Journal of Philosophy*, Supplementary Volume 18, 167-178.

Lewis, C.I. (1929), *Mind and the World Order,* New York: Scribner.

_____, (1946), *An Analysis of Knowledge and Valuation,* La Salle, Illinois: Open Court.

Quine, W.V.O. (1936), "Truth by Convention," in O.H. Lee, ed., *Essays for A.N. Whitehead,* New York: Longmans; reprinted in Quine (1966).

_____, (1943), "Notes on Existence and Necessity," *Journal of Philosophy*, 40: 113-127.

_____, (1947), "The Problem Interpreting Modal Logic," *Journal of Symbolic Logic*, 12: 43-48.

_____, (1948), "On What There Is," *Review of Metaphysics*, 2: 21-38.

_____, (1951a), "On Carnap's Views on Ontology," *Philosophical Studies*, 2: 65-72.

_____, (1951b), "Two Dogmas of Empiricism," *Philosophical Review,* 60: 20-43; second, revised, edition, 1980.

_____, (1953a), *From a Logical Point of View*, Cambridge: Harvard University Press.

_____, (1953b), "Reference and Modality," in *From a Logical Point of View*.

_____, (1953c), "Three Grades of Modal Involvement," *Proc. XI International Congress of Philosophy,* 14: 65-81.

_____, (1956), "Quantifiers and Propositional Attitudes," *Journal of Philosophy,* 53: 177-187.

_____, (1960), *Word and Object*, Cambridge: MIT Press.

_____, (1966), *The Ways of Paradox*, New York: Random House.

_____, (1968), *Journal of Philosophy*, 64: 185-212; reprinted in *Ontological Relativity and Other Essays,* New York: Columbia University Press, 1969.

_____, (1969), "Reply to Chomsky," in Davidson and Hintikka (1969).

_____, (1975), "On Empirically Equivalent Systems of the World," *Erkenntnis*, 9: 313-328.

_____, (1992), *In Pursuit of Truth*, Cambridge: Harvard University Press.

Richard, Mark (1987), "Quantification and Leibniz's Law," *Philosophical Review*, 96, 555-578.

Soames, Scott (1987), "Substitutivity," in J.J. Thomson, ed., *On Being and Saying*, Cambridge: MIT Press.

_____, (1999), "The Indeterminacy of Translation and the Inscrutability of Reference," *Canadian Journal of Philosophy,* 29, 1999, 321-370.

_____, (2002) *Beyond Rigidity,* New York: Oxford University Press.

_____, (2003a), *Philosophical Analysis in the Twentieth Century*, Vol. 1, Princeton: Princeton University Press.

_____, (2003b), *Philosophical Analysis in the Twentieth Century*, Vol. 2, Princeton: Princeton University Press.

_____, (2012), "Two Versions of Millianism," in W. Kabasenche, M. O'Rourke, and M. Slater eds., *Reference and Referring: Topics in Philosophy, Vol. 10*, Cambridge, MIT Press.

Williamson, Timothy (2007), *The Philosophy of Philosophy,* Oxford: Blackwell.