

THE MYTH OF LOGICAL BEHAVIOURISM AND THE ORIGINS OF THE IDENTITY THEORY

Sean Crawford

The identity theory's rapid rise to ascendancy in analytic philosophy of mind during the late 1950s and early 1960s is often said to have constituted a sea change in perspective on the mind-body problem. According to the standard story, logical or analytical behaviourism was analytic philosophy of mind's first original materialist-monist solution to the mind-body problem and served to reign in various metaphysically extravagant forms of dualism and introspectionism. It is understood to be a broadly logico-semantic doctrine about the meaning or definition of mental terms, namely, that they refer to dispositions to engage in forms of overt physical behaviour. Logical/analytical behaviourism then eventually gave way, so the standard story goes, in the early 1960s, to analytical philosophy's second original materialist-monist solution to the mind-body problem, the mind-brain identity theory, understood to be an ontological doctrine declaring states of sensory consciousness to be physical states of the brain and wider nervous system. Of crucial importance here is the widely held notion that whereas logical behaviourism had proposed an identity between the meanings of mental and physical-behavioural concepts or predicates—an identity that was ascertainable *a priori* through conceptual analysis—the identity theory proposed an identity between mental and physical properties, an identity that could only be established *a posteriori* through empirical scientific investigation. John Searle (2004) has recently described the transition thus:

[logical behaviourism] was gradually replaced among materialist-minded philosophers by a doctrine called “physicalism,” sometimes called the “identity theory.” The physicalists said that Descartes was not wrong, as the logical behaviourists had claimed, as a matter of logic, but just as a matter of fact. ... The identity theorists were anxious to insist on the contrast between their view and behaviourism. Behaviourism was supposed to be a logical thesis about the definition of mental concepts. The identity thesis was supposed to be a factual claim, not about the analysis of mental concepts, but rather about the mode of existence of mental states. 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. The identity theorists said no, the model is not definitions, but rather empirical discoveries of identities in science (pp. 54-5).

Searle cites Gilbert Ryle and C. G. Hempel as “famous” logical behaviourists (he could easily have added Rudolf Carnap as another) and U. T. Place, J. J. C. Smart and Herbert Feigl as identity theorists. The alleged insight of the great triumvirate of identity theorists is that materialism, if true, cannot be an *a priori* knowable semantic thesis about the meanings of mental and physical terms, but must take the form of an empirical ontological thesis about the mental and physical realms.

While there is a grain of truth in this very familiar historical take on the transition from logical behaviourism to the identity theory—it was indeed a shift away from a focus on language and concepts characteristic of analytic philosophy’s more general “linguistic turn,” to a focus on ontology—the story is much more an instance of what Richard Watson (1993) calls “shadow history of philosophy”:

The shadow history of philosophy is a kind of received view consisting of stories of philosophy that most philosophers accept even though they know that these stories are not really quite precisely right. ... The presumption is that they are basically right, that the pictures they present display the important logical or conceptual guts of history like a medical diagram, in an ideal way, without the mess of the real thing. ... They are important as the bases from which philosophers derive their systems, either by development or opposition. ... These shadow positions ... are at least as substantial and influential in the development of philosophy as the “true” positions they are shadows of. Shadow histories provide indispensable foundations for philosophy (pp. 97, 107-108, 109).

The received view of logical behaviourism just outlined is a shadow position and the story of its overthrow by its successor, the mind-brain identity theory, is shadow history in Watson’s sense—at least so I shall argue. More specifically, I will argue that the

difference between one form of what is misleadingly called “logical behaviourism”—namely, the logical positivists’ logical behaviourism—and the identity theory, has been misunderstood and its significance consequently overrated and exaggerated. In the next section I will try to demonstrate this in detail by looking closely at some of the works of the logical positivists, Carnap’s in particular, and by placing these works in the larger system of their thought. After setting the historical record straight, I will go on to explore in the following section the origins of the identity theory and the seldom discussed difference between the two different versions that succeeded logical behaviourism. I will focus on Herbert Feigl’s less familiar and rather puzzling identity theory because, in the first place, compared with Smart’s and Place’s, it has received very little critical attention and, secondly, its striking difference from Place’s and Smart’s theory is, I believe, a manifestation of a deep and perennial opposition in modern philosophy of mind, one that we are witnessing today in the debate over the so-called hard problem of phenomenal consciousness.¹

1. The Myth of Logical Behaviourism

1.1 Two Logical Behaviourisms

According to the shadow history of philosophy of mind, retold to generations of undergraduates, and reported in countless textbooks, anthology introductions, and encyclopaedias, there was something called “logical behaviourism” and it was

¹ My scope in this chapter is thus very limited. More comprehensive recent surveys of the history of analytic philosophy of mind, covering much more than the mind-body problem, can be found in Burge (2007) and Patterson (2008).

overthrown at the end of the 1950s by the mind-brain identity theory. Together with Ryle's *The Concept of Mind* (1949), Carnap's 'Psychology in Physical Language' (1933) and Hempel's 'The Logical Analysis of Psychology' (1935) are taken virtually universally to be the three canonical texts of logical behaviourism. Now, to begin with, it should be obvious on reflection that the idea that there was a single doctrine known as logical behaviourism and that it was an early twentieth century form of materialism is obviously a shadow position. For logical behaviourism was in fact associated with two very different movements in analytic philosophy, ordinary language philosophy and logical positivism. Moreover, in neither form was it a type of materialism. In the eyes of both movements, the mind-body problem is a pseudo-problem to be dissolved or replaced by linguistic analysis. All traditional "-ism" solutions to it—dualism and materialism alike—are metaphysical pseudo-doctrines. So logical behaviourism is not really a form of materialism in the traditional sense. To be sure, it was for the logical positivists a form—or more accurately, an application—of what they called "Physicalism."² But, again, the Physicalism of the logical positivists is not a metaphysical doctrine; on the contrary, it is an anti-metaphysical doctrine. The logical positivists were explicit and adamant about this, repeating it tirelessly in an often vain effort to avoid misunderstandings.

"Physicalism," Carnap (1935/1963) tells us, "... has nothing to do with any such theses as monism, dualism or parallelism"; and although Physicalism "is allied to that of *Materialism* ... the agreement extends only as far as the logical components of *Materialism*; the metaphysical components, concerned with the question of whether the essence of the world is material or spiritual, are completely excluded from consideration"

² The term was coined by Neurath (1931).

(p. 459). Physicalism is a doctrine about the language of science. It is the thesis that the only kind of language known to be capable of providing the necessary inter-subjective, inter-sensory and universal confirmation base for empirical science is physical language. So understood Physicalism (with a capital ‘P’) is crucially part of the “Unity of Science” movement initiated by the logical positivists in the early 1930s, the central aim of which was to which ensure that all of the theories and hypotheses of all empirical sciences were subject to rigorous inter-subjective confirmation or testing.³ In effect, logical behaviourism for the logical positivists is simply Physicalism applied to the empirical science of psychology. The fact that the logical behaviourism of the logical positivists was simply part of their grand goal of a “Unified Science,” and hence was no different in principle from the corresponding “physicalization” of biology and other special sciences, indicates that it must have a very different character from the logical behaviourism of the anti-scientistically inclined ordinary language philosophers—and, more importantly, a very different character from the received view that it proposed analytic meaning equivalences between mental and physical-behavioural statements.

In the hands of the logical positivists, logical behaviourism is strongly reductive; for the ordinary language philosophy, especially Ryle, it is non-reductive. The former reductive version explicitly attempts to describe behaviour in purely “physical language,” hence in non-mentalistic terms, while the non-reductive version does not. As the most cursory browsing in *The Concept of Mind* amply demonstrates, Ryle makes no attempt whatsoever to purge his behavioural-dispositional analyses of psychological terms. His

³ As Hardcastle (2007) has rightly emphasized, the Unity of Science movement was also a reaction against the prevailing view of the time in Germany that there is a fundamental difference between the methods of the *Naturwissenschaften* (natural sciences) and those of the *Geisteswissenschaften* (human sciences). This point is made very forcefully in the first early statements of Physicalism by Neurath (1931), Carnap (1932) and Hempel (1935).

analyses of mentalistic discourse are saturated with mentalistic terminology.⁴ Moreover, very early on in *The Concept of Mind*—in the most reprinted chapter, ‘Descartes’s Myth’—he explicitly states his agreement with the view that “a person’s thinking, feeling and purposive doing cannot be described solely in the idioms of physics, chemistry and physiology” (p. 18).⁵ In contrast to this, there is no question but that the logical positivists explicitly intended a person’s psychological phenomena to be described precisely in the idioms of physics, chemistry and physiology—that just is the thesis of Physicalism applied to psychology.

Although this difference between the two versions of logical behaviourism—that the one is reductive and the other non-reductive—has been noted by commentators less inclined to shadow history, its full significance has still not been appreciated. For just as it is the very non-reductive nature of Ryle’s logical behaviourism that allows it to be logical or “analytical” in character, so too it is the very reductive nature of the logical positivists’ logical behaviourism that prevents it from being logical or analytical (see §1.4 below). But before we turn to this, let us have before us some prominent examples of the shadow understanding of logical behaviourism.

1.2 Shadow (Analytic) Behaviourism

In his highly influential and often anthologized critique of logical behaviourism, ‘Brains and Behaviour’ (1963), Hilary Putnam writes that “The Vienna positivists in their

⁴ Something Burge (2007) fails to appreciate, including as he does Ryle among those philosophers he alleges “shared a tendency to think that theorizing in psychology or philosophy of mind should dispense with mentalistic vocabulary, or interpret it in nonmentalistic terms, as far as possible” (p. 441).

⁵ In agreeing with this part of the Cartesian Myth, Ryle’s central point is of course that it does not follow from this truth that a person’s thinking, feeling and purposive doing are to be described in a “counterpart idiom” referring to “occult processes” running in parallel to physical processes.

‘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*” (p. 326). He goes on to say that logical behaviourism so understood “implies that all talk about mental events is translatable into talk about actual or overt potential behaviour” (*ibid.*). Putnam’s aim is not to praise but to bury logical behaviourism (once and for all). But since he considers the Vienna Circle’s version too extreme to need burying, he purports to inter a weaker and more plausible form, according to which “There 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.” Putnam says that he “shall call these *analytic entailments*” (p. 327).⁶

Along similar lines, Jerry Fodor (1968) has stated that “To qualify as a behaviourist in the broad sense of that term that I shall employ, one need only believe that the following proposition expresses a necessary truth: For each mental predicate that can be employed in a psychological explanation, there must be at least one description of behaviour to which it bears a logical connection” (p. 51).⁷ Furthermore, claims Fodor, “one of the more important differences between ... behaviourism and ... materialism [is] that while both maintain the identity of each mental state with some nonmental [*sic*] states, the propositions that enunciate the ... behaviourists’ reductions of mental to behavioural predicates are supposed to be analytic. ... By contrast, the materialist’s identifications of

⁶ Putnam does not cite a single work of the Vienna Circle, or indeed of any logical positivist, in which is to be found an endorsement of even an approximation of this, let alone of the previous extreme version of logical behaviourism. This is not surprising since, as I shall argue, no such semblance is there to be found.

⁷ Cf. Cornman (1971), pp. 132ff, esp. p. 140 and Kim (1971), p. 328.

mental with physical states are presumably enunciated by contingent propositions' (p 155n6). Since analytic truths and entailments are supposed to be knowable *a priori*, it follows that the analyses or translations or entailments in question between psychological statements and behavioural statements proposed by logical behaviourists are, according to Putnam and Fodor, supposed to be knowable *a priori*. Moreover, *pace* the logical behaviourists, the materialists, that is, the identity theorists, maintain that any link between the mental and the physical must be empirical in character, not conceptual or semantic, and hence knowable only *a posteriori*.

1.3 The Real (Synthetic) Behaviourism of the Logical Positivists

When one turns to the two canonical logical positivist texts of logical behaviourism, however, one finds a very different story. We can begin by noting a curious and rather blatant tension in standard accounts of the logical positivists' logical behaviourism. It is common for critics to maintain both that the logical positivists' thesis of logical behaviourism claims that the links between mind and behaviour are analytic and that such meaning links are forged by the logical behaviourists' verificationism. It is rarely noticed that in order for these two tenets to be true together, the statements of the conditions of verification (or confirmation) for psychological sentences must be analytically linked to them and hence determinable *a priori*. But when one turns to the texts one finds appeals to verification, confirmation or test conditions that cannot possibly be known *a priori* and are in no way analytically connected with psychological sentences.

Take Hempel's (1935/1972) oft-quoted and much-derided attempt at a logical behaviourist "translation" of the psychological statement 'Paul has a toothache'.

According to Hempel, the conditions under which this statement would be verified include not only Paul's verbal utterances, gestures and other overt behaviour, but also his internal physiological and neurological states, such as "Paul's blood pressure, digestive processes ..." and "such and such processes occur[ing] in Paul's central nervous system" (p. 122). Now, it is obvious—and it was surely obvious to Hempel—that any connection between Paul's toothache and his blood pressure, digestive and neural processes is empirical and established *a posteriori*; hence that it cannot be conceptual or analytical. At best, it will be a lawful empirical correlation and hence clearly synthetic.

Turning now to Carnap's (1933/59) very similar example of the sentence 'Mr. A is now excited' (P_1), Carnap asks "*what does sentence P_1 mean?*" and answers as follows:

The viewpoint which will here be defended is that P_1 has the same content as a sentence P_2 which asserts the existence of a physical structure characterized by the disposition to react in a specific manner to specific stimuli. In our example, P_2 asserts the existence of that physical structure (*micro-structure*) ... of Mr. A'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. (p. 172, my emphasis).

Very few of these physical-behavioural characterizations—if any—can be considered analytically linked with excitement. One year earlier Carnap had claimed, in the "formal mode," that "all psychological statements can be translated into physical language" (1932/34, p. 28), adding that the equivalent claim in the misleading and dangerous "material mode" is that all psychological statements "refer to physical events (viz. physical events in the body, *especially the central nervous system* ...". (p. 71; my emphasis). Five years later he wrote:

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 (1938/91, p. 401).

We shall return to Carnap’s distinction between definition and reduction below (§1.5.5) as misunderstandings of it seem to have played a key role in corrupting the logical positivists’ logical behaviourism into the received shadow position. The point for present purposes is that both physicalist definition and physicalist reduction statements of psychology are synthetic.

It is true that shadow historians sometimes draw attention to the fact that both Hempel and Carnap, in their canonical statements of “logical behaviourism,” go well beyond or rather behind overt behaviour and curiously include internal neurophysiological conditions in what are allegedly supposed officially to be overt-behavioural translations of psychological sentences. The typical reaction to this surprising discovery is to conclude that Hempel’s and Carnap’s translations are faulty because no such internal conditions could possibly be analytically linked to any psychological terms or sentences.⁸ Others refrain at the outset from describing the logical positivists as behaviourists at all and characterize them rather as holding a “semantic” or “logical” or “translation” form of materialism according to which psychological statements are translatable *a priori* on the basis of conceptual analysis into physical *simpliciter* statements rather than into physical *behavioural* ones.⁹ David Rosenthal, for example, has described a strong form of materialism that he calls the “translation view,” associated with the thesis of the unity of science, which “could be established without a detailed study of psychological beings”

⁸ See, e.g., Patterson (2008) p. 532 and Kim (2011), p. 70. But cf. Kim (2003) where the opposite, correct conclusion is drawn, namely, that Hempel and Carnap were operating with very different notions of translation, definition and meaning.

⁹ See, e.g., Rosenthal (1971/87), Beckermann (1992) and Stoljar (2010).

and which “could be shown [to be true] by examining simply what we mean by the words we use.” He notes how strange this view is, as it seems to follow from it that “it would then be possible to defend the unity of science, which is a claim about the results of future scientific investigation, without appealing to any such results.” In contrast, a weaker form of materialism, of which the identity theory is one variety, “does not result in peculiarities of this sort ... [for it] can only be established on the basis of results from future scientific study.”¹⁰ While the recognition that logical positivists’ Physicalistic translations were never intended to be restricted to overt behavioural terms is salutary, it is unfortunately accompanied by the erroneous view that the translations in questions are still meant to provide analytical equivalences between mental and physical statements.

We should pause to note that Carnap does countenance behaviouristic reductions that are explicitly only about overt molar behaviour:

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 (1938/91, p. 402).

But even here, with an explicitly overt-behavioural proposal, Carnap does not say that such a molar-behaviouristic reduction sentence for anger will be analytic. Indeed, the “laws” referred to, connecting inner states with outer behavioural reactions or symptoms, are obviously intended to be empirical physical laws. Notice also that Carnap takes the inner state to *cause* the outer behavioural reaction, and so anticipates and pre-empts by

¹⁰ Rosenthal (1971/87), p. 4. Beckermann (1992) seems to misinterpret Carnap’s Physicalism in essentially the same way on pp. 2-7, as does Stoljar (2010) on pp. 117-118.

more than two decades both Putnam's celebrated "Super-Spartan" objection to shadow logical behaviourism, which formed the basis of its internment in 'Brains and Behaviour', as well as Putnam's distinction between an inner state and its outer symptoms, backed by his well-known polio analogy. It is further noteworthy that Carnap here anticipates both the causal critique of behaviourism pressed by Jerry Fodor and David Armstrong (that mental states are not identical with behaviour but are the causes of behaviour) and the causal-functional analysis of mental concepts. This latter point has been noted by Patterson (2008, p. 531) and Kim (2003, p. 275), but it needs to be handled carefully lest we wrongly re-foist shadow logical behaviourism onto Carnap. Contrary to what Kim says, it would be incorrect, I think, to associate Carnap with analytic functionalism, which is a development of Rylean logical/analytical behaviourism, according to which the 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 realizers of the definitions. Carnap's proto-functionalism is more akin, I believe, to an empirical psycho-functionalism (see Block (1978)), in which empirical science is involved at the first stage; that is, the functional definitions of mental terms are themselves specified *a posteriori* by scientific theory. This seems to me to fit the texts and the spirit of 'Logical Foundations' and 'Psychology in Physical Language' better, to make better sense of the strong analogy Carnap draws between concept formation in the physical sciences and in the sciences of psychology, and to gel better with Carnap's procedure of "physicalization" (to which we shall turn below).¹¹

¹¹ Moreover, as Hempel (1952, 1954) and Pap (1958, ch. 11) point out, if there is more than one conditional definition (e.g., a set of two or more bi-lateral reduction sentences) for a given term, as Carnap

Be that as it may, let us assume that the logical positivists' physical translations were never intended to be restricted to descriptions of physical behaviour but were intended to include physical descriptions of inner neurophysiological states, and pursue the question whether such broadly construed physical translations were intended to be analytically or conceptually true.

Throughout his early writings of the 1930s Carnap speaks of the “rules of inference,” “rules of transformation,” and “rules of translation” of the physical language in which the analyses or reductions are to be carried out. Although he is not always entirely explicit about it, it is pretty clear even in the earliest of these writings that not all of these “rules” are laws of logic and that some of them are intended to be laws of nature. For example, in ‘Unity of Science’ (1932) he wrote of the “the rules of transformation inside the physical language (*including the system of natural laws*)” (p. 88, my emphasis; cf. p. 92). This became much clearer in *The Logical Syntax of Language* (1934) and ‘Testability and Meaning’ (1936-7) in which Carnap explicitly distinguishes between the L-rules and the P-rules of a scientific language on the basis of which inferences or “transformations” may be validly carried out: the former are logical laws and the latter 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 physicalistic analysis or reduction.

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

clearly expected there to be in many cases, including psychology, then one can derive a non-analytic empirical statement from them, from which it follows that at least one of the conditional definitions must be non-analytic and hence empirical or synthetic. Conditional definition/reduction is discussed in §1.5.5 below.

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) 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 on the basis of physical laws. Carnap explicitly allowed a psychological sentence, Q_1 , and a physical translation of it, Q_2 , to be P-equipollent, 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). Carnap took 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” (*ibid.*). In a letter he wrote to Herbert Feigl in 1933 Carnap is more committal and expressly states that the two sentences are not analytic. His example is ‘N. has a visual image of a house’ (A) and he offers two translations: ‘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). Carnap’s comments on this are highly instructive:

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.

It is noteworthy that while B_1 , unlike B_2 , is claimed by Carnap to be L-equivalent to A , it is not behavioural¹²—in fact, it is not even physical. I would conjecture that it is intended as 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 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 had a strong great influence on Carnap.¹³ At any rate, it should by now be clear that the real logical behaviourism of the positivists was in fact far less extreme than even Putnam’s two-decades older weakened version. For it never claimed mental events to be logical constructions of overt behaviour and never claimed to offer analytically true logical constructions of “mind talk” into either (overt or covert) “behaviour talk” or “physical talk.”

1.4 Analytic Behaviourism vs. Synthetic Behaviourism/Materialism

The philosophical behaviourism of those ordinary language philosophers who were behaviourists, Ryle in particular, is properly called ‘analytic’, since it did indeed attempt to give *a priori* conceptual analyses of (some) mentalistic sentences in behavioural-dispositional terms. Such behavioural “definitions” or “hypotheticals” were supposed to give the ordinary meaning of mentalistic sentences. Moreover, the behaviour they adverted to was indeed of the “outer” and “overt” variety, it was purely behaviouristic; and it was so because it was on the basis of overt behaviour, not on the basis of internal

¹² Cirera (1993) also notes this.

¹³ Such adverbial strategies for avoiding intentional language are strongly criticized by Chisholm (1955-56).

neurophysiological states, that ordinary mentalistic language was learned and applied in everyday situations by ordinary language speakers. But this pure overt behaviour was described using an abundance of mental terms and so the logical behaviourism produced was flagrantly non-reductionist. On reflection, this should not be very surprising. For if there are going to be analytically true behavioural analyses of psychological sentences, these are bound to contain mental terminology, as they will draw out conceptual connections between mental states and behaviour described as *intentional action*. (If we had to have a name for it, we might call this kind of Rylean ordinary-language logical behaviourism *pure non-reductive analytic behaviourism*: “pure” because it adverts to overt behaviour only; “non-reductive” because it employs mentalistic terms in its *analysans*; and “analytic” because the connections between the mental and the behavioural are supposed to be analytically true and knowable *a priori*). In contrast, the logical positivists’ various “definitions”, “reductions,” “transformations” and “translations” were indeed explicitly couched—or intended to be so in a programmatic spirit—in non-mentalistic terminology and so were (supposed to be) truly reductive. But very few—perhaps none—of these translations were supposed to be analytic but rather were intended to be synthetic, arrived at empirically, in fact experimentally in many cases.¹⁴ Moreover, they did not advert only to purely overt behaviour but made explicit reference to inner “central” neuro-physiological processes, states and structures. (Again, if we needed a name, perhaps *impure reductive synthetic behaviourism* or, given the last mentioned fact, *synthetic semantic materialism*, would do.)

¹⁴ I am grateful to my colleague Thomas Uebel for pointing out after reading an earlier draft of this chapter that Cirera (1993), of which I was unaware, independently makes this crucial point. When following up citations of Cirera’s work, I was subsequently led to Kim (2003), who also makes the point independently.

1.5 Origins of the Shadow Doctrine

How, then, did the idea that the logical positivists endorsed the shadow doctrine that there are analytic entailments between psychological sentences and either behavioural sentences or physical sentences couched in non-mentalistic vocabulary (*pure reductive analytic behaviourism* or *analytic semantic materialism*, as we might put it) get started? Perhaps many philosophers simply ran the two versions together producing a homeless and unstable fusion that was easy to refute and that served as a foil for their own allegedly superior positions. In other words, as Watson suggests, shadow history provided the necessary opposition by which to define and build one's own position. While there must be some truth in this, I think that there are more concrete and interesting reasons.

1.5.1 *The Extensionality of Translation*

We can trace the origin of the shadow position partly to the positivists' highly technical and, especially for us today, counter-intuitive use of the expressions 'translation', 'meaning', 'synonymy', 'definition' and their cognates.¹⁵ None of these terms is used today in anything like the way the positivists, especially Carnap and Hempel, were using them in the 1930s and even to some extent in the 1940s. Many of these terms and their cognates have strong modal implications for us now that they did not have back then for the positivists. Indeed, Carnap and Hempel 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—or Carnap's (1936-7) later

¹⁵ Actually, the technical use ran against the grain even during the first half of the twentieth century. Ducasse (1941, ch. 7), e.g., complains that what Carnap (1935/63) calls translation is not truly translation.

“reduction sentences,” of which more presently—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. 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.¹⁶ From the very beginning, Carnap fully acknowledged the empirical character of the proposed physical definitions of psychological concepts and the translations of psychological sentences into physical sentences. As he says,

Sentence P₁, “A is excited” cannot, indeed, today be translated into a physical sentence P₃ 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” (1933/59, p. 175).

Physical “translation” draws on the empirical knowledge available at the time.

1.5.2 Empirical Physicalization

Indeed, in both ‘Unity of Science’ and ‘Psychology in Physical Language’—which are among Carnap’s earliest treatments of psychology and considered as canonical texts of logical behaviourism—Carnap outlines an empirical, experimental procedure he calls “physicalization.” A primary example of physicalization occurs when a non-physicalistic sentence reporting a quality (e.g., a colour or sound) is correlated with a physicalistic sentence reporting a measurable quantity (light or sound wave frequency). For example, one can physicalize a protocol (observation) sentence reporting a “qualitative

¹⁶ Carnap (1956) is especially clear about this. Philosophers of science, including Carnap (1936/7), soon began to realize that natural laws, disposition statements, and the counterfactual conditionals associated with them cannot be formalized using an extensional logic. See Carnap (1956), Hempel (1954) and Suppe (1977) for some discussion of this.

determination,” such as a statement about colour (‘Green here now’), by correlating it with a sentence reporting a “quantitative determination” (wavelength of such and such frequency). The procedure involves an experimenter varying various physical conditions (e.g., wave frequencies and oscillations) in order to discover which ones correlate with a subject’s utterance of a protocol sentence. Carnap gives an interesting, if rather quirky, example of how this might work in the psychological sub-field of graphology, the study of the relation between handwriting and personality. Carnap describes the third stage of the physicalization of graphology as follows:

the basic empirical task of graphology ... consists of the search for the correlations which hold between the properties of handwriting and those of character. ... The problem of systematization here is to determine the degree of correlation of the two properties by a statistical investigation of many instances of script of the type in question and the characters of the corresponding writers” (1933/1959, p. 189).

He also believes the same kind of physicalization can be carried out on psychological sentences describing actions: “The class of arm-movements to which the protocol-designation “beckoning motion” corresponds can be determined, and then described in terms of physical concepts” (p. 182). The interesting point for present purposes is not that Carnap thinks such physicalization is plausible or even possible in principle. No doubt it is not and Carnap was characteristically overly optimistic here. As many have pointed out time and again, it is most unlikely that all the arm movements constituting the class of beckoning or waving have anything purely physical in common, and hence most unlikely that such a class of actions can be correlated with any single physical property picked out (non-trivially, that is, non-disjunctively) by any purely physical concept. The interesting point is that the physicalization of actions proposed by Carnap is entirely empirical in character and matter of painstaking experimental work. Moreover, granted that the

physicalization in question is empirically highly unlikely, it seems no more unlikely than an identity theory's proposed type-identification of a raw feel with a brain state.

1.5.3 Physicalization vs. Identity

It is worth dwelling briefly on this last point. In his recent survey of contemporary physicalism, Daniel Stoljar discusses “the semantic version of physicalism associated with Carnap and Neurath,” which he characterizes as the thesis that “every statement or predicate is synonymous with some physical statement or predicate” (2010, p. 117). One of the many reasons for denying the “semantic view,” according to Stoljar, is that there will be many meaningful predicates that are not synonymous with any physical predicate. His example is the predicate ‘has no soul’ in the sentence ‘Otto has no soul’. Stoljar claims that in order for the semantic view of Carnap and Neurath to be true, this sentence “would have to be equivalent in meaning to a physical statement” and that “this is extremely unlikely” (p. 118). He then presses a further “simple-minded objection” that “translation is a singularly difficult business”—justifying this by noting that “translating Proust into English is something that people are still arguing about, and while Proust might be a special case, it remains advisable that one should not associate physicalism with translation too closely” (*ibid.*). He continues by stating that “It was for these and similar reasons that many philosophers in the 1950s and 1960s turned from a semantic to a non-semantic formulation of physicalism (e.g., Smart 1959)” (*ibid.*).

Although he is not explicit about it, it is clear that by “synonymy” and “translation” Stoljar has in mind a strongly modalized notion, according to which the bi-conditional translation of ‘Otto has no soul’ into some candidate physical sentence is analytic and

therefore knowable *a priori*—translators of Proust, after all, argue with each other from the armchair, as it were, and not from laboratories. In other words, in Carnap’s terminology, Stoljar is assuming that Carnap and Neurath claim that ‘has soul’ is L-synonymous with a physical predicate and ‘Otto has no soul’ is L-equipollent (or L-equivalent) with a physical sentence. Stoljar is undoubtedly correct that any such analytic translation is indeed extremely unlikely. But there are a couple of problems with Stoljar’s account here. First, neither Carnap nor Neurath nor Hempel ever held the shadow doctrine of semantic physicalism; as we have seen Carnap’s claim is that ‘has soul’ is P-synonymous with a physical predicate and ‘Otto has no soul’ P-equipollent with a physical sentence. Consequently, Stoljar has mischaracterized the difference between the physicalism of the logical positivists and the physicalism of the identity theory. Second, once it is recognized that the logical positivists’ translations and synonymies are synthetic, the claim that the psychological predicate ‘has soul’ is “synonymous” with a physical predicate—that is, *nomologically co-extensive* with some physical predicate—is no less plausible than a type identity theorist’s claim that the property of having soul is *identical* with some physical property. Indeed, one might maintain that the logical positivists’ claim is in fact more plausible than the identity theorists’, as it rests with a simple correlation between predicates, which is much weaker than a claim of identity between the properties designated by the predicates, and all that can arguably be established empirically—while the identity theorist must somehow convert the predicate-correlation into an property-identity on non-scientific, or at least non-empirical, grounds (of parsimony, say, or abduction).¹⁷

¹⁷ One original point of disagreement between Place (1956) and Smart (1959) was that Smart considered the conversion of a correlation into an identity at least partly a philosophical and not completely scientific

1.5.4 The Term ‘Logical Behaviourism’

Returning to our question of the origin of shadow logical behaviourism, we should note that the term ‘logical behaviourism’ was coined by Hempel in passing, parenthetically, in ‘Logical Analysis of Psychology’ (so far as I know, Carnap never used the term to describe his position, and of course neither did Ryle). To our ears—the ears of the latter half of the twentieth century—the expression ‘logical behaviourism’ irresistibly suggests a doctrine according to which mind talk is logically or conceptually or analytically equivalent to behaviour talk, and hence necessarily linked with it, in contrast to being merely contingently connected with it. The suggestion is only exacerbated by the occurrence of the terms ‘translation’, ‘meaning’ and ‘definition’. But the sense of ‘logic’ that both Carnap and Hempel had in mind was that of *logical analysis*, specifically the logical analysis of science, or simply ‘logic of science’, as Carnap called it. Logic of science is the analysis and study of the linguistic expressions of science, their kinds and relations, and how they are ordered and structured into systems known as scientific theories, all abstracted from the psychological and social conditions of working scientists. There are two important terminological points here. First, a “logical” investigation or study of science is intended to contrast with a psychological and sociological study (and a philosophical study, presumably, if philosophy is understood as speculative metaphysics). Second, since many of the relations between scientific sentences studied by the logical analysis of science, particularly between the theoretical sentences and the protocol or observation sentences that confirm them, will be contingent and empirically established, the “logic” of science includes the study of synthetic sentences—in particular the

matter.

synthetic sentences that describe the “translations” of sentences of the empirical sciences (as opposed to mathematics) into the physicalistic sentences that constitute the inter-subjective confirmation basis of a unified science. Hempel’s parenthetical coinage was intended further to contrast the logical positivist’s physicalization of psychology, essentially a logico-linguistic affair, with the psychological behaviourism of J. B. Watson and his followers, which was a thesis about the methods and aims of the empirical science of psychology.

The “logical behaviourism” of the logical positivists is simply part of their overall project of the physicalization of all of empirical science, which in turn is an implication of the goal of unified science. When combined with the claim that a physicalistic language is the only known inter-subjective (as well as inter-sensory and universal) language—not as a matter of necessity, but only contingently, something Carnap was at pains to point out from the very beginning¹⁸—the unity of science thesis becomes “Physicalism.” The doctrine’s emphasis on language is characteristic of the early days of analytic philosophy. When applied to the science of psychology, Physicalism, understood as the linguistic doctrine that only a physicalistic language is capable of serving as an inter-subjective confirmation base for empirical science, becomes the “logical behaviourism” of Hempel and Carnap. It is in this light that one must view Carnap’s general thesis that “all statements of Science can be translated into physical language” and the relevant sub-thesis for psychology that “all psychological statements can be translated into physical language.” This sub-thesis is no different, in principle, from the relevant sub-thesis for biology, namely, that “every statement of Biology can be

¹⁸ *Pace* Smith (1986, p. 60), who erroneously claims that Carnap did not view Physicalism as contingent. Carnap (1931, pp. 60ff and 96) contradicts Smith’s claim.

translated into physical language” (1932/1934, p. 70) and not many philosophers are tempted to view the physicalization of biology as project in *a priori* conceptual meaning-analysis.

1.5.5 The Confusion of Definition/Reduction with Analytic/Synthetic

A major source of confusion has to do with the already mentioned orthodox understanding of the transition from logical behaviourism to the identity theory. The confusion seems to originate or at least stem primarily from Feigl—ironically, as we shall see, given his letter from Carnap quoted above—and is perpetuated and carried into contemporary philosophy of mind’s self-image through the Feigl-Putnam-Fodor line of influence. Early on in ‘The “Mental” and the “Physical”’, Feigl makes clear the famous *anagnorisis* of the identity theorists:

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 shadow view that early reductionistic logical behaviourism was purely overt-behavioural, excluding reference to inner neurophysiological states, while his own early view included them, Feigl runs together two crucially different things: *analyticity* and *definability*. 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.¹⁹ But both of 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.²⁰ (For the logical positivists, of course, the defined expressions will be so-called “theoretical” terms and the primitive expressions the “observation” terms.) Now, Carnap (1936-7) 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 translation 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 conditionals with material bi-conditionals as consequents.²¹ 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

¹⁹ Cf. Feigl (1958), pp. 427, 447, as well as Feigl (1963), p. 251 and Feigl (1971), p. 302. Pap (1952, p. 210) and Smith (1986, p. 53) also seem to hold these mistaken assumptions.

²⁰ This is an oversimplification: strictly speaking, only the ultimate definition in a definition chain will have only undefined primitive terms in the *definiens*. See Carnap (1936-7) and Hempel (1952).

²¹ See Carnap (1936-7), §10 and Hempel (1952).

definitions”—they did not allow the terms to be eliminated and replaced and hence they could not form the basis for translations.²²

The important point to notice 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). Recall Carnap’s statement quoted earlier (in §1.3) about the lack of necessary and sufficient physiological conditions for anger preventing the latter’s definition. The majority of these material biconditionals are not analytic, they are neither conceptually true purely in virtue of meaning nor *a priori* knowable, but are supposed to be physical laws discovered on the basis of experimental investigations.²³ Thus, contrary to what Feigl and others seem to suppose, the failure of explicit definition and hence translation is not at all tantamount to the failure of *a priori* analytic definition and translation.²⁴

The mistaken assumption that explicit definitions are analytic 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 (1933) and Hempel (1935) state that the two philosophers no longer hold the strict definability thesis and have

²² Contrary to Suppe’s (1977) assertion that reduction sentences were claimed to be analytic sentences (p. 23), Carnap (1936-7, §8) explicitly states that some reduction sentences may be analytic and some may be P-valid. Indeed, as was pointed out in note 11 above, it can be proved that some reduction sentences are synthetic.

²³ Cf. Cirera (1993) and Kim (2003).

²⁴ A fuller treatment of this issue would need to discuss the distinction between analyticity in the narrow sense, viz., logical truth, and analyticity in the broader all-bachelors-are-married sense.

replaced it with the more flexible reducibility thesis.²⁵ 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, Kim 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 proffered physical-behavioural conditions are analytically entailed by his sample psychological sentence ‘Paul has a toothache’ (as we saw in §1.3 above). 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

²⁵ See Carnap’s 1961 addenda to the reprinting of Carnap (1932) in Alston and Nakhnikian (1963) and Hempel’s 1977 prefatory note in Block (1980). As Hempel notes in his addendum, physicalization was liberalized even further with the later introduction of “hypothetical constructs” connected to the observation language via “correspondence rules.” See also Carnap’s 1957 addendum to the reprinting of Carnap (1933) in Ayer (1959) and Carnap (1956) and Hempel (1954).

physical terms and his point is that he has now moved to the more liberal thesis of reduction. Carnap's addendum makes exactly the same point.²⁶

1.6 The Physicalism of Logical Positivism vs. the Physicalism of the Identity Theory

What, then, is the real nature of the shift from the physicalism of the logical positivists to the physicalism of the identity theory, if both in fact proposed empirical, contingent connections between the mental and the physical? The simple answer is that for the logical positivists physicalism was essentially a linguistic doctrine about the language of science, or about the language of the inter-subjective confirmation basis of science, combined with the view that the mind-body problem was a metaphysical pseudo-problem to be ignored and replaced by a logico-linguistic analysis of the place of psychology in a unified physical science. This is one strand in the great period of the "linguistic turn." The identity theorists' physicalism (or materialism), in contrast, was an ontological doctrine and they certainly did not view the mind-body problem as a pseudo problem. I submit therefore that the real change in approach to the mind-body problem that occurred in the 1950s and 60s was simply the rejection the earlier view, shared by logical positivists and ordinary language philosophers, that it was a pseudo-problem.²⁷ The radical difference of approach is succinctly summed up in Carnap's response to Feigl's identity theory:

²⁶ Another source of shadow logical behaviourism, which space limitations prevent me from discussing here, is probably the often-drawn analogy with phenomenism. If one thinks of phenomenism as the doctrine that material-object statements are analytically equivalent to sense-data statements, and one thinks logical behaviourism is like phenomenism, then one will likely arrive at a shadow understanding of logical behaviourism. But, as 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'." Phenomenism, in other words, cast its own influential shadow. But that is another story.

²⁷ Cf. Feigl (1960).

The identity statement mentioned [that a certain psychological process *P* is identical with a certain neurophysiological process *N*] is a sentence of the object language; this fact may mislead the reader into believing that the controversy about the identity view concerns a question of fact. ... It seems preferable to me to formulate the question in the metalanguage, not as a factual question about the world, but as a question concerning the choice of language. ... Those facts Feigl proposes as evidence for the identity view are perhaps better regarded as reasons for preferring a monistic language ... in this language the predicates “*P*” and “*N*”, though not L-equivalent, are P-equivalent ... I am willing to call my position an identity conception in the following sense: in agreement with Feigl I prefer the monistic language, and like him I believe that the evidence available today provides good reasons for the assumption that this language will also function well in the future (Carnap (1963), pp. 885-6).

Although Carnap wrote (or at least published) this statement in 1963, this was his position from the very beginning, thirty years earlier in the heyday of logical positivism. In 1935 Carnap had said that the “pseudo-object” identity-sentence of the material mode, ‘The evening-star and the morning-star are identical’ is to be replaced by the syntactical formal-mode sentence ‘The words “evening-star” and “morning-star” are synonymous’ (1935/63p. 447); and in the earlier more rigorous treatment (1934/2002, §75), it is clear that he means P-synonymous. Throughout his life, Carnap remained committed to his radical positivistic empirico-pragmatic view that the only legitimate philosophical problems that were not straightforward empirical scientific questions were questions of language choice. But according to Carnap himself, his position on the mind-body problem can, and always could, be expressed in the misleading material mode as a synthetic identity theory.

Kim has questioned whether Carnap’s and Hempel’s “overly generous” notion of translation “can underwrite the kind of physicalism/behaviourism that these philosophers wanted to formulate and defend” (2003, p. 273). He argues that it cannot because a nomological correlation between two predicates does not license the claim that they designate the same property. Such correlations are consistent with a variety of dualisms and “Physicalism, however it is formulated, must exclude dualisms” (p. 268). But this

criticism assumes that physicalism is a metaphysical doctrine and so flatly begs the question against Carnap and Hempel. To be sure, lawful correlations between mental and physical predicates are not sufficient to establish the kind of metaphysical physicalism that Kim and the identity theorists wish for. But the question is whether they are sufficient to establish physicalism as Carnap and Hempel understood it, namely, as anti-metaphysical Physicalism; and whether, if so, the anti-metaphysicalism of Physicalism is a better approach to the mind-body problem than Kim's preferred metaphysical physicalism. Neither question can be treated in detail here. Suffice it to say, with respect to the first question, that it is not at all clear that Physicalism cannot be underwritten merely by lawful correlations between the predicates of the various special sciences and the predicates of a physical language. Physicalism is, after all, intended to be a contingent thesis, indeed, a working empirical hypothesis. With regard to the second question, Kim offers no argument for the superiority of physicalism to Physicalism. Carnap and Hempel reject the call for an explanation of psycho-physical correlations because on their view any such alleged explanation would be metaphysical, hence impossible to verify or confirm, hence unscientific. And one might even argue that Carnap's anti-metaphysical instincts have been spectacularly borne out by the subsequent history of failed attempts to establish metaphysical physicalism as a solution to the mind-body problem. Place, Feigl and Smart, for example, all argued for the identity theory on the basis of psychophysical correlations; but there remains little consensus and much skepticism about whether these arguments are successful. Whether it is possible to convert psychophysical correlations into identities remains a highly controversial question and various arguments based on simplicity and abduction have so far garnered few adherents. Kim (2011) himself has

offered devastating criticisms of several arguments that attempt to derive identities from correlations. The second wave of identity theorists, in particular David Armstrong (1968) and David Lewis (1966), offered different but no less controversial arguments for psychophysical identity. Many more recent attempts again rely on controversial and rarely defended claims about the “causal closure of the physical domain.” Or take the concept of supervenience, introduced into philosophy of mind in the early 1970s as a possible position on the relation between the mental and the physical. For a time it was thought that supervenience offered a significant (non-reductive) materialist solution to the mind-body problem. But a decade of intensive research showed that most, if not all, mental-physical supervenience relations were consistent with various different and incompatible metaphysical positions on the mind-body relation, including even substance dualism!²⁸ I dare say Carnap would not have been the least surprised by these developments. None of this proves of course that the anti-metaphysic of Physicalism is to be preferred to the metaphysic of physicalism. But it does suggest, I think, that when the logical positivists’ anti-metaphysic is applied to the mind-body problem, it is not to be lightly dismissed.

²⁸ See Kim (1998), ch. 1.

2. Two Identity Theories as a Reflection of Two Philosophies

2.1 Introduction

The late 1950s and early 1960s saw the rise to prominence of the celebrated mind-brain identity theory propounded by U. T. Place, Feigl and J. J. C. Smart, in their respective classic articles ‘Is Consciousness a Brain Process?’ (1956), ‘The “Mental” and the “Physical”’ (1958), and ‘Sensations and Brain Processes’ (1959).²⁹ The theory was actually knocking around at least a decade earlier—in fact, as we shall see, several decades earlier. Hilary Putnam remarked in print a year before Feigl’s article and two years before Smart’s that

“Physicalism,” expressed as a working hypothesis, amounts to this: subjective experience (e.g., a particular feeling of anger) is a particular kind of physical state of the organism. This is of course a synthetic identity, if true (as Feigl has very well pointed out). Philosophers are quite right in saying that ‘the sensation blue’ cannot *mean* a physical state. But they are wrong when they maintain that it cannot *be* a physical state. (Thus, ‘the morning star’ cannot *mean* ‘the evening star’. But the morning star *is* the evening star—for both are identical with the planet Venus, to use the familiar example (Putnam 1957, p. 97).

The work by Feigl that Putnam refers to is a 1950 article entitled ‘The Mind-Body Problem in the Development of Logical Empiricism’, in which Feigl defends a proposal according to which there is an identity between the mental and the physical, an identity to be established empirically, like other scientific identifications. He too uses Frege’s famous example as an analogy, as well as other examples of “theoretical identities” that have entered into the mainstay of analytic philosophy of mind, such as heat’s being

²⁹ While Smart was following Place’s lead—the two were colleagues at the time at the University of Adelaide in Australia—Feigl was working independently at the University of Minnesota. Feigl’s long essay was reprinted as a book with an accompanying postscript written decade later: *The ‘Mental’ and the ‘Physical’*. *The Essay and a Postscript* (Minnesota, 1967).

(identical with) molecular motion and visible light's being a certain frequency of electromagnetic waves.

As Putnam makes clear, the mind-brain identity theory in question was in fact a more restricted identification of sensory consciousness with physical states. Place and Smart, having studied at Oxford, were under the influence of Gilbert Ryle's *The Concept of Mind* (1949) and so were largely sympathetic to a behaviourist-dispositional view of intentional mental states, such as beliefs, wants, fears and expectations. But they were concerned that the so-called qualitative mental states associated with sensory consciousness, the having of after-images and sensations of pain, for example, were not dispositional in nature, but were occurrent episodes, and so were unlikely to yield to behavioural analyses.³⁰ They proposed to identify this non-dispositional "mental residue" of sensory consciousness with neural events and processes in the brain. Interestingly, Feigl too independently wished to identify only sensory consciousness with brain states. But this was not owing to any reservations about the possibility of dispositional analyses of it. Rather, he viewed the idea of a physical reduction of intentional mental states to brain states as confused, a kind of category mistake, since the intentionality of intentional mental states was properly a logico-semantic problem and was to be dealt with by theories of reference and designation in philosophy of language.³¹ More importantly, however, Feigl was coming out of a very different empiricist tradition that viewed the phenomenal elements of conscious awareness as both the paradigm of the mental and as

³⁰ Pace Livingston (2004, ch. 4), neither Place nor Smart thought that Ryle himself claimed that sensations were behavioural dispositions. They just noticed that his behavioural analyses could not be extended to sensations.

³¹ The view that intentionality was ultimately a semantic phenomenon to be explained linguistically was defended at the same time by Feigl's colleague Wilfrid Sellars against the contrary view of Roderick Chisholm. See the debate between Sellars and Chisholm in the Appendix to *Minnesota Studies in the Philosophy of Science II* (1958).

the epistemological foundation for empirical knowledge. This empiricist tradition included the phenomenalism of the early days of logical positivism (especially Carnap's *Aufbau*) and the epistemologically oriented neutral monism of William James and Bertrand Russell.³² Both can be traced back to the phenomenalism of the physicist-philosopher Ernst Mach and ultimately of course to the phenomenalistic empiricism of Mill and Hume. Most crucial of all for Feigl, however, was the virtually unknown 1918 identity theory of Moritz Schlick, which was itself part of the nineteenth-century German and Austrian tradition of psychophysical parallelism (Heidelberger 2003).

It has sometimes been noted there were really two different identity theories originally developed in the 1950s, the Austrian Schlick-Feigl version and the Australian Place-Smart version, and that the latter is more materialistic and reductionistic—even eliminativistic—than the other.³³ In a recent article Leopold Stubenberg argues that “The Australian approach to the problem of identifying mind and body represents a stunning reversal of the Austrian approach” (1997, p. 136). He adds that “Whether this be idealism or panpsychism is not clear to me. But it is clear to me that this view is deeply antimaterialistic in spirit.” Stubenberg goes on to argue for the superiority of the Schlick-Feigl version. His valuable discussion brings out an important and deep rift running through post-1950 analytic philosophy of mind—a rift which has begun to widen even more under contemporary analytic philosophy of mind's turn-of-the-millennium obsession with phenomenal consciousness. I will argue that the Schlick-Feigl version of

³² This early period of analytic philosophy of mind, roughly 1900-1930, which was much more epistemologically centred than philosophy of mind is today, has received little scholarly attention. For some edifying discussion of it, see Hatfield (2002, 2004).

³³ Cornman (1971), pp. 30, 125, Armstrong (1993), p. xiii and Borst (1970), p. 20. Cf. Feigl (1975), p. 15.

the identity theory is at least as problematic as the Place-Smart version and that there is a more fundamental rift between them than that recognized by Stubenberg.

2.2 Schlick and the Origins of the Identity Theory

Most of the crucial elements of Feigl's theory are in fact derived directly from Schlick, who had proposed in his *General Theory of Knowledge* (1918; 2nd edition 1925) what seems to be truly the first twentieth-century mind-brain identity theory, as Feigl (1958, p. 80n) notes. Kim has disputed this, claiming—contrary to the editors of the English translation, one of whom is Feigl—not to be able to find in this work a “reasonably clear and unambiguous statement of the mind-brain identity theory” (2003, p. 276), and that it is anyway doubtful that Schlick could have worked out an identity theory at this time, because it is only after 1930 that the Vienna Circle adopted physicalism. Kim thus suggests that it is in Schlick's later article ‘On the Relation between Psychological and Physical Concepts’ (1935), written during his positivist period, that we find a genuine psychophysical identity theory—though even here, apparently, it is a psychobehavioural, and not a psychoneural, identity theory. But all this is, I think, a confusion, stemming again from Kim's failure to appreciate that physicalism for the logical positivists, that is, Physicalism, is not a metaphysics and hence does not purport to offer a metaphysical solution to the mind-body problem, *a fortiori* not a materialist one. Schlick's (1935) “physicalism” is simply Physicalism, as he makes clear: “every psychological proposition can be translated into an expression in which physical concepts alone occur” (1935, p. 399). On the contrary, it is precisely because Schlick's *General Theory of Knowledge* is written before his positivist period that it can enunciate an identity theory. We can in any

case settle the matter by turning directly to the text. I submit that in §§33-35 Schlick does indeed propound an identity theory. This is one of the crucial passages:

The ... relation between immediately experienced reality and the physical brain process is [not] one of causal dependency but of simple *identity*. What we have is one and the same reality, not “viewed from two different sides” or “manifesting itself in two different forms,” but designated by two different conceptual systems, the psychological and the physical.” (p. 299, italics in original; cf. pp. 310-312).

If this is a statement of the mind-brain identity theory, or rather, a statement of *a* mind-brain identity theory, as it certainly seems to be, it follows that that Place (n. d.) too is mistaken when he claims that “The earliest statement of the identity theory under the title was in the psychologist E. G. Boring’s book *The Physical Dimensions of Consciousness* published in 1933.”³⁴ Assuming Boring was the first to state the identity theory back in 1933, Place offers a plausible explanation for why it was not accepted more widely until the 1950s. First, as far as psychologists go, psychology is dominated by behaviourism during this period; second, when it comes to the philosophers, Frege’s logical work on identity statements, with its central claim that two expressions which differ in meaning (sense) can nevertheless be discovered empirically to refer to the same thing, is not yet well known among philosophers, but is a crucial part of any defensible identity theory.

It is certainly true that Feigl and Smart in their presentations of the identity theory both invoke Frege’s idea that expressions with different meanings can have the same referent and that it may take an empirical discovery to reveal this. But, as we have seen, Carnap was well aware of Frege’s works and the distinction between sense and reference and re-described it with his own L- and P-concepts in the 1930s. Carnap had in fact attended Frege’s lectures in Jena during 1910-1914, and his lecture notes for the Winter Semester of 1910-11 contain Frege’s famous astronomical example and his explanation of the

³⁴ Carnap uses the phrase ‘identity theory’ in §22 of the *Aufbau* (1928).

distinction between sense and reference.³⁵ Moreover, given that Schlick had already formulated an identity theory at least a decade earlier, and did not draw on Frege's work on sense and reference at all in order to do so, the question remains why, given Schlick's prominence and influence in philosophical circles at the time in Vienna, it still took another thirty years for it to make any significant impact. The answer must be that Schlick came under the strong spell of Wittgenstein and Carnap, and other members of the Vienna Circle, all of whom, while well aware of Frege's distinction between sense and reference, were wont to dismiss the mind-body problem as a metaphysical pseudo-problem to be replaced by the project of constructing a Physicalistic language into which to translate psychological sentences.³⁶ In short, Schlick became a logical positivist, took the linguistic turn, and ascended to the formal mode, leaving his original material-mode metaphysical identity theory behind. Any hopes (however vain) of Schlick returning to his original pre-positivist position were dashed by his murder in 1933. It was left to Feigl to carry the torch to the United States and refine the theory into a more sophisticated version, using some of the tools from the rapidly developing philosophies of logic and science—but only after liberating himself from the strictly positivist position *vis-à-vis* the mind-body problem.³⁷

³⁵ See Reck and Awodey (2003), p. 71 (p. 24 of Carnap's notes for the Winter Semester 1910-11).

³⁶ While Wittgenstein and the Vienna Circle shared the view that traditional metaphysical problems were pseudo-problems, the relationship between the Vienna Circle (especially Carnap), Wittgenstein and Physicalism is much more controversial. For discussion, see Uebel (1995).

³⁷ Feigl (1956) claims that Feigl (1934) defends a typical form of unity-of-science Physicalism.

2.3 The Schlick-Feigl Identity Theory

When thus liberated Feigl came to employ Russell's distinction between knowledge by acquaintance and knowledge by description to articulate and update Schlick's identity solution to the mind-body problem:

the physical sciences consist of knowledge-claims-*by-description*. That is to say that the objects (targets, referents) of such knowledge claims are "triangulated" on the basis of various areas of observational (sensory) evidence. What these objects are acquaintancewise is left completely open as long as we remain within the frame of *physical* concept formation and theory construction. But, since in point of empirical fact, *I* am directly acquainted with the qualia of my own immediate experience, I happen to know (by acquaintance) what the neurophysiologist refers to when he talks about certain configurational aspects of my cerebral processes (p. 450).

It is here, in connection with Feigl's idea that we are directly acquainted with the qualia of our own experiences that the peculiar nature of his identity theory emerges. It is perhaps the peculiar nature of the Schlick-Feigl theory that led Kim to claim that Schlick never articulated such a theory. There is, at any rate, a certain tension in Feigl's formulations of his identity theory that are a clue to its peculiar nature. On the one hand, Feigl offers fairly straightforward statements of identity; for example, he writes that "The identity thesis which I wish to clarify and to defend asserts that the states of direct experience which conscious human beings 'live through' ... are identical with certain ... aspects of the neural processes in those organisms" (p. 446). Along the same lines, he speaks of "The identification of raw feels with neural states" and how the identification in question "identifies the referents of subjective terms with the referents of certain objective terms" (p. 448). But on the other hand, he curiously tends to prioritize the mental side of the identity, as in: "The 'mental' states or events (in the sense of raw feels) are the referents (denotata) of both the phenomenal terms of the language of introspection, as well as of certain terms of the neurophysiological language" (p. 447). And in a later paper, he writes that "I take these referents [of mental and

neurophysiological terms] to be the immediately experienced qualities” (1960, p. 38; cf. 1963, pp. 262, 257). Now, one might think that these latter formulations are not to be taken too seriously and that they are just careless ways of re-stating the identity thesis. But Feigl is careful never to engage in the opposite prioritization of the physical over the mental and say such things as “neurophysiological processes are designated by raw feel terms” or that “the common referents of both neurophysiological terms and raw feel terms are neurophysiological processes.” That there can be no mistake is evident, I think, from the following statement: “According to the identity thesis the directly experienced qualia and configurations are the realities-in-themselves that are denoted by the neurophysiological descriptions” (1958, p. 457; cf. 474). So Feigl appears to hold a mentalistic form of the identity theory.

Unfortunately, it is not clear that such a position is coherent. Identity, after all, is a symmetric relation: if a given mental phenomenon is strictly identical with some physical phenomenon then surely that phenomenon can be no more mental than it is physical. To claim that a given psychophysical identity theory is mentalistic or idealistic would be like saying that a Hesperus-Phosphorus identity theory is Phosphoristic in that it somehow gives prominence to the “Phosphorus side” of the identity; or like saying that while heat is identical with molecular motion, one less reductive version of the identity favours the “heat side,” while the other more reductionist version promotes the primacy of the “molecular” side of the identity. None of this makes any sense.

Perhaps a psycho-physical identity theory could be said to be more mentalistic than a materialistic rival if it were a form of panpsychism, that claimed that everything in existence, including things we do not ordinarily think of as mental in any way, were also

mental, such as plants and stones and sub-atomic particles. And it is here indeed that we seem to come to the difference between Feigl's identity theory and Smart's and Place's. Strikingly, for Feigl, "sentience (qualities experienced, and in human beings knowable by acquaintance) and other qualities (unexperienced and knowable only by description) [are] the basic reality" (1967, p. 107). The concepts of theoretical physics "denote realities which are unknown by acquaintance, but which may in some way nevertheless be not entirely discontinuous with the qualities of direct experience" (*ibid.*, p. 40; cf. 1971, p. 308). According to Feigl's own understanding of the doctrine, however, it is not strictly speaking panpsychism, "for the simple reason that nothing in the least like a psyche is ascribed to lifeless matter" (1960, p. 39). Stones do not have "selves" then. This is fair enough but it is potentially misleading. For Feigl is prepared to apply the label "pan-quality-ism" to his view that the phenomenal qualities we are directly acquainted with may well be instantiated throughout all of nature—not just in brains and nervous systems. Russell (1956) similarly viewed the question of "pan-quality-ism" as wide open: "since we know nothing about the intrinsic quality of physical events except when these are mental events that we directly experience, we cannot say either that the physical world outside our heads is different from the mental world or that it is not."³⁸ Feigl very consciously drew inspiration from both Schlick and Russell in this regard.³⁹

Nevertheless, while pan-quality-ism expands the range of physical things that are mental, it does not make this universal class of mental beings any more mental than they are physical. Identity is a symmetric relation. There remains the possibility that the "identity" at stake in the Schlick-Feigl identity theory is intended to be, like Place's

³⁸ See also Russell (1927).

³⁹ Feigl (1975) explores the similarity between the two philosophers.

(1956), the so-called “identity of composition.” As Stubenberg notes, Feigl “seems to say that matter is made of qualia” (p. 143). Since composition is an asymmetric relation perhaps it is a way to make sense of Feigl’s claim that phenomenal qualities are the “basic reality.” It is not clear, however, that this interpretation sits very well with Schlick’s and Feigl’s Fregean way of stating their identity theory, that is, with the claim that physical and mental terms refer to the same things. Suffice it to say that if this puzzle at the heart of the Schlick-Feigl theory remains unresolved, the Austrian version of the identity theory is in danger of lapsing into incoherence or collapsing into its Australian rival. Let us set this problem aside and look at some further interesting differences between the two views.

2.4 Reconceiving the Physical

As Stubenberg emphasizes, the two versions of the identity theory target different sides of the mind-body problem as the source of the difficulty of accepting the identification. The Australian version—as well as its radicalized offspring, the Feyerabend-Rorty eliminativism and the later deflationary, neo-Rylean behaviouristic instrumentalism of Daniel Dennett—finds the mind-side of the dichotomy to be the culprit and seeks to deflate it, either by exposing its committal of phenomenological fallacies, its involvement in conceptual incoherence, or by re-interpreting it in a “neutral” (usually causal-functional) form more amenable to physicalist reduction. In contrast, the material or physical is relatively unproblematic. The Austrian version, however, sees the physical side of the dichotomy as the villain and is accordingly more prepared to question

conceptions of the physical. As Feigl remarks, “I am convinced that it is primarily the concept of the ‘physical’ that requires reinterpretation and reconstruction” (1967, p. 142).

The problem, in Feigl’s eyes, is that we wrongly think of the physical in misleadingly “intuitive” terms, by which he appears to mean primarily in terms of images, usually visual ones. The physical states, events and processes with which the identity theory identifies phenomenal states, events and processes should not be thought of imagistically or pictorially as literally grey brain matter or nervous tissue. We are not acquainted with the physical; we have only theoretical knowledge by description of it. Schlick had already offered precisely this diagnosis: “The worst mistake that can be made in viewing the psychophysical problem—a mistake that, strangely enough, is made time and again—is, without noticing it, to substitute for the brain processes themselves, which are to be regarded as identical with the mental processes, the perceptions or images of the brain processes” (1925/1979, p. 300; cf. pp. 311-313). This too is Feigl’s view:

even sophisticated analytic philosophers tend to confuse the meaning of physical concepts with the perceived or imaged appearance of physical things. No wonder then that we are told that the identity of certain neurophysiological states (or features thereof) with raw feels is a logical blunder. If the denotatum of “brain process (of a specified sort)” is thus confused with the appearance of the gray mass of the brain as one perceives it when looking into an opened skull, then it is indeed logically impossible to identify this appearance with ... raw feels (p. 454; cf. 1963, p. 258).

According to Feigl, then, the mind-body problem—more accurately, the sentience-body problem—is to be solved by realizing that the physical is nothing like what we intuitively think it is. To suppose that the identity theory implies that one person could literally see the “raw feels” of another person, their pains and experiences of red, because the first person can see the brain states of the second person they are identical with, is to confuse the data of sensory experience, which serves as the evidentiary confirmation base for physical theory, with what the data are evidence *for*, namely, the *true referents* of

neurophysiological terms, which can only be conceived of abstractly and theoretically by description and not visualized on the basis of acquaintance. Feigl's view thus seems to be that the alleged features of the physical that make it problematic to identify raw feels with, are not really features of the physical at all. They are the sensory evidence for the physical; hence they are in fact mental features—in short, they are the raw feels, the qualia of which we are directly aware. Once the physical has been thus “thinned out” (in Stubenberg's nice phrase) there can no longer be any objection to identifying “its referents with something directly given and knowable by acquaintance” (Feigl, 1958, p. 454).

So on Feigl's view, what happens is something like this. We look at a subject's exposed brain, or at some of his nervous tissue under a microscope, and see what we take to be its various features: its grey colour, spatial expanse, bumpy and folded texture, etc. An identity theorist then says that the subject's experiences of redness when he looks at a ripe tomato, his pains and after images and other raw feels, are in fact the very physical processes in his brain that we are now looking at. We balk at this, wondering how a pain or experience of red could be a brain process, an occurrence in the grey, bumpy, moist object we are looking at: the subject's experience is red but his brain is grey, after all—to identify them is a logical blunder! Feigl, following Schlick, replies that we have confused sensory evidence for brains with brains themselves. Brains and other neurophysiological phenomena—indeed, the entire physical realm—do not really have these features: they are not grey, they are not bumpy, they are not moist; on the contrary, these are mental features mistakenly projected by us onto the true referent of the neurophysiological term ‘brain’, which has none of these features. The mistake we have made is to try to identify

one raw feel, a visual experience of red say, with another, a visual experience of grey, which we take to be a feature of the brain we are looking at. But now that the true physical nature of the brain has been sufficiently abstracted, by relocating its problematic and “identification-resistant” features to the mind, there is no problem with identifying all the raw feels in question with processes going on in the brain, that is, with processes going in the unvisualizable and unintuitable real physical brain.⁴⁰

2.5 Austria vs. Australia: Nil, Nil

In Stubenberg’s view, “It is the Austrian Version’s ability to better satisfy the requirement of phenomenological adequacy that makes it more deserving of our acceptance [than the Australian]” (p. 143). According to him, the view “combines a profound respect for phenomenology with an unconditional acceptance of a rational, scientific view of the world. And nothing less will do. For all theories that slight phenomenology are simply false, and we know that.” He concludes that “if we want to avoid blatant falsity and lunacy we must, somehow, combine phenomenology with science without short-changing either. And Schlick and Feigl have shown us how to do just that” (p. 144).⁴¹

I am not so sure. Perhaps Schlick and Feigl have shown how to reconcile phenomenology with the deeply counter-intuitive nature of fundamental theoretical physics at its quantum and cosmological scales. Indeed, for Feigl, there do appear to be two exhaustive categories of being: the mental as phenomenal and whatever is described

⁴⁰ Cf. Schlick (1925/1979), pp. 311-13.

⁴¹ Cf. Hatfield (2004) who admires the “respect for the phenomenal” found in Russell, William James and various theoretical physicists in the first two decades of the twentieth century, while deploring the disrespect for it found in Smart and company.

by theoretical physics (Feigl 1971, p. 309). But the status and nature of the subject matter of the higher-level special sciences, such as biology and psychology, is rather harder to ascertain on the Austrian model. Phenomenology and physics may have been saved and reconciled. But we seem to have lost the world of ordinary experience in which we live and breathe. According to Schlick and Feigl, it is a grave mistake to think of brains and other macro-physical objects as actually possessing the features they appear to possess when we are looking at them and touching them (this is what blocks acceptance of the identity theory by creating logical blunders). The visual and tactual qualities of the brain are merely the brain surgeon's sensory *evidence* for the real brain, the "thing-in-itself," whose true nature is only to be revealed by the theoretical descriptions of micro-physics. Now, while the phenomenology of after-images, floaters and double-vision clearly presents them as subjective features of our individual consciousnesses, colours, sounds, textures and the rest of the so-called secondary properties certainly are presented as "out there" on the surfaces of macro-physical objects—at least the majority of the time. Perceptual reference, in other words, is more often than not to ordinary macro-physical objects, such as brains, sofas, and mountains. It is very hard to see how the Schlick-Feigl view can accommodate the mundane intentionality of phenomenology. So it is not immediately clear that phenomenology has not in fact been at least a little short-changed on the Schlick-Feigl view.

I think that despite his avowed rejection of the kind of phenomenalism associated with the early days of logical positivism, Feigl retained a quasi-phenomenalistic, or perhaps quasi-Kantian philosophy.⁴² Kantian language pervades both Schlick's and Feigl's

⁴² Aune (1966) critically discusses Feigl's phenomenalism in the context of his identity theory.

writings on the mind-body problem. They often invoke the phrase “things-in-themselves,” even if contrary to Kant, they think “things-in-themselves” are knowable, knowable, that is, only by highly abstract and mathematized theoretical description and inference. It is clear that ordinary macro-objects, such as brains, are not things-in-themselves. It is not clear what has become of them. Moreover, both Feigl and Schlick seem to view not just the traditional secondary qualities of classical phenomenalism as mind-dependent but the space within which these qualities exist too.⁴³ According to them, part of the solution to the mind-body problem necessarily involves distinguishing between two different kinds of space, physical space, effectively a metric space, and phenomenal space, effectively a field of qualities. It quite unclear what these spaces are ontologically speaking and especially how they are related. Even aside from the Kantianism, Feigl’s phenomenalistic tendencies are evident in his unquestioning acceptance of Russell’s view that the only things we can be acquainted with are the sensory qualities present in immediate experience. Everything else, including what we take to be ordinary physical objects, we know only by description, and by a highly abstracted and mathematized theoretical description, presumably only understood by physicists, at that. It is very hard to make room in Feigl’s view for the idea that we might be acquainted with ordinary physical objects too; that we can make perceptual reference to brains and other macro-physical objects. On the Australian version of the identity theory, if a person’s feeling of pain is literally identical with a certain neural event or process in his brain, and one were small enough to observe that neural process, then one

⁴³ Schlick (1916/79). Kim (2003) discusses this peculiar idealistic feature of Schlick’s view.

would literally be seeing that person's feeling of pain.⁴⁴ Rather than bite the bullet and accept the consequence that we can literally see another person's raw feels, because we can see the brain processes they are identical with, Feigl ends up in the equally (if not more) counter-intuitive position that we do not ever really see brains or any rate that brains are nothing like what we take them to be. To my mind, it is not at all clear that the Austrian theory has, as Stubenberg maintains, avoided "blatant falsity and lunacy."

2.6 Two Perennial Philosophies

Whether or not this is a faithful account of Feigl's view, there can be little doubt that crucial to the Austrian view is a particular epistemology, associated with certain forms of Cartesian empiricism: we are directly and indubitably aware of data presented to us in conscious experience (Feigl 1958, *passim*, 1975, p. 15; Schlick, 1916/79, p. 197).

Phenomenal properties are simply given to us in experience. This is a first principle beyond dispute. It is something we know before we engage in science and philosophy and it cannot be overthrown by subsequent scientific or philosophical theorizing.

Phenomenal qualities exist, pure and simple, and this datum is non-negotiable. This Austro-Cartesian tradition, according to which the phenomenal is unproblematically given to us in experience, and our focus must fall on re-conceiving the nature of the physical in order to accommodate it, is continued to this day in the work of several contemporary philosophers, all united by the idea that, in the face of the undeniable existence of phenomenal qualities, the key to solving the mind-body problem may be to

⁴⁴ A consequence embraced and made much of by Perry (2001).

re-conceive the physical.⁴⁵ These latter-day Feiglans have no truck with any lingering phenomenalist epistemology and thus, I believe, make an advance on the earlier Austrian tradition. Crucially, however, like the Austrians, they maintain that the physical, but not the phenomenal, is allowed to be re-conceived because, unlike the phenomenal, it is not something whose intrinsic nature we are directly acquainted with and can therefore claim to know.

Smart's and Place's background philosophy is very different and I think that at least in the case of Smart, it would be fair to describe it as a kind of pure untainted scientism, influenced by Quine's extreme scientism. From Smart's point of view, there is no "first philosophy" prior to scientific theorizing; certainly there is no legitimate Cartesian-style epistemological-phenomenological prioritizing of philosophy over science. According to Quine and Smart, there is only shifting and ever-developing theory—no pre-theoretic or pre-scientific intuitions that must be preserved at all costs. Contrary to the Austrian tradition, the Australian tradition has it that our current conception of the physical is in fine working order and the problem lies with the mental, in particular with various phenomenological fallacies and conceptual incoherencies we are prone to succumb to when we think of it and try to describe it. The flames of this profoundly anti-Cartesian philosophy-as-continuous-with-science approach, which seeks to inflate the physical and deflate the mental, were fanned and taken to the limit by Richard Rorty's Wittgenstein- and Sellars-inspired view that the so-called philosophical "intuition" that we are presented with immediately given phenomenal qualities—"respect for the phenomenal,"

⁴⁵ Maxwell (1978), Lockwood (1981), Chalmers (1996), Unger (1999), Stoljar (2001), Strawson (2006). It is interesting to note in this regard that Sellars (1981), a close collaborator of Feigl's, and the famous enemy of the "the given" in experience, was squarely on Feigl's side in this respect, positing phenomenal qualities as a basic and irreducible part of reality.

in Hatfield's (2002) phrase—is nothing more than a collection of deeply ingrained linguistic habits we acquired when we were taught the language-game of incorrigible sensation reports.⁴⁶ The tradition is expanded in scope to include propositional attitudes as well as sensory consciousness, in the work of the later eliminative materialists of the 1980s.⁴⁷ The most sophisticated development of this tradition is found in the work of Daniel Dennett (1969, 1979), work which combines speculative empirical theorizing simultaneously with ingenious attempts to expose alleged conceptual incoherencies in our ordinary mentalistic notions.⁴⁸

The disagreement between the two identity theories over the viability of a materialism based on our current conception of the physical is thus the product of a deeper disagreement about the nature, status and role of pre-theoretical phenomenological intuitions in philosophical theorizing about the mind. It is unlikely that we will see in our lifetimes a victor in this great battle in contemporary analytic philosophy of mind between Cartesian phenomenology and deflationary scientific naturalism—or indeed in the lifetimes of the next few generations of philosophers of mind.

⁴⁶ See Rorty (1965, 1970, 1979, 1981).

⁴⁷ E.g., Churchland (1981) and Stich (1983).

⁴⁸ Rorty (1981) discusses the opposition between the Wittgenstein-Ryle-Sellars-Dennett tradition, according to which we have “no intuitions, no ‘initial facts’, which all theorizing must always respect, about the mind” (p. 343), and the Descartes-Broad-Nagel-Searle tradition, according to which we do have such intuitions. The two identity theories are also a reflection of this opposition and the recent revival of panpsychistic forms of materialism, such as Strawson (2006), and their opponents, is the latest manifestation of this opposition.

REFERENCES

- Alston, William and George Nakhnikian (eds.) *Readings in Twentieth Century Philosophy* (New York: The Free Press of Glencoe, 1963).
- Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.
- Armstrong, D. M. 1993. 'Preface to the Paperback Edition', *A Materialist Theory of the Mind*, revised ed. London: Routledge.
- Aune, Bruce. 1966. 'Feigl on the Mind-Body Problem', in Paul K. Feyerabend and Grover Maxwell (eds.) *Mind, Matter, and Method* (Minneapolis: University of Minnesota Press).
- Ayer, A. J. (ed.) 1959. *Logical Positivism*. New York: Glencoe.
- Beckermann, Ansgar. 1992. 'Introduction — Reductive and Non-Reductive Physicalism', in A. Beckermann, Hans Flohr, and Jaegwon Kim (eds.) *Emergence or Reduction?* Berlin: Walter de Gruyter.
- Borst, C. V. 'Editor's Introduction' to Borst (ed.) 1970. *The Mind/Brain Identity Theory*. Macmillan, St. Martin's Press.
- Burge, Tyler. 2007. 'Philosophy of Mind: 1950-1990', in his *Foundations of Mind*. Oxford: Clarendon Press.
- Block, Ned (ed.) 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, Ned. 1980. *Readings in the Philosophy of Psychology*, Vol. 1 (Cambridge, Mass.: Harvard University Press)
- Carnap, Rudolf. 'Unity of Science'. 1932/34. Reprinted and translated by Max Black (London: Kegan Paul, 1934) from the original German article 'Die physikalische Sprache als Universalsprache der Wissenschaft', *Erkenntnis* 2 (1932): 432-465.
- Carnap, Rudolf. 1933/1959. 'Psychology in Physical Language', trans. George Shick in Ayer (1959) from the original German article 'Psychologie in physikalischer Sprache', *Erkenntnis* 3 (1933): 102-142.
- Carnap, Rudolf. 1934/2002. *The Logical Syntax of Language*, trans. Amethe Smeaton (Chicago: Open Court, 2002), which is a reprinting of the original 1937 translation of the 1934 German original.

- Carnap, Rudolf. 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, Rudolf. 1936-37. 'Testability and Meaning', *Philosophy of Science* 3 (1936): 419-471 and 4 (1937): 1-40.
- Carnap, Rudolf. 1938/1991. 'Logical Foundations of the Unity of Science', *International Encyclopedia of Unified Science*, vol. 1, no. 1 (Chicago: University of Chicago Press, 1938). Reprinted in Richard Boyd, Philip Gasper and J. D. Trout (eds.) *The Philosophy of Science* (MIT Press, 1991).
- Carnap, Rudolf. 1956. 'The Methodological Character of Theoretical Concepts', *Minnesota Studies in the Philosophy of Science* I.
- Carnap, Rudolf. 1963. 'Herbert Feigl on Physicalism', in Schilpp (1963).
- Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.
- Chisholm, Roderick. 1955-56. 'Sentences about Believing', *Proceedings of the Aristotelian Society* 56: 125-148.
- Cirera, Ramon. 1993. 'Carnap's Philosophy of Mind', *Studies in the History and Philosophy of Science* 24: 351-58.
- Churchland, Paul. 1981. 'Eliminative Materialism and the Propositional Attitudes', *The Journal of Philosophy* 78: 67-90.
- Cornman, James. 1971. *Materialism and Sensations*. New Haven and London: Yale University Press.
- Dennett, Daniel. 1969. *Content and Consciousness*. London: Routledge.
- Dennett, Daniel. 1979. *Brainstorms*. Cambridge, Mass: MIT Press.
- Ducasse, Curt. 1941. *Philosophy as a Science. Its Matter and its Method*. New York: Oskar-Piest.
- Feigl, Herbert. 1934. 'Logical Analysis of the Psychophysical Problem: A Contribution to the New Positivism', *Philosophy of Science* 1.4.
- Feigl, Herbert. 1950. 'The Mind-Body Problem in the Development of Logical Empiricism', *Revue de Internationale de Philosophie* 4. Reprinted in in H. Feigl and M. Brodbeck (eds.) *Readings in the Philosophy of Science* (New York: Appleton-Century Crofts, 1953).

- Feigl, Herbert. 1958. 'The "Mental" and the "Physical"', *Minnesota Studies in the Philosophy of Science* II.
- Feigl, Herbert. 1960. 'Mind-Body, *Not* a Pseudo-Problem', in Sydney Hook (ed.) *Dimensions of Mind*. New York, NY: Collier Books.
- Feigl, Herbert. 1963. 'Physicalism, Unity of Science and the Foundations of Psychology', in P. A. Schilpp (ed.) *The Philosophy of Rudolf Carnap*. La Salle, Illinois: Open Court.
- Feigl, Herbert. 1967. 'Postscript After Ten Years' in H. Feigl, *The 'Mental' and the 'Physical'* (Minneapolis: University of Minnesota Press, 1967).
- Feigl, Herbert. 1971. 'Some Crucial Issues of Mind-Body Monism', *Synthese* 22: 295-312.
- Feigl, Herbert. 1975. 'Russell and Schlick: A Remarkable Agreement on a Monistic Solution of the Mind-Body Problem', *Erkenntnis* 9: 11-34
- Feyerabend, Paul. 1963. 'Materialism and the Mind-Body Problem', *Review of Metaphysics* 17: 49-66
- Fodor, Jerry. 1968. *Psychological Explanation*. New York: Random House.
- Goodman, Nelson. 1963. 'The Significance of *Der Logische Aufbau Der Welt*', in Schilpp (1963).
- Hardcastle, Gary L. 2007. 'Logical Empiricism and the Philosophy of Psychology', in Alan Richardson and Thomas Uebel (eds.) *The Cambridge Companion to Logical Empiricism*. Cambridge: Cambridge University Press.
- Hatfield, Gary. 2002. 'Sense-Data and the Philosophy of Mind: Russell, James and Mach', *Principia* 6.2: 203-30.
- Hatfield, Gary. 2004. 'Sense Data and the Mind-Body Problem', in Ralph Schumacher (ed.) *Perception and Reality. From Descartes to the Present*. Mentis, Paderborn.
- Heidelberger, Michael. 2003. 'The Mind-Body Problem in the Origin of Logical Empiricism', in Parrini, *et. al.* (2003).
- Hempel, Carl G. 1935/1972. 'The Logical Analysis of Psychology', reprinted in Ausonio Marras (ed.) *Intentionality, Mind, and Language* (University of Illinois Press, 1972). Translated from the French original in *Revue de Synthese* (1935) by Wilfrid Sellars in H. Feigl and W. Sellars (eds.) *Readings in Philosophical Analysis* (Appleton Century Crofts, 1949). Reprinted with prefatory note in Block (1980).

- Hempel, Carl G. 1952. *Fundamentals of Concept Formation in Empirical Science*. Chicago: University of Chicago Press.
- Hempel, Carl G. 1954. 'A Logical Appraisal of Operationism', *Science Monthly* 1.79: 215-20
- Kim, Jaegwon. 1971. 'Materialism and the Criteria of the Mental', *Synthese* 22: 323-345.
- Kim, Jaegwon. 1998. *Mind in a Physical World*. Cambridge, Mass.: MIT Press.
- Kim, Jaegwon. 2003. 'Logical Positivism and the Mind-Body Problem', in Parrini, *et. al.* (2003).
- Kim, Jaegwon. 2011. *Philosophy of Mind*, 3rd ed. Oxford: Westview Press.
- David Lewis (1966). 'An Argument for the Identity Theory', *Journal of Philosophy* 63:17-25.
- Lockwood, Michael. 1981. 'What *Was* Russell's Neutral Monism?', *Midwest Studies in Philosophy* VI: 143–158
- Livingston, Paul M. 2004. *Philosophical History and the Problem of Consciousness*. Cambridge: Cambridge University Press.
- Maxwell, Grover. 1978. 'Rigid Designation and Mind-Brain Identity', in C. Savage (ed.) *Perception and Cognition: Minnesota Studies in the Philosophy of Science* 9.
- Neurath, Otto. 1931. 'Physicalism: The Philosophy of the Viennese Circle', *The Monist* 41: 618-623.
- Parrini, Paolo, Wesley C. Salmon and Merrilee H. Salmon (eds.). 2003. *Logical Empiricism. Historical and Contemporary Perspectives*. Pittsburgh: University of Pittsburgh Press.
- Pap, Arthur. 1952. 'Semantic Analysis and Psycho-Physical Dualism', *Mind* 61: 209-221.
- Pap, Arthur. 1958. *Semantics and Necessary Truth*. New Haven: Yale University Press.
- Patterson, Sarah. 2008. 'Philosophy of Mind', in Dermot Moran (ed.) *The Routledge Companion to Twentieth Century Philosophy*. London: Routledge.
- Perry, John. 2001. *Knowledge, Possibility, and Consciousness*. MIT Press.
- Place, U. T. 1956. 'Is Consciousness a Brain Process?', *British Journal of Psychology* 47: 44-50.

- Place, U. T., 'Identity Theories'. n. d. *A Field Guide to the Philosophy of Mind*, online at: <http://host.uniroma3.it/progetti/kant/field/>.
- Putnam, Hilary. 1963/1975. 'Brains and Behaviour', in R. Butler (ed.) *Analytical Philosophy, Second Series* (Oxford, 1963). Reprinted in Hilary Putnam, *Philosophical Papers*, vol. 2. (Cambridge: University of Cambridge Press, 1975).
- Reck, Erich H. and Steve Awodey (eds. and trans.) 2003. *Frege's Lectures on Logic. Carnap's Student Notes 1910-14*. La Salle, Illinois: Open Court.
- Rorty, Richard. 1965. 'Mind-Body Identity, Privacy, and Categories', *Review of Metaphysics* 19: 24-54.
- Rorty, Richard. 1970. 'Incorrigibility as the Mark of the Mental', *The Journal of Philosophy* 67: 399-424.
- Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Rorty, Richard. 1981. 'Contemporary Philosophy of Mind', *Synthese* 53: 323-348.
- Rosenthal, David M. 1971/87. 'Introduction' to D. Rosenthal (ed.) *Materialism and the Mind-Body Problem*. Hackett, 1987 (Englewood Cliffs: Prentice-Hall, 1971)
- Russell, Bertrand. 1927. *Analysis of Matter*. London: Routledge and Kegan Paul.
- Russell, Bertrand. 1956. 'Mind and Matter', in Russell's *Portraits from Memory*. London: George Allen and Unwin, 1956.
- Schilpp, P. A. (ed.) *The Philosophy of Rudolf Carnap*. La Salle, Illinois: Open Court.
- Schlick, Moritz. 1916/1979. 'Ideality of Space, Introjection and the Psycho-Physical Problem', in Schlick, *Philosophical Papers*, vol. 1. Dordrecht: Reidel.
- Schlick, Moritz. 1925/1974. *General Theory of Knowledge*, trans. Albert E. Blumenberg. New York: Springer-Verlag. Translated from the 2nd original German edition of *Allgemeine Erkenntnislehre* (1925). 1st edition published in 1918.
- Schlick, Moritz. 1935/1949. 'On the Relation Between Psychological and Physical Concepts', in H. Feigl and W. Sellars (eds.) *Readings in Philosophical Analysis*. New York: Appleton-Century Crofts. Translation from the French original 'De la relation entre les notions psychologiques et les notions physiques', *Revue de Synthese* (1935).
- Searle, John. 2004. *Mind: A Brief Introduction*. Oxford: Oxford University Press.
- Sellars, Wilfrid. 1981. 'Is Consciousness Physical?', *Monist* 64: 66-90.

- Smart, J. J. C. 1959. 'Sensations and Brain Processes', *Philosophical Review* 68: 141-56.
- Smart, J. J. C. 1963. 'Materialism', *Journal of Philosophy* 60: 651-62
- Stich, Stephen. 1983. *From Folk Psychology to Cognitive Science. The Case Against Belief*. MIT Press.
- Stoljar, Daniel. 2001. 'Two Conceptions of the Physical', *Philosophy and Phenomenological Research* 62.
- Stoljar, Daniel. 2010. *Physicalism*. London: Routledge, 2010.
- Strawson, Galen. 2006. 'Realistic Monism: Why Physicalism Entails Panpsychism', *Consciousness Studies* 13: 3-31.
- Stubenberg, Leopold. 1997. 'Austria vs. Australia: Two Versions of the Identity Theory', in K. Lehrer and J. C. Marek (eds.) *Austrian Philosophy, Past and Present: Essays in Honour of Rudolf Haller*. Dordrecht: Kluwer.
- Suppe, Frederick. 1977. 'The Search for Philosophic Understanding of Scientific Theories', in Suppe (ed.) *The Structure of Scientific Theories*, 2nd ed. University of Illinois Press.
- Uebel, Thomas. 1995. 'Physicalism in Wittgenstein and the Vienna Circle', in K. Gavroglu et al. (eds.) *Physics, Philosophy, and the Scientific Community: Essays in the Philosophy of Science and Mathematics in Honor of Robert S. Cohen*. Dordrecht: Kluwer.
- Unger, Peter. 1999. 'The Mystery of the Physical and the Matter of Qualities: A Paper for Professor Shaffer'. *Midwest Studies in Philosophy* 23:75-99.
- Watson, Richard. 1993. 'Shadow History in Philosophy', *Journal of the History of Philosophy* 31: 95-109.