Available online at www.sciencedirect.com
SCIENCE @ DIRECT®

Stud. Hist. Phil. Sci. 36 (2005) 649–666

Studies in History
and Philosophy
of Science

www.elsevier.com/locate/shpsa

Introspection and change in Carnap's logical behaviourism

Allard Tamminga

*Vakgroep Theoretische Filosofie, Faculteit der Wijsbegeerte, Rijksuniversiteit Groningen,
Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands*

Received 2 November 2004; received in revised form 15 March 2005

Abstract

In the 1930s, Carnap set out to incorporate psychology into the unity of science, by showing that all cognitively meaningful sentences of psychology can be translated into the language of physics. I will argue that Carnap, relying on his notion of protocol languages, defends a physicalistic philosophy of psychology that shows due appreciation of 'introspection' as a strictly subjective, but reliable way to verify sentences about one's own mind. Second, I will point out that Carnap's philosophy of psychology not only takes into account overt behaviour, but must comprise neurophysiological processes as well. Last, I will show that Carnap aims to develop a philosophy of psychology that does justice to the ongoing changeability of scientific knowledge.

Keywords: Logical positivism; Philosophical behaviourism; Protocol sentences; Introspection; Theory change.

1. Introduction

In the early 1930s, several members of the Vienna Circle applied their unified science project to the philosophy of psychology. The result of this application came to be known as *logical behaviourism*.[1] According to Carnap and Neurath, the logical empiricist project of unified science could only be realized by means of a philosophical programme termed

E-mail address: A.M.Tamminga@rug.nl (A. Tamminga)

[1] The term was coined by Hempel (1949 [1935]), p. 381. Carnap cites Hempel's term as he quotes a passage from the unpublished German manuscript of Hempel's article in Carnap (1932), p. 187.

0039-3681/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsa.2005.08.004

'physicalism'. Hence, in characterizing logical behaviourism, we will have to start with a brief discussion of the neo-positivists's global empiricist programme, since we would miss the crux of their philosophy of psychology by treating it as an isolated position.

As we know, the neo-positivists aimed to rid the language that we use to formulate our knowledge of all cognitively meaningless elements, in order to curb the knowledge claims of metaphysicians and other philosophists once and for all: on closer inspection, their claims would turn out to be cognitively meaningless. In drawing up a criterion for cognitive meaningfulness, the Viennese philosophers relied on the theory of meaning of the early Wittgenstein, who noted in his *Tractatus*: 'To understand a proposition means to know what is the case if it is true' (4.024). The neo-positivists reformulated Wittgenstein's theory of meaning into a criterion for cognitive meaningfulness. In an early version their criterion was roughly as follows: a sentence is cognitively meaningful if and only if that sentence implies a non-empty set of sentences that state an elementary experience. Bona fide scientific statements simply should be able to pass this cognitive meaningfulness test, regardless of the scientific discipline to which they belong. In short, whether they stem from physics or from psychology, cognitively meaningful statements should always be confirmed in the same intersubjective manner:

> 'Unity of science' ... means essentially a unity of the confirmation basis of all factually cognitive (i.e., non-analytic) statements of the natural and the social sciences. (Feigl, 1963, p. 227)

After several attempts at formulating a sound criterion for cognitive meaningfulness were seen to have problematic consequences, Carnap tried to counter these problems by means of a universal and intersubjective system language, into which precisely all cognitively meaningful statements could be translated. This universal and intersubjective system language, Carnap argued in his 'Die physikalische Sprache als Universalsprache der Wissenschaft' (Carnap, 1995 [1931]), is the *physicalistic* language, the language of physics.

The neo-positivist programme of a physicalization of psychology was, first, a contribution to the debate on the supposedly essential differences between the sciences (*Naturwissenschaften*) and the humanities (*Geisteswissenschaften*).[2] This topic had been hotly debated in German philosophy since the 1890s. Dilthey and others held that the sciences and the humanities differed fundamentally and that, just as mechanics were considered as the basis of the sciences, so psychology had to be seen as the foundation of the humanities. A succesful physicalization of psychology would show such a position to be untenable.

Second, a successful physicalization of psychology would free scientific psychology from its precarious epistemological predicament. In his pamphlet *Einheitswissenschaft und Psychologie* (Neurath, 1987 [1933]), Neurath outlined the awkward position of empirical psychology at the time: every school in the new psychology uses its own terminology, in which metaphysical expressions keep popping up, in spite of its empirical and sometimes even explicitly anti-metaphysical aspirations. Four years earlier, Carnap, Hahn, and Neurath wrote: 'The linguistic forms which we still use in psychology today have their origin in certain ancient metaphysical notions of the soul' (Carnap, Hahn, & Neurath, 1973 [1929], p. 314). Metaphysical haggling between empirical psychologists talking at cross-purposes was

[2] See, for instance, Carnap (1995 [1931]), pp. 31–37, and (1963), p. 52.

commonplace. To put an end to this waste of time and energy, Neurath recommended that psychology should be incorporated into unified science, or that all sentences of psychology should be translated into the physicalistic language.

According to Carnap and Neurath, this translation ploy would separate the wheat from the chaff: cognitively meaningless psychological sentences are exposed as metaphysical word constructions, while for every cognitively meaningful sentence there exists a translation into the language of physics. Once this philosophical programme has been successfully completed, not only will the language of psychology be purged of metaphysics, but empirical psychologists of every persuasion will be able to concentrate on a fruitful exchange of opinions and arguments without being hindered by difficulties of translation:

> Behaviourists, gestalt psychologists, reflexologists, individual psychologists, and psychoanalysts will soon see their theories before them in the unified language of physicalism and will at last be able to compare them successfully. (Neurath, 1987 [1933], p. 22)

The initial enthusiasm with which the members of the Vienna Circle greeted early American behaviourism[3] was somewhat tempered during the 1930s. In 1933, Neurath denounced what, according to him, was the exaggerated opposition of the American behaviourists to 'inner observation', for 'it is hard to see why perceptions of our stomachs or other "inner" structures [Gebilde] should not in principle be just as admissible as the perceptions of our eyes or ears' (ibid., p. 16).[4] Neurath also condemned the tendency of some behaviourists to want to infer norms and values from the results of empirical research in psychology. He pointed out that a possible explanation for this propensity of Watson *cum suis* was that American behaviourists resorted to formulations that do not stand up to the test of physicalistic language criticism.[5] Likewise, Carnap regretted that the fundamental question of behaviourism 'is often formulated in the material mode of speech as a pseudo-object-question (e.g. "Do mental processes exist?", "Is psychology concerned only with physical behaviour?", and so on)' (Carnap, 1971 [1934], p. 324).[6] In his 'Analyse logique de la psychologie' (1949 [1935]), Hempel raised a more fundamental objection against

[3] In 1929, Carnap, Hahn, and Neurath emphasized the similarities between the American behaviourists's philosophy of psychology and their own position: 'The attempt of behaviorist psychology to grasp the psychic through the behavior of bodies, which is at a level accessible to perception, is, in its principled attitude, close to the scientific world-conception' (Carnap, Hahn, & Neurath, 1973 [1929], p. 315). However, in his *Behaviorism and logical positivism*, Laurence Smith concludes that 'the alliance of behaviorism and logical positivism was based not so much on genuine intellectual understanding as on relatively superficial convergences of opinion on broad issues and on matters of rhetoric and propaganda' (Smith, 1986, p. 18).

[4] Most logical empiricists considered 'introspection' to be fully compatible with physicalism. See Carnap (1959a [1932]), p. 192; Hempel (1949 [1935]), p. 381; Reichenbach (1970, [1938]), pp. 236–238; Neurath (1983a [1941]), p. 221.

[5] The Viennese neopositivists used the term 'behaviourism "in the wider sense"' to designate their philosophy of psychology. Neurath suggested replacing that term by 'behaviouristics', so as not to have constantly to justify the views of the American behaviourists (Neurath, 1987 [1933], p. 13). Apparently, only Neurath himself followed that suggestion. See, for instance, Neurath (1983a [1941]), pp. 214, 226.

[6] Carnap conceded, though, that if the 'fundamental question of behaviorism' is formulated in the formal mode, 'it will be seen that here again the question is one of reducibility of the psychological concepts; the fundamental thesis of behaviorism is thus closely allied to that of physicalism' (Carnap, 1971 [1934], pp. 324–325).

American behaviourism: he could not imagine that the truth of the American behaviourists's philosophy of psychology depended on its empirical success.[7] Therefore, a more thorough epistemological approach was required. The logical empiricists claimed that the translation of the languages of psychology into the universal and intersubjective language of physics would remedy these problems.

The present article proposes a systematic interpretation of the logical behaviourism of Carnap, who was the first logical empiricist to seriously undertake a detailed epistemological analysis of psychology. I will argue that in his philosophy of psychology, Carnap shows due appreciation to 'introspection' as a strictly subjective, but reliable way to verify sentences about one's own mind. Second, I will point out that Carnap's philosophy of psychology not only takes into account overt behaviour, but *must* comprise neurophysiological processes as well. Last, I will show that Carnap couples full awareness of the changeability of scientific knowledge with the aspiration to develop a philosophy of psychology that does justice to this changeability.

In order to illustrate Carnap's positions, I will occasionally draw from the work of other members of the Vienna Circle. It seems, however, less obvious to try and wrench a coherent, well-argued philosophy of psychology from the writings of Hempel, Neurath, or Schlick, because that work is mainly of an introductory or even propagandistic nature. Feigl is a somewhat special case. His philosophy of psychology was closely related to that of Carnap during the early 1930s. Later, Feigl's views evolved substantially—he moved to the United States in 1930—and culminated in his 1958 monograph *The 'mental' and the 'physical'* one of the benchmarks of twentieth-century philosophy of mind. A review of Feigl's philosophy of psychology is outside the scope of this article, as I will focus mainly on the philosophy of psychology of the Vienna Circle in the 1930s, and especially on the articulation and interpretation of Carnap's philosophy of psychology.

2. Carnap's philosophy of psychology

At the basis of Carnap's philosophy of psychology of the 1930s lies the distinction between *system languages* and *protocol languages*. A *system language* is a language used to articulate the states of affairs of which a particular science speaks; a *protocol language* is the language used by a particular subject S to record his 'experiences, perceptions, and feelings, thoughts' (Carnap, 1995 [1931], p. 43).[8] Carnap thinks that it is necessary for his epistemological investigations, though practically ineffective, to exclude the theory-laden statements from such a protocol language. Protocol sentences serve as the touch stone for sentences from a system language. A subject S can test (but never prove!) every cognitively meaningful sentence p, whether it is general or singular, by checking whether the protocol sentences that can be derived from p are among the protocol sentences that the subject S holds true.

[7] Hempel writes: 'It seems ... that the soundness of the behavioristic thesis ... depends on the possibility of fulfilling the program of behavioristic psychology. But one cannot expect the question as to the scientific status of psychology to be settled by empirical research in psychology itself. To achieve this is rather an undertaking in epistemology' (Hempel, 1949 [1935], p. 375). Cf. also Carnap (1959a [1932]), p. 181.

[8] Carnap fails to draw a clear distinction between a (formal) language and a theory expressed in that (formal) language.

The reader should notice that, in his writings on the philosophy of psychology, Carnap sticks to a subjectivistic and phenomenalistic interpretation of protocol sentences: protocol sentences bridge the gap between language and reality because they are the elementary sentences of the language used by a subject S to articulate the contents of his subjective experiences. Protocol sentences account for 'the simple fact, that everybody in testing any sentence empirically cannot do otherwise than refer finally to his own observations' (Carnap, 1936, pp. 423–424).[9] Protocol languages are subject-dependent, since 'protocol languages of various persons are mutually exclusive' (Carnap, 1995 [1931], p. 88).[10] Carnap had long been partisan to such a subjectivistic foundation of meaning and truth,[11] though in earlier studies he explicitly acknowledges the possibility of a physicalistic foundation.[12]

The *content* of a system sentence p for a subject S is the set of protocol sentences from S's protocol language which follow logically from p, given a sufficiently large set of already accepted singular physicalistic sentences, the inference rules of the system language involved, and the laws of nature.[13] If no protocol sentence of S can thus be derived from p, then p is for S a (cognitively) meaningless sentence: '[i]n such a case S cannot understand the statement p, for to "understand" means to know the consequences of p, i.e. to know the statements of the protocol language which can be deduced from p' (Carnap, 1995 [1931], p. 51). Given a group G of subjects $S_1, \ldots, S_n$, a system sentence p is intersubjectively meaningful (*intersubjektiv sinnvoll*) for G, if the cognitive content of p for every subject S_i in G is not empty. Hence, a system sentence p can be intersubjectively meaningful even if the non-empty contents which p has for different subjects in G do not coincide: one and the same system sentence p may have different contents for different subjects.[14] A system language is *intersubjective* for G if every sentence in this system language is intersubjectively meaningful for G.[15]

[9] Compare Carnap (1967 [1928]), § 2; Carnap (1987 [1932]), pp. 468–469 and Richardson (1998), p. 24.

[10] In his 'Über Protokollsätze' (1987 [1932]), Carnap defends his use of the expressions 'one's own protocol sentences' and 'others' protocol sentences' against Neurath's and Zilsel's objections (Carnap, 1987 [1932], p. 463). On the next page, Carnap even refers to protocol languages as 'private languages'. Compare Carnap (1932), p. 180.

[11] In Carnap's early work, which was conceived independently of the Vienna Circle, there are already the seeds of what, through the constitution theory in *Der logische Aufbau der Welt* (Carnap, 1967 [1928]), would develop into the physicalism on a phenomenalistic basis defended by Carnap at the beginning of the 1930s. In his 'Über die Aufgabe der Physik' (1923), Carnap outlines, as a *terminus ad quem* for contemporary research, an ideal, completed physics, consisting of three parts: (*a*) a set of axioms; (*b*) a phenomenalistic-physicalistic dictionary; (*c*) the description of the physicalistic state of the world at two arbitrary, but different moments in time (Carnap, 1923, pp. 96–103). The dictionary should bridge the gap between, on the one hand, the language which contains the phenomenalistic sentences with which we express the contents of our subjective experiences, and, on the other, the physicalistic language, which consists of the sentences with which the physicalistic states of affairs are formulated.

[12] See for example Carnap (1967 [1928]), § 59.

[13] Note that any subject S can only assess the content of a system sentence p relative to a set of previously accepted physicalistic sentences. See Carnap (1935), p. 44. The laws of nature are mentioned separately, since, in the early 1930s, Carnap still considered laws of nature as inference rules rather than as sentences to be used as premisses in an inference.

[14] Let p be a system sentence, π_1 and π_2 different protocol sentences and $G = \{S_1, S_2\}$. If content$S_1(p) = \{\pi_1\}$ and content$S_2(p) = \{\pi_2\}$, then p is intersubjectively meaningful for G, because the content of p for every subject S_i in G is not empty.

[15] Compare Carnap (1995 [1931]), pp. 51, 64.

Two system sentences p and q are mutually *translatable*, 'if there are rules, independent of space and time, in accordance with which q may be deduced from p and p from q' (Carnap, 1959a [1932], p. 166). This definition implies that if two system sentences p and q are mutually translatable, then for every subject S the content that p has for S is identical to the content that q has for S.[16] (Note that for each subject the contents of p and q may be identical, without it being the case that all subjects attach the very same content to p and q.[17]) Apparently, the converse holds as well:

> The possibility of such a deduction of protocol sentences constitutes the *content* of a sentence. ... If the same sentences may be deduced from two sentences, the latter two sentences have the same content. They say the same thing, and may be translated into one another. (Ibid.)

This passage requires some interpretation, since our above considerations showed that the content of a sentence is always the content of a sentence *for a subject* S. However, the quote does not make sense if for *only one* subject 'the latter two sentences have the same content'. To see this, suppose that only for one subject S the content of a system sentence p is identical to the content of a system sentence q. Then it is still possible that the system sentences in question are *not* mutually translatable: consider two subjects, S_1 and S_2, each with his own protocol language, so that the contents of p and q are identical for S_1 but not for S_2. If p and q were mutually translatable, they should have the same content for S_1 as for S_2, which in this example is obviously not the case. Hence, to interpret our passage correctly, we must require that p and q have the same content for *every* subject S. Summing up our renderings of Carnap's position on mutual translatability, *two system sentences p and q are mutually translatable if and only if for every subject S it holds that the content which p has for S is identical to the content which q has for S.*[18]

Note that mutual translatability in terms of subject-independent contents implies mutual translatability in terms of subject-dependent contents, but not vice versa. As we shall see, Carnap needs this weaker form of mutual translatability to account for the difference in verification procedures for sentences about our own mind and for sentences about other minds.

[16] Let the system sentences p and q be mutually translatable and let the protocol sentence π from the protocol language of subject S follow from p. Then π follows from p and p follows from q. Therefore, π follows from q. So, Carnap writes, translating 'gehaltgleich' with 'équipollent': '[i]f two sentences p and q have the *same content*, i.e., are mutually deducible, then obviously the protocol sentences [énoncés de contrôle] for p are identical to those for q' (Carnap, 1935, p. 44). This refutes a claim by Ramon Cirera: 'the rules of translation (or reduction) that the Carnapian construction uses do not observe any restriction of analyticity ... In other words, they do not retain the meaning of the sentences, which Carnap called the epistemic value (*Erkenntniswert*) in the *Aufbau*, but only the logical value (*logische Wert*), or the truth value' (Cirera, 1993, p. 355).

[17] Consider the following situation, figuring only two subjects, S_1 and S_2: content$S_1(p) = \{\pi_1\} =$ content$S_1(q)$, while content$S_2(p) = \{\pi_2\} =$ content$S_2(q)$. In this case, it still holds that for every subject S the content that p has for S is identical to the content that q has for S.

[18] This interpretation is confirmed by an example given in Carnap (1935), pp. 45–46. Moreover, it solves one of Ayer's problems with Carnap's philosophy of psychology: 'it seems strange to say that a statement has the same meaning for A and B when they attach different meanings to part of what it states' (Ayer, 1963, p. 276). Kim does not even mention subject-dependent protocol languages in his discussion of Carnap's position on translation (Kim, 2003, p. 273). Hence, Kim fails to notice that Carnap sets out to give an account of introspection within the framework of a physicalist philosophy of psychology.

A language L_1 is translatable into another language L_2 'if, for every expression of L_1, a definition is presented which directly or indirectly (i.e., with the help of other definitions) derives that expression from expressions of L_2' (Carnap, 1959a [1932], p. 167). A language is *universal* if every (cognitively meaningful) sentence from every other language can be translated into it.

The *physicalistic language*, by which Carnap means the language of physics, is, as yet,[19] constructed from elementary sentences of the type $F(x,y,z,t) = [a,b]$. Such an elementary sentence expresses the attribution of a value interval $[a,b]$ to a state quantity F of a point-instant (x,y,z,t)—for example, 'The temperature of point-instant (x,y,z,t) lies in the interval $[a,b]$'. Now that we have become familiar with the conceptual apparatus, we can acquaint ourselves with the central claim of Carnap's physicalism:

> It will be proved ... that the physical language is inter-subjective and can serve as a *universal* language, i.e. as a language in terms of which all states of affairs could be expressed. Finally, an attempt will be made to show that the various protocol languages also can be regarded as partial languages ... of the physical language. (Carnap, 1995 [1931], p. 52)

2.1. Psychology in physical language

The central thesis of Carnap's and Neurath's physicalism is that for each cognitively meaningful sentence in the system language of every special science there can be found a translation into the intersubjective and universal system language of physics. This obviously also applies to psychology, whose task will eventually be 'to describe systematically the (physical) behavior of living creatures, especially that of human beings, and to develop laws under which this behavior may be subsumed' (Carnap, 1959a [1932], p. 189). Consequently, Carnap argues that 'a definition may be constructed for every psychological concept (i.e. expression) which directly or indirectly derives that concept from physical concepts' (ibid., p. 167). As a result, all psychological 'laws' could also be expressed in the physical language, and hence become laws of physics. In spite of his 'concept reductionism', Carnap still thinks it is possible—although he would rather put his money on the opposite—that the physicalistic laws to which the organic domain obeys are *not* derivable from the physicalistic laws which apply to the anorganic domain: '[t]his question of the deducibility of the laws is completely independent of the question of the definability of concepts' (ibid.). Concept reduction does not automatically lead to theory reduction.[20]

Furthermore, Carnap explicitly does not consider himself a language reformer. He does not wish to outlaw the use of all psychological concepts and to replace them with purely physicalistic concepts.[21] Rather, he aims, given our current state of knowledge, to grasp the cognitive meaning of our psychological concepts as precisely as possible

[19] Carnap writes: 'We wish however to interpret the term "the physical language" so widely as to include not only the special linguistic forms of the present merely but also such linguistic forms as physics may use in any future stage of development' (Carnap, 1995 [1931], p. 54).

[20] See also Carnap, Hahn, & Neurath (1973 [1929]), p. 314; Carnap (1995 [1931]), pp. 69, 97–98; Carnap (1971 [1934]), p. 324.

[21] See Carnap (1995 [1931]), p. 95.

by anchoring these concepts via definitional procedures in the firm ground of the physicalistic language. According to Carnap, it is possible to bridge the gap between the psychological and the physicalistic language by the subject-dependent protocol languages, since they are the ultimate basis for assessing the meaning and truth of any system sentence.

The content of a universal sentence is completely determined by the contents of the singular sentences that can be deduced from it. Hence, in his logical analysis of psychological sentences, Carnap may focus, without loss of generality, on the investigation of singular psychological sentences. These singular sentences attribute a certain quality to a particular person at a particular moment, such as 'Yesterday afternoon Mr *B* was angry'. From an epistemological point of view, it is consequential which person utters such a singular psychological sentence. Hence, Carnap subdivides singular psychological sentences into two classes: sentences about other minds and sentences about one's own mind.

2.1.1. Sentences about other minds

According to Carnap, a sentence *p* about an other's mind, such as 'Mr *B* is angry now', can only be verified by deriving it from a universal sentence *O* expressing a certain law or regularity and a number of protocol sentences $\pi_1, \ldots, \pi_n$ that formulate observations about Mr *B*'s physical state or about his behaviour. (Carnap calls attention to the non-monotonous character of this type of inference: new protocol sentences can force us to retract an earlier drawn conclusion that *p*, for example, when we discover that Mr *B* is play-acting.)[22] The criterion for cognitive meaningfulness tells us that the content of *p* consists of the set of protocol sentences which we can infer from *p* and a set of physicalistic sentences, using logic and laws of nature. These protocol sentences can refer only to *our* observations of *B*'s physical state and behaviour. Given Carnap's above-cited claim that the subject-dependent protocol languages are also translatable into the physicalistic language, it must for each subject *S* be possible to exhaustively render the cognitive meaning of *p*—'Mr *B* is angry now'—by a physicalistic sentence *q*, that has for every subject exactly the same content as *p*. This physicalistic sentence *q* postulates 'a physical structure characterized by the disposition to react in a specific manner to specific physical stimuli. In our example, *q* asserts the existence of a physical structure (microstructure) of Mr *B*'s body (especially of his central nervous system) that is characterized by a high pulse and rate of breathing, which, on the application of certain stimuli, may even be made higher, by vehement and factually unsatisfactory answers to questions, by the occurrence of agitated movements on the application of certain stimuli, etc.' (Carnap, 1959a [1932], p. 172).

According to the criterion for cognitive meaningfulness, it must be possible to translate without remainder all psychological sentences about an other's mind into the language of physics. For, if we were to reject the principled possibility of translating our sentence *p*, via the set of protocol sentences deducible from *p*, into a physicalistic sentence *q*, then *p* would immediately degenerate into a metaphysical pseudo-sentence. Thus Feigl writes in his 'The psychophysical problem' (1934): 'To ascribe to our fellow men consciousness *in addition* to

[22] Cf. Hempel (1949 [1935]), p. 379. Compare Ayer (1963), pp. 277–278.

overt behavior and discoverable physiological processes implies . . . a transcendence, an introduction of empirically unverifiable elements' (Feigl, 1934, p. 424).[23]

2.1.2. Sentences about one's own mind

The use of the notion of a subject-dependent protocol language allows Carnap to account for the putative asymmetry between verification procedures for sentences about other minds and those for sentences about one's own mind: on the one hand, if we want to test a psychological sentence about someone else, we have to rely on the physical state of that person and the behaviour he displays, but on the other, we do not need to apply this indirect procedure in some cases involving the testing of a psychological sentence about our own mind.[24] Carnap elaborates these observations in his article 'Les concepts psychologiques et les concepts physiques sont-ils foncièrement différents?' (1935), making use of the following example.[25]

Let us use 'angry$_\psi$' to refer to the mental state of anger, the feeling of anger, or the state of consciousness called 'anger', and use 'angry$_\varphi$' to refer to the class of physical states known from experience in which a person's body is if and only if that person is angry$_\psi$. Consider the following sentences:

p 'Miss A is angry$_\psi$ now.'
q 'Miss A is angry$_\varphi$ now.'
r 'A person is angry$_\psi$ at time-point t if and only if this person is angry$_\varphi$ at time-point t.'

Carnap claims that his definition of the notion 'angry$_\varphi$' implies that r is a true *empirical* law.[26] Let $C = \{\pi_1, \ldots, \pi_n\}$ be the set of protocol sentences that can be deduced from q. For Miss A, the content of the system sentence p consists of the set of protocol sentences which are deducible from p: that is, in the first place, the sentence p itself (as Miss A can assess the truth of sentence p immediately through 'introspection', p for her is a protocol sentence[27]), but also the set C of protocol sentences that can be inferred from q—even though this indirect verification procedure may be somewhat unusual for Miss A—because q follows logically from p and r. Carnap writes:

> In this case, p and the sentences in the set C are for A protocol sentences [énoncés de contrôle] of p. Indeed, first, A can verify the the sentence p directly (by introspection, as one would say); and second, [she] can verify it indirectly as well, although [she]

[23] A little further on, Feigl argues: '[t]hat peculiar "plus," that "something more" which, it appeared, is needed in order to endow . . . "mere bodily behavior" with psychological relevance is factually meaningless' (Feigl, 1934, p. 425), and '[e]xperience which is in every respect "as if" it were dependent upon an apprehension of a transcendent reality is strictly identical with an experience which is "really" so. Similarly, an organism behaving in every way "as if it had a mental life" is simply identical with what we can possibly mean by an organism "really" having mental life' (Feigl, 1934, p. 426). Cf. Haller (1993), p. 194. See also Hempel (1949 [1935]), p. 379, and remember the Turing test.

[24] Cf. Carnap (1937), pp. 10–11.

[25] See Carnap (1935), pp. 45–47.

[26] Ayer remarks: '[h]ow Carnap knows that this empirical generalization is true he does not say. Neither does he discuss the philosophical difficulties which arise when one considers how it might be proved' (Ayer, 1963, p. 272).

[27] Note that Carnap usually distinguishes between system sentences and protocol sentences, even when their spellings are identical. Compare Carnap (1995 [1931]), p. 87.

> normally would perhaps not do so, since [she] has the possibility of a direct verification. (Carnap, 1935, p. 47—quote adapted to the present situation)

In short, for Miss A, the content of p consists of $\{p,\pi_1, \ldots ,\pi_n\}$. Conversely, the content of the system sentence q for Miss A does not only consist of C, but also of p, because p for her is a protocol sentence and because p follows logically from q and r. So for Miss A, the contents of p and q are identical and are equal to $\{p,\pi_1, \ldots ,\pi_n\}$.

For someone else, for example for Mr B, p is *not* a protocol sentence, because Mr B has no direct access to Miss A's mind: '[t]he sentence [p] is not directly verifiable by B; hence it is not a protocol sentence [énoncé de contrôle], although it is indirectly verifiable' (ibid.). For the verification of sentence p, Mr B has to rely on the set of protocol sentences $C = \{\pi_1, \ldots ,\pi_n\}$. Hence, for Mr B the content of p consists only of $\{\pi_1, \ldots ,\pi_n\}$, because $\{\pi_1, \ldots ,\pi_n\}$ is the set of protocol sentences that can be inferred from q and because q follows from p and r. The same holds for the content of q for Mr B: p is not a protocol sentence for Mr B, so for him the content of q only consists of $\{\pi_1, \ldots ,\pi_n\}$. Hence, for Mr B the contents of p and q are identical and are equal to $\{\pi_1, \ldots ,\pi_n\}$. Therefore, for Miss A and Mr B the contents of the sentences p and q are identical, even though their content for Miss A is not the same as their content for Mr B.[28] 'The difference', Carnap writes, 'only consists in that the sentence [p] is directly and indirectly verifiable by A, whereas it is only indirectly verifiable by B' (ibid.).

In short, Carnap used his subject-dependent protocol languages so that he could save 'introspection' within his physicalistic framework as an autonomous and independent method of verification for sentences about one's own mind. In the long run, however, this inventive solution would not survive the outcome of the debate on protocol sentences.

2.1.3. The debate on protocol sentences

At the time when the members of the Vienna Circle developed their logical behaviourism, they also debated about the truth and meaning of protocol sentences. In their view, protocol sentences formed the basis of all empirical knowledge: they reduced the truth and meaning of each system sentence p to the truth and meaning of protocol sentences that could be derived from p. Though, in their manifesto of 1929, Carnap, Hahn, and Neurath agreed that protocol sentences served to articulate 'the given',[29] they would later cross swords over the nature of the exact relation between protocol sentences and the given. Carnap preferred to ground protocol sentences in a phenomenally given, to be obtained via introspection. Neurath had serious reservations about the methodical solipsism championed by Carnap and about Carnap's corresponding interpretation of protocol sentences

[28] It seems that here, Carnap commits himself to a somewhat hybrid position with regard to protocol sentences: on the one hand there are purely subjective protocol sentences, and on the other hand intersubjective (objective?) protocol sentences. This hybrid position might be related to Fechner's theory of *psychophysical parallelism*: '[t]he theory suggests that each human being has double access to, or has two perspectives of, himself: When I am aware of myself in a way in which no one else can be aware of me, I am aware of mental processes. When I am aware of myself in a way in which other persons can also perceive me (for example, when I see myself in a mirror), then I see the same processes in a physical, objective form; I appear to myself as a physical, material being' (Heidelberger, 2003, p. 238). See Carnap (1967 [1928]), §§ 166–169.

[29] Carnap, Hahn, and Neurath wrote: 'Since the meaning of every statement of science must be statable by reduction to a statement about the given, likewise the meaning of every concept, whatever branch of science it may belong to, must be statable by step-wise reduction to other concepts, down to the concepts of the lowest level which refer directly to the given' (Carnap, Hahn, & Neurath, 1973 [1929], p. 309).

as 'statements needing no justification [Bewährung] and serving as foundation [Grundlage] for all the remaining statements of science' (Carnap, 1995 [1931], p. 45).[30] According to Neurath, this methodological position implies that the supposedly infallible foundation of all science eludes any attempt at intersubjective testing. Neurath rejected this methodological position[31]—and with that the entire notion of subject-dependent protocol languages—and instead argued for the assumption of a single intersubjective language, a language in which all cognitively meaningful sentences can be expressed:

> It is the physicalist language, *unified language*, that all science is about [das Um und Auf aller Wissenschaft]: no 'phenomenal language' beside the 'physical language'; no 'methodical solipsism' beside another possible standpoint ... ; *only unified science* with its laws and predictions. (Neurath, 1983c [1931], p. 68)

Even though Carnap did give in to Neurath's arguments, he initially still tried to reconcile both parties—phenomenalists and physicalists—by trying to show that 'this is a question, not of mutually inconsistent views, but rather of *two different methods for structuring the language of science both of which are possible and legitimate*' (Carnap, 1987 [1932], p. 457). Finally, but gradually, Neurath managed to convince Carnap to completely abandon his phenomenalistic interpretation of the basis of our knowledge. To begin, Carnap replaced, in his 'Testability and meaning' (1936 and 1937), his multitude of subject-dependent protocol languages with one single intersubjective 'thing-language'—a more secure basis for truth and knowledge.[32] The content of every system sentence is now assessed in relation to this unique intersubjective thing-language. Thus, the cognitive meaning of system sentences is no longer subject-dependent, as Carnap erroneously stated earlier, but subject-independent, and thus more or less 'objective'. The relations between psychological and physicalistic sentences suffer a similar fate, as Carnap assesses the relations between these system sentences through their cognitive meaning. If subsequently Carnap's famous distinction between the material and the formal mode of speech is disregarded, an identity theory in the vein of Feigl is within reach.[33] Indeed, Feigl uses the results of the debate on protocol sentences in an attempt to reconstruct logical behaviourism into a philosophy of psychology in which the notion of subject-dependent protocol languages no longer plays a part and in which therefore 'introspection' no longer ranks as an autonomous and independent method of verification for statements about one's own mind:

[30] Compare Carnap (1987 [1932]), p. 463.

[31] Neurath attacks this position mainly with sceptical arguments, though as early as 1934, he comes close to Popper's thesis of the theory-ladenness of perception: '[i]f one considers that in protocol statements ... terms of perception occur that are highly imprecise, that furthermore the content of protocol statements depends on the definition of these terms in the competent sciences, the ambiguity will not surprise us from the start' (Neurath, 1983b [1934], p. 106).

[32] However, in 'Testability and meaning', Carnap still sticks to the possibility of interpreting psychological predicates phenomenalistically: the distinction between (1) sentences about other minds and (2) sentences about one's own mind reappears as the difference between respectively a physicalistic and a phenomenalistic usage or interpretation of psychological predicates.

[33] In his 'Logical positivism and the mind–body problem', Kim concludes that 'we find in Carnap ... a form of physicalism that anticipates important later developments, in particular, functionalism and psychoneural identity theory based on the functionalist approach' (Kim, 2003, p. 277). Cirera remarks that Carnap's philosophy of psychology 'seems to point to a theory of psycho-physical identity' (Cirera, 1993, p. 357). Compare Carnap (1935), pp. 50–51.

> I claim to show ... that the strict identity of the 'mental' life with certain processes in the 'physical world' ... is not a matter of belief or Weltanschauung (dogmatic monism) but a truth capable of logical demonstration. In other words, the Duality of Mind and Matter does not imply two realities, or two 'aspects' of reality, but is merely a duality of languages or conceptual systems. (Feigl, 1934, pp. 420–421)

2.2. Time present and time future

Suppose that a critic of logical behaviourism raises the familiar objection that Miss A, through introspection, can know that she is angry, while for other observers remaining apparently unmoved. Evidently, the sentence 'Miss A is angry$_\psi$' can be accepted by Miss A and rejected by all others. Must we not, then, reject Carnap's claim that every psychological sentence can be translated into a physicalist sentence? Carnap's answer would be negative. First, he could point to the fact that although mutual translatability implies intersubjective meaningfulness, it does *not* imply intersubjective *validity*. Hence, in the example discussed earlier, figuring Miss A and Mr B, the sentence 'Miss A is angry$_\psi$' and its physicalist translation 'Miss A is angry$_\varphi$' are intersubjectively meaningful, but in the situation sketched by the critic they are not intersubjectively valid, since both sentences are accepted by Miss A and rejected by all others. This reply fits in with Carnap's analysis of the content of system sentences in terms of sets of subject-dependent protocol sentences. However, this cannot be the whole story. If it were, it would commit Carnap to the awkward position that there are some psychological statements that can be known to be true by a one person only. This would show his programme of physicalizing psychology to be fundamentally flawed.[34] Carnap explicitly denies such a position: 'every reasonable man, if not corrupted by philosophy, understands that these two statements [to wit, 'Miss A is not angry$_\psi$ now' and 'Miss A is angry$_\varphi$ now'] contradict each other, and that, as a consequence, it is impossible that they both are true at the same time' (Carnap, 1935, p. 52).

There is a second and more promising line of argument to counter the critic's objection: '[o]ur ignorance of physiology can ... affect only the mode of our characterization of the physical state of affairs in question. It in no way touches upon the principal point: that sentence p refers to a physical state of affairs' (Carnap, 1959a [1932], pp. 175–176). At present, so Carnap claims, we can indeed find a physicalist translation q for every psychological sentence p, such that for every subject S it holds that the content which p has for S is identical to the content which q has for S. As long as different subjects attach different sets of protocol sentences to p (and, hence, to q), it still remains possible that one and the same system sentence has different truth-values for different subjects. However, the more experimental-psychological research enables us to assess the necessary and sufficient truth conditions for the predicate 'x is angry$_\varphi$', the more the subject-dependent contents will coincide and, hence, the less differences of opinion there will be. Finally, when experimental research will have provided us with 'the exact conditions for anger$_\varphi$ referring to facial

[34] Compare Carnap (1987 [1932]), p. 468. In this passage, Carnap reformulates the asymmetry between verification procedures for sentences about other minds and those for sentences about one's own mind by giving a physicalist translation of the sentence stating the asymmetry: '"Only S is immediately aware of his hunger" means: "Only S is able to make the statement 'S is hungry' directly on the basis of hunger, i.e., with no physical causal connection with processes outside of S's body"' (Carnap, 1987 [1932], p. 368). Note that this physicalist translation does not rule out the situation sketched by our critic.

expressions, to gestures, to pulsations of the heart, to respiration, to the nervous system, etc.',[35] the physicalization of our psychological concept is complete. As 'physical determinations are valid intersubjectively' (Carnap, 1995 [1931], p. 65), we will then be able to expose every (conscious or unconscious) impostor. The previous discussion shows that Carnap is committed to the incorporation of micro-physiological processes in his philosophy of psychology. A logical behaviourism confining itself to overt, publicly observable behaviour would, as we have argued, seriously compromise physicalism.

Carnap envisages the search for a physicalistic sentence q with the same cognitive content as the given psychological sentence p roughly as follows: (1) We define q_1 as the sentence 'The state of Miss A's body is part of the class of physical states K', K being the currently known class of physical states in which Miss A finds herself if and only if she is angry$_\psi$. Because q_1 only contains physicalistic predicates, it is a physicalistic sentence. Moreover, q_1 has *by definition* the same cognitive content as p, as q_1 reflects the current state of our psychological knowledge. (2) Ongoing psychological research into the class of physical states in which Miss A's body is if and only if Miss A is angry$_\psi$, may require us to adjust q_1 to this new state of psychological knowledge. This adjustment will result in a more accurate sentence q_2. Hence, given this new state of knowledge, there still is a physicalistic sentence with exactly the same cognitive content as p. (Note that the content of p not only depends on logic and the laws of nature, but on a sufficiently large set of already accepted sentences as well. It is plausible that the new state of knowledge necessitates alterations in this set of sentences or in the laws of nature, thereby changing the content of p.) By carefully following this procedure, we will finally find a sentence q_n that exactly states the necessary and sufficient physical conditions for p: the physicalization of p has been completed. Let us now take a closer look at these two points.

Ad (1). If Carnap is to be believed, it would already in the 1930s have been possible to find for every psychological sentence p a physicalistic sentence q with exactly the same cognitive content, regardless of the relatively inadequate state of knowledge in empirical psychology:

> Even today every sentence of psychology *can* be translated into a sentence which refers to the physical behavior of living creatures. In such a physical characterization terms do indeed occur which have not yet been physicalized, i.e. reduced to the concepts of physical science. Nevertheless, the concepts used *are* physical concepts, though of a primitive sort. (Carnap, 1959a [1932], p. 183)[36]

Apparently, Carnap is enough of a pragmatist not to insist on definitions that will attain unscathed the end of times. He simply claims that it is already possible to state—for the time being—adequate definitions of psychological terms using 'primitive' physicalistic terms: '[w]e maintain that these definitions can be produced, since, implicitly, they already

[35] Carnap (1935), p. 45.

[36] See also Carnap (1959a [1932]), p. 175. Hempel and Schlick maintain more or less the same thing. Hempel writes: 'In order for logical behaviorism to be acceptable, it is not necessary that we be able to describe the physical state of a human body which is referred to by a certain psychological statement . . . down to the most minute details of the phenomena of the central nervous system' (Hempel, 1949 [1935], p. 381). Schlick writes: 'we do not need to consider the events in the nervous system—which are for the most part unknown—for it is sufficient to pay attention to his expression, his utterances, his whole deportment. In these processes . . . we have the facts by which feelings are expressible in the physical language' (Schlick, 1949 [1935], p. 402).

underlie psychological practice' (ibid. p. 167).[37] Measured against the present state of knowledge, all proposals for such definitions are simply either accurate or inaccurate.[38]

It is true that the present (incomplete) knowledge of psychology does not yet allow for an exhaustive characterization of psychological terms in terms of neurophysiological states of, for example, Miss *A*. That, however, is no reason to deny our psychological sentences any content; Carnap argues that, for the time being, we will have to settle for a characterization of their content in terms of the information that is presently available about bodily movements, facial expressions and linguistic utterances—in short, the behavioural dispositions—of Miss *A*. This characterization is provisional and will, of course, be much less precise than the final characterization, which will incorporate all information about the neurophysiological states of Miss *A* which are not yet available. Ultimately, scientific research will enable us to replace all our present 'primitive, intuitive concept formations' by (constructions of) physicalistic state quantities, just as has already happened with former pre-theoretical proto-physicalistic concepts such as 'warm' and 'green'.[39]

Ad (2). Advancements in empirical research will force us to continually adapt the physicalistic sentence q to the latest state of knowledge. However, it is rather impractical to restate sentence q after each significant advancement in the relevant sciences: '[i]f we now were to state a definition, we should have to revoke it at a new stage of the development of science, and to state a new definition, incompatible with the first one' (Carnap, 1936, p. 449). As Carnap shows in a famous analysis, such adjustments of what was already known about the physicalistic content of the psychological sentence p should be seen as *additions* rather than as *corrections*.

The starting point of this analysis is Carnap's claim that all current psychological properties are *dispositions*.[40] It is well-known that a disposition term '$Q(x)$' (for example, 'x is soluble in water') cannot simply be defined as '$R_1(x) \rightarrow R_2(x)$' ('If x is put in water, then x will dissolve').[41] In 'Testability and meaning', Carnap develops a logical instrument—the reduction sentence—to deal with disposition terms more adequately. A reduction sentence for our disposition term '$Q(x)$' is:

$$R_1(x) \rightarrow [R_2(x) \rightarrow Q(x)]$$

Unlike the definition of $Q(x)$ as $R_1(x) \rightarrow R_2(x)$, this reduction sentence does not tell us anything about the truth value of $Q(x)$ in cases in which $\neg R_1(x)$ holds. Opposing his 1932 defense of the explicit definability of psychological concepts and the corresponding

[37] Cf. Carnap (1995 [1931]), pp. 85–86.

[38] von Kutschera rightly remarks: 'Carnap [claims] that psychological terms are *explicitly definable* in physical terms. These definitions cannot be nominal definitions, in which the definiendum is a novel expression getting meaning only via the definition, as psychological expressions already do have a meaning. Hence, in opposition to nominal definitions, the definitions considered here are not stipulations, but they are right or wrong' (von Kutschera, 1991, p. 307).

[39] Cirera argues that '[Carnap's] words suggest that neurophysiological reduction would be the most suitable; and that behavioural reduction ... is no more than a consolation prize, a purely provisional resource, given our cognitive limits' (Cirera, 1993, p. 356). Kim writes that 'Carnap claims that as our knowledge of neurophysiology grows ... , the behavioral definitions of psychological properties and states will give way to physiological definitions' (Kim, 2003, p. 275).

[40] Carnap asserts: 'Every psychological property is marked out as a disposition to behave in a certain way' (Carnap, 1959a [1932], p. 186).

[41] The material implication is the culprit. In fact, for all objects that have *not* been put in water, the definiens is true, and therefore so is the definiendum.

sequence of corrections of earlier definitions, Carnap suggests in 'Testability and meaning' that we analyse the specification of the meaning of the disposition term '$Q(x)$' by means of a growing set of reduction sentences:

> Suppose that we introduce a predicate 'Q' into the language of science first by a reduction pair and that, later on, step by step, we add more such pairs for 'Q' as our knowledge about 'Q' increases with further experimental investigations. In the course of this procedure the range of indeterminateness for 'Q', i.e. the class of cases for which we have not yet given a meaning to 'Q', becomes smaller and smaller. (Carnap, 1936, p. 448)
>
> Only if we reach, by adding more and more reduction pairs, a stage in which all cases are determined, may we go over to the form of a definition. (Ibid. p. 450)

In short, a physicalist *definition* of a psychological predicate is untenable as long as developments in the relevant sciences have not yet yielded all the required reduction pairs. Hence, Carnap considerably nuances his original translation programme for psychology with the introduction of reduction sentences. This conceptual tool enables Carnap also to better answer the objection raised at the beginning of this section. Armed with his reduction sentences, Carnap can now grant that our current state of psychological knowledge does not warrant physicalist definitions of psychological terms. New results from empirical-psychological research can now, in the form of reduction sentences, smoothly be incorporated in our previous body of partial physicalizations of psychological sentences. Earlier findings concerning relations between psychological sentences and sentences describing overt, generally accessible behaviour can simply be retained. In Carnap's new theoretical account of scientific progress in psychology, definitions of psychological terms hardly play a role: if definitions of psychological terms are not relegated to the realm of fantasy, then to a Peircean *final opinion* in some distant future.[42]

3. Conclusion

Carnap's logical behaviourism takes into account both the current state of knowledge and a *final opinion* in the distant future. His philosophical psychology allows for three physicalistic research programmes: two static programmes and one dynamic programme. Since the beginning of the twentieth century a dynamic conception of science in all its diversity has become generally accepted. Nevertheless, traditional epistemology, with its static ideal of knowledge, in which necessity is still conceived as a characteristic of truth, largely dominated this century's philosophy of mind. It is little wonder then, that twentieth-century philosophy of mind has focused its attention mainly on both static programmes.

With the first static programme, one may try to show that even nowadays, each psychological sentence can already be reduced to a set of sentences describing the subject's overt,

[42] Carnap remarks: 'The so-called thesis of *Physicalism* asserts that every term of the language of science—including beside the physical language those sub-languages which are used in biology, in psychology, and in social science—is reducible to terms of the physical language. Here a remark … has to be made. We may assert reducibility of the terms, but not—as was done in our former publications—definability of the terms and hence translatability of the sentences' (Carnap, 1936, p. 467).

generally accessible behaviour, instead of his or her neurophysiological states. Ryle, Quine and Dennett are more recent representatives of this philosophical school.

In the second static programme, which has been crucial for twentieth-century philosophy of mind, the central question is how psychological and physicalistic sentences—or, in the material mode of speech, how psychological and physicalistic phenomena—will relate to each other once we have at our disposal perfect and complete physicalistic theories with which to explain and predict behaviour. Almost everyone works within this second research programme, whatever position they take up in the debate on reductionism, functionalism or eliminativism.

On the other hand, the dynamic research programme couples full awareness of the changeability of scientific knowledge with the aim to develop a philosophy of psychology that does *justice* to this changeability. Over the last twenty years, Nagel's standard account of reduction of completed and axiomatized theories has been replaced by accounts of reduction of changing and provisional theories.[43] As we have seen, Carnap already defends a comparable dynamic programme in the 1930s.[44] Within the overall framework of physicalism, he constructed a set of conceptual tools to make intelligible the permanent revision of physicalist interpretations of psychological terms on the basis of the latest scientific developments.[45] At the same time, Carnap attempted to retain 'introspection' as an independent source of knowledge. This latter attempt ultimately ran aground on the outcome of the debate on protocol sentences, with which the first-person perspective disappeared beneath the horizon of logical behaviourism. Out of sight, out of mind?

Acknowledgements

I wish to thank Michel ter Hark, Leon Horsten, Huib Looren de Jong, Jeanne Peijnenburg, and an anonymous referee of this journal for their comments. An earlier version of the present article was published in Dutch as 'Het logische behaviorisme van Rudolf Carnap', *Tijdschrift voor Filosofie, 65* (2003), 541–565. Walter van der Star made a translation of the Dutch text into English, which was revised and extended by the author.

References

Ayer, A. J. (1963). Carnap's treatment of the problem of other minds. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 269–281). La Salle, IL: Open Court.

Bickle, J. (1998). *Psychoneural reduction*. Cambridge, MA: The MIT Press.

[43] See, for instance, Wimsatt (1976), McCauley (1986), Bickle (1998).

[44] Twenty years later, Carnap gives up his dispositional interpretation of terms from scientific psychology. Though he thinks theoretical terms unsuitable for dispositional analysis (see Carnap, 1956), he proposes to interpret the terms from scientific psychology as theoretical terms. Such interpretations will never be complete (Carnap, 1959 [1957], p. 197). A few years later, Carnap claims that everyday psychological terms should be analysed as theoretical terms as well (Carnap, 1998 [1961], p. XXI).

[45] In his 'How to define theoretical terms', David Lewis defends a type of functionalism that is similar to Carnap's dynamic programme. In Lewis's proposal, however, interpretations of T-terms—the theoretical terms of a theory T—cannot be partial. Hence, the question 'If T is ... partially reduced and partially falsified, or revised for any other reason, do T-terms retain their meanings?' (Lewis, 1983, p. 94) becomes pressing. Notice that Carnap's partial interpretations of a disposition in terms of a set of reduction sentences allow for consistent extensions of meaning in which the original meaning of the disposition is retained.

Carnap, R. (1923). Über die Aufgabe der Physik und die Anwendung des Grundsatzes der Einfachstheit. *Kant-Studien*, *28*, 90–107.

Carnap, R. (1932). Erwiderung auf die vorstehenden Aufsätze von E. Zilsel und K. Duncker. *Erkenntnis*, *3*, 177–188.

Carnap, R. (1935). Les concepts psychologiques et les concepts physiques sont-ils foncièrement différents? Revue de Synthèse, *10*, 43–53.

Carnap, R. (1936). Testability and meaning. *Philosophy of Science*, *3*, 419–471.

Carnap, R. (1937). Testability and meaning—continued. *Philosophy of Science*, *4*, 1–40.

Carnap, R. (1956). The methodological character of theoretical concepts. In H. Feigl, & M. Scriven (Eds.), *The foundations of science and the concepts of psychology and psychoanalysis* (pp. 38–76). Minneapolis: University of Minnesota Press.

Carnap, R. (1959a). Psychology in physical language (G. Schick, Trans.). In A. J. Ayer (Ed.), *Logical positivism* (pp. 165–197). New York: The Free Press. (First published as Psychologie in physikalischer Sprache. *Erkenntnis*, *3* (1932), 107–142)

Carnap, R. (1959b). Remarks by the author. In A. J. Ayer (Ed.), *Logical positivism* (pp. 197–198). New York: The Free Press. (First published 1957)

Carnap, R. (1963). Intellectual autobiography. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 1–84). La Salle, IL: Open Court.

Carnap, R. (1967). *The logical structure of the world*. Berkeley: University of California Press. (First published as *Der logische Aufbau der Welt*. Berlin: Weltkreisverlag, 1928. Reprinted 1998)

Carnap, R. (1971). *The logical syntax of language*. London: Routledge & Kegan Paul. (First published as *Logische Syntax der Sprache*. Vienna: Springer, 1934. Reprinted 1968)

Carnap, R. (1987). On protocol sentences. *Noûs*, *21*, 457–470. (First published as Über Protokollsätze. *Erkenntnis*, *3* (1932), 215–228)

Carnap, R. (1995). *The unity of science* (M. Black, Trans.). Bristol: Thoemmes Press. (First published as Die physikalische Sprache als Universalsprache der Wissenschaft. *Erkenntnis*, *2* (1931), 432–465)

Carnap, R. (1998). Vorwort zur zweiten Auflage. In R. Carnap, *Der logische Aufbau der Welt* (pp. XVII–XXIV). Hamburg: Felix Meiner Verlag. (First published 1961. Translated as Preface to the second edition. In R. Carnap, *The logical structure of the world* (pp. V–XI). Berkeley: University of California Press, 1967)

Carnap, R., Hahn, H., & Neurath, O. (1973). The scientific conception of the world: The Vienna Circle. In M. Neurath, & R. S. Cohen (Eds.), *Otto Neurath: Empiricism and sociology* (P. Foulkes, & M. Neurath, Trans.) (pp. 299–318). Dordrecht: Reidel. (First published as *Wissenschaftliche Weltauffassung: Der Wiener Kreis*. Vienna: Artur Wolf, 1929)

Cirera, R. (1993). Carnap's philosophy of mind. *Studies in History and Philosophy of Science*, *24A*, 351–358.

Feigl, H. (1934). Logical analysis of the psychophysical problem. *Philosophy of Science*, *1*, 420–445.

Feigl, H. (1963). Physicalism, unity of science and the foundations of psychology. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 227–267). La Salle, IL: Open Court.

Haller, R. (1993). *Neopositivismus: Eine historische Einführung in die Philosophie des Wiener Kreises*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Heidelberger, M. (2003). The mind–body problem in the origin of logical empiricism. In P. Parrini, W. C. Salmon, & M. H. Salmon (Eds.), *Logical empiricism: Historical and contemporary perspectives* (pp. 233–262). Pittsburgh: University of Pittsburgh Press.

Hempel, C. G. (1949). The logical analysis of psychology (W. Sellars, Trans.). In H. Feigl, & W. Sellars (Eds.), *Readings in philosophical analysis* (pp. 373–384). New York: Appleton-Century-Crofts. (First published as Analyse logique de la psychologie. *Revue de Synthèse*, *10* (1935), 27–42)

Kim, J. (2003). Logical positivism and the mind–body problem. In P. Parrini, W. C. Salmon, & M. H. Salmon (Eds.), *Logical empiricism: Historical and contemporary perspectives* (pp. 263–278). Pittsburgh: University of Pittsburgh Press.

Lewis, D. (1983). How to define theoretical terms. In D. Lewis, *Philosophical papers, Vol. 1* (pp. 78–95). New York: Oxford University Press. (First published 1970).

McCauley, R. N. (1986). Intertheoretic relations and the future of psychology. *Philosophy of Science*, *53*, 179–199.

Neurath, O. (1983a). Universal jargon and terminology. In R. S. Cohen, & M. Neurath (Eds.), *Otto Neurath: Philosophical papers 1913–1946* (pp. 213–229). Dordrecht: Reidel. (First published in 1941)

Neurath, O. (1983b). Radical physicalism and the 'real world'. In R. S. Cohen, & M. Neurath (Eds.), *Otto Neurath: Philosophical papers 1913–1946* (pp. 100–114). Dordrecht: Reidel. (First published as Radikaler Physikalismus und 'wirkliche Welt'. *Erkenntnis*, *4* (1934), 346–362)

Neurath, O. (1983c). Sociology in the framework of physicalism. In R. S. Cohen, & M. Neurath (Eds.), *Otto Neurath: Philosophical papers 1913–1946* (pp. 58–90). Dordrecht: Reidel. (First published as Soziologie im Physikalismus. *Erkenntnis*, *2* (1931), 393–431)

Neurath, O. (1987). Unified science and psychology. In B. McGuinness (Ed.), *Unified science* (pp. 1–23). Dordrecht: Reidel. (First published as *Einheitswissenschaft und Psychologie*. Vienna: Gerold, 1933)

Reichenbach, H. (1970). *Experience and prediction*. Chicago: The University of Chicago Press. (First published 1938).

Richardson, A. W. (1998). *Carnap's construction of the world*. Cambridge: Cambridge University Press.

Schlick, M. (1949). On the relation between psychological and physical concepts. In H. Feigl, & W. Sellars (Eds.), *Readings in philosophical analysis* (pp. 393–407). New York: Appleton-Century-Crofts. (First published as De la relation entre les notions psychologiques et les notions physiques. *Revue de Synthèse*, *10* (1935), 5–26)

Smith, L. D. (1986). *Behaviorism and logical positivism*. Stanford: Stanford University Press.

von Kutschera, F. (1991). Carnap und der Physikalismus. *Erkenntnis*, *35*, 305–323.

Wimsatt, W. C. (1976). Reductionism, levels of organization, and the mind–body problem. In G. Globus, G. Maxwell, & I. Savodnik (Eds.), *Consciousness and the brain* (pp. 199–267). New York: Plenum Press.

Wittgenstein, L. (1963). *Tractatus logico-philosophicus* (D. Pears, Trans.). London: Routledge & Kegan Paul. (First published in 1922)