Notre Dame Journal of Formal Logic
Volume 28, Number 1, January 1987

Thoughts

DAVID BELL

In August 1919 Bertrand Russell wrote to the author of the *Tractatus* to ask, among other things, about the nature of thoughts and their constituent elements. Wittgenstein replied: "I don't know *what* the constituents of a thought are, but I know *that* it must have such constituents which correspond to the words of Language. Again the kind of relation of the constituents of thought and of the pictured fact is irrelevant. It would be a matter of psychology to find out" ([14], p. 72). Returning to the topic some thirty years later, Wittgenstein wrote, with more than a little irony:

> "Can one think without speaking?"—And what is *thinking*?—Well, don't you ever think? Can't you observe yourself and see what is going on? It should be quite simple. You do not have to wait for it as for an astronomical event and then perhaps make your observation in a hurry. ([15], p. 327)

Now these remarks, and many others like them scattered throughout Wittgenstein's writings, raise a number of very general and, I think, profoundly difficult problems; among them: the problem of how we are to account for the relation of thought to the language which expresses it; of how to account for the relation of thought to the reality it sometimes concerns; and, even more generally, of how to provide a genuinely *philosophical* theory of what can easily seem to be a phenomenon of primarily, if not exclusively, psychological concern. What can philosophy legitimately tell us about thoughts, that it would *not* be "a matter of psychology to find out"?

In what follows I shall try to set out some of the most general, but therewith most pressing, demands that an acceptable philosophical theory of thoughts must meet. I shall then try to establish the extent to which Frege's own theory successfully meets them. In a sense, however, there are not two tasks here but only one, for in specifying the general constraints within which a philosophical theory of thought should work, one is already merely recapitulating what is perhaps Frege's most lasting and revolutionary contribution to our understanding of the matter. In an important sense the very subject is Frege's—and our first task will be to get a bird's-eye view of it.

Received May 16, 1986

1 As a rough initial approximation we can, I think, distinguish four problematic areas associated with the notion of a thought. There are problems concerning *interiority*, *objectivity*, *reflexivity*, and *rationality*. As we shall see, thoughts become most refractory, least amenable to philosophical treatment, at precisely those points where the requirements dictated by considerations from within these different areas meet and overlap.

First, then, *interiority*. Whatever else thoughts might be, and regardless of the subsequent demands we might wish to make of them, one thing is surely clear: thoughts can be *had*. Thoughts can be grasped, apprehended, entertained, or understood; which at this point is to say no more than that, quite simply, we can think. Were this not so, then of course the entire topic would lose its interest and importance. Dummett has written:

> Thought differs from other things also said to be objects of the mind, for instance pains or mental images, in not being essentially private. . . . It is of the essence of thought that it is transferable, that I can convey to you exactly what I am thinking. . . . I do more than tell you what my thought is like—I communicate to you that very thought. Hence any attempt to investigate thoughts which culminates in a study of what is in essence private, that is, of inner mental experience, must have missed its mark. ([2], pp. 116–117)

There is doubtless a sense in which this is right, and I have chosen the ugly word "interiority" rather than, say, "subjectivity" to avoid any suggestion that a thought is some sort of private, mental entity. Nevertheless one can go too far in this direction. After all, thinking *is* a mental act; grasping a thought is, paradigmatically, an event that occurs in the arena of the mind—and a philosophical theory of thoughts must surely recognize this. And yet as a topic its absence from twentieth-century analytic philosophy has been conspicuous. We might attribute this absence to the influence, say, of Frege's anti-psychologism; of the behaviorism of Watson and Quine; of American pragmatism; of the apparent antipathy of the later Wittgenstein to what is "private"—but at all events contemporary philosophers have on the whole felt themselves unable or unentitled to inquire into the phenomena associated with interiority, and have approached such notions as thought, understanding, meaning, and judgment as though objectivity, reflexivity, and rationality were the only areas of legitimate philosophical concern. I shall return later to what I take to be a genuinely philosophical problem related to interiority.

In addition to such noncommital acts as entertaining or grasping a thought, acts which are without prejudice as to the truth or falsity of what is grasped or entertained, there are also those which, like asserting, denying, doubting, accepting, and judging, *do* involve some commitment to truth or falsity. Although it might seem plausible at this point to admit three basic kinds of propositional attitude—noncommittal (having a thought), positive (taking a thought to be true), and negative (taking a thought to be false)—Frege argues persuasively that in fact only two are necessary: "To make a judgment *is* to make a choice between opposite thoughts. Accepting one of them and rejecting the other is one act. So there is no need [in a perspicuous notation] for a special sign for the rejection of a thought. We only need a special sign for negation as such ([10], p. 185). Denial, then, does not stand contrasted with judgment; for to deny that

p is merely to judge that *not-p*. With Frege we can also demand that one and the same thought can now be grasped noncommitally, now judged to be true: "Making a judgment does not alter the thought that is recognized to be true" ([10], p. 251).

The second category of demands we might make on a theory of thoughts concerns their *objectivity*, and here three separate but interrelated families of issues need to be taken into account. First, as Dummett notes in the above quotation, thoughts are essentially intersubjective: two or more people can have one and the same thought. In other words, the identity conditions of thoughts, unlike those of, say, sensations or mental images, make no essential reference to the identity of the person who has them. The second requirement is that thoughts be expressible: they can be put into words. Now when I express a thought I do not, as it were, describe my thought from the outside; rather I *embody* my thought in language: the relation between a thought and its linguistic expression is not an external but an internal relation—a truth that both Frege and Wittgenstein attempted to capture by *identifying* a thought with the sense of the declarative sentence expressing it.

Under the general rubric of the "objectivity" of thoughts there is a further and vitally important family of issues: what we think about might exist, and what we think might be true. Frege himself typically expresses these requirements in terms of the conditions of the *existence* of thoughts: "In thinking", he writes, "we do not produce thoughts, we grasp them. For what I am calling thoughts stand in the closest connection with truth. What I acknowledge as true, I judge to be true quite apart from my acknowledging its truth or even thinking it. . . . Therefore that truth cannot have come to be only upon its discovery" ([9], p. 25). There is much that is highly contentious here (some of which I shall defend below), but at least part of Frege's point can be expressed in more plausible terms: thoughts possess truth conditions which are sometimes fulfilled, sometimes not, but those truth conditions typically make no essential reference to any person who has, or who might have, the thoughts in question. (The precise relation between this point which concerns truth conditions, and the analogous point about identity conditions made in the preceding paragraph has yet to be examined.)

The constraints which concern *reflexivity* can also be dealt with briefly. In the class of things about which we can think and judge there are of course thoughts and judgments. This apparently incontrovertible observation creates a host of intractable problems concerning, e.g., intensional entities, referential opacity, and the provision of an adequate semantics for expressions in *oratio obliqua*. Fortunately such problems are not of present concern. There is one point to note in passing, however, as it will be significant in what follows. Although reflexive thought is certainly possible, we must distinguish sharply between the function of thoughts as the *contents* of acts of thinking, and thoughts as the *objects* of such acts; between *having* a thought and *thinking about* a thought.[1] Clearly not all thinking can be reflexive.

Finally I come to those constraints on a theory of thought which have to do with *rationality*. Thoughts are gregarious things: they come in trains; they stand to one another in relations of compatibility, incompatibility, relevance and entailment; they get together to form arguments, and so on. In general, of

course, it is to their structure that we look for an explanation of such phenomena, and an important ingredient in any philosophical theory of thoughts will be the provision and justification of a procedure of analysis that will enable us to recognize such structures and to isolate the repeatable elements which go to make up our thoughts. We shall require, in other words, a procedure for carving up compound thoughts into their component thoughts, and atomic thoughts into their component concepts.

So far, then, I have suggested that it is at least *prima facie* reasonable to require of a philosophical theory of thought and judgment that it meet the following condition of adequacy:

(A) It should allow that thoughts can sometimes be

(1) had (in thinking) (2) taken to be true (in judging)	**Interiority**
(3) expressed in language (4) intersubjective (5) objectively true/false	**Objectivity**
(6) about thoughts (7) about judgments	**Reflexivity**
(8) complex, i.e., containing component thoughts (9) complex, i.e., containing component concepts (10) logically related one to another	**Rationality**

and preferably, of course, it should explain such possibilities.

There is here a second, overriding requirement, namely

(B) It should employ a univocal notion of thought throughout (1)–(10) above.

The thoughts we entertain are the very thoughts that we assert, that we express in language, that are true or false, that stand in logical relations with each other, and so on.

The third requirement I shall call *the Principle of Spontaneity*. The constraints it places on possible theories of thought are entirely formal: they amount to no more than an insistence that any such theory shall not generate a vicious regressive infinity of conditions on the performance of an act. The principle is this:

(C) If the performance of an act of type T is learned or rule governed, then it cannot be a general requirement of my performing an arbitrary act of type T that I have already performed an act of that type or, indeed, of any type that, in its turn, requires the performance of an act of type T.

So this principle outlaws any theory according to which, for example, criteria are always applied on the basis of criteria, or concepts are in general understood in terms of other concepts, or judgments can only be made once we have satisfied ourselves that they are warranted, and so on. Crass infringements of the principle are rare, needless to say, yet Wittgenstein for one was adept at uncover-

ing the incoherence of views whose contravention of the principle was very far from obvious. Thus the *Investigations* opens with a quotation from Augustine's *Confessions* which offers a naive but attractive account of how we learn to talk. Perhaps the major failing of this account is that "Augustine describes the learning of human language as if the child . . . already had a language, only not this one. Or again: as though the child could already *think*, only not yet speak. And 'think' here would mean something like 'talk to itself' " ([15], p. 32). Augustine's account thus clearly contravenes the Principle of Spontaneity by implying that mastery of a language is a prior condition of aquiring mastery of a language. The principle may seem uncontentious, even trivial. Indeed, I think it is—though it is also, I believe, a principle whose consistent application yields a number of striking results; for it underlies and justifies many of Wittgenstein's conclusions about ostensive definition, and about what it is to be able to follow a rule. It is, for example, an instance of the Principle of Spontaneity which forces the conclusion that "When I obey a rule I do not choose, I obey the rule *blindly*".[2]

2 Frege's theory of thoughts is ambitious. Of the ten requirements listed under (A) above, it is intended to meet eight maximally, that is, to comprise an *explanation* of the phenomena involved. And it is intended to meet the remaining two demands minimally, in that the theory is claimed to be at least compatible with their possibility, even though in the last analysis it offers no account of them.

Given Frege's vehement antipathy to all attempts to introduce psychological considerations into logic, it is hardly surprising that the two phenomena for which he provides no theoretical account are those constitutive of what I've called "interiority". One of the specific aims of the late essays, and in particular of "Thoughts", however, was precisely to argue that insuperable difficulties for his theory are not to be expected from this quarter. Even if thoughts are, as he believed, abstract (and hence nonmental) entities, still they *can* be grasped: "This is a process in the inner world of a thinker which may have further consequences in this inner world, and which may also encroach on the sphere of the will and make itself noticeable in the outer world as well" ([9], pp. 28–29). Into the precise nature of such events, processes, and states, however, Frege like Wittgenstein thinks it no part of his brief to enquire: "Both grasping a thought and making a judgment are acts of a knowing subject, and are to be assigned to psychology" ([10], p. 253).

Frege's theory of thought is not only ambitious in scope, it is quite remarkably elegant in structure. The keystone of the edifice is the identification of a thought with the sense of a declarative sentence. Indeed Frege goes further, and identifies the *structure* of the thought with the structure of the sentence which expresses it, on the grounds that "even a thought grasped (. . .) for the very first time can be put into a form of words which will be understood by someone to whom the thought is entirely new. This would be impossible, were we unable to distinguish parts in the thought corresponding to the parts of the sentence, so that the sentence serves as a model of the structure of the thought" ([7], p. 55). This principle is then supported, on the one side, by a number of theses concerning sentence structure, and on the other side, by a number of theses concerning the nature of senses. Thus the essential elements of Frege's theory of thought can, I believe, be captured in terms of the following six theses:

Thesis 1 *A thought is isomorphic with the sentence whose sense it is.*

Thesis 2 *Only in the context of a sentence does a word mean something.*

Thesis 3 *Every unambiguous sentence has a unique function-argument analysis.*

Thesis 4 *The sense of an expression is the condition which anything must meet if it is the reference of that expression.*

Thesis 5 *The reference of an expression is that in virtue of whose identity expressions can be intersubstituted for that expression,* salva veritate, *throughout any context of the appropriate kind.*

Thesis 6 *Assertoric force is to be sharply distinguished from predicative power.*

Taken together (and duly expanded) these six doctrines yield a theory which in essence meets all the desiderata that we have so far laid down—or which at least appears to do so.

Although, as I mentioned above, Frege was generally speaking unconcerned with problems arising in connection with "interiority", he nevertheless saw clearly that Theses 1 and 6 together provide a powerful account of the difference between merely grasping or entertaining a thought, and judging or asserting its truth. The model for the difference between these mental acts is the distinction between an assertoric and an unassertoric occurrence of a sentence; and crucial to this distinction is the difference between the role performed by a predicate within a sentence, and any force possessed by the sentence as a whole. For if assertoric force belongs to some subsentential element (whether predicate, copula, or indeed any other), then a sentence containing that element will *ipso facto* possess that force. And yet there is simply no element present within an assertoric sentence "P", which is absent from that sentence when it occurs unasserted, say, in "$P \supset Q$". Both occurrences of "P" must possess the same sense, and the same truth-value, otherwise of course modus ponens fails. Frege concluded that a thought linguistically presented contains no assertoric element or component, and possesses a truth-value independently of its making an assertion. Moreover one and the same thought can be expressed by indefinitely many different sentence tokens. The need for and the justification of these doctrines and distinctions as they concern thoughts (senses) linguistically presented are then extended by Frege to thoughts as they are mentally presented in acts of thinking and judging: one and the same thought can be both grasped, nonassertively, and taken to be true in an act of judgment. Moreover, one and the same thought can be the content of indefinitely many different mental act tokens. As these mental acts can be performed by indefinitely many different thinkers, we have here one of the ingredients of Frege's account of the objectivity of thoughts, i.e., of their intersubjectivity.

In fact we earlier distinguished three aspects to the problem of the objectivity of thoughts, under the respective heads of "expressibility", "intersubjectivity", and "truth". Now Thesis 1 solves at a stroke problems to do with expressibility, i.e., with the relation in which a thought stands to a sentence such

that the latter can embody and communicate the former; for a thought just *is* the sense of a declarative sentence. As yet, however, this is little more than a slogan. It needs to be enriched by a fuller account of sentential sense.

Thesis 2 determines that our theory of sense will indeed be, *ab initio*, a theory of sentential sense; and Thesis 4 tells us that this theory will be articulated by appeal to the conditions on being the reference of a sentence. Thesis 5, in turn, tells us exactly what those conditions are. In normal contexts of direct use (I ignore for the moment contexts involving modality, direct or indirect quotation), a sentence is intersubstitutable *salva veritate* with any other which is materially equivalent to it. According to Thesis 5, therefore, its truth-value is the reference of such a sentence; and Thesis 4 now yields the result that the sense of a sentence is the condition on its possessing the truth-value it does possess—its "truth condition" for short. Frege famously expressed it thus: "Every such name of a truth-value *expresses* a sense, a *thought*. Namely, by our stipulations it is determined under what conditions the name denotes the True. The sense of this name—the *thought*—is the thought that these conditions are fulfilled" ([5], pp. 89–90). But as its sense is what we understand when we know what a sentence means, we can say more elegantly with Wittgenstein: "To understand a sentence means to know what is the case if it is true" ([13], 4.024).

We are now in a position to offer a fuller explanation both of the intersubjective accessibility of thoughts, and of the possession by them of an objective truth-value. Indeed, surprisingly, within a Fregean framework it turns out that these two issues are in fact one and the same.

The possibility of the intersubjectivity of thoughts is guaranteed, Frege believed, by the nature of their *identity* conditions: "A thought does not have to be owned by anyone. The same thought can be grasped by several people" ([10], p. 251). For a thought to be intersubjective, in other words, it is necessary that its identity condition contain no specification of the identity of the person who has it. As we noted earlier, this is perhaps the most crucial difference between thoughts and, say, sensations or mental images. On the other hand, a thought is objectively true (or false) insofar as it possesses a *truth* condition which can obtain (or not) independently of the existence of, or any subjective state of, the thinker. Now intuitively these two matters seem quite distinct, and yet time and again Frege appears to run them together: "In order to be true, thoughts . . . not only do not need to be recognized by us as true: they do not have to have been thought by us at all. A law of nature is not invented by us, but discovered, just as a desolate island in the Arctic Ocean was there long before anyone set eyes on it" ([10], p. 133). ". . . the thought expressed by the Pythagorean theorem is timelessly true, independently of whether anyone takes it to be true. It needs no owner" ([7], p. 17). "In thinking we do not produce thoughts, we grasp them. For what I am calling thoughts stand in the closest connection with truth. What I acknowledge as true, I judge to be true quite apart from my acknowledging its truth . . . therefore that truth cannot have come to be only upon its discovery" ([7], p. 25). What, one might ask, has the independent existence of thoughts got to do with their possessing a content that can be true or false? Surely my thinking that snow is white is a mental act which, as such, depends for its *existence* upon me, the person who performs it; and yet this seems to be a distinct issue from questions concerning the *content* of that

act, and in particular as to whether the content of that act is capable of being objectively true or false. It seems to me, however, that Frege is quite consistent on this point. Because a thought is an abstract object, because it has no material properties, the condition on its existence is the same as its condition of identity, which in turn just *is* a truth-condition. Quite simply, there is nothing more to a Fregean thought than its content. But as this content is exhaustively specified in terms of a truth condition (which makes no essential reference to an owner or bearer), this in turn comprises its identity condition. For Frege, the possibility of intersubjective accessibility and possession of objective truth conditions are one and the same.

I turn now to the way in which Frege's theory of thought satisfies the requirements listed above under the heading of "reflexivity". The outline of his theory is familiar, so I shall be brief. Theses 4 and 5 tell us that the sense of an expression is to be explained by appeal to the condition which its reference must satisfy, and that an expression's reference is to be explained in terms of its intersubstitutability *salva veritate.* Now typically we think about a thought (as distinct from straightforwardly having it) either by predicating something of it ("the thought that p is an interesting one"), or by ascribing it to someone ("so-and-so thinks that p"). Thesis 5 determines that the reference of an expression in *oratio obliqua* is its normal sense, for it is identity of sense that guarantees intersubstitutability *salva veritate* in such contexts. But the sense of a sentence is a thought, so the reference of a sentence in an oblique context is the thought which that sentence would normally express. The sense of such a sentence is, via Thesis 4, to be elucidated in terms of the identity condition of its normal sense. But because of the fact that, for thoughts, identity conditions and truth conditions collapse into one another, it follows that a sentence in *oratio obliqua* has the same sense as that sentence in a context of direct use. Frege's theory therefore captures and explains two important intuitions: first, that the reference of an expression in an oblique context is different from its reference in a normal context (the rules for substitutivity change); but, secondly, that if one already has a grasp of the normal sense of some expression, E, then one requires nothing further to understand occurrences of E within oblique contexts.[3] The word "cats" has exactly the same sense in the sentence "cats purr" as it does in the sentence "John thinks cats purr"; otherwise of course the first sentence would not express *what* John thinks.

Finally I come to Frege's contributions to our understanding of the problems mentioned above concerning "rationality". Thesis 3 is crucial in this respect, for Frege analyzes both molecular thoughts into their component thoughts, and atomic thoughts into their component concepts[4] in terms of the relation between a function and its arguments. The function-argument model distinguishes between subsentential functions ($(\xi)^2$; the capital of (ξ)), sentential functions ((ξ) is red; (ξ) = 2), and truth-functions (If (ξ) then (ζ)), between one- and more-placed functions, and between first- and higher-level functions. The one model thus provides a unified account of the logical form and semantic role of singular terms, predicates, relations, logical connectives, and quantifiers. In a sense this is the engine which drives virtually all of Frege's philosophy; but as it defies brief summary I shall here assume that the details of the theory are known.

In summary, then, Frege's six theses are intended to introduce a univocal notion of thought that will satisfy requirements (3)–(10) maximally, and requirements (1) and (2) minimally. The theory claims, in other words, to explain how thoughts are related to reality, to language, and to one another. The marriage of such explanatory power to so simple and elegant a theory would indeed have been a magnificent achievement. I want now to examine, however, a number of points at which the theory breaks down.

3 Theses 1 and 3 together imply that every (unambiguous) thought has a determinate function-argument structure, and that this corresponds to the function-argument structure of the sentence whose sense that thought is. This doctrine is consistent, however, only as long as sentences with different structures cannot express one and the same thought. Unfortunately, as Frege himself repeatedly acknowledged, there are good reasons for believing that the latter is in fact very far from impossible. "Let us never forget", he wrote, "that two different sentences can express the same thought ([10], p. 143).

There are three interpretations of this last remark under which it is true: two of them are unproblematic for Frege's theory of thoughts; the third, however, is an acute embarrassment. First, then, the remark is true in the weakest possible sense, i.e., insofar as different sentence tokens of the same type can express the same thought. As we noted earlier, this claim is so weak as to be unobjectionable, and we have already incorporated it into our list of plausible demands. Secondly, two sentences of different types can express the same thought if, say, both have the same deep structure or fundamental logical form, so that their differences result merely from some superficial transformation. The example which Frege typically cites in this context is the difference between a sentence in the active voice on the one hand, and its corresponding passive form on the other. The two sentences

(1) Brutus killed Caesar

and

(2) Caesar was killed by Brutus

can be allowed to express the same thought for they have identical truth conditions, and would indeed receive the same translation into a logically perspicuous notation. Frege contends, surely rightly, that the differences between sentences such as (1) and (2) are insufficient to threaten his theory: ". . . in speech the same thought can be expressed in different ways, by making now this proper name, now that one, the grammatical subject. No doubt we shall say that these differing phrasings are not equivalent. This is true. But we must not forget that language does not simply express thoughts; it also imparts a certain tone or colouring to them. And this can be different even where the thought expressed is the same" ([10], p. 193).

Before dealing with the third kind of case in which different sentences can express one and the same thought, we can usefully remove a potential source of confusion. Frege writes:

> If several proper names occur in a sentence, the corresponding thought can be analyzed into a complete and an unsaturated part in different ways. The sense of each of these proper names can be set up as the complete part over against the rest of the thought as the unsaturated part. ([10], p. 192)

The possibility Frege seems to be invoking here is that *one and the same sentence* may possess more than one function-argument structure; and if this is so, then it clearly contradicts Thesis 3. The sort of case Frege envisages is one in which a sentence such as (1) can be analyzed into a complete component "Brutus", and an unsaturated function-name "(ξ) killed Caesar"; but which can equally be analyzed into a complete component "Caesar", along with the function-name "Brutus killed (ξ)". In fact, however, there is no inconsistency here, for both these analyses are merely *partial.* Thesis 3 should be taken to claim no more than that, in Wittgenstein's words, "A proposition has one and only one *complete* analysis" ([13], 3.25; my emphasis). So although a given sentence may be decomposed into function and argument in a number of ways, in the last analysis it nevertheless possesses a determinate function-argument structure. In the example above the sentence consists determinately of two complete signs or proper names, "Brutus" and "Caesar", and a two-place, first-level function-name: "(ξ) killed (ζ)".[5]

I come now, thirdly, to the sense in which the consistency of Frege's theory of thought is threatened by the possibility of different sentences expressing one and the same thought—namely, the case in which the sentences in question have radically different function-argument structures. This is embarrassing not only because, in general, there are apparently sound intuitive reasons for allowing this possibility, but also, more specifically, because a number of Frege's own procedures and doctrines depend crucially upon it. I shall give just three examples. In the *Grundlagen* Frege writes:

> The judgment "line a is parallel to line b, or using symbols: $a//b$, can be taken as an identity. If we do this, we obtain the concept of direction, and say: "the direction of line a is identical with the direction of line b". Thus we replace the symbol // by the more generic symbol =, through removing what is specific in the content of the former and dividing it between a and b. We carve up the content in a way different from the original way, and this yields us a new concept. ([6], pp. 74–75)

Evidently, then, Frege at this time saw the necessity of allowing that

(3) Line a is parallel to line b

and

(4) The direction of a = the direction of b

are sentences with the same conceptual content. And this doctrine was not given up with the introduction of the distinction between sense and reference; for in *Grundgesetze* ([5], p. 36) the notion of a *Wertverlauf* is likewise introduced via the move from

(5) The functions $\phi(\xi)$ and $\psi(\xi)$ have always the same value for the same argument

or

(5′) $\overset{\mathfrak{a}}{\smile}\phi(\mathfrak{a}) = \psi(\mathfrak{a})$

to

(6) The function $\phi(\xi)$ has the same *Wertverlauf* as the function $\psi(\xi)$

or

(6′) $\grave{\epsilon}\phi(\epsilon) = \grave{\alpha}\psi(\alpha)$.

And, one last example, according to Frege "affirmation of existence is in fact nothing but denial of the number nought" ([6], p. 65). In other words,

(5) There exist unicorns

and

(6) The number of unicorns is not zero

express the same thought.

It would appear, then, that there is a fundamental inconsistency in Frege's theory of thought: it entails that certain thoughts both do and do not possess a determinate structure, a unique function-argument analysis. Frege might perhaps console himself (as he did when confronted by another contradiction) with the thought *solatium miseris, socios habuisse malorum*: for indeed the problem *is* a quite general one, and not one which results merely from idiosyncratic features of Frege's own theory. It results, that is, from our wanting to hold two equally plausible but apparently incompatible principles, namely: (i) that we can distinguish parts in the thought corresponding to the parts of the sentence expressing it, so that the sentence serves as a model of the thought, and (ii), that one and the same thought can be expressed by quite different sentences. Something, it seems, must be given up here. As I believe that both (i) and (ii) have sufficient intuitive plausibility to warrant their being saved, my own preference is to discard requirement B, the requirement that there be a single, univocal notion of thought capable of simultaneously satisfying each of the conditions we have introduced.[6]

I turn now to the second incoherence which is generated, I believe, by a Fregean theory of thought. It involves the incompatiblity of, on the one hand, a condition on the *objectivity* of thoughts with, on the other hand, a condition on the possibility of *interiority*, specifically with condition (C) above, which I have called the Principle of Spontaneity. In its fully Fregean form the incoherence arises as follows. Thinking is grasping or apprehending a thought, and a thought is an object. Although Frege himself nowhere explicitly asserts that thoughts are objects, this follows immediately from his identification of a thought as the reference of a singular term of the form: "The sense of the sentence '*S*' " ([4], p. 59). And yet, of course, a thought is also the *sense* of a name of an object, i.e., the sense of a sentence whose reference (if it has one) is a truth-value. It is about these two roles that Frege advances incompatible theses; for it is impossible for an *object* to fulfill the role of a *sense* ([4], p. 64). According to Frege, to have an object in mind *is* to have grasped the sense of some expression which has that object as its reference. But a vicious infinity of such

acts of grasping is generated immediately if we maintain, with Frege, that this sense is, in its turn, merely an object we have in mind; for in this case the sense would likewise have to be grasped *via* the sense of some some expression, which in its turn, as an object, would have to be grasped *via* the sense of some expression . . . and so on. Clearly this theory contravenes the formal constraints expressed in the Principle of Spontaneity: such thoughts would be unthinkable. In fact the incoherence here is one we guarded against earlier (p. 3): it is the incoherence of construing a thought (albeit tacitly) as what we think *of*, that is, as the object rather than as the content of an act of thinking.

Now it might be thought that there is little of contemporary interest here, insofar as this difficulty seems to arise from a number of peculiarly Fregean (and perhaps independently objectionable) theses, such as, e.g., that sentences are names of objects called truth-values, or that senses are themselves abstract objects. It might seem, then, that the problem can be circumvented by our simply abandoning such theses as these. This appearance is, I believe, misleading, for the difficulty is by no means peculiarly Fregean but is, rather, one that threatens any attempt to reconcile: (i) what is required if thoughts are to be objective, with (ii) what is required if interiority is to be possible.

If objectivity of thought is to be possible, then thought must conform to what Kant called "the universal condition of rules ([11], p. 179 ($A135 = B174$)), that is, its content must be determined solely by its conformity to those rules and conditions which are, precisely, constitutive of objectivity. On the other hand, however, if interiority is to be possible, then thinking must be spontaneous, in the sense that it must conform to the purely formal Principle of Spontaneity. In and of themselves these conditions are, I believe, entirely uncontentious; but when we try to amalgamate them, to apply them simultaneously, we are presented with a situation that appears paradoxical. We seem to have to take seriously, in other words, the claim that thinking must manifest a *rule-determined spontaneity*—and that looks like an oxymoron.

The strangeness of this situation, the tension which exists between the demand for spontaneity on the one hand, and the requirement that thought be rule-governed on the other, was discovered, and precisely characterized, by Kant:

> If understanding in general is to be viewed as the faculty of rules, judgment will be the faculty of subsuming under rules; that is, of distinguishing whether something does or does not stand under a given rule. . . . General logic contains, and can contain, no rules for judgment. . . . If it sought to give general instructions how we are to subsume under these rules, that is, to distinguish whether something does or does not come under them, that could only be by means of another rule. This in its turn, for the very reason that it is a rule, again demands guidance from judgment. And thus it appears that, though understanding is capable of being instructed, and of being equipped with rules, judgment is a peculiar talent which can be practiced only, and cannot be taught. ([11], p. 177 ($A132 = B171$))

Now the impossibility of there being, in general, rules for the following of rules has, I believe, profound and far-reaching consequences for our understanding of such notions as thought, judgment, understanding, meaning, objectivity, and the like. But in the present, Fregean context I shall examine only one such

consequence: it concerns the form which the analysis of acts of thinking and judging should take and, in particular, how we should construe the notion of *an object of judgment*. This is what Frege says:

> We are probably best in accord with ordinary usage if we take a judgment to be an act of judging, as a leap is an act of leaping. . . . Judging, we may say, is acknowledging the truth of something; what is acknowledged to be true can only be a thought. . . . If a judgment is an act, it happens at a certain time and thereafter belongs to the past. With an act there also belongs an agent, and we do not know the act completely if we do not know the agent. ([8], p. 42n)

The passage is interesting for it implicitly contains two incompatible models of thinking. If we take the analogy with leaping seriously, then the model will clearly be intransitive: just as there are no "objects of leaping", so there will be no objects of thought. When I leap I do not enter into a relation with an independently existing object called "a leap", and analogously, when I entertain a thought or pass judgment, I do not enter into a relation with something that exists independently of my acts of thinking and judging, something called "a thought". In this case the appropriate schema for the analysis of judgment will be:

$$agent + act.$$

On the other hand, however, Frege here seems tacitly to assume the applicability of a quite different model according to which we should take the notion of an *object of thought* quite literally, and adopt the schema:

$$agent + act + object.$$

Although in the *Begriffsschrift* Frege adopted the former, intransitive model, he eventually came to believe that objectivity could only be protected by adoption of the latter model, which accordingly came to dominate his account of thinking. He concluded that "when [a person] grasps or thinks a thought he does not create it but only comes to stand in a certain relation to what already existed" ([9], p. 18n). But it is precisely this model which is untenable, for it is this model which is incompatible with the possibility of interiority, with the possibility, that is, of *thinking*. If a thought were an object, then thinking would be, as Frege insisted, an act that refers to or picks out such an object: "Although the thought does not belong with the contents of the thinker's consciousness, there must be something in his consciousness that is aimed at the thought" ([9], p. 26). But this imposes structurally the same vicious regressive infinity of conditions on successful thinking as the directly analogous account of sentential sense imposes on the successful *expression* of thought; for if we claim that a sentence always refers to its sense then, as Wittgenstein argued in the *Tractatus*, this would imply that "whether a sentence had a sense would depend on whether another sentence were true", and "in that case we could not sketch out *any* picture of the world (true *or* false) ([13], 2.0211–2.0212; my emphasis).

The moral here would seem to be this: adoption of a transitive model of thinking, one, that is, according to which an ineliminable role is to be ascribed to the *object* of thought, will fail to meet the Principle of Spontaneity, and hence

will fail to allow for the possibility of what I have called the "interiority" of thought. It will be incompatible with the possibility that thoughts can be *had*.

If the objections which I have suggested are valid, then of the original requirements outlined above (p. 4), Frege's theory of thought and judgment fails to meet (*A*) (1) and (2), (*B*), and (*C*). And of these, the failure to fulfill condition (*C*) is the most serious, for it is this which explains its failure to meet the others.

NOTES

1. In case this point seems too obvious to deserve mention, it may be worth noting that the list of philosophers who have, albeit tacitly, contravened this requirement is long and distinguished: it includes Husserl, Moore, Russell, and indeed Frege himself (see below).
2. See [15], p. 219. For justification of these remarks, see [1], pp. 133–139.
3. For a similar conclusion, though on different grounds, see [3], pp. 96–97.
4. Clearly I am here using the term "concept" in its normal, i.e., non-Fregean sense, to mean roughly: nonpropositional component of a proposition.
5. This seems to dispose of Ramsey's "paradox", see [12], pp. 118–119.
6. I have explored elsewhere some of the further reasons for, and some of the consequences of, distinguishing two quite different notions of thought (or sense). See [1], pp. 112–125.

REFERENCES

[1] Bell, David, *Frege's Theory of Judgement*, Clarendon Press, Oxford, 1979.

[2] Dummett, M. A. E., *Truth and Other Enigmas*, Duckworth, London, 1978.

[3] Dummett, M. A. E., *The Interpretation of Frege's Philosophy*, Duckworth, London, 1981.

[4] Frege, G., "On Sense and Reference," in *Translations from the Philosophical Writings of Gottlob Frege*, edited by P. T. Geach and M. Black, Blackwell, Oxford, 1960.

[5] Frege, G., "Grundgesetze der Arithmetik," in *The Basic Laws of Arithmetic*, translated by M. Furth, University of California Press, 1964.

[6] Frege, G., "Die Grundlagen der Arithmetik," in *The Foundations of Arithmetic*, translated by J.L. Austin, Blackwell, Oxford, 1974.

[7] Frege, G., "Compound Thoughts," in *Logical Investigations*, edited by P. T. Geach, Blackwell, Oxford, 1977.

[8] Frege, G., "Negation," in *Logical Investigations*, edited by P. T. Geach, Blackwell, Oxford, 1977.

[9] Frege, G., "Thoughts," in *Logical Investigations*, edited by P. T. Geach, Blackwell, Oxford, 1977.

[10] Frege, G., *Posthumous Writings*, edited by H. Hermes *et al.*, translated by P. Long and R. White, Blackwell, Oxford, 1979.

[11] Kant, Immanuel, *Critique of Pure Reason*, translated by N. K. Smith, Macmillan, London, 1933.

[12] Ramsey, F.P., *The Foundations of Mathematics*, edited by R. B. Braitwaite, Routledge & Kegan Paul, London, 1931.

[13] Wittgenstein, L., *Tractatus Logico-Philosophicus*, translated by D. F. Pears and B. F. McGuinness, Routledge & Kegan Paul, London, 1961.

[14] Wittgenstein, L., *Letters to Russell, Keynes and Moore*, edited by G. H. von Wright, Blackwell, Oxford, 1974.

[15] Wittgenstein, L., *Philosophical Investigations*, translated by G. E. M. Anscombe, Blackwell, Oxford, 1974.

Department of Philosophy
Sheffield University
Sheffield S10 2TN
Great Britain