Philosophical Constraints on Analytical Definitions in Frege

[Abstract: Five Fregean philosophical principles are presented as constituting a framework for a theory of logical or conceptual analysis. These principles are operative in Frege's critique of other views and in his constructive development of his own view. The proposed conception of analytical explication is also connected to Frege's notion of analytical definition, and may form its core. It is also proposed that these conditions of analytic explication form a basis for what is required of a reduction of one domain to another, if it is to have the philosophical significance many reductions allegedly have.]

Although Frege's interests were obviously focused in Logic and the philosophy of mathematics (mainly in the foundations of arithmetic) his way of handling them is marked by broad and deep philosophical views, and by being involved in some general philosophical positions and doctrines, sometimes only hinted at and sometimes more fully developed. These, without in any way depriving Frege's technical achievements of their significance, may outlast the technical validity of his logistic project and are independent of it.

A central such interest, which in fact is the heart of what Frege thought he was doing, is logical analysis – its goals, its importance, and the constraints governing it. Frege was relatively explicit on the goal of logical analysis: The goal and criterion for logical analysis of a concept (or of proposition about it), which Frege compares to chemical analysis of a substance into its elements\(^1\), is that it, and the definition based on it, enable us to prove or expose implications involving the concept, which we approve, but couldn't prove, or didn't even see, before; Hence also its great importance: it not only reveals new implication links, but also brings to the fore the concepts on which the analyzed relies.

This idea of logical analysis finds expression already in the Introduction to Begriffsschrift of 1879 and persists throughout Frege's writings. A clear late expression from his "logic in Mathematics" of 1914 is:

"For it may not be possible to prove a truth containing a complex constituent so long as that constituent remains unanalysed; but it may be possible, given an analysis, to prove it from truths in which the elements of the analysis occur" (PW 209/226).

I shall call this feature of logical analysis the principle of implications enrichment.

It should be noted that this is also important for Frege's notion of consistency (e.g. in his debate with Hilbert), which, for Frege, is a property of a set of thoughts, not of
uninterpreted sentences. If these thoughts are not fully analyzed, the risk of inconsistency, he thought, is always there.

This feature of his conception of logical analysis is in fact rooted in a general conception of meaning expounded already in *Begriffsschrift* of 1879 (B), that what he called there "the conceptual content" (*begriffliche Inhalt*) of a proposition is constituted by its implication relations: Two propositions have the same conceptual content if they have the same implication relations, if they imply (perhaps with other premises) and are implied by the same propositions. The implication relations are conceived here as constitutive of the content of a proposition. This may also explain the deep philosophical reason for Frege's calling his logical language *Begriffsschrift* – concept script.² For, a logical language is designed to conspicuously display the implication relations of a proposition as a function of its inner structure; and to achieve this, the concepts of the proposition must be fully analyzed. Frege was, however, much less explicit on further constraints governing such analyses and definitions, to some of which we now turn.

A central thesis of Frege's *Foundation of Arithmetic* of 1884 (FA) is that numerical propositions like 'There are three apples on the table' "contains an assertion about a concept" [*eine Aussage von einem begriffe enthalte*, FA sections 46, 55]. This formulation, unlike many expositions of Frege's view, does not speak of a property of a concept, but is phrased in accord with the context principle in terms of propositions. Moreover, it is phrased in terms of the notion of about, which is a propositional notion – a relation between a proposition and things it is about.³ This, as I argued elsewhere, is a fundamental notion in Frege, posited at the basis of a fundamental Fregean principle, namely that any meaningful proposition is about something(s).⁴ Frege later proposed that the things – whether objects or concepts – a proposition is about are the references of its constituent terms,⁵ which also implied that a proposition cannot be about something not explicitly referred to by a constituent of it.⁶ I shall treat both claims as one thesis and call it "Frege's principle of about".

It is stated by Frege as a basic premise for his construal of quantified propositions, like universal propositions of the form 'All As are Bs', as well as of numerical propositions like the above, as consisting in an assertion about concepts.⁷ In the Introduction to Basic Laws of Arithmetic of 1893 (BL) p. 5/ix Frege presents this as his main discovery in FA. It also lies at the basis of his opposition to "implicit definitions" and of his polemics with Hilbert: One of Frege's main claims there is that such definitions do not determine meaning uniquely, and that "One surely needs to know what one is talking about" (*wovon man
aussagt, PW 213/230). This talk of the constituents of a proposition involves substantial views about its logical form, which determines these constituents, on which, Frege thought, vernacular language and traditional grammatical analyses may be profoundly misleading.

Frege's principle of about is often a quite effective constraint on explicating the sense of a proposition when we know what objects (or concepts) the proposition is about, like when we deal with regular empirical objects such as tables and apples, colors and persons. In such cases knowing the objects – their nature and ways of identification, standard ways of justifying claims about them, etc. – may pose constraints on explicating the senses (Sinne) of the terms concerned, for these senses should express ways in which these things (supposedly known to us) are given. It is for this reason that, say, modeling propositions about tables and apples in arithmetic would not count as explicating their senses or as analytic definitions of them (or of the terms referring to them; more on this later). It is much less effective, however, and in fact quite problematic when we deal, e.g. with abstract objects like numbers. If we don't know what objects the numbers are, we may lack enough constraints on explicating their senses. And this, of course, is the pivotal question of FA: "When we make a statement of number, what is that of which we assert something?" [von wem durch eine Zahlangabe etwas ausgesagt werde, FA section 45].

Somewhat more pertinently, this is also the reason why Frege did not even consider founding the analyticity of geometry on the possibility of modeling it in arithmetic, and ultimately in logic. Frege was of course aware of such possibilities. One might think that they should enable him to turn the reduction of arithmetic to logic into a reduction of geometry to logic (or the logical definitions of arithmetical terms into logical definitions of the geometrical ones), and thus establishing the analyticity of geometry. But he never proposed it. The reason for this is stated in another context in FA: "But surely, everything geometrical must be given originally by intuition" (alles Geometriche muss doch wohl ursprünglich anschaulich sein, FA 75). And an arithmetical or logical modeling of geometry would not respect the ways the geometrical things are given to us by intuition. This amounts to a philosophical requirement of an analysis and of analytical definition to the effect that it respects the way (or the standard way) the defined is given to us – its sense. The principle of about is connected to Frege's core idea of sense, on which more will be said in the sequel.

We said that the principle of about may constrain explicating the sense of propositions that deal with objects, like tables and apples, whose nature we know, but that it seems problematic with regards to other objects, like numbers, where we cannot presume such
knowledge. Frege's way of overcoming this difficulty, if only partially, was to reverse the order of explanation, namely to explicate the senses of these objects – how they are given to us – by their being constituents of the senses of propositions about them, which are independently explicated, and when this is achieved satisfactorily we may get a grip on what these objects are, or can be. Thus, by explicating the sense of many numerical propositions of the form "There are n Fs" and of what they are really about, we can get a grip on the meaning of n. This, being the general course of argument of FA, is admittedly put here in a quite general and programmatic way, but I trust it that anyone familiar with FA can guess the direction in which a fuller account would proceed.

Frege understood the notion of about realistically: A proposition cannot be about something (whether an object of a function) that does not exist. Holding this he had to show that the reference of any referring term exists. Frege in fact proves it for all referring terms in BL, but in FA he appeals to a broader view relying on a strong version of the context principle, according to which if a proposition containing these terms has a clear sense, which determines its truth conditions, this suffices for securing the terms references:

"It is enough if the proposition taken as a whole has a sense; it is this that confers on its parts also their content." [Es genügt, wenn der Satz als Ganzes einen Sinn hat; dadurch erhalten auch seine Theile ihren Inhalt (p. 71; p. 73).]

This was emphasized by Dummett, who construed Inhalt here as reference. Many scholars object here and proclaim that this goes against Frege's view that a term may have sense but lack reference.

I believe quite the opposite – that it is this possibility of a referenceless sense that goes against Frege's principles. Although some of Frege's formulations (especially in SR) seem to admit this possibility, they do not, I believe, express his better and considered view. It is rather this possibility which is incoherent with Frege's basic principles, like his adherence to the bivalence principle of classical logic, and his view that Thought is essentially true or false, his view that sense is a way a reference is given to us, and that a meaningful proposition is about the references of its terms. And indeed, most of his pronouncements on the topic go in this direction. Whenever a term lacks reference, this shows that it only seems to have a sense, but in fact doesn't. Strictly, sense for Frege is supervenient on reference. Yet, he conceived of certain extensions here in trying to account for our seeming understanding of expressions which only seem to have sense. These extensions
relate, according to Frege, mainly to vernacular expressions in mythical or poetical contexts, which he regarded as devoid of sense in the strict sense. When sentences are used to express genuine thoughts (which are their senses), they cannot contain constituents that are devoid of reference. This is a much discussed exegetical issue into which I shall not get further here.

The sense of a proposition is built up by the senses of its constituents, and these senses express ways in which their references are given to us or presented to us – their Art des Gegebenseins. This is Frege's main characterization of sense – its core idea. It should be distinguished from another one, made part of the Fregean lore especially by Dummett, construing sense as a "route to the reference" or a "way the reference of a term is determined". This, I think, has a rather slim basis in Frege. And it differs from Frege's main characterization of sense in its orientation and overtones: Frege's core idea is epistemic, directed as it were form the world to us; the latter (Dummett's) is anchored in the philosophy of language, directed as it were from language to the world. The former hardly makes sense for terms lacking reference; the latter can tolerate such an idea more naturally. However, by favoring the former I am not suggesting that Frege's notion is detached from the philosophy of language. On the contrary – they are tightly linked. The links, however, are more roundabout and sophisticated than suggested by Dummett's formulation and often presented by others. Things in the world are given to us in ways that are constrained and displayed by the meanings of the linguistic expressions that refer to them. Frege's core idea of sense may therefore be briefly put thus: A sense of a term is a way its reference is given to us as this is expressed by the (linguistic) meaning of the term referring to it.

This appeal to language as constituting ways things in the world are given to us marks Frege's divergence from Kant, and one of his main revolutions in philosophy. But this appeal to language should not undermine the general epistemic orientation of Frege's notion of sense as expressing a way something in the world is given to us. Explicating the sense of a term should therefore respect the way its reference is given us. It is for this reason that Frege was not satisfied with the mere possibility of modeling arithmetic in logic, but devoted much effort to arguing that numbers, as they are used in numerical propositions and in arithmetic, are logical objects and are given to us as such.

This notion of sense, encapsulating ways in which things in the world are given to us, has justificatory significance, particularly in forming an epistemic, non-deductive justification of basic truths (axioms) which cannot be derived from other more elementary
truths. I have defended elsewhere its ascription to Frege and argued that it is crucial for understanding his notions of analyticity and apriority, which he insists concern only justification, in applying them to basic truths (Bar-Elli, 2010). I shall not repeat it here, but add another aspect of it that has not been given its due attention. In his extensive attack on formalistic approaches in *Grundgesetze der Arithmetik* vol. II Frege contrasts what he calls "meaningful arithmetic" with "formal arithmetic". In attacking the latter, Frege's main arguments (and worries) concern the possibility of accounting for the applicability and the justification of a system of rules and axioms– arithmetic in that case. These he argues must do with senses and references and are therefore not available to formalists.

He argues at length that a formalistic approach cannot account for the applicability of arithmetic: "Arithmetic with no thought as its content will also be without possibility of application (*Anwendung*)" [...] Why can arithmetical equations be applied? Only because they express thoughts [...] It is applicability alone which elevates arithmetic from a game to the rank of a science" (Gg II §91, TPF 187). This applicability he emphasizes "cannot be an accident", and calls for grounding and account, which the formalist cannot provide (ibid.§89). In § 92 there he argues that since arithmetic and numbers have applications in various domains, its applicability is an internal arithmetic issue: "To this end it is necessary above all that the arithmetician attach a sense to his formulas".

However, it is not only the universal applicability of arithmetic which necessitates appeal to sense and reference; it is also the need to justify its base - rules and axioms: "**It is necessary that formulas express a sense and that the rules be grounded** (*Begrundung finden*) **in the reference** of the signs". (ibid. §92 p. 188; cf. § 94). In §101 he continue to argue that in meaningful arithmetic "We start from what is known about integers, from connexions grounded in their nature (*in deren Wesen begrundet sind*)". It is evident therefore that for Frege rules (and axioms) of arithmetic (and any genuine theory) are grounded and justified by the senses and references of the signs (which, he claims, are unavailable to the formalist).

We have briefly surveyed some philosophical doctrines, which, even if Frege, as some would claim, did not express as clearly as one could wish, are fundamental in his philosophy. Chief among them are: 1) the principle of implications enrichment by logical analysis; 2) the principle of about, including its corollary that understanding a proposition involves knowing what it is about; 3) the strong context principle, including its corollary that explicating the sense of a proposition secures reference to its constituents; 4) the core idea of sense and its corollary that explicating the sense of a term should answer to the
ways its reference is given to us; and 5) the justificatory role of sense – the view that the sense of a term – a way its reference is given to us – may have a justificatory role, particularly in justifying basic truths containing it, which cannot be derived from other, more elementary ones. Clearly, each of these principles deserves further clarification and elaboration, which partially I have tried to carry out elsewhere. The point I want to stress here is that this cluster of ideas forms a basis for a theory of logical analysis and of philosophical constraints on analytical definitions, to which I shall now turn.

**Definitions** are of course crucial for Frege's project, and he was well aware of that to the point he could write that arithmetic is derivable from definitions by means of logical laws (rather than from logic by means of definitions; see e.g. his letter to Liebmann, PMC 92-94). He repeatedly criticized definitions given by other mathematicians and logicians, and posed some stringent conditions for proper definitions (FA §§94-103; Frege expands on it in *Grundgesetze*, vol II §§55-164, tr. in part in FR 258-290 and in TPF). To mention some: a definition should be **conservative** – logically, no thought should be provable from it that is not provable without it, although psychologically it may be indispensable, and proofs and thinking without it would be intolerably cumbersome;¹⁴ it should be **complete** – Frege objected to partial or piecemeal definitions determining meaning only in some contexts, leaving open the meaning of the term in others; This includes the demand that a definition of a concept should be "sharp" – it should determine for any object whether or not the concept applies to it; It should determine the meaning of the defined **uniquely**; It should be given in terms of **Merkmale** – concepts of the same level as the defined, keeping strictly the distinction between first and second level predicates. The ultimate reason for these requirements is to prevent or minimize risk of inconsistencies. Actual non-appearance of contradiction cannot suffice here: "Rigoure of a proof remains an illusion, however complete the chain of inferences may be, if the definitions are only justified retrospectively by the non-appearance of any contradiction" (FA, Introduction IX; cf. also FA §94, and BL vol. II, §55).

But Frege was not only well aware of the importance of formal requirements of definitions, but also of their philosophically problematic nature and of philosophical requirements they should satisfy. Already in the Introduction to FA he writes:

"To those who feel inclined to criticize my definitions as unnatural, I would suggest that the point here is not whether they are natural but whether they **go to the root of the matter** [den Kern der Sache treffend] and are logically beyond criticism" (FA, p. xi).
"Natural" probably means here that anyone would recognize the definition as giving the
sense of what is defined, or even as offering a synonym. As if anticipating later
objections, Frege is well aware here that his definitions are not statements of synonymy
or of an even weaker meaning relation that would make them "natural". He is also
claiming that satisfying the formal-logical requirements, being "logically beyond
criticism", is of course necessary but not sufficient. Definitions should also "go to the
root of the matter". What does that mean if synonymy and naturalness are too strong,
modeling and even sameness of extension – too weak, and being logically beyond
criticism – insufficient?

Frege often distinguishes between definitions within a constructed system and
explanations of the senses of terms, which are preparatory to the system, and serve,
especially for primitive terms, in smoothing understanding and communication. He calls
these explanations Erlauterungen (translated in FG II, as elucidations). But within a system
one can also encounter definitions of terms with a long standing use. And this is the case
with most definitions in Frege's own systems. In his posthumous "Logic in Mathematics",
Frege distinguished "constructive" (aufbauende) definitions from "analytical" (zerlegende)
definitions. The former are in fact stipulations for shorts of complicated expressions.
They are, he says, like "receptacle for the sense... we conceal a very complex sense as
in a receptacle – we also need definitions so that we can cram this sense into the
receptacle and also take it out again" (PW 209/226). The latter, analytical definitions,
are definitions of terms in use, whose meanings are at least partially known and need in
some way to be maintained and clarified in the definition. These are of course the focus of
our (and his) interests, for they are the results of what he calls "logical analysis". For them
the distinction between the propaedeutic analysis and the definition of a term in use within
the system is somewhat blurred, for the former is evidently the rationale for the latter.

Frege then distinguishes two main sorts of analytical definitions. In the first, the
agreement between the complex sense of the definition and the sense of the "long
established simple sign [...] can only be recognized by an immediate insight"
[unmittelbaren Einleuchten]. In such a case, he says, the definition should be better termed
an axiom. Frege seems to assume here that if the senses of A and B are clearly recognized
one cannot be in doubt on whether or not they agree (ibid. 211/228). But what of the other
cases where such a doubt can arise?

Frege's answer divides into two parts, one of which is in fact what has come to be
known, following Carnap, as "explication". This amounts to a suggestion to replace a
term in use, whose meaning may be obscure, vague, only dimly or partially grasped, by a new term, precise and well defined within a theory. For such replacement to count as explication or analytic definition there should be a significant overlap in the use conditions of the two terms, which will include the paradigmatic and clear-cut cases of the former. This may be considered an extensional condition, and the whole idea, as Carnap emphasized, a **pragmatic suggestion** for replacing, for specific purposes, a term in use by a new precisely defined one. This has no pretension of being correct or of the latter providing logical analysis of the former, but only of being pragmatically expedient. Frege uses *Er lä uterungen* here, which was translated in PW as "illustrative examples". Whether or not the translation is a happy one, this seems to be only one kind of *Er lä uterung*. Another kind, sometimes translated as "elucidation", is e.g. "*sentences containing an empty sign, the other constituents of which are known*", which, as Frege says, may "provide a clue to what is to be understood by the sign", PW, 214/232).

However, in the second part of his answer Frege goes beyond this and writes:

"If this [the agreement of sense] is open to question although we can clearly recognize the sense of the complex expression from the way it is put together, then the reason must lie in the fact that we do not have a clear grasp of the sense of the simple sign, but that its outlines are confused as if we saw it through a mist. **The effect of the logical analysis of which we spoke will then be precisely this – to articulate the sense clearly** [Das Ergebnis jener logischen Zerlegung wird dann eben sein, dass der Sinn deutlich herausgearbeitet worden sein]. Work of this kind is very useful." (ibid. 211/228)

This passage is of great importance, both in itself and in understanding Frege's attitude towards his own definitions in FA (as well as in B and BL which are almost all of that kind). First, it implies that our subjective conviction that A and B differ in sense may be mistaken – it may arise even when in fact their senses agree. Secondly, Frege's ambitions seem here obviously much higher than Carnap's. For, he implies, what Carnap tried to avoid, that an analytical definition is the result of logical analysis of the old term and may amount to a **conceptual analysis** of it (and its concept), where the question of the correctness of this analysis is pertinent. To distinguish it from Carnap's explication, I shall therefore call it "**analytical explication**". But unfortunately Frege stops there and does not answer his initial question, except for the self-evident case, and does not say how we can judge the correctness of such an analytical definition, and to what such a judgment amounts. We can thus say that Frege here answers one aspect of the "paradox of analysis" and explains that analysis can be important and informative in giving a clear
and clearly recognized expression for an unclear and unclearly recognized concept. He seems to presume here that such an analysis is meaningful and expresses an objective thought. But he doesn't state what this thought is and what are the constraints on its correctness.

Some form of the above "extensional condition" is obviously necessary, so that analysans should meet at least paradigmatic cases of the use conditions of the analysandum. But Frege, as suggested above, aimed higher. As the term "analytical definition" suggests and as Frege makes explicit in the above passage, it amounts to a logical (conceptual) analysis, and as we saw before it should "go to the root of the matter". Frege poses the question of how we can know, or what it means that such an analysis and the definition resulting from it are correct. But he doesn't really answer it except for the case that it is immediately evident. If the problem was not extremely difficult and one of philosophy's perennial ones, this would be most surprising and almost inexcusable. For, as stated above, almost all of Frege's important definitions are of that kind and beset this problem. So, without at least some hints at the answer we might feel at lost in understanding the heart of his enterprise – what he thought he was doing in his logical analyses.

Evidently, I don't pretend to have a full answer. But I would propose that our above principles suggest philosophical constraints and conditions for achieving this: Analytical definition, analytical explication and logical (conceptual) analysis in general should expose and enable to prove implication relations we intuitively wish to hold concerning the analysandum (Implication enrichment principle); they should retain and clarify what propositions including the terms concerned are understood to be about (principle of about); they should articulate the sense of these propositions in a way that secures their constituents reference in accordance with the strong context principle; they should explicate this sense in a way that is responsive to the way its reference is given us (the core idea of sense); and they should articulate the senses concerned – ways their references are given to us – in a way that forms a basis for justifying basic truths and axioms about these references.

Philosophical constrains on definitions is a major concern in philosophy at least since Plato and Aristotle. It was a major topic in the debate between realism and nominalism and between extensionalists and essentialists throughout the ages. Frege had many reservations about classical forms of definition and did not participate in these debates in these terms. But his views about definitions, his polemics against many other approaches and his own
constructive way in FA and other writings, suggest a new path in this celebrated history, and our five principles are the basis of it.

These principles are the basis of much of Frege's criticism of other approaches. The principle of implication enrichment is often appealed to in his criticizing various analyses and definitions based on them, which do not show up in proofs and do not reveal any implication relation, and therefore are either senseless or of no use (see for instance, his long critiques of Weierstrass in "logic in Mathematics" 221/239 ff.). The principle of about is basic, for instance, in his criticism and rejection of formalistic approaches. This and the context principle are basis of many of his polemics against empirical and psychological approaches. The justificatory role of sense is, I believe, important in understanding his departure from Kant on the notions of analytic and a priori. But, perhaps even more importantly, these principles together with the core idea of sense are basic in the constructive development of his own argument in FA.

Now, as stated above, although Frege doesn't say so explicitly, most of his own definitions e.g. in FA, are what we called analytical explications. And the above principles are implicit in the course of argument of FA. The extensional condition is met by the proof that the definitions concerned offer a logical model of the basic truths of arithmetic (the so called "Peano axioms"). The principle of about is basic in the rational of Frege's doctrine of logical objects, for, logical truths, like any proposition, must be about some things. It is also appealed to (mainly in FA) in regarding the universal generality of arithmetic as a reason for regarding it as part of logic (and this, as we have seen, was the basis of Frege's account for the universal applicability of arithmetic). Accordingly, the definitions also explicate and clarify what the arithmetical propositions are "really" about – namely, logical objects and functions defined over them. They explicate the sense of the arithmetical propositions in terms of logical notions whose sense is beyond doubt, thus securing the arithmetical terms references, which render the propositions containing them objectively true or false, in accordance with the strong context principle. These definitions also reveal how the arithmetical objects are given to us – namely, as logical objects given by basic features of our ability to think. They thus form a basis for epistemic, non-deductive justification of basic truths couched in their terms, which is required for regarding them as analytic. A parallel argument works e.g. for the apriority of geometry, but we shall not follow it here.

As Frege makes clear on several occasions his main concerns in FA, and since the Nachwort for BL, his concerns with extensions and the problem his basic law V
encounters, were ultimately rooted in these philosophical aspects of his explicatory definitions. For, alternative courses of bypassing these problems would seem to him pointless if they did not form a basis for analytical explications of the nature of arithmetical and logical objects (i.e. extensions) and of how they are given to us. And the ultimate reason for this is the philosophical principles of an analytical explication, some of which were proposed above. And although I have talked of these principles separately, I hope it is clear that there are deep inter-connections between them, to the point we can talk of them together as constituting a skeleton of a theory of logical (philosophical) analysis.

As a sort of addendum let me make some general remarks on the relevance of the above conception of analytic explication to a common notion of reduction.

In many areas in mathematics, science and philosophy people talk of reductions – of reducing a "high level" theory – call it H – to a "lower level" one – L. Thus, logicians talk of reducing arithmetic to logic, mathematicians talk of reducing arithmetic and other branches to set theory, physicists talk of reducing thermodynamics to statistical mechanics etc. What is often meant is that the concepts of H can be defined in terms of L so that the truths of H are then derivable from the truths of L. Such claims often presume that H is "closed" in some way: that it is, e.g. a logical closure on a finite set of axioms (as in many mathematical cases). But this model of reduction is applied also when there is no such closure. If H is closed, a reduction can be strictly proved, but the philosophical merits of reductions tempt many to apply it also to "open" areas (like the mental) in which such proofs are unavailable and perhaps impossible.

This is a common notion of reduction, though there are several others. I shall focus in the sequel on definitions, which operate in most of them. My main claim (which I shall not back here except by hints) is that these definitions must satisfy stringent conceptual conditions in order to carry the philosophical burden people often ascribe to reductions. The main problem is not only that these conditions are only seldom if at all satisfied, but that they are not clear in themselves, and setting them is a big philosophical task. I would suggest that what we called analytic explication and the five conditions associated with it belong to this task.

People often also talk of reducing concepts so that the concepts of H are reduced to those of L. Though this is often conceived as a part of the reduction of a theory, it is important to keep the distinction between the two clear in view. One can speak of reducing concepts (say biological to physico-chemical) with no commitment to reducing
a whole biological theory. On the other hand, in many cases one can speak of reducing one theory to another just on the basis of modeling the one in the other, with no commitment to defining each concept of the one in the other. In some other conceptions people talk of reduction when L accounts for the empirical base of H, without deriving the theory H itself (with its concepts). Moreover, in many interesting cases, particularly of "open" fields (like the mental), where the set of true propositions is not closed and the concepts are open-textured and in a permanent change, there is no theory there to reduce. So the distinction between reducing a theory and reducing concepts may be important.

When reduction seems impossible, but H still is in some way dependent on L, people often talk of H as emerging from L. This usually applies to properties, i.e. when a property in H is indefinable in terms of the properties in L, and yet there is some causal connection between them, or when H supervenes in some other way on L. Yet in many such cases reduction still seems a desired goal and a working program.

No doubt, reductions are often important achievements, and sometimes amount to real breakthroughs within a field. This is indisputable. But very often they are also claimed to have philosophical significance, showing a certain field or theory to be "really" of a different nature from the one it is usually supposed to be. Thus by reducing arithmetic to logic people have claimed to show that arithmetic is "really" logic, and numbers are "really" some logical entities. Similar claims have been likewise raised regarding many other reductions. In particular, many brain scientists have recently claimed that many psychological and mental concepts, like seeing, feeling (an emotion), thinking, willing, imagining etc. are "really" not other than certain brain processes. The scientific, technical and medical merits of such reductions put aside, their philosophical motivation is often a belief that L is on a more solid or better known grounds than H, and therefore a reduction shows H to be in fact epiphenomenal. It is these philosophical claims that concern us here.

Definitions, sometimes quite sophisticated, are crucial for almost any reduction – they are its heart. A high-level theory H is couched in terms of some concepts C\textsubscript{H} and its reduction is achieved by defining these concepts in terms of those of the lower-level – C\textsubscript{L}. The exact nature of these definitions and the conditions they should satisfy pose severe problems. First the definitions, sometimes called "bridge rules" or "connecting rules", may take various forms, sometimes being empirical statements rather than definitions proper.\textsuperscript{21} People then have wondered whether in some of these forms they can carry the burden of reduction. But even waiving for the moment these difficulties aside,
the above, though a common formulation, is however imprecise and may be misleading. A more precise description is that some concepts are defined within \(L - C_{*L}\) – which are then claimed to be somehow, or in some way \textbf{equivalent} to \(C_H\). But these hedges ("somehow" etc.) raise questions – what exactly is this equivalence and what is the relationship between \(C_{*L}\) and \(C_H\)? In most (perhaps all non-trivial) cases they are definitely \textbf{not synonymous} or \textbf{logically equivalent}: after all, \(C_{*L}\) are new technical concepts defined within \(L\). We may call them by the names of \(C_H\); we may call them by whatever names we like. But we should not be misled by that to think that we were really defining the concepts of \(H - C_H\).

This, let me remark is a philosophical worry, and I leave here aside formal problems concerning acceptable definitions, as whether they should be explicit, conservative, constructive, etc. There are many attitudes to these formal aspects and I shall not go into it here. I shall focus on conceptual conditions that substantiate this claim of equivalence.

Two main routes should be considered for seeing the above relationship between \(C_{*L}\) and \(C_H\). One, which is perhaps the more common, is based on the idea that \(C_{*L}\) is claimed to be equivalent to \(C_H\) \textbf{because} by accepting the equivalence the reduction gets through. This may seem to amount to a \textbf{trivial circularity}: we claim to reduce \(H\) to \(L\) by means of defining \(C_H\) in terms of \(C_L\), i.e. by means of the equivalence of \(C_{*L}\) and \(C_H\), and we then justify the claim for this equivalence by the reduction. Let me put it in other words. For some reason or another we suspect \(H\) and its concepts. We try to improve things by reducing it to \(L\), which we consider to be on safer grounds. Such a reduction may be interpreted as showing \(H\) and its concepts to be epiphenomenal and superfluous; it may likewise be interpreted as validating \(H\) and showing it to be as solid and safe as \(L\). Be it as it may, a key move in this process is the claim of the equivalence of \(C_{*L}\) and \(C_H\). But when coming to justify this claim our main (perhaps only) reason is that it enables the reduction. When thus stated it seems flatly circular.

If \(H\) is closed and is proved to have a model in \(L\), the circularity may seem to be less troublesome, but it is much more acute when no such proof is known or even possible, as in the case of the mental. Even in the case of modeling \(H\) in \(L\) it could be argued that there is a difference between knowing that there is a model of one theory in another and actually constructing one. It is the latter, it could be argued, which is really needed. To this it could be replied that although for the actual scientific work and progress this difference is all important, for the general philosophical claims we are concerned with,
this difference is not crucial and the mere knowledge of the existence of a model might look to have significance not much less than the actual construction of one.

The above circularity, though a real defect in many expositions of reductions, may, when looked deeper, be more complicated and interesting. One aspect of the complication has to do with the relationship between a theory (a set of propositions) and its concepts. This brings us back to the difference between reducing a theory and reducing concepts, which we noted at the beginning. A landmark here was the debate between Frege and Hilbert about implicit definitions and Hilbert's procedure in the *Foundations of Geometry* (see also his correspondence with Frege) in which a concept (like point, straight line, plane, the relation of being in between etc.) is supposed devoid of meaning apart of a theory, and is (implicitly) defined by the axioms of the theory (Hilbert's axiomatization of Euclidean geometry in that case). As long as these axioms are proved consistent, they completely determine their concepts, and there is nothing to the meaning of these concepts beyond their role in the axioms. Frege argued that this is mistaken (in fact absurd). An axiom, he argued, is not a definition. It is a truth (or in any case a meaningful proposition) whose meaning presupposes that of its concepts and does not constitute it. The meaning of a proposition, for him, is built up and determined by the meaning of its constituents (the "compositionality thesis"), and Hilbert, he thought, along with many others, turn things upside down. Propositions including a concept are of course relevant to its meaning – they can (and in many cases should) serve as constraints on determining its meaning, but they cannot serve as definitions of it, as if it were devoid of meaning independently of them.

The controversy goes deep and concerns the nature of logic and its key notions. Although the history of mathematics in the 20\textsuperscript{th} century with a sharp distinction between syntax and semantics, formal systems and their models, can be claimed to go generally on Hilbert's side, and various holistic conceptions in philosophy were inspired by it. But whether there are convincing answers to Frege's complaints is still not clearly settled.

Likewise, one may argue in our case, the concepts of \(C_H\) are implicitly defined by \(H\), so that reducing \(H\) to \(L\) fixes their meaning in terms of \(C_L\). It is worth noticing that apart of Frege's general opposition to implicit definitions, the situation here and in his debate with Hilbert is more radical than in a regular "local" implicit definition, where a single term is implicitly defined by a proposition whose other terms are well understood. In contrast, in Hilbert's system no (non-logical) term in the axioms is understood, and all are simultaneously defined by all the axioms. Similarly in our case, the idea may presumably
be that all concepts of $C_H$ are simultaneously (implicitly) defined by $H$, and the reduction to $L$ gives it a solid basis. The intelligibility of this may be questioned even by those who don't share Frege's opposition to implicit definitions.

Things get more complicated (and perhaps even with an ironic twist) by the fact that it was Frege who proclaimed the basic principle that may be behind Hilbert's position – Frege's famous "context principle". This principle (proclaimed as basic in his FA of 1884, and made a cornerstone in his philosophy) says that one should never look for the meaning of a word in isolation, but only in the context of a proposition. This principle is the threshold of modern philosophy of language and of analytic philosophy, and was absorbed in most positions to the point of being almost unnoticed. It has been interpreted in many ways, in some of which it looks as anticipating and founding a general basis for endorsing implicit definitions such as Hilbert's. One can then wonder how could Frege object to this position and how his context principle can be reconciled with his firm adherence to the compositionality thesis. I cannot delve on this issue here (I have elaborated on this, proposing a solution in chapter 5 of Bar-Elli 1996), allowing myself just to proclaim that I believe that when both principles are properly understood Frege's position is seen coherent and even persuasive.

The above remarks pertain to any reduction including those "hard cases" in which there is a theory $H$ and it is even closed in some way. They are evidently more pertinent to concepts in fields like the mental, in which one can hardly talk of theories at all. When we talk of reducing mental concepts like thinking, imagining, believing etc to physical ones, the mental concepts are not part of or governed by a well defined theory (like geometry, arithmetic, thermodynamics and branches of biology). So one cannot talk here of a set of propositions that define, or even constrain the meaning of the relevant concepts. Usually what we have in these cases of alleged reduction is partial statistical correlations of applications, and some hints at causal relations. Hence, the difficulties in establishing the claim of equivalence are much severer.

The other route is based on Frege's notion of explication (Erlaeuterung), taken over (without mentioning its origin in Frege) and made famous by Carnap. A central idea there is that $C^*_L$, though having some advantages over $C_H$, should have a significant overlap of extensions or applications with $C_H$. As both Frege and Carnap emphasized, in most cases the overlap cannot be expected to be complete (so that $\text{Ext}(C^*_L) \supseteq \text{Ext}(C_H)$) and "significant overlap" here is not a precise criterion but a pragmatic directive, depending on what one wants to achieve by the reduction, and for what purposes.
But even when there is an extensional equivalence (or $\text{Ext}(C^*_{1}) \supseteq \text{Ext}(C_{1})$) this is still too weak to sustain the philosophical significance people attach to such definitions and reductions. Some of the reasons for this are spelled out above, but there are others on which I have elaborated elsewhere. To mention one purely logical point: for any consistent first order theory there is a model in arithmetic. This means that for any such theory there are "definitions" of its concepts (whether we can construct them or not) in arithmetic. But I suspect that few will see this as sufficient for reducing these concepts to numbers. Similarly the mere possibility of modeling a theory or domain $H$ in another is insufficient for a significant reduction – "significant" in the sense of having the philosophical meaning people claim it to have, as alluded to above.

When all this is taken into account one may doubt the philosophical significance people ascribe to many reductions. In order to have such significance, stringent conditions, both formal and conceptual, on the definitions of $C^*_{1}$ must be satisfied, which in many reductions are in fact simply ignored. Spelling out these conditions is a heavy philosophical task and amounts to endorsing a theory of conceptual analysis and of analytic definition. Whether the Fregean notion of what we called "analytic explication" and of "analytic (zerlegende) definition" (his term) can be strengthened to yield a basis for such a theory is, I believe, an important question (unfortunately quite neglected in the literature). I suggested above at least a general direction for a positive answer, based on some philosophical principles scattered, and sometimes fairly implicit in his writings.
Notes (numbers to the right of a slash refer to the German edition)

1  Frege uses the analogy on various place, e.g. "Boole's Logical Calculus and My Concept Script" (PW, 36/40), "On concept and Object" (FR, 182/193), "Logic in Mathematics" (PW, 208/225), "Foundation of Geometry II", Collected Papers (CP, 302/303).

2  I expanded on this in Bar-Elli 1986. Some of Frege's early remarks suggest this to be in line with the Leibnizean notion of lingua characterica, and he probably took the term Begriffsschrift from an article by Adolf Trendelenburg about Leibniz. However, in his late "Notes for Ludwig Darmstadtter" of 1919 (PW, 253/273) Frege emphasized the difference between his conception and Leibniz's, and even regretted choosing the term for its allusions to Leibniz's conception.

3  This is missed in some presentations of Frege's view. For instance in his "Frege's Theory of Number" (reprinted as ch. 6 in (1983) Parsons presents the thesis by saying: "Having a certain cardinal number is a property of a concept". Frege's formulation is not only more cautious in avoiding parson's somewhat unclear "having a number", but more importantly in being explicitly phrased in terms of proposition and the notion of about.

4  Bar-Elli (1996), ch. 7. I suggested there, against Dummett and many others, to found the realistic basis of Frege's notion of reference on that of about, rather than on the naming relation – a suggestion that seems to me better both philosophically and exegetically. The centrality of the idea of about occurs already in the early "Dialogue with Puenjer" (PW, 66-7/59-60) and persists throughout Frege's writings.

5  Cf. Dummett 1981, p. 196 ff. Though Dummett there seems to assume that reference is primary and we explicate about in terms of it, whereas I suggested in Bar-Elli 1996, ch. 7 the opposite direction.

6  This explicitly appears in SR (58/28). In FA it is explicit regarding objects: "It is impossible to speak of an object without in some way designating or naming it" (60), and implicit in the course of the general argument there regarding anything. It is re-affirmed in many of Frege's later writings. For example, in "Logic in Mathematics" of 1914 he writes: "[Myth and fiction apart] a proper name must designate something and in a sentence containing a proper name we are making a statement about that which it designates, about its meaning {Bedeutung}" (PW 225/243).

7  Bar-Elli 1996 in particular 7.D.

8  This has been sometimes called "identity-dependent" proposition; see e.g. Blackburn (1984).

9  This is a substantial thesis of BL and occurs in many of his writings after that. But it is already clearly stated in his "17 Key Sentences on Logic" (PW 174), which is probably an early piece (see Dummett 1991, pp. 66, 77). In article 10 he states: "The sentence 'Leo Sachse is a man' is the expression for a thought only if 'Leo Sachse' designates something (etwas bezeichnet)".
According to the strong interpretation, the logical formula displaying the analyzed form of e.g. "there are 3 apples on the table" secures sense to the proposition, thus conferring reference to "3", which doesn't occur in the formula. But there are weaker interpretations in which only the constituents of the formula are conferred reference. For an early such objection see e.g. Parsons ibid, p. 156. See also Dummett (1981), pp. 160-171. Dummett emphasizes that Frege, though admitting it, saw it as a defect in a natural language, and that he thought that semantics cannot allow sentences lacking truth value, but less cautious ascriptions of the view to Frege persists still today.

I shall not be strict here in distinguishing sense of a term from sense of its reference and trust it that the context makes clear which is meant. Likewise in talking of definitions I shall use "the defined" for a term and for its reference or sense.

In BL and later Frege often talks of sense as a constituent of a thought (which is the sense of a proposition). This, though having the merit of being in accord with the context principle, can be interpreted in various ways and can hardly settle the issue between the two characterizations discussed in the text.

In explaining that in thinking and talking we are often unaware of the sense of our words, Frege writes: "So if from a logical point of view definitions are at bottom quite inessential, they are nevertheless of great importance for thinking as this actually takes place in human beings" ("Logic in Mathematics", PW, 209/226).

The claim was raised by Husserl shortly after the publication of FA. Parsons cites Follesdal as claiming that in his Aufsätze and Rezensionen, 11-12 Husserl developed a distinction parallel to Frege's sense/reference one independently of and prior to Frege (see Follesdal's article cited by Parsons in From Kant to Husserl, Harvard, 2010, n. 22 p.145). Let me remark in this connection that although Frege's systematic Sinn/Bedeutung distinction appears in his papers of 1891-1892, the basic idea of the distinction appears already in section 8 of Begriffsschrift of 1879 (confined there to identity statements), and in section 62 of FA, both of which we know Husserl to have read shortly after their publication. So it is still possible that Husserl's distinction was inspired or influenced by his reading these Fregean sources before 1890.

The English "Analytical" here should not be confused with his notion of analyticity or analytical judgment (see Bar-Elli 2010,).

This is distinct, though related to Frege's use of Erlaute rung (translated as "explication" in FG II, 300/301, which is closer to Carnap's "explanation of the explicandum". Carnap attended Frege's course "Logic in Mathematics" in 1914 (see his "Intellectual Autobiography" in Schilpp, ed. (1963). The article in PW consists of the notes for this course, so there can be little doubt that Carnap took the idea of explication in ch. 2 of his Foundations of Probability from Frege.

See his Review of Husserl's Philosophie der Arithmetik (FR 224), in which, following Husserl, he gives a clear formulation of what has come to be known as the paradox of analysis. Notoriously, he doesn't answer it there.

Mind note 12 above, for strictly, analysis is in most cases of a sense rather than of a concept.
Frege had qualms about extensions prior to learning of Russell's Paradox and admitted his axiom V to be not as evident as other logical axioms. But after 1903 he gradually became more and more skeptical about the appeal to extensions, and the reduction of arithmetic to pure logic. This is a fairly well known story I shall not repeat here. See a review of this e.g. in Parson's ch. 5 of *From Kant to Husserl*, Harvard 2010.

This has been a large issue in the philosophy of science. A landmark in its early stages was Carnap's theory of "reduction sentences" in *Testability and Meaning*, and it continues today in current debates about various features of what is called "the Nagel model" of reduction.
References

Frege's Writings:


*Foundations of Arithmetic* (1884), tr. J. Austin, Blackwell, 1951 – FA


*Grundgesetze der Arithmetik* (vol. I 1893, vol. II 1903), Olms Verlag, 1998 – Gg


*Posthumous Writings*, tr. P. Long and R. White, Blackwell 1979 – PW


*Translations From the Philosophical Writings of Frege*, ed. P. Geach, and M. Black, Oxford, 1952 – TPF

"On Concept and Object", tr. P. Geach in TPF (also in FR) – CO.

"On Sense and Reference", tr. M. Black in TPF (also in FR) – SR.

Other Writers:


Husserl, E.: *Aufsaeze and Rezensionen*, 11-12.
