

ON THE LOGICAL POSITIVISTS' PHILOSOPHY OF PSYCHOLOGY: LAYING A LEGEND TO REST*

Sean Crawford

Abstract. The received view in the history of the philosophy of psychology is that the logical positivists—Carnap and Hempel in particular—endorsed the position commonly known as “logical” or “analytical” behaviourism, according to which the relations between psychological statements and the physical-behavioural statements intended to give their meaning are analytic and knowable *a priori*. This chapter argues that this is sheer legend: most, if not all, such relations were viewed by the logical positivists as synthetic and knowable only *a posteriori*. It then traces the origins of the legend to the logical positivists' idiosyncratic extensional or at best weakly intensional use of what are now considered crucially strongly intensional semantic notions, such as “translation,” “meaning” and their cognates, focussing on a particular instance of this latter phenomenon, arguing that a conflation of *explicit definition* and *analyticity* may be the chief source of the legend.

1. The Legend

Near the beginning of his recent Carnapian book *Constructing the World*, David Chalmers rightly remarks that “A surprising and often-overlooked feature of the *Aufbau* is that Carnap there requires only that definitions be extensionally adequate” (Chalmers 2012, p. 4). Chalmers then points out that in the 1961 preface to the second edition of the *Aufbau* (Carnap 1928/1961), Carnap “says that the definitions should be held to a stronger, intensional, criterion of adequacy” (Chalmers 2012, p. 5). Chalmers goes on to give “analyticity, apriority and necessity” as examples of stronger criteria (obviously intending the necessity in question to be logical or metaphysical necessity). It is clear from the preface, however, that Carnap did not demand such strongly intensional criteria, but only that the coextensiveness of the *definiendum* and *definiens* not be merely “accidental, but necessary, i.e., it must rest either on the basis of logical rules or on the basis of natural laws” (Carnap 1928/1961,

* Dedicated to the memory of Rachel Jardine.

p. ix).¹ Since natural laws are obviously not analytic or knowable *a priori*, it is clear that Carnap did not think that all philosophical definitions are required to meet the very strong modal criteria of intensionality that Chalmers requires for his constructional project. Hannes Leitgeb (2011, pp. 267, 269-270) also draws attention to the extra demand stated by Carnap in the preface, but, unlike Chalmers, notes explicitly that the intensionality in question may be merely one of nomological necessity. However, it is important to note — and this something that neither Chalmers nor Leitgeb mention, as they are concerned solely with exploring the possibility of actually carrying off a suitably revised *Aufbau*-like construction project — that the stronger requirement of some degree of intensional adequacy for definitions is acknowledged by Carnap very soon after the *Aufbau*.

Just such an intensional adequacy condition on definitions short of analyticity (and hence a priority) finds expression, for example, four years later in *Unity of Science* (Carnap 1932). One of the aims of that work is to show that protocol sentences about subjective experience can be physicalized, that is, translated into sentences employing only physical vocabulary (which is the only vocabulary known to be inter-subjective, inter-sensory, and universal, which are the requirements for any suitable confirmation base for empirical science). It is evident that Carnap there envisages such “translations” to be based partly on natural laws and hence neither logically necessary nor analytically true nor knowable *a priori*. The same is true of the broader goal of physicalizing the science of psychology (one element of unified science). The proposed physical definitions of psychological terms and physical translations of psychological sentences are required to meet (to borrow Chalmers’ apt phrase) stronger-than-extensional conditions of adequacy — but, contrary to Chalmers’s own constructional requirements, only up to *nomological* intensionality. That is to say, the proposal is that it is only nomologically necessary that the physical *definiens* and the psychological *definiendum* have the same extension, not logically necessary (let alone that the two be synonymous).

¹ Carnap also refers the reader to his reply to Goodman in the Schilpp volume (Carnap 1963b) where he says that the identity of extension must fulfill an additional requirement, which “consists in the condition that the correspondence hold, not merely accidentally, but on the basis of general regularities, e.g., physical laws or empirical generalizations” (p. 946).

Just as the original extensionality of the *Aufbau* construction is often overlooked², it is a virtually universally overlooked feature of Carnap's philosophy of psychology of the 1930s — the period in which he is alleged to have been a logical behaviourist — that his proposed physical-behavioural definitions of mental terms are only required to be intensionally adequate up to nomologicality— which means, contrary to popular belief, that Carnap was not a logical behaviourist. In so far as Hempel and fellow logical positivists agreed with Carnap on this, which they certainly seemed to, an even more general claim is warranted, namely, that the logical positivists were not logical behaviourists.

Yet Carnap and Hempel and their fellow logical positivists are considered to be examples of logical behaviourists *par excellence*. For example, in his celebrated 1963 critique of logical behaviourism, "Brains and Behaviour", Hilary Putnam wrote that

"the Vienna positivists in their 'physicalist' phase (about 1930) ... [produced] the doctrine we are calling *logical behaviourism*—the doctrine that, just as numbers are (allegedly) logical constructions out of *sets*, so *mental events* are logical constructions out of actual and possible *behaviour events*" (Putnam 1963, p. 326).

Logical behaviourism so understood was an "extreme thesis" implying that "all talk about mental events is translatable into talk about actual or overt potential behaviour" (*ibidem.*). Putnam then claimed that "In the last thirty years, the original extreme thesis of logical behaviourism has gradually been weakened to something" which, he said, "a great many" philosophers at the time would have accepted. The central tenet of the weakened thesis is that "[t]here exist entailments between mind-statements and behaviour-statements; entailments that are not perhaps analytic in the way that 'All bachelors are married' is analytic, but that nevertheless follow (in some sense) from the meanings of mind words" (*id.*, p. 327). Putnam proposed to call these entailments "*analytic entailments*."³ He eventually went on, of course, to offer the undergraduate philosophy student's favourite refutation of logical behaviourism: the fictional community of "super-spartans" (or "super-stoics"). These legendary people, as everyone knows, are capable of suppressing all pain behaviour even while suffering "the agonies of the damned"—thus supposedly demonstrating that the connection

² Even by Quine, to some extent, for as Nelson Goodman (1963, p. 555n5) has pointed out, "The avowed extensionalism of so outstanding a monument of phenomenism and constructionism as the *Aufbau* would seem to confute Quine's recent charge [in 'Two Dogmas'] that the notion of analyticity is a 'holdover of phenomenistic reductionism'."

³ Cf. (Putnam 1969).

between statements about pain and statements about pain behaviour cannot be a matter of analytic meaning equivalence or even one-way analytic entailment of any kind. The connection between the two is at best entirely synthetic—or, as we would put it nowadays, contingent.

As anyone familiar with philosophy of mind textbooks and anthology introductions can attest to, Putnam's understanding of the logical positivists' logical behaviourism is the orthodox one in the analytic tradition and continues to be so. Forty years later, for example, we find John Searle characterizing logical behaviourism as "a logical thesis about the definition of mental concepts. ... The model for the behaviourists was one of definitional identities. Pains are dispositions to behaviour in a way that triangles are three-sided plane figures. In each case it is a matter of definition" (2004, pp. 54-5), and citing C. G. Hempel as an example of a philosopher endorsing this kind of behaviourism.⁴ Since the "definitional identities" Searle refers to, as well as Putnam's "analytic entailments," are supposed to be knowable *a priori*, it follows that the analyses or translations or entailments proposed by logical behaviourists, as holding between psychological statements and behavioural statements, are supposed to be knowable *a priori*. We can, then, sum up the received view in analytic philosophy of the logical positivists' logical behaviourism as the thesis that the relation between psychological statements and the behavioural statements intended to give their meaning is an *a priori* knowable analytic entailment or equivalence. This is of course why logical behaviourism is also known as "analytical behaviourism."

As I have already suggested, when one turns to the two canonical logical positivist texts of logical behaviourism, one finds a very different story. Neither Rudolf Carnap's "Psychology in Physical Language" (1932) nor C. G. Hempel's "The Logical Analysis of Psychology" (1935) espouses the received view of logical behaviourism just adumbrated; nor do any of their subsequent writings on philosophy of psychology. First, neither Carnap nor Hempel restrict their reductive analyses of psychological statements to statements describing *behaviour*; on the contrary, they both included in their analyses reference to internal neurophysiological states. Secondly, in the eyes of both Carnap and Hempel, most of the logico-semantic relations between psychological statements and both the behavioural statements and

⁴ Soon after Putnam, similar claims were made by Fodor (1968, pp. 51, 155n6), Cornman (1971, pp. 132ff, esp. p. 140), and Kim (1971, p. 328).

the physical statements intended to give their meanings are synthetic and knowable only *a posteriori*.

Let us call the historical claim that the logical positivists maintained that the connections between “mind talk” and “behaviour talk,” or between “mind talk” and “physical-thing talk,” were analytic, the *analytic entailment interpretation* of logical positivism’s logical behaviourism. The fact is that the analytic entailment interpretation is as much a fiction as Putnam’s super-spartans. But if so, how did this extraordinary and ubiquitous legend get perpetuated throughout the analytic tradition? My chief aim here is to try to make some progress toward answering this question. I shall first establish (in §2) that the analytic entailment interpretation is indeed nothing more than a legend.⁵ Second (in §3), I will make some conjectures about why it became so widely accepted.

2. Laying the Legend to Rest

In his Physicalist writings of the 1930s Carnap refers to “rules of transformation” (sometimes “rules of inference” and “rules of translation”) of the physical language in which definitions and translations are to be carried out in accordance with the programme of Unified Science. Although he is not always entirely explicit about it, it is evident to a careful reader that even in the earliest of these writings from his Physicalist phase that not all of these inference rules are laws of logic and that some of them are intended to be laws of nature. For example, in ‘Unity of Science’ (1932/1934) Carnap writes of the “the rules of transformation inside the physical language (*including the system of natural laws*)” (p. 88, my emphasis; cf. p. 92). This becomes much clearer by the time of *The Logical Syntax of Language* (1934) and ‘Testability and Meaning’ (1936-37) in which Carnap explicitly distinguishes between the L-rules and the P-rules of a scientific language on the basis of which transformations may be validly carried out: the former are logical laws and the latter empirical physical laws. (Carnap also defines various correlative notions, such as L-validity and P-validity, L-equipollence and P-equipollence, and L- and P-synonymy.) Both kinds of “translation rules” are to be employed in physicalization.

⁵ The only writers I know of who explicitly do not hold the analytic entailment interpretation are Alston and Nakhnikian (1963, p. 391), Hempel (1958, 1969), Cirera (1993) and Kim (2003). I am grateful to my colleague Thomas Uebel for drawing my attention to Cirera’s important article as well as to Hempel (1969) and much other relevant literature, as well as for many edifying discussions of Carnap and logical positivism, which prompted several changes to this chapter.

In *Philosophy and Logical Syntax* (1935/1963), Carnap claims that “every sentence of any branch of scientific language is equipollent to some sentence of the physical language, and can therefore be translated into the physical language without changing its content” (p. 455; cf. *The Logical Syntax of Language* §82). Carnap is very clear in this work (as well as in *The Logical Syntax of Language* §§51, 82) that there can be two concepts of equipollence, that is, equivalence, in the physical language: logical equipollence (L-equipollence) and physical equipollence (P-equipollence). Two sentences are L-equipollent when they are mutually derivable solely on the basis of logical laws; two sentences are P-equipollent when they are mutually derivable only on the basis of physical laws as well.

Given this distinction between L- and P-concepts, in order for the legendary analytic entailment interpretation to be true, Carnap would have to maintain that all physical translations of psychological sentences are L-equipollent to them. But he did not maintain this. He explicitly allowed a psychological sentence, Q_1 , and a physical translation of it, Q_2 , to be P-equivalent, as Q_1 could be transformed into Q_2 on the basis of “a scientific law, that is, a universal sentence belonging to the valid sentences of the scientific language-system” (1935/1963, p. 456). He took some pains to point out that, in his view, this universal sentence “need not be analytic; the only assumption is that it is valid. It may be synthetic, in which case it is P-valid” (*ibidem*).⁶ Similarly, with respect to the translation of the psychological statement ‘Mr. A is now excited’ (P_1), Carnap (1932/1959) says that it

⁶ In a letter he wrote to Herbert Feigl in 1933 (translated by and quoted in Feigl 1963, p. 255), Carnap explicitly states that the two sentences are not analytic. He offers two translations of ‘N. has a visual image of a house’ (A), ‘The organism of N. is in the state of house-imagining’ (B_1) and ‘In the organism of N. there is an electrochemical condition of such a kind (described in terms of electrochemistry)’ (B_2), and then remarks that:

Both B_1 and B_2 are translations of A. According to my recently adopted terminology, I assert: A is equivalent (“*gehaltgleich*”) to both statements ... ; viz., L-equivalent (*logically* equivalent) with B_1 ; but P-equivalent (*physically* equivalent) with B_2 , i.e., mutually translatable (derivable) using besides the logical laws also natural laws as rules of inference, incorporated as transformation rules in the scientific language. You are therefore right in saying that B_2 is only synthetically equivalent with A.

As Ramon Cirera (1993) importantly points out, while B_1 , unlike B_2 , is claimed by Carnap to be L-equivalent to A, it is not behavioural—in fact, it is not even physical. Neither Feigl nor Cirera say what the point of B_1 is. One possibility is that it is an adverbial analysis of (A) intended to avoid commitment to the intentional object apparently designated by the phrase ‘visual image of a house’, and hence to avoid intentional language, thus making the ultimate goal of a physical translation into B_2 easier. Such adverbial techniques were sometimes employed by Russell in order to avoid commitment to intentional objects (and by some of the American New Realists in a quasi-behaviourist spirit) and Russell of course influenced Carnap. Chisholm (1955-56) famously criticized such adverbial strategies for avoiding intentional language but I know of no response by Carnap to Chisholm on this point.

“cannot, indeed, today be translated into a physical sentence P_3 of the form “such and such a physico-chemical process is now taking place in A’s body” (expressed by a specification of physical state-coordinates and by chemical formulae). Our current knowledge of physiology is not adequate for this purpose” (1932/1959, p. 175).

It goes without saying that Carnap did not view knowledge of physiology as *a priori*. It is lack of *a posteriori* knowledge of physiology that prevents P_1 from being translated into P_3 . It follows, of course, that the relation between P_1 and P_3 is synthetic, not analytic.

It might be objected that it is simply not true that, in order for the analytic entailment thesis to hold, all physical translations of psychological sentences must be L-equivalent to them. It is only required that the physical-*behavioural* ones be L-equivalent. What Carnap has in mind here is the case in which Q_2 is a non-behavioural, neurophysiological sentence. This of course can only be P-equivalent. But that is entirely consistent with claiming that a behavioural sentence, Q_3 , must be L-equivalent with the psychological sentence of which it is a translation. Moreover, this fits nicely with the logical positivists’ view (emphasized by Feigl (1958)) that the progress of the science of psychology will unfold in two stages: an initial “peripheralist” black-box behaviouristic stage and a later “centralist” neurophysiological stage.

Carnap does indeed countenance physicalizations that are explicitly only about overt behaviour. In ‘Psychology in Physical Language’, he says that P_1 (‘Mr. A is now excited’) may be inferred from p_1 , which is a sentence “about the behaviour of A, e.g. about his facial expressions, his gestures, etc. or about physical effects of A’s behaviour, e.g. about characteristics of his handwriting” (*id.*, p. 171). The question is whether Carnap intends P_1 and p_1 to be L-equivalent, as the analytic entailment interpretation requires, or whether they are supposed to be merely P-equivalent — just as P_1 and P_3 are. Although he does not employ L- and P-concepts explicitly in ‘Psychology’, it is nonetheless clear that the logical relation between P_1 and p_1 is synthetic and therefore that the two are intended to be at best P-equivalent.⁷ For he tells us that P_1 is to be derived from p_1 on the basis of the “major premise O,” which states that “When I perceive a person to have this facial expression and hand-writing

⁷ And perhaps not even that, owing to the fact that the connection between the two may not even be nomological, as we shall see presently.

he (usually) turns out to be excited. (A sentence about the expressional or graphological signs of excitement.)” (*ibidem.*). Earlier in the paper, Carnap draws a distinction between singular and general scientific — including psychological — sentences (*id.*, p. 168), taking pains to emphasize (against what he calls “phenomenology” and its objectionable *a priori* or at least non-inductive methods) that general sentences are discovered inductively and are therefore to be considered empirical hypotheses (*ibidem.*). He then explicitly says that the major premise O, on the basis of which P_1 is derived from p_1 , is such a general sentence (*id.*, p. 171). O’s synthetic status (as an empirical inductive hypothesis) is therefore beyond question. It follows that the connection between P_1 and p_1 is synthetic too; generalizing, the connection between psychological sentences and their behavioural protocols is, contrary to the analytic entailment interpretation, synthetic.

This is further highlighted by what Carnap goes on to say:

“The cited relationship between P_1 and p_1 may also be seen in the fact that under certain circumstances, the inference from p_1 to P_1 may go astray. It may happen that, although p_1 occurs in a protocol, I am obliged, on the grounds of further protocols, to retract the established system sentence P_1 . I would then say something like, ‘I made a mistake. The test has shown that A was not excited, even though his face had such and such an expression’” (*ibidem.*).

Similarly, five years later in ‘Logical Foundations of the Unity of Science’, he writes:

“Let us take as an example the term ‘angry’. If for anger we knew a sufficient and necessary criterion to be found by a physiological analysis of the nervous system or other organs, then we could define ‘angry’ in terms of the biological language. The same holds if we knew such a criterion to be determined by the observation of the overt, external behaviour. But a physiological criterion is not yet known. And the peripheral symptoms known are presumably not necessary criteria because it might be that a person of strong self-control is able to suppress these symptoms. If this is the case, the term ‘angry’ is, at least at the present time, not definable in terms of the biological language. But, nevertheless, it is reducible to such terms. ... The logical nature of the psychological terms becomes clear by an analogy with those physical terms which are introduced by reduction statements of the conditional form. Terms of both kinds designate a state characterized by the disposition to certain reactions. In both cases the state is not the same as those reactions. Anger is not the same as the movements by which an angry organism reacts to the conditions in his environment, just as the state of being electrically charged is not the same as the process of attracting other bodies. In both cases, that state sometimes occurs without these events which are observable from outside; they are consequences of the state according to certain laws and may therefore under suitable circumstances be taken as symptoms for it; but they are not identical with it (Carnap, 1938, pp. 56-57, 59).⁸

⁸ Hempel’s discussion in ‘The Logical Analysis of Psychology’ (1935, esp. §V) is considerably less clear about this, and this lack of clarity may well have contributed to the legend, especially given that

We shall return to Carnap’s important distinction between definition and reduction below (in §3) because it is highly relevant to explaining the origins of the legend. For the moment, however, let us note that Carnap here anticipates and pre-empt’s Putnam’s super-spartan objection by more than two decades, explicitly stating that he does not view the behavioural dispositions associated with anger as identical with it; on the contrary, the behavioural “symptoms” of anger are (normally) caused by anger. This is already clear in ‘Psychology’ for, as we saw above, he considers the general sentence O to state behavioural “signs” of excitement — in other words, “symptoms,” that is, causal effects of excitement, noting that they may be present even when excitement is not.⁹ In this he also anticipates the later causal critiques of behaviourism levelled by Jerry Fodor and David Armstrong (that mental events are not identical with behaviour but are the causes of behaviour). The central point, again, is the fact that even here, with an explicitly overt-behavioural proposal, Carnap does not declare that such an overt-behaviouristic reduction sentence for anger will be analytic. The “laws” referred to, connecting inner states with outer behavioural reactions or symptoms, are empirical physical laws and so are intended to be P-valid (at best) and so synthetic.

Jaegwon Kim (2003, p. 275), however, sees Carnap as anticipating a causal-functional analysis of mental concepts and so as lending support to the idea that while Carnap may have viewed the nomological correlation of psychological sentences with neurophysiological sentences as P-equivalent, he viewed the correlation of psychological sentences with *behavioural* sentences as L-equivalent and hence analytic. Now, it is true, I think, that Carnap does here anticipate—remarkably—a

Hempel’s article is more widely reprinted than Carnap’s ‘Psychology’ (it appears, e.g., in the highly influential collection *Readings in Philosophical Analysis* (Feigl and Sellars 1949)). Hempel there confusingly claims that it is *logically contradictory* to say that all the symptoms obtain but the psychological state does not. This seems to be because, first, unlike Carnap, he is using the term ‘symptoms’ (sometimes putting it between inverted commas) to cover not only the external behavior but also the internal physiological processes associated with the psychological state, and, second, he is heading off a dualist objection. Still, given that Hempel maintains that all these “symptoms” are discovered empirically (cf. note 21 below), he cannot really mean that their presence with the absence of the psychological state is logically contradictory. Rather, the sentence describing such a situation would be (at best) what Carnap (1934, §52) calls *P-contravalid* (i.e., nomologically impossible). Hempel’s confusion here may be of a piece with the one Feigl and others make about the nature of definition, as discussed below in §3.

⁹ The inner event of excitement that is the cause of the outer behavioural symptom is eventually to be identified with the inner neurophysiological state that is the cause of the behavioural symptom, *à la* later causal-role functionalism. See immediately below for more on this.

kind of functionalism. However, contrary to what Kim says, it is, I think, incorrect to associate Carnap with the *analytic* functionalism of David Lewis and others, which is really a development of Rylean ordinary-language logical/analytical behaviourism. According to analytic functionalism, the causal-functional definitions of mental terms are specified *a priori* by conceptual analysis of commonsense psychology, and the only role for empirical science is to discover *a posteriori* which inner states are the actual physical realizers of the definitions. Carnap's proto-functionalism is, I believe, more akin to an thorough-going empirical psycho-functionalism (Block 1978), in which empirical science is involved at the first stage; that is, the functional definitions of mental terms are themselves in many cases specified empirically by scientific investigation.

In a word, Carnap's proto-functionalism is what one might call *synthetic functionalism*. This fits the text and the spirit of 'Psychology' and 'Foundations' better, to make better sense of the strong analogy Carnap draws between concept formation in the natural sciences and in the sciences of psychology, and to gel better with Carnap's (1932) procedure of physicalization.¹⁰ This was certainly Hempel's view of Carnap, who maintained that for the latter "those behavioural symptoms which are generally associated with a given psychological feature will often be determined by empirical investigations leading to empirical laws rather than by an aprioristic reflection upon the meaning of the psychological terms in question" (Hempel, 1969, pp. 179-180).¹¹

¹⁰ I discuss Carnap's empirical procedure of physicalization in slightly more detail in Crawford (2013).

¹¹ Moreover, as Carnap himself later pointed out (1952, p. 71) — and as is noted by Hempel (1951, p. 72; 1952, p. 28) and Arthur Pap (1958, ch. 11) — if there is more than one partial or conditional definition, e.g., a pair of (either unilateral or bilateral) reduction sentences for a given term, as obviously Carnap expected their to be for theoretical terms of behavioural psychology, then one can derive a synthetic statement from them, from which it follows that at least one of the definitions must be synthetic. See Carnap (1936-7) for the notion of a reduction sentence. I discuss the difference between reduction and definition in §3 below. Carnap (1952) ingeniously goes on to suggest a procedure to overcome the fact that pairs of reduction sentences introducing a theoretical term will have synthetic consequences, by taking a weaker (material) conditional sentence, whose antecedent is a statement of the empirical content of the reduction pair (the "representative sentence," as he (1936-7) called it) and whose consequent is the reduction pair, as the "meaning postulate" introducing the theoretical term, because none of its logical consequences containing only the original defining (basic) terms are synthetic. However, it is important to note for present purposes that while none of these logical consequences are synthetic, they are analytic only in the narrow sense, that is, they are logical truths. Such meaning postulates cannot therefore underwrite behavioural definitions for psychological terms in the spirit of textbook logical/analytical behaviourism, which obviously requires a broader notion of analyticity. See immediately below for further relevant discussion of this point.

In short, Carnap went the whole hog: most — perhaps all — of the semantic relations between *both* psychological sentences and *neurophysiological* sentences, and psychological sentences and *physical-behavioural* sentences, were synthetic.

Indeed, there is strong reason to doubt that in the period with which we are concerned — circa 1930-1940 — Carnap could even have held onto *any* analytic entailments between psychological and behavioural statements, even if he had wanted to. To see this, we need only remember the distinction between the narrow and the broad notion of analyticity, famously emphasized by Quine (1951) in “Two Dogmas”. In the narrow sense, an analytic truth is a logical truth, that is, a truth based solely on the meanings of the logical constants, such as the tautology “If p then p”; such a logical truth remains true under all “re-interpretations” of the meanings of its non-logical terms (such logical truths are now more commonly called *logically valid* sentences or formulae). An analytic truth in the broad sense is a truth based solely on the meanings of the logical terms and the non-logical or “descriptive” terms; an “all-bachelors-are-unmarried” kind of truth, as it were. The analyticity of these latter “truths of essential predication”, as Quine (1963) calls them (apparently after Mill), lies in the fact that they can be transformed into logical truths by substitution of definitions or synonymies. It is, of course, the broad notion of analyticity that is at stake here; no one ever claimed that logical behaviourists thought the entailments between psychological and physical-behavioural sentences were tautological or trivially logically true. But this crucial distinction between the broad and the narrow senses of analyticity seems to have occurred to Carnap only after 1940 and some late remarks by him (Carnap 1964/1994, p. 259) seem to suggest that it was Quine who woke him from his dogmatic Wittgensteinian slumbers (according to which analyticity just is logical truth) that he began to distinguish explicitly between logical truth and analyticity in the broader sense and attempt to define the latter as well as the former.¹²

I cannot be wholly definite here. The received view has it that it was only later, spurred by Quine’s attack on the analytic/synthetic distinction in “Two Dogmas” and his drawing there of the distinction between logical truth in the narrow sense and analyticity in the broader all-bachelors-are-married sense, that Carnap undertook the

¹² Actually, as Quine points out in “Two Dogmas” (§1, p. 41), even in the modal-logic phase of *Meaning and Necessity* (Carnap 1947) in his later “semantic period”, when he defined analytic truth semantically as truth in all state descriptions, Carnap’s definition was still only of the narrower notion of analyticity as logical truth.

task of characterizing analyticity in the broader sense. Accordingly, Carnap first proposed to define the broader sense of analyticity for an artificial observational language in “Meaning Postulates” (Carnap 1952), which is repeated in chapter 27 of *An Introduction to the Philosophy of Science* (Carnap (1966/94), and for an artificial theoretical language in chapter 28 of that same work; and his proposal for natural languages first appears in “Meaning and Synonymy in Natural Languages” (Carnap 1955; see also Carnap’s reply to Hempel in the Schilpp volume).¹³ Be that as it may, however, it suffices for the present argument that during the time in which Carnap formed his views on the philosophy of psychology under discussion here (roughly, the late 1920s to the late 1930s) he did equate analyticity with logical truth and so could not have endorsed textbook logical behaviourism.

3. Origins of the Legend

So how on earth did the analytic entailment interpretation ever get started in the first place?

One source of the legend, which time limitations prevent me from exploring in any detail, is probably the often-drawn analogy with phenomenalism and logical constructionism. We have already seen Putnam claim that logical behaviourism was a logical constructionist thesis akin to the construction of numbers out of sets—and obviously the latter construction will involve exclusive use of L-rules. Similarly, if one thinks of phenomenalism as the doctrine that material-object statements analytically entail sense-data statements—as for example in Chisholm (1957, Appendix)—and one thinks logical behaviourism is like phenomenalism, then one will no doubt arrive at the analytic entailment interpretation. But as we have seen, it is simply a mistake to view the logical positivists’ logical behaviourism as like phenomenalism, or more generally logical constructionism, in this sense.¹⁴

¹³ Awodey (2007, p. 244n 30) endorses this take on the matter, which does indeed seem to be Carnap’s own view, at least in some of his later writings (e.g., Carnap 1964/1994, p. 259). It is absolutely clear, however, despite what Carnap says in these writings, that he was alive to the importance of the distinction between narrow logical truth and broad analyticity as far back as 1943 (and probably earlier). See, e.g., the letter Carnap wrote to Quine on 21 January, 1943 (printed in Creath 1992, pp. 303ff). The nature and development of Carnap’s views on how to define formally the difference between narrow analyticity as logical truth and broad analyticity as essential predication has not to my knowledge been studied in detail let alone resolved in a fully satisfying manner. I hope to discuss it in future work.

¹⁴ Unless—ironically—one is explicitly thinking of *Carnap’s* (as opposed, e.g., to A. J. Ayer’s) phenomenalism and logical constructionism. See note 2 above. So while there is a parallel between *Carnap’s* phenomenalism/logical constructionism and logical behaviourism, it is precisely the opposite

More importantly, we can trace the origin of the analytic entailment interpretation largely to the positivists' highly technical and idiosyncratically use of the expressions "translation," "meaning," "synonymy," "definition" and their cognates. None of these terms is used today in anything like the way the logical positivists, especially Carnap and Hempel, were using them in the 1930s and even to some extent in the 1940s.¹⁵ Most of these terms and their cognates have strong modal implications for us now that they did not have back then for the positivists, namely a degree of intensionality up to at least logical necessity and perhaps even hyperintensionality. Carnap and Hempel, however, were working with a background extensional logic. When they claim that "mind talk" can be *translated* into "physical talk," what they mean is that one can construct material bi-conditionals with mind talk on the left-hand side and physical-thing-language talk on the right-hand side. These material bi-conditional "translations" were just that—*material* bi-conditionals, containing the straightforward truth functional connective symbolized by the horseshoe. Most of these material bi-conditionals (and reduction sentences) were clearly understood at the outset to be synthetic statements of lawful correlations discovered by empirically through scientific investigation. Carnap (1956) is especially clear about this. Of course, philosophers of science, including Carnap himself (1936/37), soon began to realize the extreme difficulties they faced in formalizing natural laws, disposition statements, and the counterfactual conditionals associated with them using an extensional logic.¹⁶ In light of this, Hempel remarks on one alternative approach that appeals to empirical causal laws: "The extensional 'if ... then ...' — which requires neither logical nor nomological necessity of connection — would therefore have to be replaced ... by a stricter, nomological counterpart which might be worded perhaps as 'if ... then, with causal necessity, ...'" (Hempel 1958, p. 188).

The most significant instance of this confusion of extensional and intensional semantic concepts as applied to the philosophy of psychology seems to occur in the work of the ambivalent and erstwhile logical positivist Herbert Feigl; and I conjecture that it is perpetuated and carried into contemporary analytic philosophy of mind's self-image through the Feigl-Putnam-Fodor line of influence. Early on in his famous

of what that parallel is usually taken to be. Carnap's phenomenalism and his behaviourism were *both* synthetic.

¹⁵ Ducasse (1941, ch. 7) complains that what Carnap (1935/63) calls translation is not truly translation.

¹⁶ See Carnap (1956), Hempel (1958) and Suppe (1977) for discussion of this.

long essay, ‘The “Mental” and the “Physical”’, Feigl explains the transition from logical behaviourism to the mind-brain identity theory:

A most important *logical* requirement for the analysis of the mind-body problem is the recognition of the *synthetic* or *empirical* character of the statements regarding the correlation of psychological to neuro-physiological states. It has been pointed out time and again that the early reductionistic logical behaviorism failed to produce an adequate and plausible construal of mentalistic concepts by explicit definition on the basis of purely *behavioral* concepts. ... I was tempted to identify, in the sense of *logical* identity, the mental with the neurophysiological ...

But if this theory is understood as holding a *logical translatability* (analytic transformability) of statements in the one language into statements in the other, this will certainly not do. ...

[T]he question which mental states correspond to which cerebral states is in *some* sense ... an empirical question. If this were not so, the intriguing and very unfinished science of psychophysiology could be pursued and completed by purely a priori reasoning. ...

... Subjective experience ... cannot be *logically* identical with states of the organism; i.e., phenomenal terms could not explicitly be defined on the basis of physical₁ or physical₂ terms. (1958, pp. 389-90).

Aside from encouraging the erroneous view that early reductionistic logical behaviourism was purely overt-behavioural, excluding reference to inner neurophysiological states, while his own early view (Feigl 1950) included them, Feigl runs together two crucially different things: *analyticity* and *definability*.¹⁷ He assumes that an explicit definition cannot be synthetic but can only be analytic and consequently assumes that abandoning the idea of explicit definition is tantamount to embracing the idea that the connection between what was originally the *definiendum* and *definiens* is synthetic.¹⁸ But these assumptions are mistaken.

According to Carnap and Hempel, if a non-primitive expression, the *definiendum*, is explicitly definable in terms of primitive expressions, then it can be eliminated and replaced by its *definiens*, by the primitive expressions. Such explicit definitions were understood by Carnap and Hempel to be the specification of necessary and sufficient conditions for the *definiendum*; that is, the construction of a material bi-conditional whose right-hand side, the *definiens*, contains only undefined primitive terms.¹⁹ For the logical positivists, of course, the defined expressions — in our case, mental ones

¹⁷ It must be conceded that Hempel (1935, §V) may not have been altogether free of this conflation either. See note 8 above.

¹⁸ Cf. Feigl (1958), pp. 427, 447, as well as Feigl (1963), p. 251 and Feigl (1971), p. 302. Pap (1952, p. 210) and Laurence D. Smith (1986, p. 53) also seem to hold these mistaken assumptions.

¹⁹ Strictly speaking, only the ultimate definition in a definition chain will have only undefined primitive terms in the *definiens*, but this does not affect the present point. See Carnap (1936-7) and Hempel (1952).

— will be so-called “theoretical” terms and the primitive expressions the “observation” terms — in our case, physical-behavioural ones. Now, Carnap (1936-37) very early on saw that the search for explicit definitions of all empirical scientific terms in the physical-thing language, on the basis of which physical translations could be carried out, was misconceived—especially in the case of dispositional terms—and consequently weakened the project to one of providing what he called “reduction sentences,” which were either material conditionals with further material conditionals as consequents or material conditions with material bi-conditionals as consequents.²⁰ These reduction-sentence conditionals linked the empirical term in question to physical conditions only under certain test circumstances. Since these physical reduction sentences were not definitions of the terms they were reducing—they were only incomplete “conditional definitions”—they did not allow the terms to be eliminated and replaced and hence they could not form the basis for translations.²¹

The important point for present purposes about this shift from definition to reduction or partial definition is that, with respect to the physicalization of psychology and other empirical sciences, it is not a shift from the category of analytic truths knowable only *a priori* to the category of synthetic truths knowable only *a posteriori*. Rather, it is a shift *within* the single category of synthetic truths knowable only *a posteriori* from complete definability (which permits elimination of the defined term) to incomplete or conditional definability (which does not permit elimination of the partially defined term). Contrary to what Feigl and others seem to suppose, the failure of explicit definition and translation is not at all tantamount to the failure of *a priori* analytic definition and translation. As we saw in §2, there never was any such latter project for psychology or indeed any empirical science.

The mistaken assumption that explicit definitions are all and only strongly intensional up to analyticity seems to have been abetted by misinterpretations of the addenda that Carnap and Hempel added to later reprintings of their respective articles. The addenda to Carnap (1932) and Hempel (1935) state that the two philosophers no

²⁰ See Carnap (1936-37), §10 and Hempel (1952).

²¹ Moreover, as Hempel makes clear, even if necessary and sufficient observational conditions for a theoretical term could be discovered inductively for a merely partially defined theoretical term introduced by reduction sentences, the bi-conditional representing this finding, ‘Q iff O’, where ‘Q’ is the theoretical term and ‘O’ the observational one, “clearly does not express a synonymy; if it did, no empirical investigations would be needed in the first place to establish it. Rather, it states that, as a matter of empirical fact, ‘O’ is co-extensive with ‘Q’, or, that O is an empirically necessary and sufficient condition for Q” (Hempel 1958, p. 192). But see note 8 above.

longer hold the strict definability thesis and have replaced it with the more flexible reducibility thesis.²² In his 1977 “prefatory note” to the reprinting in Block (1980), Hempel tells us that he had reservations about agreeing to the reprinting because he no longer held the “narrow translationist form of physicalism [there] set forth” but “yielded to Dr. Block’s plea that it offers a concise account of an early version of logical behaviourism” (p. 14). On Kim’s interpretation, this implies that “Hempel was in agreement with Block’s assessment that logical behaviourism was the position advocated in his 1935 paper” (2003, p. 266). Since Kim understands logical behaviourism as the thesis that psychological sentences analytically entail physical-behavioural sentences—in other words, since Kim holds the analytic entailment interpretation—he is claiming that Hempel is implying that he (Hempel) advocated the latter thesis in his original article. Kim goes on to point out how problematic Hempel’s note is *so interpreted* because hardly any of Hempel’s (1935) proffered physical-behavioural conditions are analytically entailed by his sample psychological sentence ‘Paul has a toothache’.²³ But there is no such implication. There is absolutely nothing in Hempel’s note to suggest he understood early logical behaviourism as the thesis that psychological sentences analytically entail physical-behavioural sentences. Kim’s interpretation can be arrived at only on the assumption that explicit definitions are analytic. But Hempel was never under any such illusion. On the contrary, he is clear that he understands his early version of logical behaviourism to be the claim that psychological concepts are explicitly definable in physical terms and his point is that he has now moved to the more liberal thesis of reduction. Both the earlier definitions and the later reductions were synthetic.²⁴ Carnap’s addendum makes exactly the same point.²⁵

²² See Carnap’s 1957 addendum to the reprinting of Carnap (1932) in Ayer (1959), Carnap’s 1961 addenda to the reprinting of Carnap (1932) in Alston and Nakhnikian (1963), Carnap’s preface to the second edition of the *Aufbau* (Carnap 1928/61), and Hempel’s 1972 ‘Author’s preamble’ to the reprinting of Hempel (1935) in Marras (1972, p. 115), which are all but identical to the 1977 ‘Author’s prefatory note’ to the reprinting in Block (1980). As Hempel notes in his addenda, physicalization was liberalized even further with the later introduction of “hypothetical constructs” connected to the observation language via “correspondence rules.” See also Carnap (1956) and Hempel (1951, 1952, 1958).

²³ See Crawford (2013) for more detail.

²⁴ Although, again, as discussed in note 8 above, Hempel (1935) is admittedly not entirely clear about this.

²⁵ Moreover, to come full circle, the first of the two main changes Carnap announces in the preface to the second edition of the *Aufbau* (the second being the one discussed by Chalmers and Leitgeb which I mentioned at the outset, namely, the shift from extensionality to either logical or nomological intensionality) is the “realization that the reduction of higher level concepts to lower level ones cannot

Echoing Donald Davidson's famous remark about language, I think we may conclude that there is no such thing as the logical behaviourism of the logical positivists, not if logical behaviourism is anything like what most philosophers have supposed.

REFERENCES

- Achinstein, P. and Barker, S.F., eds. (1969) *The Legacy of Logical Positivism. Studies in the Philosophy of Science*. Baltimore: The John Hopkins Press.
- Alston, W. and Nakhnikian, G., eds. (1963) *Readings in Twentieth Century Philosophy*. New York: The Free Press of Glencoe.
- Awodey, S. (2007) "Carnap's Quest for Analyticity: the *Studies in Semantics*", in M. Friedman and R. Creath, eds., *The Cambridge Companion to Carnap*. Cambridge: Cambridge University Press.
- Ayer, A. J., ed. (1959) *Logical Positivism*. New York: Glencoe.
- Block, N. (1978) "Troubles with Functionalism", in C. W. Savage, ed., *Minnesota Studies in the Philosophy of Science*, vol. 9 Minneapolis: University of Minnesota. Reprinted in Block (1980).
- Block, N., ed. (1980) *Readings in the Philosophy of Psychology*, Vol. 1, Cambridge, Mass.: Harvard University Press.
- Carnap, R. (1928/2003). *Der Logische Aufbau der Welt*. Berlin: Weltkreis-Verlag. Trans. *The Logical Structure of the World in The Logical Structure of the World and Pseudoproblems in Philosophy*, Chicago: Open Court.
- Carnap, R. (1932/1934). "Die physikalische Sprache als Universalsprache der Wissenschaft", *Erkenntnis* 2: 432-465. Trans. *Unity of Science*, London: Kegan Paul.
- Carnap, R. (1932/1959) "Psychologie in physikalischer Sprache", *Erkenntnis* 3: 102-142. Trans. 'Psychology in Physical Language' in Ayer (1959), pp. 165-198.
- Carnap, R. (1934/2002) *Logische Syntax der Sprache*, Vienna: Springer. Trans. *The Logical Syntax of Language*, Chicago: Open Court, 2002.

always take the form of explicit definitions. ... The positivist thesis of the reducibility of thing concepts to autopsychological concepts remains valid, but the assertion that the former can be defined in terms of the latter must now be given up and hence also the assertion that all statements can be translated into statement about sense data. Analogous considerations hold for the physicalist thesis of the reducibility of scientific concepts to thing concepts and the reducibility of heteropsychological concepts to thing concepts" (Carnap 1928/1961).

- Carnap, R. (1935/1963). *Philosophy and Logical Syntax*. London: Routledge and Kegan Paul. Reprinted, with addenda added by Carnap and terminological improvements suggested by him, in Alston and Nakhnikian (1963).
- Carnap, R. (1936-37). "Testability and Meaning", *Philosophy of Science* 3: 419-471 and 4: 1-40.
- Carnap, R. (1938/1991) "Logical Foundations of the Unity of Science", in Neurath et al, *Encyclopedia and Unified Science*, Chicago: University of Chicago Press, pp. 42-62. Reprinted in Richard Boyd, Philip Gasper and J. D. Trout (eds., *The Philosophy of Science*, Cambridge, Mass.: MIT Press, 1991.
- Carnap, R. (1939) *Foundations of Logic and Mathematics. Encyclopaedia of Unified Science*, vol. 1, no. 3. Chicago: University of Chicago Press.
- Carnap, R. (1942) *Introduction to Semantics*. Harvard: Harvard University Press.
- Carnap, R. (1947) *Meaning and Necessity*. Chicago: University of Chicago Press.
- Carnap, R. (1952) "Meaning Postulates", *Philosophical Studies* 3: 65-73.
- Carnap, R. (1955) "Meaning and Synonymy in Natural Languages", *Philosophical Studies* 6: 33-47.
- Carnap, R. (1964/1994) *An Introduction to the Philosophy of Science*, ed. by Martin Garder. New York: Dover.
- Carnap, R. (1956) "The Methodological Character of Theoretical Concepts", in H. Feigl and H. Scriven, eds., *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minneapolis: University of Minnesota Press, pp. 38-76.
- Carnap, R. (1963a) "Herbert Feigl on Physicalism", in Schilpp (1963), pp. 882-886.
- Carnap, R. (1963b). "Nelson Goodman on *Der logische Aufbau der Welt*", in Schilpp (1963), pp. 944-947.
- Chalmers, D. (2012) *Constructing the World*. Oxford: Oxford University Press.
- Chisholm, R. (1955-56) "Sentences about Believing", *Proceedings of the Aristotelian Society* 56: 125-148.
- Chisholm, R. (1957) *Perceiving. A Philosophical Study*. Ithaca, New York: Cornell University Press.
- Cirera, R. (1993) "Carnap's Philosophy of Mind", *Studies in the History and Philosophy of Science* 24: 351-58.
- Cornman, J. 1971. *Materialism and Sensations*. New Haven and London: Yale

- University Press.
- Crawford, S. (2013) "The Myth of Logical Behaviourism and the Origins of the Identity Theory", in M. Beaney (ed.) *The Oxford Handbook of the History of Analytic Philosophy*. Oxford: Oxford University Press.
- Creath, R., ed., (1992) *Dear Carnap, Dear Van: The Quine-Carnap Correspondence and Related Work*. Berkeley: University of California Press.
- Ducasse, C. (1941) *Philosophy as a Science. Its Matter and its Method*. New York: Oskar-Piest.
- Feigl, H. and W. Sellars, eds., (1949) *Readings in Philosophical Analysis* New York: Appleton Century Crofts.
- Feigl, H. (1950/1953) "The Mind-Body Problem in the Development of Logical Empiricism", *Revue de Internationale de Philosophie* 4. Reprinted in H. Feigl and M. Brodbeck, eds., *Readings in the Philosophy of Science*, New York: Appleton-Century Crofts, pp. 612-616.
- Feigl, H. (1958) "The 'Mental' and the 'Physical'", in H. Feigl, M. Scriven and G. Maxwell, eds., *Concepts, Theories and the Mind-Body Problem*. Minneapolis: University of Minnesota Press, pp. 370-497.
- Feigl, H. (1963) "Physicalism, Unity of Science and the Foundations of Psychology", in Schilpp (1963), pp. 227-268.
- Feigl, H. (1971) "Some Crucial Issues of Mind-Body Monism", *Synthese* 22: 295-312.
- Fodor, J. (1968) *Psychological Explanation*. New York: Random House.
- Goodman, N. (1963) "The Significance of *Der Logische Aufbau Der Welt*", in Schilpp (1963), pp. 545-558.
- Hempel, C.G. (1935/1972) "Analyse logique de la psychologie", *Revue de Synthèse* 10: 27-42. Trans. "The Logical Analysis of Psychology" in Feigl and Sellars (1949), pp. 373-384. Reprinted with prefatory note in Marras (1972) and Block (1980).
- Hempel, Carl G. (1951) "The Concept of Cognitive Significance: A Reconsideration", *Proceedings of the American Academy of Arts and Sciences* 80: 61-77.
- Hempel, Carl G. (1952) *Fundamentals of Concept Formation in Empirical Science*. Chicago: University of Chicago Press.
- Hempel, C.G. (1958) "The Theoretician's Dilemma", in Feigl, Scriven and Maxwell (eds.) *Minnesota Studies in the Philosophy of Science*, vol. 2; reprinted in Hempel (1965), pp. 173-228.
- Hempel C.G. (1969) "Logical Positivism and the Social Sciences", in Achinstein and

- Barker (1969), pp. 163-194.
- Hempel, C.G. 1965. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: The Free Press.
- Kim, J. (1971) "Materialism and the Criteria of the Mental", *Synthese* 22: 323-345.
- Kim, J. (2003) "Logical Positivism and the Mind-Body Problem", in P. Parrini, W. Salmon and M. Salmon, eds., (2003) *Logical Empiricism. Historical and Contemporary Perspectives*. Pittsburgh: University of Pittsburgh Press, pp. 263-280.
- Leitgeb, H. (2011) "New Life for Carnap's *Aufbau*?", *Synthese* 180: 265-299.
- Marras, A., ed., (1972) *Intentionality, Mind, and Language*. Chicago: University of Illinois Press.
- Pap, A. (1952) "Semantic Analysis and Psycho-Physical Dualism", *Mind* 61: 209-221.
- Pap, A. (1958) *Semantics and Necessary Truth*. New Haven: Yale University Press.
- Patterson, S. (2008) "Philosophy of Mind", in D. Moran, ed., *The Routledge Companion to Twentieth Century Philosophy*. London: Routledge.
- Putnam, H. (1963) "Brains and Behaviour", in R. Butler, ed., *Analytical Philosophy, Second Series*, Oxford: Blackwell, pp. 211-235..
- Putnam, H. (1969) "Logical Positivism and the Philosophy of Mind", in Achinstein and Barker (1969), pp. 211-227.
- Quine, W.V.O. (1951) "Two Dogmas of Empiricism", *The Philosophical Review* 60: 20-43.
- Quine, W.V.O. (1963) "Carnap on Logical Truth", in Schilpp (1963), pp. 385-406.
- Schilpp, P.A., ed., (1963) *The Philosophy of Rudolf Carnap*. La Salle, Illinois: Open Court.
- Searle, J. (2004) *Mind: A Brief Introduction*. Oxford: Oxford University Press.
- Smith, L.D. (1986) *Behaviourism and Logical Positivism. A Reassessment of their Alliance*. Stanford: Stanford University Press.
- Suppe, F (1977) "The Search for Philosophic Understanding of Scientific Theories", in Suppe, ed., *The Structure of Scientific Theories*, 2nd ed. University of Illinois Press, pp. 3-241.

Sean Crawford
Philosophy, School of Social Sciences
University of Manchester
Manchester M13 9PL
U.K.
sean.crawford@manchester.ac.uk