



Carnap, Knowledge of Other Minds, and Physicalism.

Document Version

Final published version

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Uebel, T. (2021). Carnap, Knowledge of Other Minds, and Physicalism. *Philosophers' Imprint*, 21(34), 1.

Published in:

Philosophers' Imprint

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



Carnap, Knowledge of Other Minds, and Physicalism

Thomas Uebel

University of Manchester

© 2021 Thomas Uebel

*This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivatives 3.0 License.
<www.philosophersimprint.org/021034/>*

1. Introduction

Logical empiricist philosophy of mind has had an unenviable reputation. Widely viewed as behavioristic in orientation with a keen eye for the fallacies of ordinary language, but without appreciation of the subtle communication it affords, its contributions to standard issues in the field are widely disregarded as irrelevant to current concerns. This paper provides a detailed reconstruction of the early treatment of the problem of knowledge of other minds by Rudolf Carnap. My aim is to correct a widespread misunderstanding of its failure and how it was overcome.

What merits the attention here is that the failure in question is highly instructive, for it prompts us to amend an otherwise convincing and indeed agenda-setting treatment of the entire history of the problem of knowledge of other minds. Anita Avramides states that “in their rush to banish traditional philosophical problems, the logical positivists simply adopt a naturalist stand from which mind has been extruded” (2001, 179). To be sure, they left the matter unfinished insofar as they did not, on this issue, sufficiently systematically challenge the Cartesian framework which is widely held responsible for enabling philosophical skepticism. Even so, Carnap did not leave “entirely unresolved” the question “how we are to think of mind’s place in nature” (ibid.). I will argue that the difficulties he faced concerned precisely the formulation of a stance that sought *not* to extrude mind from the natural order. How Carnap ultimately made significant headway deserves more notice than it has received so far.

So much to motivate digging deeply in seemingly barren ground. A basic point to be noted right away is that the theorists of the Vienna Circle were far from ignorant of the doctrine of intentionality and accepted it in its psychologically descriptive version as characterizing mental phenomena by their object-directedness or aboutness. Appearances to the contrary are due to the fact that they also sought, with varying degrees of success, to naturalize the phenomenon in question (Uebel 2020). This naturalization, pursued under the heading of “physicalism,” is not always easy to fathom. In light of the Circle’s

dismissive attitude towards metaphysics in general, the deflationary pronouncements by Carnap and fellow physicalist Otto Neurath on what knowledge of other minds is knowledge of, and how it is justified, were widely misunderstood. But far from depreciating it, they were concerned with its integration into the fabric of the sciences as a whole, into “unified science”: Neurath’s interest lay with the social sciences at large (Uebel 2019a), Carnap’s with individual psychology. Here the focus is on Carnap and his “physicalization” of psychological discourse.

I begin with Carnap’s replacement of appeals to unvarnished intuition by rational reconstructions that appeal to a common object domain shared by all the sciences. This strategy brings into focus troubling questions raised by his alleged logical behaviorism. Already this was not what it seemed to be, and I argue for a non-reductive interpretation of Carnap’s approach to psychological terms. I then go on to investigate his 1928 account of knowledge of other minds in detail and show it to be far more robust in principle than is at first apparent. Yet the account also exhibits a debilitating shortcoming which is only overcome in an advanced version of his physicalism that rendered his non-reductionist ambition operational. I will argue that in developing this response Carnap had to anticipate, albeit in a limited way and in his own “scientific” idiom, certain later moves in the Wittgensteinian vein. This story takes us from 1928 to 1935 and, as readers will note, also turns standard objections to physicalism upon their head.

2. Unified Science’s Rejection of Rationally Unreconstructable Intuition

Carnap’s dual campaign to exhibit the rational ground of all empirical knowledge claims by “rational reconstruction” and reject bare intuition as a legitimating instance of scientific reason when it remained unreconstructable can be clearly discerned in his first major work, *Der Logische Aufbau der Welt* (*The Logical Structure of the World*, 1928a/1967, hereafter *Aufbau* or *A*). There he developed a theory of “constitutional systems,” systems of concepts analyzed so as to render explicit what

remains implicit in cognitive practice: the evidential basis of their application. “The constitutional system is a rational reconstruction of the entire formation of reality, [a process] which, in cognition, is carried out for the most part intuitively” (*A* §100, 158; cf. §179, 289).¹ While the *Aufbau* shows no concern with justifying individual knowledge claims, it provides the logical foundations for doing so by outlining a general theory of such constitutional systems and developing one in relative detail. It provided a “genealogy of concepts” that indicates their “constitution” by simpler constituents, ultimately by logical iteration and permutation of one primitive relation. One aim of Carnap’s complex project was to explain and substantiate the objectivity claim of science in a new way; another was to “overcome the separation of unified science into unrelated special sciences” by constructing a “unified system of all concepts” that reconceptualized different “types of objects” as belonging to “different levels” of ascending complexity (*A* §§2 and 4, 7, and 9). No provision of a foundationalist epistemology was intended.

The concept of other minds — “the heteropsychological” — was explained briefly in the *Aufbau*, but knowledge of other minds was discussed further only in his *Scheinprobleme der Philosophie* (“Pseudo-problems in Philosophy,” 1928b/1967, hereafter *Scheinprobleme* or *S*), a slim volume intended to provide an accessible presentation of rational reconstruction in application. “The aim of epistemology,” understood as “the formulation of a method for the justification of cognitions” (*S* §1, 305), was explored by way of the examples of the problem of other minds and the realism dispute. The conceptual framework developed in the *Aufbau* was presupposed in broadest outline, but the reasoning no longer concerned the constitution of concepts and objects of

1. I follow the restoration adopted in Alan Richardson’s forthcoming revised translation of the *Aufbau* of “constitution” and “constitutional system” as translations of Carnap’s *Konstitution* and *Konstitutionssystem*. Note that I include, for references to *Aufbau* and *Scheinprobleme*, the section numbers before the page numbers of the 1967 translation. Square brackets in quotations are by the present author, as are translations from sources where none is indicated in the bibliography.

thought but knowledge claims and their justification. Verificationism moved to the forefront.

What attracted Carnap's attention to the issue of knowledge of other minds was not skepticism — he always rejected the "philosophical" question of other minds as cognitively meaningless (*S* §11, 335–336 and 1963, 888). What prompted him were the methodological claims made by opponents of the doctrine of the unity of science who believe that a sharp epistemological (if not also ontological) difference had to be drawn between different types of empirical science. This dispute was briefly commented on in the manifesto of 1929, *The Scientific World Conception: The Vienna Circle*, written mainly by Carnap and Neurath. Stressing that "unified science is envisaged as a goal," it dismissed claims about the power of empathetic knowledge, of *Verstehen* via *Empfindung*, under the heading of "intuition":

Intuition (*Intuition*) which is especially emphasized by metaphysicians as a source of knowledge, is not rejected as such by the scientific world-conception. However, it requires that all intuitive knowledge claims are followed by a rational justification step by step. Any method may be used by those who search; but what has been found must stand up to testing. The view that attributes to intuition a superior and more penetrating power of knowing, which is capable of leading beyond the contents of sense experience and which must not to be confined by the shackles of conceptual thought — this view is rejected. (Carnap, Hahn, and Neurath 1929/2012, 83–84)

What brought on the rejection of intuition as an autonomous source of knowledge was the distinction between, on one side, knowledge claims of everyday life that can be justified by making explicit the reasoning implicit in them and, on the other side, claims for a separate type of knowledge — namely one seeking to escape "the shackles of conceptual thought" — that cannot be so justified. According to the Circle's manifesto, the former claims can be retained while the

latter must be rejected. These include not only claims for metaphysical intuition but also certain types of claims for knowledge of other minds — yet by no means all.

Legitimate claims to knowledge of other minds, Carnap held, are those that can be rationally reconstructed. Such reconstructions involve the relative justification of claims, the demonstration that they can be inferred from another, more basic claim, together with knowledge already acquired (*S* §2b, 310). Knowledge of other minds, Carnap stated, is inferential knowledge: it is justified either by inference from reports issued by another person (E_1) or from observation of another person's expressive motions or acts (E_2) or, occasionally, from one's knowledge of the other person's being in certain external conditions (E_3). "There is no other way to gain knowledge of the heteropsychological." Most notably, "in each of the cases, E_1 , E_2 , E_3 , the cognition of the heteropsychological is connected with the perception of physical facts" (*S* §4, 317; cf. §5, 319–320). So in each case justification was achieved by inference from intersubjectively observable facts, behavior, or conclusions previously arrived at in this way. Clearly, for such inferences to work, the behavior must stand in some kind of indicator relation to the psychological state identified. Only if those relations obtained was knowledge of other minds possible. Of what nature then were those indicator relations on which the cognition of other minds depended?

3. The Issue of Logical Behaviorism

Carnap's analysis raises the specter of behaviorism, threatening the significance of knowledge of other minds altogether. To be sure, neither the "psychological" behaviorism of Ivan Pavlov and John B. Watson nor the "philosophical" or "logical" variety of behaviorism deny the very existence of mental states outright. Yet behaviorist psychologists deny any explanatory relevance to introspection and intentional phenomena, something logical behaviorists were loath to do even though they hold that talk of people's mental states reduces to talk of their behavioral manifestation. For them, the cognitive content or meaning of the former is exhausted by the latter. *Scheinprobleme* and papers by

Carnap — and by Herbert Feigl (1934) and Carl Hempel (1935), which are disregarded here — from the early to mid-1930s seem to lend themselves to such a reading readily. Thus Hilary Putnam (and many other philosophers following him) characterized “the Vienna Positivists in their ‘physicalist’ phase” as holding that “mental states are logical constructions out of actual and possible behavior events,” a view which “implies that all talk about mental events is translatable into talk about actual or potential overt behavior” (1965, 2). The trouble with this view — even if it forgoes full translations and merely requires, as Putnam noted, “analytic entailments” — is that it falls victim to the so-called Super-Spartan objection according to which on this view no pain could be ascribed to people who suppressed all behavior indicative of pain (*ibid.*, 9). Implausibly, logical behaviorism requires types of state of mind to be strictly correlated with types of behavior.² But was Carnap a behaviorist of this sort?

Already in the *Aufbau* (§164) Carnap signaled agreement with Bertrand Russell’s *Analysis of Mind* (1921). Unlikely to have extended to his ontological concerns or his psychologistic refashioning of propositional attitudes, this agreement concerned Russell’s naturalistic account of mind which contradicted neither the existence nor the causal relevance of mental states but questioned the importance of consciousness for intentional phenomena and regarded them as keyed to stimuli and behavior. Russell’s self-confessed “behaviorism” differed from that of Watson and, in turn, inspired the “behaviorism” that featured in the discussions of the Vienna Circle which, in about 1930, morphed seamlessly into discussions of what Neurath dubbed “physicalism.” But was Carnap’s version of it, to repeat, an instance of logical behaviorism?

Important help is found in the literature, but no full answer. Sean Crawford (2013, 2015) provided a careful analysis that debunked “the

2. Putnam’s diagnosis was elaborated by a systematic critique of behaviorism in Fodor (1968, Ch. 2), a foundational text for the philosophy of cognitive science. Unlike in Putnam’s paper, logical positivists were not even mentioned but were clearly in the frame. Logical positivism or empiricism and logical behaviorism remained closely associated in the literature ever since.

legend” of Carnap’s logical behaviorism in the 1930s, thereby offering essential clarification.³ In light of the fact, however, that Crawford did not address the question of Carnap’s logical behaviorism prior to the 1930s, and that Avramides read the *Aufbau*’s methodological solipsism as an attempt to “avoid the trap of behaviourism” (2001, 173), presumably as one then betrayed in *Scheinprobleme*, another close look at the issue recommends itself — especially as the latter is central to my concern here.⁴

Scheinprobleme claimed that the recognition of “heteropsychological occurrences” depends on “perceptions of physical events” (*S* §5, 320–321), which typically are behavioral events. However, the *Aufbau* also allowed for the possibility, in principle, of constructing heteropsychological objects (mental states of others) on the basis of “brain events” (*A* §140, 216), yet Carnap only returned to this proposal in the early 1930s in the debates about physicalism (e.g., 1932c/1959, 159 and 175) when he re-invoked it but repeated that, due to the stage of current science, it was not yet a practicable method of justifying the ascription of mental states to others. Instead Carnap then helped himself to ascribing states like “seeing red now” wholesale to the bodies of persons and characterizing these states again in terms of dispositions to behavior (1932b/1934, 85–87). This may look like logical behaviorism, but

3. Previously, Kim (2003) had called this traditional categorization into question but muddied the waters with a metaphysical interpretation of Carnap’s physicalism and elsewhere referred to Carnap (1932c) as “important” for logical behaviorism without qualification (2011, 87). Still earlier, Cirera (1993) had challenged as unsustainable the analyticity imputed by the logical behaviorist interpretation to Carnap (1932b, 1932c), a point wrongly disputed by Tamminga (2005, 654n.16). Finally, Kitchener also noted (1999, 401n.5 and 404) Carnap’s indicator relations to be based on empirically established verification conditions, but his later “the thesis of physicalism or logical behaviorism claims that every psychological event is logically equivalent to a physical concept (or word)” (2004, 43) again muddied the waters. Needless to say, I have withdrawn my own attribution of residual logical behaviorism (2010, 205–206).
4. Notably, Carnap’s terse answer to Ayer’s related criticisms also does not help, for it focuses on correcting the claim that his views on other minds had never changed and does not address whether Ayer was right about his early views (see Ayer 1963 and Carnap 1963, 886–889).

the fact alone that Carnap also entertained the “reduction” of mental state talk to talk of brain states gives a first indication that he was not the logical behaviorist Putnam portrayed him as.

To be sure, the *Aufbau* does treat heteropsychological objects as logical constructions out of behavior (A §§57–58, 92–94): their reconstruction out of brain events remained an unactualized constructional possibility. So how does this sit with my denial? The *Aufbau* states that “the heteropsychological is (even intuitively) apprehensible only as the meaning of an utterance (of an expressive motion or a sign production). *The meaning of an utterance is a unique function of the physical properties of the utterance* (‘function’ in the mathematical, not in the psychological sense)” (A §143, 221, orig. emphasis). So the question is whether “the function” in question expresses an analytical entailment or an empirical fact: What accounts for the correlation between the motion and the meaning? Note that it follows from Carnap’s assumption of psychophysical parallelism — as an empirical fact obtaining, not as a metaphysical thesis (A §22, 39) — that such correlation exists. Far from begging philosophical questions (alleged by Ayer 1963, 272), this is an example of the fact that Carnap required his logical reconstructions to be informed, at least in outline, by empirical results (e.g., A §103, 162–163; S §3b, 214). Consequently, the behavioral “definitions” of heteropsychological phenomena envisaged (none were actually given, only the general route of arriving at them was outlined) must be regarded as based on empirical hypotheses about the correlation of two sets of assertions. The analyticity of these characterizations in the *Aufbau* was only apparent and entirely derivative.

Already in the *Aufbau* then Carnap regarded the indicator relations involving either behavioral dispositions or brain states as inductively established generalizations. Note, moreover, that from *Scheinprobleme* onwards Carnap allowed that psychological ascriptions on the basis of sign productions and expressive motions “may rest upon an error,” for the report relied upon “can either be a lie or an error” and “pretense is always a possibility” (S §5, 320; cf. 1932c/1959, 171). So the indicator relations on which Carnap based the ascription of mental states to

another were recognized as fallible and far from error proof.⁵ Since it makes little sense to speak of definitions as fallible, this confirms that the indicator relations were not regarded as analytical in nature. In sum, Carnap’s relevant assertions from *Scheinprobleme* onwards contradict the attribution of logical behaviorism as an analytic and *a priori* meaning theory for mental state terms. It is not surprising then that in due course Carnap himself deployed what amounts to the Super-Spartan objection when he spoke of “a person of strong self-control” as “able to suppress these symptoms,” i.e., behaviors, and stated that psychological states and observable events “are not identical” (1938, 57 and 59, noted by Crawford 2013; see also Carnap 1935a, 92–94).

What accounts for the fact that the *Aufbau*’s provision of necessary and sufficient conditions for its constituted objects does not conflict with my claim that Carnap did not endorse logical behaviorism is the *Aufbau*’s extensionalism. Note that what the *Aufbau* calls “translations” and “definitions” only demand extensional equivalence (A §50, 83–84). (Rational reconstructions do not seek to render the full meaning of the statements under analysis or give a psychologically realistic picture of the cognitive processes involved.) Carnap retained this usage throughout most of the 1930s, e.g., in his discussion of what sentences about other minds “mean” (1932c/1959, 172). Even the tightened demand for nomological adequacy of these equivalences did not involve him, as a Humean, in modal notions to sustain the correlations between behaviors and mental states and thereby spoil his extensionalism. We must also not be misled by Carnap’s retrospective talk of psychological terms as “explicitly definable” (1963, 886): this has to be understood as referring to his mistaken understanding of disposition terms as definable in observational terms that was corrected only in

5. In the English translation of Carnap’s paper, the reactions characterizing a bodily state of seeing red, say, are said to be “usually regarded as necessary and sufficient criteria for anyone to be ‘seeing red now’” (1932b/1934, 86). This misleading turn of phrase is due to the translator Max Black. The original reads “gewöhnlich als Kennzeichen dafür, dass jemand ‘jetzt rot sieht’” (1932b, 457) and says nothing about necessary and sufficient conditions but is compatible with the quasi-criterial interpretation I develop below.

1935 (see his 1936 and 1936–1937). Carnap's extensionalism, his exclusive concern with "logical value" (A §50, 84), is consistent with his reconstructions not trading in analytic definitions or entailments as would be required by logical behaviorism as a meaning thesis.

To be sure, we must differentiate between the *Aufbau* and the works that came after: while much remained the same until the embrace of comprehensive physicalism (physicalism without an autopsychological net) in late 1932, one significant change was introduced already by *Scheinprobleme*. As noted, the primary task of the *Aufbau* was the development and exemplification of constitution theory, the general theory of the constitution of objects. Its epistemological task was limited to reflecting in its example system's order of constitution what Carnap deemed to be the order of epistemic priority (and to illustrating the thesis of the structural nature of knowledge). By contrast, *Scheinprobleme* and the early explorations of physicalism (until late 1932: see §5 below) directly pursued the epistemological problems of justification according to the *Aufbau*'s schema of a conceptual organization (S §6; 1930/1959, 144; 1932b/1934, 42–44). Yet importantly, even when the *Aufbau*'s schema was so presupposed, allowances had to be made for the vicissitudes of empirical inquiry. This meant that the epistemological explorations from *Scheinprobleme* onwards could not rely on the *Aufbau*'s specific idealized determinations. Epistemological application demanded the explicit recognition of the inductive origin of what the *Aufbau* reconstructed simply as definitions, for such application had to allow and account for the fallibility of our cognition of other minds noted in *Scheinprobleme* and presumed since. It follows that the behaviors and related dispositions serving as indicators had to be recognized as merely typically correlated with the mental states in question.

It may be wondered, of course, how such "merely typical correlations" between behavioral dispositions and mental states sit with the presumably exceptionless correlations presupposed by the hypothesis of psychophysical parallelism. The answer is that the latter are presumed to hold between states of a person's nervous system the details of which are as yet unknown (1932b/1934, 85) and mental states. The

gappiness of correlations between descriptions of behavior and mental states is due to the shortfall in precision incurred by the shortcut of using behavioral rather than neurophysiological descriptions and is typically glossed over by talk of the body of *S* being in a state of red-seeing (ibid., 87). (At this time Carnap thought of natural laws as strictly universal — "rules independent of space and time" [1932c/1959, 166] — and so was unable to entertain the notion of *ceteris paribus* laws.)

One worry remains: What are we to make of Carnap's talk of "reduction" (*Zurückführung*), so clearly in evidence throughout these years? The answer is that, as with "translation," such talk is idiosyncratic by our contemporary standards of use. Rather than specifying a meaning-theoretical dimension (let alone an ontological one), reduction talk had the function of indicating an epistemological relation, namely that of statements being linkable to an observation statement for purposes of testing in line with Carnap's verifiability criterion of meaningfulness as spelled out in *Scheinprobleme* (S §7). (To "reduce" a mental state attribution to its indicators was merely to indicate its testability.) Carnap's seeming behaviorism, we can see now, was an attempt to "operationalize" mental state attributions in this epistemological sense without forcing them into the Procrustean bed of full meaning equivalence with behavioral dispositions: eliminative reduction was not the aim, only the overcoming of a mistaken conception of mental phenomena as inaccessible to ordinary empirical investigations in principle (e.g., as emanations of a non-physical substance).

Worried about whether such a tentatively non-reductive approach (in our contemporary sense) to physical characterizations of mental phenomena can be brought off in general, we may note that the discussions of Wittgenstein's remarks about the criteria governing mental state ascriptions in his *Blue Book* and the *Philosophical Investigations* raise concerns not dissimilar to ones we can ask about Carnap's seeming behaviorism. (How strict are criteria and what logical relation do they stand in to what they are criteria of?) Concerning Wittgenstein's remarks on the matter, it has been argued (Albritton 1966) that to understand characteristic behaviors as indicative of certain types of

mental states is a “fact of nature” (or, perhaps better, of human “natural history”) but that no necessity attaches to this (the characteristic behaviors could have been different) and that, most importantly, “no entailment” obtains between such true descriptions of behavior and the possession of associated mental states, however close the connection established by the “use” of the “grammar” of the relevant expressions.

Clearly such a criterialist position is incompatible with physicalism as it is understood nowadays, namely as asserting any of various forms of mind-body identity or supervenience: there is no such thing as *ceteris paribus* identity or supervenience. Wittgenstein, of course, kept his distance from such “scientistic” theses as physicalism, so his position is unaffected, but what about Carnap? Here it is important to remember that for him physicalism was a metalinguistic thesis. In its settled form it held that the physical language was the only “universal” language such that every other language was translatable into it (up to extensional equivalence). Carnap’s physicalism did not assert identities but claimed instead that to every true sentence about a psychological state of affairs it was possible to coordinate a true sentence about a physical state of affairs. Since Carnap’s physicalism asserted correlations of types of descriptions (not first-order states), it could accommodate the idea that these correlations were not exceptionless but held only *ceteris paribus*. In principle then, Carnap’s physicalism, like his earlier position in *Scheinprobleme*, was supported by a criterial understanding of the physical translations of mental state ascriptions as their defeasible indicators. (Whether Carnap in fact extended this understanding beyond the behavioral also to the yet to be discovered neurophysiological descriptions is questionable, given his assumption of psychophysical parallelism as presumably lawful.) Yet we must guard against assimilating Carnap’s position too closely to Wittgenstein’s. Importantly, Wittgenstein questioned the causal nature of the relation between mental and dispositions to behavior, but Carnap did not (1938, 59; 1963, 887). Accordingly, I will speak of Carnap’s “quasi-criterial” understanding of the behavioral translations of mental state ascriptions.

Before looking deeper into Carnap’s early treatment of other minds, we must elaborate one significant shortcoming already noted in passing. Even if it is granted that eliminative reductions of mental states to behavior or behavioral dispositions were neither intended nor effected, it remains the case that until about 1935 Carnap’s understanding of dispositions was mistaken. Dispositions of all kinds resist the definitional reductions to observation statements of the sort apparently envisaged in the *Aufbau* and during the early stages of Carnap’s physicalism — to say nothing about truly “theoretical” terms. Of this failing Carnap cannot be absolved. But at least the specter of an undue behaviorism was banned. Along with this an important role of Vienna Circle anti-metaphysics is cast into relief: whatever “translations” and “reductions” were employed to support mind-body parallelism as an empirical thesis in Carnap’s *Aufbau* period, or later to underpin Carnap’s and Neurath’s doctrines of physicalism, were designed to let their scientific world conception comprehend whatever was comprehensible about *Geist*, to *integrate* the mental into the natural world that science investigates, not to exclude it.

4. Knowledge of Other Minds in *Scheinprobleme*

I return to *Scheinprobleme* to consider in detail how knowledge of other minds was accounted for by Carnap in the late 1920s. Fully in line with the conceptual architecture of the *Aufbau* here presupposed, Carnap distinguished sharply the autopsychological from the heteropsychological language: accounting for our knowledge of other minds meant accounting for how claims of the latter were justified in the former. Carnap’s use of indicator relations suggests that he followed a traditional route to conceive of knowledge of other minds as inferential — albeit without assuming the mind to be non-physical. Given the other person shows certain behaviors, I infer on the basis of an analogy between what behaviors “inner” states seemingly prompt in me that the other is equally possessed by such states when exhibiting like behavior. Such an argument was endorsed, albeit elliptically, by Carnap in a lecture of 1929: “The fact that I express by the sentence

'Mr. N feels joy' is not directly perceivable. But I can infer it from perceptions, e.g., from the observation that the facial expressions of Mr. N presented a certain look" (2004, 58). There is a fact of the matter of other minds, but that fact is not directly perceivable. Carnap also specified that knowledge of the heteropsychological can obtain only under the condition that correlations between behavior and mental states have been observed previously (*S* §4, 318). Since a stable correlation of behavior and mental states is presupposed by the argument from analogy as well, it makes sense to read Carnap as appealing to it in order to support *bona fide* knowledge claims about other minds. (The *Aufbau* employed analogies in the constitution of intersubjective worlds [*A* §§146–148], but the use of analogy in cognition, as in *Scheinprobleme*, is different.)

Yet Carnap's exposition of knowledge of other minds in *Scheinprobleme* is far from unproblematic. Critics point out that Carnap falls victim to the epistemological asymmetry that his own distinction between auto- and heteropsychological languages casts into stark relief: first-person mental state attributions are directly understood and verifiable without mediation, whereas third-person attributions are understood and verifiable only mediately via the recognition of that person's behavior. With a certain understanding of verificationism, other minds then drop out of the picture. The difficulty is to make sense of the idea of "getting at" the properly mental properties of others' states "through" their behavior. Clearly, we cannot do so if there is nothing more to mental states than dispositions to behavior. Accordingly, Carnap is also viewed as denying that we can know other *minds*.⁶

Carnap does have a case to answer. Merely stating that he was no logical behaviorist does not resolve a particular puzzle *Scheinprobleme* presents to the reader. Having noted that "in each particular case the recognition of the heteropsychological goes back to the recognition

6. See, e.g., Ayer (1963, 273–274), Avramides (2001, 175), and most recently Ambrus (2020, 359). For readings of Carnap as a logical behaviorist that seem to suggest but do not spell out this consequence, see, e.g., Lycan (1990, 4) and Chalmers (2002, 4).

of physical occurrences" and that "one could translate any statement about a given heteropsychological occurrence, for example 'A is now joyful', into a statement which mentions only physical occurrences, namely expressive motions, acts, words, etc.," Carnap summarized this state of affairs as follows: "we are confronted with two different languages, one of them the psychological and one physical; we maintain that they both express the same theoretical content" (*S* §11, 334–335). To the objection that "in the statement 'A is joyful' we express more than in the corresponding physical statement," he replied:

This is indeed the case. Aside from having the advantage of much greater simplicity, the psychological language also expresses more than the physical language, but this more does not consist of additional theoretical content; it expresses only accompanying representations; these are merely object representations, that is, representations which do not stand for any fact, and hence cannot form the content of a statement. ... [B]y saying "A is joyful" and not merely "A shows facial expressions of such and such a form" I express that I have a *representation of a feeling of joy*, although a feeling of joy in the autopsychological sense, since I cannot know any other. However, to believe that by using the psychological instead of the physical language, that is to say, by using the expression "joy" instead of "facial expression of such and such a form", we express a fact which goes beyond the physical state of affairs, is to confuse the theoretical content of the statement with an accompanying representation. ... No fact is even conceivable or stateable which could connect the representation "feeling of joy" (in the autopsychological sense) with the behavior of A. (*S* §11, 335, orig. emphasis)

It is hard to avoid the conclusion that Carnap in this passage denied that psychological attributions say more than behavioral statements,

for what is added does not amount to anything factual. This would negate his claimed recognition of other minds.

A lot needs explaining here. Evidently Carnap rejected the idea that projecting what is a representation of my own mental state (“a feeling of joy in the autopsychological sense”) onto another body should count as knowledge of another’s mental state: this is but a conceptual confusion. But what follows from it? Certain worries can be quickly allayed. When Carnap claimed that we “cannot know any other” than our own “feeling[s] of joy in the autopsychological sense,” he can be read as stating correctly that one cannot literally experience somebody else’s joy. And when he claimed that nothing can “connect the representation ‘feeling of joy’ (in the autopsychological sense) with the behavior” of another, he can be read as stating correctly that one’s own feeling has no bearing on what another’s behavior is indicative of. It also is clear that Carnap’s argument is directed against a use of the argument from analogy that tried to legitimize knowledge claims about minds of a radically non-physical, dualistic sort (“a fact which goes beyond the physical state of affairs”). Yet one cannot help but wonder whether his argument still allowed for any knowledge of other minds at all. Does it not deny that third-person mental state ascriptions “express more” theoretical content than the corresponding physical statement, and are we not then left, as regards the other, with statements about behavior alone? To see what is the case, we must look in greater detail at the rational reconstruction offered and what is meant by “theoretical content.” We need to understand Carnap’s taxonomy of experiential content.

As in the *Aufbau*, “rational reconstruction” in *Scheinprobleme* also does not aim to portray “the actual experience” but only “a certain logical dependency between certain constituents of the experience” (S §2b, 310). The distinctions drawn and constructions offered are not phenomenologically descriptive but serve purely analytical purposes. And as in the *Aufbau*, “translations” only aim for extensional adequacy. To say of two statements in the physical and the psychological languages

that they have the same content is only to say that they are true of the same things: no identity of meaning is claimed.

From such a reconstructive standpoint, experience possesses two kinds of content, “theoretical content” and “accompanying” or “object representations” (S §8, 329–330). Theoretical content is propositional and truth-valuable; by contrast, accompanying or object representations are non-propositional and not truth-valuable (but represent a grab bag of anything from bare sense impressions to personal memories or random thoughts to ideas of abstract concepts). If we are concerned to explain causally a person’s behavior, it is often to these accompanying representations that we must turn in addition to the contents of beliefs and preferences (S §11, 338). But the epistemologist is only interested in the theoretical content of experiences expressible in linguistic form.

Theoretical content equates to factual content. “Factual content” and related notions are defined as follows:

If a statement p expresses the content of an experience E , and if the statement q is either the same as p or *can be derived from p and prior experiential knowledge, either through deductive or inductive inferences*, then we say that q is “supported by” the experience E . A statement is said to be “testable” if conditions can be indicated under which an experience E would occur which supports p or the contradictory of p . A statement p is said to have “factual content”, if experiences which would support p or the contradictory of p are at least conceivable, and if their characteristics can be indicated. (S §7, 327, trans. amended, emphasis added).

This, of course, is Carnap’s and the Vienna Circle’s first formal verificationist criterion of meaningfulness (announced well over a year before Wittgenstein made his famous pronouncements on the matter

to Moritz Schlick and Friedrich Waismann).⁷ By Carnap's criterion, factual content is in principle, but not necessarily in practice, testable (and certainly not necessarily conclusively so). Note also that due to the condition here italicized, Carnap's factual content can far exceed what is directly observable. It follows that we must not allow claims such as "All statements must be reducible to perceptions or else are meaningless" (2004, 59) to mislead us to assume that only observation statements are meaningful.

The analysis of the theoretical content of an experience has "logical" and "epistemological" aspects. Logically speaking, theoretical content splits into "sufficient" and "dispensable" parts. What makes a constituent *b* dispensable "relative to" the sufficient constituent *a* is "that *b* does not give me any information that is not already contained in *a* together with my prior knowledge" (S §2b, 310). Three things need to be noted here. First is that a dispensable part remains cognitive and retains factual content. Being dispensable does not equate to being an object representation. Second is that the inference through which the content of a dispensable part can be found in a rational reconstruction typically appeals to one's prior knowledge and does not spell out a mere entailment of the sufficient part on its own. The ability to furnish such a reconstruction provides the criterion of the correctness of the logical analysis. Third is that "this logical analysis is frequently ambiguous" such that "the same experience can be analyzed in different ways" (S §2a, 309). Depending on the case at hand, I may also be able to start the analysis with *b* and derive *a* in conjunction with prior knowledge. Then *b* would be the sufficient part and *a* the dispensable one.

An epistemological analysis of experience does, unlike its logical analysis, not have this reconstructive freedom. According to it, theoretical content splits into a "nucleus" which carries the justificatory burden and a "secondary part." That is, the inferential relation in

question is now fixed in its direction so as to be able to legitimate knowledge claims. The logically sufficient part turns out to constitute the epistemological nucleus, while the logically dispensable part constitutes what is epistemologically secondary, but only one of the possible logical analyses can provide the blueprint for the epistemological analysis. Accordingly, knowledge of other minds has as its nucleus a perception of the other's behavior, yet as in the logical analysis, the inferences from nucleus to secondary part involve prior knowledge (and do not only spell out analytic entailments of the nuclei).

The epistemological analysis of experience is not only more constrained than its logical analysis but also subject to a modality unknown to the latter. Since the behavioral descriptions must be quasi-criterially understood, the epistemological analysis yields at best a contingent truth. This creates problems for Carnap's taxonomy of experiential content, as we shall see presently.

Note first that Carnap was happy to speak of "the theoretical content of *b*," with *b* representing a secondary part. This makes explicit what we discerned already: that Carnap did not mean to deny factual content to third-person mental state attributions. He allowed for what we could call (but Carnap did not) "secondary theoretical content" to distinguish it from the "primary theoretical content" possessed by the nucleus. So far, so good, but now note that Carnap's taxonomy brings us up short, for it follows from it that, given the fallibility of inferences from *a* to *b*, there is something about *b*—normally one would say: about its content—that differs from *a*. However, as we saw, Carnap was adamant that what *a* and *b* are placeholders for, physical and psychological sentences about the same episode ("Mr. A is angry"), have the same content. Something did go wrong.

A similar problem confronts us when we find Carnap speaking of the overdetermination of the content of our experience in a way that squarely seems to undermine his explanation of how we could come to know other minds: "We experience more than is necessary in order to gain the knowledge that can be obtained. This is to say, we can leave certain constituents of experiences unevaluated (fictional expression:

7. "The sense of a proposition is [the method of] its verification." See McGuinness (1979, 47, 79, and 97, inserted phrase only in the second occurrence) for these remarks of 22 December 1929, 2 January 1930, and 22 March 1930.

these constituents could disappear from our experience) and our knowledge would not be diminished" (S §2c, 311). Does this claim not suggest that psychological descriptions add nothing to descriptions of behavior? It does and it is hard to disagree with it if it is offered as a comment about "the logical character of the theoretical content of our experiences." However, considered in the light of an epistemological analysis, this claim is inadequate.

Confronted by the problematic claim that a psychological formulation does not possess "additional theoretical content" over and above the relevant behavioral description (S §11, 335), it would, of course, be correct to recall that rational reconstructions only aim for extensional adequacy, but this does not help when we ask how Carnap can account for and conceptualize the fallibility of third-person mental state ascriptions. (If a sentence *s* serves only fallibly as an indicator of the truth of another sentence *p*, then the contents of *s* and *p* cannot be identical.) Given how *Scheinprobleme* allowed for the specification of content, namely by indicating their verification conditions, it is impossible to specify the theoretical content of the secondary part, namely the mental content seemingly expressed by psychological sentences, as at all different from that of the nucleus, the sentences describing indicative behavior. Indeed, since Carnap claimed that any attempt to draw such a distinction is guilty of confusing cognitive factual content with non-cognitive accompanying object representations (S §11), it is not surprising that critics read him as a logical behaviorist.

We must conclude that the inability to specify differences between the primary and secondary theoretical contents that the demand of the fallibility of behavioral indicators requires fatally undermines the non-reductive ambitions of Carnap's conception of knowledge of other minds in *Scheinprobleme*. To assess this failure, we need to see what caused it. As Carnap was not a logical behaviorist, it is arguable that it was not rooted in his conception of other minds. The culprit appears to be rather the radically anti-psychologistic animus of his extensionalist method of rational reconstruction. It left Carnap without a conceptual vehicle for what his projected account of knowledge of

other minds required, namely something that distinguished secondary from primary theoretical content in the face of their extensional equivalence. What his account required was precisely what the constitutional system of the *Aufbau* was unable to supply. As Carnap put the matter there, "This is an essential characteristic of the constitutional method: as regards object names, statements, and propositional functions, it is concerned exclusively with logical, not with epistemic value; it is purely logical, not psychological" (A §50, 84, orig. emphasis). Needless to say, this approach is not helpful when we deal with epistemological questions and are concerned with epistemic value. It is precisely on their epistemic value that the intuitive difference in content between the nucleus and the secondary part turns — a difference Carnap clearly recognized. To dispose of it as merely psychologically relevant object representations was a very serious misstep.⁸

What we find then is that with the means at his disposal, inherited from the *Aufbau*, no substance can be given to the distinction between primary and secondary theoretical content, and the fallibility of inference from behavior to mental states perforce remains unsubstantiated (see especially S Appendix, 342). It is a difficult question whether Carnap's preference for the extensional method of analysis was justified in the *Aufbau* by relegating senses and intensional statements to the domain of psychology, namely as concerned with "a concept as the content of a representation or thought" (A §45, 77) and likewise by categorizing concern with "epistemic value" as merely "psychological, not logical in nature" (A §50, 84). But even if it was justified there, it was no longer justified in *Scheinprobleme*, which unlike the *Aufbau* engages directly in epistemology. For by denying the relevance of sense or intension and not providing a functional equivalent for it, Carnap

8. What appears to be a related mistake wreaked havoc with Ayer's version of logical positivism in his *Language, Truth, and Logic* (1936) and was diagnosed by John Foster as the identification of the core idea of verificationism as the "content principle" rather than the "evidence principle" (1985, 22), that what is meaningful must be observable rather than be suitably evidentially related to what is observable; for discussion, see Uebel (2021).

robbed himself of just what he needed to substantiate his taxonomy of experience and his epistemology of heteropsychological objects.

Confronted with an inconsistency between theses to which a work commits its author, how is an interpreter to decide which one to give priority to as closer to the authorial intention? The principle of minimal mutilation is not necessarily correct but is certainly defensible when applied appropriately in light of longer-term developments. Thus I take the significance of Carnap's early liberal verificationism and of his specific fallibilism with regard to mental state ascriptions to outweigh that of his austere early extensionalism. Accordingly, I will not blame the former for the resultant inconsistency but the latter. Having to hand nothing but his concept of factual content for purposes of non-psychologistic talk of meaning, it follows that the very difference is inexpressible, which the objector to the claim that physical and psychological sentences have the same content wants to express. But for Carnap neither his liberal verificationism nor his quasi-criterialist approach to mental state ascriptions were negotiable in the long term. Less than twenty years later, both were still endorsed by him — his verificationism was still further liberalized in 1935 to finally settle the issue just raised (see §6 below) — but his austere extensionalism was not (see Uebel 2020, Appendix).

That said, Carnap's *Scheinprobleme* model of knowledge of other minds remains very badly damaged. True, Carnap could correctly have claimed to have made a good start in preparing the ground for a non-reductive understanding of psychological statements about other people with his notion of quasi-criterial indicators and his taxonomy of theoretical experiential content. But his extensionalism prevented him from articulating what is importantly different about psychological statements compared to physical including behavioral ones and this forced him into incoherence: to declare third-person mental state attributions fallible in light of their behavioral indicators being satisfied yet refuse them theoretical content over and above that possessed by their behavioral indicator statements. So Carnap's analysis must be faulted not only for its underlying account of disposition statements.

Though it was not eliminativist towards other minds in intent, the promise that his taxonomy of theoretical content held out was not redeemed. The recognition of other minds does indeed involve their fallible inference from what the *Aufbau* called "expressive events" involving a body other than mine (*A* §140, 216), but the austere extensionalism retained from it rendered epistemic value, precisely what made psychological attributions distinctive and therefore important, utterly invisible.

A separate but not wholly unrelated worry is this. Carnap left unelucidated the remark that "it is a more and more widely accepted *insight* that the autopsychological and the heteropsychological have an entirely different epistemological character; at the present time this fact can be denied only if one holds to certain metaphysical *standpoints*" (*S* §4, 316). Carnap meant to draw attention to the difference between the direct way in which one can gain knowledge of one's own mental states and the indirect, inferential way in which one gains knowledge of other minds. Yet as we saw, the distinction is not as unproblematical as Carnap seems to have assumed. It flags up what Avramides has called the "conceptual problem" of other minds, as opposed to the standard epistemological one (2001, 218–230). What ensures that first- and third-person mental state ascriptions have the same meaning? How can we justify our assumption that they do? What would ensure that the secondary theoretical content of third-person mental state ascriptions is suitably similar to the primary theoretical content of first-person psychological reports? That *Scheinprobleme* is silent on this matter would standardly be taken to confirm the diagnosis that Carnap endorsed an untenable logical behaviorism, but as I've argued, this is wrong. Nevertheless, the problem remained.

It might be thought that the problem is even worse. For is it not the case that the epistemological asymmetry already on its own brought a difference in meaning in its train — and therewith right away dashed the hope to have vindicated the claim to have knowledge of other minds broadly like our own? Given some version of the verificationist theory of meaning, this is a very reasonable worry. (Indeed, presently

we'll see it activated by Carnap's early 1930s thinking about knowledge of other minds.) However, in 1928 this worry did not yet get traction — for the simple reason that the verification theory of meaning, deriving from Wittgenstein's remarks at the end of 1929, was not yet currency for Carnap until 1930. (In *Scheinprobleme*, Carnap put forward a verificationist conception of *meaningfulness*.) Until Wittgenstein's conception became currency, it remained possible for Carnap to think of the difference between first- and third-person mental state attributions solely in terms of the routes of their testability.

Finally, a point in Carnap's favor. Note how my diagnosis of the failure of Carnap's account relates to what many readers (including one of his editors: Patzig 1966, 121–124) found extremely disconcerting in *Scheinprobleme*. Carnap there also imagined two scientists, "one of them a solipsist, the other a non-solipsist idealist or realist," who agree "on the basis of empirical criteria of psychology whether A's joy is real or only simulated (empirical reality)" but disagree whether "A really has consciousness." Carnap concluded, "The divergence between the two standpoints occurs beyond the factual, in a domain where in principle no experience is possible; hence according to our criterion, they have no scientific significance" (S §11, 336). Clearly here the emphasis lies on imputing a sense of reality not addressed by the indicator conditions that already are assumed to be satisfied. For Carnap, such imputations are as in vain in making sustainable knowledge claims as intuitive object representations. Note also that his conclusion would hold even if he had been able to sustain the notion of secondary theoretical content, for what Carnap disputed the meaningfulness of here was not the question whether A has consciousness but rather the question whether, given that A has it has already been established by empirical means, A "really" has consciousness. That, coupled with his strict extensionalism, Carnap's verificationism got the better of secondary theoretical content does not redeem the emptiness of the metaphysical talk here opposed. Of course, that the difference in theoretical content of mental state descriptions from that of their indicators did vanish in his account contributed to the confusion: it misled many

readers to mistake his anti-metaphysical argument for an anti-mental-ist and eliminationist one.

5. The Transition to Mature Physicalism

It is striking that, in retrospect, Carnap called his position in *Scheinprobleme* a representative of "an early phase of physicalism" (1961/1967, x). As is known from analyses of his protocol-sentence debate with Neurath, Carnap's early phase of physicalism was far from unproblematic. Since *Scheinprobleme* endorsed the broad conceptual framework of the constitutional system of the *Aufbau* (S §6), it cannot have diverged with regard to the assumption with which physicalism took issue (methodological solipsism), so the question in just what sense *Scheinprobleme* represents "an early phase of physicalism" does not have an obvious answer. Yet we are unlikely to go far wrong if we take Carnap's remark to point to the crucial role that was assigned to behavior, to physical facts, now foregrounded by the focus on intersubjectively available evidence for knowledge claims about other minds.

The question arises, of course, whether the new emphasis was compatible with continued adherence to methodological solipsism. Here we must recall that the rational reconstruction of our ordinary conceptual system by the constitution system of the *Aufbau* was offered as an extensionally equivalent *simulation* of it (see A §45 and §§50–51). The simulation was intended to bring out certain structural features of human knowledge but not to recreate it *in toto*. Indeed, it could not do so as all the objects constituted in the *Aufbau* "consist merely in the reorganisation of the given" (A §148, 229, orig. emphasis). In particular, "the entire experience sequence of the other person consists of nothing but a rearrangement of my own experiences and their constituents" (A §140, 215, orig. emphasis; see also §144, 222, §145, 223, §147, 226, §160, 255). Given methodological solipsism, there was no transcending the phenomenal given.

Now imagine a critic who picks up on this limitation and challenges that knowledge of other minds involves a complexity that cannot be accounted for in Carnap's methodologically solipsist model. This

critic would argue that without object transcendence we cannot distinguish between ascriptions where only the behavioral indicators are satisfied and ones where also the mental description applies. This line of argument is *prima facie* suggestive. But that Carnap lacked a representational vehicle for secondary theoretical content with the help of which a misfiring cognition of the other could be simulated was due to the extensionalism of his reconstructive method. Since it is not obvious that methodological solipsism itself demands purely extensionalist analyses of language, it seems prudent to consider unproven the particular charge that methodological solipsism is responsible for the problem at hand.

Carnap, in any case, was not aware of the deficit of his account I have demonstrated, so he did not worry (as yet) about the cogency of applying his simulationist rational reconstruction to assessing concrete knowledge claims on its account. As it happens, however, the late 1920s and the early 1930s saw him responding (independently of the troubles of his account of knowledge of other minds) to various challenges that attempted to break down the apparent self-sufficiency that his reconstructive methodology seemed to enjoy by focusing on shortcomings allegedly owed to his methodological solipsism. In response, Carnap's epistemological views shifted more than once: a brief review of the doctrines in play and their development is in order since the changes also affected his account of knowledge of other minds — ultimately for the better.

As noted, for Carnap, physicalism was a meta-linguistic thesis stating the “universality” of the physical language, that “all” other languages can be translated into it. To characterize Carnap's development, four versions of physicalism must be distinguished (they will be further clarified in the text below): they all construe the universality in question and/or its relation to methodological solipsism somewhat differently.

Physicalism₁ was promoted from 1930 throughout 1931 and holds that all the languages of intersubjective science are translatable into the physical language, i.e., the language of physics (see Carnap 1930).

(Carnap's remark notwithstanding, given the lack of its specificity, it is unclear whether *Scheinprobleme* even falls under physicalism₁.)

Physicalism₂ was embraced in early 1932 and holds that all languages dealing with empirical matters without exception are translatable into this physical language (see Carnap 1932b and 1932c).

Physicalism₃ was embraced at the end of 1932 and holds, unlike in earlier versions, that for epistemological purposes it is no longer mandatory, but still possible, to employ a methodologically solipsistic protocol language (see Carnap 1932d).

Physicalism₄ was introduced in 1936. It holds that for purposes of the analysis of the language of science, methodological solipsism is abandoned and that all the languages dealing with empirical matters are “reducible” — in the special non-eliminative way of his misleadingly called “reduction sentences” — to the so-called thing language that speaks of intersubjectively observable objects, events, processes, properties, or relations (see Carnap 1936–1937, 10 and 1963, 869 and 944–945).

Note that one of the changes involved concerned the designation of “the physical language.” It changed from the coordinate language of mathematical physics as in the *Aufbau* and still preferred in *Unity of Science*, via the pragmatically motivated qualification that “[q]uantitative determination can also be replaced by qualitative, as is usual in science as well as in everyday life, for reasons of brevity and ease of understanding” on the condition that “they can be understood as determinations of physical states of affairs or occurrences” (1932b/1934, 53), to the ordinary “thing-language.” This latter position was further liberalized in the 1950s by the recognition of the irreducibility of the theoretical language and its terms (physicalism₅, as it were).

Turning to the other changes in a little more detail, we may first note that the development was set off by the critical question whether the methodologically solipsist constitution system of the *Aufbau* could account for the intersubjectivity of science.⁹ With scientific statements

9. This issue was raised, probably in December 1929, by a student member of

requiring translation into the autopsychological language in order to be understood and justified, different scientists did not really communicate, and science was robbed of its intersubjective basis. As long as autopsychological claims remained untranslatable back into the physical language — as indeed they were on the operative *Aufbau* model¹⁰ — intersubjective testing, the hallmark of scientific inquiry, dissolved into individualist given-gazing. In response Carnap adopted a new position, physicalism. Distinguishing “positivist” constitution systems with an autopsychological basis from “materialist” ones with a physical basis, he noted:

[They] do not contradict one another. Both are correct and indispensable. The positivist system corresponds to the epistemological viewpoint because it proves the validity of knowledge by reduction to the given. The materialist system corresponds to the viewpoint of the empirical sciences, for in this system all concepts are reduced to the physical, to the only domain which exhibits the complete rule of law and makes intersubjective knowledge possible. (1930/1959, 143–144)

Note not only the foundationalist outlook (alien to the *Aufbau* and abandoned again in 1932) but particularly that now both constitutional

the Circle, Heinrich Neider. For the argument, see Haller and Rutte (1977, 29–30); for its dating and discussion, see Uebel (2007, 130–136).

10. Already in the Circle, Edgar Zilsel remarked on this asymmetry of translatability in the *Aufbau* (1932, 145–146). That it is not noted more often may be due to a failure to separate the *Aufbau*'s general theory of constitution from the specific constitutional system developed. That Carnap noted that a constitutional system of concepts could also be erected on a physical basis (*A* §62) says nothing about the specific constitutional system developed in the *Aufbau*. Likewise, for the assertion that “every statement about a psychological object is translatable into a statement about physical objects” (*A* §57, 92): while it is possible in principle that both are, in the system developed in the *Aufbau*, given its architecture with an autopsychological base, only heteropsychological statements are translatable into physical ones. Related diagnoses of the asymmetry of translatability are indicated but not further elaborated in Feigl (1950/1981, 289), Kim (2003, 269), and Ryckman (2007, 95); for some discussion, see Uebel (2014).

systems are declared “indispensable.” This marks a change of emphasis from the *Aufbau* where the system with a physical basis was mainly recognized as “advantageous” because “most appropriate” for representing the conceptual order operative in empirical science itself, and only the small print stated that “science as whole ... needs both an experiential and a materialistic derivation of all concepts” (§59, 95–96). What rendered the physicalist constitution system indispensable for reconstructing empirical science was that “its basic domain was the only one to exhibit determinate nomological relations of its processes” (§59, 95); notably, the need to account for intersubjective knowledge was not yet seen as a reason for its indispensability until 1930. By implication, moreover, physical languages were now also recognized as meaningful on their own account.

It might be wondered why Carnap did not appeal at this juncture (late 1929/early 1930) to the inherent limitation of his simulation strategy in response to charges that the intersubjectivity reconstructed in the *Aufbau* was not “real” intersubjectivity. His reason appears to have been that he wanted his reconstruction to have some purchase also on cognition in the wild. So rather than claim heroically that his methodologically solipsist model simulated all cognitive activities — as the intersubjectivity objection presupposed — Carnap pulled back and clarified its range of applicability in a new two-languages model. (For documentation and discussion, see Uebel 2007, 191–200.) In unpublished drafts he now entertained a dualism of “universal” languages where the autopsychological language no longer had to account for all uses of the physicalist language and where the physical language had its universality limited to languages expressing states of affairs that are “intersubjectively recognizable”: the domain of the autopsychological language was excluded from its reach. For Carnap at this stage, the meaning of autopsychological statements was not captured by their translation into the physical language. His conception of their different offices explains why: the physical language provided no reduction to the phenomenal given. Carnap’s assumption of the epistemic order

assigned to the autopsychological language a significance which the physical language lacked.

Yet Carnap's model of two universal languages proved unstable, given the new role that verificationism came to play in the wake of Wittgenstein's then recent pronouncements to Schlick and Waismann. According to these, the process of verification was constitutive of meaning itself and required to be strict and conclusive. (Wittgenstein glossed this as "No, if you can never verify the sense of a proposition completely, then I cannot have meant anything with the proposition either. Then the proposition signifies nothing whatsoever" [in McGuinness 1979, 47].) Now whether Carnap fully accepted Wittgenstein's conception of verificationism, like Schlick and Waismann, and if so for how long, is debatable. What is clear, however, is that Wittgenstein's strictures also proved troublesome for Carnap's epistemology. They spelled trouble not only for all universal sentences and, given methodological solipsism, for any sentence of the physical language but also for statements about other minds. Recall that on Carnap's scheme, behavioral indicators only furnished fallible evidence for mental states: this fallibility now undercuts the very status of other mind talk as *bona fide* meaningful.

As it happened, this difficulty was solved when in 1931 Carnap decided to reject strict verificationism (1936–1937, 37n.; 1963, 57). With the fallibilism restored, other mind ascriptions once again became meaningful in their own right, for meaningfulness was again extended to any sentence from which a protocol sentence could be derived inductively. This meant that the phenomenal autopsychological language no longer needed to be the basis of all meaningful assertion and the physical language was again intelligible independently of it, as presumably intended by the two-languages model all along. Yet the difficulty of sustaining the intersubjectivity of evidence for assertions of the physical language was unresolved as long as methodological solipsism was retained for epistemological purposes and autopsychological statements remained untranslatable into physical ones. Carnap had to change further, and from the turn of 1931/32 onwards,

physicalism meant for him that the physical language is *the* universal language such that *all* other languages can be translated into it. This marked the start of physicalism₂.

Carnap overcame his previous objections against the translatability of the autopsychological language by introducing the distinction between the "material" and the "formal mode of speech": "The first speaks of 'objects', 'states of affairs, of the 'sense', 'content' or 'meaning' of words, while the second refers only to linguistic forms" (1932b/1934, 38). At issue was the proper form of metalinguistic discourse. Carnap asserted that while the material mode of speech was not strictly speaking wrong, it tended to mislead the unwary into metaphysical confusion and was best avoided. It prompted "pseudo-questions concerning the essence or reality of the objects mentioned in the definition of a language" (1932b/1934, 40). In the formal mode, the meaning or "content" of sentence was determined instead by the set of sentences derivable from it (1932b/1934, 87–88; 1932c/1959, 106).

The supposed non-translatability of the autopsychological language, upheld until then, was a case in point. As Carnap once put it to Neurath (who had urged him for some time to drop methodological solipsism altogether): "Only due to the sharp distinction and the rejection of the material mode has the elimination of the dualism of the two languages become possible."¹¹ Carnap's thought was that once talk of meaning (except in the guise of the formal mode) was considered misleading (like the object representations of old), any adversion to some supposed surplus meaning of first-person psychological reports was easily overruled as trading on misleading associations. Thus Carnap now argued that, on pain of remaining meaningless to others, autopsychological sentences had to be translatable into the physical language (1932b/1934, 77–81; 1932c/1959, 192–194). However, this did not yet mean that the epistemological priority of the autopsychological over

11. Carnap to Neurath, 2 March 1932, Rudolf Carnap Papers, Archive of Scientific Philosophy, Hilman Library, University of Pittsburgh, 029-12-60/61, p. 2, quoted with permission; cf. Carnap (1932b/1934, 74n.). The dualism at issue is that of the physical and autopsychological languages (see Uebel 2007, 243).

the physical language was abandoned as well (1932c/1959, 191). After continued opposition by Neurath, Carnap made this crucial change only in another step, in the Fall of 1932, when Carnap revoked the epistemological privilege of the constitution system with an autopsy-psychological basis: according to the not yet so-called Principle of Tolerance, the adoption of that system was no longer mandatory for the pursuit of properly understood epistemological inquiries (see Carnap 1932d). This marked the inception of physicalism₃.

Clearly, these changes altered Carnap's general epistemology of science, but how was knowledge of other minds affected? It may be thought that once Carnap had embraced physicalism₂ — a move which integrated first-person reports into the intersubjective language of science — he was able to resolve the worries that *Scheinprobleme* left us with. However, matters were not that simple. To see this, let's review how the problem of sustaining his non-reductive ambitions concerning other minds presented itself to him at this point.

6. Knowledge of Other Minds in Physicalism

Consider how knowledge of other minds was characterized in "Psychology in Physical Language." The fallibilist inferentialist conception of knowledge of other minds first outlined in *Scheinprobleme* was reaffirmed in the new physicalist setting. To be justified, a third-person psychological sentence must be shown to be derivable from a "perception sentence" describing physical behavior and a "major premise" which asserts a correlation between a person's perceived behavior and their mental state (1932c/1959, 171).

At least on the face of it, the old problems continued. Critics of Carnap's supposed logical behaviorism focus on his claim that a sentence about another mind 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 physical stimuli" (1932c/1959, 172; cf. 1932b/1934, 91). In this they see an inadvertent admission that talk of other minds does not amount to more than talk about other people's actual and potential behavior, and so, knowingly

or unknowingly, they reassert the charge that had already been raised against *Scheinprobleme*: that mind disappears. Carnap's continued invocation of the alleged confusion of theoretical content and accompanying object representations in defense of his assertion did little to quell such criticism — nor did Carnap's replies to various other objections.

Against the objection that neuroscience was as yet too undeveloped, Carnap noted that our present way of speaking is already physicalistic: it concerns what has intersubjectively discernible traces, otherwise we would not understand one another. Against the objections that an argument from analogy or from the use of empathy or from reliance on testimony or simply from the meaningfulness of behavior itself would support a non-physical interpretation of statements about other minds, Carnap argued that for such statements to be meaningful they have to be testable and for them to be testable there had to be criteria the satisfaction of which must be discernible intersubjectively — which undercuts the supposedly non-physical interpretations. Finally, the objection that the possibility to confirm a statement about my mind by introspection shows that no physicalist interpretation is required was met by the observation that for scientific purposes and for other people to understand such a statement it was precisely a physicalist interpretation that was required, and that for scientific purposes the interpretation by introspection is largely irrelevant. Yet none of these rebuttals — however good¹² — address the inability to distinguish the content of psychological from physical sentences in the face of their extensional equivalence.

To be sure, Carnap again does not appear to have acknowledged the problem — after all, he repeated, the "mere object representations" defense — but this did not make his problem less severe. In the *Aufbau* he

12. These are serious arguments that Carnap's critics tended to dismiss or misunderstand, like the argument against the traditional deployment of the argument from analogy. Carnap's rejection of the alleged conclusion "That person is angry" as "meaningless" was expressly qualified by "if its physical interpretation is rejected" (1932c/1959, 176): Carnap rejected the dualist but not the physicalist employment of the argument from analogy, a point that's often misunderstood. Instead, he was and continues to be read to deny the mentality involved, e.g., Weinzwieg (1962, 252) and Ambrus (2020, 364).

had relegated sense and epistemic value to the psychological domain, and even following his embrace of the semantic conception of truth in 1935 he stuck with a purely extensional construal of truth conditions until *Meaning and Necessity* in 1947. Facing the problem of distinguishing the content of extensionally equivalent expressions, which nowadays is met by invoking intensional notions of one variety or other, Carnap appears to have been badly stuck. Yet partly due to fortuitous circumstances he was about to find an at least temporary solution to his problem of other minds which did not involve appeal to Fregean senses or intensions as such. Conceptual resources were becoming available to allow him to specify the secondary theoretical content that his intended non-reductionist conception of mind required.

To begin with, the innovation of physicalism₂ – the translatability of the autopsychological protocol language into the physical language – paved the way for reworking the twice-severed connection between first- and third-person mental state ascriptions by also overcoming the obstacles presented by the epistemological asymmetry between them. For a start, a psychological language the intelligibility of which cannot plausibly be denied – the one in which I speak of my own experiences – was now deemed translatable into the physicalist language. This opened the door to exploration of the inferential relations in which first-person psychological statements, in their physical translation, could be seen to stand. But while this was necessary, it was far from sufficient for coming to terms with the epistemological asymmetry of first- and third-person psychological sentences such that the obstacle which it posed was overcome.

Another promising idea already noted was that according to Carnap's new formal-syntactic definition, the content of a sentence was conceived as the class of sentences that it is possible to derive from it (1932b/1934, 91). Importantly, such derivations made use of both logical and empirical laws. Relevant here for our purposes were statements of correlations deemed nomological between psychological and physical states. (Above I noted that Carnap treated behavioral placeholders more leniently than the neurophysiological states ultimately aimed for,

so we must differentiate between practice and what it seeks to approximate.) Given suitable empirical laws, it became possible in principle to hold that certain first- and third-person mental statements can share an identical class of sentences that follow from them. Sharing their content would then amount to sharing their meaning. However, what still barred this integration of first- and third-person psychological talk in practice was the hold that verificationism exercised. Even as liberalized to accept merely probable confirmation, it continued to reinforce the epistemological asymmetry between first- and third-person psychological sentences: being verified in so radically different ways, how could they possibly mean the same? Even if, like Carnap, one was not enamored of verificationism as a theory of what meaning consisted in, this asymmetry cast doubt on the supposed identity of meaning.

Help for this problem came from a perhaps unexpected quarter. Recall the innovation of physicalism₃: the demand that epistemological reasoning be rationally reconstructed with reference to the autopsychological language was dropped. Carnap revoked what had been his presumption all along, namely that methodologically solipsist reconstructions mirrored the epistemic order and that epistemological reconstructions had to respect this. Note what this change brought along in its train: verification lost its intrinsically first-person perspective and that perspective lost its epistemological privilege. Statements in the physical language could now be justified without mediation of the autopsychological language and, given further conditions, serve as test or protocol statements themselves. Now it became possible to make use of the idea that the same statement could be verified in different ways. Applied to psychological sentences, this idea offered a way of removing the obstacle posed by the epistemological asymmetry of first- and third-person psychological sentences.

Recall that, since physicalism₂, the concept of equality of content as equipollence had provided Carnap with a formal mode of speech replacement for the material mode notion of sameness meaning (in its extensional dimension). Given physicalism₃, he could develop a formal mode approximation of the notion of sense, as it were. Carnap

now set to differentiate the ways in which cognitive access was gained from distinct perspectives to the theoretical content that co-extensional sentences share: thereby he was able to capture the difference in the epistemic value of first- and third-person psychological sentences. Upholding his anti-psychologism, Carnap's extensionalism softened in one crucial respect: presumably having realized the inadequacy of his earlier response, he now was prepared to find a way to make the notion of epistemic value respectable.

To do so Carnap only had to effect revisions in concepts he had been using already. Even so, his solution was not published until his contribution to a themed issue of *Revue de Synthèse* in 1935, in which the challenge that the epistemological asymmetry posed to the treatment of the psychological language was addressed for the first time.¹³ The first revision was that the thesis of physicalism was said to be rendered "more precise" by replacing the condition of translatability with that of equipollence: "For every scientific sentence there exists an equipollent sentence in the physical language" (1935b, 45–46). This alone seems to amount to little more than a redescription of existing doctrine intended to redirect attention away from any intuitive notion of meaning associated with the notion of translatability and replace it with a precisely specifiable formal notion. That equipollence, the sameness of content of two sentences, was now specified as the shared class of *non-analytic* sentences derivable from them (1934a/1937, §49), does not make much of a difference, nor that, depending on whether only logical or also empirical laws were involved in that derivation, Carnap now spoke of L- or P-equipollence (*ibid.*, §52, 184–185). Ever since physicalism₂, the sameness of content of certain physical and psychological sentences of the auto- and heteropsychological languages was

13. Carnap's new way of facilitating intersubjective psychological discourse also received a brief mention, without any fanfare, later that year in his London lectures (1935a, 90–91). Indications are, however, that Carnap began to think along the lines of this proposal already in 1933 as he wrote in *Philosophy of Science* that "a proposition ... can be transformed in more than one way with equal content" (1934b, 16). That this stems from 1933 is suggested by the faulty listing of *Logische Syntax* to 1933 in the paper's bibliography.

a matter of P-equipollence, as the law involved in their mutual derivation was empirical. Even so, the focus on equipollence in place of translation was indispensable for the logical operationalization, as it were, of the notion of epistemic value.

It worked in conjunction with a corresponding second change, one made to the verificationist conception of meaning. As noted, Carnap's own earlier verificationism only served as a criterion for meaningfulness, and he showed as little interest in the "essence question" of meaning (what it consists in) as in all other essence questions (*A* §§20–21), and his endorsement of Wittgenstein's verificationism was short lived. Yet now he adapted his dictum and put it to new use. For Carnap the idea that "the meaning of a sentence is the method of its (possible) verification" provided a legitimate "logical concept of 'sense'" ("le concept logique du 'sens'") (1935b, 46, emphasis added). Note first that here Wittgenstein's verificationism was understood in the framework of the formal mode of speech: that "sense" was spelled out "logically" meant spelling it out in terms of verification conditions (not in psychological terms of ideation). Carnap sought to develop a notion to cohere with and complement the "syntactic" one of meaning as the set of non-analytic consequences.

Yet the ideas behind Wittgenstein's dictum also required a very important further expansion.

[O]ne should not, as is common, speak of "the" method of verification of a sentence, but of the class of possible methods of verification or better conditions of verification. For typically (maybe even always) a sentence can be checked in different ways; there are various verification conditions for it. By a verification condition for a sentence we shall mean a control sentence, which under certain circumstances may be a protocol sentence, which is derivable from that sentence with the help of scientifically recognized laws. Wittgenstein's meaning criterion must therefore be expanded such that the meaning of a

sentence is determined by the class of its verification conditions. (Ibid., 46–47)

Note how talk of classes of verification conditions and the syntactic explication of meaning coheres. With individual verification conditions identified with control sentences, speaking of different verification conditions for the same sentence occasions no conflict with speaking of the meaning of a sentence as the set of its non-analytic consequences, for clearly more than one control sentence can typically be derived from a meaningful sentence. Most importantly, first- and third-person psychological sentences can now be shown to have the same meaning as the same sets of control sentences can be derived from them (with the help of scientific laws) even though the control sentences their verification focuses on are different. It was by granting that one sentence can have different verification conditions and that one can distinguish between them that a solution to both the epistemological and the conceptual problems of other minds became available.

Suppose we have two sentences, one in the psychological language, (1) “I am angry now_{ps}” where “I” refers to Mister A, and one in the physical language, (2) “Mister A is angry now_{ph}” which specifies the bodily or behavioral conditions obtaining when (1) is true. The sentences (1) and (2) are P-equipollent because of the bridge-law (3): “(1) \leftrightarrow (2).” Clearly, however, the verification of (1), a first-person psychological sentence, is different from that of (2), the physical sentence that corresponds to it: (1) is verifiable by introspection and so verifiable only by the bearer of the putative state, A, whereas (2) can be verified by any observer B of A’s body. While the paths of verification for (1) and (2) are different, the class of verification conditions is identical for them.

The modified criterion leads to the same result for both sentences (1) and (2): the class of verification conditions of (1) coincides with that of (2). Every tester, whether it is A himself or another person B, can derive from (1) the same control sentences as from (2). Let C be the class of singular sentences about processes in the body of A

which can be derived from (2). Then (1) and sentences of class C are control sentences for (1) for A for the following reason. A can verify (1) directly (by introspection, as one puts it) but also indirectly, even if he is unlikely to do so given the possibility of direct verification. The indirect test would consist of A deriving (2) from (1) with the help of the law sentence (3) and then deriving from (2) the sentences C which he can test directly by observation. On the other hand, for A all control sentences for (1) are also control sentences for (2) as (1) is derivable from (2) with the help of (3). So for A the class of control sentences for (1) agrees with that for (2). According to the modified criterion therefore, (1) and (2) have the same meaning. Now for B the sentences of C belong to the control sentences of (2). (1) is not directly testable for B and so is not a control sentence, but it is indirectly testable. For B can derive (2) from (1) with the help of (3) and control sentences from (2), the sentences of C. So for B as well the class of control sentences for (1) agrees with that for (2). In consequence also for B the sentences (1) and (2) have the same meaning. Our consideration shows that the commonly expressed view that (1) is verifiable only for A but not for B is incorrect. The difference consists only in this that sentence (1) can be tested by A directly and indirectly, but by B only indirectly. But this difference does not amount to a fundamental one with regard to the logical nature of the two sentences. (Ibid., 47)

Now Carnap was able to admit the difference in method of verification without having to worry that this counted as a difference in meaning. Substituting P-equipollence for translatability in the formulation of what physicalism amounts to demands broadening the concept of verification from individual methods of testing to classes thereof, and this then allowed explaining how it was possible for the psychological

predicates in first-person and third-person ascriptions, indeed the auto- and the heteropsychological language, to share their meaning despite the epistemological asymmetry between them. Their content, conceived in terms of the set of non-analytical sentences that followed from them, was identical, but self- and other-ascribers verified or confirmed them in different ways and followed different derivational paths to their respective verification conditions.

No doubt, some eyebrows will be raised at this point: this account of different methods of verification as turning on different control sentences, different sentences within the set of non-analytic consequences, does not appear to be a purely structuralist account. To this the answer is that the account is about as structural as interpretation is syntactic, or better, it is as little purely structural as interpretation is purely syntactic. In other words, Carnap's "structural" solution of the problem of how to differentiate the content of co-extensional statements is as impure as his "syntactic" reconstruction of the language II of arithmetic in *Logical Syntax*. Of his reconstructive methods there, it is well known that they go beyond the resources of syntax alone and use what is now recognized as semantic reasoning (see Coffa 1977 and 1987). Carnap himself stated calmly, "The interpretation of a language is a translation and therefore something which can be *formally represented*; the construction and examination of interpretations belongs to formal syntax" (1934a/1937, §62, 228, orig. emphasis). So just as in his reconstruction of arithmetical languages, Carnap helped himself to semantic reasoning under the cover of syntax (illegitimately so if we employ our current understanding of the terms), so he helped himself to pragmatic ideas under the same cover to advance his conception of knowledge of other minds. It was in the use of the structurally representable content that speaker and hearers of first- and third-person mental attributions differed, not in the mere possession of it.

The epistemological asymmetry was explained and defused — rendered formally discernible — by the idea that speakers and hearers of first- and third-person psychological attributions make use of different control sentences that nevertheless belong to the same set of

non-analytic consequences derivable from the attributing sentence given suitable laws. With content conceived in formal logical terms to start with, such specifications of control sentences delivered the required differentiation without calling upon concepts radically alien to the analysis already underway.

To repeat, to make this advance towards understanding knowledge of other minds, Carnap's acceptance of physicalism₂ with its insistence that the autopsychological language was fully translatable into the physical language was a necessary condition, but it was not sufficient. What was required also, though easily overlooked, is the contribution made by physicalism₃. With concern about other minds an epistemological preoccupation par excellence, it was clearly vital that Carnap's old injunction that epistemological inquiries have to be conducted in a methodologically solipsist protocol language be repealed. This was a precondition for the further liberalization of verificationism that de-problematized the previously troublesome asymmetry between justifying mental state ascription to oneself or others. So it turns out that while methodological solipsism alone cannot be held responsible for Carnap's trouble with other minds in *Scheinprobleme* (see §5 above), it was only its overcoming that made a resolution of these troubles possible.¹⁴

7. Further Developments and Conclusion

All's well that ends well? Only in the fullness of time: while the (noted) failure of his analysis of disposition statements was also corrected by the middle of the 1930s, the failure to recognize the *bona fide* theoretical nature of mental state attributions was fully remedied only in the 1950s. To be sure, Carnap's basic conception of knowledge of other minds as first attempted in 1928 — namely as knowledge of psychological states broadly like mine in content but associated with another

14. Physicalism₄ clarified the nature of the empirical base by canonizing the use of the thing language and the practical impossibility of reconstructing the language of science on a methodologically solipsist base generally (1936–1937, 464) but added nothing to the account of knowledge of other minds.

person's body that is gained by empirical inductive inferences, including analogy, mainly from behavior — did not alter again because of this further change either (see, e.g., 1963, 888–889). As we saw, it was years after physicalism was first embraced, in print only by 1935, that Carnap was able to provide an account of the sameness of meaning of first- and third-person psychological statements that allowed extensionally equivalent statements to have different verification conditions. Yet this delay notwithstanding, it must be noted that, far from preventing the formulation of a proper account of knowledge of other minds, physicalism enabled Carnap's account finally to work as intended.

We may also note that for Carnap the much later change to recognizing psychological terms as theoretical instead of dispositional ones (1954, 1956) was not as momentous as it may appear. It did not mean granting their referents an epistemological status they did not possess before (from partially observable to unobservable, say) but rather doing better justice to the one that he had intended their ascription to possess all along. Thus Carnap noted that “the interpretation of scientific terms as pure dispositions cannot be easily reconciled with certain customary ways of using them,” for “a scientist, when confronted with a negative result of a test for a concept, will often maintain that it holds, provided he has sufficient positive evidence to outbalance the one negative result” (1956, 68). However, defining a concept in dispositional terms (as “pure dispositions”) has the consequence that on arriving at a negative test result for it, the ascription has to be withdrawn — no ifs and buts. By contrast, if it was viewed as a theoretical term, “the result ... of any observations, external or internal, is not regarded as absolutely conclusive evidence for the state in question” (ibid., 71), a position which was much to be preferred. Importantly, again, when he embraced the theoretical interpretation of psychological terms, Carnap did not abandon a previously held logical behaviorism, but by so re-classifying them, he further clarified his understanding of mental state ascriptions. It also allowed him to think of the argument from analogy that his account of knowledge of other minds

relied on as a species of inference to the best explanation and so avoid standard objections to it.

This is not to say, of course, that not much else had changed in Carnap's philosophy by the 1950s, beyond the full embrace of physicalism, since *Scheinprobleme*. To begin with, there is Carnap's embrace of semantics, later in 1935, that finalized his rejection of verificationism as a theory of meaning, even in the tenuous, non-essentialist and liberalized fashion that he had endorsed still at the beginning of that year (when he adopted Wittgenstein's dictum for expository purposes), and its retention as a criterion of meaningfulness only. Then there is, throughout the 1940s, his increasing acceptance of pragmatics as a component of the theory of meaning broadly construed. What is most important, however, and still further changed his already radical approach to philosophy, was that by the time Carnap abandoned the mandatory appeal to methodological solipsism for epistemological purposes, he also underwent the significant conversion to embrace the already mentioned “logical tolerance”: “*We are not concerned to issue prohibitions but wish to arrive at conventions*” (1934/1937, §17, orig. emphasis). No longer did he advocate rational reconstructions as determinations of what were deemed to be *the* facts of the philosophical subject matter under consideration, as he did from the *Aufbau* until “Psychology in Physicalist Language”; now he regarded rational reconstruction as a pluralist venture putting forth logico-linguistic explications of contested concepts and puzzling conceptions that answered, beyond fairly minimal constraints imposed by the original explicanda, to nothing but pragmatic concerns. “Everyone may construct their logic, i.e. their form of language, as they wish. But if they want to have discussions with us, they must clearly indicate their design and give syntactic specifications instead of philosophical elucidations” (ibid.). In light of this metaphilosophical development on Carnap's part, it may be wondered what status is possessed by what I advertised as his solution to the problem of other minds.

The answer is short. Now Carnap refrained from promoting his preferred treatment of knowledge of other minds as *the* solution to the problem (and none of my formulations above should be read contrarily). Philosophy offers logico-linguistic proposals and draws up possible ways of conceiving and speaking of different subject matters; it does not unveil the “true reality” of things. Carnap’s physicalist model of knowledge of other minds in 1935 is one such proposal, nothing more. But this does not mean, of course, that penetrating thought did not go into making it or that Carnap considered this proposal as good as any other. It made sense as far as it went – later he was to improve upon it, as noted – but already when it was put forward it made more sense of the phenomenon it treated than others, given the naturalistic starting point of the “scientific world-conception.” To begin with, it accounted for knowledge of other minds as interpersonal understanding without unexplained and unexplainable appeals to intuition. Moreover, Carnap agreed that the physicalist model was the only one that makes sense of our other minds talk. “The phenomenal or phenomenological language ... can neither refer to material objects nor to other minds” (1963, 869). These are weighty advantages.

When considering Carnap’s position on knowledge of other minds, it must also be noted, of course, that the very idea that such knowledge is theoretical at all has come under attack in recent decades. This attack discards both the argument from analogy and inference to best explanation in favor of a more direct quasi-perceptual route (see, e.g., Avramides 2009, Stout 2010, Smith 2015). In light of these developments, Carnap’s achievement in the mid-1930s may not seem to amount to much. But note that at least his inductivism on this issue was up to date at the time. (And still today it has not been rejected universally by philosophers, least of all by those reticent to accept unvarnished intuition as a legitimating source of knowledge claims.) Moreover, since he was not a logical behaviorist but all along regarded relevant behavior as a fallible indicator of mental states and also was happy to include interpreted speech behavior under that rubric, Carnap’s quasi-criterial behaviorism may also be considered “non-reductive” in the

sense of Gilbert Ryle’s later logical behaviorism. A remarkably undogmatic position emerged here from the Vienna Circle.

Last but not least, it must be noted that even before he hit upon his 1935 solution to it, Carnap showed himself sensitive to the problem of whether first- and third-person ascriptions still use the same psychological predicates. It is of course from this issue that the so-called conceptual problem of other minds takes its start that much occupied Wittgenstein and prompted Peter Strawson’s important proposal (1959, Ch. 2) that the concept of a person as subject to both psychological and physical descriptions was “primitive” in our conceptual scheme. To be sure, Carnap never strayed far into philosophical psychology and issues of what constitutes personhood, and even his recognition of the conceptual problem came slowly. It is also true that for some time Carnap appears not to have noted that the very distinction in *Scheinprobleme* of the auto- and heteropsychological languages posed the problem very starkly. But matters changed not only in 1935 when an account was provided of how first- and third-person ways of verification can be equivalent in determining the same content. More significant in this context is that as early as 1932 Carnap provided an account of how mental self-ascriptions are learned such that their use is responsive to public criteria.

A tired child says “Now I am happy to be in bed”. If we investigated how the child learnt to talk about the states of his own mind, we would discover that, under some circumstances, his mother said to him, “Now you are happy to be in bed”. ... Learning to talk consists of B’s inducing a certain habit in A, a habit of “verbalizing” (as the behaviorists put it) in a specific manner in specific circumstances. And indeed one tends so to direct this habit so that the series of words produced by speech movements of the child A coincides with the sentence of the appropriate intersubjective physical language which not only describes the appropriate state of A, but – and this is the

essential point — describes *A*'s state as *B* perceives it, that is, the physical state of the body. The example of the child shows this especially clearly. The sentence, "You are happy", spoken by the mother, is a sentence about somebody else's mind, and thus, according to our earlier analysis, can designate nothing but some physical state of affairs. The child is thus induced to develop the habit of responding to specific circumstances by uttering a sentence which expresses a physical state observed by some other person (or inferred by some other person from observed signs). (1932c/1959, 196–197)

While not answering all questions that could be asked, these remarks about how the psychological language is acquired surely go some way towards explaining how the unicity of first- and third-person ascriptions could come to be learned.¹⁵

Carnap's naturalization of mental talk did not, then, quite scale the heights that some current accounts aspire to, namely that of rendering groundless the skeptical stances engendered by Cartesianism concerning mind and knowledge, but he moved a fair bit towards it. Moreover, the advance Carnap made by his fully mature physicalism is still highly significant in its own terms. For note that it was only then that Carnap could do justice to the possibility of actual empathetic knowledge. In *Scheinprobleme* he was still reduced to statements such as this:

Empathy (*Einfühlung*) is not cognition; it does not produce any theoretical content or anything that can be stated; it is doing, not cognizing; it is a doing which establishes contact with the other and thus leads to a different practical orientation and consequently to different

15. Compare Wittgenstein (1953, #244): "Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behavior."

external actions. Ethical values come into play, but there is no connection with truth and falsity. (S §11, 337–338)

There is insight here, I think, but it was jeopardized by the failure of Carnap's account of knowledge of other minds at the time. Only the embrace of physicalism and the liberalization of the verificationism (before abandoning its meaning theory variant wholly for semantics) made it possible to justify imaginative projections of mental state descriptions onto others as a legitimate heuristic that had to be independently confirmed: there had to be grounds to believe that there was (to use his earlier example) a (possibly *ceteris paribus*) bridge-law like (3) that linked the states of affairs expressed by sentences like (1) and (2). Contrary to widespread presumption then, the very possibility of rationally reconstructing empathetic knowledge was only redeemed once Carnap managed to comprehend knowledge of other minds physicalistically.

Carnap's physicalism remained unspecific as to the physical conditions of mental states ("S's body is red-seeing") — W. V. Quine once called it "facile" but whole-heartedly approved of it nevertheless (1957/1976, 243) — yet it was clearly a step towards the naturalization of mind *talk*. This latter qualification is essential if we are to not lose sight of what continues to make Carnap's physicalism different — and difficult. Whereas Quine had no qualms about understanding physicalism as an ontological thesis spelling out identities of states of mind and body, Carnap's understanding of philosophy as logic of science abjured object-level pronouncements and proceeded exclusively at the metatheoretical level of logico-semantic or pragmatic analysis. Carnap never moved from his essentially metalinguistic understanding of physicalism. Even in later years Carnap resisted Feigl's attempts (1963) to recruit him for the mind-body identity theory. Instead of affirming an object-language identity statement — "the [psychological] process $P(a,t)$ is the same as the [neurophysiological] process $N(a,t)$ " — Carnap affirmed the metalinguistic statement that in his preferred physicalistic language "the predicates 'P' and 'N', though not L-equivalent, are

P-equivalent, i.e., equivalent on the basis of the postulates and rules." And, about the psychological language containing predicates like *P*, Carnap concluded that "the evidence available today provides good reasons for the assumption that this language will also function well in the future" (1963, 885–886). Carnap's scrupulous abstention from ontology was in line with his principle of tolerance, whose metaphilosophical deflationism many philosophers to this day have found unnerving — a matter I must leave unaddressed. My aim was simply to show that Carnap's very own kind of naturalization of knowledge of other minds was neither as simple a matter as he thought early on nor as hopeless an endeavor as it has sometimes been held to be by his critics. That it takes an advanced version of physicalism and the rejection of methodological solipsism to begin to be able to comprehend our knowledge of other minds is surely a lesson of lasting value.¹⁶

References

- Albritton, R. 1966. Appendix to "On Wittgenstein's Use of the Term 'Criterion.'" In G. Pitcher (ed.), *Wittgenstein: The Philosophical Investigations*, London: Macmillan, 247–250.
- Ambrus, G. 2020. "Carnap and Wittgenstein on Psychological Sentences." In Fissette, Fréchette, and Stadler 2020, 253–286.
- Avramides, A. 2001. *Other Minds*. London: Routledge.
- . 2009. "Other Minds." In A. Beckermann, B.P. McLaughlin, and S. Walter (eds.), *The Oxford Handbook of Philosophy of Mind*, Oxford: Oxford University Press, 727–741.
- Ayer, A.J. 1936. *Language, Truth, and Logic*. London: Gollancz. 2nd ed. 1946.
- . 1963. "Carnap's Treatment of the Problem of Other Minds." In Schilpp 1963, 269–282.
- (ed.). 1959. *Logical Positivism*. New York: Free Press.
- Carnap, R. 1928a. *Der logische Aufbau der Welt*. Berlin: Weltkreisverlag. Trans. *The Logical Structure of the World: Pseudoproblems in Philosophy*, University of California Press, 1967, v–300.
- . 1928b. *Scheinprobleme in der Philosophie: Das Fremdpsychische und der Realismusstreit*. Berlin: Weltkreisverlag. Trans. *The Logical Structure of the World: Pseudoproblems in Philosophy*, University of California Press, 1967, 301–343.
- . 1930. "Die alte und die neue Logik." *Erkenntnis* 1: 12–26. Trans. "The Old and the New Logic," in Ayer 1959, 133–146.
- . 1932a. "Überwindung der Metaphysik durch logische Analyse der Sprache." *Erkenntnis* 2: 219–241. Trans. "The Elimination of Metaphysics through Logical Analysis of Language," in Ayer 1959, 60–81.
- . 1932b. "Die physikalische Sprache als Universalsprache der Wissenschaft." *Erkenntnis* 2: 432–465. Trans. *Unity of Science*, London: Kegan Paul, Trench, Trubner, 1934.
- . 1932c. "Psychologie in physikalischer Sprache." *Erkenntnis* 3: 107–142. Trans. "Psychology in Physicalist Language," in Ayer 1959, 165–198.
- . 1932d. "Über Protokollsätze." *Erkenntnis* 3: 215–228. Trans. "On Protocol Sentences," *Noûs* 21 (1987): 457–470.
- . 1934a. *Die logische Syntax der Sprache*. Vienna: Springer. Trans. *The Logical Syntax of Language*, London: Kegan Paul, Trench, Trubner, 1937.
- . 1934b. "On the Character of Philosophic Problems." *Philosophy of Science* 1: 5–19.
- . 1935a. *Philosophy and Logical Syntax*. London: Kegan Paul, Trench, Trubner.
- . 1935b. "Les concepts psychologiques et les concepts physiques sont-ils foncièrement différents?" *Revue de Synthèse* 10: 43–53.
- . 1936. "Über die Einheitssprache der Wissenschaft. Logische Bemerkungen zum Project einer Enzyklopädie." In *Actes du Congrès Internationale de Philosophie Scientifique, Sorbonne, Paris 1935, Facs. II, "Unité de la Science," Paris: Herman & Cie, 60–70.*

16. Sincere thanks to my colleagues Sean Crawford, Fraser MacBride, Joel Smith, and Graham Stevens for discussions and critical comments and two anonymous referees for suggestions.

- . 1936–1937. “Testability and Meaning.” *Philosophy of Science* 3: 419–471 and 4: 1–40.
- . 1938. “Logical Foundation of the Unity of Science.” In O. Neurath et al., *Encyclopedia and Unified Science*, Chicago: University of Chicago Press, 42–62.
- . 1939. *Foundations of Logic and Mathematics*. Chicago: University of Chicago Press.
- . 1947. *Meaning and Necessity*. Chicago: University of Chicago Press. 2nd ed. 1956.
- . 1954. “On Belief Sentences: Reply to Alonzo Church.” In M. Macdonald (ed.), *Philosophy and Analysis*, Oxford: Basil Blackwell, 128–131. Repr. in Carnap, *Meaning and Necessity*, 2nd ed., 1956, 230–232.
- . 1956. “The Methodological Character of Theoretical Concepts.” In H. Feigl and M. Scriven (eds.), *The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, Minneapolis: University of Minnesota Press, 38–76.
- . 1961. “Vorwort zur zweiten Auflage.” In Carnap, *Der logische Aufbau der Welt*. 2. Auflage, Hamburg: Meiner. Trans. “Preface to the Second Edition,” in Carnap, *The Logical Structure of the World: Pseudoproblems in Philosophy*, University of California Press, 1967, v–xiv.
- . 1963. “Intellectual Autobiography” and “Replies and Systematic Expositions.” In Schilpp 1963, 1–84 and 859–1015.
- . 2004. “Von Gott und Seele. Scheinfragen in Metaphysik und Theologie.” In Carnap, *Scheinfragen in der Philosophie. Metaphysikkritische Schriften* (ed. T. Mormann), Hamburg: Meiner, 49–61.
- Carnap, R., Hahn, H., and Neurath, O. 1929. *Wissenschaftliche Weltauffassung. Der Wiener Kreis*. Vienna: Wolf. Trans. “The Scientific World Conception: The Vienna Circle,” in F. Stadler and T. Uebel (eds.), *Wissenschaftliche Weltauffassung. Der Wiener Kreis. Hrsg. vom Verein Ernst Mach (1929)*, Vienna: Springer, 2012, 75–116.
- Chalmers, D. J. (ed.). 2002. *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.
- Cirera, R. 1993. “Carnap’s Philosophy of Mind.” *Studies in History and Philosophy of Science* 24: 351–358.
- Coffa, A. 1977. “Carnap’s Sprachanschauung circa 1932.” In *PSA 1976*, vol. 2, East Lansing: Philosophy of Science Association, 205–241.
- . 1987. “Carnap, Tarski and the Search for Truth.” *Noûs* 21: 547–572.
- Crawford, S. 2013. “The Myth of Logical Behaviourism and the Origins of the Identity Theory.” In M. Beaney (ed.), *The Oxford Handbook of the History of Analytic Philosophy*, Oxford: Oxford University Press, 621–655.
- . 2015. “On the Logical Positivists’ Philosophy of Psychology: Laying a Legend to Rest.” In M. C. Galavotti et al. (eds.), *New Directions in the Philosophy of Science*, Dordrecht: Springer, 711–726.
- Feigl, H. 1934. “Logical Analysis of the Psychophysical Problem.” *Philosophy of Science* 1: 420–445.
- . 1950. “The Mind-Body Problem in the Development of Logical Empiricism.” Repr. in Feigl, *Inquiries and Provocations* (ed. R. S. Cohen), Dordrecht: Reidel, 1981, 286–301.
- . 1963. “Physicalism, Unity of Science and the Foundations of Psychology.” In Schilpp 1963, 227–268.
- Fisette, D., Fréchette, G., and Stadler, F. (eds.). 2020. *Franz Brentano and Austrian Philosophy*. Cham: Springer.
- Fodor, J. A. 1968. *Psychological Explanation: An Introduction to the Philosophy of Psychology*. New York: Random House.
- Foster, J. 1985. *Ayer*. London: Routledge.
- Haller, R., and Rutte, H. 1977. “Gespräch mit Heinrich Neider.” *Conceptus* 28–30: 21–42.
- Hempel, C. G. 1935. “Analyse logique de la psychologie.” *Revue de Synthèse* 10: 23–43. Trans. “The Logical Analysis of Psychology,” in Hempel, *Selected Philosophical Essays* (ed. R. Jeffrey), Cambridge: Cambridge University Press, 2000, 165–180.
- Kim, J. 2003. “Logical Positivism and the Mind-Body Problem.” In P. Parrini, W. C. Salmon, and M. H. Salmon (eds.), *Logical Empiricism: Historical and Contemporary Perspectives*, Pittsburgh: University of Pittsburgh Press, 263–278.
- . 2011. *Philosophy of Mind*. 3rd ed. Boulder: Westview Press.

- Kitchener, R.F. 1999. "Logical Behaviorism." In W. O'Donohue and R. Kitchener (eds.), *Handbook of Behaviorism*, New York: Academic Press, 399–418.
- . 2004. "Logical Positivism, Naturalistic Epistemology, and the Foundations of Psychology." *Behavior and Philosophy* 32: 37–54.
- Lycan, W.G. (ed.). 1990. *Mind and Cognition: A Reader*. Oxford: Basil Blackwell.
- McGuinness, B. (ed.). 1979. *Ludwig Wittgenstein and the Vienna Circle: Conversations Recorded by Friedrich Waismann*. Oxford: Basil Blackwell.
- Patzig, G. 1966. "Nachwort." In R. Carnap, *Scheinprobleme in der Philosophie*, Frankfurt a. M.: Suhrkamp, 83–136.
- Putnam, H. 1965. "Brains and Behavior." In R. Butler (ed.), *Analytical Philosophy: Second Series*, Oxford: Basil Blackwell, 1–18. Repr. in Putnam, *Mind, Language and Reality*, Cambridge: Cambridge University Press, 1975, 325–341.
- Quine, W.V. 1957. "The Scope and Language of Science." *British Journal for the Philosophy of Science* 8: 1–17. Repr. in Quine, *The Ways of Paradox*, 2nd ed., Cambridge, Mass.: Harvard University Press, 1976, 228–245.
- Russell, B. 1921. *The Analysis of Mind*. London: Allen & Unwin.
- Ryckman, T. 2007. "Carnap and Husserl." In M. Friedman and R. Creath (eds.), *The Cambridge Companion to Carnap*, Cambridge: Cambridge University Press, 81–105.
- Schilpp, P.A. (ed.). 1963. *The Philosophy of Rudolf Carnap*. LaSalle, Ill.: Open Court.
- Smith, J. 2015. "The Phenomenology of Face-to-Face Mindreading." *Philosophy and Phenomenological Research* 90: 274–293.
- Stout, R. 2010. "Seeing the Anger in Someone's Face." *Proceedings of the Aristotelian Society, Supplementary Volumes* 84: 29–43.
- Strawson, P.F. 1959. *Individuals*. London: Methuen.
- Tamminga, A. 2005. "Introspection and Change in Carnap's Logical Behaviourism." *Studies in History and Philosophy of Science* 36: 650–667.
- Uebel, T. 2007. *Empiricism at the Crossroads: The Vienna Circle's Protocol-Sentence Debate*. Chicago: Open Court.
- . 2010. "Opposition to 'Verstehen' in Orthodox Logical Empiricism." In U. Feest (ed.), *Historical Perspectives on Erklären und Verstehen*, Dordrecht: Springer, 291–309.
- . 2014. "Carnap's *Aufbau* and Physicalism." In M.C. Galavotti, E. Nemeth, and F. Stadler (eds.), *European Philosophy of Science*, Dordrecht: Springer, 45–56.
- . 2019a. "Neurath on Verstehen." *European Journal of Philosophy* 27: 912–938.
- . 2019b. "Verificationism and (Some of) Its Discontents." *Journal for the History of Analytical Philosophy* 7 (4): 1–31.
- . 2020. "Intentionality in the Vienna Circle." In Fissette, Fréchette, and Stadler 2020, 135–168.
- . 2021. "A Logical Positivist's Progress: A Puzzle about Other Minds in Early Ayer Resolved." In A.T. Tuboly (ed.), *The Historical and Philosophical Significance of Ayer's Language, Truth and Logic*, Basingstoke: Palgrave Macmillan, 191–247.
- Weinzweig, M. 1962. "Our Knowledge of Other Minds: A Pseudo-Problem?" *Philosophy and Phenomenological Research* 23: 250–255.
- Wittgenstein, L. 1953. *Philosophische Untersuchungen / Philosophical Investigations*. Oxford: Basil Blackwell.
- Zilsel, E. 1932. "Bemerkungen zur Wissenschaftslogik." *Erkenntnis* 3: 143–161.