

Preprint of a paper appearing in
Philosophical Studies (Special Issue on Definitions)
Volume 69 (1993), pp. 113--153.

Frege on the Psychological Significance of Definitions

John F. Horty

Philosophy Department and
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742
(Email: horty@umiacs.umd.edu)

Contents

1	Introduction	1
2	Background	3
2.1	Indirect discourse	3
2.2	Some constraints on sense	6
3	Definitions	12
3.1	Definition and concept formation	12
3.2	Fruitfulness and sense identity	16
4	Sense and psychological significance	24
4.1	The conflict	24
4.2	Resolving the conflict	27
5	Conclusion	36

1 Introduction

Ever since Frege's introduction of senses as the theoretical representatives of meanings, it has been thought that there should be a close correlation between sentential senses (thoughts, propositions) and certain psychological states, the propositional attitudes. This has led a number of people whose concerns lie primarily with the philosophy of mind to take an interest in the formal semantic theories whose inspiration can be found in Frege; and it has led to a preoccupation among philosophers of language, and even some linguists, with certain problems belonging more properly to philosophical psychology. The strongest thesis in the vicinity is that the correlation between senses and psychological states should be exact. As an example, take the psychological states involving *belief*. Then the strong thesis is that two sentences ' S_1 ' and ' S_2 ' express the same sense just in case the psychological state of believing that S_1 is identical with the psychological state of believing that S_2 .

About twenty years ago, work emerging in the philosophy of language began to show that this strong thesis could not be right—at least, not when psychological states were understood as they must be to figure in the explanation and prediction of actions, and senses were understood in such a way that the sense of an expression determines its extension in each possible world. The problems with the thesis were already implicit in Saul Kripke's treatment of proper names, but it was Hilary Putnam in (1975) who first put the point in exactly this way, with an emphasis on psychological states.

Putnam focuses on natural kind expressions, such as 'water', and develops a kind of counterfactual argument that is now familiar. Suppose Arthur believes what he expresses through the sentence 'The seas are full of water', and consider his counterpart in a possible world like ours except that every bit of H_2O is replaced there by the distinct but indistinguishable substance XYZ. By hypothesis, since these two substances are indistinguishable, Arthur and his counterpart would then have to be in the same psychological state as far as the explanation and prediction of their actions is concerned; but they do not believe the same propositions. What Arthur believes is a proposition true in any world just in case the seas in that world are full of H_2O ; this proposition is the sense he assigns to the sentence 'The seas are full of water'. In its place, Arthur's counterpart assigns to this sentence, as its sense, a proposition true in any world just in case the seas in that world are full of XYZ;

and this is the proposition he believes.

Although the overall concept of meaning is multifaceted, we can introduce a notion corresponding explicitly to whatever aspect of this overall concept correlates with psychological states: let us say that a sentence ‘ S_1 ’ has the same *psychological significance* as ‘ S_2 ’ just in case the state of believing that S_1 is identical with the state of believing that S_2 (and likewise for the other propositional attitudes). What Putnam’s argument shows, then, is that natural kind expressions force a distinction between sense and psychological significance; the sentence ‘The seas are full of water’ has the same psychological significance for Arthur and his counterpart, but it carries a different sense for each of them. The psychological significance of this sentence depends only upon internal features of these two individuals, but its sense is determined also by external features of the worlds in which they are embedded.

A few years after this work, David Kaplan (1977) and John Perry (1977) argued that the presence in a language of indexical expressions leads to a similar kind of break between sense and psychological significance. And the observation was generalized even more broadly by Tyler Burge (1979), who showed, first, that the distinction between sense and psychological significance is not limited to those expressions that tend to force a *de re* reading (names, indexicals, natural kind terms), and second, that the external features affecting an expression’s sense can include social as well as physical conditions.

The central idea in this line of research—that the psychological significance of an expression cannot be identified with its sense—has had an important impact on contemporary philosophy of language, and also on our appreciation of Frege, the historical figure. I believe, however, that the way in which this idea has been argued for and developed in the literature captures only one dimension of the phenomenon. Most of these arguments follow a common pattern, apparent already in Putnam’s work. Individuals whose internal psychological states are supposed to be identical are imagined in different external circumstances; because their internal states are identical, it is concluded that their words must carry the same psychological significance, even though those aspects of sense determined by external factors might vary. At the very least, this pattern of argument relies crucially on a kind of contingency, the possibility that those aspects of the world that contribute to determining sense might be otherwise. More generally, the arguments seem to suggest that the break between sense and psychological significance results entirely from the fact that certain aspects of an ex-

pression's sense are determined by factors external to the individual. Indeed, the arguments are often characterized in this way, as arguments against *individualism* in semantics.

I do not intend to discuss these counterfactual arguments in any detail, but only the suggestion—which I do feel is implicit in much of the literature—that the distinction between sense and psychological significance springs entirely from anti-individualistic concerns. What I show in this paper is that the phenomenon is more general: there is reason to distinguish sense from psychological significance even apart from concerns about the external factors that might influence meaning, purely on the basis of individualistic considerations, and even in situations where the kind of contingency that drives the standard counterfactual arguments is absent. In fact, the motivation for the distinction set out here between sense and psychological significance can be found in Frege's own writings—in his treatment of definitions. Because of this, and because this distinction applies naturally to mathematical languages, it has more bearing on Frege's own concerns than the distinction arrived at through the standard counterfactual arguments; and it may have more bearing also on the real concerns of psychology.

The paper is organized as follows. Section 2 reviews the connection in Frege between sense and indirect discourse, as well as some other background constraints on the notion of sense. Section 3 is devoted to a study of defined expressions within Frege's semantic framework, and especially, to developing what is described there as a weak interpretation of his idea that definitions should be fruitful. These threads come together in Section 4: given the constraints set out earlier on the notion of sense, we will see that even the weak interpretation of fruitfulness forces a distinction between sense and psychological significance.

2 Background

2.1 Indirect discourse

In contemporary work, the connection between semantic content and psychological significance is usually established against the background of two general principles. The first is simply the principle of compositionality—the idea that the content of a compound expression is determined by its syntactic form together with the contents of its parts. The second is the principle that the semantic content of a sentence determines its truth value.

Together, these two principles allow us to formulate a familiar substitutional criterion for content identity between expressions. If the expressions ‘ E_1 ’ and ‘ E_2 ’ are identical in content, compositionality tells us that any two sentences ‘ $\Phi(E_1)$ ’ and ‘ $\Phi(E_2)$ ’—resulting from the respective placement of ‘ E_1 ’ and ‘ E_2 ’ in some sentential context ‘ $\Phi(\dots)$ ’—must likewise share the same content; and in that case, since content determines truth value, the two sentences must agree in truth value as well. In order to show that two expressions do *not* coincide in overall semantic content, therefore, it is enough to find some sentence in which the substitution of one expression for the other affects the truth value of the result.

Given this general substitutional criterion, the linkage between semantics and psychology can then be established, almost as a side-effect, by focusing on a particular range of sentences, those describing the propositional attitudes. If ‘ E_1 ’ and ‘ E_2 ’ carry the same content, we should be able to conclude, in particular, that a sentence of the form ‘Karl believes that $\Xi(E_1)$ ’ shares its truth value with ‘Karl believes that $\Xi(E_2)$ ’, that ‘Susan hopes that $\Psi(E_1)$ ’ shares its truth value with ‘Susan hopes that $\Psi(E_2)$ ’, and so on. If we can find any counterexample to this pattern—if it turns out, say, that ‘Janet is afraid that $\Omega(E_1)$ ’ is true, while ‘Janet is afraid that $\Omega(E_2)$ ’ is false—then the general substitutional criterion forces us to conclude that the expressions ‘ E_1 ’ and ‘ E_2 ’ must differ in some aspect of their overall semantic content; they differ, at least, in psychological significance.

This style of argument is often attributed to Frege, but in fact, he never did try to motivate his introduction of senses in exactly this way. In his own discussions, Frege tended to rely, not on truth value, but rather on a distinction he perceived in the “cognitive value” of different sentences. Some, he thought, could properly be classified as informative, or possessing cognitive value; others must be classified as entirely uninformative, or self-evident. In place of the contemporary principle that semantic content determines truth value, then, Frege’s own arguments were driven instead by the analogous requirement that content should determine cognitive value: if two sentences agree in semantic content, they are either both informative or both uninformative.

As an example, consider the most familiar of these arguments, from the beginning of (1892), where he focuses on a true identity of the form ‘ $a = b$ ’. Frege claims that this identity is informative, that it holds some cognitive value, unlike ‘ $a = a$ ’; these two identities, he says, are “obviously sentences of differing cognitive value” (p. 25). Because content determines

cognitive value, it then follows that the two identities cannot coincide entirely in semantic content; and so from compositionality, that there must be some component of content in which the expressions ‘*a*’ and ‘*b*’ differ as well. It is these distinct components that Frege describes as their *senses*.

Now, our contemporary interest in semantic content is largely focused on the matter of psychological significance. We are primarily interested in the problem of specifying truth conditions for sentences describing psychological states—our beliefs, hopes, and fears—rather than Frege’s problem of trying to account for differences in cognitive value. It is legitimate to wonder, therefore, why the notion of sense developed in Frege’s writings should be thought to bear any contemporary relevance. Why should we bother with a conception of semantic content keyed explicitly to cognitive value if what we really want is a notion that can be used in computing the truth value of sentences describing our psychological states? The answer, of course, is that these two concepts are supposed to coincide. Although he came at it from a different angle, what Frege was looking for in his notion of sense is exactly what we want today from a notion of psychological significance.

This coincidence between Frege’s concerns and our own is often just assumed, but in fact it needs argument. The easiest way to establish the coincidence is to show that expressions can be classified as identical in sense according to Frege’s principles just in case they can be classified as identical in psychological significance according to our contemporary standards.

One direction of argument is straightforward. Suppose the expressions ‘ E_1 ’ and ‘ E_2 ’ fail to satisfy Frege’s criterion for sense identity: there is some sentential context ‘ $\Phi(\dots)$ ’ such that ‘ $\Phi(E_1)$ ’ is informative while ‘ $\Phi(E_2)$ ’ is not. It then follows that the two expressions will fail also to satisfy our contemporary criterion for identity of psychological significance. There will have to be some propositional attitude sentence in which the replacement of one of these expressions by the other affects truth value—for by assumption (almost), the sentence ‘It is informative that $\Phi(E_1)$ ’ will be true, but the sentence ‘It is informative that $\Phi(E_2)$ ’ false.¹

¹The reason for the qualification, of course, is that the slide between representing the concept of informativeness as a metalinguistic predicate and representing the same concept as a sentence forming operator is not quite trouble free: it is not quite obvious that a sentence ‘ S ’ should be classified as informative just in case ‘It is informative that S ’ is true. In logic, it sometimes does make a great deal of difference whether a particular concept (such as necessity) is represented formally as a predicate or an operator. But I do

The other direction of argument is more complicated, since there seems to be no direct route from the premise that two expressions fail to satisfy our contemporary criterion for identity of psychological significance to the conclusion that they should fail also to satisfy Frege’s criterion for sense identity. Even if we suppose that the replacement of ‘ E_1 ’ by ‘ E_2 ’ does affect the truth value of some sentence describing our psychological states—say, ‘Karl believes that $\Xi(E_1)$ ’—it is hard to see how we could conclude directly from this that there should be some sentence also in which the replacement of ‘ E_1 ’ by ‘ E_2 ’ affects cognitive value. At various points throughout his discussion in (1892), however, Frege affirms four additional principles that allow us to connect our contemporary notion of psychological significance with his conception of sense through a more roundabout route. The principles are: (i) that the referent of a sentence is its truth value (p. 34); (ii) that the referent of a sentence is determined by the referents of its parts (p. 35); (iii) that expressions in contexts of “indirect speech” take their “indirect” referents (p. 28); and (iv) that the indirect referent of an expression is its sense (p. 28). Using these principles, we can conclude that ‘ E_1 ’ and ‘ E_2 ’ must differ in sense just because they fail to meet our contemporary criterion for content identity, even if we cannot show directly that they fail to meet Frege’s criterion. Because they yield sentences with different truth values when substituted into the context ‘Karl believes that $\Xi(\dots)$ ’, we can conclude from (i) and (ii) that ‘ E_1 ’ and ‘ E_2 ’ take different referents in this context. Because the context is one of indirect speech, we can conclude from (iii) that the indirect referents of ‘ E_1 ’ and ‘ E_2 ’ must be distinct; and so from (iv), that they must differ in sense.

2.2 Some constraints on sense

Psychological limitations

Even according to Frege’s own principles, then, differences of psychological significance are supposed to be reflected in differences of sense; and so he *could* have tried to motivate his introduction of senses in a way more akin to our contemporary methods, by focusing on propositional attitudes. However, he never actually did so, perhaps because of the additional not think that this difference between the two ways of representing informativeness would have mattered much to Frege; and although he does usually represent the concept as a metalinguistic predicate, there are occasions (1914, p. 224; XIV/11, p. 126) on which he seems to treat it instead as an operator.

complications that would have been involved in appealing explicitly to the principles (i) through (iv). Each of Frege’s own arguments follows the course set out in (1892); each relies on his crucial contrast between informative and uninformative sentences—between those possessing cognitive value (XIV/11, p. 126; XV/14, p. 152; XV/18, p. 164), which he describes also as containing “valuable extensions of our knowledge” (1892, p. 25; see also VIII/12, p. 80) or “increas[ing] our knowledge” (1914, p. 224), and those without cognitive value, which he describes also as “self-evident” (XV/14, p. 152; 1914, p. 224).

Although the analysis of this contrast that will guide our work here is familiar, it is worth setting out explicitly. We will suppose that a true sentence ‘*S*’ is informative, possessing real cognitive value, if it is possible to understand ‘*S*’ without knowing that *S*; and that it should be classified as uninformative or self-evident otherwise, if understanding ‘*S*’ entails knowing that *S*.² Even if this analysis is correct, however, it is still only schematic. As it stands, the analysis cannot be used to determine the status of particular sentences: it does not tell us, for example, whether ‘ $2 + 3 = 5$ ’ is supposed to possess cognitive value, or whether it is self-evident. Because of this, the analysis cannot be used to settle particular questions of identity between senses or psychological states: it does not tell us whether ‘ $2 + 3$ ’ and ‘ 5 ’ are supposed to share the same sense, or whether the state of believing that $5 = 5$ is identical with the state of believing that $2 + 3 = 5$. In order to arrive at a more concrete understanding of Frege’s notion of sense, therefore, we must see how the underlying ideas of cognitive value and self-evidence might be deployed in particular cases.

As it turns out, the classification of a range of sentences as informative or uninformative can be seen as reflecting a background conception of the speaker’s psychology—particularly, the degree and kind of intelligence that we attribute to the speaker in our judgments of knowledge and understanding. If we suppose that speakers possess greater intelligence, we are more likely to presume that understanding entails knowledge: where ‘*S*’ is some true sentence, we are more likely to treat evidence that a speaker does not know that *S* as evidence that he simply does not understand the sentence, and so we are more likely

²This analysis agrees with Dummett’s view of the uninformative truths as those that he defines as “trivially true” (1973, p. 289). It is worth noting that the analysis treats the ideas of informativeness and unformativeness as use-mention hybrids, relating sentences to their contents. The hybrid nature of these concepts may help to explain why Frege wavered between treating them as metalinguistic predicates or sentential operators.

to classify ‘*S*’ as self-evident. If our judgments are based on a conception of speakers as somewhat less intelligent, we tend to accept a looser connection between understanding and knowledge: we are less likely to include the requirement that a speaker know that *S* among our standards for judging simply that he understands ‘*S*’, and so we are more likely to classify the sentence as one possessing cognitive value.³ Now let us ask: what conception of the speaker’s psychological abilities lay behind Frege’s own ideas of cognitive value and self-evidence?

One option is to suppose that Frege had been relying implicitly on a conception of speakers as creatures, like those described by A. J. Ayer (1936, pp. 85–86) or Hans Hahn (1933, p. 159), whose reasoning is perfectly accurate, comprehensive, and instantaneous. Just by understanding a language, ideally intelligent creatures like these—Ayer-Hahn monsters—would have to know all of its a priori truths, all the truths expressible in the language that could be discovered through reasoning alone. In that case, our schematic analysis would force us to classify each sentence expressing an a priori truth as self-evident; and we can assume also, for the sake of simplicity, that any sentence expressing an a posteriori truth could be classified as possessing cognitive value.⁴

An interpretation along these lines would naturally carry with it very loose standards for identity between the senses of particular expressions, and therefore, between the psychological states defined in relation to these senses. By Frege’s criterion, there is reason to distinguish the senses of two expressions only if their exchange in some context leads from a self-evident sentence to one possessing cognitive value; and so any two a priori equivalent expressions would have to be assigned the same sense—since self-evidence is identified with a priori truth, and there is no way to shift the status of a sentence from a priori to a posteriori by exchanging a priori equivalent expressions. We would have to agree with Hahn that ‘5’ has the same meaning as ‘ $2 + 3$ ’, or with Ayer when he says that ‘7189’ and ‘ 91×79 ’ are synonymous. If ‘7189’ carries the same sense as ‘ 91×79 ’, compositionality tells us that the

³A similar connection has been pointed out by Fodor (1979, p. 107) between the intelligence (“optimality of functioning”) attributed to a system and the transparency of its propositional attitudes.

⁴There is really nothing to prevent us from stipulating that certain a posteriori sentences also should be classified as self-evident; we might decide that they express truths so fundamental that a speaker could not be said to understand these sentences without knowing that they are true. However, this possibility leads to unrelated complications, and I ignore it.

two sentences ‘ $7189 = 7189$ ’ and ‘ $91 \times 79 = 7189$ ’ would have to express the same sense as well; and so the psychological state of believing that $7189 = 7189$ would have to be identified with the state of believing that $91 \times 79 = 7189$. Or to take a more extreme example, we would have to classify the four color theorem as self-evident, since it is true a priori.

Of course, from a contemporary perspective, this interpretation of self-evidence as a priori truth, along with the accompanying treatment of senses and psychological states, seems problematic for all the familiar reasons: because it is based on a conception of speakers as ideally intelligent creatures, the interpretation often leaves us at a loss in describing creatures like ourselves—at least some of whom once seemed to understand the statement of the four color theorem without knowing that it was true. But Frege was not necessarily driven by our concerns. Although it seems problematic for us, it is conceivable that Frege, like some of the logical positivists, might have been willing to accept the interpretation of self-evidence as a priori truth; and his contrast between the informative and the uninformative sentences is often interpreted in this way, as a contrast between a posteriori and a priori truths, by writers whose central focus is the semantics of empirical languages.⁵ Could the interpretation be correct?

In fact, there is some textual evidence that Frege’s notion of sense was based on a view of self-evidence as a priori truth. The evidence occurs in an unfortunately prominent position, just at the beginning of (1892), where he contrasts a priori sentences with those that may “contain very valuable extensions of our knowledge”—suggesting that the sentences which may extend our knowledge, those possessing cognitive value, must be a posteriori (p. 25). But this evidence is overridden at once: at the end of the same paragraph, he describes another a priori truth, a sentence of geometry, as one that does contain “actual knowledge” (p. 26). And throughout the rest of his work, he provides a number of examples even of very simple a priori equivalent expressions which, he says, differ in sense. In (1893, p. 7), for instance, we learn that ‘ 2^2 ’ and ‘ $2 + 2$ ’ carry different senses; we learn in (1914, p. 224) that ‘ $2 + 3$ ’ does not share its sense with ‘ 5 ’. Peano is told in (XIV/11, p. 128) that ‘ $5 + 2$ ’ differs in sense from ‘ $4 + 3$ ’. And Russell is told once in (XV/14, p. 152) that the sense of ‘ $2^3 + 1$ ’ is different from that of ‘ 3^2 ’, and then reminded again, two years later in (XV/18, p. 163), that ‘ 7 ’ and ‘ $4 + 3$ ’ have different senses.

⁵See, for example, Salmon (1986, p. 57).

It is clear from what Frege says about the senses of these individual expressions that he could not have accepted the view of self-evidence as a priori truth. Since this view goes hand in hand with the conception of speakers as the kind of ideal intellects described by Ayer and Hahn, he could not have accepted that conception either. Instead, the ideas of self-evidence and cognitive value underlying Frege's notion of sense must have been based on a conception of speakers as something less than perfect intellects; and it is most natural to take them as creatures like ourselves, subject to intellectual limitations like our own.

Sense and structure

It is important to remember that natural languages, in Frege's eyes, were supposed to be terribly defective instruments: even apart from their vagueness and ambiguity, he believed that they were incapable of expressing thoughts in a manner suitable for a precise specification of the inferential relations among them (1879, p. 5–6). Because of the flaws he saw in natural languages, Frege concentrated throughout his life on the development of an alternative formalism that he hoped to be free from these defects; and one of his central aims in the design of such a language was the establishment of a close correspondence between expressions and their senses. Frege thought of senses, like expressions, as structured entities; and there is strong evidence that he wished to require of an ideal language, not only that each of its expressions should carry a unique and determinate sense, but that these expressions should correspond in structure to their senses.

This evidence is scattered throughout Frege's logical work, from the earliest period to its final stages. One of his original objections to natural languages as a vehicle for scientific investigation was the lack of any structural correspondence between expressions drawn from these languages and the concepts they are supposed to represent. This concern extended even to the level of lexical compounding, a matter that is still not well understood today:

There is only an imperfect correspondence between the way words are concatenated and the structure of the concepts. The words 'lifeboat' and 'deathbed' are similarly constructed though the logical relations of the constituents are different. So the latter isn't expressed at all, but is left to guesswork (1880/81, pp. 12–13).

Frege's explicit goal in his early work was to remedy this defect by designing a formalism

in which “content is rendered more exactly than is done by verbal language” (p. 12), a formalism that represents in its syntax those logical relations among conceptual constituents that are left to guesswork in natural languages; he motivates this goal by comparing it to Leibniz’s ideal of “a system of notation directly appropriate to objects themselves” (1879, p. 6; see also 1880/81, p. 9). Later, after drawing the distinction between sense and reference, he is able to describe his goal more exactly. The “objects themselves” to which the system of notation is supposed to be directly appropriate are senses, not referents; and one of the ways in which an expression is supposed to be “appropriate” to a sense is by reflecting its structure:

We can regard a sentence as a mapping of a thought: corresponding to the whole-part relation of a thought and its parts we have, by and large, the same relation for the sentence and its parts (1919, p. 255).

And this requirement of a correspondence in structure between expressions and senses is echoed repeatedly throughout Frege’s later work (1923, p. 36; 1914, p. 225; VIII/12, p. 79).

Nevertheless, there are some cases in which Frege seems to violate his own requirement of structural correspondence, insisting at times that certain structurally distinct expressions belonging to an ideal language must carry the same sense. For our purposes, the most important of these cases concerns expressions introduced through stipulative definition.⁶ Frege explains in (1893, pp. 44–45) how to introduce defined expressions into the language described in that work. Where ‘*P*’ is some expression belonging to the language already, it is possible to introduce a new simple expression—say, ‘*Q*’—using his double-stroke turnstile of definition, writing

$$||- P = Q.$$

In such a case, ‘*P*’ and ‘*Q*’ tend to differ in structure, since ‘*Q*’ must be simple but ‘*P*’ will generally be complex. Still, Frege claims that the two expressions should carry exactly the same sense:

We introduce a new name by means of a *definition* by stipulating that it is to have the same sense and the same referent as some name composed of signs that are familiar (1893, p. 82).

⁶Some of the other cases are described in Chapter 17 of Dummett (1981).

And of course, what Frege says here is intuitively very appealing: if ‘ Q ’, a meaningless symbol, is *stipulated* to mean exactly what ‘ P ’ does, then how could it possibly carry a distinct sense?

However, although the language of (1893) is supposed to be an ideal formal language (in fact, it is Frege’s canonical example), and it does contain the facility for introducing defined expressions, one might imagine all the same that this case should not count as a real counterexample to Frege’s requirement of structural correspondence between the expressions of an ideal language and their senses. One might imagine that, even though the language does happen to contain a facility for introducing defined expressions, this facility is, somehow, not an essential part of the language—that it is present only as a matter of notational convenience. This is a standard view of definitions, endorsed by a number of writers, such as Whitehead and Russell (1910, p. 11), for instance; and there is a good deal of evidence that the view should be attributed also to Frege, such as his introductory remark that:

The definitions [in this work] do not really create anything, and in my opinion may not do so; they merely introduce abbreviated notations (names), which could be dispensed with ... (1893, p. vi).

But for Frege, as I will try to show, this simple picture of definitions as nothing but notational abbreviations is badly misleading.

3 Definitions

3.1 Definition and concept formation

The best way to see the importance of definitions for Frege is to look at his work from a somewhat broader perspective, focusing on the notion of *analyticity* in his philosophy of mathematics. One of the first things we learn when we study Frege is that he hoped to show, contrary to Kant, that the truths of arithmetic are analytic. But what does this mean: what is it that Frege hoped to show, exactly, about the truths of arithmetic?

Kant had characterized the analytic truths as those subject-predicate statements in which the predicate concept belongs to the subject concept; and from this characterization, he was able to derive certain other traits of these statements. He concluded, for instance, that the

analytic truths are a priori—in our terminology, that they would be self-evident to an ideal intellect, such as the Ayer-Hahn monster. The reason is simple: since these statements are true solely in virtue of relations among concepts, they can be discovered through conceptual analysis alone, apart from experience. More important for our purposes, Kant seemed to conclude also that the analytically true statements would have to be self-evident even to more limited intellects, such as ourselves; even creatures like ourselves, just by understanding an analytic truth, bringing to mind the concepts on which it is based, would have to know that it is true. And again, the reason for this conclusion is easy to see: concepts, for Kant, were supposed to be such simple things that any statement whose truth could be discovered through conceptual analysis alone would have to be almost obvious.

But does any of this tell us how to understand *Frege's* claim that arithmetic is analytic? Does he mean that the truths of arithmetic are analytic in Kant's sense? Not exactly—for he denies that these truths must be statements whose subject concepts contain their predicate concepts: most, he says, are not even statements of subject-predicate form (1884, pp. 99–100). And he denies also that, for creatures like ourselves, they must be self-evident: “propositions which extend our knowledge can have analytic judgments for their content” (p. 104). Evidently, Frege had shifted away from Kant's account of analyticity, and his alternative characterization is not hard to find: early in (1884), he describes the analytic propositions as those that can be derived from general laws of logic and definitions alone (p. 4). In claiming that arithmetic is analytic, then, Frege does not mean that we think the predicate concept of an arithmetical truth whenever we think its subject concept, or that all of arithmetic is self-evident. He means that the truths of arithmetic can be derived in his formal system of logic (or in the right logic, anyway), supplemented only with definitions.

Now, this new criterion of analyticity does represent a real change from Kant's.⁷ It is important to see this change, but it is important, also, not to overestimate its significance. Certainly Frege thought of himself as working still with Kant's original notion; he says that his new characterization is intended “only to state accurately what earlier writers, Kant in particular, have meant” (p. 3). And there is a good deal of continuity between Kant's account of analyticity and Frege's. Put roughly, both view the analytic statements as those whose truth can be discovered entirely through conceptual analysis. Of course,

⁷As emphasized by Benacerraf (1981, p. 26).

this is obvious in the case of Kant. To see that it is true also for Frege, we must see that he treats definition—the introduction of a defined expression into a language—as a kind of concept formation, and deduction as an analogue to Kant’s process of rendering explicit what is contained in a concept.

Frege’s most extensive discussion of concept formation occurs in (1880/81), where he compares Boole’s logical notation and deductive machinery with his own. He argues here for the superiority of his own concept-script, in large part, because it allows for the formation of more complex and scientifically fruitful concepts than Boole’s: “[i]t is in a position to represent the formations of the concepts actually needed in science, in contrast to the relatively sterile multiplicative and additive combinations we find in Boole” (p. 46). Working within Boole’s language, we can form new concepts only by taking the logical sums, products, and complements of already existing concepts. Frege explains, for example, how to form the concept *homo* as the product of the already existing concepts *rationalis* and *animal* (p. 33). He illustrates the familiar way in which, if these two existing concepts are represented by regions in a plane, the new concept can be represented by their intersection; and then, alluding to his illustration, he describes the sterility of these Boolean techniques for the formation of new concepts:

In this sort of concept formation, one must assume as given a system of concepts, or speaking metaphorically, a network of lines. These really already contain the new concepts: all one has to do is to use the lines that are already there to demarcate complete surface areas in a new way. It is the fact that attention is primarily given to this sort of formation of new concepts from old ones, while other more fruitful ones are neglected, which surely is responsible for the impression one easily gets in logic that for all our to-ing and fro-ing we never really leave the same spot (1880/81, p. 34).

In Frege’s formalism also, new concepts are supposed to be constructed by definition out of old ones, but they are constructed by means of a richer set of definitional techniques. These techniques play an essential role in allowing him to define such fruitful concepts as that of a continuous function, for example (p. 24). And when we compare the concepts defined through these new techniques with those that can be constructed using Boolean methods alone, Frege writes, we find that in the new case:

there is no question of using the boundary lines of concepts we already have to form the boundaries of the new ones. Rather, totally new boundary lines are drawn by such definitions—and these are the scientifically fruitful ones (1880/81, p. 34).

Although this discussion from (1880/81) is concerned only with Boole, what Frege says there about the limitations inherent in the Boolean techniques of concept formation should apply also to Kant's; and in fact, his comparison in (1884) between Kant's view of concept formation and his own is closely parallel to his earlier treatment of Boole. Frege begins with the claim that Kant “seems to think of concepts as defined by giving a simple list of characteristics in no special order; but of all ways of forming concepts, that is one of the least fruitful” (1884, p. 100). By contrast, he writes:

If we look through the definitions given in the course of this book, we shall scarcely find one that is of this description. The same is true of the really fruitful definitions in mathematics, such as that of the continuity of a function. What we find in these is not a simple list of characteristics; every element in the definition is intimately, I might almost say organically, connected with the others (1884, p. 100).

And he goes on to describe the difference between Kant's view of definitions and his own through the same geometrical metaphor that occurs in his earlier treatment of Boole: if we think of concepts as figures on a plane, then Kant's techniques allow us to construct new concepts only by using the boundary lines of those already given in a new way, but with his own techniques, it is as if we could draw “boundary lines that were not previously given at all.”

At this point, however, Frege's discussion of Kant moves beyond his earlier discussion of Boole: he draws a new conclusion, and one that I want to emphasize. As we have seen, Kant was able to view the truths arrived at through conceptual analysis as self-evident, without cognitive value, because of his very simple picture of concepts. But Frege, in presenting us with a more complicated picture of concepts, also revises Kant's judgment of the truths that can be discovered through their analysis. Once we have defined such a complex, scientifically fruitful concept, he says:

What we shall be able to infer from it, cannot be inspected in advance; here, we are not simply taking out of the box again what we have just put into it (1884, pp. 100–101).

And he continues by drawing an explicit contrast between his own classification of what can be discovered through the analysis of concepts and Kant's:

The conclusions we draw from it extend our knowledge, and ought therefore, on Kant's view, to be regarded as synthetic; and yet they can be proved by purely logical means, and are thus analytic (1884, p. 101).

In order to describe the situation, Frege turns to a new metaphor: the conclusions we draw can still be thought of as “contained in the definitions, but as plants are contained in their seeds, not as beams are contained in a house.”

3.2 Fruitfulness and sense identity

I hope that what I have said here is enough to show why it is mistaken to think of definitions, for Frege, as nothing but notational conventions. They play a much more important role in his thought than that. Frege claims that arithmetic is analytic, characterizing the analytic truths as those that follow from general laws of logic and definitions alone. And it is in part his view of definitions that enables him both to see a connection between his characterization of analyticity and Kant's, and to break the connection Kant sees between analyticity and self-evidence.

This picture of definition as a means of building complex, valuable concepts suggests a certain requirement on a semantic theory of definition, which I will call the requirement of *fruitfulness*: an adequate account must allow for the possibility that some definitions, at least, are worthwhile—that the introduction of defined expressions can make new discoveries, new proofs, possible. We have seen the suggestion of this requirement in Frege's images of boundary lines, boxes, and seeds; and at times, he spells it out directly:

Definitions show their worth by proving fruitful. Those that could just as well be omitted and leave no link missing in the chain of our proofs should be rejected as completely worthless (1884, p. 81).

But even if definitions are not just notational conventions, even if they play some more important role, they are at least notational conventions, they play this role at least; and so it seems that a semantic theory of definition should be subject also to a requirement of *sense identity*: the theory should represent an expression containing defined symbols as identical in sense with the expression that results when each of these symbols is eliminated through definitional expansion. The remarks cited at the end of Section 2 provide evidence enough for attributing this requirement to Frege, and he is even more explicit about the matter in the following passage:

In fact it is not possible to prove something new from a definition alone that would be unprovable without it If the *definiens* occurs in a sentence and we replace it by the *definiendum*, this does not affect the thought at all. It is true we get a different sentence if we do this, but we do not get a different thought. Of course we need the definition if, in the proof of this thought, we want it to assume the form of the second sentence. But if the thought can be proved at all, it can also be proved in such a way that it assumes the form of the first sentence, and in that case we have no need of the definition (1914, p. 208).

Evidently, these two requirements of Frege's—fruitfulness and sense identity—stand in at least an apparent conflict. According to the fruitfulness requirement, definitions should make new proofs possible; but according to the requirement of sense identity, the introduction of a defined expressions cannot really make it possible to establish thoughts that would be unprovable otherwise. Now it may seem tempting to dismiss this conflict on the grounds that it results from taking together two requirements that are supposed to apply, separately, to stipulative and explicative (or analytic) definitions. Frege himself may not have been sharply aware of the distinction between these two kinds of definition in his earlier writings, including (1884); but it is plain in his later work. Once he had isolated the idea of explicative definition, Frege was driven by a problem much like the standard paradox of analysis to conclude that, although they might be fruitful, these definitions are not required to preserve sense.⁸ And just as explicative definitions can be fruitful, but are not required to preserve

⁸Frege's considered treatment of explicative definitions is found in (1914, pp. 209–211); an earlier, more confused discussion can be found in (1894, pp. 318–321).

sense, one might suppose that, although stipulative definitions do satisfy the requirement of sense identity, there is no reasonable way of seeing how they might be fruitful.⁹

Such a reading would be mistaken, however. Within Frege’s framework, there are, in fact, two different ways in which stipulative definitions might be thought of as fruitful. The first, which has been extensively explored by Michael Dummett (1981, pp. 336–342; 1984, pp. 221–223; 1989, pp. 6–7; 1991a, pp. 41–32) involves a very robust understanding of fruitfulness. Imagine that the predicate ‘*OP*’ (“odd prime”), for example, is introduced into an arithmetical language through the definition

$$\| - [\xi > 2 \wedge \forall y(y|\xi \supset (y = 1 \vee y = \xi))] = OP(\xi).$$

According to the strong interpretation, this definition is supposed to confer upon ‘*OP*’ its own sense, which is then to appear as a constituent in the sense of any sentence containing this new predicate. The thought expressed by the sentence ‘*OP*(17)’, for instance, might be represented as a pair containing the sense of ‘*OP*’ together with that of ‘17’. On this interpretation, definitions are fruitful by allowing us to express genuinely new senses. Unless the original language happened already to contain a predicate equivalent in sense to ‘*OP*’, this thought simply could not have been expressed prior to the introduction of this new predicate; the sense of

$$17 > 2 \wedge \forall y(y|17 \supset (y = 1 \vee y = 17)),$$

which results from ‘*OP*(17)’ by definitional reduction, is itself a more complicated structure, isomorphic to this more complicated sentence.

This strong interpretation of the fruitfulness requirement respects Frege’s idea that the sense of a compound expression should correspond in structure to the expression itself; but of course, the interpretation is irreconcilable with the requirement of sense identity between expressions containing defined symbols and the expressions that result when those symbols are eliminated. Although this way of understanding Frege’s fruitfulness requirement is important, and certainly accounts for much of what he says about fruitfulness in his earlier writings, including (1880/81) and (1884), it did not figure prominently in his work

⁹Joan Weiner, for example, argues (1984, pp. 66–68) that there can be no non-trivial requirement of fruitfulness for stipulative definitions (she calls them “mathematical” definitions).

after he explicitly introduced the notion of sense. I wish to concentrate instead, therefore, on a second way of understanding fruitfulness for stipulative definitions—a much weaker, psychological reading, according to which definitions are supposed to be fruitful only by aiding our thinking. This aspect of fruitfulness is often thought of as unimportant, even by Dummett (1991, p. 23; 1991a, pp. 33–34). And in a well-known paper, Paul Benacerraf also dismisses the psychological interpretation as follows:

Definitions are not *simply* conventions of abbreviation; for if they were, the requirement of fruitfulness cited above would make little sense. The fruitfulness would be a matter only of psychological heuristic and not something to which Frege would attach much importance (1981, p. 28).

Taken in its context, this remark of Benacerraf’s seems to suggest two things: first, that for Frege to care about it, fruitfulness would have to be more than a matter of psychological heuristic; and second, that if fruitfulness were merely psychological, there would be little theoretical difficulty reconciling it with the requirement of sense identity, viewing definitions simply as conventions of abbreviation. These are, in any case, the two theses I want to discuss in exploring the weak interpretation of fruitfulness (whether or not they do accurately capture the content of Benacerraf’s remark). I will argue that both are mistaken: the remainder of this section deals with the first of these theses, establishing Frege’s concern; the following section shows that even on the weak, psychological interpretation, there is still a conflict between fruitfulness and sense identity.

The evidence that Frege would be concerned with fruitfulness even if it were only a matter of psychological heuristic occurs in another passage from (1914), in which he discusses the point of stipulative definitions. Just after the very strong statement cited earlier of his sense identity requirement, according to which stipulative definitions are simply abbreviations, allowing us in principle neither to express nor to prove any new thoughts, he writes:

It appears from this that definition is, after, all, quite inessential. In fact considered from a logical point of view it stands out as something wholly inessential and dispensable (1914, p. 208).

But then he continues:

...I want to stress the following point. To be without logical significance is by no means to be without psychological significance (1914, p. 209).

And he goes on to describe in detail the psychological significance of stipulative definitions, imagining a situation in which, through a series of definitions, perhaps, we have introduced into some formal language a simple expression with a very complex sense. Frege takes as his example the expression ‘integral’—defined, say, in pure set theory, or pure course-of-values theory:

If we tried to call to mind everything appertaining to the sense of this word, we should make no headway. Our minds are simply not comprehensive enough. We often need to use a sign with which we associate a very complex sense. Such a sign seems, so to speak, a receptacle for the sense, so that we can carry it with us, while being always aware that we can open this receptacle should we have need of what it contains. It follows from this that a thought, as I understand the word, is in no way to be identified with a content of my consciousness. If therefore we need such signs—signs in which, as it were, 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 take it out again. So if from a logical point of view definitions are at bottom quite inessential, they are nevertheless of great importance for thinking as it actually takes place in human beings (1914, p. 209).

Frege seems to be concerned, both in this passage and elsewhere in (1914), chiefly with limitations on the complexity of thoughts capable even of being grasped, or entertained, by creatures like ourselves. The limitations are supposed to result from the fact that our minds are not sufficiently “comprehensive”—or as he says later, that we “simply do not have the mental capacity to hold before our minds a very complex logical structure so that it is equally clear to us in every detail.” Although this observation seems very reasonable, and is surely in accord with our everyday intuitions, it is not easy to see how it can be accommodated within Frege’s framework: since he tells us so little about the mechanism through which we are supposed to grasp thoughts, it is hard to locate any source at all for complexity constraints on the thoughts we can grasp. In fact, even in this manuscript, Frege presents his point about our mental limitations using metaphors of containment, suggesting

that these very complicated thoughts are somehow just too large to fit directly into our minds. And it may seem that these metaphors are particularly inappropriate, since it is so central to his philosophy that thoughts (and senses generally) are not supposed to be *in* our minds at all.

One of the most interesting things to notice about the passage from (1914, p. 209), however, is that it shows Frege exploring the consequences of a *representationalist* view of thinking—a view according to which we are supposed to grasp thoughts and other senses, not directly, but rather through the mediation of linguistic items with which they are associated. This kind of view is popular today, but it is not necessarily something that one would have expected to find in Frege. Still, strictly speaking, the idea is not even new with (1914). Frege had remarked on several earlier occasions, such as (1882, p. 84) and (1897, p. 143), simply that we do think in language; but these earlier remarks have the character of asides, and it is not until his later writings that he seems to have reflected on the consequences of this idea in any detail. His concern with the role of language in our grasp of senses is especially evident in (1918), where he says at one point that a thought become graspable by being “clothed in the perceptible garb of a sentence” (p. 61), and elsewhere apologizes to the reader for not being able to present him with a thought directly, but only “wrapped up in a perceptible linguistic form” (p. 66, n. 6). He returns to this point again at (1924/25, p. 269), writing that it is necessary for us human beings that any thought of which we are conscious should be “connected in our mind with some sentence or another.” And it seems that the idea is at work also in the cited passage from (1914, p. 209), where he speaks, for example, of linguistic expressions as “receptacles” for senses.

I believe that it is by taking seriously this representational idea of Frege’s—as he himself seems to have done in his later works—that we can best understand both the source of the constraints he recognized in (1914) on our ability to grasp and reason with complex senses, and also the way in which definitions are supposed to help us get around these constraints. Against the background of this representational picture, it is natural to interpret Frege’s metaphors of containment as applying, not exactly to thoughts themselves, but instead, to the linguistic items through which these thoughts are presented to us: it is these items that can reasonably be described as contained in our minds, subject to the limitations on our mental capacity. If we tried to grasp a very complicated thought by presenting ourselves with

a sentence that faithfully encoded each of its structural details, we might find that simply attending to this sentence would exhaust our limited mental capacity. One of the ways in which definitions are fruitful, then, is by enabling us to get around these limitations—giving us the means to present a thought to ourselves through an expression that conceals some of these details, and so allowing us both to grasp and, presumably, to establish thoughts that would have been inaccessible otherwise.

This interpretation of fruitfulness can be illustrated by elaborating Frege's own example of an integral. Let us suppose that along with some numerals and terms denoting certain functions, the conventional symbol for an integral has been properly introduced into a pure set theory, or a pure course-of-values theory. Now consider even a simple sentence containing these new expressions—a sentence of the sort that might appear in a high school text, such as ' $\int_0^3 x^2 dx = 9$ '. Since the defined expressions in this sentence have been properly introduced, they will be eliminable; and the result of eliminating them will be a sentence—' P ', say—from the base language, provable in principle just in case ' $\int_0^3 x^2 dx = 9$ ' is provable in the extended language. However, ' P ' will be a very complicated sentence, expressing a correspondingly complicated thought. And what Frege seems to believe, quite sensibly, is that creatures like ourselves, with our psychological limitations, would either be unable to think such a complex thought by presenting ourselves with the sentence ' P ', or even if we were able to do so, that merely attending to this sentence, which represents the thought in complete detail, would exhaust so much of our mental capacity that we would find it impossible to hold this thought in mind along with others, or to manipulate its representation in the ways that constitute reasoning. For most ordinary purposes, much of the structure contained in the thought that P will be irrelevant anyway; definitions allow us to pack away this irrelevant structure, representing to ourselves only the complexity called for by the purposes at hand, and so conserving our limited mental capacity.

Frege's emphasis on metaphors of mental capacity might suggest that the sheer size of a representation is the only issue, and that definitions aid our thinking only by allowing us to represent a complex thought to ourselves more compactly. But there is more to it than that. In general, definitions allow us, not only to suppress structure in our representation of thoughts, but also to modify that structure. The sentences ' $\int_0^3 x^2 dx = 9$ ' and ' P ', for example, will not share a common structure even at the highest level of syntactic analysis.

We cannot expect to find constituents of ‘ P ’ corresponding to the grammatical constituents of ‘ $\int_0^3 x^2 dx = 9$ ’. There will be no single, isolable constituent of ‘ P ’ corresponding to the occurrence in this sentence of the the function symbol ‘ f ’, for example, or even to the occurrence of ‘9’. When we present the thought that P to ourselves through the sentence ‘ $\int_0^3 x^2 dx = 9$ ’, then, we are representing this thought in a way that is not only more compact than its representation as ‘ P ’, but also essentially different—through a sentence with a different logical form, constructed from new constituents, in a new manner. And for creatures like ourselves, representing the same thought through sentences with essentially different structures may aid our thinking in a way that goes beyond the conservation of mental capacity; it may be that certain logical linkages among thoughts are more easily noticeable depending on the structure of the sentences through which they are represented.¹⁰ Of course, in the passage I have cited from (1914), Frege does not discuss *this* particular manner in which definitions can be helpful. He speaks instead only about conservation of mental capacity, and no wonder: the idea that thoughts might be represented through sentences with essentially different (as opposed to simply: less) structure clashes much more severely with his view that sentences from an ideal language should mirror in their structure the thoughts they express.

In any case, whether we are concerned only with the conservation of mental capacity, or also with the advantages to be gained through different structural representations, the example described here allows us to illustrate one important aspect of fruitfulness for stipulative definitions, one way in which the introduction of defined expressions might enable us to discover thoughts that we could not have discovered without them. It is perfectly reasonable to suppose that, as a matter of psychological fact, a creature with our cognitive limitations would never have been able to discover the thought that P working only in the base language of set theory (or course-of-values theory), where he could present this thought to himself only through the sentence ‘ P ’. It is at least in part because we are able to present this thought to ourselves through the sentence ‘ $\int_0^3 x^2 dx = 9$ ’ that we can discover its truth.

¹⁰This seems to be the kind of thing Wittgenstein had in mind when he argues, as in (1956, pp. 85–87), that definitions are primarily useful, not for providing us merely with “shortened” notation (p. 86), but for allowing us to discover what he calls “a new *aspect*” (p. 85).

4 Sense and psychological significance

Let us review. We saw in Section 2 that Frege was committed to the idea that sentential senses, or thoughts, should correlate with psychological states, and also that his notion of sense was itself closely bound up with a view of speakers as creatures subject to psychological limitations. In Section 3, we turned our attention to Frege's theory of definition, focusing on an apparent conflict between his two requirements of fruitfulness and sense identity; and there, in interpreting the notion of fruitfulness, we concentrated on developing a weak, psychological reading, according to which definitions are supposed to be fruitful simply by aiding the thinking of creatures like ourselves.

Now it may seem that this weak reading of fruitfulness is theoretically uninteresting, and in particular, that it is consistent with Frege's requirement of sense identity. I argue in the present section, however, that this is not so. Even on the weak interpretation, the idea of fruitfulness forces a conflict in Frege's thought—between the requirement of sense identity, and the correlation he sought to establish between senses and the psychological states of creatures with limited intellects. After describing this conflict, I explore one way in which Frege might have resolved it; and I compare that method of resolution with some similar themes in contemporary work.

4.1 The conflict

It is easiest to see the tension in Frege's thought by considering a particular example; so let us return to the situation described in the previous section. We supposed there that the language of pure set theory had been definitionally enriched with conventional symbols for the numbers and certain functions, including the definite integral. The enriched language contained the sentence ' $\int_0^3 x^2 dx = 9$ ', whose definitional reduction into the base language, we assumed, was the sentence ' P '. And we assumed that, as a matter of psychological fact, a creature with our cognitive limitations might be able to establish the sentence ' $\int_0^3 x^2 dx = 9$ ' in the enriched language, even though he would not have been able to establish the sentence ' P ' working in the base language alone.

Now imagine that Karl is a creature with our psychological limitations, and that working in the extended language, using the appropriate proof or calculational techniques introduced

along with the defined expressions, he does manage to establish the sentence ‘ $\int_0^3 x^2 dx = 9$ ’. Does Karl then believe that $\int_0^3 x^2 dx = 9$? Well, we can suppose that, as a result of having established this sentence, he is now willing to assent to it, even to place bets on its truth. In that case, it seems that our ordinary standards would force us to the conclusion that Karl believes that $\int_0^3 x^2 dx = 9$; and so, that the sentence ‘Karl believes that $\int_0^3 x^2 dx = 9$ ’ is true.

But must we conclude also that Karl believes that P ? Since the sentence ‘ P ’ follows by definitional reduction alone from one that Karl has already accepted, we can assume that he possesses a procedure for discovering its truth. But our ordinary standards allow a distinction between possessing a procedure for discovering the truth of some sentence and actually believing the proposition it expresses. Once Karl has learned elementary arithmetic, for example, he can be said to possess a procedure also for discovering the truth of ‘ $91 \times 79 = 7189$ ’, but our ordinary standards do not force us to conclude that he knows that $91 \times 79 = 7189$. The definitional reductions involved in arriving at ‘ P ’ from the sentence Karl has already accepted are complicated, and he simply may not have carried them out. Lacking any independent reason to accept this sentence, he might withhold his assent or refuse bets on its truth; and in that case, it seems most closely in accord with our ordinary standards to describe the situation as one in which Karl does not believe that P , or in which ‘Karl believes that P ’ is false.

If this story is consistent, our ordinary standards for judging such things seem to allow a situation in which ‘Karl believes that $\int_0^3 x^2 dx = 9$ ’ is true while ‘Karl believes that P ’ is false; Karl might be in a psychological state characterized as believing that $\int_0^3 x^2 dx = 9$ without also being in a state characterized as believing that P . Because of the correlation that Frege sought between senses and psychological states, we would have to conclude from this that the two sentences ‘ $\int_0^3 x^2 dx = 9$ ’ and ‘ P ’ possess distinct senses. But of course, this conclusion clashes with his requirement of sense identity for defined expressions, since the second of these sentences is a definitional reduction of the first. Even the weak, psychological aspect of fruitfulness leads to a conflict with Frege’s requirement of sense identity: it forces a clash between sense identity and the idea that senses should correlate with psychological states.¹¹

¹¹Readers familiar with Kripke (1979) will recognize in the preceding paragraphs an appeal to something like his principles of disquotations. If Karl assents to ‘ $\int_0^3 x^2 dx = 9$ ’, we can conclude that he believes that

In a way, this kind of conflict should really come as no surprise. To say that definitions are psychologically fruitful, that their fruitfulness is due to our psychological limitations, is simply to say that a creature without these limitations, such as the Ayer-Hahn monster, would not find the introduction of defined expressions helpful in the way that we do. When we present the thought that P to ourselves through the sentence ‘ $\int_0^3 x^2 dx = 9$ ’, we are—precisely because of our limitations—in a psychological state significantly different from that in which we present this thought to ourselves through the sentence ‘ P ’. More of our cognitive capacity remains free, and we are more likely to see the connections between this thought and certain others; different predictions are warranted about the sentences we will assent to, and about other aspects of our behavior as well. Because of this, there seems to be a legitimate distinction, for us, between the psychological state characterized as believing that $\int_0^3 x^2 dx = 9$ and the state characterized as believing that P .

Since there are no bounds on the cognitive capacity of the Ayer-Hahn monster, and he sees all logical connections at once anyway, he would be in exactly the same psychological state no matter how he is presented with the thought that P ; and so for such a creature, there is no point in drawing a distinction between the state of believing that $\int_0^3 x^2 dx = 9$ and the state of believing that P . But as we saw earlier, the Ayer-Hahn monster would be in the same state also, and for exactly the same reasons, whether he was presented with the sentence ‘ $2 + 3 = 5$ ’, the sentence ‘ $91 \times 79 = 7189$ ’, or even a statement of the four color theorem. It is just because of our psychological limitations that we find ourselves in different psychological states when we are presented with these different sentences; so it is just because of these limitations that these sentences must be assigned distinct senses. And likewise, if our limitations place us in different psychological states when we are presented with the sentences ‘ $\int_0^3 x^2 dx = 9$ ’ and ‘ P ’, then it would seem—as long as sentential senses

$\int_0^3 x^2 dx = 9$ using the simple principle of disquotation (p. 248); but to conclude that Karl does not believe that P from his refusal to assent to ‘ P ’ we must appeal to the strengthened principle (p. 249). In fact, if we are able to think of definitional reduction as a *translation* from a language containing defined expressions into its base fragment, then a conflict similar to that set out here could be derived from Kripke’s principles rather than Frege’s. Because Karl assents to the sentence ‘ $\int_0^3 x^2 dx = 9$ ’, simple disquotation tells us that ‘Karl believes that $\int_0^3 x^2 dx = 9$ ’ is true. From this, Kripke’s additional principle of translation (p. 250) would allow us to conclude that ‘Karl believes that P ’ is true as well; but we must also conclude from strengthened disquotation that ‘Karl believes that P ’ is false, since he will not assent to ‘ P ’.

are to correspond to our psychological states, not those of the Ayer-Hahn monster—that these two sentences should be assigned distinct senses as well.

4.2 Resolving the conflict

Through his treatment of indirect discourse, Frege was committed to the idea that sentential senses, or thoughts, should correlate with psychological states—so that sentences like ‘Karl believes that S_1 ’ and ‘Karl believes that S_2 ’, for example, should agree in truth value whenever ‘ S_1 ’ and ‘ S_2 ’ express the same sense. In addition, he explicitly endorsed a requirement of sense identity for stipulative definitions, according to which an expression containing defined symbols should share its sense with the expression that results when those symbols are eliminated. But if the state of believing that $\int_0^3 x^2 dx = 9$ is to be distinguished from the state of believing that P —if ‘Karl believes that $\int_0^3 x^2 dx = 9$ ’ can be true while ‘Karl believes that P ’ is false—then Frege cannot maintain both of these two things. Which should he give up?

In setting out the conflict between fruitfulness and sense identity, we hinted at the option of retaining the correlation between senses and psychological states while abandoning the requirement of sense identity, allowing a sentence containing defined symbols to express a different thought from the sentence that results when those symbols are eliminated. However, in some very interesting remarks, offered almost in passing, Frege suggests that he prefers the other route: insisting that a sentence containing defined symbols and the sentence resulting from the elimination of those symbols must express the same thought, but allowing that this thought might not entirely determine the psychological state of a speaker who entertains it.

The first of these remarks sets the stage for the passage that motivated our weak interpretation of fruitfulness:

When we examine what actually goes on in our mind when we are doing intellectual work, we find that it is by no means always the case that a thought is present to our consciousness which is clear in all its parts. For example, when we use the word ‘integral’, are we always conscious of everything appertaining to its sense? I believe that this is only very seldom the case. Usually just the word is present to our consciousness ... (1914, p. 209).

The second occurs within the cited passage, but I will repeat it here, for convenience. Right in the midst of his argument, in the course of explaining how definitions can be useful for thinking as it actually takes place in human beings, he says:

It follows from this that a thought, as I understand the word, is in no way to be identified with a content of my consciousness (1914, p. 209).

Now, most often when he contrasts thoughts with “contents of consciousness,” as at (1918, pp. 68–75), for example, Frege intends to make his familiar point that thoughts are objective, external to the mind. But the context shows that this cannot be what he means here, in the second of these remarks, since that observation is supposed to be justified by the preceding discussion (“It follows from this that . . .”), and he is not even talking about objectivity. He is talking, instead, about the usefulness of being able to present thoughts to ourselves in different ways, by means of more compact formulas; and his point, I think, is plain. Throughout this passage, Frege seems to characterize contents of consciousness in much the same way that we would characterize psychological states; it is the contents of our consciousness that help to explain the inferential connections we will recognize in mathematical research, and also, presumably, the sentences we will assent to. Evidently, the contents of our consciousness are supposed to be determined by what is “present to our consciousness,” and this, as Frege says in the first of these remarks, need not be a thought, but might be only a linguistic expression (“a word”). What Frege seems to mean, then, when he claims here that thoughts are not to be identified with contents of consciousness is simply that we can be in different psychological states while presenting the same thought to ourselves—as we are when we present the thought that P to ourselves, first, through the sentence ‘ P ’ itself, and then again, through the sentence ‘ $\int_0^3 x^2 dx = 9$ ’. Our psychological states are determined, not only by the thoughts we entertain, but also by the way in which we entertain them, the particular expressions we use to present these thoughts to ourselves.

As far as I can tell, this passage from (1914) is the only place in which Frege considers the idea that, because of defined expressions, sentential senses, or thoughts, might not correlate with psychological states; and of course, he does not really explore the idea in any detail here. If he had, I think that he would have been forced also to restructure some important components of his overall semantic theory. As we have seen, Frege’s correlation

between sentential senses and psychological states—his identification of sense with psychological significance—results from taking together his idea that the truth value of a sentence is determined by the referents of its parts, and his idea that the referent of an expression in indirect discourse is its sense. Now suppose Frege had explicitly admitted that ‘Karl believes that $\int_0^3 x^2 dx = 9$ ’ might be true while ‘Karl believes that P ’ is false, so that ‘ $\int_0^3 x^2 dx = 9$ ’ and ‘ P ’ must differ in psychological significance, even though they express the same sense. In that case, he would have been forced to conclude either that the truth value of a sentence is not determined compositionally by the referents of its parts, or more likely, that the referent of an expression in indirect discourse must be something other than its sense, such as its psychological significance. I think it is not inconceivable also that he might have come to regard a sentence like ‘ $(\int_0^3 x^2 dx = 9) = P$ ’ as informative, even though it is an identity between expressions carrying the same sense, because these expressions differ in psychological significance.

On such a picture, some of the functions assigned to sense would have to be allocated instead to a separate notion of psychological significance, much as Frege had divided his earlier notion of “conceptual content” in arriving at the distinction between senses and referents. It would be the psychological significance of a sentence that determines its informativeness, and also acts as its indirect referent at least in propositional attitude contexts. It would be the sense of an expression—still a more fine-grained aspect of meaning than its referent—that is preserved by definitional reduction, and perhaps also, that determines the ideal route to its referent. The requirement of a correspondence in structure between expressions and their senses would have to be dropped; but it is plausible to impose a similar requirement for psychological significance.

Rather than speculating further, however, on what the resulting theory—which Frege did not develop—would have been like if he had developed it, I now want to consider some analogies between the picture suggested here and two more recent developments in the philosophy of language.

Indexicals

Most contemporary arguments for distinguishing sense from psychological significance involve considerations rather far removed from Frege’s own concerns; but the arguments set

out by David Kaplan (1977) and John Perry (1977) rely only on indexicals, something that Frege did explicitly discuss in (1918). Since he did not himself separate sense from psychological significance, one would expect to find some tension in this discussion. Perry documents it. Kaplan provides a formal semantics in which these two aspects of meaning are provided with different model theoretic reifications: sense is represented as a function from worlds to extensions; psychological significance, as a function from contexts to senses.¹²

Both Kaplan and Perry argue that sense cannot be identified with psychological significance in languages containing indexicals: the sense of an expression is supposed to stand for a particular way of representing its referent; its psychological significance is supposed to stand for something more like a particular way of representing its sense. At times, however, both seem to suggest that this kind of break between sense and psychological significance can come about *only* through the presence of indexicality. Kaplan writes, for example:

So long as Frege confined his attention to indexical free expressions ...it is not surprising that he did not distinguish objects of thought (content) [sense] from cognitive significance (character) [psychological significance], for that is the realm of *fixed* character and thus ...there is a natural identification of character with content (1977, p. 64).

And Perry:

There will be no conflict [between sense and psychological significance], when one is dealing with eternal sentences The need for distinguishing [psychological significance] from [sense] will not be forced to our attention, so long as we concentrate on such cases (1977, p. 495).

I do not know how seriously to take these suggestions; but even if they are interpreted very broadly, so that “indexicality” is supposed to encompass all the ways in which environmental contingencies might affect sense, they are surely mistaken. The presence of a facility for introducing defined expressions forces a break between the sense of an expression and its

¹²The vocabulary in this area can be confusing. The aspect of a sentence’s meaning that I describe as its “sense” is characterized by Kaplan as its “content” and by Perry as the “thought” expressed by that sentence. The aspect of meaning that I describe as “psychological significance” is characterized by Kaplan as “character” and by Perry as “sense.”

psychological significance even in languages containing only indexical-free, eternal sentences; this is what Frege recognized in (1914).

Nevertheless, it would be a mistake to overemphasize this analogy between the kind of distinction between sense and psychological significance resulting from indexicals and that resulting from defined expressions. There are at least a couple of important differences. The first concerns structural features of the two phenomena. Indexicals allow a two way slippage between sense and psychological significance: not only can expressions distinct in psychological significance carry the same sense, but expressions identical in psychological significance can be used in different contexts to carry different senses. In the case of definitions however, the slippage seems to work only one way: the same sense can be carried by expressions distinct in psychological significance, but expressions identical in psychological significance must agree also in sense. The second difference concerns the overall character of the two phenomena, and the appropriate mode of analysis in each case. Kaplan's work shows that there is a plausible model theoretic explication of the distinction between sense and psychological significance that is introduced by indexicals. However, it is hard to see how the distinction introduced by defined expressions could be accommodated within a model theoretic framework, since this distinction seems to depend upon those syntactic features of expressions that model theory is designed to ignore.

Language learning

The second analogy is stronger. Towards the end of Section 3, I suggested that Frege had been tentatively exploring a representationalist picture of thinking, according to which we are supposed to grasp senses through the mediation of linguistic items. As mentioned there, this kind of view is popular today; one theorist who adopts it is Jerry Fodor. Although Fodor is best known for advocating a philosophy of mind that centers around the role of an internal language in our cognitive processes, his overall view is actually very close to that being ascribed to Frege.¹³ And it turns out that the theory developed by Fodor in his well-known (1975) runs into problems like those we have already seen in Frege's treatment

¹³This is clear from the discussion in (1978, pp. 200–202), where Fodor imagines that our relations to “propositions” much like Frege's senses might be mediated by items of our internal language; and the idea is spelled out in more detail in (1981, p. 260).

of defined expressions.

In order to see these problems, we must first review two of Fodor's central ideas.

The first is his treatment of the propositional attitudes, which are explicitly analyzed as involving relations to sentences. For each individual propositional attitude—believing, hoping, and so on—there is supposed to be some particular relation between speakers and sentences; the speaker is then thought to bear that attitude toward a proposition just in case he stands in the appropriate relation to some sentence expressing the proposition (1975, pp. 75–77; see also 1987, p. 17). As an example, suppose that R_{Bel} is the relation corresponding to the attitude of belief. The account then tells us that:

- (*) A speaker believes that S_1 if and only if there is some sentence ' S_2 ' such that
(i) the speaker stands in the relation R_{Bel} to ' S_2 ', and (ii) ' S_2 ' means that S_1 .

Of course, there is more to Fodor's overall treatment than this: he argues, for example, that the sentence ' S_2 ' in this schema is supposed to belong to the speaker's internal language, and also that the relation R_{Bel} is supposed to be a computational relation. But there is no need to consider these additional aspects of the theory, since they do not affect the phenomena that we are concerned with here.¹⁴

It is important to note, however, that this analysis of Fodor's forces a correlation, like Frege's, between psychological states and whatever aspect of meaning is supposed to figure in clause (ii) of (*). If ' U ' and ' V ' agree in this aspect of meaning, then the speaker will believe that U just in case he believes that V . For suppose the speaker believes that U . Then according to (*), there must be some sentence ' W ' such that (i) the speaker bears the relation R_{Bel} to ' W ', and (ii) ' W ' means that U . But if ' U ' and ' V ' agree in the relevant aspect of meaning, we must conclude from (ii) also that ' W means that V '; and so from (*) again, that the speaker believes that V .

The second of Fodor's ideas that we need to review is a particular picture of language learning, according to which a speaker is supposed to arrive at an understanding of a new predicate by learning some generalization that determines its extension (1975, p. 59). For

¹⁴In fact, when these additional aspects of the overall theory are left out of account, the simple analysis of propositional attitudes given above represents a core view shared also by several other writers, such as Field (1978).

Fodor, this involves mastering what he calls a “truth rule” for the new predicate. If the new predicate is ‘ F ’, for example, the appropriate truth rule would have the form

‘ Fx ’ is true if and only if Gx ,

where ‘ G ’ is some (possibly complex) predicate already understood by the speaker. It is necessary that ‘ G ’ should be already understood, since otherwise, the truth rule cannot tell the speaker what it needs to about the new predicate ‘ F ’.

From this picture of language learning, Fodor draws some notorious conclusions. He first observes that, because the speaker can learn new predicates only through truth rules linking them to predicates he already understands, he must already know a language “rich enough to express the extension” of any predicate he can learn; and the observation is then paraphrased as follows:

...one can learn what the semantic properties of a term are only if one already knows a language which contains terms having the same semantic properties (1975, p. 80).

Of course, this paraphrase involves a noticeable slide—from the idea that the original language must already be rich enough to express the extension of any new predicate, to the conclusion that it must already contain a predicate with the same “semantic properties.” It may appear that such a slide is unwarranted, since there is more to meaning than extensionality. However, although Fodor explicitly discusses only extensions in this passage, he states earlier that his conclusions are to hold also for intensional aspects of meaning (p. 60); and he reemphasizes this point in the present context, writing that “[i]ntentionalist theories lead to precisely the same conclusions as I have just drawn, and do so by precisely the same route” (p. 82, n. 19; see also p. 80, n. 18). So it seems that we are to take the displayed paraphrase literally, as applying to *all* semantic properties, intensional as well as extensional.

Now the picture at work here of the way in which we are supposed to extend our language through truth rules is very close to a definitional picture. Rather than explicitly specifying the truth conditions for sentences of the form ‘ Fx ’, the speaker could just as easily have introduced the predicate ‘ F ’ into his language through a stipulative definition. From this point of view, Fodor’s idea that new predicates must share their semantic properties with

those already present is similar to Frege's requirement of sense identity for stipulative definitions. And in fact, he goes on (pp. 82–84) to describe an apparent “paradox” that is much like the tension we have already seen in Frege's theory between sense identity and fruitfulness.

Fodor considers a situation in which a speaker learns a new language, or systematically extends a language he already knows. Of course, he is primarily interested in the special case of a speaker who originally knows only his internal language of thought, and is learning some natural language, but nothing depends on that. We could just as well suppose that the original language is a natural language and the speaker extends it by learning some specialized subject, such as chemistry or corporate law; or that the original language is set theory and the speaker extends it by introducing symbols appropriate for number theory and analysis. Fodor wants to say that there is “*some* sense” in which, as a result of extending the original language in this way, incorporating within it a new conceptual system, the speaker is able to think thoughts that he “could not otherwise entertain” (p. 84). But he notes that his own picture of language learning seems to rule out this possibility, since according to this picture, “nothing can be expressed” in the definitionally enriched language that cannot already be expressed in the original.

Like Frege, Fodor must find some way of explaining how the introduction of defined symbols could actually allow the speaker to entertain new thoughts, even though they do not allow for the expression of genuinely new propositions; and in fact, the solution he proposes is very much like the weak interpretation developed here for Frege's fruitfulness requirement. Although the enriched language can be no more expressive than the original in principle, for an ideal reasoner, Fodor argues that it might nevertheless allow speakers like ourselves to entertain new thoughts precisely because we are not ideal reasoners, because of our psychological limitations. This suggestion can be seen in the following passage, which again focuses on the special case of a speaker extending his language of thought to incorporate items from some natural language:

True for every predicate in the natural language it must be possible to express a coextensive predicate in the internal code. It does not follow that for every natural language predicate *that can be entertained* there is an *entertainable* predicate of the internal code If terms of the natural language can become incorporated into

the computational system by something like a process of abbreviatory definition, then it is quite conceivable that learning a natural language may increase the complexity of the thoughts that we can think (1975, p. 85).

Fodor’s explanation of the utility of defined expressions, like Frege’s, relies crucially upon the consideration of our psychological limitations, or “performance parameters” (p. 86). And just as we highlighted the role of these considerations in Frege’s account by emphasizing the worthlessness of definitions for speakers, without these limitations, such as the Ayer-Hahn monster, Fodor also reinforces his point by focusing on another kind of ideal reasoner:

If an angel is a device with infinite memory and omnipresent attention—a device for which the performance/competence distinction is vacuous—then, on my view, there’s no point in angels learning Latin; the conceptual system available to them by virtue of having done so can be no more powerful than the one they started out with (1975, p. 86).

So both the tension in Fodor’s account of language learning and his way of resolving it are similar to what we have already seen in Frege: for Fodor also, the introduction of new symbols into our language allows us to entertain new thoughts only in a weak, psychological sense, by enabling us to overcome some of our cognitive limitations. In the case of Frege, we saw that even this weak notion of fruitfulness forced a conflict between his requirement of sense identity for defined expressions and the idea that senses should correlate with psychological states. As we have seen, Fodor also is committed both to a correlation like Frege’s between some aspect of meaning and psychological states, and to a requirement, like Frege’s requirement of sense identity, that an expression containing defined symbols must share its semantic properties with the expression that results from definitional reduction. It should come as no surprise, therefore, that even the weak interpretation of fruitfulness should force a conflict in Fodor similar to that found in Frege.

Again, this conflict is seen most easily through an example, and so we return to the case of Karl, a creature with psychological limitations like our own. As before, we suppose that Karl has enriched his base language of set theory with numerals and symbols from analysis—allowing him to form, for example, the sentence ‘ $\int_0^3 x^2 dx = 9$ ’, which reduces to ‘ P ’ in his base language. And we suppose as before that, because of his psychological limitations,

a situation might arise in which our ordinary standards allow us to characterize Karl as believing that $\int_0^3 x^2 dx = 9$, but not as believing that P . From this, as we have seen, it follows that ‘ $\int_0^3 x^2 dx = 9$ ’ and ‘ P ’ must differ in *whatever* aspect of meaning is supposed to figure in clause (ii) of the analysis of propositional attitudes that Fodor sets out in (*). But of course, this conclusion clashes with the idea that ‘ $\int_0^3 x^2 dx = 9$ ’ must share *all* its semantic properties with ‘ P ’.

There are several options for modifying Fodor’s theory to avoid this conflict; but the one that seems most promising follows the line explored earlier with Frege. Fodor must weaken his claim that new expressions share all of their semantic properties with expressions from the original language. They may be required to share extensional properties, and even some properties commonly regarded as intensional—just as Frege requires defined symbols to share the senses of their defining expressions. But expressions containing defined symbols cannot be required to share all of their semantic properties with expressions already present; and in particular, they cannot be required to share those aspects of meaning that figure in Fodor’s own analysis of the propositional attitudes.

5 Conclusion

In contemporary work, the distinction between the proposition expressed by a sentence and its psychological significance is usually motivated by a familiar kind of counterfactual argument; and the discussion of these issues usually centers around the role of external factors in determining the meaning of our words. My primary goal in this paper has been to show that a similar, though not identical, distinction between two aspects of meaning can be developed entirely on the basis of considerations internal to language users—their cognitive limitations. To make this point, I have focused on symbols introduced through stipulative definitions. In a language containing such symbols, certain expressions and their definitional reductions will seem to differ in psychological significance for creatures with limited intellects, and so in any aspect of meaning that is supposed to correlate with psychological significance; but it seems also that there is some important aspect of meaning that they share.

I have argued that a distinction in meaning like this—between sense and psychological significance—should be drawn even in the kind of languages of most concern to Frege,

and that his failure to do so led to tensions in his thought. Of course, this observation only touches on the many issues involved in interpreting Frege's theory of definition more generally. I have not tried to describe here, for example, the ways in which the weak interpretation of fruitfulness might interact with the more robust interpretation mentioned earlier; I have only mentioned Frege's views on explicative definitions and the paradox of analysis, and failed even to mention either his treatment of contextual definition, or his peculiar objections to conditional definitions. I do want to emphasize, however, that the distinction drawn here is not simply a matter of Frege scholarship, but that it has some contemporary relevance as well. As we have seen, Frege's semantic goals often coincide with our own; and a number of contemporary writers are explicitly concerned, like Frege, to construct a semantic theory that is able to account for differences in meaning among logically equivalent expressions. Any such theorist should recognize a distinction like that drawn here between sense and psychological significance, and should avoid subjecting an account of one notion to constraints appropriate only for the other.

Acknowledgments

This paper is based on my doctoral thesis, completed in 1987 at the University of Pittsburgh. I am deeply grateful to my thesis advisor, Nuel Belnap, for the countless hours that he spent with me during this project, for his encouragement and critical attention, and for his friendship. I would also like to thank the other members of my thesis committee—Robert Brandom, John Haugeland, and Richmond Thomason—for their help and generosity at the time; and Georges Rey and Jamie Tappenden for some more recent comments and corrections.

References

Ayer, A. J.

- 1936 *Language, Truth and Logic*. Victor Gollantz, Ltd. (1936), 129 pp. Second edition reprinted by Dover (1952), 160 pp.; pagination refers to this edition

Benacerraf, P.

- 1981 Frege: the last logicist. In *Midwest Studies in Philosophy, Volume VI: The Foundations of Analytic Philosophy*, P. French, T. Uehling, and H. Wettstein (eds.), University of Minnesota Press (1981), pp. 17–35.

Bynum, T. W.

- 1972 *Conceptual Notation and Related Articles*. Oxford University Press (1972), xii+291 pp.

Dummett, M.

- 1973 *Frege: Philosophy of Language*. Harvard University Press (1973), xlix+697 pp. Second edition, 1981.
- 1981 *The Interpretation of Frege's Philosophy*. Harvard University Press (1981), xviii+621 pp.
- 1984 An unsuccessful dig. In *Frege: Tradition and Influence*, C. Wright (ed.), Basil Blackwell (1984), pp. 194–226.
- 1987 More about thoughts. *Notre Dame Journal of Formal Logic*, vol. 30 (1989), pp. 1–19.
- 1991 Frege and the paradox of analysis. In M. Dummett, *Frege and Other Philosophers*, Oxford University Press (1991), pp. 17–52.
- 1991a *Frege: Philosophy of Mathematics*. Harvard University Press (1991), xiii+331 pp.

Field, H.

- 1978 Mental representation. *Erkenntnis*, vol. 13 (1978), pp. 9–61.

Fodor, J.

- 1975 *The Language of Thought*. Harvard University Press (1975), x+214 pp.
- 1978 Propositional attitudes. *The Monist*, vol. 61 (1978). Reprinted in J. Fodor, *Representations: Philosophical Essays on the Foundations of Cognitive Science*, The MIT Press (1981), pp. 177–203; pagination refers to this printing.
- 1979 Three cheers for propositional attitudes. In *Sentence Processing: Psycholinguistic Studies presented to Merrill Garrett*, W. Cooper and E. Walker (eds.), Lawrence Erlbaum Associates (1979). Reprinted with extensive revisions in J. Fodor, *Representations: Philosophical Essays on the Foundations of Cognitive Science*, The MIT Press (1981), pp. 257–316; pagination refers to this printing.
- 1981 The present status of the innateness controversy . In *Representations: Philosophical Essays on the Foundations of Cognitive Science*, The MIT Press (1981), pp. 257–316.
- 1987 *Psychosemantics*. The MIT Press (1987), xiii+171 pp.

Frege, G.

- 1879 *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle (1879). Translated as *Begriffsschrift, a Formula Language, Modeled upon that of Arithmetic, for Pure Thought*, in *From Frege to Gödel: a Source Book in Mathematical Logic, 1879–1931*, J. van Heijenoort (ed.), Harvard University Press (1967), viii+664 pp; pagination refers to this translation.
- 1880/81 Booles rechnende Logik und die Begriffsschrift. Manuscript (between 1880 and 1881). Translated as “Boole’s logical calculus and the concept-script” in Frege (1979); pagination refers to this translation.
- 1882 Über die wissenschaftliche Berechtigung einer Begriffsschrift. *Zeitschrift für Philosophie und philosophische Kritik*, vol. 81 (1882), pp. 48–56. Translated as “On the scientific justification of a conceptual notation” in Bynum (1972); pagination refers to this translation.
- 1884 *Die Grundlagen der Arithmetik*. Breslau (1884). Translated as *The Foundations of Arithmetic*, Basil Blackwell (1950), xviii+119 pp.

- 1892 Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, vol. 100 (1892), pp. 25–50. Translated as “On sense and meaning” in Frege (1984).
- 1893 *Grundgesetze der Arithmetik*, Volume 1 Jena (1893). Partially translated in Frege (1964).
- 1894 [Review of E. Husserl’s *Philosophie der Arithmetik*]. *Zeitschrift für Philosophie und philosophische Kritik*, vol. 103 (1894), pp. 313–332. Translated in Frege (1984).
- 1897 Logik. Manuscript (1897). Translated as “Logic” in Frege (1979); pagination refers to this translation.
- 1914 Logik in der Mathematik. Manuscript (1914). Translated as “Logic in mathematics” in Frege (1979); pagination refers to this translation.
- 1918 Der Gedanke: eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus*, vol. 1 (1918), pp. 58–77. Translated as “Logical investigations I: thoughts” in Frege (1984).
- 1919 [Notes for Ludwig Darmstaedter]. Manuscript (1919). Translated in Frege (1979); pagination refers to this translation.
- 1923 Logische Untersuchungen, ditter Teil: Gedankengefüge. *Beiträge zur Philosophie des deutschen Idealismus*, vol. 3 (1923), pp. 36–51. Translated as “Logical investigations III: compound thoughts” in Frege (1984).
- 1924/25 Erkenntnisquellen der Mathematik und der mathematischen Naturwissenschaften. Manuscript (between 1924 and 1925). Translated as “Sources of knowledge of mathematics and the mathematical natural sciences” in Frege (1979); pagination refers to this translation.
- 1964 *The Basic Laws of Arithmetic*. M. Furth (ed.), University of California Press (1964), lxiii+144 pp.
- 1979 *Posthumous Writings*. H. Hermes, F. Kambartel, and F. Kaulbach (eds.), University of Chicago Press (1979), xiii+288 pp.

- 1980 *Philosophical and Mathematical Correspondence*. G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, and A. Veraart (eds.), University of Chicago Press (1980), xviii+214 pp.
- 1984 *Collected Papers on Mathematics, Logic, and Philosophy*. B. McGuinness (ed.), Basil Blackwell (1984), vi+412 pp.
- VIII/12 Letter to P. Jourdain (undated). Translated in Frege (1980); pagination refers to this translation.
- XIV/7 Letter to G. Peano (29 September 1896). Translated in Frege (1980); pagination refers to this translation.
- XIV/11 Letter to G. Peano (undated). Translated in Frege (1980); pagination refers to this translation.
- XV/14 Letter to B. Russell (28 December 1902). Translated in Frege (1980); pagination refers to this translation.
- XV/18 Letter to B. Russell (13 November 1904). Translated in Frege (1980); pagination refers to this translation.
- Hahn, H.
- 1933 *Logik, Mathematik und Naturerkennen*. Vienna (1933). Translated as “Logic, mathematics, and knowledge of nature” in *Logical Positivism*, A. Ayer (ed.), The Free Press (1959), viii+455 pp.; pagination refers to this translation.
- Kaplan, D.
- 1977 *Demonstratives: an essay on the semantics logic, metaphysics and epistemology of demonstratives and other indexicals*. Manuscript, Philosophy Department, University of California at Los Angeles (1977), 119 pp.
- Kripke, S.
- 1979 A puzzle about belief. In *Meaning and Use*, A. Margalit (ed.), D. Reidel Publishing Company (1979), pp. 239–283.

Perry, J.

- 1977 Frege on demonstratives. *The Philosophical Review*, vol. 86 (1977), pp. 474–497.

Putnam, H.

- 1975 The meaning of ‘meaning’. In *Language, Mind, and Knowledge*, K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, vol. VII, University of Minnesota Press (1975).

Salmon, N.

- 1986 *Frege’s Puzzle*. The MIT Press (1986), xi+194 pp.

Weiner, J.

- 1984 The philosopher behind the last logicist. In *Frege: Tradition and Influence*, C. Wright (ed.), Basil Blackwell (1984), pp. 57–79.

Whitehead, A., and Russell, B.

- 1910 *Principia Mathematica*, Volume 1. Cambridge University Press (1910).

Wittgenstein, L.

- 1956 *Bemerkungen über die Grundlagen der Mathematik*. Basil Blackwell (1956).
Translated as *Remarks on the Foundations of Mathematics*, The MIT Press (1967); pagination refers to this translation.