Chapter 1

Semantic paradox

1.1 Three paradoxes

Suppose I write on the board the following expressions:

pi

six

the sum of the numbers denoted by expressions on the board.

What does the third expression denote? Well, suppose it denotes the number k – that is, k is the sum of the numbers denoted by expressions on the board. But if the third expression denotes k, then, since the first denotes π and the second denotes 6, the sum of the numbers denoted by expressions on the board is $\pi+6+k$. But then $k=\pi+6+k$, and we obtain a contradiction. So, on pain of contradiction, the third expression is defective, and it cannot denote any number.

But we need not stop here. Since only the first and the second expressions denote numbers, it follows that the sum of the numbers denoted by expressions on the board is $\pi+6$. But here is the third expression again, so it does denote a number, namely $\pi+6$. And if that is so, then the sum of the numbers denoted by expressions on the board is $\pi+6+(\pi+6)$. But then the third expression denotes $\pi+6+(\pi+6)$. And we can iterate the reasoning, and obtain the absurdities that $\pi+6 = \pi+6+(\pi+6) = \pi+6+(\pi+6+(\pi+6)) = \ldots$.[1]

This is an example of a semantic paradox – a paradox of *denotation*. It's related to what are traditionally called the *paradoxes of definability*, associated with Richard, Berry and König.

The semantic relation between a denoting or referring expression and its denotation or referent is one of our basic word-world relations, so it is unsettling to find that we are led so quickly to paradox. And there are parallel paradoxes for our other basic semantic concepts – for example, the *extension* of a predicate, and the *truth* of a sentence.

Consider a paradox for extension, related to Russell's paradox. Now I write on the board these two predicates:

moon of the Earth

unit extension of a predicate on the board

(where a unit extension is an extension with just one member). Let E_1 be the extension of the first predicate. Let's suppose that the second predicate has a well-determined extension, E_2. Since E_1 is a unit extension, E_1 is a member of E_2. Either E_2 is a member of itself or it isn't. If it is a member of itself, then it has two members (E_1 and E_2); but then it isn't a unit extension of a predicate on the board, and so it isn't a member of itself – contradiction. So suppose on the other hand that it isn't a member of itself. Then it has one member (E_1), so it is a unit extension of a predicate on the board – so it is a member of itself, and we have a contradiction again. Since we're landed in contradiction either way, we must conclude that the second predicate is defective, and does not have a well-determined extension.

But again we can keep going. If the second predicate has no extension, and *a fortiori* no unit extension, then E_1 is the only unit extension of a predicate on the board. But here is the second predicate again, and so it *does* have a determinate (unit) extension, with sole member E_1. And then this extension is a unit extension of a predicate on the board, and so it is a member of itself. So we are led to the absurdity that the extension of the second predicate has no members, one member, and two members.

A liar paradox is generated by the following sentence, written on the board:

(L) The sentence written on the board is not true.

If L is true, then it isn't – contradiction. If L is false, then it's not true – which is what it says, so it's true. Either way we reach a contradiction. So L is defective – it cannot be given truth conditions.

But if L is defective, then the sentence written on the board is not true. But this is just L again. So L is true. And now we're caught in an endless cycle of contradictions: L is true, so L is not true, so L is true, … .

Each of these paradoxical discourses establishes that an expression is paradoxical - but they go further. The defectiveness or pathology of the expression becomes a new premise in the reasoning; we reason past pathology. A remarkable feature of this later stage of the reasoning is that the paradoxical expression re-emerges: we find ourselves asserting that the sum of the numbers denoted by expressions on the board is π+6, or that E_1 is the only unit extension of a predicate on the board, or that L is not true. There is nothing technical or recherché about the paradox-producing reasoning – ordinary speakers with the notions of denotation, extension and truth in their repertoire can easily follow it, and readily appreciate the challenge that the paradoxes present. Accordingly, any adequate solution to these paradoxes must respect the naturalness of the reasoning, and provide an account of these discourses in their entirety. The solution may identify an unwarranted assumption or find a logical flaw in the reasoning; it may show that what seems contradictory really isn't; it may instruct us to accept the contradictions. But whatever the solution, it must take the reasoning seriously. An adequate solution cannot block the reasoning in an *ad hoc* way, or simply set it aside.

I am after a unified account of these particular paradoxes and, more generally, of the paradoxes of definability, Russell's paradox for extensions, and the Liar. It's clear that the notions of reference, predicate application, extension and truth form a family: the name 'Napoleon' refers to a man, the predicate 'man' is true of Napoleon and Napoleon is in its extension, and the sentence 'Napoleon is a man' is true. Attention has been lavished on truth and the Liar paradox; the paradoxes of definability have taken a back seat, and Russell's paradox has largely been treated in the setting of sets and classes, not extensions. But the notions of denotation, predicate-application and extension are arguably no less fundamental than truth, just as referring expressions and predicates are arguably no less fundamental than sentences. I shall argue that the expressions 'denotes', 'extension', 'true' and 'true of' are all susceptible to the same kind of analysis, as befits the members of such a close-knit family. To give it a name, I shall be developing a *singularity theory* of these terms. The guiding idea of the singularity theory was briefly (and tantalizingly) expressed by Gödel this way:

> "It might even turn out that it is possible to assume every concept to be significant everywhere except for certain 'singular points' or 'limiting points', so that the paradoxes would appear as something analogous to dividing by zero. Such a system would be most satisfying in the following respect: our logical intuitions would then remain correct up to certain minor corrections, i.e. they could then be considered to give an essentially correct, only somewhat 'blurred', picture of the real state of affairs."[2]

1.2 Ramsey's division

I will be proposing a unified account of the paradoxes of denotation and the Liar paradox, along with a version of Russell's paradox (in terms of extensions). This prompts some adjustments to Ramsey's well-known division of the paradoxes. According to Ramsey, the paradoxes

> "fall into two fundamentally distinct groups, which we will call A and B. The best known cases are divided as follows:-
>
> A. (1). The class of all classes which are not members of themselves.
> (2). The relation between two relations when one does not have itself to the other.
> (3). Burali-Forti's contradiction of the greatest ordinal.
>
> B. (4). 'I am lying'.
> (5). The least integer not nameable in fewer than nineteen syllables.
> (6). The least indefinable ordinal.
> (7), Weyl's contradiction about 'heterologisch'."[3]

Ramsey continues,

> "Group A consists of contradictions which, were no provision made against them, would occur in a logical or mathematical system itself. They involve only logical or mathematical terms such as class and number, and show that there must be something wrong with our logic or mathematics. But the contradictions of Group B are not purely logical, and cannot be stated in logical terms alone; for they all contain some reference to thought, language, or symbolism, which are not formal but empirical terms".[4]

Peano anticipated Ramsey's division, claiming that Richard's paradox (a variant of (5) and (6)) belongs to "linguistics" rather than mathematics.[5] Nowadays, Ramsey's division is often drawn in terms of the "logical" paradoxes versus the "semantic" paradoxes.[6]

I believe that it *is* sensible to distinguish two types of paradox here, though not quite in the received terms. On the one hand, there are paradoxes that arise within mathematical language, in particular the language of set theory, and these turn on technical notions like set, cardinal number, ordinal number. I would resist calling them logical paradoxes, since it is far from clear that sets are logical objects; 'set-theoretical paradoxes' is a better label. On the other hand, there are those paradoxes that arise from semantical terms, like 'denotes' or 'defines', or 'true' and 'true of'. These are terms of natural language, not terms drawn from a technical, mathematical language. So the paradoxes arise in two very different settings, the setting of a formal, mathematical language and that of a natural language such as English. Here, then, is a principle of division for Groups A and B.

This is not to deny that there may be structural similarities between the paradoxes of the two groups. They may have certain common features, perhaps self-reference, or circularity, or a shared diagonal structure. But if they are couched in different kinds of language, we should be prepared to find differences in what will count as an adequate solution. What is required of an adequate response to the set-theoretical paradoxes of Group A? Perhaps the development of a contradiction-free set theory, which is expressed in a formal, precise language, and which provides a suitable foundation for mathematics. For the semantic paradoxes of Group B, on the other hand, we are after an account of our familiar notions of reference and truth – not an artificial, formal theory that fails to respect our employment of these concepts. Tarski himself was suspicious of bringing formal methods to bear on natural language -- for example, he rejected a 'Tarskian' solution to the Liar as it occurs in natural language. It is doubtful, he thought, that natural language, once regimented into a series of object languages and metalanguages, "would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalized languages".[7] An adequate solution to the Group B paradoxes must not replace our ordinary concepts with sanitized, artificial surrogates, or fail to capture semantic concepts available to the ordinary speaker, or fail to respect the reasoning we carry out using these concepts.

Consider, for example, Russell's paradox. Russell's paradox arises in *both* settings. Couched in terms of sets, and turning on the set of exactly those sets that are non-self-members, Russell's paradox belongs to Group A. What should we expect of a solution to Russell's paradox in this setting? Consider the claims sometimes made for Zermelo-Fraenkel set theory. Not only is it a formally precise axiomatic theory that avoids the contradiction (by restricting the Axiom of Comprehension), it also provides a natural conception of set (the iterative conception), and diagnoses and explains the fault in the reasoning that generates Group A paradoxes.[8] And, moreover, ZF provides a suitable foundation for mathematics. If all this is true of ZF, then it has a claim to be a solution to the set-theoretical paradoxes.

But as we have seen, Russell's paradox also arises for *extensions*. We should be careful to distinguish extensions and sets - or so I shall argue in Chapter 5. Extensions are tied to predication - extensions are the extensions of predicates. And so Russell's paradox for extensions arises in the setting of natural language. Accordingly, this version of Russell's paradox belongs to Group B. Here our concern is not to develop a formal set theory that will serve as a foundation for mathematics, but rather to provide an account of the relation between our predicates and their extensions, and explain the reasoning that appears to lead to contradiction. Resolutions of Group B paradoxes will be centrally concerned with everyday, non-technical notions, and certain discourses in natural language that turn on these notions. A resolution will be adequate to the extent that it captures those ordinary notions and provides a plausible analysis of the discourses. Technical maneuvering will be quite out of place here.

Should we think of the paradoxes of Group B as semantic paradoxes? This is surely reasonable, for Russell's paradox as well as the others. It is plausible to take the relation between a predicate and its extension to be a semantical relation (after all, an object is in the extension of a predicate if and only the predicate is true of the object). Still, here we should be open to the idea that the distinction between the semantic and the logical paradoxes may blur. Frege took the move from a predicate to its extension to be a purely logical move (more on this in Chapter 5). If we were to follow Frege here, extensions are logical objects and Russell's paradox is a logical paradox.

So we can make some adjustments to Ramsey's division – we can call the paradoxes in Group A ‘set-theoretical’ rather than ‘logical’, add a version of Russell's paradox to Group B, and call the paradoxes of Group B semantic (though it is arguable that the distinction between *semantical* and *logical* blurs in the case of extensions). But the division is sound, and what primarily distinguishes the two groups is the kind of language in which the paradoxes arise, mathematical language of set theory in the case of Group A, natural language in the case of Group B. And this division places different adequacy conditions on a solution to a given paradox, depending on which side of the division the paradox falls.

1.3 Universality

The adequacy conditions on a solution to the Group B paradoxes will depend on the special features of natural languages. Tarski remarked that a characteristic feature of natural language is its *universality* (we shall look more closely at Tarski's remarks in Chapter 4). When he speaks of the universality of a natural language, Tarski has in mind its expressive power, its flexibility and open-endedness, its capacity to evolve and expand. A natural language has the potential for saying anything that can be said in any language. In contrast, a mathematical or scientific language - such as the language of ZF - is expressively restricted. Unlike natural languages, these languages have a limited vocabulary and a limited subject matter.

Moreover, Tarski has shown that a suitably regimented language cannot express its own semantic concepts. If L is a classical formal language, it cannot contain a predicate 'true-in-L' applying exactly to the true sentences of L, on pain of contradiction. L cannot contain its own truth predicate, or its own denotation predicate, or its own satisfaction predicate.[9] If we want to express the concept of truth-in-L, we must ascend to an essentially richer metalanguage M for L, where M contains the predicate 'true-in-L'. However, M cannot express its own truth concept - the predicate 'true-in-M' belongs in turn to a metalanguage for M. In this way, "we arrive at a whole hierarchy of languages".[10]

The situation is quite different with natural languages. English, for example, contains the semantic predicates 'true', 'denotes', 'true of', 'extension', and it has the resources for explaining how these words are used. The very idea of an ascent to a metalanguage makes little sense for a natural language. We cannot move from English to an "essentially richer" metalanguage – that metalanguage will just be more English, or a language in principle translatable into English. In Tarski's terminology, English is *semantically universal* or *semantically closed*. Tarski identified this feature of natural languages as the primary source of the semantic paradoxes. And according to Tarski, this feature presents an insuperable difficulty for anyone seeking a definition of 'true' (or any semantical term) in English. If we bring formal methods to bear on a semantically universal language, then contradictions will inevitably arise, contradictions associated with the semantic paradoxes. Accordingly, Tarski turned away from natural languages, and defined truth for formal languages only.

In my view, we should respect Tarski's intuitions about natural languages. And so any account of the semantic paradoxes must be sensitive to the idea that natural languages are universal, and in particular semantically universal. We must reject any solution that regiments away the characteristic features of natural language.

For example, suppose we adopt a naive hierarchical solution to, say, the paradoxes of denotation. We start with a fragment L_0 of English free of the term 'denotes'. We obtain the metalanguage L_1 by adding to L_0 its denotation predicate 'denotes-in-L_0'. The extension of this two-place predicate is composed of ordered pairs of denoting phrases of L_0 and their denotations. In turn, we obtain the metalanguage L_2 by adding to L_1 *its* denotation predicate, 'denotes-in-L_1'. And so on, through the hierarchy. Now consider the paradox-producing phrase: *the least positive integer which is not denoted by an expression of English containing fewer than thirty five syllables*. (This phrase contains thirty four syllables.) According to the hierarchical account, any occurrence of 'denotes' is assigned a level - the phrase itself is a phrase of a language at some level of the hierarchy. So the phrase will contain the predicate 'denotes-in-L_i' for some level i, and the phrase is represented by:

> the least positive integer which is not denoted-in-L_i by an expression of L_i containing fewer than thirty five syllables.

Now this expression is a phrase of L_{i+1}, not L_i. And so the paradox is avoided, since the expression itself is not among the expressions of L_i to which it makes reference. Indeed, in principle a denotation can be found for the expression, by a suitable survey of the denoting expressions of L_i containing fewer than thirty five syllables.

The paradox is avoided, but at a high price. English appears to be a single language containing a single denotation predicate 'denotes'. According to this hierarchical account, however, English is stratified into a series of distinct languages, each with its own denotation predicate. We have left natural language behind. Here it does seem that English has been regimented away - we have merely described an artificial structure in which paradox does not arise, a structure with little or no bearing on English.

Further, consider the language in which we present the hierarchical solution. In this language we describe the hierarchy and its levels. But surely this language is just more English. If we can describe the hierarchy, we can do so in English - here again is the intuition that natural languages are universal. But if we can describe the hierarchy in English, we have the resources within English to articulate new paradoxes, 'revenge paradoxes' couched in the very terms of the theory. Consider for example the phrase 'the least ordinal number not denoted by any expression at any level of the hierarchy'. On the plausible assumption that the ordinals outrun the number of denoting expressions in the hierarchy, this phrase appears to denote a number. If every phrase appears in a language at some level of the hierarchy, so must this one. But the phrase denotes a number different from any expression in the hierarchy - and we have a contradiction again. If our task was to resolve the denotation paradoxes *for English*, then we have failed.

More generally, the universal character of English poses a challenge to any would-be theory of the semantic terms of English. The language of the theory will be expressible in English, and so there is the threat of new revenge paradoxes, expressed in English and employing the terms of the theory itself. The theory must meet this threat. If it doesn't, then it is not an adequate account of the semantic terms *of English*.

Here then are a number of desiderata for an account of the semantic paradoxes. Since the paradoxes arise in natural language, we must take into account the special character of natural language. The account must not be "harsh and highly artificial", to use Russell's phrase;[11] it must not force on, say, English an unnatural regimentation. And it must respect the apparent universal character of English – in particular, the fact that English contains its own semantic terms, and the resources for explaining their use. A consequence of universality is that any theory of the paradoxes will itself be expressed in English, or at least be translatable into English – and so if the theory itself generates new revenge paradoxes, then the paradoxes as they arise in English remain unresolved.

1.4 The plan of this book

In developing the singularity theory I make two main claims: first, that the semantic expressions ‘denotes’, ‘extension’ and ‘true’ are context-sensitive, and second, that these expressions are significant everywhere except for certain singularities. In Chapters 2-4, I lay out the main ideas of the singularity theory, framed by a close analysis of the simple paradoxes introduced above in 1.1. Chapter 2 defends the claim that ‘denotes’, ‘extension’ and ‘true’ are context-sensitive expression, in the light of recent work by philosophers, semanticists and linguists on the kinematics of context-change. Chapter 3 introduces the notion of a singularity, and Chapter 4 develops the central notions that will allow us to identify singularities.

Chapter 5 extends the scope of the singularity theory beyond the simple paradoxes, to the traditional paradoxes of definability and a variety of forms of the Russell and the Liar. In the case of the Russell paradoxes, I’m particularly interested in the distinction between sets and extensions – I argue that *extension* and *set* are two distinct and mutually irreducible notions, and provide two very different settings for Russell's paradox. In the case of truth, I pay special attention to the so-called *strengthened liar*, which has received a good deal of discussion in the literature.

Chapters 2-5 set the stage for the general, formal theory of singularities, presented in Chapter 6. The formal theory is pitched at a sufficiently high level of generality, so that it treats the paradoxes associated with *denotation*, *extension*, *truth* and *truth-of* in a single, unified way.[12] The singularity theory is put to work in Chapter 7. I present a number of paradoxes that are of interest in their own right, and I show how they're resolved by the singularity theory. These paradoxes include a transfinite paradox of denotation, various versions of the Liar that have figured in recent discussions (including the Truth-Teller and the Curry paradoxes), and new paradoxes of denotation and extension that, like Yablo's version of the Liar, do not involve self-reference or circularity.

Any solution to the paradoxes must take on the revenge phenomenon, and revenge is the focus of Chapters 8 and 9. We've already seen one form of revenge exhibited by the simple paradoxes of 1.1. In each case, the defective expression on the board re-emerges in the course of the reasoning, apparently intact and with a semantic value. The paradoxical expressions exact their revenge – our declaration that they're defective seems to lead only to their rehabilitation! Another form of revenge turns the theory against itself. For example, if a theory of truth introduces truth value gaps, then a revenge liar takes the gaps on board ("This sentence is false or gappy'); and similarly with a hierarchical solution ('This sentence is not true at any level') or a contextual solution ('This sentence is not true in any context').

In Chapter 8, I examine the revenge problem for several prominent theories of truth –

Kripke's theory, paracomplete theories, especially Field's, and dialetheist theories, especially Priest's. I argue that Kripke's theory fails to deal with the revenge problem, that Field's theory is too distant from natural language, and that Priest's theory is itself subject to revenge, despite its embrace of true contradictions. In Chapter 9, I turn to the revenge problem for contextual theories in general, and the singularity theory in particular. I present the singularity theory's response to revenge.

In Chapter 10, I draw out the consequences of the singularity theory for deflationary accounts of truth, principally disquotationalism, Horwich's minimalism and the prosentential theory. I argue that if the singularity theory is correct, then we cannot take a deflationary view of truth -- or of denotation or predicate-application. The semantic paradoxes have a direct bearing on the nature of our semantical concepts.

Notes to Chapter 1

1. This paradox was first presented in Simmons 2003.

2. Gödel, in Schilpp 1944, p.150.

3. Ramsey 1925, p.20.

4. *ibid*.

5. Peano 1906, pp.357-358.

6. For just one example, see Mendelson 1964, pp.2-3.

7. Tarski 1933/1986, p.267

8. For a recent defense of ZF as a solution to the set-theoretical paradoxes, see Giaquinto 2002, pp.214-8, and for discussion, see Simmons 2004, pp.172-175.

9. Satisfaction is the converse of the 'true of' relation: x satisfies y iff y is true of x.

10. Tarski 1944, in Blackburn and Simmons 1999, p.126.

11. Russell 1903, p.528. Ironically enough, this was Russell's reason for rejecting a type-theoretic response to the paradoxes, a response that he was later to endorse.

12. The specific theory of truth that I develop in Simmons 1993 is similar in spirit to the general theory that I develop here. But the present theory is at a higher level of abstraction and generality, and also differs from the earlier theory in a number of more specific features.