New Life for Carnap's *Aufbau*?

Hannes Leitgeb

March 2008

Abstract

Rudolf Carnap's *Der logische Aufbau der Welt* (*The Logical Structure of the World*) is generally conceived of as being the failed manifesto of logical positivism. In this paper we will consider the following question: How much of the *Aufbau* can actually be saved? We will argue that there is an adaptation of the old system which satisfies many of the demands of the original programme. In order to defend this thesis, we have to show how a new "*Aufbau*-like" programme may solve or circumvent the problems that affected the original *Aufbau* project. In particular, we are going to focus on how a new system may address the well-known difficulties in Carnap's *Aufbau* concerning abstraction, dimensionality, and theoretical terms.

1 Introduction

Rudolf Carnap's (1928) classic *Der logische Aufbau der Welt* (*The Logical Structure of the World*) was abandoned at least twice: at first when the Vienna circle turned from logical positivism to logical empiricism or from epistemology to philosophy of science; secondly when philosophy of science moved from its understanding of being a "logic of science" towards its new emphasis on naturalistic-pragmatic-historical-sociological features of science.

More recently, the *Aufbau* has attracted attention from philosophers who question its traditional interpretation as being the modernized upshot of British empiricism. While these reinterpretations of the *Aufbau* have initiated a renewal of interest in its content, its assessment as a famous and perhaps even notorious failure has remained unchanged.

In this paper, we will deal with the *Aufbau* not from a historical but from a systematic point of view.[1] We are going to argue that the old *Aufbau* has a core part which might actually be *saved*: although the original programme cannot be restored itself, there is hope for a "new *Aufbau*" which shares several important properties with its predecessor.

This is the plan of the paper: In section 2 we start with an exposition of the old *Aufbau's* aims, and we constrast them with the weakened intentions that lie behind the development of the new system. Then we turn to the problems of the original *Aufbau*. If any attempt of introducing a new *Aufbau* is to be successful, it has to demonstrate how the problems that affected Carnap's *Aufbau* are either circumvented or solved. We will concentrate our efforts on two representative problem sets: Goodman's problems of abstraction and dimensionality (section 4) and Quine's problem of holism and theoretical terms (section 5). Both problems can only be explained satisfyingly if what is called the "basis" in the *Aufbau* is outlined beforehand: this will be done in section 3. In sections 6, 7 and 8 we will finally introduce the new system and see how it addresses Goodman's and Quine's worries. Section 9 will end up with a summary of what has been achieved and with an outlook of future work on a new *Aufbau*.

2 A New Epistemological Project

According to the traditional interpretation – exemplified by Quine (1951), Goodman (1963), and, retrospectively, by Carnap himself (1963) – the aim of the *Aufbau* is to support the following thesis:

- Old thesis: Every scientific sentence can be translated via explicit definitions into another sentence that consists solely of logical signs and terms that refer to "the given", such that in each of the underlying definitions the defined expression and the defining expression necessarily have the same extension.

The new interpretation by Friedman (1999), Richardson (1998) and a few others ascribes an even stronger claim to the *Aufbau*:

- Old thesis: Every scientific sentence can be translated via explicit definitions into another sentence that is purely structural, i.e., which consists solely of logical signs, such that in each of the underlying definitions the defined expression and the defining expression necessarily have the same extension.

Whilst the first interpretation considers the *Aufbau* as the result of applying the then new logical means of Whitehead & Russell's *Principia Mathematica* to the traditional empiricist-phenomenalist programme, the second one understands the *Aufbau* as being influenced by the Neo-Kantian tradition and emphasizes its neutrality with respect to all traditional epistemological positions. While the intention of the *Aufbau*, according to its traditional interpretation, is to show how scientific claims may ultimately be reduced to claims about the contents of our immediate subjective experience, the more recent interpretation has it that science is ultimately about the structure of experience, where 'structure' is supposed to denote something that is intersubjective rather than subjective.

Let us consider the two theses from above in more detail now:

In the first thesis, 'given' denotes what is given by experience, in particular, by sense experience. Indeed, for the rest of this paper, sense experience will be the only form of "data" that we are interested in.

'scientific sentence' refers to any sentence in a language of any scientific discipline that uses its terms in a clear and non-ambiguous way.

A translation is to be regarded a mapping from "the" set of scientific sentences to itself. The two theses claim that that there are translation mappings of a particular and distinguished kind: (a) they are induced by a system of definitions in the way that a scientific sentence A is translated to another scientific sentence $tr(A)$ if and only if the stepwise replacement of the defined terms in A by their defining (and ultimately) primitive terms yields $tr(A)$; (b) the corresponding primitive vocabulary conforms to the syntactic restrictions that are explained in the theses – logical terms and terms that refer to the given in the first case, only logical terms in the second one; (c) finally, the transition from a defined expression to its defining one is to preserve extension necessarily.[2]

As Carnap explains in §50 of the *Aufbau*, the translations of sentences and terms are claimed to preserve what Carnap then called "logical value", i.e., *extension*. In the preface of the second edition of the *Aufbau*, Carnap clarifies his view by pointing out that what he actually demands is the *necessary* preservation of extension, i.e.: the translation image of an expression ought to have the same extension as the translated one *by logical rules or by laws of nature*. In particular, if a sentence A is translated to a sentence $tr(A)$ by substituting a defined expression by its defining expression, then the defined expression should necessarily have the same extension as the defining one and consequently A is to be necessarily materially equivalent to $tr(A)$. As far as the translation of sentences is concerned – and this is what Carnap aims at ultimately – the goal is thus more than just the preservation of truth values; rather it is the preservation of truth conditions. Indeed, demanding only the preservation of truth values for sentences would seem to be too weak, because any translation function which maps all true sentences to, say, $\forall x\, x = x$, and all false sentences to $\neg\forall x\, x = x$ would meet this criterion. However, even in order to set up a translation like this, one would have to know which scientific sentences are true and which are false, which is certainly beyond human capabilities, and which certainly is not presupposed by the *Aufbau*. Indeed, according to the *Aufbau* programme, the translation mappings whose existence is claimed by the two theses above should be definable *a priori* – before any empirical investigation into the truth or falsity of scientific hypothese even commences. Carnap is well aware of the fact that definitions are normally demanded to preserve sense or (in *Aufbau* terminology) "Erkenntniswert" rather than truth conditions, but he argues that the preservation of truth conditions is in fact all that is needed for scientific purposes as opposed to, e.g., aesthetic purposes. If tr is a translation that is based on definitions and which preserves truth conditions, and if A is translated to $tr(A)$, then Carnap holds that A can be replaced by $tr(A)$ in all scientific contexts without any scientifically significant loss.

Now we turn to what could be the aims of a *new Aufbau*. When we say that the old *Aufbau* has a core part that may actually be saved, this really amounts to the claim that a thesis sufficiently close to the two theses above is true. When we say that the original programme itself cannot be restored, this means that the new thesis has to be weaker than the two theses from above. Here is the corresponding thesis that guides our new attempt at an *Aufbau*-like system:

- New thesis:
  - Every scientific sentence can be translated to an *empirically* equivalent one which consists solely of logico-*mathematical* signs and terms that refer to *experience*, such that
  - the translation image expresses a *subject-invariant* constraint on experience.

We have highlighted the differences between the new thesis and the old ones in italics: First of all, if A is translated to $tr(A)$, then the two sentences are no longer demanded to be materially equivalent, let alone necessarily materially equivalent; instead, A and $tr(A)$ should be empirically equivalent. More particularly, we want $tr(A)$ to express the empirical content of A, i.e., to use a phrase of Quine: $tr(A)$ is to describe the difference the truth of A would make to possible experience (cf. Quine 1969). For reasons of space, we will not be able to offer an independent analysis of the notion of empirical content in this paper, but we will rather take 'empirical content' to be a primitive term here while simply presupposing that it is sufficiently understood; hopefull, to some extent, what we have in mind should become clear from the investigations below.[3] Given the broad agreement among philosophers of science that the truth of scientific theories may be underdetermined empirically, sentences A and $tr(A)$ may thus differ in truth value according the new *Aufbau* even though their empirical contents are required to be the same.[4] Accordingly, A may not be replaced by $tr(A)$ for all scientific purposes whatsoever. Note that our new thesis does not presuppose any form of verificationism according to which the meaning of a sentence is identified with its empirical content. Moreover, since the translated sentences will normally have truth conditions which differ from those of their translation images, the translations in question should not be regarded as subserving any sort of "ontological reduction" of physical objects to sense experience, or the like.[5]

At second, the translation mappings we claim to exist are no longer supposed to be definable by a system of explicit definitions alone. As we are going to point out later, our translation will be partially based on contextual definitions, which is anticipated by Carnap in the *Aufbau* when he accepts "definitions in use" as legitimate means of reduction by definition (see also Quine 1969). Moreover, each of the explicit or contextual definitions that we will propose is only meant to hold up to empirical equivalence; as mentioned before, we do not demand coextensionality, necessary coextensionality, or synonymy between defining and defined expressions. However, our translation manual will still turn out to be *a priori* in the sense that it is possible in principle to set it up before empirical investigation.

A further difference between the new thesis and its precursors consists in our reference to mathematical signs as being additional to logical ones. At the time of the *Aufbau*, Carnap still subscribed to logicism along the lines of Frege and Russell. But logicism – at least in its traditional form – does not work, and the denial of the existence of genuinely mathematical concepts and sentences is no longer part of our "enlightened" programme. In particular, we do not insist that the set-theoretic membership sign that will be used later is a logical symbol.

As far as the empirical aspects of our translation mappings are concerned, we have replaced the term 'the given' by 'experience': this indicates that our new *Aufbau* system does not rely on any phenomenalistic conception of what the basis of our subjective experience consists in. In fact, the new system will be open both to a phenomenalistic *and* a physicalistic interpretation. Experiences may be the contents of particular mental states or they may be particular mental states themselves; mental contents and mental states may turn out be identical to occurrences in the brain or to brain states.

Finally, the goal of having the translations of scientific sentences express subject-invariant constraints on experience is our substitute for the "structural" intentions of the original *Aufbau*, as highlighted by the second more recent interpretation of the two interpretations considered above. In the following, however, we we will not deal with this part of our new thesis, but we will concentrate just on the rest of it.

Since every translation that preserves truth conditions may be assumed to preserve empirical content as well, our new thesis is weaker than the two "old" theses that we have discussed. But the new thesis is still reasonably close to the old ones. When Carnap uses the term 'necessary' in the preface of the second edition of the *Aufbau* in order to express the goal of the necessary preservation of extension, he circumscribes this in the following way: the extension of a defined expression within the phenomenalistic language that is associated with a subject S should be identical to the extension of its defining expression, independent of what the experience of S is like, as long as S's senses function "normally" and as long as "unfavourable circumstances" are excluded. As far as sentences are concerned, this is actually very close to saying that $tr(A)$ is to describe the difference the truth of A would make to *possible experience* of S.

Before we turn to the problems notoriously affecting the old system and to the details of a new *Aufbau*-like system, we want to point out very briefly why the development of a new *Aufbau* is still a worthwhile epistemological endeavour. In other words: Why should we care about a new "weakened" *Aufbau* programme at all?

- It may cast new light on where and why the old *Aufbau* *really* failed: We claim that each of the problems that have been ascribed to the original *Aufbau* fall into one of three categories: (a) they do not even apply to the original *Aufbau* (although they might apply to other aspects of the Vienna circle philosophy) – these are the "pseudo-problems"; (b) they did affect the *Aufbau* but they may be solved in a new system by adapting the original construction in ways that are still acceptable from the point of view of the old programme – these are the "feasible problems"; (c) they did affect the *Aufbau* but they may be circumvented in the new system by lowering the intentions of the latter – these are the "serious problems". The construction of a new *Aufbau* will give us some information on which problems of the old *Aufbau* belong to the third, and philosophically most relevant, category. In particular: as we will see, what we call Goodman's problems below ought to count as feasible, while what we call Quine's problem is serious.

- It may deepen our understanding of the empirical content of terms and descriptive sentences: Although the meaning of an expression is not identical with its empirical content, the latter is certainly one relevant component of its meaning. Indeed, if experience is understood in terms of a subjective basis that is relativized to a particular cognitive agent, then the so-determined empirical meanings may be considered to be among the internalist meaning components of linguistic expressions – the meaning components that are "in" this agent's mind – which are additional to externalist (referential) ones.

- It may fill the gap between subjective experience and the intersubjective basis of scientific theories: After the protocol sentence debate in the early 1930s, philosophers of science more or less decided to conceive of the observational basis of science as being intersubjective right from the start; observation terms and observation sentences were meant to refer to observable real-world objects and to their observable space-time properties. While this move is perfectly acceptable from the viewpoint of philosophy of science, it leaves an interesting epistemological topic out of consideration: the relation of this intersubjective "observational" basis to the subjective act of observation and its experiential content. The new *Aufbau* addresses this latter topic by relocating empirical contents into the observer. In this way, given an analysis of 'experience' in terms of q subjective basis for a cognitive agent, it is possible to study what difference the truth or falsity of a statement about common sense observable objects and properties makes to an agent's private experience.

- It may also relate questions in cognitive science and the philosophy of cognition to questions in epistemology and philosophy of science; as Glymour (1992), p. 367, puts it, "Carnap wrote the first artificial intelligence program" when he introduced his phenomenalistic construction system in the *Aufbau*. E.g., an answer to the question of whether the empirical contents of scientific terms and sentences are generally computable might be an interesting spin-off. Or: How parsimonious can the expressive resources of a language be such that the empirical or "experiential" contents of sentences, as being given by a subjective basis, may still be expressed in it with sufficient accuracy?

- Finally, a new *Aufbau* may refine our understanding and assessment of structuralist claims: since the days of the *Aufbau*, structural realism (cf. Worrall 1989) has evolved into a serious competitor for an adequate description of scientific progress and its limits. As Demopoulos & Friedman (1985) have shown, some of the problems that are claimed to affect present-day structural realism are among the difficulties that Carnap faced as well when he dealt with the reducibility of scientific expressions to "structural descriptions" in the *Aufbau* (see §11–16, 153–155).

It should have become clear by now that this is not a metaphysical project but rather a semantic and epistemological one, with possible applications to the philosophy of science, the philosophy of language, and the philosophy of cognition. Whether it can be carried out successfully, depends on how it comes to terms with the well-known problems that affected the "old" Aufbau. In the next section we are going to concentrate on two of these problems which we refer to as 'Goodman's problems'. In order to explain the gist of Goodman's problems, we have to start with an outline of what is called the "basis" in the *Aufbau*.

3 The Basis of the "Old" *Aufbau*

Carnap's *Aufbau* may be viewed as consisting of two parts: (a) the phenomenalistic constitution or construction system that is described in §106–155, which is nothing but an extensive list of definitions, and (b) a philosophical metatheory that analyzes, justifies, and applies this constitution system and compares it to alternative ones. As every finite system of definitions, a constitution system presupposes a choice of primitive, i.e., undefined terms; (i) the set of interpretations of these terms together with (ii) the members of the intended universe of discourse of the system are referred to as "the basis" of the constitution system in the *Aufbau*. While (ii) constitute the "basic elements" of the system, (i) gets referred to as its "basic properties and relations"; we call the corresponding predicates that express the basic properties and relations "basic predicates". In the case of the phenomenalistic constitution system of the *Aufbau*, this basis is, of course, *phenomenalistic*: it consists of

- (Old) Basic elements: elementary experiences (erlebs) of a given and fixed subject S within a given interval of time;

- (Old) Basic relations: the membership relation $\in$ and the relation Er of "recollected similarity".

The intended universe and the intended interpretation of the basic terms of the phenomenalistic constitution system in the *Aufbau* can be explained extra-systematically:

An elementary experience or erleb (this is Goodman's term in *The Structure of Appearance*, 1951) of a subject S within an interval of time is a total momentary slice through S's stream of experience, i.e., the sum of all visual, auditory, tactile,... experiences that S has at a subjectively experienced moment of time, where the moment is assumed to be included in the given interval of time.

The membership relation is just the standard mathematical relation that holds between the members of a set and the set itself. The underlying set theory of the *Aufbau* was actually a version of simple type theory in which '$\in$' was not really primitive but rather contextually eliminable in favour of higher-order quantification. However, for our purposes it is more convenient to consider the set theoretical system of the *Aufbau* as a version of modern set theory with a given universe of urelements. The urelements are just the basic elements as described above, i.e., elementary experiences.[6]

'Er' is a binary predicate that expresses a relation between erlebs: it is the case that $x\,Er\,y$ if and only if x is recollected by S as being part-similar to y. E.g., if S experiences in x a particular light-red spot in the left-upper part of her visual field and if a little later S has an elementary experience y in which she experiences a dark-red spot in the left-middle part of her visual field, then x and y have "parts" that are similar to each other; this is what S is aware of, if x stands in the Er-relation to y. We can express this more formally by presupposing – as Carnap does – that every elementary experience can be described by reference to pairwise disjoint quality spaces which come equipped with distance functions (metrics). Indeed, these quality spaces may be regarded as mathematical entities which get realized or instantiated when S has experiences of some sort, and Carnap's basis can be explained by exploiting this correspondence to a mathematical structure. E.g., instead of saying that S experiences in x a particular light-red spot in the left-upper part of her visual field, we may say equivalently that the erleb x realizes a particular point in the "light-red and left-upper" region of S's visual quality space, if only this "realization" or "instantiation" relation is explained in a way such that the equivalence between the two statements holds by definition.[7] The part-similarity of x and y then corresponds to the fact that there are quality points p, q in a single sensory quality space (visual, auditory, tactile,...) – in this example the visual one – such that (i) p and q are metrically "close" to each other, i.e., they have a distance that is less than or equal to some given and fixed real number ϵ, and (ii) x realizes p while y realizes q. In the case of the visual quality space, the closeness of p and q amounts to the fact that p and q represent colours-at-places where the colours resemble each other, and where the places resemble each other, too. If the part-similarity of two erlebs x and y is recollected by S in the sense that S compares a memory image of the past erleb x with her current erleb y, then this is precisely what gets expressed by '$x\,Er\,y$'. Er thus has a qualitative *and* a temporal component. In particular, if $x\,Er\,y$, then the erleb x occurred before y.[8]

4 Problem Set 1: Goodman's Problems

Carnap's main goal in the first part of his constitution system – the so-called "auto-psychological domain" (§106–122) – is to show that the meager basis of this system suffices for the definition of various kinds of terms by which one may express and analyze S's experiences qualitatively. In particular, Carnap wants to define a general term 'phenomenal quality point'[9] the extension of which should be the set of phenomenal counterparts of visual, auditory, tactile,... quality points as described above. While the quality points are just points in some mathematical spaces that come associated with sense modalities, the phenomenal counterparts to these quality points – call them *phenomenal quality points* – are set-theoretic constructs on erlebs: a quality point p is meant to induce a phenomenal quality point in the sense that the latter is the set of all erlebs in which p is realized. The set of phenomenal quality points is therefore the class of all sets of erlebs which are induced in this way. However, while this is the intended interpretation of the predicate 'phenomenal quality point', Carnap has to show that its extension may be defined, whether directly or indirectly, solely in terms of the basic relations $\in$ and Er. The way in which he tries to accomplish this is, roughly, (i) by defining a similarity relation Sim of erlebs as the reflexive symmetric closure of Er, (ii) by abstracting from Sim the phenomenal counterparts of spheres in quality spaces (call them *phenomenal spheres*), and finally (iii) by defining the members of the extension of 'phenomenal quality point' in terms of these phenomenal spheres. Step (i) is intended to have the result that $x\,Sim\,y$ if and only if x and y realize quality points in a common closed quality sphere of diameter ϵ, i.e., x and y realize quality points that have a distance less than or equal to ϵ.[10] The steps (ii) and (iii) constitute Carnap's method of quasianalysis, a method of abstraction that generalizes Frege's and Russell's method of abstracting equivalence classes from equivalence relations. Phenomenal spheres are thus supposed to have a mediating role between erlebs and phenomenal quality points. Analogously to the case of phenomenal quality points, a quality sphere Q is meant to induce a phenomenal quality sphere in the sense that the latter should be the set of all erlebs which realize some quality point in Q. The set of phenomenal quality spheres is then the class of all sets of erlebs which are induced in this way. The first part of quasianalysis is intended to define the extension of 'phenomenal quality sphere' to be this class, where the definition is to be spelled out solely in terms of '$\in$' and 'Er'.

After having defined 'phenomenal quality point', Carnap's strategy is to introduce a definition of a new similarity relation which is defined on the basis of the similarity relation for erlebs but which applies to the newly defined phenomenal quality points. He is especially interested in the connectivity components of this new similarity relation: these components are expected to be exactly the phenomenal counterparts of quality spaces, because if S has experiences which are sufficiently varied it is likely that phenomenal quality points which correspond to visual quality points are never qualitative "neighbours" of, say, phenomenal quality points that correspond to auditory quality points. Carnap then shows how "dimension numbers" may be assigned to the connectivity components, which seems to be possible because he assumes that every subjective quality space has a well-defined dimensionality. In particular, the visual quality space is supposed to be the only five-dimensional quality space: a five-dimensional subset of the Euclidean space $\mathbb{R}^5$, where the first two coordinates correspond to the x- and the y-coordinates of places in the two-dimensional visual field, and where the other three coordinates represent the hue, brightness, and saturation of the colour spots that sit at these places. Every colour-at-a-place thus corresponds to a unique quality point in a five-dimensional space that is usually depicted as a cone-like mathematical object (the "colour cone"). Accordingly for all other sense classes – e.g., the auditory quality space may be assumed to be a two-dimensional subset of $\mathbb{R}^2$, and so forth. In this way, Carnap would be able to identify the visual sense modality by its dimension, such that on this basis he could define the phenomenal counterpart of the visual quality space as well as the counterparts of all the other quality spaces that are associated with the remaining sense classes.

Unfortunately, this strategy of defining phenomenal quality points and distinguishing phenomenal quality spaces is affected by two serious shortcomings. As Goodman (1951, 1963, 1971) has shown,

- Carnap's method of abstracting phenomenal quality spheres and phenomenal quality points from a relation of similarity for erlebs is deficient;

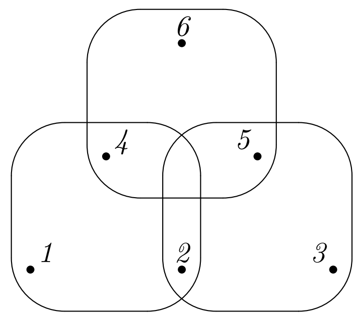
- Carnap's method of determining the visual phenomenal quality space by dimensional analysis fails if the set of erlebs is of finite cardinality.

We will now deal with these two problems in more detail. We focus first on quasianalysis: By definition, Sim is a reflexive and symmetric relation on the given set of elementary experiences. If X is a set of erlebs, let X be called a *clique* with respect to Sim if and only if for all $x, y \in X$: $x \, Sim \, y$. Here is the main idea of the first step of quasianalysis: consider some set X of erlebs which realize a quality point within a fixed quality sphere Q of diameter ϵ, i.e., of radius $\frac{\epsilon}{2}$.[11] E.g., Q might be the set of visual quality points that have distance $\frac{\epsilon}{2}$ or less from the quality point that represents a particular tone of red located at a particular spot in the visual field. X will certainly be a clique with respect to similarity, since every two members of X are part-similar; this is because every two members of X realize points of Q and thus points which are metrically close, i.e., which have a distance that is less than or equal to ϵ from each other. Let X' now be a superset of X, such that every erleb in X' still realizes some quality point in Q: then X' is again a clique with respect to Sim and thus X' is a clique that is larger than X. X' seems to be a better approximation of the phenomenal counterpart of Q than X was. Accordingly, Carnap suggests to define the phenomenal counterparts of quality spheres to be *maximal* cliques with respect to Sim, where X is a *maximal clique* with respect to Sim if and only if X is a clique with respect to Sim and there is no set Y of erlebs, such that $X \subsetneqq Y$, and Y is also a clique with respect to Sim. However, this method does not work in each and every case: sometimes the intended phenomenal quality spheres are not introduced by quasianalysis, since they cannot be separated with respect to the similarities that they induce – Goodman calls this the "companionship difficulty" – or they are introduced unjustifiedly because several erlebs are pairwise similar without there being a single quality sphere in which all of them realize a point – this is referred to by Goodman as the "difficulty of imperfect community".

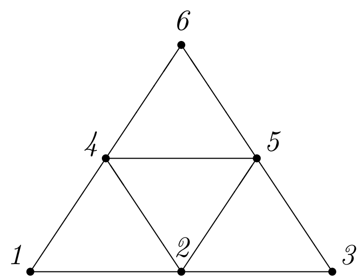
Here in an example of imperfect community (many more examples can be found in Leitgeb 2007, together with a detailed analysis of the problems and merits of Carnap's quasianalysis):

Example 1 *(Imperfect Community)*
For a given set of six erlebs $1, \ldots, 6$*, let us assume:* $1, 2, 4$ *realize a quality*

point in a sphere Q_1 (and no other erleb does), $2, 3, 5$ *realize a quality point in a sphere Q_2 (and no other erleb does), and* $4, 5, 6$ *realize a quality point in a sphere Q_3 (while no other erleb does), and we suppose again that these are all spheres in which points are realized. So the phenomenal counterparts of quality spheres are:*



The graph that depicts the similarity relation which corresponds to this distribution of realized quality spheres is:



If the first step of quasianalysis is applied, a "new" triangle $\{2, 4, 5\}$ *is defined to be a member of the extension of 'phenomenal quality sphere' because* $\{2, 4, 5\}$ *is a maximal clique with respect to similarity. However,* $\{2, 4, 5\}$ *is not the phenomenal counterpart of any of the actual quality spheres.* $2, 4, 5$ *are indeed pairwise similar, but in each case for a different "reason". As Goodman expresses this type of problem, they form an "imperfect community".*

As we have just seen, the first step of quasianalysis in the *Aufbau* may fail. But let us assume for the moment that the set of maximal cliques with respect to *Sim* would indeed coincide with all and only the phenomenal counterparts of quality spheres: how could the set of phenomenal counterparts of quality *points* be defined in terms of the latter? As a first approximation, Carnap discusses the possibility of defining phenomenal quality points as maximal non-empty intersections of phenomenal spheres, just as quality points correspond bijectively to maximal non-empty intersections of quality spheres. However, this method of defining phenomenal points on the basis of phenomenal spheres will not do, because there may be maximal non-empty intersections of phenomenal spheres which do not coincide with *any* phenomenal point: Carnap refers to this as the problem of "accidental intersection" (§80–81 in the *Aufbau*). The difficulty is that an erleb may realize points in many different quality spheres at the same time; therefore, the phenomenal counterparts of two quality spheres might either intersect because the two quality spheres themselves have a non-empty intersection in the quality space and this gets reflected by their phenomenal counterparts – the unproblematic case – or a single erleb realizes points in two quality spheres although the two spheres do not intersect – this is the case where the corresponding phenomenal quality spheres intersect "accidentally". In order to overcome this difficulty, Carnap includes a quantitative condition which essentially says (simplifying just a bit): look for maximal intersections of phenomenal spheres by taking intersections in a step-by-step manner, but do only take an intersection step if the set-theoretic overlapping of a phenomenal sphere with the previously generated intersection is not "too small" compared with the number of elements of the previous intersection. This constitutes the second step of quasianalysis. As Goodman and others have shown, even this more elaborate method does not avoid accidental intersections and hence does not always give the intended results.

Carnap himself was aware of these problems. The reason that he was not worried about them is that he regarded the situations in which these problems do occur as exceptional (Moulines 1991 argues in a similar manner). As we show in Leitgeb (2007), the problems are in fact *serious*: it is extremely likely that a cognitive agent such as our given subject S has experiences of a kind that lead to extensions of 'phenomenal quality sphere' and 'phenomenal quality point' which differ significantly from the actually intended sets of phenomenal quality spheres and phenomenal quality points. Moreover, in the extreme case of "varied experience" in which the formal structure of Carnap's basis actually coincides with the formal structure of the mathematical entities that it corresponds to, the problems do in fact *not* vanish. So Goodman was right after all, even though it needs more elaborate formal investigations into the problem in order to see that this is actually so.

So we can turn to the second of Goodman's problems – the dimensionality problem. When Carnap defines the dimension of his phenomenal quality spaces, i.e., of the connectivity components of the similarity relation for phenomenal quality points, he relies on Menger's classic topological definition of dimension for topological spaces or on a variant of it (§115–119).[12] The similarity relation functions as a "neighbourhood" relation on the phenomenal quality points, which is all that is needed in order to define a topology on its connectivity components. What Carnap overlooked when doing so, but what Goodman did observe, was that every *finite* topological space is in fact *zero*-dimensional (where we call a topological space 'finite' if and only if its underlying point set is finite). But Carnap assumes explicitly that the given set of erlebs is finite, as he points out in §180 of the *Aufbau*. Hence also the set of phenomenal quality points, which are nothing but sets of erlebs, is finite. Therefore, every phenomenal quality space, including the visual phenomenal quality space, is actually zero-dimensional, and Carnap's plan of identifying the visual sense class by its dimension fails.

One way of avoiding this problem would be to give up the presumption that the set of erlebs is finite. However, the resulting constitution system would be dubious from a phenomenalistic point of view: in a phenomenalistic system, the subject should in principle have cognitive access to the basic elements of the system; if there are infinitely many basic elements, this does not seem to be possible, at least if *simultaneous* access to the basic elements is needed. The situation would change if a system were set up which were meant to have a *physicalistic* (but still epistemic) interpretation instead: just as a mechanical system may have infinitely many possible states, the set of possible contents or states of experience for a subject, or neural system, S might be infinite. If such a set were chosen to be the set of basic elements of a physicalistic constitution system, Carnap's original strategy might be put to work.

In our new *Aufbau*, we will follow a different line of reasoning. As mentioned before, our system will be open to a phenomenalistic *and* to a physicalistic interpretation. Accordingly, we are going to leave open what the cardinality of the set of basic elements is like. Since we will nevertheless take up Carnap's idea of characterizing phenomenal quality spaces in terms of their dimension, we will have to show how dimension numbers may be assigned to them, independently of whether there are finitely or infinitely many basic elements. We will suggest a solution to this problem, as well as a solution to Goodman's first problem, in section 7.

In the next section we are going to turn to another notorious difficulty that has been ascribed to Carnap's *Aufbau*: the problem of holism and the non-definability of theoretical terms.

5 Problem Set 2: Quine's Problem

After having introduced phenomenal quality points, the similarity relation for them, and the different phenomenal quality spaces, several other definitions in the *Aufbau* system just fall into place: e.g., Carnap is able to define the set of phenomenal colour qualities, which is a set of sets of visual quality classes; the set of places in the visual field; a neighbourhood relation for these places; the set of visual sensations, where the latter are ordered pairs $\langle x, X \rangle$ of an erleb x and a visual phenomenal quality point X, such that X occurs within x, i.e., $x \in X$. Moreover, the transitive closure of the given basic relation Er can be used as a "preliminary time order" for erlebs. Indirectly, Carnap is thus able to define phrases such as 'x is the place of the visual sensation y', 'x is the phenomenal colour quality of the visual sensation y', 'visual sensation x occurs before visual sensation y', and so forth. All of these definitions deal solely with the auto-psychological domain.

Carnap's first attempt to link experiences to *physical* properties – or rather to the phenomenal counterparts thereof – was his "definition" of the function *col* which is to assign phenomenal colour qualities to points of four-dimensional space-time. The idea was to project the phenomenal colour qualities that occur in visual sensations "outwards", i.e., to map phenomenal colour qualities – along lines of sight that originate in places of the visual field – to points in $\mathbb{R}^4$. This should be done in a way, such that (i) the temporal and neighbourhood relations between visual sensations are respected, (ii) the phenomenal colour qualities "travel" on segments of continuous world-lines through space-time, and (iii) certain maxims of intertness are satisfied: the colours on world-lines should change as slowly as possible, the curvature of their world-lines should be as small as possible, the colours should move along world-lines as slowly as possible, world-lines should preserve their spatial distances as much as possible, and the like.

However, in contrast to the very precise and detailed exposition of the definitions in the auto-psychological domain, Carnap does not state an explicit definition of the colour assignment *col* in terms of $\in$, Er, and the already defined terms, but leaves the issue with a general outline of the desiderata. It might seem that this is just a matter of abridgement rather than a problem that affects the transition from the autopsychological to the physical domain fundamentally. Quine (1951) famously thought otherwise:

> Carnap did not seem to recognize... that his treatment of physical objects fell short of reduction not merely through sketchiness, but in principle. Statements of the form 'Quality q is at point-instant $x; y; z; t$' were, according to [Carnap's] canons, to be apportioned truth values in such a way as to maximize and minimize certain over-all features... I think this is a good schematization... of what science really does; but it provides no indication... of how a statement of the form 'Quality q is at point-instant $x; y; z; t$' could ever be translated into Carnap's initial language of sense data and logic. The connective 'is at' remains an added undefined connective; the canons counsel us in its use but not in its elimination.

According to Quine, it is not a mere coincidence that Carnap did not spell out an explicit definition of the colour mapping: he simply could not have done so. While from the viewpoint of later philosophy of science, '*col*' would maybe count as a basic observational term that was not even in need of a definition, within a system such as the *Aufbau* '*col*' is the first instance of a *theoretical* term. That is: it is theoretical relative to the extremely parsimonious basis of the *Aufbau*. Its extension is pinned down in terms of a little theory which consists of certain principles or maxims that contain the basic terms $\in$ and Er as well as '*col*' itself. If all terms that are theoretical with respect to the basis of the *Aufbau* turned out to be definable just in terms of $\in$ and Er alone (apart from logical expressions), then these theoretical terms would have a meaning of their own that could be conveyed through primitive experiential or logico-mathematical terms. Accordingly, all sentences which would involve terms such as *col* would have a content of their own. This is precisely what Quine denies: only whole theories have content and only theories as wholes can be empirically confirmed or disconfirmed. This is Quine's doctrine of holism: meaning holism on the one hand and confirmational holism on the other.[13] In a nutshell: the transition from concepts for sense experience to concepts for the physical domain involves theoretical terms which cannot be defined in terms of the given experiential basis.[14] In section 8 we will see how this problem can be approached in new *Aufbau*-like setting. The next section is devoted to the basis of the "new" *Aufbau*.

6 The Basis of the New *Aufbau*

In some sense, it is not so surprising that Carnap's phenomenalistic constitution system is affected by the problems that were outlined by Goodman.

Carnap's basis is minimalistic, indeed *too minimalistic*: (i) Er is weak: since the similarity of erlebs is a notion of part-similarity, too many erlebs may turn out to be (part-)similar to too many other erlebs. E.g., a single common red spot on a particular location in the visual field suffices to let two erlebs come out to be similar. (ii) Er does not allow for "respects of similarity": there is no way of distinguishing cases in which two erlebs x and x' are similar in the very same respect in which two further erlebs y and y' are similar, from cases in which this is not so. (iii) Er does not support "gradations" of similarity: the similarity of an erleb x to an erleb y is an all-or-nothing affair; a comparative notion of resemblance would be more fine-grained and perhaps more plausible from a phenomalistic point of view.

Thus, the first step of avoiding Goodman's problems is to change the basis of the system. However, the solution is not just, say, to presuppose a primitive ternary relation of similarity of the form 'x is similar to y in a respect in which z is neither similar to x nor to y' (Eberle 1975 has suggested this as a solution to Goodman's problem). The main reason for the problems that affect quasianalysis is neither a flaw in the method nor the restriction to *binary* similarity, but rather that the content of information that is coded by a set of phenomenal quality spheres or by a set of phenomenal quality points simply cannot be coded by a similarity relation of erlebs with fixed finite arity (see Leitgeb 2007). This does not entail that the constitution of phenomenal quality spheres or quality points from similarity is absolutely impossible: if similarity is e.g. assumed to be a relation which is both "contrastive" and has *variable* finite or infinite arity, a substitute of quasianalysis can be found that is always adequate (this was suggested by Lewis 1983). Alternatively, if the domains of similarity structures are extended beyond the original domain of erlebs and if at the same time a numerical concept of similarity is used, phenomenal qualities can be constituted again (see Rodriguez-Pereyra 2002). Of course, none of these options tells us anything about how to approach Goodman's second problem.

The basic relations that we are going to presuppose in our new system are qualitative and still of fixed arity.[15] None of our basic relations is a similarity relation; instead, similarity will be defined later in terms of the new basis:

- (New) Basic elements: experiential tropes instantiated by the erlebs of a given and fixed subject S within a given interval of time;[16]

- (New) Basic relations: the membership relation $\in$, the temporal "before" relation $<$, and the relation Ov of "qualitative overlap".

Our new basic elements are tropes, i.e., property bits or property instances, which in our case we take to have an extended temporal "location" (see Mellor&Oliver 1997 for a collection of classic articles on tropes). A standard example of a trope would be *the red of the pencil that has been right in front of me for the last three seconds*. Our basic elements, however, are property bits which are exemplified by erlebs rather than physical entities; so an example would be more like *the red-colour-range in the left-upper part of my visual field that has been instantiated by my last few erlebs*. Note that erlebs themselves are not among the basic elements of our system; we only refer to them when we explain extra-systematically what the variables of the statements of our new constitution system are intended to range over.

Just as Carnap's erlebs correspond formally to sets of quality points – the sets of quality points that they realize – we assume our new basic elements to correspond formally to pairs $\langle C_q, C_t \rangle$ where (i) C_q is a bounded, extended, closed convex[17] set of quality points in a sensory quality space (visual, auditory, tactile,...), (ii) C_t is a bounded, extended, closed convex set of temporal instants on the real "time" axis, i.e., a compact interval of finite length, and (iii) there is an erleb of S which instantiates some quality point in C_q within the interval C_t. We will return to this formal representation below. Except for stating these necessary conditions, we leave open which pairs $\langle C_q, C_t \rangle$ among those that satisfy (i), (ii), (iii) actually *do* correspond to our basic elements, but it is clear that the more basic elements there are in our intended universe of discourse, and the more varied their temporal and qualitative relationships, the more the formal structure of our set of basic elements will approximate the formal structure of the set of *all* pairs $\langle C_q, C_t \rangle$ of convex sets with the described properties. In any case, we want to emphasize that the basic elements of our new system are *not* convex sets of points in a Euclidean space *themselves* but only that they can represented as such, such as locations on the surface of the earth can be represented by purely mathematical entities without coinciding with them.

$\in$ is of course again the set-theoretic membership relation. We use some standard first-order set theory (say, of the strength of ZFC) with urelements, where the urelements are our basic elements. This project is by no means a nominalistic one, and neither was its predecessor; standard mathematical resources are indeed crucial for its execution.

'$<$' is a binary predicate which expresses a relation of basic elements, such that $x < y$ if and only if x occurs "completely" before y, where 'completely' is meant to imply that x and y do not overlap temporally. According to the intended formal representation of our basic elements, if x is represented by $\langle C_q^1, C_t^1 \rangle$ and y is represented by $\langle C_q^2, C_t^2 \rangle$, then x stands in the

$<$-relation to y if and only if every member of C_t^1 is before every member of C_t^2 (which implies that $C_t^1 \cap C_t^2 = \varnothing$). Although our basic elements correspond temporally to compact intervals of $\mathbb{R}$ and thus to subsets of what is usually regarded as the formal model of *physical* time, one should not mix up $<$ with the order relation of real numbers. The latter holds between points in a non-denumerable continuum; the former is a relation of possibly finitely many experiential tropes that have a temporal extension.

The intended interpretation of the primitive term Ov can also be explained extra-systematically: 'Ov' is a unary predicate that applies to sets X of basic elements. It is the case that $Ov(X)$ if and only if the members of X have a common qualitative overlap. In terms of the formal model that we have introduced above, if $X = \{Y_i : i \in I\}$ and if each Y_i is represented by $\langle C_q^i, C_t^i \rangle$, then $Ov(X)$ if and only if $\bigcap_{i \in I} C_q^i \neq \varnothing$. Note that the overlap of two basic elements x and y is a special case of our general overlap relation, since binary overlap can be expressed easily by '$Ov(\{x, y\})$'. Accordingly, although 'Ov' is a unary predicate, we will often speak of Ov as an overlap *relation*, because it can be viewed as a relation that holds between the members of every set to which it applies.

Let us compare this new basis with Carnap's in the *Aufbau* and with Goodman's in his *The Structure of Appearance* (Goodman 1951). Carnap's idea was to start from erlebs and to define phenomenal quality spheres as an intermediate step in order to be able ultimately to state his intended definition of phenomenal quality points. Goodman's basic elements correspond roughly to Carnap's phenomenal quality points; his basic relations, which hold for these phenomenal quality points, are chosen in a way that makes it easy for him to compose complex phenomenal entities from the given atomic phenomenal units.[18] Finally, the basic elements of our new system are on a level of abstraction that corresponds to the level of phenomenal quality spheres: they are neither total momentary slices through S's stream of experience nor can they be regarded as "point-like" qualities, but they rather lie somewhere in between. In some respects, they resemble what Whitehead (see Grünbaum 1953 for an overview) and Russell (1954, 1961) referred to as extended "events".[19] From a phenomenalistic point of view, it is questionable whether "point-like" basic elements are subjectively accessible; points seem more likely to be abstractions from extended basic elements which are more easily accessible for a cognitive being, which might be an attractive feature of our new basis.

While Carnap's basic objects are concrete entities and Goodman's basic elements are abstract ones, the basic elements of our system share properties with both of them: like the former they can only occur within particular intervals of time; just as the latter they are instantiated in the same way as properties or types are instantiated by their bearers or tokens. It is a matter of terminology of whether our basic elements should thus be called 'concrete' or 'abstract'. Either way the basic elements that we presuppose are actual entities, i.e., our set of basic objects is not meant to include mere possibilia.

Here are some further remarks on the choice of our basis:

– Are we relying too much on the "metaphysics of tropes" here in order for this to be a "properly" Carnapian project? Not really. It is clear that every choice of a basis amounts to laying down an ontology for its corresponding constitution system; in this case, it is an ontology of experiential tropes and sets thereof, and three basic relations. But of course we do not claim in any sense that this is the "only" ontology to use, or the "right" one, or the "most fundamental one", or the like, which would be truly against the Carnapian spirit. Hopefully, our basis is just one that serves our purposes.

– Why demand that our basic elements correspond to pairs of *convex* sets? Convex sets have been suggested by Gärdenfors (1990, 2000) as plausible candidates for "natural" regions in quality spaces, i.e., the qualitative representations of "natural kinds" or "natural properties". Gärdenfors presents several arguments in favour of this suggestion: The quality space interpretations of classical examples of non-projectible predicates such as 'grue' (Goodman's new riddle) or 'non-black' (Hempel's paradox) are non-convex sets, in contrast with 'green' or colour predicates in general. Convex sets are not closed under complement and union, but the intersection of two convex sets in the same quality space is again a convex set; natural properties seem to obey the same closure conditions. While a bounded convex set can be ascribed a "center of gravity" which might be regarded as a prototype that corresponds to it, non-convex sets do not have this property; so convex sets subserve prototype representations. There is one further feature of convex sets that we want to add to Gärdenfors' list and which is of particular relevance in the context of the *Aufbau*: convex sets may be regarded as *respects of similarity* – if p is similar to r in a particular respect (say, Q) and q is qualitatively between p and r, then it seems to be necessary that p and r are similar to q in the same respect Q. But this is just the closure condition for convex sets, whence convex sets seem to be plausible candidates for qualitative respects of similarity.

– We do not assume that the subject S *perceives* basic elements; in fact, we regard the old "sense data" theory of perception as false. What we presuppose is that while S perceives physical objects and their properties, she *has* certain experiences. Sentences which involve our basic predicates may be used to describe which sense experiences S has. These descriptions of S's experience in terms of basic predicates are not necessarily S's "first-person" descriptions, but they might just as well be a neuroscientist's "third-person" descriptions. S is not assumed to be consciously aware of her sense experiences either, i.e., our basis is open to the existence of unconscious sense experience.

– Since the basis of our system – and the same holds for Carnap's – involves at the same time basic elements and basic relations, the basis is, in a sense, propositional from the start. It is a given that some set of basic elements has non-empty qualitative overlap or that one basic element occurs before another one does; what is given here is of propositional form. The sentences that can be formed in our restricted first-order language on the basis of '$\in$', '$<$', and 'Ov', are meant to express these "given" propositions. But we do not subscribe to any sort of epistemological foundationalism: sentences involving our basic terms are not necessarily certain or self-justifying; S might think that they are true or we might think that they are true but in fact they are false. As far as their justification is concerned, their status might differ only gradually from the status of sentences about the physical world. It is not even our primary goal to justify sentences about the physical world on the basis of sentences that can be formulated in the language of our new constitution system. The latter may indeed play some role in the analysis of empirical confirmation, but it is not obvious what this role actually consists in. In particular, empirical equivalence should not be mixed up with evidential equivalence, neither in the case where 'empirical equivalence' is explained in terms of a physical basis nor if it is understood in terms of a subjective basis: if A and $tr(A)$ are empirically equivalent, this does not by itself entail that whatever counts as evidence in favour of A is also evidence for $tr(A)$ and vice versa (see the discussion in Ladyman 2002). It should be kept in mind that it is even questionable whether Carnap's original *Aufbau* programme was a foundationalist one. The proponents of what we called the second interpretation of the *Aufbau* put forward very good arguments that it was not. In any case, nothing like Sellars' "myth of the given" applies to our new "Aufbau-like" system.

– We are not committed to any particular way in which $<$ and Ov are caused to hold between basic elements. It is clear that what S perceives is to play a role, but if some of S's *theoretical* beliefs also do so, this is fine with the new system. Our choice of basic elements and basic relations reflects the choice of a level on which S's experiences are described. We leave open to what extent these experiences are causally influenced by external input and to what extent they are shaped by internal mechanisms. What we call 'experience' is simply whatever is to be found on our chosen level of S's cognitive "life".

– The basis of our system has both an "enlightened" *phenomalistic* interpretation (as Carnap's in the old *Aufbau*) and a subjective *physicalistic* interpretation (as Quine's envisioned naturalization of the *Aufbau* in Quine 1969, 1993, 1995). We say 'enlightened' because of what we have pointed out above concerning sense data perception and epistemological foundationalism. One physicalistic way of viewing our basic elements is to think of them in terms of neural activation patterns of perceptual detector units: a pattern that corresponds formally to a pair $\langle C_q, C_t \rangle$ is generated by a detector if and only if an external stimulus is detected that overlaps qualitatively with the range C_q while overlapping temporally with the range C_t. Even if such a physicalistic interpretation is adopted, the basis is still subjective in the sense that the basic elements and the basic relations make up a subject S's experience. It is just that experience is now conceived from a naturalistic point of view. Carnap himself mentioned in the *Aufbau* the possibility of constitution systems other than the phenomenalistic system that he had chosen to work out in detail.

– It can be shown that the unary basic predicate 'Ov', which applies to *sets* of basic elements, could be replaced by a sevenary overlap relation of *basic elements*. Thus, we do not really rely on the fact that Ov applies to sets, although this choice is convenient from an expositional point of view. It may also be shown that no overlap predicate of lower arity could be employed if the definitions that we are going to introduce below are to be preserved.[20]

– The empirical contents of sentences, which we want to preserve by our translation mapping tr, will only be given relative to our choice of basic elements and basic relations; 'empirical content' in our sense is short for 'empirical content relative to the basis ...', where '...' is to be replaced by a description of our new basis. But of course there are other possible choices concerning basic elements and basic relations. Our basis might actually be constituted in terms of the basis of a different system, just as Carnap's basis turns out to be reconstructible in our own system. It might even be the case that the basic elements and basic relations of two systems are in some sense interdefinable. A basis with more primitive relations might correspond to a more fine-grained notion of empirical content, but perhaps the rather economical basis that we have chosen suffices in order to express the empirical contents of sentences in a non-trivial and satisfying way. Furthermore, the choice of a basis is always guided by extra-systematic *empirical* considerations on the system that would be determined by the basis. E.g., Carnap's choice was clearly motivated, and to some extent justified, by *Gestalt* theories of perception. Our own choice is inspired and – hopefully – also somewhat justified by theories in cognitive science, such as Gärdenfors' theory of natural regions in conceptual spaces, although we cannot say much about these background theories in this paper. However, it should be clear that every attempt of rational reconstruction such as Carnap's or the present one presupposes some amount of idealization. In this respect, it is helpful to think of the given subject S not as a human being but rather as an artificial cognitive agent. E.g.: If it turns out empirically that the visual space of humans cannot be considered as a five-dimensional quality space, then we might still assume our artificial subject S to have a visual space of the intended kind. We would then argue that if the empirical contents of scientific sentences relative to such an artificial agent can be analyzed within our constitution system, something similar might be achieved for an actual human agent on the basis of a sufficiently adapted system.

7 How to Solve Goodman's Problems

We are now going to introduce a sequence of definitions which is a part of our new constitution system. As explained at the beginning, the idea behind such a system of definitions in this context is that it determines a corresponding translation mapping for sentences. The final goal of the definitions in this section is to have a procedure at hand by which sentences about phenomenal quality points and their temporal and qualitative relations can be turned into sentences that are formulated just on the basis of '$\in$', '$<$', and 'Ov'. The strategy by which we want to approach Goodman's problems will be to consider first the dimensionality problem and only then the problem of defining phenomenal quality points. The change of basis, together with the change of the definitional procedure, will enable us to avoid the difficulties of companionship, imperfect community, accidental intersection, and collapse of dimensionality.

We start with the definition of 'set of basic elements', or briefly, 'Bas'. The members of the members of the extension of 'Ov' are definitely basic elements. Moreover, for every basic element x the set $\{x\}$ is certainly a member of the extension of 'Ov', because x has non-empty overlap with itself. Therefore, the following definition, by which all and only the members of the members of Ov are collected together, assigns the intended extension to 'Bas':

- Constitution of *set of basic elements*:

  $Bas =_{df} \bigcup Ov$.

Now we are going to make use of our basic relation $<$. At first, we can define a binary relation of temporal overlap for basic elements:

- Constitution of *temporal overlap*:

  $Ov_{temp}(x,y) \leftrightarrow_{df}$ (i) $x, y \in Bas$, (ii) $x \not< y$ and $y \not< x$.

This definition is justified in view of the fact that if a basic element x is neither totally before another basic element y nor totally after it – where the after-relation is just the converse of the before-relation – then x and y must overlap temporally. The reason why we did not start outright with a *basic* relation of temporal overlap is that subjective time does not only have an overlap structure – as the qualitative spaces have – but also an order structure, which we are going to exploit below.

Once we have temporal overlap, we can define time instants and a betweenness and order relation on them. Time instants are simply defined as maximal sets of basic elements that have pairwise overlap. The definition is related to Carnap's system in two respects: time instants have the same function in our system as the (then primitive) erlebs did in the original *Aufbau*; they include all instances of experience at a time. Secondly, our definition of time instants follows Carnap's strategy of defining phenomenal spheres, i.e., the first part of quasianalysis. Does the definition thus fall prey to the same shortcomings? No – in our case, every basic element corresponds temporally to a compact (i.e., bounded and closed) real interval. It can be shown that if every two intervals of a set of compact intervals have non-empty intersection, then the members of the set have a joint non-empty intersection.[21] The definition of betweenness below is unproblematic because our basic elements correspond formally to convex sets, which are by definition closed under betweenness. The definition of temporal order for time instants is simply the result of lifting our basic relation $<$ to the next higher level of abstraction. So we define:

- Constitution of *time instant* (or *erleb*):

  x is a time instant $\leftrightarrow_{df}$
  (i) $x \subseteq Bas$, (ii) for all $y, z \in x : Ov_{temp}(y,z)$,
  (iii) there is no $x' \subseteq Bas$, s.t. $x \subsetneq x'$ and for all $y, z \in x'$ $Ov_{temp}(y,z)$.

  Let $P_{temp} =_{df} \{x | x \text{ is a time instant}\}$.

- For all $a \in Bas, x \in P_{temp}$:

  a is at time $x \leftrightarrow_{df} a \in x$.

- Constitution of *betweenness of time instants*:

  For all $x, y, z \in P_{temp}$:

  $B_{temp}(x, y, z) \leftrightarrow_{df}$
  for all $a \in Bas$: if a is at x and a is at z, then a is at y.

- Constitution of *order of time instants*:

  For all $x, y \in P_{temp}$:

  $x <_{temp} y \leftrightarrow_{df}$ there are $x' \in x, y' \in y$, such that $x' < y'$.

Now we turn to the qualitative aspects of experience. We have already remarked that we want to define the dimensionality of phenomenal quality spaces before we define phenomenal quality points. Following Carnap, we can define the phenomenal counterparts of quality spaces as connectivity components, but not connectivity components with respect to a similarity relation but rather with respect to the given relation Ov of qualitative overlap. E.g.: Basic elements which correspond qualitatively to convex subsets C_q of the visual quality space do not stand in the Ov-relation to basic elements that correspond qualitatively to convex subsets C'_q of the auditory quality space. On the other hand, we may assume that the convex sets of quality points that our basic elements correspond to are distributed over their quality space in a sufficiently uniform way, such that every two of these convex sets in a common quality space can be connected by a chain of pairwise overlappings. Note that we invoke considerations on the exclusion of *unfavourable circumstances* here, just as Carmap did in the *Aufbau*, but in our case one can show that proper "variedness" of basic elements actually excludes such circumstances, in contrast with Carnap's own case.

This amounts to:

- For all $x \subseteq Bas$:

  x is a connectivity component $\leftrightarrow_{df}$

  - for all $y_1, y_2 \in x$ there are $z_1, \ldots, z_n \in Bas$ ($n \geqslant 0$), such that $Ov(\{y_1, z_1\})$, $Ov(\{z_1, z_2\})$,..., $Ov(\{z_{n-1}, z_n\})$, $Ov(\{z_n, y_2\})$;
  - for all $y_1 \in x$, for all $y_2 \in Bas$: if there are $z_1, \ldots, z_n \in Bas$ ($n \geqslant 0$), such that $Ov(\{y_1, z_1\})$, $Ov(\{z_1, z_2\})$,..., $Ov(\{z_{n-1}, z_n\})$, $Ov(\{z_n, y_2\})$, then $y_2 \in x$.

Now that we have defined connectivity components, we can turn to the question of how to assign dimensions to them. Here we make use of the auxiliary notion of k-Hellyness, which is defined as follows:

- For all connectivity components $x \subseteq Bas$, for all $k \in \{1, 2, \ldots\}$:

  x is k-Helly $\leftrightarrow_{df}$
  for every $y \subseteq x$ the following two conditions are equivalent:
  (a) for all $z \subseteq y$ with $|z| \leqslant k$: $Ov(z)$
  (b) $Ov(y)$.[22]

The dimensionality of connectivity components may be defined in terms of 'k-Helly'. By the famous theorem of Helly (cf. Matousek 2002), every class of closed, bounded, convex subsets of $\mathbb{R}^n$ is $(n+1)$-Helly relative to overlap in terms of non-empty intersection, where 'k-Helly' is defined analogously to the above. Moreover – in a non-degenerate case – a class of closed, bounded, convex subsets of $\mathbb{R}^n$ is not n-Helly.[23] E.g., the set of compact real intervals can be regarded as a degenerate subset of $\mathbb{R}^2$ in the sense that it can be regarded as a subset of $\mathbb{R}^2$ but that it can also be regarded as a subset of a space with lower dimension, i.e., of $\mathbb{R}$. We assume that the convex subsets of the five-dimensional visual quality space that our basic elements correspond to are distributed over it in a non-degenerate manner, i.e., their overlapping patterns may not be realized in a space with lower dimension; accordingly for all other quality spaces. Fortunately, the cardinality of the set of basic elements does not play a role here, since Helly's theorem also applies to finite classes of convex sets. So we have:

- Constitution of *k-dimensionality*:

  For all connectivity components $x \subseteq Bas$, for all $k \in \{1, 2, \ldots\}$:

  x is k-dimensional $\leftrightarrow_{df}$
  x is $(k+1)$-Helly, but not k-Helly.

Sense classes can thus be identified by dimensionality, which solves Goodman's second problem. In particular:

- Constitution of *visual phenomenal space*:

  $vs =_{df} \iota x$ (x is a connectivity component and x is 5-dimensional).

A visual basic element is simply a member of vs. Finally, within a sense class, quality points can be defined as maximal sets that have non-empty common overlap, which solves Goodman's first problem. Carnap's problem of "accidental" intersection does not occur, because rather than intersecting sets of erlebs, which may simultaneously realize points in different qualitative regions, we consider the overlapping of our basic elements, which correspond to such regions themselves. E.g., in the case of the visual phenomenal space:

- Constitution of *visual phenomenal quality point*:

  x is a visual phenomenal quality point $\leftrightarrow_{df}$
  (i) $x \subseteq \wp(vs)$, (ii) $Ov(x)$,
  (iii) there is no $x' \subseteq \wp(vs)$, s.t. $x \subsetneqq x'$ and $Ov(x')$.

  Let $P_{vis} =_{df} \{x|x \text{ is a visual phenomenal quality point}\}$.

In fact, within an n-dimensional sense class, quality points could be defined as maximal sets of $(n+1)$-fold overlappings, i.e., in (ii) and (iii) we could restrict ourselves to demanding that $Ov(\{y_1, \ldots, y_{n+1}\})$ for all $y_1, \ldots, y_{n+1} \in x$ (respectively, x'). This is again a consequence of Helly's theorem. Note that if we had defined the phenomenal quality points that belong to an n-dimensional quality space in terms of $(n+1)$-fold overlappings, it would have been crucial that the definition of dimensionality for phenomenal quality spaces had been achieved *before* the definition of their corresponding phenomenal quality points.

The set of visual phenomenal quality points can be equipped easily with a metric notion of similarity. The more uniformly distributed the quality regions and points in the visual space to which our visual basic elements and visual phenomenal quality points correspond, the more this metric will correspond to the actual metric on visual quality points:

- Constitution of *similarity metric on phenomenal visual quality points*:

  For all $x, y \in P_{vis}$:

  $d_{vis}(x,y) =_{df} |\{z \in vs\,|(z \in x \wedge z \notin y) \vee (z \notin x \wedge z \in y)\}|$.

d_{vis} measures the degree of separability of x and y in terms of visual basic elements. It can be shown that d_{vis} is a metric on P_{vis}.

Furthermore, we are able to define phenomenal quality spheres, a relation of part-similarity for time instants or erlebs, a betweenness relation for visual phenomenal quality points and accordingly for all other sense modalities, and many further interesting concepts, such as different types of qualitative or comparative similarity. All of Carnap's terms for the qualitative analysis of sense experience can be expressed on the basis of '$\in$', '$<$',

'Ov'; in particular, we can state a definition of the set of phenomenal colour qualities, the set of visual sensations, the set of places in the visual field, the neighbourhood relation for these places, and so forth (cf. section 5).

Summing up: Why is it that we were able to avoid Goodman's problems in our new setting? Our basic elements are already situated on the level of Carnap's phenomenal quality spheres, so we did not have to take the first step of quasianalysis; the difficulties of companionship and imperfect community simply do not arise. Accidental intersections are taken care of by our selection of basic elements and of Ov as the given relation of overlap. Since a binary notion of overlap would not suffice, we conceive of Ov as a class of sets, although a sevenary relation would actually do as well. In the case of temporal overlap, a binary relation, which is definable in terms of '$<$', is sufficient, since time is one-dimensional. By Helly's theorem, connectivity components are guaranteed to receive their intended dimension numbers, such that we are able to identify the different sense classes by their dimensions. This is achieved by exploiting just the overlap relation for our basic elements; the definition does not depend on a previous definition of phenomenal quality points. All of the stated definitions yield at least approximately the intended interpretations of the defined terms if only very mild assumptions on the overall experience of our subject S are satisfied; these assumptions can be made explicit extra-systematically, and – in contrast with Carnap's definitions of phenomenal quality spheres and phenomenal quality points – if the formal structure of our phenomenal basis coincides with the formal structure of the mathematical entities that it corresponds to, then all defined terms do receive exactly their intended interpretations.

Why do we have reason to believe that our definitions subserve the aim of determining a translation mapping that preserves empirical content? We tried to make sure that the extension of every defined term in our system is the phenomenal counterpart of its quality space preimage. If we were successful in doing so, then the formal structure of the actual quality space entities will show up in their phenomenal counterparts. E.g., the order structure of subjective time instants will be a coarse-grained image of the actual order structure of time, the dimensional structure of connectivity components will be a coarse-grained image of the actual dimensional structure of quality spaces, the metric structure of phenomenal visual quality points will be a coarse-grained image of the actual metric structure of visual quality points, and so forth. If $tr(A)$ is based on our definitions, it is therefore going to describe – though maybe in a coarse-grained fashion – the difference that the truth of A makes to possible experience: While A is a description of quality spaces and how they get realized or instantiated by experience,

$tr(A)$ is a description of the coarse-grained phenomenal copies of quality spaces – of how the formal structure of quality spaces "imprints" on the phenomenal structure of experience. Hence, at least approximately, $tr(A)$ should preserve the empirical content of A.

8 How to Solve Quine's Problems

In the following we build on work which originated with Ramsey (1931) and which was developed further by Carnap (1959, 1966a, 1966b) and Lewis (1970).

Let us reconsider Carnap's colour assignment sign '*col*' in the *Aufbau* as an example of a theoretical term. The procedure of setting up a translation mapping for sentences that contain '*col*' can be divided into two steps:

Step 1: Axiomatize Carnap's (implicitly stated) theory for the *primitive* colour-assignment function sign '*col*'.[24] Let $A[col]$ be the sentence which axiomatizes this theory; so $A[col]$ will include clauses of the form '$\ldots col(x, y, z, t) = c \ldots$', '*col* is such that...', and so forth.

The actual details of this axiomatization are tedious, because Carnap's maxims involve several auxiliary notions. Essentially, what one has to do is to define what we call the *set of colour assignment tuples*, where a colour assignment tuple collects the different components that Carnap refers to in his informal exposition. Formally, a colour assignment tuple is an octuple $\langle pv, dv, et, dev, lv, wlf, ca, ca_2\rangle$ where (i) pv is a mapping that tracks a possible point of view of S, (ii) dv is a possible main-direction-of-view mapping of S, (iii) et maps erlebs to points of time, i.e., to real numbers, (iv) dev represents a possible local-deviation-of-the-direction-of-view-mapping for S, (v) lv is a possible line-of-view function that is associated with S, (vi) wlf is a family of world-lines, i.e., of continuous trajectories through four-dimensional space-time, (vii) ca is a partial mapping from space-time to the set of S's phenomenal colour qualities – it is intended to be the colour assignment for points of space-time that are seen by S – and (viii) ca_2 is a mapping of the same type as ca but it is devoted to the assignment of colours to points of space-time which are unseen by S. The different components have to satisfy various conditions in order to let $\langle pv, dv, et, dev, lv, wlf, ca, ca_2\rangle$ be a colour assignment tuple. Some of these conditions ensure that the different mappings harmonize with each other – e.g. the line-of-view mapping has to "match up" with the point-of-view mapping, the main-direction-of-view mapping, and the local-deviation-of-the-direction-of-view mapping. Other conditions connect the mappings with S's actual experience; in particular, *et* has to preserve the temporal ordering of erlebs, *dev* has to respect the neighbourhood relation for places in the visual field, *ca* assigns points in space-time to phenomenal colour quality points according to S's visual sensations and S's line of view, as well as according to the assumed world-lines wlf along which colours are supposed to "travel"; finally, ca_2 fills in the "gaps" that are left by *ca*. All of these conditions are implicitly contained in Carnap's specification of the colour assignment mapping in §126–127 of the *Aufbau*. If expressed in our language, Carnap assumes that there are $pv, dv, et, dev, lv, wlf, ca, ca_2$, such that $\langle pv, dv, et, dev, lv, wlf, ca, ca_2\rangle$ is a colour assignment tuple and *col* is the result of "putting" the two partial mappings *ca* and ca_2 together[25]. However, being the fusion of the two last components of a colour assignment tuple is only a necessary condition for being Carnap's actual colour assignment *col*. Carnap's maxims in §126 may be reconstructed in the way that the colour assignment tuple $\langle pv, dv, et, dev, lv, wlf, ca, ca_2\rangle$ to which *col* belongs is maximally "inert" among all colour assignment tuples. This can be made precise by introducing measures of inertness on the set of colour assignment tuples: a colour change index (the higher the index, the less the total number of colour changes), a curvature change index (the higher the index, the less the total sum of curvature changes), a velocity index (the higher the index, the less the total sum of velocities), and a neighbourship preservation index (the higher the index, the higher the spatial neighbourship preservation for world lines). Each index maps a given colour assignment tuple to a particular number. Finally, based on these numbers, an inertness preorder for colour assignment tuples can be introduced by which one may express that one colour assignment tuple is less-than-or-equally-inert as another.[26] What Carnap's theory of colour assignment finally amounts to is this: there are $pv, dv, et, dev, lv, wlf, ca, ca_2$, such that (a) $\langle pv, dv, et, dev, lv, wlf, ca, ca_2\rangle$ is a colour assignment tuple, (b) *col* is the result of taking the unions of the two partial mappings *ca* and ca_2, and (c) $\langle pv, dv, et, dev, lv, wlf, ca, ca_2\rangle$ is maximal with respect to the inertness preorder on colour assignment tuples. $A[col]$ is precisely this statement.[27]

Step 2: On the basis of this axiomatization, we offer three main options of solving Quine's problem by setting up translations of sentences involving '*col*', i.e., sentences of the form $B[col]$:

Option 2.1: Translate $B[col]$ into the so-called Ramsey sentence[28]

$$\exists x(A[x] \wedge B[x])$$

Since the only descriptive terms in $B[col]$, except for 'col', are '$\in$', '$<$', 'Ov' and terms which are defined on the basis of them, the resulting Ramsey sentence only contains descriptive terms that can be reduced to '$\in$', '$<$', 'Ov'. Furthermore it is easy to see that the Ramsey sentence has the same logical consequences as the sentence $A[col] \wedge B[col]$, as far as sentences are concerned which solely consist of '$\in$', '$<$', 'Ov' (and logical terms); the two sentences are thus empirically equivalent at least in the syntactic sense of 'entailing the same observation statements'. The idea of the translation mapping is that if someone claims $B[col]$ to be true, he implicitly claims $A[col] \wedge B[col]$ to be true, because the extension of 'col' is given by the theory $A[col]$. But $A[col] \wedge B[col]$ may be regarded empirically equivalent to $\exists x(A[x] \wedge B[x])$. Ramsification can be viewed as a method of contextual definition so that the empirical content of 'col' is only explained in, and dependent on, the sentential contexts.

Ramsification is put forward sometimes as a means of making either the instrumentalist view of theoretical terms or the structuralistic view of scientific theories precise: according to the former, the only function of theoretical terms is that they help "ordering" or keeping track of our experiences in a neat way. The transition from sentences with theoretical terms to their corresponding Ramsey sentences seems to preserve precisely this aspect of theoretical terms. At the same time, the Ramsey sentences seem to subserve the aims of structural realists who want to show that the transition from former empirically successful but false theories to our current improved theories preserves "structural content"; the Ramsey sentences that are associated with theories are supposed to express their structural content. However, our intention of using Ramsey sentences is neither tied to an instrumentalistic picture of scientific discourse nor to a structuralistic account of scientific progress. As far as the first is concerned, we do not claim that $A[col] \wedge B[col]$ is just a short-hand for $\exists x(A[x] \wedge B[x])$ or that the two have the same meaning or pragmatic function. Our goal is simply to set up a translation mapping for scientific sentences that maps sentences to other sentences, such that (i) the latter are directly or indirectly composed of our basic terms, and (ii) the translation preserves empirical content. Ramsification is just a manner of achieving this goal. Concerning structural realism, Newman's observation (see Demopoulos & Friedman 1985), which is usually regarded to contradict the structural realists' aspirations of relying on Ramsey sentences in order to clarify the notion of 'structural content', is irrelevant for our project. Newman showed that a Ramsey sentence which consists solely of observational and logical expressions is roughly as strong as the set of all observational consequences of the original "unramsified" theory together with a cardinality assumption on its universe of discourse. Put differenty: the only "structure" which the Ramsey sentence adds to the observational part of the theory is a cardinality claim (see Ketland 2004 for the more precise model-theoretic statement). While this runs counter to the intentions of structural realists, it leaves our new *Aufbau* untouched; for our concerns, the translation of sentences in terms of Ramsey sentences only has to preserve empirical content and this is what we get. The additional cardinality constraint is irrelevant since our intended universe of discourse is assumed to include the whole set theoretic hierarchy anyway. The Ramsification of a theory with respect to a particular theoretical term only expresses what the structure of the extensions of the other terms has to be like if the theory is to come out as true. In our case, "the other terms" are just our basic experiential terms, such that the Ramsification of Carnap's theory of colour assignment with respect to the theoretical term '*col*' expresses what the structure of S's experience has to be like if the colour assignment theory is to be true.

Other criticisms of Ramsification do not apply to our system either: in particular, we do not regard Ramsification as subserving a particular theory of truth or meaning. E.g., as Glymour (1980) observes, while the inference from $P[t]$ and $Q[t]$ to $P[t] \wedge Q[t]$ is logically valid, the Ramsified inference from $\exists x P[x]$ and $\exists x Q[x]$ to $\exists x(P[x] \wedge Q[x])$ is not; but this is only a problem if the Ramsey sentences are supposed to determine or reveal the truth conditions of the original sentences. In our case, Glymour's observation amounts to an observation about the properties of the translation mapping tr that we are after. He shows that tr is not compositional: $tr(B[col]) = \exists x(A[x] \wedge B[x])$ and $tr(C[col]) = \exists x(A[x] \wedge C[x])$, however $tr(B[col] \wedge C[col]) = \exists x(A[x] \wedge B[x] \wedge C[x])$ rather than $tr(B[col] \wedge C[col]) = \exists x(A[x] \wedge B[x]) \wedge \exists x(A[x] \wedge C[x])$. While this is a fact that is interesting in itself, it certainly does not preclude tr from being the translation mapping that we were looking for in our section 2.

Option 2.2: Define '*col*' by a Lewis-style definite description (cf. Lewis 1970, Papineau 1996):

$$col =_{df} \iota x A[x]$$

If we pursue this option, our intended translation mapping is actually given by a definition (where 'x' runs over sets, including set-theoretic functions). However, if we decide to make use of Russell's theory of definite descriptions, this definition gives rise to a contextual elimination procedure again which resembles the one of the last option, the only difference being that now an additional uniqueness claim is included in the translation image. This has the following effect: assume that $B[col]$ is an atomic sentence; then $tr(B[col]) = B[\iota x A[x]] = \exists x(A[x] \wedge \forall y(A[y] \rightarrow y = x) \wedge B[x])$, so $B[col]$ does not precisely have the same logical consequences in the language given by '$\in$', '$<$', 'Ov' as $tr(B[col])$, since $B[col]$ does not imply $\exists x(A[x] \wedge \forall y(A[y] \rightarrow y = x) \wedge B[x])$ although $tr(B[col])$ does (trivially). However, as Lewis argues, if someone claims $B[col]$ to be true, (i) he implicitly claims $A[col] \wedge B[col]$ to be true, because the extension of '*col*' is given by the theory $A[col]$, and (ii) *additionally it is tacitly presupposed that* $A[col]$ *specifies the reference of 'col' uniquely.* If so, the slight increase of empirical content that happens to characterize the transition from the Ramsey sentence $\exists x(A[x] \wedge B[x])$ to the Lewis sentence $\exists x(A[x] \wedge \forall y(A[y] \rightarrow y = x) \wedge B[x])$ is acceptable.

A more serious concern about translation mappings according to option 2.2 is the question of how likely sentences such as $tr(B[col])$ are *true*. After all, 'x' runs over a set-theoretic universe; therefore, if $A[x]$ is not of a particularly restricted form, there will be "many" – in fact, infinitely many – values of 'x' which satisfy $A[x]$. Even worse, there might be instances of formulas $A[x]$ that are not satisfied uniquely, independently of what the extensions of '$<$', 'Ov' are like, i.e., independently of the qualitative features of S's experiences.

One way of avoiding this is to restrict the quantification in translation images to "natural experiential sets (relations, functions)": Not every member of our set-theoretic universe would count as a "natural" object. Although there may be many sets that satisfy $A[x]$, there is hope that there is just one natural set among them. Lewis (1970) uses precisely this "trick", although in his case the restriction is to natural *physical* kinds and relations. The suggestion can be made precise by introducing two types of variables, such that variables of one type would take arbitrary basic elements and sets as their values, while the range of the variables of the other type would be restricted. If x is a variable of the first kind and a is a variable of the second kind, then the definition above should actually be changed into: $col =_{df} \iota a A[a]$. Alternatively, one might introduce an additional unary predicate 'Nat' the intended interpretation of which is the class of all natural sets. The corresponding definition of '*col*' would thus be: $col =_{df} \iota x(A[x] \wedge Nat(x))$. Although both of these options are viable in principle, they come with a cost: in the first case, 'a' should no longer be regarded as a member of the *logical* vocabulary of the language of our constitution system (cf. Schurz 2006); it is a descriptive sign with a genuinely empirical content. Accordingly, in the second case, 'Nat' is another descriptive sign that is additional to '$\in$', '$<$', 'Ov'; in contrast with them, the extension of 'Nat' is unclear and cannot simply be explained extra-systematically in terms of examples and a formal model. In both cases, the new signs would have to be counted as further basic terms of the system.

Yet another way of dealing with the uniqueness problem is to include additional clauses which are supposed to ensure that the definiens is satisfied uniquely. In a nutshell, the idea is to define '*col*' by definite description with conventional choice. E.g., if all the x that satisfy $A[x]$ could be well-ordered, such that this well-ordering were definable in terms of '$\in$', '$<$', 'Ov', then the following definition would do: $col =_{df} \iota x \exists y (A[y] \wedge \wedge x \ is \ least \ w.r.t. \ \ldots)$ (where '...' is to be replaced by the defining clause of the well-order). Moreover, if such a well-ordering is not expressible – which is likely to be the case – then one might adopt the following strategy: for every colour assignment tuple, define its "coarsening", i.e., a tuple of coarse-grained versions of the components of the former. E.g., let the coarsening of a colour assignment tuple include mappings ca' and ca'_2 which assign colours to, say, *cubical regions* of space-time; a region would be mapped to a phenomenal colour quality c if and only if the mappings ca and ca_2 of the original colour assignment tuple map the measure-theoretic majority of points in the region to c (there are several possible variations of this recipe). The point of the coarsening is that if it is done in the right way, there will be *finitely* many coarsenings, as long as space-time gets shrunk to a sufficiently large sphere; the inertness indices that we have introduced above could be defined directly for coarsenings; finally, a well-ordering of coarsenings may be introduced, since there are definable enumerations of cubical regions, of time instants, of the set of phenomenal colours, of the set of visual sensations, and thus of the finite set of colour assignment coarsenings. Hence we can define: $col =_{df} \iota x \exists y (A[y] \wedge Coarsening(x, y) \wedge x \ is \ least \ w.r.t. \ \ldots)$ (where '...' is now to be replaced by the defining clause of the well-order for coarsenings). In this way, uniqueness can be guaranted without making use of quantification over natural classes. While the approximation of colour assignment tuples by their "coarse-grained" counterparts is certainly reflected by a change of meaning as far as the translation of sentences with '*col*' to sentences without '*col*' is concerned, the empirical content of the original sentences is likely to be unaffected. Our observer S may certainly be assumed to have finite capacities of discrimination herself, thus an assignment of colours to regions rather than points is all that is asked for if one is only interested in the preservation of empirical content. The underlying thought of each of these variants of option 2.2 is that the intended uniqueness of definite descriptions can be guaranteed if there is a manner of expressing a unique selection method for the objects that satisfy the description. The choice itself is conventional in the same sense as it is a matter of convention whether we choose Kuratowski's definition of ordered pairs in axiomatic set theory or a different one as long as the characterizing axiom for ordered pairs is satisfied. The drawback of this translation method is that the empirical contents of theoretical sentences would be determined only up to convention, but this is perhaps excusable. Carnap's Aufbau itself may be regarded as a conventionalistic project (cf. Runggaldier 1984).[29]

Option 2.3: Define '*col*' by a Hilbert-style epsilon term (see Zach 2003 for an overview):

$$col =_{df} \epsilon x A[x]$$

This is like defining '*col*' in terms of a definite description, however the uniqueness presupposition of iota-terms, and hence the problem that we have just dealt with, can be avoided: $\epsilon x A[x]$ denotes *an* object that satisfies $A[x]$ if there is one; otherwise, $\epsilon x \alpha[x]$ is undefined (and so are all sentences containing it). The logic and semantics of epsilon terms has been studied intensively since the days of the Hilbert school, and Carnap himself suggested to analyze theoretical terms as epsilon terms (cf. Carnap 1959, 1966b; see also Psillos 2000). Since defining expressions on the basis of epsilon terms is nothing but the singular term counterpart to Ramsification – both express *existence* claims – the former preserves empirical content just as the latter does.[30]

We suggest that one of these options can be applied in order to translate sentences with theoretical terms into sentences in the language of our constitution system, such that this translation preserves empirical content. According to either of these options, scientific sentences will normally be translated to rather "longish" sentences that include various fragments of scientific theories (thoug maybe stated in a form in which predicate constants have been replaced by variables for sets). In this sense, some of Quine's holistic aspirations are indeed satisfied by our translation mappings. As Quine points out,

> If we can aspire to a sort of *logischer Aufbau der Welt* at all, it must be one in which the texts slated for translation into observational and logico-mathematical terms are mostly broad theories taken as wholes. [...]
> The translation of a theory would be a ponderous axiomatization of all the experiential difference that the truth of the theory would make... we may, following Peirce, still fairly call this the empirical meaning of theories" (Quine 1969).

Since – as we claim – the extensions of our theoretical terms are typically given by certain theoretical modules or building blocks rather than by "the" scientific theory in total, our translation mappings only conform to a *partial* sort of holism. Furthermore, Quine seems to have overlooked the possibility of using these theory fragments in order to set up term-to-term and sentence-to-sentence translations which preserve empirical content. This solves Quine's problem as far as our new *Aufbau* project is concerned.

Three final remarks on our method of approaching Quine's problem:

– If we presuppose option 2.2 for the moment, then the definition of theoretical terms may involve our basic terms as well as terms – including theoretical terms – that have already been defined. This leads to a system of levels of terms, such that the definition of a term of level n only involves terms on levels below n. Friedman (1999) poses the question how such a system of constitutional levels is supposed to come to terms with the phenomenon of *revision*: E.g., the subjective colour assignment that is at first based solely on the immediate qualitative experience of our subject S has to be revised subsequently on the basis of the reports of other subjects on the one hand and on the basis of hypotheses about scientific regularities on the other; but our knowledge of other subjects and of scientific regularities presupposes our subjective colour assignment. Accordingly, the ultimate rational reconstruction of *col* seems to depend on the definition of concepts applying to other subject's reports and on further scientific concepts, whilst the definition of these other concepts seems to presuppose the definition of '*col*'. We submit that this circle can be broken by introducing new "high-level" theoretical terms as refinements of theoretical terms that have already been defined on lower levels. Our definition of '*col*' would e.g. be the definition of a preliminary colour assignment. On the basis of '*col*' and other primitive or defined terms, definitions of further scientific terms can be given. On the basis of the latter, a new term '*col**' may be introduced the extension of which may be regarded as a refinement of the original colour assignment *col*; and so forth. Carnap himself hints at this procedure when he defines what he calls a "preliminary time order" in §120 of the *Aufbau*.

– §155 of the *Aufbau* is devoted to an application of our option 2.2 in order to *define* what was originally meant to be Carnap's *basic* relation of recollection of similarity. The latter is defined as *the* binary relation that satisfies a particular high-level condition that is supposed to be characteristic of the relation of similarity recollection. In fact, Carnap uses the variant of 2.2 from above in which we suggest to make use of an additional predicate 'Nat', or in Carnap's terminology, '$Found$'. The extension of '$Found$' is supposed to be the class of "founded" or "experienceable" or "phenomenally natural" relations (cf. §154). Thus, Carnap's definition of the basic empirical predicate of his system is an early instance of Lewis' (1970) idea of defining theoretical terms by definite description, the only difference being that where Carnap makes the intended quantification over natural relations explicit in his object language, Lewis leaves it implicit in the metalinguistic interpretation of the variables he employs. As Demopoulos & Friedman (1985) and Friedman (1999) argue convincingly, 'Nat' or '$Found$' are not *logical* terms, therefore the strong structuralistic thesis that was presented in section 2 when we dealt with the second interpretation of the *Aufbau* is not supported by the existence of definitions of this sort. This failed structuralistic claim is of course not a part of our own thesis.

– Why is it that we have to make use of contextual definitions, or explicit definitions on the basis of additional logical resources such as iota-terms or epsilon-terms, in the transition from the autopsychological domain to the physical domain, while we have been able to restrict ourselves to standard explicit definitions in the former? The exact answer to this question would need more elaboration, but our hypothesis is that the approach in the last section is actually not as different from the one in this section as it may seem at first glance. Terms such as 'quality point', 'visual place', 'visual sensation' and so on, are theoretical terms themselves – their extensions are given by little theories on cognition. However, their corresponding definitions in terms of, say, Lewis' definite descriptions can be turned into equivalent explicit definitions which are of the same, or of a similar, form as the explicit definitions that we have given in our section 7. In contrast to expressions about the immediate qualitative properties of our experience, the empirical extension of '*col*' is too complex to be cast into a standard explicit definition on the basis of our primitive terms alone without recourse to logical devices such as iota-terms or epsilon-terms.

One final remark: we did not cover dispositional terms in this section since they are not be regarded as theoretical terms. Disposition terms constitute a separate and important problem for an *Aufbau*-like programme, but not a problem that we can deal with in this paper (cf. footnote 5). More generally, the translation of *modal* expressions, in particular function signs for objective single-case probability measures, and also of *statistical* expressions, needs special treatment that has to be postponed to a different occasion.

9 Summary and Outlook

We have finally arrived at a scheme for translating scientific sentences A to sentences $tr(A)$ where the latter consist solely of logico-mathematical signs (logical connectives, quantifiers, '$=$', '$\in$', 'ι', 'ϵ') and terms that refer to experience ('$<$', 'Ov'). In the case of the "autopsychological" terms that we dealt with in section 7, tr was given by standard explicit definitions. In section 8 we made several suggestions of how to translate sentences that include the colour assignment function sign '*col*' (and accordingly for other theoretical terms): either to apply Ramsification, or to define '*col*' in terms of a definite description, or on the basis of an epsilon term. Either way the translation is set up, logical and mathematical signs are always left invariant by translation.

As far as the preservation of empiricial contents is concerned, we took care that the extensions of all autopsychological terms are defined as to have the phenomenal counterparts of qualitative objects as their intended interpretations. This should guarantee that every definiendum of a definition in section 7 is empirically equivalent to its corresponding definiens. Finding solutions to Goodman's problems was a necessary prerequisite for achieving this. The translations of sentences that involve '*col*' in terms of Ramsey or Lewis or Hilbert/Carnap sentences can be shown to preserve empirical content while doing justice to Quine's holistic concerns about the corresponding passage of the original *Aufbau*. Since the extension of '$\in$' may be assumed to be fixed – at least from a Platonistic point of view on mathematics – each translation $tr(A)$ expresses a constraint only on the extensions of '$<$' and 'Ov'. As the last section has shown, this constraint might be a fairly complex one. E.g., $tr(A)$ might say that there is a mapping which is defined on a set in our set-theoretic hierarchy such that some condition that is expressed in terms of '$\in$', '$<$', 'Ov' is satisfied; the existence of such a function might correspond to a situation in which S has experiences which instantiates some complex pattern of temporal succession and qualitative overlap. Mathematical expressions are needed for two reasons: (i) they are necessary to set up the definitions of autopsychological terms; (ii) they occur in scientific theories; therefore, according to the methods of translation that we discussed in the last section, mathematical terms will show up in the translation of sentences with theoretical terms. In either case they enable us to express constraints on experience that could not be expressed on the basis of '$<$' and 'Ov' alone.

The translation mapping that is induced by our choice of basis in section 6, our choice of explicit definitions in section 7, and finally our choice of explixit or contextual definitions in section 8 is relativized to empirical theories in three ways: the basis is selected extra-systematically according to theoretical considerations; the definitions in the autopsychological domain only assign the intended extensions to their definienda if certain empirical hypotheses about S and her experiences are satisfied – e.g., S has sufficiently varied experiences, basic elements are distributed qualitatively in a sufficiently uniform manner, and so forth; the explicit or contextual definitions of '*col*' and of other theoretical terms include theory fragments. Thus, it is certainly not the case that our translation mapping is given by unrevisable rules of correspondence in the traditional sense of the word. Instead, every revision of our empirical theories may lead to a corresponding revision of the translation mapping. The choice of our translation mapping depends on empirical theories and so does the notion of 'empirical content' that we have used.

At least for all sentences A which solely consist of the linguistic expressions that we have investigated in this article, the thesis put forward in section 2 has been defended, except for one part: we still need to show that $tr(A)$ expresses a subject-invariant constraint on experiences. We leave this part for another paper.

Acknowledgements: This article has quite some history. Its first version was written up during my stay at Stanford as a visiting scholar in 2004/05, supported by a Erwin-Schrödinger Fellowship (J2344-G03) by the Austrian Research Fund FWF; in particular, I want to thank Michael Friedman and Ed Zalta for extensive discussion. Drafts were discussed in seminars in Salzburg, Leuven, Bristol, Stanford, Paris (IHPST), London (LSE), and at my Vienna Circle Beth Evening Lecture in Amsterdam in 2006. I would like to thank the participants for several helpful suggestions. An early written version was presented at the Formal Epistemology Workshop 2005 in Austin; I am grateful to several participants, including James Justus who commented on the paper, for their hints and remarks. I took up the theme again in 2008 (see Leitgeb 2008) when I gave related presentations on empirical content at the Theoretical Frameworks and Empirical Underdetermination Workshop in Düsseldorf, in a seminar in Salzburg organized by Edgar Morscher, and in a seminar at the ANU in Canberra organized by Dave Chalmers. Writing up the final version was supported by a generous grant by the Leverhulme Trust. I am very grateful to Reinhard Kleinknecht, Leon Horsten, James Ladyman, Samir Okasha, Alexander Bird, Gerhard Schurz, Ioannis Votsis, and Jeff Ketland for many helpful discussions, and to an anonymous referee for advice on this paper.

Notes

[1]There is not a lot of recent systematic work which aims to continue, extend, or modify Carnap's programme in the *Aufbau*. Moulines (1991), Mormann (1991), and the "Canberra Plan" (see Chalmers and Jackson 2001) are important exceptions.

[2]In the *Aufbau*, Carnap states some examples of what the translations $tr(A)$ of some concrete sentences A are like. For reasons of space, we will not be able to do so in this paper, but we will have to restrict ourselves to just a sketch of what such translations will be like in the new constitution system that we are going to develop below.

[3]However, we give a detailed model-theoretic analysis of empirical content in an unpublished draft; cf. Leitgeb (2008).

[4]In philosophy of science it is common understanding these days to explain the expressions 'empirically equivalent' and 'experience', as used in our new thesis, in terms of intersubjectively observable properties of physical entities. However, before the classical protocol sentence debate, it was simply a matter of choice whether one would analyze experience in terms of a physical basis or a subjective basis; different choices would be appropriate for different purposes, or so is Carnap's claim in the *Aufbau*. Without wanting to elaborate on this claim, we still regard it as true. Consequently, we want to leave open at this point how to understand 'empirically equivalent' and 'experience' exactly. Later on we will opt for a subjective basis that is to serve a similar epistemological purpose as Carnap's in the *Aufbau*, i.e., to reconstruct scientific expressions on the basis of terms that are epistemically prior to them: terms for subjective experience.

[5]Since our new thesis refers to a translation mapping that is not necessarily supposed to preserve meaning but only empirical content, one should maybe use a term different from 'translation' in this context. However, in order to compare the old theses to the new one, it is handy to use the same term. So 'translation' ought to be taken in an abstract sense as a mapping from one language into another which satisfies some preservation conditions that are specified separately and which do not necessarily concern meaning.

[6]In our context, the philosophical and mathematical differences between different suggestions for axiomatic systems of set theory are not of major importance.

[7]If one prefers to do so, one may just as well replace 'realize' by ''represent' – no demaning metaphysical views get expressed here.

[8]This mixture of qualitative and temporal components was rightly criticized by Moulines (1991) for having some counterintuitive consequences. In our new system, qualitative aspects will be separated conceptually from temporal ones by reserving one basic relation for each of them.

[9]For simplicity, we will not always be using Carnap's original terms.

[10]It can be shown that Carnap's definition of 'Sim' does not always subserve this intention. However, for the sake of the argument, we will ignore this additional problem of Carnap's procedure.

[11]In a one-dimensional quality space, a quality sphere of diameter ϵ is simply an interval of length ϵ.

[12]Carnap discusses this notion of dimensionality in his *Abriss der Logistik* (Carnap 1929).

[13]We will not go into details how meaning holism and confirmational holism differ from each other or what their logical relationship looks like. Moreover, Quine's view on this topic is not completely clear itself and was subject to subtle changes throughout the years.

[14]Carnap of course dealt with an undefinability problem himself when he studied the difficulty of defining dispositional terms on the basis of observation terms (Carnap 1936–

1937). But this topic should not be mixed up with the problem concerning '*col*': disposition terms are not theoretical terms as their extensions are not given by theories; they stand somewhere "in between" observation terms and theoretical terms. We will not be able to deal with Carnap's problem here.

[15]The extension of our new basic predicate '*Ov*' will actually be a set of *sets* of basic elements, which can be seen as a formal reconstruction of a Lewis-style relation of basic elements with variable arity. However, as we will point out below, one could in principle dispense with this basic predicate in favour of a sevenary predicate that applies to basic elements directly.

[16]Once again, all sets of such experiential tropes, all sets of sets of experiential tropes, and so forth, will be members of our intended universe of discourse, too.

[17]In our case, convex sets will always be subsets of some Euclidean space $\mathbb{R}^n$. A subset X of $\mathbb{R}^n$ is called *convex* if and only if for every $x, y \in X$, the straight line segment between x and y is included in X, i.e., for all $\lambda \in [0,1] : \lambda x + (1-\lambda)y \in X$. In the case of $n = 1$, convex sets simply coincide with bounded or unbounded real intervals, and hence bounded closed convex sets coincide with bounded closed intervals. Informally, a convex set is closed under "betweenness": if p and r are members of a convex set Q and q is between p and r, then q is a member of Q as well. By 'extended' we simply mean non-empty and not "point-like", i.e., neither identical to the empty set nor to a singleton set.

[18]In the preface of the second edition of the *Aufbau*, Carnap notes that he would now have opted also for phenomenal quality points as basic elements.

[19]Indeed, several of the definitions in the next section are inspired by Russell's (1954, 1961).

[20]There is nothing "magical" about the number seven: the overlap predicate for basic elements would need to have an arity that is at least of magnitude *highest dimension of quality space involved plus two* in order to let our definitions be adequate. Since the five-dimensional visual quality space is supposed to be of largest dimension, this yields an arity of seven.

[21]This follows from Helly's theorem, one of the classic results in Convex Geometry; see our definition of dimensionality. The point of assuming that our basic elements correspond temporally to *bounded and closed* convex sets was that Helly's theorem for infinite sets of convex regions only applies if these convex regions are compact, i.e., bounded and closed.

[22]$|z|$ is the cardinality of z. See Berge (1989) for the notion of k-Hellyness.

[23]As in the temporal case, we have assumed our basic elements to correspond to *closed and bounded* convex regions for the reason that Helly's theorem does not apply to arbitrary infinite classes of convex regions, without compactness being assumed as well.

[24]Strictly, '*col*' will not be a function sign in the first-order language sense, but rather an individual constant. However, since the individual constant '*col*' is intended to denote a particular function which is a member of the set-theoretic universe that we presuppose, one may still conceive of it as a function sign, except that in contrast to proper function signs, predicates may be applied to it.

[25]This is possible since the domain on which ca is defined is disjoint from the domain on which ca_2 is defined.

[26]One way of introducing such an inertness preorder is to rank the index functions by priority and to order the colour assignment tuples lexicographically according to the priority ranks.

[27]We have actually written up this statement in the language of our new constitution system. However, for the sake of brevity, we cannot reproduce it here.

[28]The original Ramsey sentences are second-order sentences. But since we presuppose set theory, second order quantifiers can be construed as first-order quantifiers.

[29]There are actually further variants of these options, which we have not discussed: e.g., Ramsification might involve quantification to natural classes, too, which would be a variant of 2.1.

[30]In Leitgeb (2008) we actually *prove* that both Ramsification and epsilon term substitution preserve empirical content, within a formal possible worlds framework in which every possible world contains an experiential substructure and where each possible world represents an internalistically accessible epistemic possibility.

10 References

Berge, C. (1989). *Hypergraphs.* (Amsterdam: North-Holland)

Carnap. R. (1928). *Der logische Aufbau der Welt*, Berlin: Weltkreis. English translation: Carnap, R. (1967). *The Logical Structure of the World*, trans. by R. George. (Berkeley and Los Angeles: University of California Press)

Carnap, R. (1929). *Abriss der Logistik.* (Vienna: Springer)

Carnap, R. (1936, 1937). Testability and Meaning. *Philosophy of Science*, 3, 420–268, and 4, 1–40.

Carnap, R. (1959). Theoretical Concepts in Science. Lecture, published in Psillos (2000).

Carnap, R. (1963). Intellectual Autobiography. (In P.A. Schilpp (Ed.), *The Philosophy of Rudolf Carnap* (pp. 3-84). (Peru, Illinois: Open Court)

Carnap, R. (1966a). *Philosophical Foundations of Physics.* (New York: Basic Books)

Carnap, R. (1966b). On the Use of Hilbert's ϵ-Operator in Scientific Theories. (In Y. Bar-Hillel et al. (Eds.), *Essays on the Foundations of Mathematics* (pp. 156–164). (Jerusalem: The Magnes Press)

Chalmers, D.J. & Jackson, F. (2001). Conceptual Analysis and Reductive Explanation. *Philosophical Review* 110, 315–361.

Demopoulos, W. & Friedman, M. (1985). Bertrand Russell's *Analysis of Matter*: Its Historical Context and Contemporary Interest. *Philosophy of Science* 52, 621–639.

Eberle, R.A. (1975). A Construction of Quality Classes Improved upon the *Aufbau.* (In J. Hintikka (Ed.), *Rudolf Carnap, Logical Empiricist* (pp. 55–73). Dordrecht: Reidel.)

Friedman, M. (1999). *Reconsidering Logical Positivism.* (Cambridge: Cambridge University Press)

Gärdenfors, P. (1990). Induction, Conceptual Spaces, and AI. *Philosophy of Science*, 57, 78–95.

Gärdenfors, P. (2000). *Conceptual Spaces*. (Cambridge: MIT Press)

Glymour, C. (1980). *Theory and Evidence*. (Princeton: Princeton University Press)

Glymour, C. (1992). Android Epistemology: Computation, Artificial Intelligence, and the Philosophy of Science. (In M.H. Salmon et al., *Introduction to the Philosophy of Science* (pp. 364–403). Indianapolis/Cambridge: Hackett.)

Goodman, N. (1951). *The Structure of Appearance*. (Cambridge: Harvard University Press)

Goodman, N. (1963). The Significance of *Der logische Aufbau der Welt*. (In P.A. Schilpp (Ed.), *The Philosophy of Rudolf Carnap* (pp. 545–558). La Salle: Open Court.)

Goodman, N. (1972). Seven Strictures on Similarity. (In N. Goodman, *Problems and Projects* (pp. 437–446). Indianapolis: Bobbs-Merrill.)

Grünbaum, A. (1953). Whitehead's Method of Extensive Abstraction. *The British Journal for the Philosophy of Science*, IV/15, 215–226.

Ketland, J. (2004). Empirical Adequacy and Ramsification. *British Journal for the Philosophy of Science*, 55, 287–300.

Ladyman, J. (2002). *Understanding Philosophy of Science*. (London: Routledge)

Leitgeb, H. (2007). A New Analysis of Quasianalysis. *Journal of Philosophical Logic*, 36/2, 181–226.

Leitgeb, H. (2008). Empirical Content and Its Presuppositions. Unpublished draft.

Lewis, D. (1970). How to Define Theoretical Terms. *The Journal of Philosophy*, 67/13, 427–446.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61/4, 343–377.

Matousek, J. (2002). *Lectures on Discrete Geometry*. (New York: Springer)

Mellor, D.H. & Oliver, A. (Eds.) (1997). *Properties*. (Oxford: Oxford University Press)

Mormann, T. (1997). Repräsentation, Struktur, Quasianalyse. Formale Aspekte einer Carnapianischen Konstitutionstheorie. (In G. Meggle (Ed.), *Analyomen 2* (pp. 431–438). Berlin: de Gruyter.)

Moulines, C.U. (1991). Making Sense of Carnap's Aufbau. *Erkenntnis*, 35, 263–286.

Papineau, D. (1996). Theory-Dependent Terms. *Philosophy of Science*, 63/1, 1–20.

Quine, W.V.O. (1951). Two Dogmas of Empiricism. *The Philosophical Review*, 60, 20–43.

Quine, W.V. (1969). Epistemology Naturalized. (In W.V. Quine, *Ontological Relativity and Other Essays* (pp. 69–90). New York: Columbia University Press.)

Quine, W.V. (1993). In Praise of Observation Sentences. *Journal of Philosophy*, 90/3, 107-116.

Quine, W.V. (1995). *From Stimulus to Science*, Cambridge: Harvard University Press.

Psillos, S. (2000). Rudolf Carnap's 'Theoretical Concepts in Science'. *Stud. Hist. Phil. Sci.*, 31/1, 151–172.

Ramsey, F.P. (1931). Theories. (In R. B. Braithwaite (Ed.) (1990), *The Foundations of Mathematics*, London: Routledge & Kegan Paul.) (Reprinted in D.H. Mellor (Ed.), *F.P. Ramsey. Philosophical Papers* (pp. 112–136). Cambridge: Cambridge University Press.)

Richardson, A. (1998). *Carnap's Construction of the World: The Aufbau and the Emergence of Logical Empiricism.* (Cambridge: Cambridge University Press)

Rodriguez-Pereyra, G. (2002). *Resemblance Nominalism.* (Oxford: Oxford University Press)

Runggaldier, E. (1984). *Carnap's Early Conventionalism.* (Amsterdam: Rodopi)

Russell, B. (1954). *The Analysis of Matter.* (London: George Allen & Unwin) (*The Analysis of Matter* appeared first in 1927.)

Russell, B. (1961). *Our Knowledge of the External World.* (London: George Allen & Unwin) (The first edition appeared in 1914.)

Schurz, G. (2006). Semantic Holism and (Non-)Compositionality in Scientific Theories. (In M. Werning et al. (Eds.), *The Compositionality of Meaning and Content*, Volume I: Foundational Issues (pp. 271–284). Frankfurt: Ontos Press.)

Worrall, J. (1989). Structural Realism: The Best of Both Worlds?. *dialectica*, 43, 99–124.

Zach, R. (2003). The Practice of Finitism. Epsilon Calculus and Consistency Proofs in Hilbert's Program. *Synthese*, 137, 211–259.