DIAGONALIZATION AND LOGICAL PARADOXES

DIAGONALIZATION AND LOGICAL PARADOXES

By

HAIXIA ZHONG, B.B.A., M.A.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2013) (Philosophy)

McMaster University

Hamilton, Ontario

TITLE: Diagonalization and Logical Paradoxes

AUTHOR: Haixia Zhong, B.B.A. (Nanjing University), M.A. (Peking University)

SUPERVISOR: Professor Richard T. W. Arthur

NUMBER OF PAGES: vi, 229

ABSTRACT

The purpose of this dissertation is to provide a proper treatment for two groups of logical paradoxes: semantic paradoxes and set-theoretic paradoxes. My main thesis is that the two different groups of paradoxes need different kinds of solution. Based on the analysis of the diagonal method and truth-gap theory, I propose a functional-deflationary interpretation for semantic notions such as 'heterological', 'true', 'denote', and 'define', and argue that the contradictions in semantic paradoxes are due to a misunderstanding of the non-representational nature of these semantic notions. Thus, they all can be solved by clarifying the relevant confusion: the liar sentence and the heterological sentence do not have truth values, and phrases generating paradoxes of definability (such as that in Berry's paradox) do not denote an object. I also argue against three other leading approaches to the semantic paradoxes: the Tarskian hierarchy, contextualism, and the paraconsistent approach. I show that they fail to meet one or more criteria for a satisfactory solution to the semantic paradoxes. For the set-theoretic paradoxes, I argue that the criterion for a successful solution in the realm of set theory is mathematical usefulness. Since the standard solution, i.e. the axiomatic solution, meets this requirement, it should be accepted as a successful solution to the set-theoretic paradoxes.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Professor Richard Arthur, for all of his helpful feedback and encouragement throughout the development of this dissertation. I would also like to thank Professor Nicholas Griffin, Professor David Hitchcock, Professor Graham Priest, Professor David De Vidi and Professor Gregory Moore for their thorough comments and helpful discussions.

I would also like to thank the Department of Philosophy at McMaster University for providing me with academic resources and financial support, without which I could not have accomplished my research.

Finally, I want to thank Meng, for putting up with me this whole time; and Elena, for giving me the courage to meet any challenge head on.

TABLE OF CONTENTS

Chapter 1: Introduction

1.1. The Topic and Scope of the Thesis

In this thesis, I will discuss the roots of and solutions to two groups of logical paradoxes. By ‘logical paradoxes’, I mean arguments which start with apparently analytic principles concerning truth, membership, etc., and proceed via apparently valid reasoning, while leading to a contradiction. Traditionally, these paradoxes are divided into two distinct families: set-theoretic paradoxes (Russell’s paradox, the paradox of ordinal numbers, the paradox of cardinal numbers, etc.) and semantic paradoxes (the liar paradox, Berry’s paradox, König’s paradox, Richard’s paradox, the heterological paradox, etc.). This classical division was made by Frank Ramsey (1926), based on what terms are used to express each paradox. There is also another kind of paradox which is closely related to them, that is, intensional paradoxes.[1] One example is the paradox of the concept of all concepts not applying to themselves. Another example can be found in Saul Kripke’s book *Philosophical Troubles: Collected Papers* (2011), that is, a paradox concerning the ‘set’ of all times when I am thinking about a ‘set’ of times that does not contain that time. However, since intensional paradoxes are based on intensional concepts, rather than linguistic expressions, I shall not include them in the scope of my thesis.

[1] Cf. Priest (1991). Gödel is also an advocator for this kind of paradoxes, though he calls them ‘conceptual paradoxes’. Gödel’s doctrine on conceptual paradoxes can be found in Wang (1996).

The two groups of paradoxes that I want to discuss, the set-theoretic and the semantic ones, are also often called ‘paradoxes of self-reference’. By this term, I mean that paradoxes in these two groups either explicitly or implicitly involve self-reference. This is achieved either by indexical terms that directly refer to the subject itself (e.g. in the liar paradox, and Russell’s paradox); or by circular use of some key notions (e.g. ‘definable’ in Berry’s paradox). It is arguable whether the feature of self-reference involved in these paradoxes is achieved in the same way. For example, in Gödel’s proof of the first Incompleteness Theorem, diagonalization is a crucial method to achieve self-reference within arithmetic. In Russell’s paradox, as well as the paradox of cardinal numbers, the role of diagonalization is also pretty clear. Then, one may ask, what is the role of diagonalization in other paradoxes of self-reference, especially the semantic paradoxes? This is a central issue for my thesis, which will be discussed intensively in Chapter 2 and Chapter 5.

Next, it is natural to ask whether all logical paradoxes are at the same time paradoxes of self-reference. The answer is ‘no’. For example, Yablo (1985) has successfully constructed a logical paradox without self-reference. Instead, it consists of an infinite chain of sentences, and each sentence expresses the untruth of all the subsequent ones.[2] Yablo’s paradox does not involve self-reference or

[2] More specifically, Yablo’s paradox can be stated as follows. For each natural number i, let’s define S_i as ‘for all $j>i$, S_j is not true’. To deduce the contradiction, first, let’s assume S_i is true for some i, then it is true that for all $j>i$, S_j is not true. Then, consider S_{i+1}, it is not true. But S_{i+1} is the sentence ‘for all $j>i+1$, S_j is not true’. Therefore, it is not the case that for all $j>i+1$, S_j is not true. Then there must be some sentence S_k ($k>i+1$) which is true. This contradicts the assumption that for all $j>i$, S_j is not true. Secondly, since we have proved

circularity. Rather, it involves the notion of infinity, and one can view it as violating some principle similar to the axiom of foundation in axiomatic set theory.

On the other hand, there are also paradoxes of self-reference which are not logical paradoxes. One example is the paradox about the knower: this sentence is not known by anyone. Although it sounds like the liar paradox, it is essentially about our notion of knowledge and depends on our epistemic evidence to discover the contradiction entailed by this claim. Therefore, it is not classified as a logical paradox.

In my thesis, I am mainly interested in logical paradoxes that involve self-reference. In particular, I shall discuss the essential feature of these paradoxes, the reason why contradictions arise, and the proper 'solution' (if any) to them. My main thesis is that the two different groups of logical paradoxes, semantic paradoxes and set-theoretic paradoxes, need different kinds of solution. For the semantic ones, I propose a functional-deflationary interpretation for semantic notions, and argue that the contradictions in the semantic paradoxes are due to a misunderstanding of the non-representational nature of semantic notions. For the set-theoretic paradoxes, I argue that the criterion for a successful solution is mathematical usefulness. Since the standard solution, i.e. the Axiomatic solution, meets this requirement, it should be accepted for a successful solution to set-

that none of the sentences S_i can be true, then for all $j>0$, S_j is not true. But this is exactly what is stated in S_0, therefore it must be true, which is again a contradiction.

theoretic paradoxes. The body of my thesis is divided into 6 chapters. The main idea of each chapter is summarized below.

1.2. Summary of Each Chapter

In Chapter 2 ‘The Diagonal Arguments’, first I summarise Cantor’s diagonal argument that there are infinite sets which cannot be put into one-to-one correspondence with the infinite set of natural numbers. Then I shall examine the diagonal method in general, especially the diagonal lemma and its role in mathematical logic. In Section 3, I briefly survey the discussion around diagonal arguments in logical paradoxes. In Section 4, I shall clarify the meaning of some important terms concerning diagonalization. Finally, I identify the features of the diagonal in three aspects: (i) that it passes through every row/element of the totality; (ii) that it is dynamic; and (iii) that it is essential to achieve self-reference in diagonal arguments.

Chapter 3, ‘The Liar Paradox: Introduction’, concerns some preliminary issues which are necessary for the discussion of the liar paradox. Firstly, there is an issue with truth bearers, which I identify as propositions. I distinguish propositions from linguistic entities (sentence types and tokens) or non-linguistic entities (such as the meaning of a sentence), and argue that a grammatically correct and meaningful sentence does not necessarily express a proposition. Since propositions are primary truth bearers, we talk about the truth or falsity of ‘sentences’ only in the derived sense, i.e. because a sentence expresses a true or false proposition. The distinction between sentences and propositions is important for my treatment of the liar paradox, because I will argue that the liar sentence is a meaningful sentence which does not express a proposition.

There are many versions of the liar paradox. In particular, we should distinguish contingent liar sentences from essential liar ones. Examples of the first kind often involve a description which denotes a sentence which happens to be the sentence itself. An example of the second kind is ‘This sentence is not true’, where ‘this sentence’ refers to the quoted sentence itself. The descriptive terms in contingent liar sentences can refer to something else in a different situation, so that the given sentence can have a truth value in that situation. For essential liar sentences, however, the referential terms cannot refer to anything else but the sentence itself. In this thesis, my primary interest is in essential Liar paradoxes.

In the third part of Chapter 3, I will briefly summarize major contemporary approaches which try to solve the liar paradox:

1. Tarskian hierarchy approach: No language can contain its own truth predicate. There is no unique truth predicate, but a hierarchy of infinitely many truth predicates, each of which is subscripted, and can only apply to sentences in a lower rank of the hierarchy.
2. Truth gap theories: There is a unique truth predicate for a language, and this language contains its own truth predicate. To avoid inconsistency, some sentences in this language cannot receive a truth value, among which

we find the liar sentence. Thus, this language contains some truth value gaps.

3. Contextualism: The truth value for the liar sentences is not stable, because the truth value should be assigned relative to a context, while the context for the liar sentence is always changing.
4. Paraconsistent approach: The basic idea is to allow the contradiction caused in the liar paradox, but to reject the thesis that everything follows from a contradiction.

I will provide a *prima facie* evaluation of these approaches based on the following three criteria for an adequate solution to the liar paradox in a natural language. First, since the aim is to explain the liar paradox found in a natural language, a proposed solution should accord as much as possible with natural 'pre-theoretic' semantic intuitions. Second, an adequate analysis of a paradox must diagnose the source of the problem in the paradoxes, and thereby help us refine the concepts involved, making them truly coherent. To design some artificial apparatus which simply circumvents the problem is not a good solution according to this standard. Third, an adequate account should provide a proper treatment for the problem called 'the revenge of the liar' (explained below).

All of the approaches mentioned above are flawed for failing to meet one or more requirements. However, though the truth gap approach has flaws too, I think there is a promising way to fix the problem. Thus, my solution to the liar paradox can be viewed as following the truth gap approach, and my major task is to provide a philosophical interpretation for the nature of truth value gaps, so that this explanation can meet the three criteria for an adequate solution.

In Chapter 4, ‘The Truth Gap Approach: Philosophical Interpretations and Problems’, I will discuss two of the most important theories within the truth gap approach, as well as their problems. In his paper ‘Outline of a Theory of Truth’, Kripke (1975) has shown how to construct a formal language which can consistently contain its own truth predicate by allowing truth-value gaps. In his construction, an interpretation of the truth predicate T is given by a ‘partial set’ (S_1, S_2), where S_1 is the extension of the truth predicate ‘T’, and S_2 the anti-extension of ‘T’, and ‘T’ is undefined for entities outside the set $S_1 \cup S_2$. Kripke proves there is at least one ‘fixed point’ for this language, where all sentences that can receive a truth value do receive a truth value at that point. However, the liar sentence cannot receive a truth value at the minimal fixed point; thus its truth value is undefined. Kripke calls such sentences ‘ungrounded’, and has provided a precise definition of this term.

Despite the mathematical precision and technical elegance, Kripke admits that the philosophical justification for such a construction of truth gaps needs to be supplied. Kripke intends to use Strawson’s ‘referential failure’ theory as the philosophical interpretation for the nature of ‘truth gaps’. According to this theory, a sentence cannot receive a truth value because the referential term in this sentence fails to refer. But Kripke does not specify the details of this explanation, nor is it clear why the referential term in the liar sentence fails to refer. After all, intuitively the referential term ‘this sentence’ in the liar sentence refers to the sentence itself.

A more important and troublesome problem for Kripke’s theory is the one called ‘the revenge of the liar’. Although truth gaps are allowed in Kripke’s language to avoid the contradiction in the liar, this treatment has generated a strengthened version of the liar, which is based on the gaps themselves. If Kripke uses ‘undefined’ to describe the status of the liar sentence, then we may ask what the truth value for the following sentence is:

This sentence is either false or undefined.

If we say that the above sentence is true, then apparently this assignment will lead to a contradiction. If this sentence is false, then since it says of itself that it is ‘either false or undefined’, this assignment will make it true. If we say that the truth value of this sentence is undefined, then again, since it says of itself that it is ‘either false or undefined’, this assignment will make it true, so that there is still a contradiction involved. Without an adequate philosophical account of the nature of truth gaps, it seems that the problem of the revenge of the liar is inevitable for all truth gap theories. Furthermore, since any truth value assignment of such a sentence will generate a contradiction in the given language, it seems that the only way out of the problem is to admit that this language cannot contain the predicate 'either false or undefined', which is equivalent to the predicate 'untrue'. Then, this language is not semantically closed[3], and we still need something like the Tarskian hierarchy in order to talk about the predicate 'untrue' for this language. However, if a truth gap theory in the end needs to resort to a Tarskian hierarchy to solve this problem, then the explanatory value of this theory is unclear.

The second part of Chapter 4 concerns Soames' theory, which is a major development of Kripke's approach. Soames wants to provide a philosophical explanation for the nature of truth gaps by using 'linguistic conventions'. According to his theory, the truth value of the liar sentence is undefined because our linguistic conventions do not say anything about its truth value. This interpretation, however, still has some intrinsic flaws. Firstly, Soames argues that, though the liar sentence cannot receive a truth value, it nonetheless still expresses a proposition. But his argument is based on examples of contingent liar sentences, while he does not explain how an 'essential' liar sentence can still express a proposition. Secondly, though Soames uses an artificial example ('smidget') to illustrate how the 'linguistic convention' works, it is not very clear whether there is any such explicit, artificial linguistic convention for our usage of the truth predicate. Finally, the definition that he provides for the truth predicate is essentially circular, and so cannot be a proper definition.

[3] A semantically closed language, as defined by Tarski (1944), is a language which contains names for its own expressions, as well as its own semantic predicates.

In Chapter 5, ‘Diagonalization and the Functional-deflationary Conception of Truth’, I try to provide my own philosophical explanation for the nature of truth gaps, which is based on the notion of diagonalization and on a distinction between representational and non-representational predicates, so that the heterological paradox and the liar paradox can be treated properly. I will also show how this account can solve problems such as the revenge of the liar. Based on Kripke’s truth gap theory, I construct a diagonal array as a simple model for Language **L**, which is a simplified version of English. I define the heterological predicate *Het* based on the diagonal function of the array. As argued in Chapter 2, the diagonal function is a dynamic notion and should not be confused with any row of cells in the diagonal array. Since *Het* is defined on the basis of the diagonal function, it is also a dynamic function and thus non-representational, in the sense that it cannot be fixed by any row of cells in the diagonal array. Consequently, the heterological paradox is solved, because there is no cell in the array which corresponds to the heterological sentence. In other words, the heterological sentence is not a proper candidate for a truth bearer. For the liar paradox, I advocate a functional-deflationary conception of truth, with the result that the truth predicate *T* should not be treated as a fixed set of cells in the diagonal array either. Consequently, there is no cell corresponding to the liar sentence in the diagonal array, which means that the liar sentence is not a proper candidate for a truth bearer either. In this way, I argue that the truth gaps associated with semantic notions are not caused by artificial linguistic rules, but are caused by the systematic features of natural language. Also, there is no problem like the revenge of the liar in this interpretation, because it is impossible to apply the truth predicate to the liar sentence. At the end of this chapter, I compare my interpretation with another approach to the liar, contextualism, and try to show that the latter violates some important intuitions associated with natural language.

In Chapter 6, ‘Paradoxes of Definability’, I extend the treatment of the liar to another kind of semantic paradox, paradoxes of definability (also called ‘paradoxes of denotation’[4]), which include Berry’s paradox, König’s paradox and Richard’s paradox. The chapter begins with a solution to this kind of paradox, which is an inference from the functional-deflationary interpretations of the heterological predicate *Het* and the truth predicate *T* developed in Chapter 5. Semantic notions are not representational. This feature is also called ‘deflationary’, for they do not have the content that ordinary expressions have. Semantic paradoxes, such as the liar, the heterological paradox, and the paradoxes of definability, are all caused by confusing non-representational terms with representational ones. Thus, I argue, that they can all be solved by clarifying the relevant confusion: the liar sentence and the heterological sentence do not have truth values, and phrases used to generate paradoxes of definability (such as that in Berry’s paradox) do not denote an object.

[4] Though, generally speaking, the word ‘define’ has a wider application than the word ‘denote’, in the context of these paradoxes they can be treated as meaning the same.

In the second section of this chapter, I defend this view further by arguing against a form of physicalism (held by Field 1972), and emphasize the distinction between representational expressions and non-representational ones.

In the third section of this chapter, I investigate another leading approach to semantic paradoxes: Priest’s dialetheism. Graham Priest (2002) argues that all logical paradoxes, including both set-theoretic paradoxes and semantic paradoxes, share a common structure, the Inclosure Schema, so they should be treated as one family. And the aim of this argument is to pave the way for his ‘uniform solution’ for all logical paradoxes, i.e. dialetheism. Through a discussion of Berry’s paradox and the semantic notion ‘definable’, I argue that (i) the Inclosure Schema is not fine-grained enough to capture the essential features of semantic paradoxes, i.e. the ‘indefiniteness’ of semantic notions; and (ii) that the traditional separation of the two groups of logical paradoxes should be retained. I shall also respond to Priest’s criticism of my argument and compare his dialetheism with my functional-deflationary solution, and argue that my explanation is preferable.

In Chapter 7 ‘Set-theoretic paradoxes’, I discuss the set-theoretic paradoxes. The main conclusion of this chapter is that the semantic paradoxes and the set-theoretic paradoxes belong to two different groups. I argue that the axiomatic solution is an adequate solution for set-theoretic paradoxes.

Through a careful examination of Cantor’s domain principle, I argue that Cantor’s philosophical argument cannot achieve his initial goal, i.e. justifying the existence of transfinite numbers while at the same time excluding the absolute infinite from his set theory. The acceptance of the notion of an infinite set in today’s mathematical practice is not due to Cantor’s domain principle, but due to the usefulness of this notion in mathematics. Therefore, the two groups of logical paradoxes should remain separated, because mathematicians and philosophers have different aims in their discussion of paradoxes. For mathematicians, their aim is to block the set-theoretic paradoxes efficiently, while the system of set theory is still strong enough to serve as a foundation of mathematics. Mathematicians need a scientific theory with useful, consistent concepts. That is why they are content with axiomatic set theory, which blocks the paradox by an extensional understanding of ‘set’.

However, for philosophers, when they deal with the semantic paradoxes, they want a theory which can explain the intuitions associated with natural language, a theory which can promote our understanding of the mechanisms of natural language. Since there are different aims for the discussion of the two different groups of logical paradoxes, the solutions of them are accordingly different.

Chapter 2: The Diagonal Argument

The family of diagonal arguments can be found in various areas of mathematical logic. It is also well-known that diagonal arguments play a central role in the set-theoretic and semantic paradoxes. In this chapter, I first introduce Cantor's original diagonal argument. In Section 2, I examine the diagonal method in general, especially the diagonal lemma and its role in mathematical logic. In Section 3, I briefly survey the discussion around diagonal arguments in the logical paradoxes. In Section 4, I clarify the meaning of some important terms used in discussing diagonalization. Finally, I identify the features of the diagonal in three aspects: (i) that it passes through every row/element of the totality; (ii) that it is dynamic; and (iii) that it is essential to achieve self-reference in diagonal arguments.

2.1. Cantor's Use of the Diagonal Argument

In 1891, Georg Cantor presented a new proof for the result that there are non-denumerable sets. A set is non-denumerable if it is an infinite, non-enumerable set. A set is enumerable if its members can be enumerated: arranged in a single list with a first entry, a second entry, and so on, so that every member of the set appears sooner or later on the list. Cantor had already established this result earlier in 1874 by a more cumbersome method.[1] The new method published in the 1891 paper is extremely simple and elegant, yet more powerful and

[1] Cantor (1874): "On a Property of the Set of Real Algebraic Numbers", in Ewald ed. (2007): 839-843.

convincing. This method, which is called ‘the diagonal method’ by later scholars, not only proves the existence of some non-denumerable sets, but also establishes a more general result, which says that, for any set X, the cardinality of its power-set P(X) is greater than the cardinality of X.

The result that Cantor wants to establish is that the collection *M* of elements $E_n = (x_1, x_2, \ldots, x_k, \ldots)$, where each x_i ($i \in \mathbf{N}$) is either *m* or *w*, is a non-denumerable set[2]. The idea of his argument is a *reductio* proof. First, suppose E_1, E_2, …, E_n, … is a complete enumeration of the set *M*, i.e. *M* is denumerable. Then, all the elements in *M* could be arranged in the following way:

Cantor’s Array: the diagonal argument for non-denumerable sets

$$E_1 = (a_{11}, a_{12}, a_{13}, \ldots, a_{1n}, \ldots)$$

$$E_2 = (a_{21}, a_{22}, a_{23}, \ldots, a_{2n}, \ldots)$$

$$\ldots$$

$$E_n = (a_{n1}, a_{n2}, a_{n3}, \ldots, a_{nn}, \ldots)$$

$$\ldots$$

Second, define an element $E_0 \in M$ as follows:

$$E_0 = (b_1, b_2, b_3, \ldots, b_i, \ldots) \ (i \in \mathbf{N}),$$
$$\text{such that, for each } k \in \mathbf{N},\ b_k = f(a_{kk}) = \begin{cases} w, \text{ if } a_{kk} \neq w \\ m, \text{ otherwise} \end{cases}$$

[2] Throughout the thesis, I want to distinguish these two terms: set and totality; and reserve the word ‘set’ in its strict, technical sense, according to ZF set theory. When I use ‘totality’, this means either it is not a set, or it awaits proof that it is a set. One may argue that, in Cantor’s proof, the *reductio* argument could be on ‘set’, instead of ‘denumerable’, but this is another issue, which I will discuss in Chapter 7. In this chapter, I shall follow the standard interpretation of Cantor’s proof, and treat the totality of real numbers as a non-denumerable set.

As shown above, the definition of the element E_0 is based on all the digits a_{kk} along the diagonal line of the array. This is why this method is called the 'diagonal method'. Since each b_k is either *m* or *w*, E_0 should belong to *M*. However, E_0 does not show up in the array above, because for any $k \in \mathbf{N}$, $b_k \neq a_{kk}$, thus for any $n \in \mathbf{N}$, $E_0 \neq E_n$. In other words, E_0 is left out by the list, which is supposed to be a complete list of the members of *M*. Since there is a contradiction derived, i.e. E_0 should belong to the array but in fact does not appear on the array, it follows that the assumption, that the array is a complete enumeration of *M*, is false.

Furthermore, there *cannot* be a complete enumeration of *M*. This is because, if one adds the new element E_0 to the sequence $E_1, E_2, \ldots, E_n, \ldots$, then the new array still can be diagonalized out by the same method. This process can keep on going without an end. Consequently, no sequence like $E_1, E_2, \ldots, E_n, \ldots$ could be a complete enumeration of all the elements in the set *M*.

2.2. The Diagonal Method in Mathematical Logic

The diagonal method created by Cantor has far-reaching consequences, not only in its original context of set theory, but also in the foundations of mathematics, computability and recursion theory (Gödel's fundamental incompleteness theorems, the halting problem, etc.), and the foundations of semantics (Tarski's theorem of the undefinability of truth). Of his diagonal proof for the non-deumerability of the set *M*, Cantor made the following comment:

> This proof is remarkable not only because of its great simplicity, but more importantly because the principle followed therein can be extended immediately to the general theorem that the powers of well-defined manifolds have no maximum, or, what is the same thing, that for any given manifold *L* we can produce a manifold *M* whose power is greater than that of *L*. (Cantor 1891, in Ewald ed. 2007: 921-2)

Cantor has correctly seen the wide application of this simple method. As quoted above, this method can show that for any given well-defined set (in his terminology, ‘well-defined manifolds’), there is a set (i.e. the power set of the given set) with a strictly bigger cardinality. This is known as ‘Cantor’s theorem’.

Other applications of the diagonal method, especially those theorems mentioned above, rely heavily on a single exceedingly ingenious lemma, the Gödel diagonal lemma. We have seen that there is an implicit feature of self-reference in Cantor’s original proof, i.e. the horizontal and the vertical index numbers for a_{11}, a_{22}, …, a_{nn}, … are the same. The feature of self-reference is more clearly manifested in the diagonal lemma, a classical version of which can be shown as follows[3]:

> Let *T* be a theory containing **Q**. Then for any formula $B(y)$ there is a sentence G such that $\vdash_T G \leftrightarrow B(\ulcorner G \urcorner)$.[4]

[3] The following version is quoted from Boolos et al. (2007): 221. In the statement, “**Q**” stands for minimal arithmetic, which has a finite set of axioms that are “strong enough to prove all correct ∃-rudimentary sentences”. (Boolos et al. 2007: 207).

[4] The formula ‘G’ surrounded by corner quotes “ ⌜ G ⌝ ” stands for the Gödel number of the formula ‘G’. The method of Gödel numbering is a systematic way of assigning to every formula ‘G’ in a language a natural number. Thus, the code “⌜ G ⌝ ” can serve as a name for that formula. The symbol ‘$\vdash_T$’ means what follows it is provable in Theory *T*.

The lemma indicates clearly that there is something like self-reference in the result (i.e. the sentence ‘G’ actually says that it itself has the property named by ‘B’). In order to show the role of diagonalization in achieving this result, we shall explore some details of the proof, which begins with the definition of ‘the diagonalization of a formula A’:

> The diagonalization of a formula A is the expression $\exists x(x = \ulcorner A \urcorner \& A)$.

Thus, we may think the formula ‘A’ is like the index number n for a_{nn} in Cantor’s proof, i.e. it is used both in the horizontal and vertical levels. The above definition is of most interest in the case of a formula $A(x)$ which has exactly one free variable. To prove the lemma, it is crucial to define ‘$A(x)$’ as: $\exists y$ (Diag(x, y) & $B(y)$). In this formula, the unbounded variable ‘x’ and the bounded variable ‘y’ range over Gödel codes for formulas. Therefore, the formula $A(x)$ (i.e. $\exists y$ (Diag(x, y) & $B(y)$)) actually says that there is a number y that is the Gödel code of a formula that is the diagonalization of the formula with Gödel code x, and that satisfies ‘B’. Then, the diagonalization of $A(x)$ becomes the sentence G: $A(\ulcorner A \urcorner)$, which is equivalent to $\exists y$ (Diag($\ulcorner A \urcorner$, y) & $B(y)$). Since G is the diagonalization of $A(x)$, then by the definition of diagonalization, $\vdash_T \forall y$ (Diag($\ulcorner A \urcorner$, y) $\leftrightarrow y = \ulcorner G \urcorner$), we obtain $\vdash_T \mathrm{G} \leftrightarrow \exists y(y = \ulcorner G \urcorner \& B(y))$, which is equivalent to: $\vdash_T G \leftrightarrow B(\ulcorner G \urcorner)$.

The trick of this proof is that, for the formula $A(x)$ which has one free variable, the diagonalization of $A(x)$ is self-referential: $A(\ulcorner A \urcorner)$. In other words, $A(x)$ is satisfied by itself. On the other hand, the formula $A(x)$ is also defined based on diagonalization, which creates self-reference in another sense: the diagonalization of *A* as a whole says that itself has the property *B*. We can use a less formal example to show what is really going on in the above proof. Let us define ‘the diagonalization of an expression’ (in the informal sense) as the result of substituting the quotation of the expression for every occurrence of the variable *x* in the expression.[5] Second, let the formula ‘*A*(*x*)’ be ‘John is reading the diagonalization of *x*’, and the formula ‘*B*’ be ‘John is reading’. Accordingly, the diagonalization of *A*(*x*) becomes:

> (*G*) the diagonalization of ‘John is reading the diagonalization of *x*’.

Beginning with the definite article ‘the’, ‘*G*’ looks like a noun phrase. But the noun phrase actually functions like the Gödel code $\ulcorner G \urcorner$, which stands for a sentence:

> (*G*) John is reading the diagonalization of ‘John is reading the diagonalization of *x*’.

Therefore, the sentence ‘*G*’ actually asserts ‘$B(\ulcorner G \urcorner)$’, i.e. John is reading the very sentence itself.

2.3. The Role of Diagonal Arguments in the Logical Paradoxes

We have seen that the diagonal method has many constructive results in set theory and logic. But this method is a double-edged sword, which has

[5] The following example is adapted from Smullyan (1994): 3-4.

destructive aspects as well. When it is used destructively, it causes various paradoxes. Cantor himself discovered a paradox several years after he established the theorem on power sets. That is Cantor's paradox of the greatest cardinal number, which was discovered in 1899. This paradox is an immediate inference from Cantor's theorem, provided that we accept the totality of all sets as a well-defined set. Let us consider the cardinal number κ of the 'set' S of all sets. On the one hand this number κ should be the greatest possible cardinal. However, if we apply Cantor's theorem of power set to Set S, we should obtain a set with a cardinal number which is strictly greater than κ. Thus we end up with a paradox.

It was also through pondering on the diagonal method that Russell discovered his famous paradox. In *The Principles of Mathematics*, Russell writes:

> When we apply the reasoning of his [Cantor's] proof to the cases in question we find ourselves met by definite contradictions, of which the one discussed in Chapter x is an example. (Russell 1903: §362)

In a footnote to this passage he adds: 'It was in this way that I discovered this contradiction'. What Russell discovered is that, if we consider the universal class[6], say U, we can have a function f such that for each element x in U:

$$f(x) = \begin{cases} \{x\}, & \text{if } x \text{ is not a class} \\ x, & \text{otherwise} \end{cases}$$

Then consider the diagonal class D, whose members are all classes not belonging to themselves. It turns out that $f(D) = D$. Then we may wonder whether D belongs

[6] Here we can understand Russell's term 'class' roughly as having the same meaning as Cantor's 'set'.

to $f(D)$ or not. The paradox is as follows: D belongs to $f(D)$ if and only if it does not. (cf. Russell 1903: §349)

Moreover, the destructive force of the diagonal method is found not only in set-theoretical paradoxes, but also in semantic paradoxes. Russell already discovered that there are some formal similarities between set-theoretic paradoxes and some prominent semantic paradoxes. This similarity is succinctly summarized by Simmons as a theorem (Simmons 1993: 25):

$$\text{(Ru)} \quad \neg\exists x\forall y(\mathrm{J}(x, y) \leftrightarrow \neg\mathrm{J}(y, y)).$$

Simmons' theorem (Ru) is developed from Thomson's discussion of semantic and set-theoretic paradoxes. In his paper "On Some Paradoxes" (1962), Thomson argues that the heterological paradox, Richard's paradox, and Russell's paradox 'can be shown to have a common structure'. (Thomson 1962: 104) The 'common structure' is identified as a theorem by Thomson as follows:

> (1) Let *S* be any set and *R* any relation defined at least on *S*. Then no element of *S* has *R* to all and only those *S*-elements which do not *R* to themselves. (ibid. 104)

This is actually the theorem (Ru) identified by Simmons above. Here it is stated in ordinary language rather than in symbols. We can consider the three paradoxes mentioned above, as well as the 'paradox' (or pseudo-paradox) of the Barber, to see how they can be shown as having the same structure:

Table 1: Thomson's Analysis of Common Structure[7]

Paradox	***S***	***Rxy***	**No element of *S* has *R* to all and only those *S*-elements which do not *R* to themselves**
The Barber	All villages (V)	*x* shaves *y*	No $x \in$V shave all and only those who do not shave themselves.
Heterological	All adjectives (A)	*x* is true of *y*	No $x \in$A is true of all and only those which are not true of themselves.
Russell's	All classes (C)	*x* is a member of *y*	No $x \in$C is a member of all and only those which are not a member of themselves.
Richard's	All names of sets of positive integers (N)	*y* belongs to the set of N named by *x*	No $x \in$N has all and only those elements *y* which do not belong to the set of N named by *y*

Based on Russell's and Thomson's analysis, Simmons (1993) examines both the constructive and destructive aspects of diagonal arguments, and summarizes the components of a diagonal argument. He shows that any diagonal argument consists of the following components: a side, a top, an array, a diagonal, a value, and a countervalue. ('Countervalue' is sometimes also called 'anti-diagonal'. To avoid confusion, I follow Simmons' terminology and use 'countervalue' throughout the thesis.) Visually, these components can be arranged as demonstrated in 'Array R' (on the next page).

The Array R has a side D_1, a top D_2. 'The diagonal' in Simmons' terminology refers to a 1-1 function from D_1 to D_2:

[7] This table is adapted from Thomson's analysis in his paper. Also, there are other kinds of analysis about the 'common structure' of logical paradoxes, which are different from the Thomson-Simmons analysis. I shall return to this issue in Chapter 6.

> F is *a diagonal on* D_1 and D_2 $\leftrightarrow_{df}$ F is a 1-1 function from D_1 to D_2. (Simmons 1993: 24)

Array R

D_2 D_1	1	2	3	…
1	▨			
2		▨		
3			▨	
…				▨ …

It should be noted that the diagonal need not to be 'the diagonal line' of an array literally (such as the line on Cantor's Array), but it has an essential feature: it must pass through every row, and must map each member of the side D_1 to a unique member of the top D_2. We shall see below that this mapping is essential to establish self-reference.

In Simmons' theory, 'the value' is defined by the diagonal function F:

> Let R be an array on D_1 and D_2, and let F be a diagonal on D_1 and D_2. Then, G is *the value of the diagonal* F *in* R $\leftrightarrow_{df}$ $\forall x \forall y \forall z(Gxyz \leftrightarrow Fxy \ \& \ Rxyz)$. (ibid.)

Thus, the value of the diagonal is actually a set of ordered triples. Based on the definition of the value, we can define 'a countervalue', which systematically changes the elements in the value to a different one[8]. Like the value, a countervalue is also a set of ordered triples on the array.

[8] In Cantor's original diagonal argument, there are only two options for a digit a_{nn}, that is, *m* and *w*.

Simmons distinguishes two kinds of diagonal arguments, good and bad, which correspond to the constructive use (which results in logical theorems) and the destructive use (which results in logical paradoxes) of the diagonal method. A bad diagonal argument assumes the well-determinedness of all the components of a diagonal argument, which is not the case. Consequently, a bad diagonal argument usually ends up with a contradiction, but we do not know where exactly the mechanism goes wrong. According to Simmons, many semantic paradoxes, including the liar paradox, are such bad diagonal arguments.[9] For a good diagonal argument, on the other hand, any component that is not a well-determined set is assumed to exist for a *reductio* proof. Therefore, in good diagonal arguments we can establish theorems by rejecting one of the assumptions, as Cantor did in his original proof.

2.4. Clarification of Important Terms

So far, we have discussed many issues around diagonalization: the diagonal method, the diagonal argument, the diagonal as a function, the value of the diagonal, a countervalue, and the diagonal array. Some of them have been defined by other authors, while others rely on our intuitive and vague understandings. Therefore, for the discussion in later chapters, it is important to

Therefore, there is only one countervalue of the diagonal. However, if there are more than two options for a digit, as that in a decimal notation, then there are more than one countervalue of the diagonal.

[9] In Chapters 5 and 6, I will discuss the diagonal argument in the liar and other semantic paradoxes in detail. Here I am more interested in the general features of a diagonal argument.

clarify the meaning of these terms and make it clear in which sense I use these terms.

A diagonal array is like that generated by Cantor. Usually, both the horizontal and the vertical dimensions of a diagonal array contain infinitely many elements. A diagonal argument is an argument which can be analyzed using such an array. 'The diagonal method' and 'diagonalization' thus are used loosely to refer to the method involved in diagonal arguments.

It is essential for a diagonal array to have a diagonal. Temporarily, let us follow Simmons' definition for the diagonal, which identifies the diagonal as a 1-1 function from the side to the top of the array. This definition will be refined in Chapter 5 for the discussion of the heterological paradox. The features of the diagonal of an array will be explored in detail in the next section.

'The value of the diagonal' and 'a countervalue' are both sets of ordered triples. For example, on Cantor's Array which is discussed at the beginning of this chapter, the value consists of all a_{11}, a_{22}, …, a_{nn}, …., i.e. the set $\{<n, n, a_{nn}>| n\in\mathbf{N} \}$. Similarly, all the shaded cells in Array R above consist in the value for Array R. It is important to note that *the diagonal* as a function is fundamentally different from *the value of the diagonal*, as I shall explain below. A countervalue is generated by systematically changing the third digit in each triple to a different one. The result of this operation is thus a new element which is different from any existent row of the array. That is, if we use b_n as the opposite digit to a_{nn}, then the new element consists of $\{<n, n, b_n >| n \in \mathbf{N} \}$.

2.5. The Features of the Diagonal Method

2.5.1. The Diagonal Passes through Every Row/element

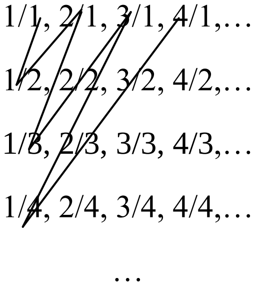
What exactly is the diagonal method? In the literature, many authors have discussed this issue. For example, Graham Priest writes:

> The essence of Cantor's proof is as follows. Given a list of objects of a certain kind (in this case, the subsets of x), we have a construction which defines a new object of this kind (in this case z), by systematically destroying the possibility of its identity with each object on the list. The new object may be said to 'diagonalise out' of the list. (Priest 2002: 119)

It is true that the diagonal method creates a new element that is different from any given object on the existent list, but this is the *result* of this method. This is a description of what has been produced by the diagonal method. The answer to the question 'what exactly is the diagonal method', on the other hand, should reveal some underlying features of this method, rather than just talking about the results produced by it. But what should these 'underlying features' be, if they are different from the result that a new element has been produced? To answer this question, it is helpful to explore some other methods that Cantor employed to show the enumerability of a given set.

The notion 'enumerable', which is defined as 'able to be arranged in a single list with a first entry, a second entry, and so on, so that every member of the set appears sooner or later on the list', means that for a given set, it has a definite beginning and there is a systematic, precise way to count all its elements. This idea underlies all the proofs for the enumerability of a given set. For example, in Cantor's proof for the enumerability of rational numbers, he employed the zigzag method:

Cantor's Zigzag Method for the Enumerability of Rational Numbers



1/1, 2/1, 3/1, 4/1,…
1/2, 2/2, 3/2, 4/2,…
1/3, 2/3, 3/3, 4/3,…
1/4, 2/4, 3/4, 4/4,…
…

Essentially, to enumerate all the elements in a set is to find a pattern in it so that all these seemingly rambling elements could be made quite 'tidy'. The pattern that Cantor found in the set of rational numbers is that every rational number can be written as a ratio of two integers. Although the array of rational numbers seems to have two infinite dimensions, i.e. each row and each column of the array can extend infinitely, it still has a definite starting point. There is a 'single continuous thread' (i.e. the zigzag line) going through every element in this set. It is tempting to think of the line which passes through the value, i.e. the set consisting of all <n, n, a_{nn}>, in Cantor's diagonal argument as such 'single continuous thread', but this is not the case. The value, as a set, simply passes through all the existent rows E_1, E_2, …, E_n, … on the array. Based on these elements, we can have a countervalue, which diagonalizes out the given array. Thus, the new element, i.e. the countervalue generated in that way, cannot be covered by the value. However, as that would be shown below, the countervalue would still be governed by the diagonal. This forces us to make an important distinction between 'the value of the diagonal' and 'the diagonal'.

2.5.2. The Dynamic Diagonal

The value of the diagonal is a fixed set. It is static, which contrasts with the dynamic diagonal. The latter is the diagonal function. From now on, I will use these three terms interchangeably: *the diagonal*, *the diagonal function*, and *the dynamic diagonal*. The details of the diagonal function will be redefined in Chapter 5. For our purpose here, it is enough to know that 'the diagonal' refers to a 1-1 function from the side D_1 to the top D_2. Every row on D_1 is governed by this function, though the diagonal does not have to pass through every column on D_2. Only the diagonal (function) can pass through every element, no matter whether it is in the array or is newly generated. For example, for the newly generated element E_0 in Cantor's proof, though it is not covered by the value of the diagonal in the existent array, it is still governed by the diagonal function. That is, when E_0 is added as a row of the array, then it is still within the domain of the diagonal function.[10]

[10] One may think that since the diagonal is a function defined on D_1 and D_2, and since that D_1 and D_2 are fixed, then the diagonal is also fixed. However, this is a misunderstanding of the diagonal. As shown above, the diagonal has the ability to generate more and more new elements, and these newly generated elements share the same structure with other elements on D_1, so when they are added to D_1, they will be still governed by the diagonal function. This is exactly why we call the diagonal 'dynamic'. Admittedly, since I stress the dynamic feature of the *Diag* function, this is a non-standard use of the word 'function'. As we shall see in Chapter 5, when I discuss the heterological paradox and the liar paradox, this non-standard understanding of 'function' is crucial

If we compare the zigzag line for rational numbers and the diagonal for real numbers, we can see the dynamic feature of the latter more clearly. For the former, both the set and the list are given once and for all, while this is not the case for the diagonal array. There is no new element created by the zigzag line. However, in the diagonal argument, there will always be new elements generated based on the elements given. The totality is thus a dynamic totality, compared with the 'static' set of rational numbers.[11] It is because of this dynamic feature that we cannot enumerate its elements. Correspondingly, the diagonal is also dynamic and can cover any newly added element, since it is defined for any element on the side D_1, no matter whether it is already existent on the array or will be produced through a countervalue.

Both the value and a countervalue generated by systematically changing the digits of the value are well-defined sets. When a countervalue is generated as a new element of the totality, it is then added to D_1 as a row. Since it is a row, it is within the domain of the diagonal function. On the other hand, the diagonal function could not be an element of the totality, nor could it be fixed by any set. The diagonal function (or briefly, 'the diagonal') is easily confused with the value of the diagonal, such as all the shaded cells in Array R. The latter could be represented as a row on that array, and a countervalue can also be added as a new

for my thesis.

[11] When I say that the totality is 'dynamic', this notion is used in the relative sense. For example, the set of real numbers is dynamic, compared with the 'static' set of natural numbers, since the former cannot be exhausted by the latter. On the other hand, the totality of all sets is dynamic compared with the set of real numbers, and the latter becomes relatively static, since it is still limited and bounded.

row of the array. But the diagonal function is not a row. This distinction is important for my discussion in Chapter 5, where I argue that the confusion between the value of the diagonal and the dynamic diagonal is the main reason for the contradiction in some semantic paradoxes. But at this moment, I only want to emphasize the dynamic feature of the diagonal in a general sense: *the diagonal is not a row*.

2.5.3. The Feature of Self-reference

Besides the two aspects discussed above, there is another important feature we need to notice about the diagonal. In his book *Gödel, Escher, Bach*, Hofstadter uses the following words to describe the diagonal method:

> The essence of the diagonal method is the fact of using one integer in two different ways—or, one could say, *using one integer on two different levels*—thanks to which one can construct an item which is outside of some predetermined list. One time, the integer serves as a *vertical* index, the other time as a *horizontal* index. (Hofstadter 1999: 423, original emphasis)

It is true that the numbers related to the diagonal have been used on two levels: horizontal and vertical. But it is more important to point out that in doing this, it results in a feature of self-reference for the elements to which the diagonal function applies. Recall the definition of 'the diagonalization of a formula *A*':

> The diagonalization of a formula A is the expression $\exists x(x = \ulcorner A \urcorner \& A)$.

As shown above, in the case when A has only one free variable, the diagonalization of $A(x)$ becomes: $A(\ulcorner A \urcorner)$, which means, A is satisfied by its own

Gödel number. Accordingly, if we draw an array similar to that in Cantor's proof, it is clear why this operation is called 'diagonalization' (see 'The Diagonal Lemma Array' on the next page).

In this demonstration, the side consists of all the one-place formulas in this language, while the natural numbers on the top stand for the G ödel codes of these formulas. The shaded cells represent sentences saying that when substituting the variable in the formula with the code of the formula, the resulting sentence satisfies the formula itself, e.g. $A(\ulcorner A \urcorner)$[12]. (See the illustration on the next page.) The feature of self-reference is also very important in various semantic paradoxes, especially the liar paradox, as we shall see in later chapters. Therefore, we should recognize the feature of self-reference as another important aspect of the diagonal.

In sum, the diagonal is a function which governs every element in the totality (i.e. it passes through every row of the array, whether existent or potential). As a function, it is dynamic and should not be confused with the value of the diagonal, or with any row on the array. Also, the diagonal is important in achieving the feature of self-reference in diagonal arguments. We will see that these features play an essential role in the semantic paradoxes, which will be discussed in Chapter 5. But, before I can provide any treatment to these paradoxes,

[12] The feature of self-reference is very important for Gödel's proof of the first incompleteness theorem, since he needs a sentence which says of itself that it is unprovable. As shown on Page 18, this essentially relies on the diagonalization of 'A': $A(\ulcorner A \urcorner)$, which is the sentence G itself. Thus, when the formula 'B' stands for the formula 'is unprovable', it follows from the diagonal lemma that this system entails the following sentence: $G \leftrightarrow B(\ulcorner G \urcorner)$. In other words, this system entails this sentence: G iff 'G' is unprovable.

there is some preliminary work which must be done in order to pave the way for the discussion in Chapter 5.

The Diagonal Lemma Array

	1	2	3	…	$\ulcorner A \urcorner$	…	
F_1	F_1 (1)						
F_2		F_2 (2)					
F_3			F_3 (3)				
…				…			
A					$A(\ulcorner A \urcorner)$		
…						…	

Chapter 3: The Liar Paradox: Introduction

In this chapter, I discuss some foundational issues which are necessary for the treatment of the semantic paradoxes, especially the liar. Firstly, there is an issue with truth bearers, which I identify as propositions. I distinguish propositions from linguistic entities such as sentence types and tokens and from non-linguistic entities such as the meaning of a sentence, and argue that a grammatically correct and meaningful sentence does not necessarily express a proposition. Secondly, I examine some different forms of the liar paradox, and distinguish ‘contingent’ liar sentences from ‘essential’ ones. In Section 3, I briefly survey major approaches to the liar paradox in the contemporary literature. At the end of this chapter, I summarize the criteria for an adequate solution to the liar paradox: (i) it respects intuitions associated with natural language; (ii) it can explain the mechanism of natural language rather than circumvent the problem; and (iii) it can provide an adequate treatment for the strengthened liar.

3.1. Foundational Issues

3.1.1. Truth Bearers: Sentences *vs*. Propositions

‘True’, as a grammatical predicate, applies to nouns and noun phrases which refer to sentences[1], statements, beliefs, assertions, propositions, assumptions, claims, etc. ‘True’ thus is also considered as a logical predicate,

[1] Throughout the discussion, ‘sentences’ should refer to declarative sentences, unless specifically explained.

which describes a certain kind of entity as having a certain property, i.e. truth. Such entities are usually called ‘truth bearers’. What is a truth bearer? When we say that John’s belief that the earth is round is true, we mean that John believes something that is true. When Mary says that Washington, D.C. is the capital of the United States, we think that Mary says something true. Thus, it is *something* that is true. Such things can be believed, asserted, claimed, stated, assumed, etc. Philosophers have a name for such things: propositions.

A proposition is the most popular candidate for truth bearer, if there is any. However, it is also notoriously difficult to characterize the status of propositions. What kind of entities are propositions? How do they exist? Are propositions linguistic entities? What is the relation between propositions and reality? Any of these questions would touch fundamental issues in metaphysics and philosophy of language, and would thus be difficult to answer. It is not my aim in this short section to provide a through philosophical theory for these issues. Instead, what I want to advocate is the following: there are some things that are true, and these things can be believed, asserted, claimed, stated, and understood by different people, even if they speak different languages. Such things are called ‘truth bearers’, and in this thesis, ‘proposition’ is simply another name for a truth bearer.

Propositions should be properly distinguished from other entities, such as sentences. It is common to say that this or that sentence is true, so it seems that a sentence and a proposition are quite similar. However, this is not the case. We may distinguish a sentence type from a sentence token. When John and Mary both say ‘I’m hungry’, they utter the same English sentence, but they make their own individual utterances. Thus, we may say that the sentence type is the same, while the tokens are different. Both sentence type and sentence token are language related. Thus, a translation of the English sentence ‘I’m hungry’ into French is a different sentence type from that sentence in English. Since different sentence types can express the same thing, and the same sentence type can express different things on different occasions, it is reasonable to regard the proposition expressed by a sentence as different from a sentence type. On the other hand, a proposition is different from a sentence token too. This is because sentence tokens are basically individualized in their physical sense, i.e. marks or sounds. Thus, although Mary’s statement that the Earth is round and John’s claim that the Earth is round express the same thing, they are different tokens. Therefore, though propositions have a close relation with linguistic entities like sentence types or tokens, the former cannot be identified with either of the latter.

However, if we regard propositions as truth bearers, while propositions cannot be identical with sentences, then does it make sense to say that this or that sentence is true? This problem could be solved by distinguishing a primary truth bearer from a truth bearer in a derivative sense. The primary truth bearers are propositions. A sentence or part of a sentence that expresses a proposition could be called a ‘truth bearer’ in a derivative sense. Thus, loosely speaking, we also say that such and such a sentence is true, especially in everyday language. A grammatically correct sentence cannot be true if it expresses no propositions. To use an example from Strawson's 'On Referring' (1950), when I extend my hand to you, and seriously claim that 'this is a fine red one', while there is nothing in my hand, you may feel confused and ask, 'what do you mean by this?' The audience feels confused because, though I stated one sentence, it is unclear what has been expressed by that sentence. Therefore, a grammatically correct declarative sentence can fail to express a proposition, so that it cannot be said to be true or false.

3.1.2. Meaning and Propositions

The second distinction that I want to make is between proposition and the meaning of a sentence. Though both of them are notoriously hard to articulate, there still could be something said concerning their difference. On many occasions, it seems that there is no difference between what a sentence means and what a sentence expresses. For example, it seems that the meaning of the sentence 'the Earth is round' and the proposition expressed by this sentence are the same. However, in some other cases, we can see the difference between these two aspects. Mary's claim 'I'm hungry' and John's claim 'I'm hungry' have the same meaning, in the sense that any competent speaker of English can understand this sentence, without knowing who utters this sentence. However, the audience cannot know what proposition is expressed by this sentence if she does not know the speaker of this sentence. Thus, in context-sensitive circumstances, the meaning of a sentence and the proposition expressed by the sentence can be clearly distinguished.

In his paper ‘On Referring’, Strawson characterizes the meaning of a linguistic expression or sentence as ‘the rules, habits, conventions governing its correct use, on all occasions, to refer or to assert.’ (Strawson 1950: 327) Following this idea, a competent speaker of a language can understand the meaning of a sentence, if she understands and has grasped the correct usage of that sentence. A sentence could be meaningful or significant, but it does not necessarily assert something that is either true or false. The minimal standard for a sentence to be meaningful is that it is grammatically appropriate. Thus, the sentence ‘this is a fine red one’, when the subject term ‘this’ fails to refer to anything (in that context), is still meaningful. Sometimes people may have stricter standards for a sentence to be ‘meaningful’. For example, if someone claims, ‘the sky is courageous’, without being in any poetic or metaphorical context, it may be considered as claiming something meaningless. This is because the predicate ‘courageous’, which is typically an adjective describing a person (or an animal), could hardly be used to describe the sky. We may think in this case the sentence fails to be meaningful according to our (implicit) understanding of linguistic conventions. I shall return to this issue at Chapter 5, but here what I want to draw from this discussion is simply that meaning and proposition are different issues, and that meaningfulness is a necessary but not a sufficient condition for a sentence to express a proposition. The notions of meaning and of proposition are distinct.

3.1.3. Context, Reference, and Propositions

From the discussion above, one may infer that whether a meaningful sentence expresses a proposition depends on context. Especially, one may think that, if a sentence contains some context-sensitive terms, then the sentence fails to express a proposition when the context-sensitive terms fail to refer to anything when used in a given context. Thus, in Strawson's example, it is because that the subject term 'this' fails to refer to anything that the sentence does not express a proposition. The issue of which proposition is expressed by a sentence cannot be determined without a look at the particular context. Although this sounds obviously true for sentences which contain context-sensitive terms such as 'I', 'it', 'this', 'today', etc. it could be very controversial in some other cases. Whether a sentence expresses a proposition largely depends on whether the referring term does refer to something. Thus, in some controversial cases, it is arguable whether the referring term actually refers. For example, in his discussion of definite descriptions, Donnellan (1966) has distinguished two kinds of use of such terms: attributive use and referential use. It is thus arguable what proposition is expressed by the sentence when the definite description in the sentence is used referentially. To use one of Donnellan's examples, when someone sees a man being kind to a young lady, without knowing that the lady is a spinster, and thus claims that ‘her husband is kind to her’, the speaker may claim something true even though there is no one to fit the description ‘her husband’ in that context. Donnellan says:

> But when we consider it as used referentially, this categorical assertion is no longer clearly correct. For the man the speaker referred to may indeed be kind to the spinster; the speaker may have said something true about that man. Now the difficulty is in the notion of "the statement." Suppose that we know that the lady is a spinster, but nevertheless know that the man referred to by the speaker is kind to her. It seems to me that we shall, on the one hand, want to hold that the speaker said something true, but be reluctant to express this by “It is true that her husband is kind to her.” (Donnellan 1966: 300)

If this argument is acceptable, then what has been expressed by a sentence in a context becomes very flexible and context-dependent. Also, the intention of the speaker for how to use the referential terms becomes critical for the proposition expressed by the sentence.

Since even when the object does not satisfy the description, the speaker still can use the description to refer to the object, then it seems that it is quite rare that the speaker uses a definite description but fails to refer to anything. That may happen, for example, when the speaker uses the description to refer to something, but there is nothing there (i.e. the speaker has some illusion in his/her head). This interpretation suggests one way to look at propositions. Although the speaker’s intention of using the language is important to determine what proposition is expressed by a sentence, a proposition is not something private, like a private mental image in the speaker’s mind. Rather, it should be understandable and communicable by others. The speaker may have something in his mind when he says ‘this is a fine red one’, but the audience cannot understand what he said because there is nothing in his hand. Though the speaker intends to use the term ‘this’ to refer, his intention is not enough to determine the reference and consequently the proposition. He cannot stipulate a referent for the term either. Propositions are not some private psychological states of the speaker. Rather, they are public, which can be accessed and grasped by any competent speaker in communication.

3.2. Some Versions of the Liar

3.2.1. Contingent *vs*. Essential

Usually, the liar paradox includes three key parts: self-referential terms, truth and negation. This paradox is also usually construed as depending on empirical facts. For example:

(1) Any sentence printed in this thesis on p. xx, line xx is not true.

Whether the sentence above is a liar sentence depends on the empirical fact of which sentence is actually printed in this thesis on p. xx, line xx. If it turns out some sentence other than Sentence (1) itself is printed in that place, then it is simply a normal sentence. It may be inferred from this instance that when self-reference is achieved by some description, then empirical facts are required to fix the reference of that description. In these cases, then, whether the relevant sentence is a liar sentence is not something that one can know *a priori*. The general form for this kind of liar could be summarized as follows:

(2) $\forall x\ (\mathrm{P}(x)\rightarrow \mathrm{Q}(x))$

Let ‘P’ be some description that could only be satisfied by this very sentence (2) itself, let ‘Q’ be the predicate ‘is not true’. Thus, the sentence in question ‘says of itself’ that it satisfies Q(x). Usually, the description ‘P’ denotes some property that depends on empirical facts, like the example in Sentence (1). The liar thus obtained is called ‘contingent liar’. There are other forms of contingent liars as well. For example, one can have a pair of sentences, the first of which says that the second sentence is true and the second of which says that the first sentence is not true. Or one can have a universal generalization like the statement, “I never tell the truth”, which includes itself in the class of items over which it is generalizing. All of them are called ‘contingent liars’ because they all depend on some empirical facts to obtain the paradox.

However, empirical facts are not necessary to establish the liar sentence. Using Gödel numbering, Gödel has shown how self-reference could be established by purely syntactic methods. Thus, if ‘is not true’ can be expressed in such a language, then the liar can be obtained by purely syntactic methods. Also, a liar sentence can be generated by the simplest method:

(α) (α) is not true.

Therefore, empirical facts are not a necessary condition to establish self-reference. Let us call liar sentences which do not rely on empirical facts ‘essential liars’. We shall see in Chapter 5 that an important difference between contingent liars and essential ones is that it is possible for the former to express a proposition, while there is no such kind of possibility for the latter. We shall also see in Chapter 5 that the reason why it is impossible for an essential liar to express a proposition reveals the underlying causes for semantic paradoxes. For these reasons, essential liars are more important and thus are the primary concern of this thesis.

3.2.2. False *vs*. Untrue

Although sentence (α) is the standard form of a liar sentence, the latter can be stated using concepts other than ‘true’. For example:

(3) The sentence I am saying is a lie.
(4) This sentence is false.

Usually, people treat ‘false’ as meaning the same thing as ‘not true’, and similarly, ‘lying’ as ‘not telling the truth’. However, it is still arguable whether sentence (4) expresses the same thing as the following one:

(α) (α) is not true.

People may argue, for example, that a sentence being false means that the condition for falsity of this sentence is fulfilled. However, it is not necessary that when we say something is ‘not true’ we mean that the condition for falsity is fulfilled. Instead, we may simply mean that the condition for truth is absent. If the condition for falsity of a sentence is fulfilled, then it follows that the condition for truth is absent, but not *vice versa*. This distinction is crucial for some approaches in contemporary solutions for the liar paradox. As we shall see below, both the contextual approach and truth gap theories rely on this distinction.

3.3. Major Contemporary Approaches to the Liar: A Survey

The liar paradox has been discussed for more than two thousand years by philosophers, and there have been numerous proposed solutions to it. In this section I do not intend to provide a survey of all solutions which have been proposed in history. Rather, what I want to do is to survey major proposals made since new mathematical logic methods became available in 20th century. My survey thus begins with Tarski, the first person who has provided a formal treatment for the semantic notion ‘true’, and then I will give a brief discussion of major post-Tarskian approaches. Also, I am not going to delve into the technical details of these theories, which are simply beyond the scope of this thesis. What I want to stress in this survey of these approaches are the philosophical assumptions and arguments for a certain treatment.

3.3.1. The Tarskian Hierarchy

According to Tarski’s analysis of the liar paradox, there are three assumptions which prove essential for generating paradoxes in a language (c.f. Tarski 1983: p. 165):

i. For any sentence which occurs in a language a definite name of this sentence also belongs to the language;
ii. Every instance of T-schema is to be regarded as a true sentence of this language;
iii. Usual laws of logic hold for this language.

A ‘T-schema’ (also called a ‘T-convention’) is a schema of the following form:

(T) *X* is true iff *p*.

where ‘*X*’ is a name for a sentence, and *p* is that sentence. One important thesis in Tarski’s truth definition is that a materially adequate truth definition for a formalized language should have all the sentences in the form of (T) as its consequence. Since any language that includes minimal arithmetic can contain the names for its sentences[2], to reject the first assumption is not an option to solve the problem. Also, as a strong advocate for classical logic, Tarski has quickly rejected the option of changing logic. Thus, the only option left for him is to reject the second assumption. Consequently, he famously built up different hierarchies for each choice of object language, with an object-language at the bottom and each successive language being the meta-language of its immediate predecessor. Thus, if level L_i is called the object language, then the level L_{i+1} is called the meta-language of L_i. Sentences in an object language can only be predicated as ‘true’ or ‘untrue’ in its meta-language or meta-meta-language, etc. In other words, for any natural numbers *m* and *n*, Tarski treats as ill-formed predicating ‘true$_m$’ of a sentence containing ‘true$_n$’ when $n \geq m$. By moving the T-sentences of an object

[2] It should be noted that if the claim is true in a general sense, then the language needs to satisfy other conditions as well, for instance, the class of well formed expressions needs to be recursively enumerable.

language to its meta-language, the application of the T-schema is restricted. There is no universal ‘true’ predicate in the hierarchy of languages, so that any instance of the T-schema should be subscripted according to the level at which it resides. That is, for a proposition p in a language L_i, the instance of the T-schema for p belongs to the meta-language of L_i.

In constructing this hierarchy of formalized languages, Tarski never intended to use it as an interpretation for natural language[3], nor could it be used as a solution for the liar paradox that is construed in natural language. In his eyes, natural language is hopelessly infected with contradiction. Thus, all his approach tries to do is simply to provide a sanitized model for languages that is appropriate for scientific usage. Tarski’s concept of an explicit hierarchy of languages has been criticized later for its artificiality. To use an example from Kripke’s criticism (1975), suppose that Dean says that all of Nixon’s utterances about Watergate are false. And Nixon also says that everything Dean says about Watergate is false. In this case, according to Tarski’s truth definition, the level for Dean’s utterance should be above the levels of all Nixon’s utterances. But Nixon’s words about Dean also require that its level should be above the levels of all Dean’s utterances. Therefore, we end up with a contradiction.

Though it is true that Tarski’s truth definition is not an adequate solution for the liar in natural language, one cannot blame Tarski for this failure. As said

[3] Nevertheless, Tarski himself also notes that the translation into colloquial language of a definition of a true sentence for a formalized language is a fragmentary definition of truth for colloquial language. See Tarski 1956: p. 165, n.2.

above, this is simply not his intention. Philosophers, on the other hand, cannot be satisfied with a solution for a formalized language, because to provide an adequate account for various intuitions associated with natural languages is an important goal for philosophers who work on semantic paradoxes, as will be discussed in Section 3.4.1 below. Thus, although they accept Tarski's work as a successful technical attempt, few of them are content with this achievement. Post-Tarskian truth theorists try to back up technical construction with some of our important intuitions about natural language, so that their proposals are not only for a formalized language, but also can serve as an adequate model for natural language, as discussed below.

3.3.2. Truth Gap Theories

Though Tarski deems the third option above, i.e. changing classical logic, as not a good candidate for solving the problem, this is generally regarded as the solution that truth gap theorists have adopted.[4] Thus, it is usually thought that truth gap theorists propose to handle the problem by denying some basic law in classical logic: the principle of bivalence. Though it may sound rather extreme to refer to the 'change of logic' as a way out of paradox, truth-gap theorists seem to have some good motives for their proposals as well. In opposition to Tarskian artificial, stratified truth predicates, truth gap theorists aim to search for a single

[4] In Chapter 4, we will see that both Kripke and Soames, two important authors for the truth gap approach, argue that there is no change of logic in this approach. In Chapter 5 I shall argue further for this point. To treat truth gap theory as changing classical logic is simply a misunderstanding of truth gap theory.

truth predicate with constant extension that applies to everything that can be said (truly) in the language. They try to argue that the status of the liar is quite special. No matter which concept they use to describe the liar (e.g. ‘undefined’, ‘ungrounded’, ‘indeterminate’, ‘categorically different’, etc.), all of them argue that the liar cannot receive a truth value evaluation like normal sentences. Also, they are more motivated by explaining natural language rather than by defining a truth predicate for a formalized language. Thus, they are more concerned with the intuitions that we have for natural language. There are many proposals for a gap theory for truth:

(1) The Kripke-Strawson theory of presupposition

In this proposal, truth gap theorists argue that some sentences suffer from a ‘truth gap’ because some essential presupposition is not fulfilled. Presupposition failure is different from falsity; thus, it could be called a ‘gap’ in truth values. For example, Kripke himself advocates the Strawson Presupposition theory as an interpretation for the nature of truth gaps:

> Under the influence of Strawson, we can regard a sentence as an attempt to make a statement, express a proposition, or the like. The meaningfulness or well-formedness of the sentence lies in the fact that there are specifiable circumstances under which it has determinate truth conditions (expresses a proposition), not that it always does express a proposition. (Kripke 1975: 699)

(2) Category Mistakes

Around 1970, there are also several other truth theorists who have come to an idea that is similar to Kripke's. Among them[5], Robert Martin (1967, 1970, 1976) advocates 'category theory', where he uses the notion 'category' to explain the nature of truth gaps.[6] Martin argues that, 'according to the category solution, every predicate of a natural language has, as one aspect of its meaning, a certain range of applicability (RA)'. (Martin 1976: 286) Thus, according to Category theory, we cannot say the liar sentence is true (or false) because the reference of the singular term is not the right sort of thing for the semantical predicate (i.e. 'true') to apply to.

(3) Gappy Predicates and Linguistic Conventions

Recently, Soames (1999) has proposed another interpretation for the nature of truth gaps, one which resorts to 'linguistic conventions'. For most sentences in a natural language, our linguistic conventions can tell us whether they are true or false. However, there are some sentences about which our linguistic conventions are silent, among which we find the liar sentence, as well as many other problematic sentences. We cannot find a definite answer to the truth values of these sentences because the predicates involved are partially defined by

[5] Others are: van Fraassen (1968), whose idea is similar to Kripke-Strawson's 'presupposition' explanation; and Herzberger (1970), whose work is mainly in the technical aspect of truth gap theory.

[6] It is said that later Martin rests little weight on the category idea, and sees it as subsumable under considerations of presupposition (reported by Burge (1979), footnote 6). Also, this approach is one of several implicit in Ryle's paper "Heterologicality", *Analysis*, XI 3 (1951): 61-69.

our conventions. Thus, we should be content with this situation and not ask about the truth values for the liar.

Although truth gap theory seems more natural than Tarski's theory, and has more concern about intuitions associated with natural language, it still faces some entrenched problems. The most notorious one is 'the strengthened liar' (sometimes also called 'the revenge of the liar'). No matter what concept the 'gap' is based upon, there can be a strengthened liar constructed based on that concept. Thus, for example, if the gap is explained as 'undefined', then there could be the strengthened liar: 'this sentence is either undefined or false.' If, on the other hand, there is a gap between 'determinately true' and 'determinately false', then there could be the strengthened liar as 'this sentence is not determinately true'. In a word, a strengthened liar can be constructed so that there is no gap between true (or 'determinately true') and its complement. Consequently, truth gap theorists should finally have to resort to something like Tarski's hierarchy, where they would distinguish object language from metalanguage, so that something which cannot be expressed in the object language (i.e. the complement of 'true' or 'determinately true') can be expressed in the metalanguage. But then, it would not be clear why we still need the truth gap explanation. As Burge objects:

> Indeed, they do little more than mark, in a specially dramatic way, the distinction between pathological sentences and sentences that are ordinarily labeled "false." (Burge 1979: 177)

I shall discuss the achievements and deficiency of these truth gap theories in Chapter 4 and 5.

3.3.3. Contextualism: Rediscovering the Tarskian Hierarchy

Since there are some fundamental problems for truth gap theories, and it seems that finally this approach cannot escape the Tarskian ghost of hierarchy, some theorists return to Tarski's idea of a hierarchy, and try to find new resources to solve the problem. They have rediscovered some reasonable factors in the thought of a hierarchy. But this time, it is not the explicit hierarchy in syntax, but in some more pragmatic elements like 'contexts', 'situations', etc. This group of proposals is generally called 'contextualism'.

There are many different forms of contextualism, for example, Parsons (1974), Burge (1979), Barwise and Etchemendy (1989), and Simmons (1993), but they share some basic intuitions and certain common features. One of the intuitions shared by contextualists is that, when the liar sentence is first stated, it is not true. It is not true because it lacks truth conditions. This is called 'falsity by default'. For example, in a cornerstone of this approach, Burge (1979) explains our intuition about the liar sentence in this way:

> In all the variants of the Strengthened Liar so far discussed, we started with (a) an occurrence of the liar-like sentence. We then reasoned that the sentence is pathological and expressed our conclusion (b) that it is not true, in the very words of the pathological sentence. Finally we noted that doing this seemed to commit us to saying (c) that the sentence is true after all. (Burge 1979: 178)

Thus, in the first step (i.e. step (a)), we conclude that the liar is not true since there is no *condition* (in some other contextual approaches, they also use 'facts') that can make it true. After establishing this result, there is also another important feature of the contextual approach. That is, there is a shift in contexts (or 'situations', as they are called by some contextualists), so that the result that has just been established is evaluated in a new context in step (b). In this new context that is just generated based on the result in step (a), there is some new 'factor' that makes the liar true. It is interesting to explore what this new 'factor' is. In Burge's theory, for example, he calls this factor 'implicature'. In another similar theory, it is called 'semantic fact'.[7] No matter what name contextualists give it, it should have this feature: it is generated by the confirmation of the 'untruthfulness' of the liar in the first step of the process, and affects the second step of the process of truth evaluation for the liar. It is because of this new 'implicature' or 'semantic fact' that the liar turns out to be true after all. But this shall not cause inconsistency, contextualists insist, because when the context shifts, it is a different truth predicate that is used to evaluate the sentence (or it is a different proposition expressed by the liar sentence that under evaluation). Therefore, the result that the liar is true in the second context is not contradictory with the result that the liar is not true in the first context. For example, Burge's approach is summarized as follows:

step (a): (I): (I) is not true

[7] For example, Barwise and Etchemendy (1989).

> Represented as: (1): (1) is not true$_i$
> Implicature: (1) is evaluated with truth$_i$ schema.
> step (b): (I) is not true (because pathological)
> Represented as: (1) is not true$_i$
> The implicature of step (a) is canceled.
> step (c): (I) is true after all
> Represented as: (1) is true$_k$
> Implicature: (1) is evaluated with truth$_k$ schema. (Burge 1979: 180-81)

All of these contextualists insist on two basic ideas: (i) that the default truth value for the pathological liar sentence in the first step is 'not true' – 'falsity by default'; (ii) that the shift in context gives the liar sentence a new truth value.

The shift of contexts keeps on going, because whenever the new evaluation is made, there is a new context generated, thus giving the liar another different truth value in the newly generated context, and so on *ad infinitum*. In this sense, it is also said that the truth value for the liar is unstable, i.e. it is always changing. Although this approach is widely advocated by philosophers, it also has its own difficulties. The most important among them is to explain the shift of contexts, and how this can affect the truth value of the liar. Also, the 'implicature' or 'semantic facts' in their theories is also very unclear. I will examine more details of contextualism in Chapter 5.

3.3.4. The Paraconsistent Approach

Although there are some differences in the approaches sketched above, they have something in common: all of them try to avoid contradiction by resorting to something like a hierarchy, either explicitly or implicitly. It is because of this feature that all the approaches above, including truth gap theories, are called 'Parameterisation' by Priest (2002):

> Even the contemporary solutions that are not explicitly parametric have to fall back in the last instance on the Tarskian distinction between object and metalanguage, and so on parameterisation. (Priest 2002: 152)

Instead of preventing or circumventing the contradiction by using something like hierarchy, Priest (2002) advocates another kind of solution: dialetheism, according to which there is at least one sentence (called a 'dialetheia') A, such that both it and its negation, ¬A, are true. His argument supporting this radical treatment consists of three steps. First, he argues that there is a common structure for all logical paradoxes, which he calls the 'Inclosure Schema':

> There are properties φ and ψ, and a function δ such that
>
> i. $\Omega = \{y: \varphi(y)\}$ exists and $\psi(\Omega)$;
> ii. if x is a subset of Ω and $\psi(x)$, then
>   a) $\delta(x) \notin x$, and
>   b) $\delta(x) \in \Omega$

Second, he advocates a principle of uniform solution: same kind of paradox, same kind of solution. (Priest 2002: 166) Third, he argues that such a 'uniform solution' could only be his dialetheic solution: to accept the contradiction in all these paradoxes. This approach is more radical than any of those discussed above, since it accepts contradiction as a legitimate part of the theory. I shall discuss Priest's approach in Chapter 6, and provide the reason why his ‘inclosure schema’ cannot guarantee a uniform solution to all logical paradoxes.

3.4. The Criteria for a Solution to the Liar Paradox

Since there are many different approaches in the contemporary literature to the liar paradox, one may wonder how to determine whether a proposed solution is a good one. To answer this question, we should first be clear what the aim of the particular theory is. Does it aim to provide a solution or at least an explanation for the liar paradox found in natural language, or does it simply try to define the concept ‘true’ in a way that could be free of contradiction? If the latter, as Tarski intended in his construction, then we cannot complain that that theory cannot accommodate intuitions in natural language. However, nowadays most philosophical discussions about the liar aim to solve the problem found in natural languages. For them, the issue related to natural language becomes an important criterion for us to evaluate whether a given theory or explanation is a good one.

3.4.1. Intuitions about Natural Language

Although in modern discussion of the theory of truth there is heavy weight laid on the side of technical ingenuity, there are also more and more theorists realizing that a philosophically satisfying theory must administer to the various intuitions associated with the natural notion of truth. For example, Kripke (1975) thinks that there are two merits in his theory:

> First, that it provides an area rich in formal structure and mathematical properties; second, that to a reasonable extent these properties capture important intuitions. (Kripke 1975: 699)

This consideration about intuition also can be found in many other authors' writings. For example:

> My objective is an account of the "laws of truth" whose application accords as far as possible with natural "pre-theoretic" semantical intuition. (Burge 1979: 170)
>
> Our goal in this book will be to provide a rigorous, set-theoretic model of the semantic mechanisms involved in the Liar, a model that preserves as many of our naive intuitions about such mechanisms as possible. (Barwise and Etchemendy 1989: 8-9).

Thus, it is a consensus that a theory for the liar paradox in natural language should take our intuitions about natural language seriously. However, there is a more fundamental problem regarding intuitions: is there any differentiation with regard to intuitions? Should all the claimed intuitions about natural language concerning truth be accepted as equally important, or is there any difference in their weight? The question could be even harder: are all claimed intuitions about natural language concerning truth equally sound and treatable as a philosophical justification? If the answer turns out to be 'no', then how could we make a judgement between different intuitions, and further, on the theories based on these intuitions? I will return to this topic in Chapter 5. At this moment, I only point out that any adequate solution for the liar found in natural language should take the issue about intuitions very seriously.

3.4.2. Explaining *vs*. Circumventing the Problem

There are two senses in which one could claim to have a ‘solution’ for a paradox: (i) that the paradox disappears after the treatment; (ii) that through the analysis of the paradoxes we understand more deeply what the roots of paradoxes and mechanism of languages are. A lot of existent treatments solve the problem in the first sense. The problem with them is that one can prevent a problem easily by setting some artificial restrictions, but it usually remains unexplained why these restrictions are reasonable and should be accepted. Usually, various ‘formal’ solutions offer no philosophical argument to back up their formal principles. However, if a treatment forces us to accept some seemingly implausible principle, or to abandon some intuitively plausible principles without providing any further reason other than the paradox itself, then such a solution cannot be said to be a successful one. It simply steps around the problem. Thus, the sense in which we need an adequate solution is the second sense, as observed by Barwise and Etchemendy:

> An adequate analysis of a paradox must diagnose the source of the problem the paradox reveals, and thereby help us refine the concepts involved, making them truly coherent. (Barwise and Etchemendy 1989: 4)

3.4.3. Adequate Treatment for the Strengthened Liar

The strengthened liar may be the hardest problem for most proposed solutions. I have mentioned this problem in the discussion of truth gap theories. However, this problem exists not only for truth gap theories, but for other approaches as well. For example, assuming a Tarskian hierarchy, one can construct the strengthened liar as follows:

(5) This sentence is not true in any level of the hierarchy.

For contextualists, the strengthened liar could be:

(6) This sentence is not true in any context of utterance.

The common issue in these approaches, as Priest has pointed out, is that all of them use some kind of parameter, so that it is possible to construct the strengthened liar based on that parameter. Therefore, an adequate solution to the paradoxes should not only be able to explain the ordinary liar, but also provide an adequate treatment for the strengthened liar. Moreover, these two levels of treatment should be consistent, in the sense that an *ad hoc* solution for the strengthened liar which does not have much to do with the explanation for the simple liar should not be accepted as a successful solution to this problem. Thus, as Burge observes: ‘The Strengthened Liar does not appear to have sources fundamentally different from those of the ordinary Liar.’ (Burge 1979: 173) And,

> Any approach that suppresses the liar-like reasoning in one guise or terminology only to have it emerge in another must be seen as not casting its net wide enough to capture the protean phenomenon of semantical paradox. (ibid.)

Other ordinary criteria for philosophical arguments, for example, Occam’s Razor, apply to these approaches as well. In the next chapter, I will use these criteria to evaluate some leading proposals following the truth gap approach, and discuss their problems. In Chapter 5, I will put forward a new version of truth gap theory which I argue satisfies these criteria.

Chapter 4: The Truth Gap Approach: Philosophical Interpretations and Problems

In this chapter, I discuss two of the most important theories in the truth gap approach, as well as their problems. In his paper 'Outline of a Theory of Truth', Kripke (1975) has shown how to construct a formal language which can consistently contain its own truth predicate by allowing truth-value gaps. Kripke intends to use Strawson's 'referential failure' theory as the philosophical interpretation for the nature of 'truth gaps', but he does not specify the details of this explanation, and it is unclear why the referential term in the liar sentence fails to refer. Since the nature of truth gaps is unclear, there is a troublesome problem for this theory – 'the revenge of the liar'. Because of this problem, this language cannot contain its own 'untrue' predicate (as well as 'undefined', 'ungrounded', etc.), and is thus not semantically universal.

Soames' gappy predicates theory, which is a major development of Kripke's approach, is an attempt to solve these outstanding problems. According to his theory, the truth value of the liar sentence is undefined because our linguistic conventions do not say anything about its truth value. This interpretation, however, still has some intrinsic flaws. Firstly, the 'linguistic convention' for the truth predicate that Soames has provided is essentially circular, thus cannot be the proper definition. Secondly, Soames argues that the liar sentence still expresses a proposition. But his argument is based on examples of 'contingent liars', while he does not explain how an 'essential liar sentence' can still express a proposition.

4.1. Kripke's Theory of Truth

4.1.1. The Intuition

Kripke's theory of truth begins with the idea of *grounding*. The intuition behind this notion is that one starts to describe the world with sentences containing only non-semantic terms, and then builds up successively more complex sentences containing semantic expressions. If a given sentence can receive a truth value during this process, then it is grounded; otherwise, ungrounded. According to Kripke's theory, all declarative sentences are thus divided into two groups: grounded and ungrounded. Grounded sentences are naturally assigned the values 'true' or 'false' in the process described above, while ungrounded ones such as the liar sentence are not. Based on this intuition, Kripke constructs a language which can consistently contain its own truth predicate.

An example would be helpful to illustrate this idea. Suppose we have to explain the notion 'true' to a person who does not yet understand it. 'We may say that we are entitled to assert (or deny) of any sentence that it is true precisely under the circumstances when we can assert (or deny) the sentence itself.' (Kripke 1975: 701) This understanding of truth is the same as the classical Aristotelian concept of truth: “to say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.” (Aristotle: *Metaphysics* 1011b25) Thus, the agent will learn that the circumstance where she is entitled to assert that ‘“snow is white” is true’ is the circumstance where she is entitled to assert ‘snow is white’. If, however, the sentence itself contains a truth predicate, then, without further instruction, the agent would still feel puzzled and would not know how to predicate ‘true’ of such sentences. Using Kripke’s own example, if one reads sentences like:

(1) Some sentence printed in the *New York Daily News*, October 7, 1971, is true.

she still would not know how to attribute truth to it. But this problem could easily be fixed. If there is at least one sentence printed in the *New York Daily News*, October 7, 1971, which does not contain the semantic notion ‘true’ and is actually true, then the agent should be able to assert that that sentence is true, by the process described above. Accordingly, by existential generalization, she would be able to assert that (1) is true. In the same way, the agent then could attribute truth to more and more complex sentences that contain ‘true’ in them. The intuitive idea is that a sentence containing the notion ‘true’ is only grounded if its truth value can ultimately be decided on the basis of the truth value of sentences which do not contain such a notion. A consequence is that the truth value of the liar cannot be determined in this way, so it is ungrounded.

4.1.2. ‘True’ as a Partially Defined Predicate

The intuitive notion of ‘grounding’ is substantiated by a truth gap construction. In Kripke’s view, a properly formed declarative sentence is always meaningful, but it is not the case that such a sentence always has a truth value (‘true’ or ‘false’). Therefore, the predicate ‘true’ in this language is only partially defined, i.e. the truth values for some sentences are *undefined*. The aim is to define the notion ‘true’ in such a way that the truth value of the liar sentence is undefined, so that one cannot meaningfully talk about the truth value of the liar. To achieve this goal, Kripke makes use of Kleene’s three-valued logic and some set-theoretic devices to construct a language, which is briefly explained below.

An interpretation of the truth predicate *T* in Kripke’s language is given by a ‘partial set’ (S_1, S_2), where S_1 is the extension of the truth predicate *T*, and S_2 the anti-extension of *T*, and *T* is undefined for items outside the set $S_1 \cup S_2$. In this language, there is no explicit hierarchy of truth predicates as in Tarski’s theory, but there is a process of interpretation of the language. In each step of this process, more and more true (false) sentences are recognized. Correspondingly, there are more and more elements added to S_1 and S_2. One may also think that there is an implicit hierarchy of languages corresponding to this process: L_0, L_1, L_2, …, L_n, …. In the first language L_0, the truth predicate is completely undefined, thus both the extension S_1 and the anti-extension S_2 of *T* are empty. This corresponds to the stage when the agent starts to learn the word ‘true’, and has not predicated truth or falsity of any sentence yet. After someone has explained to her the meaning of ‘true’ (e.g. Aristotle’s concept of truth), she is able to predicate truth of a vast number of sentences which do not contain semantic notions, and this corresponds to the language L_1, where both the extension and the anti-extension of the truth predicate have been greatly enlarged. And then she starts to learn how to predicate ‘true’ of sentences like (1), i.e. sentences that contain ‘true’ in themselves. This process keeps on going, and the agent’s ability to predicate truth becomes stronger and stronger.

A feature of this hierarchy is that any sentence that is true (or false) in L_n remains true (or false) in all the later steps. Thus, no truth value which is previously established changes in later steps. In other words, for any sentence *s*, the following condition holds:

s is true (false) in L_n iff T (<*s*>) is true (false) in L_{n+1}.

However, previously undefined sentences could receive truth values in later steps, since the agent’s ability to predicate truth increases. Consequently, one natural question that arises is whether this process will just keep on going indefinitely, or whether there will be any stage at which all sentences that can receive a truth value do receive a truth value. Kripke has successfully shown that there is such a stage, which is called a ‘fixed point’.

4.1.3. Fixed Point and the Formal Definition for ‘Grounding’

The intuitive idea of the notion ‘fixed point’ is that, at this stage, anything which could be described as ‘true’ or ‘false’ is fully recognized, so that it is stable in some sense. In other words, this language would contain its own truth predicate. It can be shown that the sequence of languages described above will eventually stabilise. In other words, there must be a fixed point where all the grounded sentences do receive a truth value at this point, so that the extension of *T* stops growing. Kripke has provided a formal proof for the existence of a fixed point. The main idea is a *reductio* proof. Suppose there is no fixed point; then there should always be a sentence which will be declared true/false for the first time at any given level. This means that for every ordinal number, there is always a new sentence in L corresponding to it. Therefore, the number of sentences in L should be equal to that of all ordinal numbers, which is non-denumerably infinite.

One might wonder why the steps of truth attribution should correspond to that of ordinal numbers, which is undenumerably infinite. This is because, with set-theoretic constructions, Kripke can not only define the hierarchy of languages corresponding to finite levels in this process: L_0, L_1, L_2, …, L_n, …, but also the hierarchy corresponding to limit levels. For example, for the first transfinite level, L_ω, we can define =($S_{1,\ \omega}$, $S_{2,\ \omega}$), where $S_{1,\ \omega}$ is the union of all $S_{1,\ \alpha}$, for all finite α, and $S_{2,\ \omega}$ is similarly the union of $S_{2,\ \alpha}$, for all finite α. In this way, the definition of the successor of a language applies to L_ω, thus giving us $L_{\omega+1}$, $L_{\omega+2}$, and so on. All the limit levels are thus defined as the union of all the previous levels. Kripke considers this feature of his language, i.e. that it can be extended to transfinite levels, as an advantage compared with Tarski's.

However, since each sentence of this language can only contain finitely many symbols, it follows that the totality of all the sentences of this language constitutes a denumerable set[1]. Therefore, since L consists only of denumerably infinitely many sentences, which means that there cannot be always a sentence to be declared true/false at a given level, the assumption that there is no fixed point must be false.

Actually, there could be more than one fixed point, but the most interesting one is the 'minimal fixed point', where all the grounded sentences receive a truth value. If L_r is the language corresponding to such a fixed point, then this language L_r should contain its own truth predicate. With these resources, Kripke then goes on to define 'grounding' precisely as follows[2]:

> Given a sentence *A* of L, let us define *A* to be *grounded* if it has a truth value in the smallest fixed point L_a; otherwise, *ungrounded*. (Kripke 1979; 706)

Since the liar sentence cannot receive a truth value at the minimal fixed point, it is ungrounded in L. Thus, the solution to the liar paradox implicit in Kripke's theory

[1] This assumes that there is only an at-most countable stock of symbols in this language.

[2] One may think that here Kripke just defines "grounded" in a way that makes the claim that all grounded sentences get truth values true, so it looks like a stipulation. However, it is a virtue of his definition that he can make the terms whose meaning is vague in natural language mathematically precise, while at the same time the definition also accords with our intuitive understanding of the word.

is that the truth value for the liar sentence is undefined, i.e. it suffers from a truth-value gap.

4.2. Philosophical Problems for Kripke's Construction

4.2.1. The Strawsonian 'Referential Failure' Interpretation

Although Kripke says that he has 'not at the moment thought through a careful philosophical justification of the proposal'; yet he actually suggests Strawson's 'reference failure' theory as the philosophical interpretation for the nature of 'truth gap':

> Under the influence of Strawson, we can regard a sentence as an attempt to make a statement, express a proposition, or the like. The meaningfulness or well-formedness of the sentence lies in the fact that there are specifiable circumstances under which it has determinate truth conditions (expresses a proposition), not that it always does express a proposition. (Kripke 1975: 699)

Thus, Kripke wants to make the distinction between a meaningful sentence and the proposition expressed by the sentence. This is the distinction that Strawson made in his paper 'On Referring' (1950):

(A1) a sentence
(A2) a use of a sentence
(A3) an utterance of a sentence

By '(A1) a sentence', Strawson means a sentence type, which is different from (A3) the utterance of a sentence (i.e. 'token'). For the type of a sentence, we can talk about the meaning of that sentence. According to Strawson, the meaning of a sentence or a linguistic expression is 'the rules, habits, conventions governing its correct use, on all occasions, to refer or to assert.' (Strawson 1950: 327) A sentence could be meaningful or significant, but it does not necessarily assert something that is either true or false. The latter is the function of the use of the sentence.

> So the question of whether a sentence or expression is significant or not has nothing whatever to do with the question of whether the sentence, uttered on a particular occasion, is, on that occasion, being used to make a true-or-false assertion or not, or of whether the expression is, on that occasion, being used to refer to, or mention, anything at all. (Strawson 1950: 327-8)

This passage shows that Strawson insists on the distinction between the meaningfulness of a sentence and the assertion or proposition it is used to make on a particular occasion. Consider a standard liar sentence:

(α) (α) is not true.

Following Strawson's suggestion, we may say that this sentence is meaningful since it is grammatically correct, but it does not assert anything that is either true or false. Given this explanation, however, one may still wonder why it fails to express a proposition. Kripke does not say anything about this problem. One possible explanation is, as suggested by Strawson, that the 'uniquely referring use' of the phrase '(α)' fails to refer to anything when this sentence is used on an occasion. If we treat '(α)' in the liar paradox as a proper name, then we may say that this name is an empty name. This problem may be considered as similar to that in the example provided by Strawson:

> I advance my hands, cautiously cupped towards someone, saying, as I do so, ‘This is a fine red one.’ He, looking into my hands and seeing nothing there, may say: ‘What are you talking about?' Or perhaps, 'But there's nothing in your hands.’ (Strawson 1950: 333)

Since one of the components of the sentence (i.e. the demonstrative pronoun ‘this’) fails to refer to anything, the sentence ‘this is a fine red one’ fails to express any proposition at all. People who hear this sentence would not say that the speaker said something false, but would rather point out the confusion in the speaker’s words.

The ‘reference failure’ interpretation, though it makes sense in Strawson’s example, could hardly explain the liar case if there is no further clarification provided. In Strawson’s example, it is the demonstrative pronoun ‘this’ in question, and there is really nothing in the speaker’s hand, so it is obvious that the word ‘this’ fails to refer to anything. In the liar case, however, there seems to be something that is denoted by the name ‘(α)’, i.e. the sentence itself. Therefore, it is not so obvious in what sense this name fails to refer.

One may argue in favour of Kripke’s theory by distinguishing sentences and propositions. As clarified in Chapter 3, propositions are primary truth bearers, and sentences are truth bears only in a derivative sense. A sentence cannot be true or false if it does not express a proposition. The name ‘(α)’ is a name which should refer to the truth bearer in that case, since the sentence predicates ‘is not true’ of what it refers to. Since (α) does not express a proposition (as claimed by Kripke), there is no truth bearer in this case. Therefore, (α) is an empty name. In this way, Kripke’s interpretation of the liar is analogous to Strawson’s interpretation of ‘this is a fine red one’. However, further explanation needs to be given of why (α) does not express a proposition, which Kripke does not give. This is a major problem for Kripke’s Strawsonian interpretation of the nature of truth gaps. The next chapter, where I develop another interpretation for truth gaps, will be mainly devoted to this problem.

4.2.2. The ‘Revenge of the Liar’ and the Expressive Power of the Language

Kripke’s theory holds that the truth value of the liar sentence is undefined, and this is the key to retaining consistency in a language which contains its own truth predicate. But in such a language, this fact cannot be stated. Although the liar sentence is categorized as ‘undefined’ in order to avoid contradiction, another related sentence could not be treated in the same way.

(β)   (β) is untrue.

In Kripke’s theory, ‘untrue’ means ‘either undefined or false’. Thus, if (β) is undefined, then it is untrue, which will make (β) true. According to the same reasoning, if it is false, then it is untrue, which will also make it true. Finally, if it is true, then this will directly lead to contradiction. It seems that no matter whether this sentence is true, false, or undefined, the result would be the same: contradiction. Therefore, on a higher level, the liar returns. This is called ‘the revenge of the liar’.

One may say that the reasoning from ‘undefined’ to ‘true’ is problematic, for if this sentence is undefined (which would be undefined in any sense), then how could such a sentence be true after all? In Kripke’s theory, ‘undefinedness’ means failing to receive a truth value (‘true’ or ‘false’) at the minimal fixed point. However, if someone understands ‘undefined’ as a third truth value, then it may be a candidate for further truth evaluation. Kripke insists that ‘undefined’ is not a third truth value (i.e. the truth value ‘neither true nor false’), as he says:

> “Undefined” is not an extra truth value… Nor should it be said that “classical logic” does not generally hold… *If* certain sentences express propositions, any tautological truth function of them expresses a true proposition. Of course formulas, even with the forms of tautologies, which have components that do not express propositions may have truth functions that do not express propositions either. … Mere conventions for handling terms that do not designate numbers should not be called changes in arithmetic; conventions for handling sentences that do not express propositions are not in any philosophically significant sense “change in logic.” The term ‘three-valued logic’, occasionally used here, should not mislead. All our considerations can be formalized in a classical metalanguage. (Kripke 1975: footnote 18)

However, his explanation is perplexing. First, he intends to use Kleene’s three-valued logic as the basis for his truth gap theory, which suggests that the gap is in the syntactic level and its existence alters classical logic. But then he says that what he has done is simply to articulate some ‘conventions for handling sentences that do not express propositions’. This suggests that the gap is at the semantic level, and it does not change the logic. It seems that Kripke would prefer that the gaps are on the semantic level, but his admission of the inference from ‘undefined’ to ‘not true’ and his reliance on Kleene’s three-valued logic are misleading. This is a problem in his theory, and I shall return to this issue in Chapter 5.

Because of the “revenge of the liar”, one is forced to conclude that this language cannot contain its own ‘untruth’ predicate, for the sake of consistency. Since this language cannot express certain concepts, its expressive power is limited. As Soames objects,

> one must then acknowledge that L_a does not contain its own untruth predicate and either ‘~’ fails to capture the sense of negation expressed by *not* or T_x fails to capture the notion of truth in L_a or both. (Soames 1999: 193)

Similarly, this language cannot express the predicate ‘undefined’ or ‘ungrounded’ either. This is because, if such predicates can be stated, then the predicate ‘untrue’ would also be expressible, for ‘untrue’ simply means ‘either undefined or false’, and other terms such as ‘false’ ‘either … or …’ have already been defined in this language. A consequence is that the words which Kripke uses to define the notion ‘ungrounded’ cannot be contained in the language that he aims to define. As objected by Simmons (1999):

> There are sentences that are intuitively grounded but are not in the minimal fixed point, for example, the grounded sentences of the metalanguage in which Kripke's paper is written. Such sentences are not captured by Kripke's definition. And the definition does not deal with sentences in which 'grounded' itself appears. This is a critical shortcoming, since the intuitive notion of groundedness itself gives rise to paradox. (Simmons 1999: 194)

There are also other problems about the expressive power of Kripke’s language, most of which are not philosophically significant so much as technically important. For example, if ‘p ≡ q’ is logically equivalent to ‘(p ⊃ q) & (q ⊃ p)’, which is intuitively correct, then it is not the case that every sentence with the form ‘*T*(A) ≡ A’ can receive a truth value at the minimal fixed point. Recently, Hartry Field (2008) has developed a modified fixed point theory to fix these technical problems. Since I am more interested in the philosophical interpretation of the nature of truth gaps than in developing some logical techniques, I will not discuss these technical issues in my thesis.

4.2.3. A Semantically Universal Language?

According to Tarski, natural language is universal, in the sense that natural language can express everything that can be expressed at all:

> A characteristic feature of colloquial language (in contrast to various scientific languages) is its universality. It would not be in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that 'if we can speak meaningfully about anything at all, we can also speak about it in colloquial language'. (Tarski 1983: 164)

It is arguable whether natural language is universal in this sense. For example, someone may question whether the native languages of some remote tribes have the resources to express matters concerning high-technology in modern society. But this is not important for our discussion here. The important issue is, whether natural language is semantically universal. By ‘semantically universal’, we mean that a natural language can be used to say all there is to be said about its own semantics. Tarski would say ‘yes’ to this question, since semantic universality is simply one aspect of universality *per se*. In another paper, Tarski (1944) uses another notion, ‘semantically closed’, to refer to the same idea. A language is semantically universal or closed, if it contains

> in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term “true” referring to sentences of this language; we have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language. (Tarski 1944: 348)

Thus, if we examine English, we do find there are names for its own expressions, as well as semantic terms such as ‘true’, ‘denote’, ‘satisfy’, etc. And it does seem to have the resources for describing the proper use of these expressions (e.g. Aristotle’s concept of truth). This feature has been identified by Tarski as the primary source of semantic paradoxes. Thus, in his formalized language, he gives up this feature, and restricts the use of the predicate ‘true’ relative to different levels, so that no level can contain its own truth predicate. Since Tarski does not aim to provide a definition of truth found in natural language, we cannot complain that he sacrifices the semantic universality of natural language. However, Kripke does try to provide a model for natural language (at least to some extent), so that the issue concerning the semantic universality of natural language could be a legitimate objection for him. As discussed above, the expressive power of Kripke’s language is limited. In particular, it cannot express its ‘untrue’ predicate. Therefore, this language is not semantically universal. And if we want to predicate ‘untrue’ of some sentence in this language, we will inevitably have to ascend to a meta-language. But if so, then what is the advantage of Kripke’s theory over Tarski's? Kripke does not have any good answer to thisquestion. He simply admits that this is a weak point of his theory:

> The necessity to ascend to a metalanguage may be one of the weaknesses of the present theory. The ghost of the Tarski hierarchy is still with us. (Kripke 1975: 714)

He also doubts whether such a theory could be given, i.e. one providing a way out of the liar paradox without sacrificing semantic universality:

> Nevertheless the present approach certainly does not claim to give a universal language, and I doubt that such a goal can be achieved. First, the induction defining the minimal fixed point is carried out in a set-theoretic meta-language, not in the object language itself. Second, there are assertions we can make about the object language which we cannot make in the object language. (Kripke 1975: 714)

Therefore, the problem about the semantic universality of natural language persists as an important problem for all truth gap theorists. Also for this reason, truth gap theory is criticized by some authors as getting the fundamental picture wrong:

> Adopting gaps and assuming universality leads to contradiction: The gaps allow the construction of a concept that, if assumed to be expressible, generates a paradox. But the point is not just that an appeal to truth gaps fails to preserve intuitions about universality. This new paradox arises out of the appeal to gaps and must be resolved in some other way. The truth-value gap theorist fails to provide a general, unified account of semantic paradox. (Simmons 1993: 46-7)

Kripke himself does not reply to this criticism. Let us see whether we can find a satisfactory answer in another interpretation proposed by Soames (1999).

4.3. Soames' Gappy Predicate Interpretation

4.3.1. Gappy Predicates and Linguistic Conventions

In his book *Understanding Truth* (1999), Soames tries to provide a new interpretation for the nature of 'truth gaps' in Kripke's formal construction. Soames' interpretation is to resort to 'linguistic conventions'. Briefly, his idea is that our linguistic conventions do not say anything about the truth value of the liar sentence: thus it is undefined. Soames illustrates this idea by an example which explains how a gappy predicate (i.e. a partially defined word) can be introduced into a language by conventions.

Suppose there are two groups of people. Group A consists of adults who are abnormally short (around four feet tall), and Group B consists of adults whose height is at the low end of the normal range (around five feet tall). Moreover, each member of group B is perceptibly taller than any member of group A. One can introduce a new word 'smidget' into the language in the following way:

> i. Every member of group A is (now) a smidget. Further, for any adult whatsoever (and time t), if the height of that adult (at t) is less than or equal to the (present) height of at least one member of group A, then that adult is a smidget (at t).
>
> ii. Every member of group B is not (now) a smidget. Further, for any adult whatsoever (and time t), if the height of that adult (at t) is greater than or equal to the (present) height of at least one member of group B, then that adult is not a smidget (at t).
>
> iii. Nonadults (and nonhumans) are not smidgets. (Soames 1999: 164)

The definition above gives a sufficient condition for something to be a smidget and a sufficient condition for something not to be a smidget. But it does not give sufficient *and necessary* condition for a thing to be called ‘smidget’. If there is an adult, say Mr. Smallman, whose height is precisely halfway between that of the tallest person in group A and the shortest person in group B, then it remains undefined whether this adult is a ‘smidget’. In short, there is a gap between the height of the tallest person in Group A and that of the shortest person in Group B. Since the convention governing the usage of ‘smidget’ remains silent about this range of height, the truth value of the sentence ‘Mr. Smallman is a smidget’ is undefined. In other words, the linguistic convention cannot tell us anything about whether a person of such a height is a smidget.

Soames argues that the status of truth gaps in Kripke’s truth theory is just like those in the smidget example. The truth values of some sentences are undefined because our linguistic conventions do not say anything about them. For atomic sentences and their negations, Soames identifies our linguistic conventions governing the usage of the predicate ‘true’ as follows:

> 3a. The predicate 'red' applies (does not apply) to an object ≡ it is (is not) red. The predicate ‘smidget’ applies (does not apply) to an object ≡ it is (is not) a smidget. (And so on, one clause for each predicate in the language)
>
> 3b. For any n-place predicate P and terms $t_1, \ldots, t_n$, $\ulcorner Pt_1, \ldots, t_n \urcorner$ is true (not true) ≡ P applies (does not apply) to the n-tuple $<o_1, \ldots, o_n>$ of referents of the terms.
>
> 3c. For any sentence S, $\ulcorner \sim S \urcorner$ is true (not true) ≡ S is not true (true). (Soames 1999: 166)

As Soames explains, ‘instruction (3b) uses the notion of a predicate applying to an object to explain what it is for an atomic sentence to be true; (3c) extends the explanation to negations of such sentences.’ (ibid. 166) For composite sentences with connectives (i.e. conjunction, disjunction, material implication, and quantification), their truth conditions could be defined using a system similar to Kleene’s three value logic.

4.3.2. Whether the Liar Sentence Expresses a Proposition

One difference between Soames’ interpretation and Kripke’s Strawsonian interpretation is around the issue whether the liar sentence expresses a proposition (which is the truth bearer) at the minimal fixed point. Kripke intends to adopt Strawson’s referential failure theory as the interpretation. As we know, Strawson argues that there are sentences which, when used on certain occasions, fail to express a proposition. Consequently, we may conclude that according to the Kripke-Strawsonian Interpretation the liar sentence fails to express a proposition at the minimal fixed point.

This point is criticised by Soames. He argues that ‘it is simply not true that all liar sentences, truth teller sentences, and other ungrounded sentences fail to express propositions at the minimal fixed point.’ (Soames 1999: 193) He supports this claim by using an argument concerning propositional attitudes. Consider the following sentence:

(2) Bill believes that this is a fine red one.

Since ‘this’ fails to refer to anything (in Strawson’s example), it follows that if Bill is rational and a competent language user, he cannot believe that this is a fine red one. In other words, there is no proposition that he can believe. On the other hand, consider the case for the liar sentence:

(3) Bill believes that (α) is not true.

It seems that this sentence is not as problematic as (2). And some theorists not only believe but also try to show that the proposition expressed by the Lair sentence is indeed not true.[3] Thus, it seems that sentence (3) could be true. But, if propositional attitude ascriptions report relations between the agent and a proposition, then the liar sentence must express a proposition. Soames therefore says:

> We already have the proposition; it is just that the proposition cannot correctly be evaluated for truth value in every possible circumstance. But then the same thing should be said about (la) and (lb)[4]. Once it is admitted that there are propositions that resist evaluation in certain circumstances, there are no longer grounds to suppose that (la) and (lb) do not express propositions in the context originally imagined. (Soames 1999: 169)

Soames understands the meaning of a sentence as ‘a compositional function from contexts of utterance to propositions expressed’ (Soames 1999: 168). If we consider the liar sentence, its meaning would determine different propositions in

[3] For example, contextualists argue that the liar sentence is false by default in the initial context, because of the absence of the condition for its truth.

[4] (1a) Mr. Smallman is a smidget. (1b) Mr. Smallman is not a smidget.

different context. Thus, in some possible context, the liar sentence can express a proposition which is true or false. Soames gives the following example:

(4) Some sentence written in place P is not true. (Soames 1999: 193)

Depending on which sentence is written in place P, the sentence above could be a liar sentence and thus fail to have a truth value. But it also could express a proposition which is either true or false. Thus, Soames argues that we cannot say that the liar sentence fails to express a proposition.

4.4. The Advantages of Soames' Theory

Soames argues that his interpretation has several advantages that the Kripke-Strawsonian interpretation does not possess. In particular, Soames argues that his theory can deal with important criticisms of the truth gap theory.

4.4.1. Is 'Undefined' a Third Truth Value?

Following Kripke, Soames insists that 'undefined' is not a third truth value. According to Soames, to say that 'undefined' (or 'ungrounded') is the third truth value is to make a kind of category mistake:

> To assert something ungrounded is to make a kind of mistake. But the mistake is not correctly described as that of saying something untrue. Rather, it is in saying something that cannot, in the end, be sanctioned by the linguistic conventions that give one's words their meaning. (Soames 1999: 172)

For the liar sentence:

(α) (α) is not true.

Soames insists that since we have no ground or justification to attribute truth value to it, we should reject both of these claims: (i) (α) is true; (ii) (α) is not true. However, rejecting one sentence does not equate to denying it. Thus, rejecting a sentence does not mean accepting or affirming its negation. This is because, if we reject one sentence, it means we do not have justification to either assert it or deny it, which means that the truth condition and the falsity condition are both absent. But if we affirm the negation of this sentence, it means that we do have justification to assert the negation of it, i.e. the falsity condition is present. Thus rejecting (i) cannot give us any justification for affirming that (α) is not true. Similarly, rejecting (ii) does not give us any justification for affirming that (α) is true. Therefore, from the rejection of (i) and (ii), we cannot draw the conclusion that (α) is neither true nor false. For the same reason, we cannot say that (α) is both true and false either. In a word, there is no 'third truth value' for such a sentence. In this way, Soames argues that his theory has provided an interpretation for the status of the truth gap which sounds more natural and reasonable:

> First, the gaps are not technical artifacts cooked up just to avoid the paradox; rather, they exist independently in language and arise from a process that applies to semantic and nonsemantic notions alike. Second, the gaps result from a plausible set of instructions for introducing the truth predicate; the gappy character of Truth Tellers and Liars is an automatic and unpremeditated consequence of these instructions. Third, the gaps provide an explanation of how we can (and indeed must) reject the claim

> that the Liars are true while also rejecting the claim that they are not true (thereby avoiding the Strengthened Liar). (Soames 1999: 176)

4.4.2. The Revenge of the Liar

As indicated in the quotation above, Soames argues that his interpretation can avoid the problem of the revenge of the liar. Recall the revenge problem for Kripke's theory:

(β) (β) is untrue.

If, according to Kripke, it is undefined, then it follows that this sentence is either false or undefined. From this, someone argues that the liar sentence turns out to be true. Therefore, the liar returns on a higher level. However, such a problem only can arise when 'undefined' is treated as a third truth value, and Kripke has not provided a detailed argument for why it is not. If, as Soames argues, 'undefined' means the absence of truth or falsity condition, and thus not a third truth value, then one cannot meaningfully ascribe any truth value to sentence (β).

Can this language express its 'untrue' predicate? Following the reasoning above, Soames argues that it can. Since 'untrue' does not equate to 'false or undefined', the revenge of the liar is thus blocked. Therefore, the liar sentence will not cause any contradiction to this language. In Kripke's interpretation, the expressive power of the language is limited because it cannot express its own 'untrue' predicate. According to Soames' interpretation, however, the strengthened liar sentence cannot cause any inconsistence in the language, so that there is no problem in expressing the predicate ‘untrue’ in that language.

4.5. Problems with Soames’ Interpretation

Soames’ interpretation relies heavily on the analogy between the non-semantic predicate ‘smidget’ and the semantic predicate ‘true’. According to Soames, both of them are gappy predicates which are defined by linguistic conventions, and both can occur in some sentences which express propositions with no truth value. If, however, the case of ‘true’ is significantly different from the case of ‘smidget’, then the force of Soames’ argument is questionable. In what follows, I shall examine several important dissimilarities between these two cases.

4.5.1. Partially Defined Concepts *vs*. Circular Concepts

As pointed out by Gupta (2002), there could be two senses provided for the phrase ‘partially defined’. The first sense is the one in the example of Smidget. That is, a predicate is partially defined when its definition is incomplete for a range of objects by *explicit* linguistic conventions. We may consider these terms as ‘partially defined’ in the strong sense. On the other hand, someone may say that 'true' is partially defined in the sense that neither 'true' nor 'not true' is correctly ascribed to some objects (e.g., the liar). But there is no explicit rule defined for the word ‘true’. The latter could be called ‘partially defined’ in the weak sense. As Gupta has pointed out, ‘the thesis that truth is partially defined, if understood in the weak way, provides no explanation of the puzzling behaviour of truth.' (Gupta 2002: footnote 10) This means, the fact that the word 'true' is partially defined in the weak sense only amounts to telling us the result that there are some problems in some cases when we use this word. It by no means tells us the underlying causes for why these problems arise.

Soames may argue that the predicate 'true' is partially defined by our linguistic conventions, just like the predicate 'smidget' is. If that is the case, then 'true' is not only partially defined in the weak sense, but also in the strong sense. To see whether this is the case, we have to examine the linguistic conventions that Soames has provided which are supposed to explain how the predicate 'true' is been defined.

Let us consider the most basic case for how positive atomic sentences can be defined as 'true'. Soames gives such conventions as the following:

> 3b. For any n-place predicate P and terms $t_1, \ldots, t_n$, $\ulcorner Pt_1, \ldots, t_n \urcorner$ is true (not true) ≡ P applies (does not apply) to the n-tuple $<o_1, \ldots, o_n>$ of referents of the terms.

The definition above gives the instruction for how to use the predicate 'true', and such instruction is given based on another notion 'apply'. Thus, if one is a beginner of this language, one may want to know the meaning of 'apply'. Soames' explanation for the convention for the use of 'apply' is as follows:

> 3a. The predicate ‘red’ applies (does not apply) to an object ≡ it is (is not) red. The predicate ‘smidget’ applies (does not apply) to an object ≡ it is (is not) a smidget. (And so on, one clause for each predicate in the language)

It should be noted that the definition of the word ‘apply’ is given by a schema, which has infinitely many cases as its instances. It is totally defined only when all the instances of the schema have been given. In other words, if one wants to know the meaning of ‘apply’, then one has to understand, in each case, how to use the word ‘apply’ in that case. Accordingly, for the predicate ‘true’, since it is also a predicate, one also has to know when one can say that this predicate applies to an object. Thus, one needs to know something like this:

> (5) The predicate ‘true’ applies (does not apply) to an object ≡ it is (is not) true.

As discussed in Chapter 3, the primary truth bearer is a proposition, so ‘the object’ here should be a proposition. But by our convention, the predicate ‘true’ also can apply to a sentence in a derivative sense. Since any proposition should be expressible by a sentence, and Soames uses ‘sentences’ as truth bearers throughout his book, we may follow his usage and say that the object mentioned in (5) is a sentence. Furthermore, since we only consider atomic sentences here, we can say that such sentence should have a form like $\ulcorner Pt_1, \ldots, t_n \urcorner$. Consequently, the above condition becomes:

> (6) The predicate ‘true’ applies (does not apply) to $\ulcorner Pt_1, \ldots, t_n \urcorner$ ≡ $\ulcorner Pt_1, \ldots, t_n \urcorner$ is (is not) true.

Putting (6) and (3b) together, we obtain:

> The predicate 'true' applies (does not apply) to ⌜$Pt_1, \ldots, t_n$⌝ ≡ the predicate 'true' applies (does not apply) to the n-tuple $<o_1, \ldots, o_n>$ of referents of the terms.

This shows that if we use one semantic notion (i.e. 'apply') to explain another semantic notion (i.e. 'true'), this will not bring us much further forward. Since the definition is essentially circular, a beginning learner of the language still does not know how to use the predicate 'true'. On the other hand, the definition for the predicate 'smidget' does tell us something about the use of that predicate. In this case the conventions given for the two predicates are essentially different, and thus it is problematic to say that the predicate 'true' has been partially defined in the sense that 'smidget' has been defined. However, this conclusion causes a serious problem for Soames' theory, since all the advantages this theory has are based on the analogy between these two cases. If the analogy breaks down, then it is very questionable to what extent Soames' interpretation can solve the problems of truth gap theory.

4.5.2. Does the Liar Sentence Express a Proposition?

Soames insists that the liar sentence expresses a proposition, just as does the sentence 'Mr. Smallman is a smidget'. According to his theory, the problem in both cases is that the proposition cannot receive a truth value on some occasions, since both predicates were introduced into English in a 'gappy' way.

Consider the smidget example. 'Mr. Smallman' is a name for a person. Since Soames advocates the theory of rigid names, this name refers to the same person, i.e. Mr. Smallman, in different possible worlds. He could be shorter or taller than the way he actually is, which would make the proposition expressed by the sentence ‘Mr. Smallman is a smidget’ receive a value ‘true’ or ‘false’. In the situation that his height is within the intermediate range, the proposition would have no truth value. No matter how the appearance of this man changes, it is always this same person, and the same concept ‘smidget’ under consideration. Therefore it is the same proposition that is under evaluation for truth. Consequently, we can conclude that the sentence ‘Mr. Smallman is a smidget’ does express a proposition, even though in some situations this proposition cannot have a truth value.[5]

On the other hand, the case for the liar sentence seems to be quite different. Soames insists that ‘it is simply not true that all liar sentences, truth teller sentences, and other ungrounded sentences fail to express propositions at the minimal fixed point.’(Soames 1999: 193) based on an example such as the following:

15. Some sentence written in place P is not true. (Soames 1999: 193)

[5] In the case of ‘smidget’, the truth value gap results from artificial definition. However, there are also many words in English that may cause truth value gaps for sentences where they occur, but that is not due to explicit definition like that for ‘smidget’. Examples of these words can be vague terms in English. As I will argue in the next chapter, though the gaps caused by vague terms are not due to explicit definitions, they nevertheless have the same status with ‘smidget’, in the sense that the corresponding sentence still expresses a proposition.

He argues, ‘Whether or not this is a liar sentence depends on which, if any, sentences are written in place P’ (Soames 1999: 193). However, in saying so, it seems that Soames has mixed up two things: sentences and propositions. He treats the meaning of a sentence as “a compositional function from contexts of utterance to propositions expressed” (Soames 1999: 168). Since the phrase ‘sentence written in place P’ is a definite description, it refers to different objects (if any) in different situations. Therefore, what sentence (15) expresses in different contexts will be different propositions, instead of different sentences. It is true that, depending on which sentence is written in place P, sentence (15) could be a paradoxical or a normal sentence. And if it is a normal one, then it does express a proposition. But from this it does not follow that the liar sentence does express a proposition. The example of sentence (15) only shows that some sentence in some context could be the liar sentence (which may be called ‘the contingent liar’), not that the (essential) liar sentence in some (or all) contexts expresses a proposition.

Also, when sentence (15) is expressed in a context where we know it is a liar sentence by looking at empirical facts, it is simply a ‘contingent liar’. As explained in Chapter 3, contingent liar sentences are not as important as ‘essential liars’. If we consider an ‘essential liar’ sentence, such as (α), then how do we show that it expresses a proposition, and what that proposition is? Soames does not deal with this problem. It is clear what is denoted by the name ‘Mr. Smallman’, but what is denoted by the name ‘(α)’? One may answer that it is the sentence (α) denoted by this name. But we have argued that what is in question here is the truth bearer, and a sentence is a truth bearer only in a derivative sense. Therefore, the name '(α)' should refer to something that is a 'primary truth bearer', i.e. a proposition. However, it is far from clear which proposition is denoted by the name '(α)'. In sum, to argue that an essential liar sentence such as (α) does express a proposition, firstly one has to show what the proposition expressed by the liar sentence is, which involves identifying the proposition denoted by the name '(α)'. But this can hardly be achieved, because the liar sentence is 'intuitively empty'.

In the next chapter, I shall argue in detail why a liar sentence such as (α) is intuitively empty and cannot express a proposition. For our argument here, since the gappy sentence which contains the predicate 'smidget' still expresses a proposition while the essential liar sentence (α) cannot, it follows that the truth predicate is fundamentally different from Soames' 'smidget'. Therefore, the analogy between the truth predicate and Soames' 'smidget' breaks.

Chapter 5: Diagonalization and the Functional-deflationary Conception of Truth

In Kripke's theory of truth, the truth predicate *T* of the language he has constructed is partially defined by a pair (S_1, S_2) of disjoint subsets of a nonempty domain, where 'S_1' is the extension of *T*, and 'S_2' the anti-extension of *T*. S_1 and S_2 are mutually exclusive but not jointly exhaustive, leaving some truth value gaps in this language. Sentences which do not belong to $S_1 \cup S_2$ at the minimal fixed point are called 'ungrounded' sentences, among which we find the liar sentence, which says of itself that it is not true.[1] We may think that S_1 is the set of all true sentences in that language, while S_2 the set of all false sentences in that language.

Kripke's construction has been criticized by many authors. In particular, Simmons (1990, 1993) analyzes the heterological paradox as a diagonal argument, and shows the flaw in Kripke's construction by using a simplified model of that language. In this chapter, I first analyze Simmons' argument and point out the deficiency in his analysis. Second, I revise his model to argue that the truth gap approach can provide a satisfactory treatment for the heterological paradox. From Section 4 to Section 7, I provide a functional-deflationary interpretation of truth, so that the nature of truth gaps can be explained, and problems such as the revenge of the liar can be treated properly within the truth gap approach. At the end of this chapter, I argue that another leading approach to

[1] In this chapter and thereafter, if there is no specific indication, I use 'the liar' to refer to essential liar sentence.

the liar paradox, contextualism, fails to deal with some important intuitions associated with the notion ‘truth’.

5.1. A Model for the Heterological Paradox

Simmons’ argument begins with the heterological paradox, which he identifies as ‘a bad diagonal argument related to the Liar’[2] (Simmons 1990: 288). A predicate is heterological if and only if it cannot apply to itself, while ‘autological’ means that a predicate applies to itself. To have a uniform terminology, we may understand the predicate ‘is heterological’ as ‘is not true of itself’, and ‘is autological’ as ‘is true of itself’, so that both of these two predicates are related to the truth predicate *T*. (In what follows, I use ‘*Het*’ to stand for the heterological predicate defined in this way, and ‘*Aut*’ for the Autological predicate.) Accordingly, let us consider a language **L**, which is a simplified version of English. Consider all the 1-place predicates in **L**. Since a predicate in English only contains finitely many letters, all the monadic predicates in **L** can be listed in an array according to their alphabetic order (see Array 1 below).

Let the side and the top of Array 1 be the ‘totality’[3] of all ordinary 1-place predicates of English. By ‘ordinary’, I mean predicates such as ‘is red’, ‘is

[2] In Simmons’ terminology, ‘any diagonal argument assumes the existence of a number of components: a side, a top, an array, a diagonal, a value, and a countervalue. In a bad diagonal argument, one or more of these sets is not well-determined.’ (Simmons 1993: 29) I will explain the relation between the heterological paradox and the liar paradox in Section 4 below.

[3] Simmons uses the term ‘set’ instead of ‘totality’, but I want to distinguish these two terms and reserve the word ‘set’ for use in its strict, technical sense. As I have clarified in Chapter 2 (footnote 2), when I use ‘totality’, this means either that it is not a set or that it awaits proof that it is a set.

hardworking', etc, which depict the subject as having some particular property. Intuitively, we may say that these predicates represent the subject as being a certain way. So we may call these predicates as 'representational'. On the other hand, predicates which denote certain semantic properties, such as 'is true', 'is not true', 'is heterological', 'is autological', do not directly represent the world as being a certain way. Instead, they describe the *relation* between signifiers, like words, phrases, signs, and symbols, and what they stand for. Therefore, though they are all 1-place predicates, let us exclude them from both the side and the top of Array 1 temporarily. The distinction between representational predicates and non-representational predicates will be explored in full detail below. Here let us temporarily be content with this intuitive understanding.

On Array 1, for each 1-place predicate P_i, we can decide which value we should assign for the box <P_i , P_j> by considering the relation between P_i and P_j. If the predicate 'P_i' is true of the predicate 'P_j', then for the box <P_i , P_j> we enter '1'. If the predicate 'P_i' is not true of the predicate 'P_j', then we enter '0' into the box <P_i , P_j>.

Array 1

	P_1	P_2	P_3	...	P_i	P_j	..
P_1	1	0	1	…			
P_2	1	0	0	…			
P_3	0	1	0	…			
…	…	…	…	…			
P_i					0	1	..
…						…	

Usually we should expect that for a given cell, there is either a '1' or a '0' in it, which means that the corresponding sentence ('P_j is P_i' or '"P_i" is true of "P_j"') is either true or false. However, since **L** is supposed to contain all the monadic predicates in English, there are some complexities in this issue. It is possible that P_j is not within the range of application of P_i. For example, let P_i be 'is sleepy', and 'P_j' be 'is a prime number'. Then for <P_i , P_j> we have the following sentence:

(1) 'Is a prime number' is sleepy.

We can hardly say that the predicate 'is a prime number' is sleepy, because normally the predicate 'is sleepy' is supposed to apply to an animal (especially a human being). In this case, we may want to put 'n.a.' into the box <P_i , P_j>. But we should do this with caution. We may understand 'n.a.' as 'P_i is not defined for P_j'. However, this is different from the 'undefinedness' for the truth value of the liar in truth gap theories. This is because there are some implicit linguistic conventions guiding the usage of these ordinary predicates, so that a competent English speaker knows that 'is sleepy' applies to an animal, while 'is a prime number' applies to a number. It is not difficult for an ordinary English speaker to find out the problem in sentences such as '"is a prime number" is sleepy'. Even if sometimes there may be some confusion at the beginning, the problem can be fixed quite easily. Consider, for example,

(2) 'Is a prime number' is black.

One may think that this sentence is as nonsensical as (1), so she will put ‘n.a.’ in the box accordingly. But another person may think Sentence (2) is true. There is no contradiction involved in this case because the first person understands the subject term of sentence (2) as a name for a predicate, and it is nonsense to say that a predicate (in the abstract sense) is black. However, the second person may think the subject of Sentence (2) is the phrase in quotation marks, and this phrase is indeed black in color. There is no contradiction involved, and a competent English speaker can immediately recognize these differences based on implicit linguistic conventions.

Therefore, based on linguistic conventions, a competent speaker knows that usually the predicate ‘is sleepy’ is not applicable to the predicate ‘is a prime number’. But there are no such linguistic conventions available to rule out cases such as the liar or the heterological sentence:

(α) Sentence (α) is not true.

(3) ‘Is heterological’ is heterological.

By linguistic conventions, the truth predicate ‘is true’ and its negation ‘is not true’ normally apply to a sentence, and ‘Sentence (α)’ is indeed a name for a sentence. Similarly, ‘is heterological’ normally applies to a 1-place predicate in English, and ‘is heterological’ is indeed such a predicate in English[4]. If we do not assume any expertise in philosophy, but simply rely on an ordinary speaker’s knowledge

[4] One may argue that *Het* is not a predicate in English. I will deal with this problem in the next section.

of English, there is nothing wrong with either (α) or (3). This shows a fundamental difference between the 'undefinedness' for paradoxical sentences (e.g. (α) and (3)) and the inapplicability of the predicate in sentences such as (1), which can easily be identified by one's understanding of linguistic conventions.

To facilitate our discussion, let us weed out those 'n.a.' sentences which can be easily recognized as ill-formed based on linguistic conventions, and consider an ideal model where each cell is bivalent.[5] That is, we enter either '1' or '0' into each box $<P_i , P_j>$. In this case, the array is given by the following function:

$$\mathrm{R}\ (x, y) = \begin{cases} 1, \text{ if } x \text{ is true of } y, \\ 0, \text{ if } x \text{ is not true of } y. \end{cases}$$

In Chapter 2, I introduced Simmons' definition for the diagonal, the value and the countervalue on an array. He defines 'the diagonal' as a 1-1 function F from the side of the array to the top, and 'the value of the diagonal' as a set of ordered triples based on the diagonal function F. Intuitively, we can understand 'the value of the diagonal' as the set of all the shaded cells (with their truth values) on Array 1, and 'the countervalue' as all these cells with the opposite truth values. Simmons then argues that the countervalue thus generated defines the predicate *Het*. He says:

[5] Simmons (1990, 1993) does not mention this aspect in his construction of the array. He simply assumes that each box in the array is bivalent.

> The countervalue is a determinate set of ordered triples, and the associated set of predicates of English is a determinate set of English predicates definable in terms of the countervalue. We can say that an English predicate P_i is a member of this set iff < P_i , P_i , 1> is a member of the countervalue. And we can talk about all this in English - indeed, that is just what we are doing. (Simmons 1993: 60)

But there is no need for the countervalue if one wants to define the totality associated with *Het*. The value of the diagonal F in Simmons' definition has already provided all the apparatus needed for both *Het* and *Aut*: *Het* is defined in terms of the totality of all P_i on Array 1 which have a '0' as the truth value for the box <P_i , P_i>, while *Aut* is defined in terms of the totality of all P_i on Array 1 which have a '1' as the truth value for <P_i , P_i>. It is not clear why we should rely on the countervalue to define *Het*. Simmons may want to make an analogy between *Het* and the countervalue in Cantor's proof discussed in Chapter 2. However, for Cantor's proof, the new element which diagonalizes out of the list should be based on all the elements in the totality, while for *Het*, it only applies to a proper subset of the predicates on Array 1. The rest is covered by *Aut*.

Although his treatment is cumbersome, it is not a big problem for Simmons' argument. Thus, he continues, since *Het* is a predicate in English, and does have a non-empty extension, then as a 1-place English predicate, it should be listed as a row of Array 1 as well. Let us call that row '*Het*'. Then there should be a cell <*Het*, *Het*> on that row. If we want to fill in the value for this cell, then it leads to a contradiction, i.e. the heterological paradox. To say that the value of this cell is ‘undefined’ cannot fix the problem, because we can construct a new predicate ‘is false or undefined of itself’, and then the heterological paradox returns, i.e. the superheterological paradox.

5.2. Possible Solutions

There are several ways to respond to Simmons’ argument: (i) *Het* is not a predicate at all; (ii) it is not a predicate in language **L**; and (iii) though it is a predicate of language **L**, it does not express a concept.

The first kind of response is not very plausible because it is simply a fact that *Het* is a predicate in English. Also, it is an abbreviation of the phrase ‘is not true of itself’ or ‘does not apply to itself’, each of which is a grammatically well-formed verb phrase consisting of words in common use in English. Even if *Het* does not exist until the point we discuss it, we still can introduce this predicate into English (and **L**) pretty easily. After all, natural language is very flexible, and people create new words every day. What is more important, a similar argument can also be formulated about the truth predicate, and one can hardly deny that ‘true’ is a predicate in English. We have been using this word for thousands of years! Therefore, according to the criteria for a proper solution towards semantic paradoxes discussed in Chapter 3, this solution is highly counter-intuitive and thus cannot be a satisfactory solution.

A related idea is to say that semantic facts are not expressible.[6] Therefore, though there are semantic notions in English, they simply cannot express these semantic facts. In other words, statements about semantic facts are simply without sense. This is what has been suggested in early Wittgenstein's *Tractatus*, which tells us that facts about semantics can only be shown, but cannot be said. However, the reason why they cannot be said is because of the paradoxes: since to state them would cause paradox, they therefore cannot be said. But this response cannot promote our understanding of semantics. What is worse, this position itself is not consistent. By telling us that semantics cannot be said, it thus says something about the things which cannot be said. Based on these reasons, one may think that solutions following this approach are not good enough.

The second kind of response is to resort to Tarskian hierarchies. Thus, one may say that *Het* does not belong to language **L**, but belongs to some meta-language. Actually, this object-language/meta-language distinction is also endorsed by Kripke (1975). Though by allowing truth gaps, his language can contain its own truth predicate, he admits that it nonetheless cannot contain semantic notions such as 'grounded', 'paradoxical', etc.:

> Such semantical notions as "grounded", "paradoxical", etc. belong to the metalanguage. This situation seems to me to be intuitively acceptable; in contrast to the notion of truth, none of these notions is to be found in natural language in its pristine purity, before philosophers reflect on its semantics (in particular, the semantic paradoxes). If we give up the goal of

[6] This idea is related to the first point discussed above, i.e. that *Het* is not a predicate at all, in the sense that both of them deny that there is a certain semantic fact or concept associated with *Het*, as opposed to the third kind of solution below.

> a universal language, models of the type in this paper are plausible as models of natural language at a stage before we reflect on the generation process associated with the concept of truth, the stage which continues in the daily life of non-philosophical speakers. (Kripke 1975: 714, footnote 34)

However, if truth gap theorists need finally to resort to a Tarskian hierarchy to solve the problem, then the value of positing truth gaps for natural language is not clear. No doubt Kripke's construction has some advantage compared with Tarski's truth definition, especially since in Kripke's definition there is no need for a hierarchy of subscripted truth predicate *in* that language. However, as a model for natural language, it remains implausible to assume that there is some meta-language above natural language. Kripke has argued against Tarski's approach at the beginning of his paper, regarding the latter as pretty artificial. There is no sign that natural language is stratified in the way suggested by Tarski. By the same token, there is no sign that there is a 'meta-language' existing above natural language either. Moreover, Kripke says that such semantic notions as "grounded" and "paradoxical" are not a part of natural language. However, intuitively, 'paradoxical' is a natural language predicate and it seems that we understand its meaning as well[7]. On the other hand, for 'grounded', the reason that it cannot be part of Kripke's object language (which serves as a model of natural language) is simply that it can cause paradoxes. Thus, Kripke's idea about meta-language/object-language is implausible because firstly, it contradicts our

[7] For example, one possible definition for 'paradox' is: an argument which starts with apparently true claims and proceeds via apparently valid reasoning, while leading to a contradiction.

intuitive view about natural language: natural language is semantically universal in the sense that it can talk about its own semantics. Secondly, the distinction between ‘grounded’ and ‘ungrounded’ expressions has no other basis than the paradoxes themselves, which violates one criterion for a good solution discussed in Chapter 3.

The third kind of response can be found in Martin (1976). In this paper, Martin argues that though this predicate *Het* is expressible in English, there is simply no ‘concept’ expressed by this predicate. In other words, the gap is not at the level of language, but at the level of ontology, as he says:

> In the second case we deny that there are such concepts as we first supposed - that is, we propose conceptual reform, involving ontological rather than linguistic restrictions. The languages are viewed as capable of saying all there is to be said - we simply judge that there is less to be said than first thought. It is misleading, then, to speak in this second case of an ‘expressibility gap’; there is a gap or discrepancy, but it is between the situations before and after analysis. (Martin 1976: 287)

This solution retains our intuitive view of natural language as semantically universal, but it gives up the intuition that there is a concept of heterologicality expressed by the predicate *Het*. This solution is also flawed because it seems that these two intuitions are equally appealing, and it is not clear why one should prefer one while giving up the other. Furthermore, as argued by Simmons, there is a ‘set’ associated with *Het*, so it is not clear why there is no such concept as ‘heterologicality’. Martin may argue that it is undecidable whether *Het* itself should belong to this ‘set’ or not. This is true, but it is not an adequate reason to support his view according to the criteria for a satisfactory solution to the liar paradox which I specified in Chapter 3. The reason for the undecidability of heterologicality is simply that it will cause paradox. Thus again, the only reason to support Martin's conclusion that there is no such concept is that it will cause paradox. But, as argued above, there should be some independent reason to support a proposed solution to a paradox, rather than the paradox itself.

It seems that these solutions are all flawed because of failure to meet one or more of the criteria discussed in Chapter 3. In the next section, I shall propose a treatment for the heterological paradox and *Het* which does not sacrifice our intuitions about natural language.

5.3. The Dynamic Nature of the Heterological Predicate

The main point I wish to make is that the semantic notion *Het* is dynamic, and this dynamic characteristic is derived from the dynamic feature of the diagonal function of the array.

In Chapter 2, I discussed the essence of the diagonal. The diagonal is a 1-1 function from elements on the side to those on the top. It governs every element in the totality. By 'governs', I mean that the diagonal does not only pass through every existent row on the array, but also will pass through any newly generated rows. It is in this sense that the diagonal is 'dynamic'. Thus, these three terms are used interchangeably: *the diagonal*, *the diagonal function*, *the dynamic diagonal*. On the other hand, the diagonal should not be confused with *the value of the* *diagonal*, which is a fixed set of triple orders and can be treated as a row on the array. Also, the diagonal is important in achieving self-reference in diagonal arguments. I mentioned Simmons' definition for the diagonal function in Chapter 2. His definition, which only defines the diagonal as a '1-1 function from the side to the top', does not mention self-reference, and thus has missed one important feature of the diagonal. For our discussion of the heterological paradox, I shall provide a refined definition for the diagonal function (*Diag*) on an array where the rows are 1-place predicates and use this definition as the basis for my discussion below:

> **Definition 5.1** *Diag* is a diagonal on a diagonal array where the rows are 1-place predicates $\leftrightarrow_{\mathrm{df}}$ *Diag* is a 1-1 function from the side of the array to the top, and for any P_{i} on the side, $Diag(P_{\mathrm{i}}) = P_{\mathrm{i}}$.[8]

Thus, the diagonal function *Diag* on Array 1 is a function which maps each P_{i} on the side to this predicate itself on the top. Through these two indexes, we can find a cell $<P_{\mathrm{i}}, Diag\ (P_{\mathrm{i}})>$ to which we can assign a truth value. Let *T* be the function which assigns a value '0' or '1' to such cells[9]:

> **Thesis 5.1** For any P_{i} on the side of the diagonal array, there is a function *T* such that either $T<P_{\mathrm{i}}, Diag(P_{\mathrm{i}})>=0$ or $T<P_{\mathrm{i}}, Diag(P_{\mathrm{i}})>=1$.[10]

[8] This definition works for Array 1 and any expansion of Array 1 by adding more and more elements to the top of the array, as we shall see in Section 5.5 below.

[9] The nature of this function *T* will be explored in depth in the next section. Here let us be temporarily content with this relatively loose description.

[10] This thesis has assumed that the array is bivalent. As explained at the beginning of this chapter, we want to consider an ideal model for language **L** so that we weed out all those 'n.a.' sentences. Therefore, there is no big problem for such an assumption. Also, it should be pointed out that '*T*' is the function that assigns '0' or '1' to the cells on the diagonal array, not any proper subset of it. Otherwise, there would be infinitely many such functions *T*, and each *T* is for a subset of the set of cells on the array. The reason for this stipulation is that we want *T* to be the truth predicate in

As analyzed in Section 1, the extension of *Het* is the totality of all the P_i on the side which have a ‘0’ as the value for the box <P_i, *Diag*(P_i)>. Let ‘Ext(x)’ stand for the extension of x. Then the heterological predicate can be defined as follows:

> **Definition 5.2** For any P_i on the side of a diagonal array, $P_i \in$ Ext(*Het*) $\leftrightarrow_{df}$ T(<P_i, *Diag*(P_i)>)=0.

In Array 1 above, the problem is where to put *Het*. If we follow Simmons’ analysis that the predicate *Het* is defined by *the countervalue*, then it should be added as a row to Array 1. But if we put it as a row on the array, just like other ordinary 1-place predicates, then we do not know what the value for T(<*Het*, *Diag*(*Het*)>) is. Both ‘0’ and ‘1’ would lead to a contradiction. If, as suggested in the solutions discussed above, this word is not a predicate of this language, or it is not a predicate at all, then this solution is highly counter-intuitive. We can easily introduce and define it in language **L**, as I have just done. Also, there is no trouble for us to understand the meaning of this word. Therefore, one plausible solution is to acknowledge its place in language **L**, but deny that it has the same status as other ordinary 1-place predicates of **L**. In other words, the alternative solution is to acknowledge that *Het* still appears on Array 1, but deny that it is a row like other ordinary 1-place predicates P_i. Thus, a natural question is: if this predicate

English, thus T should be general enough to govern all (properly formed declarative) sentences in English, rather than just a portion of it. Consequently, when manifested on the array, we want T to apply to all cells on the array, rather than just a subset of these cells.

Het is not a row of the array, then where is it? The answer is: it is defined based on the diagonal function of the array.

In Chapter 2, I argued that *Diag* is a function which has the potentiality to govern any row of the array. Since the definition of the predicate *Het* is based on *Diag*, it thus is also dynamic. As a function, *Diag* should be distinguished from the value of the diagonal (i.e. all the shaded cells in Array 1). The latter could be represented as a row of the array, but it is impossible to represent the diagonal function as a row of the array. Correspondingly, the heterological predicate cannot be represented as a row either.

Definition 5.2 may give one the impression that the extension of *Het* is a fixed set, but this is not the case. What this definition has told us is that, for any given predicate P_i, we can use it to determine whether P_i belongs to the extension of *Het* or not. It is by no means follows that the extension of *Het* is a fixed set; otherwise we would have a hierarchy of heterological predicates, rather than a single one.

If we treat *Het* in the fixed way, then for Array 1, we have a Het_1, whose extension is the *set* (not ‘totality’) of all the P_i such that $T(<P_i, Diag(P_i)>)=0$. Thus,

$Het_1 = \{ P_2, P_3, \ldots, P_i , \ldots \}$

But since this Het_1 is also monadic, we should add it to both the side and the top of Array 1 to obtain a new array 1*. There is also a cell $<Het_1, Diag(Het_1)>$ on the row Het_1. We should put '0' as the value for this cell, because the predicate Het_1 is not included in the set Het_1 defined above. Then, for this new array 1*, we obtain a new predicate Het_2, whose extension is the set Het_2:

$$Het_2 = \{ P_2, P_3, \ldots, P_i , \ldots, Het_1, \ldots\}$$

Following the same line, we should put '0' into the cell $<Het_2, Diag(Het_2)>$, so that a new monadic predicate Het_3 can be generated, and then we have a third array 1** and we should put '0' for the cell $<Het_3, Diag(Het_3)>$ as well… This process will keep on going, and the result is that we cannot have a unique heterological predicate *Het*, but we have a hierarchy of predicates: Het_1, Het_2, Het_3, … This can avoid contradiction. But, as argued in Section 2 above, it contradicts our intuitive understanding about natural language that there is a unique heterological predicate *Het*.

Therefore, *Het* is defined based on the diagonal function of the array, rather than on the value of the diagonal (or the countervalue): this is exactly the difference between my treatment of *Het* and Simmons'. As defined by the function[11] *Diag*, *Het* is thus a dynamic notion, in the sense that, no matter what predicate has been added to the side as a row, it is then covered by *Het*. But *Het* itself is not a row. In other words, Array 1, which is a simplified model of natural

[11] As clarified in Footnote 10 in Chapter 2, when I say that the *Diag* function is dynamic, it is a non-standard way to use the word 'function'.

language, is *indefinitely extensible*, continually expanding to cover more and more new predicates. The predicate *Het* varies and extends itself with the array as well, since for any new P_i, if it satisfies that $T(<P_i, Diag(P_i)>)=0$, then P_i should be included in the extension of *Het*. Since this interpretation avoids the hierarchy of heterological predicates, it reflects our intuitive understanding of natural language that there is only one heterological predicate in English.

The dynamic nature of *Het* shows its special status compared with ordinary 1-place predicates in language **L**. Its status as defined based on the diagonal function reveals its role in our linguistic system. *Het* is not fixed by any set of cells, which is the way that other 1-place predicates are presented (e.g. P_i can be represented by a set of cells on the row of P_i[12]). We may understand each cell on Array 1 as *representational*,[13] in the sense that it describes the world as being some way, and is true if the world is that way. In Section 1, I mentioned the intuitive distinction between 'representational predicates' and 'non-representational predicates'. Now we can define a 1-place predicate as being 'representational' in a more rigorous way:

> **Definition 5.3** If a 1-place predicate P_i is fixed by a row of cells on a diagonal array, then such a predicate is called 'representational'.

[12] Strictly speaking, this is not quite right, since the row gives only the extension and anti-extension among the set of predicates. It does not mention other objects to which the predicate applies and to which the predicate does not apply. However, in principle, P_i can be represented by a row of the array by including more and more kinds of objects on the top. In Section 5.5, I shall show how the array can be expanded to include names for sentences and individuals.

[13] Each cell on the array corresponds to a sentence, and every cell is representational. However, it is not the case that every sentence in this language is representational.

Admittedly, there are some controversies about this term. For example, according to my analysis, statements in mathematics and logic are also representational, as are predicates in these areas. But it is arguable in what sense they could be said to be ‘representational’. Nonetheless, it is beyond the scope of this thesis to discuss the metaphysical status of mathematical and logical objects. Therefore, I shall put this issue aside, and shall mainly consider ordinary sentences and predicates.

It is a notable feature of *Het* that it is not representational. *Het* is defined by the function *Diag*. As a function, *Diag* is not representational. Since the predicate *Het* is defined essentially based on such a function, it is not representational either. This result can be generalized; thus we have the following thesis:

> **Thesis 5.2** If x is defined as a function, or defined based on a function, on Array 1, then x is not representational.

I shall defend this thesis further in the next section, but here the intuitive idea is that, if an expression is representational, then it tells us something about the world. On the other hand, if an expression is a function, or defined based on a function, on the array, then it tells us something about other linguistic expressions, i.e. it is not directly about the world, but about cells or rows of cells on the array. For example, when the function *Diag* is applied to P_i, it maps P_i to itself on the top, so that we can have a cell <P_i, *Diag*(P_i)>. This is very important information, since it tells us that P_i is an ordinary 1-place predicate, and occurs as a row on Array 1. The heterological predicate, which is based on *Diag*, also tells us something about those representational predicates. If a predicate P_i is heterological, we know that there is a certain relation between P_i and itself. Analyzed in this way, we may think that *Het* should be construed as a second level notion. Although it itself is also a 1-place predicate, it is actually *about and governing* all other ordinary 1-place predicates. This is what Ryle suggests:

> 'Self-epithet' and 'non-self-epithet' convey no philological information about words. They are specially fabricated instruments for talking *en bloc* about the possession or non-possession by philological epithets of whatever may be the philological properties for which they stand. (Ryle 1951: 66)

If we recognize the non-representational nature of *Het*, then it is clear that this language **L** can contain its own *Het* predicate without leading to a contradiction or to a hierarchy of heterological predicates. The contradiction can only arise when we forget about this feature of *Het* and thus treat it as a row of Array 1. In that way, there is a cell <*Het*, *Diag*(*Het*)> on the row for which we do not know how to assign a truth value. But as argued above, since *Het* should not be treated as a row, then there is no such cell <*Het*, *Diag*(*Het*)> to which we can assign a value. This explains why truth gap theorists insist that there is a truth value gap for this sentence:

(3) 'Is heterological' is heterological.

The truth value gap for this sentence is not due to some artificial stipulation, but due to the systematic feature of our natural language. There is no cell <*Het*, *Diag*(*Het*)> for which we can assign a truth value, which means that sentence (3) is not a proper candidate for truth evaluation. This reminds us of the distinction between sentences and propositions, but I shall leave the issue about the nature of truth gaps to Section 5 below, after I have discussed the truth predicate *T*.

On the other hand, as based on a function on Array 1, the predicate *Het* is contained in language **L**. Therefore, we do not need any meta-language to express the heterological predicate of **L**. We may consider Wittgenstein's 'standard meter' example for an analogy. In *Philosophical Investigations* §50 Wittgenstein writes:

> There is one thing of which one can say neither that it is one metre long, nor that it is not one metre long, and that is the standard metre in Paris. – But this is, of course, not to ascribe any extraordinary property to it, but only to mark its peculiar role in the language-game of measuring with a metre-rule. (Wittgenstein 1953: §50)

We need the standard meter to measure and tell us the length of a particular object, just as we need functions and semantic notions which are based on such functions to tell us the property of each element (i.e. cells, predicates, etc.) of Array 1. There is no doubt that the standard meter also has a length, but it is inappropriate to ask whether it is one meter long or not. Similarly, the *Diag* function and the semantic notion *Het* are also in Language **L**, but it is inappropriate to treat them also as a row of the array. Consequently, it is inappropriate to ask whether the heterological predicate is heterological or not. This analysis confirms another intuition about natural language: it is semantically universal. Natural language can contain and actually contains its own semantic notions. There is no meta-language for a natural language such as English. All that one needs to do is to recognize that natural language is extremely flexible and indefinitely extensible. Correspondingly, its semantic notions are dynamic, and truth value gaps caused by semantic notions are due to their dynamic nature.

5.4. The Functional-deflationary Conception of Truth

So far everything I have said concerns the heterological predicate *Het* and the heterological paradox. Now how about the Truth Predicate *T* and the liar paradox? As I have argued above, *Het* is defined based on the diagonal function. By the same token, the autological predicate *Aut*, which is intuitively understood as ‘is true of itself’, can also be defined based on the diagonal function:

> **Definition 5.4** For any P_i on the side of a diagonal array, $P_i \in \mathrm{Ext}(Aut) \leftrightarrow_{\mathrm{df}} T{<}P_i, Diag(P_i){>}{=}1$.

Though the truth value assignment for the following sentence does not lead to a contradiction, this sentence nonetheless is pathological:

(4) ‘Is autological’ is autological.

Following the treatment of Sentence (3) in the previous section, we can say that there is no such cell as <*Aut*, *Diag*(*Aut*)> on Array 1 either, so there is no proposition expressed by sentence (4). In other words, *Aut* is not representational either, and Sentence (4) also suffers from truth value gaps, because of the dynamic nature of the semantic notion *Aut*.

Comparing Definition 5.4 with Definition 5.2, we notice that both of them rely on the diagonal function *Diag*. But they also both rely on another function: *T*. In the previous section, *T* is loosely construed as the function which assigns value ‘0’ or ‘1’ to cells such as <P_i, *Diag*(P_i)>. One may immediately recognize that *T* is related to the truth predicate. But construed as such, *T* is a function. According to Thesis 5.2 proposed in the previous section, if something is a function on the array, then it is not representational. Consequently, we may wonder whether this thesis still holds for *T*.

As a 1-place predicate, it is natural to regard *T* as a row of Array 1, just like other 1-place predicates P_i. But this is a misunderstanding of *T*, and we will see that it easily leads to the liar paradox. If, on the other hand, we understand *T* as a function on the array, then just like *Diag* and *Het*, it cannot be fixed by any row of cells. Nonetheless, *T* is still contained in language **L**, so that we do not have to ascend to a metalanguage to talk about *T*. This may sound like a good solution to the liar paradox, but one may still hesitate to accept all these claims, since one may suspect that it is simply a stipulation that *T* is a function on the Array and that a function cannot be a row. To dispel such doubts, in what follows I shall develop a kind of deflationary conception of truth, which I call ‘functional-deflationary conception of truth’, so that it can provide reasonable support for the claims about functions on Array 1 and the validity of Thesis 5.2.

Kripke (1975) himself has not mentioned any specific theory about truth (e.g. the correspondence theory of truth, the coherence theory of truth, the deflationary theory of truth, the pragmatic theory of truth, etc.). In his truth definition, which is constructed in a formal language, a truth value assignment for ordinary sentences which do not contain semantic concepts is simply one part of the normal interpretation of that formal language. However, as he also intends to take this formal language as a model for natural language, there should be some explanation for how we assign truth values to ordinary sentences when we try to explain the predicate 'true' in natural language. Kripke describes the intuitions behind this process as follows: 'we may say that we are entitled to assert (or deny) of any sentence that it is true precisely under the circumstances when we can assert (or deny) the sentence itself' (Kripke 1975: 701). This description has a remarkable resemblance to Tarski's T-convention:

(T) *S* is true iff *p*.

where '*S*' is a name for *p*. It remains unclear, however, what the role of the T-convention in Kripke's truth definition is. It could, as in Tarski's theory, serve as a requirement that a materially adequate definition of truth must meet. Although this convention is not by itself committed to any specific theory of truth, it could be developed into a correspondence theory of truth in terms of the concepts of reference and satisfaction, as has been shown by Field (1972). It could also, as Horwich (1998, 2001) argues, be considered as a full-fledged theory of truth, i.e.

all that one can say about the conception of truth is the T-convention. Horwich's view, which he calls 'minimalism', is a deflationary theory of truth. There are many authors debating the issue of whether the Truth Gap Approach for the liar paradoxes is compatible with the deflationary understanding of truth (especially in the form of Horwich's minimalism).[14] Although their answers to this question are mainly negative (because their targets are usually minimalism), it nonetheless should not prevent us from exploring whether there are other forms of deflationism which can be integrated into the truth gap approach.

'Deflationism' is a general label for a group of views about truth. On the one hand, as Horwich's minimalist theory suggests, the biconditional in the instances of the T-convention is all there is to say about the conception of truth. It is a long tradition that truth theorists think there is some 'underlying nature' of truth which stubbornly resists philosophical elaboration, but Horwich argues that 'there is simply no such thing' (Horwich 1998: 5). According to his view, there is nothing mysterious or special about the word 'true'. On the other hand, one may think that a T-convention does reveal a non-trivial, special status for the word 'true'. In his argument for the undefinability of truth, which is also an argument against the correspondence theory of truth, Frege says:

> And any other attempt to define truth also breaks down. For in a definition certain characteristics would have to be specified. And in application to any particular case the question would always arise whether it were true that the characteristics were present. So we should be going round in a

[14] For example, Glanzberg 2003c, Holton 2000.

> circle. So it seems likely that the content of the word ‘true’ is *sui generis* and indefinable. (Frege 1918-9: 353)

Frege’s view is usually characterized as the redundancy theory of truth, another form of deflationism which regards the truth predicate as totally empty and redundant. However, it is doubtful whether this interpretation is correct. Instead of saying that the content of the word ‘true’ is completely empty, Frege says that its content is *sui generis*. In a reflection on this issue, he asks, ‘may we not be dealing here with something which cannot be called a property in the ordinary sense at all?’ (ibid. 354-5). This suggests one possible way to interpret Frege’s words ‘*sui generis*’, that is, the truth predicate T is *categorically* different from other ordinary 1-place predicates P_i. In the argument quoted above, Frege seems to suggest that the understanding of truth has been presupposed in any assertion in a language. For example, suppose we use ‘correspondence’ to define truth, say, p is true iff p corresponds to some existing state of affairs. Then, Frege wants to argue that in order to assert ‘p corresponds to some existing state of affairs’, one should already have the understanding of the concept ‘true’. Furthermore, this argument can be generalized. For any characteristic which is supposed to define the word ‘true’, one cannot assert that the claimed characteristic *is* a characteristic of ‘true’ without an understanding of the notion ‘true’ beforehand. That is why Frege thinks that any definition for this word must be essentially circular. If this understanding is reasonable, then we may say that it is because of the understanding of T that any assertion and truth value assignment for a sentence become possible. In this sense, truth is *conceptually prior* to all ordinary notions.

This is why Frege concludes that this word cannot be defined in turn by other notions. Understood in this way, his view is clearly different from the redundancy view about truth. On the contrary, the predicate *T* is better construed as a function which assigns truth values to each assertion in this language. This confirms what I have suggested in Thesis 5.1:

> For any P_i on D_1, there is a function *T* such that either $T<P_i, Diag(P_i)>=0$ or $T<P_i, Diag(P_i)>=1$.

Moreover, *T* not only governs the value assignment for cells related to the diagonal, but also governs all the cells on the array. As discussed in Chapter 2, one essential feature of the dynamic diagonal is that it passes through every row of the array (i.e. every ordinary predicate in this language). For each predicate P_i, there is a cell $<P_i, Diag(P_i)>$ on the row P_i. According to Thesis 5.1, the function *T* assigns '1' or '0' to such cells. But it is not the case that *T* only governs one cell on each row. When we know how to assign a value for *one* cell on the row labelled by 'P_i', it means that we know what it is like for P_i to apply to an object, and then we *can* assign value to *any* box on the row P_i.[15] That is to say, being able to assign one value means that we have understood the condition for how to apply

[15] Admittedly, here it is arguable in what sense we say that one 'knows' what it is like for P_i to apply to an object. For example, one may know that the proposition expressed by the sentence "'is round' is round" is false, since she knows that 'is round' signifies having a shape of some sort and that 'is round' is not the sort of thing that has a shape. But she may not yet have learned what sort of shape is roundness. However, for our argument here, I want to leave the epistemology issue aside and use the word 'know' in a stronger sense, i.e. the person knows the truth condition of the relevant sentence. Therefore, for a person to know that "'is round' is round" is false, it is not enough to know only that the predicate 'is round' is not a kind of shape so that the sentence is false, but she should also know that what sort of shape is roundness.

this predicate 'P_i' to any given object.[16] Thus the understanding of *T* is tantamount to the understanding of the condition for any predicate on the array, whether it is applicable to a given object or not. This is why Frege treats *T* as a primary concept, which is *conceptually prior* to any ordinary predicate.

But, as a function on the array, *T* is not representational, which means that it cannot be construed as any row of cells on the array. Compare the following two sentences:

(α) Sentence (α) is not true.
(5) Sentence (5) is an English sentence.

Both of them are self-referential, grammatically well-formed and meaningful sentences according to our linguistic conventions, while (5) is true and the other causes paradox. This is because the predicate 'is an English sentence' is representational, while 'is not true' is not. Correspondingly, Sentence (5) is representational, while (α) is not. One may wonder whether any sentence which contains some non-representational predicate is thus also non-representational. However, this is not the case. To decide whether such a sentence is representational or not, we should examine the role of non-representational predicates in such sentences further. Consider, for example:

(6) The first sentence written on this page is not true.

[16] There are some special cases, though. One is concerning vague predicates, like 'heap', 'bald', etc. I shall deal with the issue of vagueness in the next chapter, and show the difference between vagueness and semantic paradoxes.

Sentence (6) is known as a ‘contingent’ liar sentence. Though it contains the predicate ‘is not true’, it nonetheless still could be representational. When its subject term (i.e. ‘the first sentence written on this page’) denotes a sentence other than itself or the liar sentence (α) or some other form of the liar, then it can be rewritten in a way that it does not contain the predicate ‘is not true’. For example, suppose it denotes Sentence (5). Then (6) can be rewritten as ‘Sentence (5) is not an English sentence’, which is a normal sentence and representational. On the other hand, if the subject term happens to denote sentence (6) itself or the liar sentence (α), then it cannot be rewritten in a way that does not contain the semantic notion ‘true’. This sentence is consequently non-representational.

In Chapter 3, I mentioned the distinction between sentences and propositions. Sentences, although they are grammatically correct and meaningful, do not necessarily express propositions. Here, we may understand ‘propositions’ as truth bearers, no matter what kind of metaphysical status they have. Thus, from the discussion about representability above, we may understand truth bearers as follows:

> **Definition 5.5** For any *p*, *p* is a truth bearer $\leftrightarrow_{\text{df}}$ *p* is the content of a well-formed declarative sentence that is representational.

By ‘well-formed’, I mean that the sentence is not only grammatically well-formed, but also meaningful according to our linguistic conventions. Then, consider sentences with subject-predicate form, where the predicate is a 1-place predicate. Since ‘representational’ for such sentences means being fixed by a cell of the array in our model, we can infer that any such sentence is a truth bearer in a derivative sense only if it can be fixed by a cell of the diagonal array.

Based on these discussions, I can summarize the version of the deflationary conception of truth advocated in this chapter, which I call '*the functional-deflationary conception of truth*', as follows. It is a thesis that the semantic notion 'true' has no representational content, but serves as a function to govern the usage of other ordinary linguistic expressions. From this aspect, we may say that the truth predicate *T* is categorically different from other ordinary 1-place predicates. By admitting that it has no representational content, this form of deflationism goes along with the redundancy theory and minimalism. However, by emphasizing the role of *T* as a function governing the use of linguistic expressions, it is different from the other two theories. In other words, we cannot say that *T* is 'redundant' or 'completely empty'. Instead, it has its special role in our use of natural language.

One may doubt whether the functional-deflationary conception of truth is compatible with the representational nature of truth bearers.[17] There is no contradiction, however. Saying that the predicate *T* is not representational is by no means to say that truth bearers are not representational. On the contrary, the latter

[17] For example, Boghossian (1990) argues that 'any proposed requirement on candidacy for truth must be grounded in the preferred account of the nature of truth'. (p.165)

must be representational[18]. As we shall see below, separating representational sentences from non-representational ones is the basis for the truth gap approach.

5.5. The Nature of Truth Gaps

The functional-deflationary interpretation of truth can reveal the nature of truth gaps in paradoxical sentences. Why are other interpretations about truth gaps not satisfactory? For these interpretations, semantic notions such as *T* and *Het* are still construed as representational, so that they have no significant difference from other ordinary 1-place predicates. According to these views, *Het* then should occupy a row of Array 1. But if it occupies a row, then there is a box <*Het*, *Diag*(*Het*)> on that row which would cause paradox. Though truth gap theorists insist that the value of this box is 'undefined', this treatment sounds *ad hoc* and the aim is simply to avoid contradiction. What is more important, it is simply unclear why we cannot treat the 'undefinedness' as the third truth value and then make the inference from 'undefined' to 'not true'. In this case, we still have the problem called 'the superheterological paradox'.

There is no such problem for the functional-deflationary view, since according to this view there is no such box as <*Het*, *Diag*(*Het*)> to whose content we can even assign a value! *Het* is based on the function *Diag*, and thus is not

[18] Glanzberg (2003c) argues for this thesis. In his paper, he uses this thesis to argue against Horwich's minimalism. At the end of his paper, he is inclined to doubt whether there is any form of deflationism that could provide principles that explain the nature of truth bearers and demarcate their domain. The functional-deflationary conception of truth is an attempt to respond to his challenge.

representational at all. Similar explanations can be given for the truth predicate *T* and the liar sentence, though there needs to be some modification for Array 1. To deal with the liar paradox and sentence (α), Array 1 should be modified so that we include sentences on the top as well. Consequently, there will be a lot of 'n.a.' sentences generated. But since such sentences can be treated in the same way as suggested in the first section of this chapter, let us put them away and consider an ideal array where each cell is bivalent. Then, the following sentence will appear somewhere at the top of this array:

(α) Sentence (α) is not true.

If we treat the truth predicate *T* as representational, i.e. as a row of the array, then its negation ~*T* should show up as a row of that array too, since ~*T* is also a 1-place predicate[19]. Then, on the row ~*T*, there must be such a box on that row: <~*T*, α>, which is sentence (α) itself. If, following other truth gap theorists' suggestions, we put 'undefined' into the cell <~*T*, α>, then it is unclear why we cannot infer from 'undefined' to 'not true', and we will encounter the revenge of the liar. Thus, this treatment cannot succeed. To solve the problem, we should treat the predicate *T* as a function, so that neither *T* nor ~*T* can show up as a row

[19] There may be some controversy about this claim, if we consider the distinction between predicate negation and sentence negation. Predicate negation involves a negative element in the predicate (as in the predicate 'does not have 5 words in English'), while sentence negation is expressed in sentences beginning with 'It is not the case that ….' For singular sentences such as (α), a difference between these two kinds of negations is that, predicate negation is the *assertion* that the thing referred by the subject term does not have a certain property, while sentence negation is the *denial* that something is the case. Though it is possible to interpret (α) in both ways, it seems more intuitive to treat it as an assertion about itself, i.e. the assertion that it is not true. (The same kind of treatment can also be found in Barwise and Etchemendy 1987: 16-18.) Consequently, I shall treat '~*T*' also as a 1-place predicate.

on the array. Thus, there is no such box <~*T*, α> existing on the array. Therefore, the fact that the liar sentence cannot receive a truth value is not due to some artificial stipulation, but is due to the functional-deflationary status of the truth predicate. We may call truth value gaps caused by such a reason 'systematic gaps'.

This contrasts with various 'artificial gaps'. For example, in his book *Understanding Truth*, Soames (1999) advocates a theory which aims to explain the nature of truth gaps in Kripke's construction. I introduced Soames' 'Gappy Predicates' theory in Chapter 4. Here let us recap the core of his theory. Soames argues that the truth gaps caused by semantic paradoxes are just like the gaps caused by the predicate 'smidget' in the following example.

Suppose there are two groups of people. Group A consists of adults who are abnormally short (around four feet tall), and Group B consists of adults whose height is at the low end of the normal range (around five feet tall). Moreover, each member of group B is perceptibly taller than any member of group A. Then one can introduce a new word 'smidget' into our language by the following conventions:

> i. Every member of group A is (now) a smidget. Further, for any adult whatsoever (and time t), if the height of that adult (at t) is less than or equal to the (present) height of at least one member of group A, then that adult is a smidget (at t).
> ii. Every member of group B is not (now) a smidget. Further, for any adult whatsoever (and time t), if the height of that adult (at t) is greater than or

> equal to the (present) height of at least one member of group B, then that adult is not a smidget (at t).
> iii. Nonadults (and nonhumans) are not smidgets. (Soames 1999: 164)

The definition above gives a sufficient condition for something to be a smidget and a sufficient condition for something not to be a smidget. But it does not give a sufficient *and necessary* condition for a thing to be called ‘smidget’. If there is an adult, say Mr. Smallman, whose height is precisely halfway between that of the tallest person in group A and that of the shortest person in group B, then it remains undefined whether this adult is a ‘smidget’ or not. In short, there is a gap between the height of the tallest person in Group A and that of the shortest person in Group B. Since the convention governing the usage of ‘smidget’ (i.e. the definition given above) remains silent about heights in this range , then the truth value of the following sentence is undefined, despite the fact that it does express a proposition:

(7) Mr. Smallman is a smidget.

Soames argues that the gap caused by the liar sentence is just like that of (7). In other words, according to Soames’ theory, the liar sentence (α) also expresses a proposition (i.e. it is a truth bearer), and this proposition cannot receive a truth value in just the same way as the proposition expressed by (7) cannot receive a truth value.

However, this analogy cannot hold. As I have argued above, the truth predicate *T*, as a function on the array, is categorically different from other ordinary 1-place predicates. Although there is also something special about the predicate ‘smidget’, i.e. it is partially defined, it still cannot enjoy the same status as the predicate *T*. If we are asked to provide an array to show the difference between these two predicates, it is clear that the predicate ‘smidget (*S*)’ should occupy a row of it. In other words, this predicate is representational:

Array 2: Smidget *vs*. True

	P_1	P_2	P_3	...	P_i	…	…	$\text{n}_{(s)}$	…
P_1	1	0	1	…					
P_2	1	0	0	…					
P_3	0	1	0	…					
…	…	…	…	…					
P_i					0				
…						…			
S								U	
...									…

Since ‘Mr. Smallman’ is a name for an individual, the array should be expanded again to include such names. As usual, let us consider an idealization of this array which excludes all the ‘n.a.’ sentences. Let us abbreviate ‘Mr. Smallman’ as ‘$\text{n}_{(s)}$’. The box <*S*, $\text{n}_{(s)}$> does not belong to the value of the diagonal because the name ‘$\text{n}_{(s)}$’ does not stand for the predicate ‘smidget’. In the situation where Mr. Smallman’s height is within the intermediate range, the truth value for the box <*S*, $\text{n}_{(s)}$> is undefined. Thus, we put ‘U’ in this cell to indicate that its value is undefined. But before the discussion about the difference between ‘Smidget’ and ‘True’, we should notice that this ‘undefined’ in the cell <*S*, $n_{(s)}$> is also different from the ‘n.a.’ sentences which I have discussed at the beginning of this chapter. For sentences such as ‘“Is a prime number” is sleepy’, we put ‘n.a.’ in the corresponding cell to indicate that this is not a well-formed sentence according to our implicit understanding of English. Normally, the predicate ‘is sleepy’ is not applicable to a predicate such as ‘is a prime number’. However, there is no such restriction for ‘Smidget’ (*S*). Normally, this predicate *S* is applied to a person, and ‘$n_{(s)}$’ is a name for a person. The truth value for Sentence (7) is not undefined because of general linguistic rules in English, but because of some more specific, artificial rules governing the use of this predicate (i.e. the rules introduced by Soames). Admittedly, the boundary between general linguistic conventions and specific rules is vague and flexible. If Soames’ rule were accepted by the majority of English speakers and became an implicit understanding of this word, then it would become a general linguistic convention. After all, we introduce new words to English every day and some of them become popular and then become part of ordinary English. Thus, the boundary between ‘U’ and ‘n.a.’ is flexible. Nonetheless, the term ‘smidget’ cannot enjoy the same status as the truth predicate *T*.

Unlike the liar sentence (α), although the value of the cell <*S*, $n_{(s)}$> is undefined for some artificial reason, Sentence (7) still occupies a cell. This shows the fundamental difference between ‘smidget’ and ‘true’: one occurs as one row of the array, while the other does not. Consequently, though sentence (7) may not have a truth value in some situation, it still expresses a proposition and is indirectly a truth bearer. Consequently, it is *possible* for this sentence to receive a truth value, while it is impossible for the liar to receive a truth value, since it cannot be a truth bearer at all. We may say that the gap in the 'smidget' case is due to an artificial reason, since it is totally artificial that the predicate 'smidget' is introduced in such a way that sentences like (7) do not receive a truth value in some circumstances. If, however, we introduce the predicate 'smidget' into English without allowing gaps, then there is no 'U' occurring on the row labelled by '*S*' at all. Also, if Mr. Smallman grows a little bit higher so that he is perceptibly taller than the shortest person in Group B, then we shall put a '0' into the cell <*S*, $n_{(s)}$>. These facts indicate that the undefinedness for the cell <*S*, $n_{(s)}$> is totally artificial and contingent. It is simply an 'artificial gap' which can be avoided if we *make* a different set of rules for the word 'smidget'. However, the truth predicate *T* is not introduced into our natural language by some artificial rules, nor could the liar sentence (α) receive a truth value if another situation were to obtain. This is why I distinguish contingent liars from essential liars in Chapter 3. For contingent liars, the sentence could express a proposition if the subject term does not refer to the sentence itself. In that case, the sentence ceases to be a liar sentence as well. However, for essential liars such as (α), no matter what kind of situation obtains, it does not express a proposition. Since the 'smidget' sentence could receive a truth value in a different situation, it is inadequate to make the analogy between the artificial gap for ‘smidget’ and the systematic gap for ‘true’. To mix up the two cases is to mix up a function on the diagonal array with the predicates governed by such a function. Problems like the revenge of the liar are caused by such confusion.

5.6. The Revenge of the Liar

As many critics have pointed out, various forms of truth gap theory are plagued with the problem called ‘the revenge of the liar’. Kripke has provided a precise mathematical definition for Truth in a formal language. His theory suffers from the problem of the revenge of the liar because it lacks an appropriate philosophical justification for such a definition. There is no such problem if we adopt the functional-deflationary interpretation of truth. Recall Array 1. The revenge of the liar could only occur when we treat *T* and thus ~*T* as one row of the array, i.e. as representational. In that case there would be a box on the row ~*T*, i.e. <~*T*, α>, for which any truth value assignment would lead to contradiction. But, as argued above, ~*T* cannot occur as a row of Array 1 at all. Thus, there is no such box that could cause the problem. In other words, if we understand the predicate ‘true’ appropriately, then we even shall not raise the question ‘what is the truth value of the liar sentence?’ Consider the liar sentence again:

(α) Sentence (α) is not true.

To receive a truth value, the content of (α) should in the first place be a truth bearer. According to the definition of a truth bearer, it must be representational.

But when we check what has been expressed by (α), we can immediately see the tension: its subject term refers to a sentence (i.e. itself) the content of which cannot be representational, since it is essentially based on the non-representational predicate *T* and cannot be rewritten in a way that does not involve the truth predicate. The self-referential feature simply leads us back to the notion *T* in an infinite loop.

Thus, (α) cannot be representational. We are inclined to think that it is because there is some confusion involved: the confusion from the linguistic analogy that, since *T* is a monadic predicate, it must also occupy a row like other monadic predicates in this language. But, as analyzed above, this is a misconception of the truth predicate, which mixes up the functional role of *T* with the representational role of other ordinary predicates.

One common objection to truth gap theory is that, from the liar's being 'undefined', we can infer that it is not true, as argued by Burge:

> Claiming that in the problem sentence the truth predicate is undefined or its application indeterminate does not help matters. For one may still reason that, if a sentence's predication is undefined or indeterminate, then the sentence is not true. This reasoning may or may not involve a broadening of the domain of discourse or a sharpening of the extension of 'true'. But it is informally quite intuitive. (Burge 1979: 175)

Such reasoning makes some sense before the nature of truth gaps has been clarified. But if we start to understand the functional role of *T*, then it is clear that this reasoning is misleading. The word 'undefined' does not mean that there is a 'third truth value' for sentences like the liar, but only suggests that talking about the truth value of such sentences is totally inappropriate. The liar sentence cannot receive a truth value, not because of an artificial stipulation, but because the systematic feature of our language makes the truth value assignment *impossible*.

Another question is about the claims in the present chapter. How do we view these statements that have been made about the truth predicate as well as other semantic predicates like *Het* and *Aut* in this chapter? Don't they belong to some meta-language, which is in a level above the language **L**? Questions like these are still due to a misunderstanding of semantic notions in natural language. The functional-deflationary interpretation of truth in this chapter is not the construction of some theory which gives a 'new' definition or convention for truth. Rather, it is simply the explication of what has been implicitly understood by every speaker when they say in our natural language that something is true. People do not have to learn this interpretation before they can use the word 'true' in their everyday conversation. But, if one is puzzled by semantic paradoxes such as the liar or the heterological paradox, then the interpretation in this chapter helps to clarify what has gone wrong in the reasoning. In a word, what has been said in this chapter is simply to clarify some misunderstandings associated with the word 'true' found in our natural language. These explanations, of course, are still part of natural language.

5.7. Truth Gaps and Non-classical Logic

According to Tarski, there are two assumptions that are essential for the liar paradox (Tarski 1944: 348):

> (I) We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term "*true*" referring to sentences of this language; we have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language. A language with these properties will be called "*semantically closed*".
>
> (II) We have assumed that in this language the ordinary laws of logic hold.

Tarski himself finds a way out of the paradox by rejecting the first assumption, while at the same time he deems rejecting the second assumption as the most unfavoured option:

> It would be superfluous to stress here the consequences of rejecting the assumption (II), that is, of changing our logic (supposing this were possible) even in its more elementary and fundamental parts. We thus consider only the possibility of rejecting the assumption (I). Accordingly, we decide *not to use any language which is semantically closed* in the sense given. (ibid. p.349)

It is commonly believed that the truth gap approach, by admitting gaps between 'true' and 'false', changes our logic, since it seems to violate the law of excluded middle. Kripke himself uses Kleene's 'three-valued logic' to handle truth values for a language containing truth gaps. Thus, his theory is also sometimes called a 'non-classical logic' approach by some critics. However, this label is inappropriate. In his paper, Kripke says:

> "Undefined" is not an extra truth value… Nor should it be said that "classical logic" does not generally hold… *If* certain sentences express propositions, any tautological truth function of them expresses a true

> proposition. Of course formulas, even with the forms of tautologies, which have components that do not express propositions may have truth functions that do not express propositions either. … Mere conventions for handling terms that do not designate numbers should not be called changes in arithmetic; conventions for handling sentences that do not express propositions are not in any philosophically significant sense "changes in logic." The term 'three-valued logic', occasionally used here, should not mislead. All our considerations can be formalized in a classical metalanguage. (Kripke 1975: footnote 18)

In this passage, Kripke distinguishes two issues: whether each proposition is either true or false, and whether each sentence expresses a proposition. In his view, logic applies to propositions, but it is not the case that every declarative sentence expresses a proposition. What he has done in his paper is to give some guide for handling sentences that do not express propositions. Thus, the title 'non-classic logic', strictly speaking, does not appropriately describe his truth gap approach.

In my interpretation, a proposition is the content expressed by a sentence, i.e. the truth bearer. There is no analogues problem that classic logic has been altered, since only when there are truth bearers can we consider their truth values. In other words, logic applies to cells on the array, and for each cell on the diagonal array, it is bivalent[20]. All we have done is simply to clarify why paradoxical sentences such like the liar sentence do not occupy such a cell. Therefore, as Kripke has correctly pointed out, there is no change of logic, but

[20] For this claim, one may raise the question about vague terms and artificially defined terms like Soames' 'smidget'. However, as argued above, since the gaps associated with these terms are due to linguistic conventions (either explicitly or implicitly defined), it is possible for us to make the corresponding sentence ungappy, by changing the relevant linguistic rules. So, in this sense, we can say that the cells on the array are bivalent.

only an explanation for why some sentences do not express propositions and are thus not truth bearers.

5.8. Intuitions about Truth

In Chapter 3, I discussed the criteria for a good solution to the liar paradox, and one of them is about intuition. It is a consensus that a philosophically satisfying solution must accommodate the various intuitions associated with the natural notion of truth. In previous sections, I have argued that the functional-deflationary conception of truth confirms our intuitions: that we can express the truth predicate in natural language, that there is no metalanguage for natural language, and that natural language is semantically universal.

For the ‘intuition’ criterion, one may argue that not every so-called ‘intuition’ is adequate and thus can serve as the justification of a theory, as Barwise and Etchemendy have recognized:

> The obvious lesson taught by the liar is that our semantic intuitions, though doubtless generally sound, need refinement. But the process of refining our intuitions requires a better understanding of the linguistic mechanisms themselves, and of how they interact, not just an assessment of the faulty principles that describe our untutored intuitions. (Barwise and Etchemendy 1989: 8)

If one regards some intuition about natural language as not adequate, then one should explain why. For example, I argued that the intuition behind the reasoning from the ‘undefinedness’ of paradoxical sentences to the conclusion that they are ‘not true’ is misleading, because it makes a false analogy between *T* and other ordinary 1-place predicates. Thus, the functional-deflationary interpretation can meet the requirement of the intuition criterion. In Section 2, I argued that some alternative approaches do not meet this requirement. In what follows, I argue that another major competing approach, contextualism, has this problem as well.

One fundamental intuition that contextualists rely on is 'falsity by default'. The basic idea of contextualism is that the truth value of the liar sentence (α) is unstable; nonetheless, it still receives a truth value in each context. Now, how do they assign a truth value to the liar in the initial context? They say that it is not true. Since they do not admit truth gaps, 'not true' is equivalent to 'false'. However, the falsity of the liar is not because its truth conditions are not fulfilled, but because it has no truth conditions in the initial context. For example[21]:

> The sentence is *not true$_i$*-not because its truth$_i$ conditions are not fulfilled, but because it has no truth$_i$ conditions. But it does have truth$_k$ conditions and indeed is true$_k$. (Burge 1979: 180)

One may think that this idea is 'quite intuitive'. However, we may ask this question: what are the truth conditions for the autological sentence (4), as well as for sentence (β), which is called 'the truth teller'?

(4) 'Is autological' is autological.
(β) Sentence (β) is true.

Can the truth teller sentence have truth conditions in the initial context? If it can, then what are the conditions? If it cannot, then shouldn't we also assign 'falsity'

[21] The same treatment can also be found in Barwise and Etchemendy 1989: 135.

to the truth teller? Burge thinks that we should treat the truth teller in the same way as the liar:

> Self-referentially intended strings like 'This sentence is true$_i$,' are not true$_i$-not because their truth, conditions are not fulfilled (they have no truth$_i$ conditions), but because they are pathological$_i$ in not applying 'true$_i$,' ('true' at the appropriate occurrence) derivatively.
> Construction 1 rules pathological the sentences that are intuitively empty or lead to paradox. But to some (including myself) it seems too stringent. (Burge 1979: 186-7)

According to the contextual analysis of the liar, the truth teller should also be analyzed as ‘not true’ in the initial context, but this contradicts our intuitive idea. Intuitively, we may think that the truth teller sentence is true. It seems that this sentence is not only true, but also necessarily true. Contextualists may argue that because they are ‘intuitively empty’, pathological sentences such as (3), (4), (α), (β) are all false. But this account can hardly be satisfactory. If we follow the terminology used in previous sections, ‘intuitively empty’ amounts to saying that the sentence is non-representational. In other words, the truth teller cannot occupy a cell either. This intuition should be respected because it reveals the fundamental distinction between the predicate *T* and other ordinary predicates. However, contextualists want to assign ‘not true’ to such sentences, so that the specialty of this semantic notion is thus totally ignored. For pathological sentences such as (3), (4), (α), and (β), contextualists do not conduct an inquiry into the issue why they are pathological. Rather, they simply efface this peculiar feature of truth and evaluate pathological sentences in just the same way as ordinary sentences. This treatment thus cannot improve our understanding of the mechanism associated with truth. The functional-deflationary interpretation advocated in this chapter, on the other hand, does supply an explanation for why pathological sentences are ‘intuitively empty’. By exploring the ‘undefinedness’ of the liar, it has been shown in previous sections that the emptiness of the liar (as well as other pathological sentences) is due to the functional role of semantic notions in our usage of language. In this sense, compared with contextualism, the functional-deflationary interpretation has not only accommodated our intuitions in a better way, but also has properly explained the reason behind these intuitive ideas about truth.

There is also another important problem for contextualism. It is not clear why the liar sentence, which has no truth conditions in the initial context, can have truth conditions in subsequent contexts. It is a common feature of the contextual approach that the context for the evaluation of the liar sentence always changes. Thus, after the first evaluation of the liar, there is a new context generated. And the result of the first evaluation, i.e. ‘the liar is not true$_i$’ immediately creates a new ‘semantic fact’ which is ‘imported’ to the second context. In this way, the liar depicts something in this new context and thus becomes true$_{i+1}$. Consequently, contextualists argue that the liar is not empty in the second level. The idea that the empty liar sentence suddenly becomes non-empty in the second level is quite suspicious. But there is something which is more suspicious when we consider this question: what has caused the shift of contexts? Contextualists may say that it is the initial evaluation of the liar and the ‘semantic fact’ generated by such evaluation. However, if that is the case, then can we generalize it and say that, whenever someone has claimed that something is (or is not) the case, it then generates a semantic fact and causes a shift of contexts? For example, let us consider an ordinary interlocution which involves two interlocutors. Each of them states one sentence in turn. According to contextualism, each statement creates a new ‘semantic fact’; accordingly a new context is generated. In other words, the context of this conversation is always changing, and the speaker may even not remember to which context her claims should belong. Does this idea reflect our ordinary understanding of natural language? If the contextualists still insist that it is intuitive, then we may wonder how, in an ever-changing context, ordinary communication could become possible. If contextualists argue that this treatment is only for the pathological sentences involved in semantic paradoxes, so that ordinary conversations are not affected, then we may suspect the significance of this treatment. After all, it is a consensus among theorists that an adequate treatment should be able to improve our understanding of semantic mechanisms. An *ad hoc* solution which can only avoid contradiction could never be a good solution.

5.9. Conclusion

In this chapter, I provided a philosophical justification for truth gaps associated with semantic paradoxes. Through a simple model for a language **L**, I argue that the heterological predicate *Het* is a dynamic notion and thus cannot be fixed by a row of cells in this model. By recognizing the functional role of *Het*, the heterological paradox is solved without resorting to a hierarchy of heterological predicates, nor need we abandon the intuitive idea that natural language is semantically universal. For the liar paradox, I advocate a functional-deflationary conception of truth, so that the truth predicate *T* should not be treated as a fixed set of cells in the model either. Their functional role shows that semantic notions such as *Het* and *T* are not representational, and this explains the nature of truth gaps. Also, there is no problem like the revenge of the liar in this interpretation, because it is impossible to apply the truth predicate to the liar sentence. In the end, I compared this interpretation with another approach to the liar, contextualism, and showed that the latter violates some important intuitions associated with natural language. Therefore, I conclude that the functional-deflationary conception of truth can deal with our semantic intuitions in a better way and thus could be an adequate solution to the liar paradox.

Chapter 6: Paradoxes of Definability

The topic of this chapter is another kind of semantic paradox: paradoxes of definability (also called 'paradoxes of denotation'[1]), which include Berry's paradox, König's paradox and Richard's paradox. The chapter begins with a solution for this kind of paradox, which is an inference from the functional-deflationary interpretation of semantic notions developed in Chapter 5. In defending this view, I also argue against a form of physicalism (held by Field 1972), and emphasize the distinction between representational expressions and non-representational ones. In Section 6.3, I investigate another leading approach to semantic paradoxes: Priest's dialetheism. This investigation is divided into two steps: first, I argue against Priest's thesis that a common structure he has identified, i.e. the 'Inclosure Schema', guarantees a uniform solution for all logical paradoxes. Second, I respond to Priest's criticism of my argument compare his dialetheism with my functional-deflationary solution, and argue that my explanation is preferable. This chapter ends with a concluding remark about semantic paradoxes.

6.1. The Functional-deflationary Solution to Paradoxes of Definability

As an example for this group of paradoxes, let us use Berry's paradox, which concerns the least natural number not definable in English in fewer than

[1] Though generally speaking, the word 'define' has a wider application compared with the word 'denote', in the context of these paradoxes, however, they can be treated as meaning the same.

nineteen words. What kind of solution should be provided for it? According to the functional-deflationary view developed in the previous chapter, the answer is very simple: this phrase simply does not denote any natural number. However, it sounds intuitive that it denotes. The task of this chapter is to explain why we have the inclination to think that it does.

6.1.1 Decimal Numerals and Abbreviations

In Berry's paradox, the supposed 'contradiction' is that the natural number (if any) both is and is not definable in English by a phrase with fewer than 19 words. However, to get a rigorous contradiction, there are some assumptions that must be fulfilled beforehand.

First, it assumes that there is a set whose members are all natural numbers definable in English in fewer than 19 words. Let us call this 'set' $\boldsymbol{DN_{19}}$. Only if such a set exists can there be a least natural number which is outside it. Second, it assumes that the phrase in Berry's paradox, i.e. 'the least natural number not definable in English in less than nineteen words', can indeed denote a natural number. There is a contradiction only when this phrase indeed denotes a natural number. However, one may ask, why should we take these two assumptions for granted?

I shall discuss the first assumption in Section 3. Here let us first consider the problem with the second assumption. To begin with, we have to ask what kind of expressions should be included in English. Should we treat decimal numerals as part of ordinary English? If so, how many English words do they consist in? Should we regard the decimal numeral '**1234567890**' as one word in English or 10 words in English, or not a word in English at all? It seems that normally we treat it as one word in English, which denotes the number *1234567890*. But if this is the case, then for any natural number, no matter how large it is, there is a phrase in English with only one word that denotes that number. Accordingly, there is no natural number that is not denotable by a phrase in English with fewer than 19 words. Therefore, the phrase in Berry's paradox does not denote any natural number.[2]

One may want to exclude decimal numerals from English. But this restriction cannot help to produce a genuine contradiction either. Let us restrict 'words' to ordinary English words which denote a number, such as 'the number of planets in the solar system'. Suppose that one (or a super machine) has just examined all such phrases in English, and found out the least natural number that is not denoted by a phrase included in the list. Accordingly, the phrase 'the natural number which I just found' should denote that number with fewer than 19 words. We can even abbreviate that phrase in only one word. We also can give it whatever name we like. Thus that number is no longer a proper candidate for contradiction. Moreover, for any such candidate, we can denote it in the same way. Therefore, just as in the case of decimal numerals, there is no natural number

[2] This argument is adapted from Shapiro and Wright 2006, page 260-2.

which cannot be denoted in English with fewer than 19 words. Again, the conclusion is that there is no contradiction involved: the phrase in Berry's paradox simply does not denote any natural number.

6.1.2. Defining 'Denotation' Based on *T*

The argument above shows the flexibility of natural languages. There is nothing that cannot be denoted in English with fewer than 19 words. If it has not been denoted in that way before, then we can simply give it a name and establish the denotation relation. Why do we have the inclination to think that the phrase in Berry's paradox should denote some natural number? This is similar to the question about truth: why we are inclined to think that the Liar sentence should have a truth value? In Chapter 5, I have distinguished two kinds of linguistic expressions: representational and non-representational. It is because of the non-representational feature of the truth predicate that the Liar cannot receive a truth value. There I argue for the following thesis:

> **Thesis 5.2** If *x* is defined as a function, or defined based on a function, on the Array, then *x* is not representational.

Since 'denote' is also a semantic notion, we may expect that it is also non-representational. To show that it indeed is, we can define this notion by the functional-deflationary *T* (the explanation of which is provided in Chapter 5):

> For any linguistic expression ‘*x*’, and for any *y*, ‘*x*’ denotes *y* iff ‘*y* = *x*’ is true. [3]
>
> iff $T<x, y>=1$[4]

Since the semantic notion ‘denote’ is defined on the basis of the semantic notion *T*, which should be understood as a function on the array, then the notion ‘denote’ is not representational either. When we say “an expression ‘*x*’ denotes *y*”, it does not describe the world as being some way, in contrast to ‘*x* beats *y*’. The semantic notion involved in the former simply tells us the role of a linguistic expression in our use of language. Thus, there is a person denoted by the phrase ‘the shortest man who has been beaten by more than 19 people between t_1 and t_2 in place *w*’, but there is no natural number denoted by ‘the least natural number not definable in English in fewer than 19 words’. We have the inclination to think that it is representational and thus that the phrase in Berry’s paradox denotes a natural number because there is some confusion involved: the confusion from the linguistic analogy that, since ‘denote’ is a 2-place relation, it must also be representational in the same way as ordinary relations, such as ‘beat’. But, as analyzed above, this is a misconception of the semantic notion ‘denote’, which mixes up its functional role with the representational role of ordinary relations.

[3] The *x* surrounded by quotes means that this expression is mentioned, while the expression without quotes means that it is then used. For example, ‘Plato’ denotes the teacher of Aristotle iff ‘the teacher of Aristotle = Plato’ is true.

[4] Here it is an extension of the use of ‘*T*’ discussed in Chapter 5, since the variables range over names rather than over predicates. In Chapter 5, if $T<P_i, P_j>=1$, it means that P_i applies to P_j. Also, if $T<P_i, n_{(a)}>=1$, it means that P_i applies to the object denoted by $n_{(a)}$. Here, if $T<x, y>=1$, it means that ‘*x*’ denotes *y*.

6.2. Argument against Tarski-Field's Physicalism

However, there is an argument supported by some authors, which amounts to saying that the semantic notion of denotation is representational. In his interpretation of Tarski's work on Truth, Field (1972) argues that Tarski's aim is to define the semantic conception of truth by non-semantic terms. According to Field's report, Tarski's philosophical stance could be characterized as a kind of 'physicalism': 'the doctrine that chemical facts, biological facts, psychological facts and semantical facts are all explicable (in principle) in terms of physical facts.' (Field 1972: 357) If a definition of semantic notions such as truth could not be given in terms of physical facts, Tarski writes, 'it would then be difficult to bring [semantics] into harmony with the postulates of the unity of science and of physicalism (since the concepts of semantics would be neither logical nor physical concepts).' (Tarski 1936: 406)

Tarski's truth definition does not achieve the goal of defining truth without employing any undefined semantic terms. Instead, argues Field, Tarski manages to show how truth can be characterized in terms of some other semantic notions, such as 'denotation' and 'satisfaction', which Field calls 'primitive' semantic notions. It is an obstacle to Tarski's physicalism if such 'primitive' semantic notions as 'primitive denotation' cannot be explicated further in physical terms. As a physicalist (at least in that paper), Field argues that they can. A promising candidate theory in his mind to achieve this goal is Kripke's causal theory of reference (developed in Kripke 1980):

> I don't think that Kripke or anyone else thinks that purely causal theories of primitive denotation can be developed (even for proper names of past physical objects and for natural-kind predicates); this however should not blind us to the fact that he has suggested a kind of factor involved in denotation that gives new hope to the idea of explaining the connection between language and the things it is about. It seems to me that the possibility of some such theory of denotation (to be deliberately very vague) is essential to the joint acceptability of physicalism and the semantic term 'denotes', and that denotation definitions like DE and DG merely obscure the need for this. (Field 1972: 367)

Let us call this view 'Tarski-Field's Physicalism'. If Field is right, i.e. if we can reduce semantic concepts to physical terms, while at the same time the definition is in accordance with our intuitive understanding about semantics, then it goes directly against my thesis about non-representational expressions developed in Chapter 5. For expressions which describe physical facts are representational, because they describe the world as being some way. If semantic notions can be reduced to such terms, then it means that they are also representational. Thus, the distinction between representational and non-representational terms vanishes, and it is fatal to my argument for the functional-deflationary truth and the inferences from it. Thus, this challenge must be treated properly before we can proceed.

According to the causal theory of reference, the fact that 'Aristotle' denotes Aristotle is to be explained in terms of certain kinds of causal networks between Aristotle and our uses of 'Aristotle'. In particular, there was an initial 'baptism' where the new-born baby and his name were introduced to other people.

In this way, the word ‘Aristotle’ was passed down to us today from the original users of the name. That is why people today can still use this name and refer to the same person as the original users of the name in Ancient Greek. However, though this is called ‘the causal theory of reference’, Kripke himself does not intend to produce a substantive theory of denotation for names in his *Naming and Necessity*. Instead, he only intends to criticize the mistakes in an alternative theory (which is usually called ‘descriptivism of names’), and ‘present just a better picture than the picture presented by the received views’ (Kripke 1980: 93). Field insists that this ‘alternative picture’ could be developed into a substantive theory, but how?

Let us try a tentative development. Since ‘denote’ is explained by the causal link between the users of the name and the object that is named, we may define this notions as follows:

> For any linguistic expression ‘*x*’, and for any *y*, ‘*x*’ denotes *y* (in a given context) iff there is a causal link *C* of a proper kind, such that *C* (the user of *x* in a given context, *y*).

The relation *C* is a 2-place relation between users of a linguistic expression and an object. To facilitate our discussion, let us narrow ‘objects’ to ‘persons’. Since the person denoted by a name is definite, the relation *C* can be characterized by a set of ordered pairs[5]:

[5] As the physicalists want the relation *C* to be a physically explainable term, it should be able to be represented extensionally.

> *C*= {<*x*, *y*>|there is a physical relation of a certain kind between the users of *x* and *y*}

It is quite doubtful whether the physicalists can specify the details of the ‘kind of physical relation’, since the physical relation between a user of a name and the name’s denotation varies from one user to another. However, for our argument here, let us simply assume that they have worked it out, so that the set *C* is used to define the semantic notion of denotation. No doubt both ‘the kind of physical relation’ and ‘the user of’ can be fully described in terms of physical facts, at least in principle. This thus seems to confirm Tarski-Field’s Physicalism. But there is a problem for this definition.

This problem is not novel. It is the plague for all semantic paradoxes: if we treat the relative semantic notion as being defined by a *set*, then it is fixed. As a fixed set, it can be diagonalized out, so that we have a new element which should belong to but does not belong to the set. For our present definition, the diagonalization can be done in the following way. Suppose we can have such a set *C*. There is no doubt that there are some human beings in the world that have not been named by any person yet (e.g. new-born babies who have not been named). Those people can be listed in an ordered way (e.g. according to their birth time). Let *a* be the first person on the list. Accordingly, this description ‘the first one on the list of those human beings who have not been named by any person’ should pick out a unique person. Let us call this person ‘Adam’. Thus, my usage of

‘Adam’ should denote the person *a*. However, the pair <Adam, *a*> is not an element in *C*, since *a* is not the second item of any ordered pair in *C*. Since *C* is supposed to define ‘denotation’, it means that there is no denotation relation between ‘Adam’ and *a*. But it is obvious that there is. Thus, there is a contradiction for this definition.

One may argue that there is no ‘such kind of physical relation’ between my usage of ‘Adam’ and the person thus named. But then it is the arguer’s burden of proof to characterize the details of such kind of physical relation and to show that my usage does not qualify. Note that Kripke himself admits that names introduced in this way are genuine names, and can genuinely denote an object. In his discussion about the name ‘Neptune’, Kripke admits that this name denoted the planet which caused such and such discrepancies in the orbits of certain other planets, even before this planet was discovered. ‘If Leverrier indeed gave the name ‘Neptune’ to the planet before it was ever seen, then he fixed the reference of ‘Neptune’ by means of the description just mentioned.’ (Kripke 1980: n. 33, p. 79) Since Field treats Kripke’s theory as a promising candidate for a physicalist’s definition of denotation, then he should admit that the name ‘Adam’ in my usage indeed denotes the person. But if he admits this, then he is forced to admit that the physicalists’ definition in terms of the set *C* and ‘a physical relation of a certain kind’ cannot define the semantic notion of denotation.

There may be other way to develop a notion of denotation in physical terms, but Field did not specify one in his paper. However, even if there is some way other than ‘proper causal link’ to cash out his version of physicalism, it should be subject to the same criticism. That is because any such characterization, if it really employs only terms allowed by physicalists, must be representational. Since it is representational, it can be characterized in terms of a *set* (of ordered pairs). Accordingly, we can apply the same method to diagonalize out such a set and generate a contradiction for such a definition. It is because of this limitation that Tarski-Field’s Physicalism is flawed.

The facts described by representational terms can be reduced to physical facts. Semantic terms do not describe that kind of fact, and therefore cannot be reduced in that way. However, this does not suggest that semantic terms describe some ‘Cartesian’ facts which are located beyond the physical world. The functional-deflationary view has no such implication. It is called ‘deflationary’, indicating that there are no ‘objects’ corresponding to semantic terms, whether ordinary or special. Instead, semantic notions express the function which governs our use of language. All that can be described in representational terms can be explicated in physical terms – if this is what is meant by ‘physicalism’, then the functional-deflationary view about semantics is totally compatible with such physicalism. There is nothing mysterious (i.e. nothing of the sort suggested by the early Wittgenstein’s *Tractatus*) about semantic notions. We use them and talk about them every day. The only thing that makes them special is that they are not representational.

6.3. What a 'Common Structure' Can Imply

In Chapter 3, I briefly introduced approaches to semantic paradoxes in contemporary literature: Tarskian Hierarchy, Contextualism, Truth gap approach, and Dialetheism. My own explanation follows the truth gap approach. I have discussed the deficiency in the Tarskian Hierarchy in Chapter 3, and difficulties for Contextualism in Chapter 5. Accordingly, there also should be some discussion about Dialetheism. I shall use Priest's dialetheism (developed in his (2002) *Beyond the Limits of Thought*) as an example of this approach. Priest's argument for a dialetheic solution to both semantic paradoxes and set-theoretic ones consists in three steps: (i) that there is a common structure for both set-theoretic and semantic paradoxes; (ii) that such a common structure implies a uniform solution to these paradoxes; and (iii) that such a uniform solution must be dialetheism.

6.3.1. Review of the Discussion on a 'Common Structure'

As introduced in Chapter 1, logical paradoxes are traditionally divided into two distinct families: set-theoretic paradoxes (Russell's paradox, the paradox of ordinal numbers, the paradox of cardinal numbers, etc.) and semantic paradoxes (the Liar paradox, Berry's paradox, König's paradox, Richard's paradox, the heterological paradox, etc.). This classical division was made by Ramsey [1926], based on what terms are used to express each paradox (in what follows, I call this 'Ramsey's content criterion'). As mentioned in Chapter 2, it has also been argued by many authors (e.g. Russell, Thomson, Simmons and Priest) that these paradoxes have a common structure, i.e. they all can be analyzed as a diagonal argument. For example, Thomson has summarized this common structure as the following theorem:

> Theorem (1): Let *S* be any set and *R* any relation defined at least on *S*. Then no element of *S* has *R* to all and only those *S*-elements which do not have *R* to themselves. (Thomson 1962: 104)

This theorem is refined by Simmons (1993) in the following symbolic form (Simmons 1993: 25):

(Ru) $\neg\exists x\forall y(\mathrm{J}(x, y) \leftrightarrow \neg\mathrm{J}(y, y))$.

Although these authors all agree that these paradoxes can be analyzed as a diagonal argument, they have different attitudes toward the implication of such 'common structure'. For example, though Russell is the first one who published this idea, he didn't at the same time say that a common structure means that they must be subject to a common resolution.[6] In fact, he proposes the type theory as the solution for his famous paradox about the set of all sets which are not members of themselves, but he does not claim that this is also a solution for the

[6] It should be noted that later on, in *Principia Mathematica*, Russell did think that there is a common resolution to these paradoxes, i.e. the theory of types. However, that is not because they have a common structure.

semantic paradoxes. Actually, he seemed also to be puzzled by the semantic paradoxes and was not willing to provide a hasty solution at that moment. For example, immediately after the analysis of the 'structure' of Cantor's paradox, the paradox he has discovered, and the paradox about propositions, Russell says:

> The only solution I can suggest is, to accept the conclusion that there is no greatest number and the doctrine of types, and to deny that there are any true propositions concerning all objects or all propositions. Yet the latter, at least seems plainly false, since all propositions are at any rate true or false, even if they had no other common properties. In this unsatisfactory state, I reluctantly leave the problem to the ingenuity of the reader. (Russell 1903: p. 368, §349)

That is to say, though Russell has identified some similarity between these various paradoxes, he has never suggested that such a 'common structure' can guarantee a unique solution. On the contrary, he recommended the theory of types as the solution for those in naïve set theory, while judging that 'the case of propositions is more difficult' (ibid. p. 367, §349)

Later, Thomson (1962) argues that the Barber 'paradox', the heterological paradox, Richard's paradox, and Russell's paradox can be shown to have a common structure. According to his analysis (see the introduction in Chapter 2), all of these four 'paradoxes' follow from a 'plain and simple logical truth' which is stated in Theorem (1). However, it should be noted that this theorem itself does not assert that there is no *S*-element which has *R* to all and only those *S*-elements which do not have *R* to themselves. Instead, it only asserts that if there is such a thing, it is not in *S.* Therefore, simply considering this theorem, there is no contradiction involved.

In the case of the Barber, we accept the solution that if there is a barber who shaves all and only the villagers who do not shave themselves, then he himself is not a member of this village. Thus, there is no paradox of the barber. In the case of the adjective ‘heterological’ in English, however, we cannot provide the same solution. We cannot simply dismiss ‘heterological’ as an adjective in English, in the same way as that is done for the Barber. It is because of this additional condition that there is a paradox: on the one hand, it is implied by (1) that there is no such adjective as ‘heterological’ in English; while on the other hand, it seems obvious that there is such an adjective in English.

Why do their treatments differ, even though both of them follow from the same theorem and have the same structure? Thomson has also asked the same question in his paper: “if the Barber and ‘heterological’ have a common structure, why should the one need so much more discussion than the other?” (Thomson 1962: 115) And he replies as follows:

> But, and perhaps more importantly, there is an obvious difference between ‘all villages’ and ‘all adjectives’. … But the number of adjectives available to those who speak English is not fixed and definite, for we invent new words - such as ‘heterological’. So someone who defined ‘heterological’ over the set of adjectives in current English use could quite naturally claim to mean all adjectives in such use at the time he introduced his definition. (ibid. 115)

That is to say, in Thomson's view, a common structure is not enough to suggest what kind of solution we should adopt to solve the paradoxes. Instead, we should look more closely into the complexities of different cases. This line of thought can also be found in Simmons' *Universality and the Liar* (1993). As I have discussed in Chapters 2 and 5, Simmons distinguishes two kinds of diagonal arguments: good diagonal arguments and bad diagonal arguments. In his analysis, a diagonal argument consists of several components: a side, a top, an array, a diagonal, a value, and a countervalue. In a bad diagonal argument (which is usually associated with logical paradoxes), the well-determinedness of its components is just assumed, and is not reduced to a contradiction in a *reductio* proof. To provide an adequate solution, then, one needs to investigate the issue where a component is not well-determined.

According to these authors, then, even if we can analyze these logical paradoxes as all having the structure of a diagonal argument, it does not automatically follow that all of them should therefore have a uniform solution. However, this point has been challenged by Priest (1994, 2002). Priest takes over the idea of a common structure from Russell, and argues that, because of the common structure, all logical paradoxes should therefore receive a uniform solution. His argument is briefly summarized below.

6.3.2. From Russell's Schema to the Inclosure Schema[7]

Priest's analysis begins with a common structure for all set-theoretic paradoxes: (Priest calls this 'Russell's Schema' (Priest 2002: 129)):

There is a property φ, and function δ such that

i. $\Omega =\{y: \varphi(y)\}$ exists; and
ii. if x is a subset of Ω, then
   a) $\delta(x)\notin x$, and
   b) $\delta(x)\in\Omega$

Generally, the contradiction emerges from substitution of 'Ω' for 'x' in (ii): (iia) $\delta(\Omega) \notin\Omega$ and (iib) $\delta(\Omega)\in\Omega$. Priest calls (iia) the *Transcendence Condition* and (iib) the *Closure Condition*. But Priest also wants to extend Russell's idea to semantic paradoxes.[8] In order to achieve that, he realizes that the above schema needs modification. He adds another property ψ to it, which requires that any subset of

[7] The content from section 6.3.2 to 6.3.6 is adapted from Zhong (2012), 'Definability and the Structure of Logical Paradoxes', *Australasian Journal of Philosophy*, Vol. 90, No. 4, pp. 779-88.

[8] In replying to my argument, Priest says: "At any rate, though Russell may never have stated the Inclosure Scheme explicitly in the general form, it certainly informs his mature theory of all the paradoxes of self-reference. It is therefore a little misleading to suggest, as Zhong does, that the formulation of the Inclosure Schema to apply to the semantic paradoxes is not Russell's." (Priest 2012: 790, footnote 5). However, it is questionable whether interpreting Russell's idea as Priest's Inclosure Schema is in accordance with Russell's original idea. For example, Landini (2009) argues, "we are left with the conclusion that Priest himself misconstrued Russell's (R) in generating his schema." (Landini 2009: 120). Nevertheless, the historical issue is not a central one for me. What I am interested in is the question whether the Inclosure Schema, as presented by Priest, can be an adequate description for the semantic paradoxes. I shall therefore leave the historical issue aside.

Ω (including Ω itself) should be definable. The structure after modification is called the 'Inclosure Schema' [ibid. 134][9]:

> There are properties φ and ψ, and a function δ such that
>
> i. $\Omega = \{y: \varphi(y)\}$ exists and $\psi(\Omega)$;
> ii. if x is a subset of Ω and $\psi(x)$, then
>   a) $\delta(x) \notin x$, and
>   b) $\delta(x) \in \Omega$

Then, taking Berry's paradox as an example, it can be tabulated as an instance of the Inclosure Schema:

Table 1

Paradox	$\boldsymbol{\delta(x)}$	$\boldsymbol{\varphi(x)}$	$\boldsymbol{\psi(x)}$	$\boldsymbol{\Omega}$
Berry's	the least natural number not in x.	x is a natural number definable in fewer than 19 words	x is definable in fewer than 14 words[10]	The set of all natural numbers definable in fewer than 19 words: DN_{19}

The explanation of Berry's paradox is similar to that of the set-theoretic ones, except that the totality $\boldsymbol{DN_{19}}$ should satisfy an additional condition: it should be definable in fewer than 14 words, i.e. $\psi(\boldsymbol{DN_{19}})$. The contradiction involved is $\delta(\boldsymbol{DN_{19}}) \in \boldsymbol{DN_{19}}$ and $\delta(\boldsymbol{DN_{19}}) \notin \boldsymbol{DN_{19}}$. Since Russell's Schema is just a special case of the Inclosure Schema, set-theoretic paradoxes certainly also can be made

[9] In his paper [1994], Priest does not use this name. There he calls the second version 'the Qualified Russell's Schema'. But in his book *Beyond the Limits of Thought* [2002], he uses the name 'the Inclosure Schema'. Since in that book he has a more detailed treatment of this issue, I shall follow his terminology there.

[10] The totality $\boldsymbol{DN_{19}}$ should be definable in fewer than 14 words, so that the least natural number that is not in $\boldsymbol{DN_{19}}$ is definable in fewer than 19 words.

instances of the latter. Therefore, Priest argues that he has successfully shown that all the logical paradoxes can be explained as instances of the Inclosure Schema, and thus have the same structure. However, it should be noted that showing that all the paradoxes have a common structure is simply the first step in his project. His purpose in arguing for a common structure is actually to advocate a uniform solution, which is manifested in the Principle of Uniform Solution (PUS): 'same kind of paradox, same kind of solution' [ibid. 166]. Furthermore, Priest wants to argue that such a uniform solution should be his dialetheism. Since this is a chain of reasoning, we need to examine whether each step is sustainable. Especially, we want to ask in what sense Priest can claim that the set-theoretic and the semantic paradoxes have a common structure and in what sense this common structure can guarantee a uniform solution.

6.3.3. Which 'Level of Abstraction' is Appropriate?

The change from Russell's Schema to the Inclosure Schema is interesting, for this seems already to display a structural difference between the two groups. There is no doubt that set-theoretic paradoxes could also be *made* into instances of the Inclosure Schema, but in that case, the satisfying of the newly added property $\psi(x)$ is quite artificial and trivial. Thus, one can argue, why couldn't the difference between semantic and set-theoretic paradoxes exactly consist in this new property ψ? In his criticism of Priest's argument, N. Smith [2000] raises the question concerning the 'level of abstraction': 'two objects can be of the same kind at some level of abstraction and of different kinds at another level of abstraction' [Smith 2000: 118]. He argues that, though at the most general level both set-theoretic paradoxes and semantic paradoxes are instances of the Inclosure Schema, at the more concrete level they are different, and thus deserve different treatments.

One problem with Smith's argument is that he does not provide the reason why the more concrete level is more important than the general level. Thus, it is quite reasonable that Priest responds to Smith by saying, 'the appropriate level at which to analyze a phenomenon is the level which locates underlying causes', where the term 'underlying causes' means 'the essential features of a situation that are responsible for something or other' [Priest 2000: 125]. No doubt the appropriate level is the one that locates the underlying causes for the contradiction. It remains controversial, however, whether the level of the Inclosure Schema is the appropriate level. To argue that it is, Priest says:

> Once one sees that a certain operation on any totality of objects of a particular kind generates a novel object of that kind, it becomes clear why applying the operation to the totality of all such objects must give rise to contradiction. (Priest 2000: 124)

Priest argues that the tension between the inclosure of a certain kind of totality and the transcendence of that totality is the essential feature of all logical paradoxes. For example, in the case of Berry's paradox, we seem to be able to use the phrase 'the set of all natural numbers definable in fewer than 19 words' to refer to a certain kind of totality, just as we seem to be able to use the phrase 'the set of all ordinal numbers' to refer to the totality of all ordinal numbers. And if we also accept that this ‘set’ is ‘definable’ and is a genuine set, then Berry’s paradox would look quite similar to the paradox of ordinal numbers. However, as discussed at the beginning of this chapter, it is simply an assumption that such a ‘set’ indeed exists. If we examine Berry’s paradox more carefully, it becomes clear that this assumption should be rejected.

6.3.4. Two Different Reasons why Totalities are not Sets

Since the discovery of Russell’s paradox, the voluminous literature on the foundations of set theory has called into question the principle on which the paradox seems to depend: the unrestricted comprehension principle. According to this principle, given any predicate (with one free variable) there exists a set the members of which are just those entities that have the property denoted by that predicate. Following Russell, we may say that for paradoxical sets such as the set of all ordinal numbers, the predicate involved cannot denote a property which has a set as its extension. At a glance, this seems parallel to the situation concerning the definability of the ‘set’ $\boldsymbol{DN_{19}}$. If we suspect whether one can legitimately talk about the ‘set’ of all the natural numbers definable in fewer than 19 words, it seems that the trouble with the paradox of all ordinals also consists exactly in the fact that the totality which is supposed as the extension of the property ‘being an ordinal number’ is not a legitimate set. Thus, the proponents of the ‘common structure’ view can still argue that both of them have the same reason for the contradiction, i.e. they both treat the problematic totality as a legitimate set. If we deny that they are sets, then there is no paradox in either case.

Roughly speaking, we may say that the problem with both paradoxes is that the totality involved cannot be accepted as a genuine set. After a careful examination, however, we can discern two different senses in which we say that the totality involved is not a set. For the totality of all ordinals, the problem lies in the fact that the size of the totality is unlimited, i.e. it is *absolutely infinite*[11] (using Cantor's terminology); for ***DN***$_{19}$, its size is limited, but the boundary of this totality is fuzzy. In other words, it's a finite but indefinite, flexible and dynamic totality. As argued in Chapter 5 and the beginning of the present chapter, these are distinctive features of all semantic notions.

In axiomatic set theory (specifically, ZF), the unrestricted comprehension principle is replaced by the axiom schema of separation: $\exists y \forall x(x \in y \leftrightarrow \exists z(x \in z \& \varphi(x)))$. This axiom effectively blocks all known set-theoretic paradoxes, since it only allows something to be a set when it is definable as a subset of something known on other grounds to be a set. Also, this axiom guarantees that no semantic paradoxes can be formulated in set theory, since there are some restrictions about the predicate φ. In his original construction of this

[11] This concept is different from 'transfinite'. A set whose size is transfinite is a legitimate object of mathematical study, without leading to any paradox. However, an absolutely infinite 'set' is one that can be mapped via a bijection to a proper class (following the terminology of von Neumann), and is paradoxical if it is treated in the same way as ordinary sets.

system, Zermelo (1967 (1908)) presented two considerations concerning this axiom:

> In the first place, sets may never be *independently defined* by means of this axiom but must always be *separated* as subsets from sets already given; thus contradictory notions such as “the set of all sets” or “the set of all ordinal numbers” . . . are excluded. In the second place, moreover, the defining criterion must always be definite in the sense of our definition in No. 4 (that is, for each single element *x* of *M* the fundamental relations of the domain must determine whether it holds or not), with the result that, from our point of view, all criteria such as “definable by means of a finite number of words”, hence the “Richard antinomy” and the “paradox of finite denotation”, vanish. (Zermelo 1967 (1908): 202)

At that time, Zermelo didn’t precisely delineate the idea that ‘the defining criterion must always be definite’. This task was completed by later researchers, the most important among whom was Skolem [1967 (1922)]. A defining criterion is definite in Skolem’s sense if it is expressible by a well-formed formula (*wff*) of a specified logical system. For example, he defined the notion ‘definite proposition’ as ‘*a finite expression constructed from elementary propositions of the form $a \in b$ or $a=b$ by means of the five operations mentioned*’ (Skolem 1967 (1922): 292-3).

Thus, by *wffs* of a first-order language in which the sole predicate constants are $\in$ and =, Skolem has characterized precisely which kind of expressions are definite and which are not. By this criterion, we can see the difference between set-theoretic paradoxes and semantic ones. The semantic notion ‘definable’ involved in the predicate ‘*x* is a natural number definable in fewer than 19 words' cannot be defined in such language. In a word, semantic notions (understood in their natural sense) resist rigorous formulation.

6.3.5. Is the Semantic Notion 'Definable' Definable?

Priest, however, seems to have a different understanding of whether a 'set' can be defined by some predicate, for he understands 'definable' as follows: 'call something *definable* if there is some non-indexical noun-phrase (of English) that refers to it' [Priest 2002: 131]. Since there is a non-indexical noun-phrase that refers to the paradoxical totality $\boldsymbol{DN_{19}}$, then according to his standard $\boldsymbol{DN_{19}}$ is definable. On the other hand, Priest also argues that 'the work of Gödel and Tarski showed how these notions could be reduced to other parts of mathematics (number theory and set theory, respectively)' (ibid. 142), so Ramsey's content criterion which relies on what terms are used is quite 'superficial'. It seems that Priest uses two standards. When accepting $\boldsymbol{DN_{19}}$ as definable (and consequently accepting the semantic notion 'definable' in it as itself definable), he uses a loose standard of 'definable'. While criticizing Ramsey's division, he employs a very rigorous standard (i.e. by referring to Gödel's and Tarski's work). Let's call the former the 'natural language standard', and the latter the 'formal language standard'. The problem is, as shown at the beginning of the chapter, that the additional condition that $\boldsymbol{DN_{19}}$ is definable is a necessary condition for the emergence of contradiction in the case of Berry's paradox. If we deny that $\boldsymbol{DN_{19}}$ is definable (in other words, is a set), then there is consequently no contradiction.

Since Priest cites Tarski's and Gödel's work (i.e. the 'formal language standard') to argue against Ramsey's content criterion, it is important for him to remain consistent and to use the 'formal language standard' both for Ramsey and for his own treatment of semantic notions.

It is well-known that Tarski's work on the definition of 'truth' is a double-edged sword. The negative aspect of his result is that if we try to define a semantic notion (e.g. 'true') for a language within that language, and if that language has the expressive power sufficient to have diagonalization, then this would inevitably lead to contradiction. On the other hand, Tarski also has provided a positive result, i.e. a rigorous and precise definition for the semantic notion 'true' in a formal language. Thus, semantic notions in the unqualified sense are not definable (according to the formal language standard), while in the qualified sense they are definable. It is only in the unqualified sense that paradox arises, for semantic notions in that sense have their peculiar characteristics: indefinite, non-representational, and dynamic. Once they are reconstructed according to the formal language standard, paradox vanishes. Take the semantic notion 'definable' as an example. While Gödel got the inspiration for the proof of his first Incompleteness Theorem from the Liar paradox, we can construct a similar proof by a careful examination of Berry's paradox, and that would yield another sentence undecidable in the formal theory involved.[12] Let us consider a

[12] For the following proof I have consulted Chaitin [1975], Boolos [1989] and Boolos et al. [2007: 227-9]. However, the version I present here is different from both Chaitin's and Boolos'. Firstly, I

notion ‘denominable’, the counterpart in the language of first order arithmetic **L** of the notion ‘definable’. To make the notion precise, it must be defined relative to a theory **T** of **L**, where **T** is a consistent, axiomatizable extension of minimal arithmetic. So, let us begin with the definition of the notion ‘denominate’ as follows:

(1) A number n[13] is *denominated* in **T** by a formula $\varphi(x)$ iff ‘$\forall x\ (\varphi(x) \leftrightarrow x=\mathbf{n})$’ is provable in **T**.

A number n is denominable in **T** if and only if there is some formula (with one free variable) $\varphi(x)$ in **T** such that n is denominated by $\varphi(x)$. Someone may worry that in terms of ‘n is denominated by some formula φ in **T**’, this treatment requires illicit quantification over properties. But this is not the case. The notion ‘denominate’ is defined by the syntactic notion ‘provable’, which in standard treatment is defined as quantification over Gödel codes of predicates. Therefore, what is required for ‘denominable’ is also quantification over the codes of predicates, which is licit. Let $\boldsymbol{\varphi}$ stand for the Gödel code of the formula φ. And let us abbreviate ‘n is denominated by $\varphi(x)$’ as ‘Den($\mathbf{n}$, φ)’. Thus:

(2) A number n is *denominable* in **T** (*simpliciter*) iff $\exists\boldsymbol{\varphi}$ (Den($\mathbf{n}$, φ)).

do not use the notion ‘complexity’, which is essential in Chaitin’s information-theoretic version. Secondly, I’ve mentioned a specific number (i.e. 10↑10) as the upper limit of the number of symbols, which I think is more analogous to Berry’s original paradox, while Boolos [1989] doesn’t mention any *specific* number. Also, for the version in Boolos et al. [2007], it seems that the authors fail to distinguish the definition of ‘a number n is denominable in **T** by a formula $\varphi(x)$’ from the definition of ‘a number n is denominable in **T** *simpliciter*’, while here I’ve paid more attention to this distinction.

[13] For any natural number n, let $\mathbf{n}$ be the expression consisting of **0** followed by n successor symbols s. Therefore, $\mathbf{n}$ stands for the number n.

The phrase in Berry’s paradox which is supposed to denote a number is ‘the least natural number not definable in fewer than 19 words’. As suggested above, we may replace ‘definable’ with ‘denominable’. And we also want to replace ‘words’ with ‘symbols’, since in a formal system, it’s more convenient to count symbols (where the lowly blank is also a symbol). Then the phrase becomes: the least natural number not denominable (in **T**) in fewer than 73 symbols. This phrase (which contains 72 symbols) is supposed to refer to a number in **T** which is undenominable in fewer than 73 symbols. To get the intended contradiction, we should translate this phrase in **T** in fewer than 73 symbols. However, there are several difficulties. The definition of ‘denominable’ stated above is based on the definition of ‘denominate’, which is itself based on the notion ‘provable’. If we want to express ‘denominable’ in **T**, first of all we should express the notion ‘provable’ in **T**, and this could hardly be done within 73 symbols. Fortunately, Gödel’s work assures us that we can express the notion ‘provable’ in **T**. With the apparatus of Gödel numbering, this goal can be achieved in principle. Although it cannot be done in 73 symbols, it would not require more symbols than those in an ordinary encyclopedia. Thus we may raise the word limit so that it can accommodate the need to express this phrase in **T**. Let’s say 10↑10, where ‘↑’ expresses the super-exponential function. Thus the number 10↑10 is an astronomically large number, which ensures that we can express the phrase ‘the least natural number not denominable in **T** in fewer than 10↑10 symbols’ in fewer than 10↑10 symbols in **T**. On the other hand, it's easy to work out an algorithm to compute how many symbols are contained in one formula. Let's use 'Sym(φ)' for the function that maps the formula φ to the number of symbols in that formula. Thus:

(3) A number *n* is not *denominable* in **T** (*simpliciter*) in fewer than 10↑10 symbols iff $\neg\exists\boldsymbol{\varphi}\ ((\mathrm{Sym}(\varphi)< \mathbf{10{\uparrow}10})\ \&\ \mathrm{Den}(\mathbf{n}, \varphi))$.

Since we need *the least* natural number that is not denominable in **T** in fewer than 10↑10 symbols, the last step of our definition is to add the part for 'the least':

(4) A number *n* is the least natural number not *denominable* in **T** (*simpliciter*) in fewer than 10↑10 symbols iff $\neg\exists\boldsymbol{\varphi}\ ((\mathrm{Sym}(\varphi)<\mathbf{10{\uparrow}10})\ \&\ \mathrm{Den}(\mathbf{n}, \varphi))\ \&\ \forall u\ (u<\mathbf{n}\rightarrow\exists\boldsymbol{\psi}\ ((\mathrm{Sym}(\psi)<\mathbf{10{\uparrow}10})\ \&\ \mathrm{Den}(u, \psi))$.

Notice that, though '10↑10' is an extraordinarily large number, there would still be natural numbers that cannot be denominated by formulas with fewer than 10↑10 symbols. So the phrase 'the least natural number not denominable in **T** in fewer than 10↑10 symbols' can uniquely denote a natural number. Let's call this number **m**. Although there has never been a single person who knows how large *m* is, and there probably won't be any one in the future interested in working out this number, nevertheless, this job can be done in principle, since there is an effectively computable process following which one (or a computer) can work out this number theoretically. Briefly, the idea is that according to their alphabetical order, one can list every formula in **T** which has one free variable and which can be satisfied by and only by a particular natural number, then check from the simplest formula, until one eventually goes through all the formulas expressed in fewer than 10↑10 symbols. The least natural number which cannot satisfy these formulas is the number we are looking for: *m*.

With all these preliminaries out of the way, now we can construct the G ödel-Berry sentence. Let the formula ‘GB (x, $\mathbf{10\uparrow 10}$)’ mean ‘x is the least natural number not denominable in **T** in fewer than 10↑10 symbols’. Then,

(5) GB ($\mathbf{m}$, $\mathbf{10\uparrow 10}$) iff
$\neg\exists\boldsymbol{\varphi}\ ((\mathrm{Sym}(\varphi)<\mathbf{10\uparrow 10})\ \&\ \mathrm{Den}(\mathbf{m},\ \varphi))\ \&\ \forall u\ (u<\mathbf{m}\rightarrow\exists\boldsymbol{\psi}$
$((\mathrm{Sym}(\psi)<\mathbf{10\uparrow 10})\ \&\ \mathrm{Den}(u,\ \psi))$.

We know that ‘GB ($\mathbf{m}$, $\mathbf{10\uparrow 10}$)’ is true. And we know that the number of the symbols needed to express ‘GB (x, $\mathbf{10\uparrow 10}$)’ in **T** is smaller than the number 10↑10. So, there is a formula in **T** with fewer than 10↑10 symbols that actually *denotes*[14] (but does not *denominate*) *m*. But this truth cannot be proved in **T**, otherwise it would be contradictory to ‘$\neg\exists\boldsymbol{\varphi}((\mathrm{Sym}(\varphi)<\mathbf{10\uparrow 10})\ \&\mathrm{Den}(\mathbf{m},\ \varphi))$’, and the system would cease to be consistent. Since we have assumed that **T** is consistent, Berry’s paradox (which is stated loosely in natural language), when it is reconstructed according to the formal language standard, becomes a *reductio ad absurdum* for the unprovability of the sentence ‘GB ($\mathbf{m}$, $\mathbf{10\uparrow 10}$)’ in **T**. Moreover, this sentence is undecidable in **T**, for the negation of it is not a theorem of **T** either. Therefore, the above demonstration shows that when the semantic notion

[14] The notion ‘denote’ is used here loosely in the sense that it doesn’t require to be specified relative to a particular formal theory.

‘definable’ has been defined rigorously (in terms of ‘denominable’), there is no paradox, but a *reductio* proof of the limits of **T**.

As Tarski [1935, 1944] has pointed out, an effective way to block the Liar paradox is to distinguish object language from metalanguage. (Note: to ‘block’ a paradox does not mean that it is automatically a good solution to a paradox, according to the criteria in Chapter 3.) Similarly, an effective way to block Berry’s paradox is to define ‘definable’ relative to a certain theory. In the above demonstration, if we use the phrase ‘the least natural number not denominable in fewer than 10↑10 symbols’ in the unqualified way, then it will have the same problem as the phrase ‘the least natural number not definable in less than 19 words’. The counterpart notion ‘denominable’ can only be well defined relative to a particular theory, say **T**. This is to set a sharp boundary for this notion. But once this is done, the phrase GB (*x*, **10↑10**) cannot denominate *m* in **T** at all, though we can say in an expanded theory **T**' that the phrase GB (*x*, **10↑10**) expressed in **T** actually *denotes* the natural number *m*. Since Priest regards Tarski’s work as the paragon of formal semantics, it’s reasonable for us to believe that our treatment of the semantic notion ‘definable’ is better than Priest’s own definition, and better meets the requirement to ‘reduce semantic notions to other parts of mathematics’. But if one accepts that, then there is no paradox, but just a metalogical result about the limits of representability of **T**.

6.3.6. The Argument against a ‘Uniform Solution’

What Priest wants to argue is that, since all the logical paradoxes can be made into instances of the Inclosure Schema, and since this schema reveals the core reason for all of them, then all these paradoxes are *essentially* the same, so that they should receive the same solution. What I want to argue is that, though they all can be made to fit one scheme, this could be just a common phenomenon for them. There are several components in the proposed structure, and one component could be vital for one kind of paradox, but trivial for another kind. If one wants to dig into the ‘underlying causes’ of one kind of paradox, one will still have to ask which component is responsible in that case. And, if one can only find that reason on a more concrete level, then, even if all of them can be generalized as ‘the same’ on some level of abstraction, that would still not be the most important level. And if we want to provide an adequate solution to these paradoxes, then this solution must be provided according to the ‘underlying causes’ of each kind of paradox.

Consequently, the argument boils down to this issue: what in the end is the root cause of the semantic paradoxes? Priest diagnoses it as being that ‘a certain operation on any totality of objects of a particular kind generates a novel object of that kind’ [Priest 2000: 124]. It is true that the problem with both the set-theoretic and the semantic paradoxes is that, if we admit that the totality in question is a genuine set, then there’s an operation that works on this ‘set’ to generate a new problematic object which causes inconsistency. Therefore, on this level, we can say that the reason for the paradoxes in both cases is the assumption that the totality in question is a genuine set. But if one asks further—What is wrong with this assumption? Why can’t we assume that the totality is a set?—then one will receive different answers for different cases. As Zermelo [1967 (1908)] has pointed out, the problem with semantic paradoxes is that the defining criterion which aims to define the totality is not definite, for there are semantic notions involved and the criterion for whether a semantic notion is applicable to a given object is indefinite. Thus, if we provide a clear and precise criterion to specify whether the notion is applicable to a given object, as we did for ‘denominable’ in the last section, then the paradox goes away. But one should keep in mind that, though the notion ‘denominable’ is supposed to be a precise formal regimentation of the notion ‘definable’, it cannot fully capture all the intuitions involved in the latter. That is why someone may still think that the formal regimentation, which is a *reductio* argument ending up with some theorem for a certain system, could not be counted as a solution to Berry’s paradox (understood in its natural language sense). An adequate solution should reveal why semantic notions cannot be definite. As argued at the beginning of this chapter, they cannot be definite because they are non-representational, so that they cannot be defined by terms which are representational. The formal construction in the last section cannot provide such a solution. Instead, it only serves as a comparison, through which one can see clearly the peculiar feature of the naïve semantic notion ‘definable’ (i.e. that it cannot be defined by representational notions) and recognize the importance of this feature for explaining semantic paradoxes.

This essential point, however, cannot be manifested in the Inclosure Schema. In the case of Berry's paradox, the new condition '$\psi(x)$' is assumed to be fulfilled, just as other components of the schema are fulfilled. Then, at this level, if one wants to know the reason for the paradoxical result, one could only say: 'there must be something wrong with *at least one* of these components in this structure'. But the structure itself cannot tell her which component is problematic. If she wants to understand the reason more deeply, then she would have to ask further questions: which component goes wrong, and why? What I have attempted to do in the previous sections is to pinpoint which component of the semantic paradoxes goes wrong. Therefore, the disagreement between my position and Priest's is, to put it informally, whether it is necessary to ask these further questions. Proponents of the common structure view would say that everything has been explained at the level of the Inclosure Schema, but I argue that it has not. The aim of the research on paradoxes is to find out the root cause of the contradiction and to understand more deeply the key notions involved. Since these further questions help to show the peculiar feature of the semantic paradoxes, and help to promote our understanding of semantic notions, I do not see why we should not ask them. And, since answers to these crucial questions will differentiate semantic paradoxes from set-theoretic ones, we should retain the traditional separation of the two groups of logical paradoxes. Therefore, a common structure at a very general level cannot guarantee a uniform solution for both groups.

6.3.7. Responses to Priest's Criticism[15]

In reply to my argument, Priest says:

> Of course, if **T** is inconsistent, the sentence may well be provable, in the same way that an inconsistent theory can prove its own "Gödel undecidable sentence". If this is an argument against dialetheism, then Zhong's simple assumption that **T** is consistent begs the question. (Priest 2012: 793)

However, this criticism does not apply to the argument presented in the previous sections. Dialetheism is the view that there is some sentence (called a 'dialetheia') A such that both it and its negation, ¬A, are true. My argument that the 'common structure' does not automatically guarantee a uniform solution for the paradoxes with such a structure is not necessarily an argument against dialetheism. As mentioned above, there are three steps in Priest's argument. Even if there is no problem with the first two steps, there still could be other alternatives for the last one. For example, one can use the common structure to argue for a uniform solution of another kind, say the axiomatic treatment for both set theory and semantics. Thus, the argument constructed above is not necessarily one against

[15] Priest (2012) has also some other criticisms for my argument, which I shall deal with in the next chapter.

dialetheism. It is simply against the first two steps, especially the second one. Therefore, it does not beg the question.

But, since I want to compare my own functional-deflationary solution for semantic paradoxes with dialetheism, and argue that the former is a better solution compared with the latter, is there any danger that my argument begs the question?

The similarity between dialetheism and my explanation is that both of them recognize the special role of diagonalization in the semantic paradoxes and the dynamic nature of the totality involved. Priest regards these totalities in logical paradoxes as both ‘closed’ and capable of being ‘transcended’, so that they are contradictory objects. That is to say, he also recognizes that there is a fundamental distinction between ordinary notions and semantic notions. The difference between dialetheism and my explanation consists in the fact that, according to dialetheism, semantic notions correspond to a special kind of object, i.e. contradictory objects; while the functional-deflationary interpretation regards them as performing the function to govern our usage of language. The latter explanation is called ‘deflationary’ because it does not acknowledge that there are certain ‘objects’ described by semantic notions, either ordinary or contradictory. On the contrary, it emphasizes that such notions are non-representational. They work together to form the framework which enables us to use language to represent objects and the world; they themselves are not objects to be represented.

It seems that there is no other source available to argue against dialetheism. Any argument should assume consistency, and thus faces the danger of begging the question. If we compare these two solutions, the only criterion that can make one superior to the other is Occam's razor. If both of them have the same explanatory power, then we should favor the one positing fewer entities. The functional-deflationary view can explain intuitions associated with natural language well, and it does not need any meta-language above natural language to avoid contradiction. Compared with dialetheism, it treats semantic notions as corresponding to certain functions governing our usage of language rather than as denoting some 'contradictory objects'. As Williamson observes, to admit the existence of contradictory objects should be the last option in our search for a solution:

> If one abjures contradictions, it is certainly harder to think about the limits of thought; but that extra difficulty may in the end produce greater depth. (Williamson 1996: 334)

6.4. Final Remarks about Semantic Paradoxes

Semantic notions, as argued in Chapter 5 and this chapter, are not representational. This feature is also called 'deflationary', for they do not have the content that ordinary expressions have. Semantic paradoxes, such as the Liar, the heterological paradox, and paradoxes of definability, are all caused by confusing non-representational terms with representational ones. Thus, they all can be solved by clarifying the relative confusion: the Liar sentence and the heterological sentence do not have truth values, and phrases in paradoxes of definability (such as Berry’s paradox) do not denote an object.

On the other hand, we should not understand the term ‘non-representational’ as suggesting that these notions ‘represent another kind of object’. This differentiates the view advocated in this thesis from Cartesianism and dialetheism. Because of their functional role in our usage of language, semantic notions do not represent the world as other ordinary expressions do. But this by no means suggests that they represent a special kind of objects (Cartesian objects, contradictory objects, etc.). If we understand ‘physicalism’ as the view asserting that all events are physical events and have a physical description, rather than the view that all facts are reducible to physical facts, then the view advocated in this thesis is totally compatible with physicalism. Also, without positing the existence of any ‘special objects’, this view is thus ontologically economical and more desirable.

One may wonder whether a similar solution could also be provided to the set-theoretic paradoxes. For example, it is very intuitive that the naïve understanding of ‘set’ is quite similar to that of semantic notions. Then shouldn’t the set-theoretic paradoxes also be solved in the functional-deflationary way? In the next chapter, I will discuss this issue and defend the standard solution (i.e. axiomatization) of the set-theoretic paradoxes.

Chapter 7: Set-Theoretic Paradoxes

In this chapter, I discuss the philosophical issues concerning the set-theoretic paradoxes, especially the philosophical justification for transfinite sets as proper mathematical objects and their difference from the absolute infinite. The chapter begins with Cantor’s domain principle, which he employed to argue for the legitimacy of transfinite sets and numbers as proper mathematical objects. I argue that the domain principle cannot justify the transfinite while at the same time safely exclude the absolute infinite from Cantor’s set theory. On the contrary, the justification for transfinite sets is their usefulness in mathematical construction. Second, I consider another philosophical argument for the transfinite, the limitation of size theory; and argue that this theory again fails to provide an independent criterion for distinguishing the transfinite from the absolute infinite. Finally, I argue that the real problem for the absolute infinite is that it is indefinitely extensible, like the semantic notions in the semantic paradoxes. However, though these two groups of paradoxes have the same underlying cause for contradiction, they still deserve different solutions, because of the different aims of research in the two different areas.

7.1. The Domain Principle

7.1.1. Potential Infinity *vs*. Actual Infinity

In the late 19th century, when Cantor was developing the theory of transfinite numbers and na ïve set theory, he faced serious doubts about the existence of such entities as transfinite numbers and the corresponding sets. Thus, he not only needed to show that the concept ‘transfinite’ could be thought consistently, but also had to provide philosophical arguments for treating the corresponding sets as legitimate objects for mathematical study. The most important philosophical argument that Cantor provided for the existence of transfinite sets and numbers relies on an important claim which is called ‘The Domain Principle’ (following Hallett’s terminology):

> Any potential infinity presupposes a corresponding actual infinity. (Hallett 1984: 7)

The two terms ‘potential infinity’ and ‘actual infinity’ stem from Aristotle. In *Metaphysics* IX, Aristotle wrote about ‘potential’ and ‘actual’ infinity as follows:

> …but the infinite is not potentially in this way, namely that it will be actually separate, but by coming into being. For it is the division’s not coming to an end which makes it the case that this actuality is potentially, and not the infinite being separated. (Aristotle: *Met.* 1048b13-18, Makin 2006: 7)

Usually, people understand this paragraph as saying that there are two kinds of infinity: potential and actual, but only the potential infinity exists, and it would never become an actual infinity. However, in *Physics III,* Aristotle says that the phrase ‘potential existence’ is ambiguous. (*Phys.*206a18) According to the usual meaning of the term ‘potential’, what is potential will become actual in appropriate circumstances, and there is nothing impossible in its being actual. If this word did not change its meaning, then it means that the potential infinite could become an actual infinity in some cases.[1] This seems to contradict what Aristotle said above. Since Aristotle's theory about potential and actual infinity had become orthodox for western intellectuals until the 19th century, and since it is usually believed that Cantor's theory on actual and potential infinity contradicts Aristotle's doctrine, it is necessary to make this issue clear: in what sense does Aristotle use 'potential' and 'actual' to describe infinity, and in what sense does Cantor use these two terms?

To answer these questions, it is helpful to begin with the examination of Aristotle's explanation about the two senses of the word 'is'. To differentiate two senses for 'is', Aristotle compares a statue with a contest:

> (We must not take 'potentially' here in the same way as that in which, if it is possible for this to be a statue, it actually will be a statue, and suppose that there is an infinite which will be in actual operation.) Since 'to be' has many senses, just as the day is, and the contest is, by the constant occurring of one thing after another, so too with the infinite. (In these cases too there is 'potentially' and 'in actual operation': the Olympic games *are*, both in the sense of the contest's being able to occur and in the sense of its occurring.) But [the infinite's being] is shown in one way in the case of time and the human race, and in another in the case of division of magnitudes. In general, the infinite is in virtue of one thing's constantly being taken after another-each thing taken is finite, but it is always one followed by another; but in magnitudes what was taken persists, in the case of time and the race of men the things taken cease to be, yet so that [the series] does not give out. (Aristotle: *Phys*. 206a20-206b2, Hussey 1993: 14)

A statue exists as an individual entity, while a contest is a process. Both of them have potential and actual existence. A wooden statue exists potentially in wood,

[1] Hintikka (1966) holds this position. In his interpretation of Aristotle, Hintikka advocates a thesis called 'the principle of plentitude', which says that every genuine possibility is sometimes actualized (Hintikka 1966: 197).

and becomes actual as a product of the sculptor's activity. On the other hand, a process is not an entity. The way of existence for a process is 'the constant occurring of one thing after another'. (Aristotle: *Phys.*206a20) When we say that a contest exists potentially, it means that it may occur, and its being actual means that it is actually occurring. Aristotle says that the way that the infinite 'is' is in the sense that a contest 'is'. However, since 'actual' still applies to a contest but there will not be an actual infinity, it seems that there still are some differences between a contest and the infinite when we think about 'potential' and 'actual'.

The key to this problem is that the infinite is neither an entity nor a process. Rather, it is an attribute of a process. When Aristotle says that the infinite exists in the same way as a contest, he means that the infinite is an attribute of a process, but it cannot be an attribute of an (actual) individual entity. This is what he means by 'it will never actually have separate existence'.

When the infinite applies to a process, this process has a property that it is endless. As with a contest, for such a process we still can distinguish potentiality and actuality. Take the process of dividing a magnitude as an example. Any actual process of dividing has an end; it stops or will stop at some time. Therefore, no endless dividing actually occurs, and there is no actual infinity associated with such a process. However, we still can say that the infinite of such a process is potential, which means:

> For any dividing process *S* and any step *x* of *S*, there *could* be a step *y* of *S* which comes after *x*.

But to say that this endless process is actual, it should be:

> For any dividing process *S* and any step *x* of *S*, there *is* (or *will be*) a step *y* of *S* which comes after *x*.

Since any actual dividing has or will have an end, it cannot be actually infinite. In other words, it cannot be actually endless.

However, there is still a problem for this account. For natural numbers, it seems that the 'abstract' process of generating more and more natural numbers is actually endless. Here by using the word 'abstract', I do not mean that a natural number is generated by a person's thinking of or saying it, but that the sequence of natural numbers exists in the abstract sense. Accordingly, the kind of infinity associated with the sequence of natural numbers should be formulated in terms of 'is':

> For any given natural number *n*, there *is* a larger natural number *n+1*.

Consequently, according to Aristotle's explanation for a process, there should be an actual infinity associated with the process of generating natural numbers, because by saying 'there is', instead of saying 'there could be', it means that all natural numbers in this process already exist.

How could Aristotle solve this problem? A quick answer is that he would not accept the claim that the sequence of natural numbers exists in the abstract sense, since abstraction for him is a mental operation of separating quantity or number from actual entities and there is no realm of abstract objects in his ontology. However, even without resorting to his general ontological theory about abstract objects, there is still a way to make his doctrine on potential/actual infinity consistent. We can find an answer for this issue from his doctrine of 'anywhere' and 'everywhere' division of lines. In his solution to Zeno's dichotomy paradox (*Phys.*263b3-9), Aristotle admits that it is indeed impossible for the runner (call him Achilles) to complete infinite tasks (e.g. traversing infinitely many points or units) within a finite period of time. However, actually, Achilles has not traversed infinitely many points, because there are not infinitely many *actual* points on a line. Aristotle distinguishes potential points from actual points. For a point in a line to be actualized, one should 'do' something at that point, such as stopping at it, or reversing one's direction at it, or dividing the line at it. Mere continuous motion is insufficient to actualize any point. For any point in the line, it is possible to be actualized, e.g. one can divide the line at it. However, it is impossible for one to actualize every point in a line. In a word, a line is divisible through and through in the sense that there is one point anywhere within the line and you can take them singly one by one (i.e. distributively), but not in the sense that you can take all of them simultaneously (collectively). (*On Generation and Corruption* 317a3-10)

Similarly, for natural numbers, Aristotle might say that, for any given natural number *n*, one can have a number *n+1* which is larger than *n*. However, it is impossible for one to take all the natural numbers simultaneously. In other words, the sequence of natural numbers is not given as a completed entity. We cannot have all the natural number once for all.

Now let us return back to the domain principle. When Cantor claims that every potential infinity presupposes an actual infinity[2], is this principle contradictory to Aristotle's doctrine? If we take the terms 'potential' and 'actual' infinity as meaning the same as those words in Aristotle's theory, then we can see that Cantor's domain principle contradicts what Aristotle has said. However, though Cantor uses the term 'potential infinity', what he actually means is a variable quantity, and by 'actual infinity' he means a domain for that variability (Cantor's theory on variability and domain will be explained in the next section), while in Aristotle's theory the potential infinity has nothing to do with variability. When Aristotle discusses the potential infinity of natural numbers, he does not mean 'an arbitrarily large natural number'. On the contrary, he means that for any *given* natural number *n*, there is a larger one. This given number *n* is both finite and definite. We only add '1' to this finite natural number to get a larger finite natural number. All this is within the realm of the finite. Therefore, the concept 'potential infinity' in Aristotelian theory is simply an extrapolation of finite natural numbers. There is no clue why this extrapolation of finite natural numbers should presuppose a concept of actual infinity, as required by Cantor's domain principle.

[2] Cantor has an argument for this principle, which is going to be discussed in the next section.

To better illustrate this idea, we may compare Aristotle's potential/actual infinity with another pair of concepts, which were discussed by Leibniz. In his discussion of infinity, Leibniz made a distinction between the 'syncategorematic infinite' and the 'categorematic infinite'. The categorematic infinite equates to what Cantor means by 'an actual infinite set'. In this sense, 'to say that there are infinitely many parts is to say that there is a number of parts greater than any finite number.'[3] Take natural numbers as an example. Let 'N' stand for natural numbers, and '>' stands for the relation 'larger than'. The categorematic infinite can be represented as follows:

> (1) $\exists x \forall y\ (\mathrm{N}y \rightarrow x > y)$: there is a number which is greater than any natural number.

In contrast, the syncategorematic infinite cannot be treated in this collective way, but only in a distributive way. According to the syncategorematic sense of infinity, to say that there are infinitely many natural numbers is to say that no matter how many terms you take, there are more. In symbols:

> (2) $\forall x \exists y\ (\mathrm{N}x \rightarrow (y \neq x\ \&\ y > x))$: for any given natural number, there is a number that is greater.

It implies:

> (3) $\neg \exists x \forall y\ (\mathrm{N}x\ \&(y \neq x \rightarrow x > y))$: there is no greatest natural number.

[3] Arthur (*forthcoming*): 'Leibniz's actual infinite in relation to his analysis of matter', page 11.

It is clear that from (2) we cannot infer (1). Leibniz only accepts the syncategorematic understanding of the infinite[4]. As pointed by Arthur (forthcoming), “The statement is called *syncategorematic* because the term ‘infinite’ occurs in it, but that term does not actually have a referent corresponding to it.”[5] It is clear that the syncategorematic concept of infinity does not presuppose the categorematic one. As argued by Arthur, Leibniz can consistently have the syncategorematic view of infinity, and could build another foundation for mathematics which is different from that given by Cantor. One may argue that there should be a domain for the variables *x* and *y* in (2) and (3), and that is Cantor’s actual infinity. However, even if we need such a domain of all the natural numbers, it does not follow necessarily that such a domain must be a set in the Cantorian sense.

Therefore, ‘potential infinity’ in the Aristotelian sense can be understood consistently, i.e. as the extrapolation of finite existence, which does not have to presuppose an actual infinity (in Cantor’s sense). If the term ‘potential infinity’ in

[4] Leibniz’ theory of infinity serves as the technical background for his account for the division of body. He insists that body is actually infinitely divided. This means that, for any assignable number of divisions, there *are* (not simply ‘could be’, but *actually are*) more divisions. This is different from Aristotle’s potential infinity. However, Leibniz still denies we can take all these divisions together as a whole. His reason is that to treat infinite terms as a whole will lead to contradiction, i.e. the whole will equal to its part. For example, intuitively, there are more natural numbers than even numbers. However, if we take all the natural numbers and all the even numbers as a whole respectively, then there would be a one-to-one correspondence between these two collections; thus it can be concluded that there are as many natural numbers as even numbers. This contradicts the part-whole axiom which states that a whole is larger than any of its proper parts. Leibniz thinks that this consequence is unacceptable; thus the Categorematic understanding of infinite terms should be avoided. For Leibniz’ theory about infinite, see Leibniz, 2001: *The Labyrinth of the Continuum: Writings on the Continuum Problem, 1672-1686* (ed. Arthur), especially ‘infinite numbers’, pp. 83-101.

[5] Arthur (*forthcoming*): ‘Leibniz’s actual infinite in relation to his analysis of matter’, page 11.

Cantor’s domain principle is understood in this Aristotelian sense, then it is not clear why we should accept this principle.

7.1.2. Variability and the Domain

It should be recognized that Cantor’s term ‘potential infinity’ has a modern sense which is different from Aristotle’s conception. In the 19th century, mathematicians understood ‘potential infinity’ as signifying a variable quantity, which is useful in various mathematical fields. For example, when we consider the series of natural numbers, we may say phrases such as ‘for any natural number *n*’. Similarly, we use the concept ‘an arbitrarily small quantity’ when we discuss limits. The ε-δ method introduced in the 19th century relies heavily on the concept ‘variability’. However, one cannot speak of variability without speaking of variability over a completed domain. Thus, in terms of ‘a variable quantity’, Cantor endowed the concept ‘potential infinity’ with another kind of meaning, which provides the possibility to relate a potential infinity to an actual infinity, i.e. a completed, infinite domain. Furthermore, Cantor argues that such a domain cannot be variable. On the contrary, it should be determinate and fixed, so that it can provide support for the variable quantity. In other words, it should be treated as a finished, definite, and infinite whole. Such a ‘whole’ is what Cantor called a ‘set’. In his own words:

> There is no doubt that we cannot do without *variable* quantities in the sense of the potential infinite; and from this can be demonstrated the necessity of the actual infinite. In order for there to be a variable quantity

> in some mathematical study, the 'domain' of its variability must strictly speaking be known beforehand through a definition. However, this domain cannot itself be something variable, since otherwise each fixed support for the study would collapse. Thus, this 'domain' is a definite, actually infinite set of values. Thus, each potential infinite, if it is rigorously applicable mathematically, presupposes an actual infinite. (Cantor 1886: 9, English translation from Hallett 1984: 25)

There are two steps in Cantor's argument: first, a variable quantity presupposes a domain; second, such a domain is a set (in the technical sense). It makes sense to say that a variable quantity presupposes a domain. We may also admit the existence of an infinite domain, at least in a loose sense. However, one still can doubt whether such a domain is a 'set' in the Cantorian sense. For, even in ordinary conversation, there is always an implicit understanding of a domain, but it does not follow that such a domain must be a set in the technical sense in Cantor's theory. For example, Kripke has recently published a paradox of self-reference concerning the 'set' of all times when I am thinking about a 'set' of times that does not contain that time. In this paradox, the quantifier quantifies over times, which are defined by my thinking, and the domain could be 'all the times when I am thinking about something'. (Kripke 2011: 373-9) There is no problem in conceding that there is such a domain, but it is very questionable why we should treat this domain as a set. Thus, though the first step in Cantor's argument can be justified, the second step needs justification. There must be some reason to establish it.

7.1.3. The Usefulness of 'Domain' in Mathematical Research

Mathematicians, of course, are not interested in domains like ‘all the times when I am thinking about something’. But they do care about domains like ‘all the natural numbers’, ‘all the real numbers’, etc., because such domains are extremely useful and important in mathematical construction. For example, Cantor has another argument for treating such completed infinite domains as sets, the so-called ‘irrational number argument’. In his time, the standard definition of ‘irrational numbers’ was based on the concept of a completed, transfinite domain. That is why Cantor claims: “the transfinite numbers *stand or fall* with the finite irrational numbers; they are alike in their most intrinsic nature; for the former like these latter (numbers) are definite, delineated forms or modifications of the actual infinite.”[6]

To illustrate what Cantor means in the above quotations, we can take Dedekind’s ‘cut’ definition for irrational numbers as an example. Dedekind points out that the series of rational numbers has the following property: for any two rational numbers, there still could be a third rational number in between. However, such a series still can be divided into two parts, so that all the rational numbers of the second part come after all rational numbers of the first part, and no rational numbers lie between them, while yet the first part has no last term and the second part has no last term. For example, let us look at the following set[7]:

[6] Cantor (1932), pp. 395-6. English translation is cited from Dauben (1979), p.128.

[7] It may beg the question if we treat this term here in the Cantorian sense. However, we can just understand it as suggesting a kind of totality in a very loose sense. To understand it as the Cantorian set is the conclusion to be established by this argument.

$A= \{p \mid p \in Q, p<0 \text{ or } p^2<2\}$

If we arrange all the numbers in this set according to a certain relation (e.g. 'greater than'), then this sequence will approach to a certain point (let us call it '√2') indefinitely while it never can reach it. On the other hand, all rational numbers which do not belong to A (they form another set Q/A) come after that point, and any number in Q/A is greater than any number in A. However, this point (√2) belongs to neither part. Therefore, Dedekind thinks that there should be a new term (number) corresponding to it. This new number serves as the 'limit' of the two sequences of rational numbers, while it belongs to neither of them. Therefore, it is an irrational number. With such numbers introduced, they together with the rational numbers form the set of real numbers R, which has a new property that the series of rational numbers does not possess. That is, no matter how you 'cut' the series of real numbers, it always can secure one part which has an endpoint, while the other part has no such point. This property is identified as the 'continuity' of the real numbers:

> **Continuity**: a series is continuous iff all the terms of the series can be divided into two parts, such that the whole of the first part precedes the whole of the second part, and either the first part has a last term or the second part has a first term, but never both.[8]

[8] This definition of 'continuity' is from Russell's interpretation of Dedekind's original definition of 'continuity', and Russell believes that this definition gives 'what Dedekind *meant* to state in his axiom' (Russell 1903: 279). Dedekind's original definition is stated as follows: "If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions." (Dedekind 1901: 5)

This property can be better illustrated by the comparison between real numbers and other kinds of numbers. As we have already seen above, the division of rational numbers cannot secure that it is *always* that one and only one part has an ending point, since when the series of rational numbers is cut by an irrational number, then neither of the two parts can have an ending point. On the other hand, for the sequence of integers, if it is cut by some point, both parts have an ending point. Consequently, neither rationals nor integers are continuous.

As we have seen, Dedekind defines irrational numbers as the point or limit lying between two sequences of rational numbers, both of which contain infinitely many terms. Thus, if we accept his definition[9], then we should also accept the concept of an 'infinite set'. One may argue that though we should accept an infinite 'set', we can apply Aristotle's concept of potential infinity or Leibniz's syncategorematic understanding of infinity, therefore avoiding commitment to Cantor's transfinite sets (i.e. collective, actual infinities). However, this argument cannot go through. First, for Aristotelian potential infinity, I have argued above that this concept is simply an extrapolation of finite natural numbers. Since an extrapolation of finite natural numbers is still finite, it cannot provide the basis for the collection of infinitely many rational numbers which is required in the definition of irrational numbers. On the other hand, Leibniz' syncategorematic

[9] One may wonder why we should accept Dedekind's definition. If there is a definition for irrational numbers which does not rely on the set of rational numbers, then Cantor's irrational number argument fails. However, it seems that all the alternative definitions (for example, Cantor's own definition, as well as Russell's segment definition) involve the set of rational numbers.

understanding of infinity cannot satisfy the requirement either, as we shall see below.

Take the '√2' example. If we treat the infinity involved in Leibniz' Syncategorematic sense, then we should understand Set A (defined on the previous page) in the distributive way:

> For any rational number p which belongs to A, there is another rational number q which also belongs to A such that $p<q<\sqrt{2}$.

But this kind of understanding cannot define √2, for it cannot distinguish √2 from another *variable* rational number which approaches √2 indefinitely. For the density of rational numbers, there is always a rational number r such that $p<q<r<\sqrt{2}$. Therefore, the syncategorematic understanding of infinity cannot serve the purpose to define √2, because it cannot make the distinction between √2 and r. In other words, since √2 follows *all* the rational numbers in Set A, it is only the whole set of all the rational numbers less than √2 that can define √2. This requires the collective understanding of infinity. The syncategorematic understanding, which treats the infinite in the distributive sense, cannot meet the requirement. This is what Cantor means when he writes that 'the transfinite numbers *stand or fall* with the finite irrational numbers', because both of them require the collective/categorematic understanding of the infinite. Therefore, the infinite set required in Dedekind's definition of irrational numbers cannot be understood either in Aristotle's potential sense, or in Leibniz' syncategorematic way, but must be treated in the Cantorian sense, i.e. as a completed infinite whole.

Since the concept of irrational numbers (and consequently real numbers) had already been accepted by the mathematical community in Cantor's time, and since the definition of such numbers requires the acceptance of a completed infinite domain, Cantor's irrational number argument is a very strong argument for the existence of actual infinity. However, this argument is independent of his domain principle. It is not about the presupposition relation between variability and a domain. Rather, it is about the usefulness of infinite domains in mathematical research. Thus, it is an argument from a pragmatic perspective. Even if we are convinced by this argument, we may still feel suspicious about Cantor's domain principle. What is worse, our doubt is not without reason. Consider set theory. When we say 'for any set *S*', it seems that we have already presupposed the domain of all sets. According to the domain principle, such a domain should also be a set. Cantor says that a 'domain cannot itself be something variable, since otherwise each fixed support for the study would collapse.' However, if we treat such a domain also as a set, then we get paradoxes.

7.1.4. The Absolute Infinite

One may agree that a variable quantity presupposes a corresponding domain (in the intuitive sense), but we still want to ask why such a domain should be deemed to be a set (in Cantor's technical sense). If every domain is taken as a set, then this principle would bring disaster to set theory, for those extremely large and problematic infinities (which Cantor calls 'the absolute infinite') would become inevitable. Take ordinal numbers as an example. Just as a variable quantity over the domain of natural numbers presupposes the set of all natural numbers, it seems quite natural to conclude from this principle that a variable quantity over the domain of ordinal numbers also presupposes the set of all ordinal numbers. In other words, the 'set' of all ordinal numbers, as the domain of ordinals, should be presupposed beforehand so that the discussion of ordinal numbers can be meaningful. However, it is notorious that if we admit such a thing as a 'set' of all ordinal numbers, then we will get a paradox (known as Burali-Forti's paradox). A much simpler example is about the concept 'set' itself. According to the domain principle, when we say 'given any set *S*', we have already presupposed a domain of all sets, and such a domain should also be treated as a set. However, such a 'set' cannot be accepted, since it results in the paradox of the largest cardinal number.[10]

Therefore, if Cantor wants to consistently rely on the domain principle to argue for the existence of transfinite sets and numbers, then he should also accept the absolute infinite. If he wants to exclude the absolute infinite from the realm of 'sets', then he has to provide a clear criterion for excluding such things as being sets. The criterion he provided is 'mathematical determination'. The transfinite can be rationally subjugated. They are fixed, definite and can be treated as mathematical objects. The absolute infinities, on the other hand, always lead to

[10] That is, if we allow the totality of all sets as a set, then it would have the largest cardinal number. However, by Cantor's theorem, the power set of this 'set' would have a strictly larger cardinal number. So we would have a contradiction.

contradiction. They are not fixed or definite, nor are they within the realm of mathematical study. Cantor claims that the absolute infinite can only be thought by God, so they are not the proper candidates in his set theory. In a letter to Dedekind, Cantor wrote:

> If we start from the notion of a definite multiplicity (a system, a totality) of things, it became clear to me that we must necessarily distinguish between two kinds of multiplicity (by this I always mean *definite* multiplicities).
>
> For on the one hand a multiplicity can be such that the assumption that *all* of its elements 'are together' leads to a contradiction, so that it is impossible to conceive of the multiplicity as a unity, as 'one finished thing'. Such multiplicities I call *absolutely infinite* or *inconsistent multiplicities*.
>
> As one easily sees, the 'totality of everything thinkable', for example, is such a multiplicity; later still other examples will present themselves.
>
> When on the other hand the totality of elements of a multiplicity can be thought without contradiction as 'being together', so that their collection into '*one* thing' is possible I call it a *consistent multiplicity* or a *set*.[11]

Even if we accept this explanation, this again provides little justification for the domain principle. The domain principle itself cannot distinguish a legitimate set from the absolute totalities. The differentiation criterion, i.e. whether they make mathematics inconsistent, is still justified from a pragmatic perspective. If we treat the domain principle as saying that any presupposed domain should be treated as a set, then it amounts to the unrestricted comprehension principle:

> For any defining criterion φ, there is a corresponding set defined by φ.

[11] Letter to Dedekind, dated 28 July 1899, in Cantor [1932], p. 443 (English translations are from Hallett 1984: 166).

For example, when we talk about natural numbers, we say ‘for any natural number *n*, …’. Cantor would argue that this usage of variability presupposes the domain of natural numbers, which is a transfinite set. However, we also can say that here ‘*n*’ stands for any object that satisfies the property ‘is a natural number’, and thus amounts to the ‘defining criterion *φ*’ in the unrestricted comprehension principle. Accordingly, to say that such a variable quantity presupposes a *set* as its domain equates to saying that this defining criterion *φ* presupposes a corresponding *set* defined by *φ*.

The unrestricted comprehension principle has been blamed by later scholars as the root of all the trouble within naïve set theory, as discussed below. On the other hand, if we do not treat every domain as a set, then the domain principle cannot justify the domain of all natural numbers as a set either. All the arguments for differentiating the transfinite from the absolute infinite are from the pragmatic perspective, i.e. what is useful for and can be subject to mathematical reasoning. Taking the domain principle alone, which can be regarded as a generalized version of the unrestricted comprehension principle, we may say that the transfinite sets *stand or fall* with the absolute ones.

7.2. The Limitation of Size Theory

There is another argument in the discussion of the foundations of mathematics, which aims to provide an adequate criterion to differentiate the transfinite from the absolute infinite. This argument is usually called ‘the limitation of size theory', according to which the problem with problematic absolute multiplicities (e.g. the totality of all sets, the totality of all ordinal numbers, etc.) is that their sizes are too big for them to be mathematically determinable, i.e. they are an 'absolute maximum'. Thus, to avoid paradox, the unrestricted comprehension principle is repaired by saying that, for anything to be a set, its size must not be 'too big'. This became a guiding principle in ZF axiomatic set theory:

> Our guiding principle, for the system ZF, will be to admit only those instances of the axiom schema of comprehension which assert the existence of sets which are not too 'big' compared to sets which we already have. We shall call this principle the limitation of size doctrine. (Fraenkel *et al*. 1973: 32)

Therefore, it is required in the ZF system that a set, if it is not introduced by an axiom, must be able to be shown to be a subset of some existent set. However, this principle has the same problem as Cantor's argument above. That is, how big is 'too big'? This is why Russell (1907) dismissed the limitation-of-size conception almost as soon as he raised it:

> A great difficulty of this theory is that it does not tell us how far up the series of ordinals it is legitimate to go. It might happen that ω was already illegitimate: in that case all proper classes would be finite. … Or it might happen that ω^2 was illegitimate, or ω^ω or ω_1 or any other ordinal having no immediate predecessor. We need further axioms before we can tell where the series begins to be illegitimate. … But our general principle does not tell us under what circumstances such a function is predicative. (Russell 1907: 44)

According to Hallett (1984), Fraenkel tried to answer the question ‘how big is “too big”’ by referring to the idea of diagonalization.[12] For any set in the ZF system, since there is a larger set, i.e. its power set, the set in question is not too big or too comprehensive. If a multiplicity cannot be diagonalized out, then it is too comprehensive to be a set. However, there are several problems with this account. First, it seems that Burali-Forti’s paradox does not involve diagonalization, but simply relies on the concept of ‘order type’. If Fraenkel insists that the differentiating criterion is diagonalization, then this term must be used in a loose and metaphorical sense, for, strictly speaking, there is no diagonalization involved in the totality of all ordinal numbers. Second, a more serious problem is that if we simply rely on diagonalization to exclude multiplicities that are too comprehensive, then how to justify the Axiom of power set?

Axiom of Power Set: $\forall x \exists y \forall z[z \in y \equiv \forall w(w \in z \rightarrow w \in x)]$

This axiom implies that for any set, there is a set with a strictly larger cardinality. In particular, when Cantor initially introduced the diagonal method as a *reductio* proof of the non-enumerability of real numbers (which was discussed in Chapter 2), why should we treat it as a *reductio* proof of the non-enumerability of the set of real numbers, rather than as showing that the totality of real numbers cannot be diagonalized out, so that it is a totality which is too comprehensive to be a set? Consider his diagonal argument (adapted from Cantor 1891):

[12] Cf. Hallett 1984: 200-5.

i. Suppose that the totality of all real numbers in a given interval is an enumerably infinite set.
ii. By this supposition, we can have a complete enumeration of all the real numbers in that interval.
iii. Based on this list, we construct a new item which cannot be identical with any element on the given list.
iv. Therefore, this list is not a complete enumeration of all the real numbers in that interval, which contradicts (ii).
v. Therefore, the set of all real numbers in an interval is not enumerable.

For this argument, one may wonder why the *reductio* should be on the term ‘enumerable’ rather than on ‘set’. Why do we still take it for granted that the totality of all the real numbers in an interval is a set, if there is some contradiction being derived from this supposition? If an absolute maximum is something that is too big, then isn’t it the case with the totality of all real numbers? After all, Cantor’s diagonal argument has shown exactly that this totality cannot be diagonalized out.

This question reminds us of the old debate around paradoxes. To deduce some contradiction is not a dangerous thing, since it is a common practice in *reductio* arguments, which are widely employed in mathematical proofs. But it is crucial to know when the contradiction indicates a *reductio* argument, and when it is really a paradox. In history, there were too many so-called ‘paradoxes’ attributed to the concept ‘infinity’. Many of them were dismissed when Cantor’s transfinite theory became available. In light of this new conception of infinity, it is shown that many *prima facie* paradoxes simply indicate some flaw or deficiency in our understanding of the relevant notions. Once the misunderstanding is clarified, or when some new theory about the concept is developed, then the paradoxical result is in turn dismissed. Therefore, if one views a paradox in this way, then one believes that the current paradoxical result is only due to our temporary ignorance and that, when new knowledge or a new perspective becomes available, the paradox will be solved. The contradiction only indicates further areas for conceptual inquiry and development; thus there is nothing that we should feel afraid of in facing a paradox.

However, there is another view, which is called the ‘bankruptcy’ view, according to which a contradiction in a paradox is really a contradiction. It does not indicate a further area for intellectual inquiry. Instead, it manifests the fundamental confusion in our understandings and the limits of our thought. In history, the ‘bankruptcy’ view usually leads to theological ideas or mysticism. When paradoxes were discovered in naïve set theory, Cantor’s reaction to the absolute infinite is a kind of bankruptcy view. Since these absolute maximums cause bankruptcy for set theory, they must be blocked. They are monstrous, which are things beyond rational (mathematical) thinking.

For the question why the contradiction in Cantor’s diagonal argument for real numbers does not indicate bankruptcy, set theorists would answer that there is still something that we can ‘lean back on’, to borrow a term from Gödel[13]. In this case, we still can lean back on the concept ‘enumerable’, so that we can conclude

[13] See Wang’s report on Gödel’s philosophical ideas about set theory and logic: “The bankruptcy view only applies to general concepts such as proof and concept. But it does not apply to certain approximations where we do have something to lean back on. In particular, the concept of set is an absolute concept [that is not bankrupt], and provable in set theory by axioms of infinity is a limited concept of proof [which is not bankrupt].” (Wang 1996: 270)

that this set is non-enumerable. Similar things can be said about Cantor's argument that all natural numbers form a set. In history, there are various objections to treating the totality of all natural numbers as a whole (i.e. a set in Cantor's sense). For example, if we treat all natural numbers as a set, then the number of all natural numbers would be equal to the number of all even numbers, because there is a one-one correspondence between them. But that is intuitively unacceptable, for the number of natural numbers should be twice as many as that of even numbers. However, what Cantor did was exactly to show that what was ridiculous was the idea that a proper subset of an infinite set had to be smaller than the original. Thus, we have a new conception of cardinality. Although initially we face a contradiction, we still have something to 'lean back on', i.e. one-one correspondence. Therefore, the pseudo-paradox with natural numbers is solved.

From these discussions, we can see that the diagonalization proposed by Fraenkel cannot differentiate the transfinite from the absolute infinite; otherwise we should treat real numbers as too big to be a set as well. However, nowadays mathematicians certainly do not want this result. They simply do not want to be expelled from the paradise created by Cantor. That is to say, there should be some other reason according to which we can draw the line between the transfinite and the absolute. In the discussion above, I use the term 'lean back on' to suggest this reason. But this is simply a metaphor, and we need a more precise characterization of this idea.

7.3. Indefinite Extensibility

7.3.1. All Things are Indefinitely Extensible?

According to Cantor, a totality is a set if it can be consistently thought of as a 'whole'. This idea may sound too liberal to Dummett, who thinks that even ω is too big to be a set. In Dummett's view, not only the absolute infinite, but also the concept of all real numbers, and even the concept of all natural numbers are all indefinitely extensible:

> An indefinitely extensible concept is one such that, if we can form a definite conception of a totality all of whose members fall under that concept, we can, by reference to that totality, characterize a larger totality all of whose members fall under it. (Dummett 1993: 441)

In terms of 'a definite conception of a totality all of whose members fall under that concept', Dummett means a conception which corresponds to a definite set of the initial concept. Thus, taking ordinal numbers as an example, 'the definite conception of a totality' refers to all the existent ordinals. But since they all can form an ordered sequence, there is an order type associated with them. Thus, by reference to this order type, we can characterize a new ordinal number and a larger totality of ordinal numbers.

According to Dummett's definition, not only are such notions as 'all ordinal numbers', 'all cardinal numbers', 'all sets' etc. indefinitely extensible, but also such notions as 'all real numbers', 'all natural numbers'. In the section above, I discussed Cantor's diagonal argument for the non-enumerability of real numbers.

This argument can show that the conception of real numbers is also indefinitely extensible in Dummett's sense. 'The definite conception of a totality' refers to all the real numbers on the list. By referring to this totality, we can characterize a larger totality of real numbers. Also, for natural numbers, the notion of 'all natural numbers' is indefinitely extensible, by referring to any finite set of natural numbers.

On the other hand, semantic notions such as 'true', 'heterological', and 'denote' all can be shown to be 'indefinitely extensible'. First, I constructed a definition for 'denominate' in Chapter 6, which is a counterpart of 'denote'. The notion 'denote' can be characterized by reference to the totality defined by 'denominate'. Second, for the heterological predicate *Het,* we saw in Chapter 5 that there is a hierarchy of heterological predicates Het_1, Het_2, Het_3, … associated with it. Every Het_i is a definite set of 1-place predicates. *Het* is indefinitely extensible compared with any such Het_i. Finally, the same construction can also be said about 'true'. Using the hierarchy of T_i in Tarski's truth definition, the semantic conception of *T* can be understood as an indefinitely extensible concept by referring to a definite T_i in the hierarchy. Thus, we can summarize these indefinitely extensible concepts in Table 1 (see next page).

Dummett did not mention semantic paradoxes. He only concludes that there is no fundamental difference between 'small' totalities such as all real numbers and 'absolute' totalities such as all ordinal numbers. But, as shown above, his definition can be easily extended to semantic notions. Then, one may ask, doesn't it suggest something like Priest's 'Uniform Solution Principle': same kinds of paradoxes, same solution? Shouldn't we have the same solution for both set-theoretic paradoxes and semantic paradoxes? Moreover, what is the basis for accepting real numbers as a set while rejecting the absolute infinite, if all these notions are indefinitely extensible?

Table 1

	Definite conception of a totality	**Indefinitely extensible concepts**
Set-theoretic	A denumerable set of real numbers	The totality of all real numbers
	A finite set of natural numbers	The totality of all natural numbers
	A transfinite[14] set of ordinal numbers	The totality of all ordinal numbers
	A transfinite set of cardinal numbers	The totality of all cardinal numbers
	A transfinite set of sets	The totality of all sets
Semantic	All the numbers that can be denominated	All the numbers that can be denoted
	All Het_{i} predicates	All heterological predicates
	All T_{i} sentences	All true sentences

7.3.2. 'Infinite Sets' as Single Objects

Let us begin with the second question. The difference between the absolute infinite and Cantorian transfinite sets is that, in the case of the former, we have no further concept that we can lean back on. Comparing the set of real

[14] Here, the word 'transfinite' is to emphasize the contrast between transfinite sets and the absolute infinities.

numbers and the absolute infinite, we may say that the contradiction in real numbers results from the concept ‘enumerable set’, so we can reject the enumerability of real numbers. But for the absolute infinite (e.g. the ‘set’ of all sets), there is only one concept involved, i.e. ‘set’, so that the contradiction could only result from the concept ‘set’. To fix this problem, it seems that the only way out of paradox is to reject the naïve conception of ‘set’, i.e. to deny that the totality of the absolute infinite is a set.

To describe the difference between the transfinite and the absolute, Cantor also used many different terms. For example, he says that the transfinite is ‘increasable’; ‘subject to numerical determination’; ‘determined in all its parts’; ‘definite and fixed’; ‘mathematizable’, ‘mathematically determinable’, ‘rationally subjugable’; etc., while the absolute infinite has the opposite properties. Though words like ‘increasable’ may suggest something relevant to the size of the totality, the underlying idea behind these terms is that the totality involved in the transfinite could be treated as a definite object, as Cantor wrote:

> When … the totality of elements of a multiplicity can be thought without contradiction as ‘being together’, so that their collection into ‘*one* thing’ is possible, I call it a consistent *multiplicity* or a *set*. (Cantor 1932: 443, English translation from Hallett 1984: 34)

In terms of ‘one thing’, it means that the totality in question is limited by some other larger totality, so that it is restricted and limited. This is what is meant by ‘increasable’, since there is something larger. As mentioned above, we say that such sets still have something to ‘lean back on’. On the other hand, since it is restricted, the set can be treated as an object. This is what Cantor means when he uses words like ‘mathematizable’, ‘rationally subjugable’, etc. The essential feature of an object is that it can be differentiated from its background. Thus, to be an object, there must be some boundary around it. For transfinite sets, such a boundary is settled by a larger set. Therefore, from the perspective that they can be treated as single objects, transfinite sets have no fundamental difference from finite sets. This is called ‘Cantorian finitism’ by Hallett (1984):

> First, it expresses a certain ‘finistic’ attitude to sets (mathematical objects) and which is what gives the theory its unity. Namely, sets are treated as simple objects, regardless of whether they are finite or infinite. Secondly, all sets have the same basic properties as finite sets. (Hallett 1984: 32)

In section 7.1.3, I discussed Dedekind’s definition of real numbers. That definition forces us to admit that the relevant infinite collections of rational numbers are single objects. Although the collections of rational numbers are highly complex – they contain infinitely many items - we simply treat them as individual objects. Therefore, although the set of real numbers and the set of natural numbers are both indefinitely extensible in Dummett’s sense, they still can be treated as single objects, because they are still restricted by some other larger ones. This feature differentiates them from the absolute, since we cannot draw any boundary for the latter.

7.3.3. The Extensionalizable Understanding of ‘Set’

Dummett’s definition of ‘indefinitely extensible concepts’ relies on the understanding of the complementary concept ‘definite’, which he does not specify.

In Cantor's characterization of sets, we also find phrases such as 'subjects to numerical determination'; 'determined in all its parts'; 'definite and fixed'. These phrases suggest another essential feature of sets: they can be treated extensionally. Recall Zermelo's initial idea for designing the axiom schema of separation. He wrote:

> the defining criterion must always be definite in the sense of our definition in No. 4 (that is, for each single element *x* of *M* the fundamental relations of the domain must determine whether it holds or not), with the result that, from our point of view, all criteria such as "definable by means of a finite number of words", hence the "Richard antinomy" and the "paradox of finite denotation", vanish. (Zermelo 1908: 202)

Although Zermelo made many efforts to characterize the notion of a 'definite' assertion, this notion was still vague in his treatment. This problem was addressed by several set-theorists, the most prominent among whom are Skolem, Fraenkel, and von Neumann. In Chapter 6, I briefly mentioned Skolem's treatment, i.e. that a notion is definite iff it can be expressed by a *wff* in a first order language containing only '$\in$' and identity. Skolem's method is now universally accepted by mathematicians and logicians because of its simplicity and generality.[15] Nevertheless, even though this idea has been shown mathematically productive, there is still a philosophical issue to be addressed. It seems too stringent to require that a definite notion must be expressible in systems such as those described by Skolem. Many ordinary terms are not paradoxical at all, but it is unclear how they

[15] Zermelo himself, however, rejected this treatment because in his view, it implicitly involves the notion of finite cardinal (natural number) which should be based on set theory. (Fraenkel *et al*. 1973: 38)

can be expressed in such systems. For example, in his reply to my discussion of this issue, Priest challenges this requirement of definiteness made by Skolem, especially the requirement of definiteness for semantic notions:

> In his axiomatization of set theory, Zermelo required the conditions used to define sets to be 'determinate'. He did not specify what this meant. Later, Skolem characterized it as being a condition expressible in a first-order language with only the predicates '=' and '∈'. The notion of definability cannot be expressed in this language. So, says Zhong, it is not definite. This reasoning can hardly be correct. Being expressible in the way Skolem suggested may be adequate for pure set-theory. But there are many perfectly good determinate (whatever that means) predicates which cannot be so expressed: 'is an electron', 'weighs exactly n grams', 'is midday January 1st, 2012, GMT'. (Note, in particular, that these are not vague predicates.) Why isn't definability like that?[16]

Admittedly, it is hard to express a determinate predicate like 'is an electron', 'weighs exactly n grams', 'is midday January 1st, 2012, GMT', etc. in a first-order language which only contains '=' and '∈'. But this is not the essential point behind the idea about 'definite'. By 'definite', Zermelo means that for a given object in the domain and a defining criterion φ, we can 'determine without arbitrariness whether it holds or not.' (Zermelo 1908: 201) In other words, there is an extensional understanding of the criterion or concept φ, by which we can determine without ambiguity whether it holds for a given object or not. If such an extensional understanding is available, and is regarded as capturing the essence of the concept φ, then we may say that the concept φ is *extensionalizable*. It is easy to see that concepts like 'is an electron', 'weighs exactly n grams' etc., can be

[16] Priest (2012): "Definition Inclosed: a Reply to Zhong", *Australasian Journal of Philosophy,* Vol. 90, No. 4, page 792.

extensionalized. However, ‘set’ in the naïve sense is an intensional understanding of ‘set’, since according to the unrestricted comprehension principle, in some circumstance such as Russell’s paradox, we cannot determine whether the relevant criterion φ holds for a given totality or not.

Can ZF axiomatization of sets guarantee that the concept ‘set’ is extensionalizable? It is known that ZF axiomatic set theory is based on the iterative concept of set. That is, a larger set is built on smaller sets by the operation ‘set of’. There are several axioms telling us how to obtain larger sets from small sets, for example, the axiom of union, the axiom of pairing, etc. However, there are two axioms which look more suspicious, since taken together they can produce very ‘large’ sets. One is the axiom of power set; the other is the axiom of Infinity[17].

Axiom of Power Set: $\forall x \exists y \forall z[z \in y \equiv \forall w(w \in z \rightarrow w \in x)]$

I have discussed Cantor’s diagonal argument above. The conclusion of this argument is that, though the power set of an infinite set is much larger, there is still a way to hold all the elements together. That is, the power set of a given set is still restricted by larger sets and thus can be treated as a single object. Therefore, it will not lead to the problem to which the absolute infinite leads. In other words, though this axiom admits sets with much larger cardinal numbers, these large sets

[17] If we talk about ZF set theory and not just Zermelo set theory, the axiom of replacement makes greater ontological commitments even than the axiom of power set and the axiom of infinity.

are still based on smaller sets and are always restricted by even larger ones. As Priest says,

> I observe that in each case, the limit is defined 'from below'; but the contradiction is produced by considering it 'from above': that is, in each case we take the limit to be itself a unity and note its properties. (Priest 2002: 120)

We may say that the axiom of power set allows sets built 'from below' based on smaller sets, and thus it is safe. We can treat such sets as extensional if the basis set is also extensional. On the other hand, if a set is defined 'from above', then it is suspicious whether it is still extensional. This leads us to consider another axiom:

Axiom of Infinity: $\exists x[\emptyset \in x \,\&\, \forall y(y \in x \rightarrow \cup\{y,\{y\}\} \in x)]$

This axiom asserts that there is at least one infinite set. Beginning with the empty set $\emptyset$, we can construct an infinite set which is equivalent to the set of all natural numbers, which is of the following form:

$$\{\emptyset,\ \ \{\emptyset\},\ \ \{\emptyset, \{\emptyset\}\},\ \ \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\},\ \ \ldots\}$$

Obviously, this axiom introduces a set which is defined 'from above'. It simply *asserts* that all the natural numbers form a set. How can we guarantee the extensionality of such a set? As argued in the first part of this chapter, this problem cannot be solved by referring to the domain principle, but only can be justified by pragmatic principles: such sets are useful constructions in mathematics, as Fraenkel *et al.* says:

> For those mathematicians who believe in the essential soundness of classical mathematics, the task posed by the antinomies is that of constructing a system in which all the notions of classical mathematics can be defined and all (or essentially all) the theorems of mathematics up to and including analysis can be derived but such that its consistency can be proved or, short of this, such that the argumentations leading to the known kinds of antinomies are effectively excluded. (Fraenkel *et al*. 1973: 12)

There is a consistent way to treat the notion ‘natural numbers’ extensionally, because there are larger totalities (e.g. rational numbers, real numbers) to restrict the totality of natural numbers, so that it can be thought as a single object. These larger totalities are accepted and required in mathematical construction. As Cantorian finitism suggests, since the totality of natural numbers can be thought as a single object, it shares some important properties with finite sets: for instance, it is definite, and restricted by some boundary. This is the justification for the extensional understanding of the axiom of infinity.

Therefore, the axioms in the ZF system can guarantee an extensionalizable understanding of the concept ‘set’. In contrast to the naïve concept of set, the notion ‘set’ defined by the ZF axioms is thus not indefinitely extensible. On the contrary, ZF sets are definite and extensionalizable, since they are constructed by the ZF axioms. This discussion brings us to the final question concerning the two kinds of paradoxes: semantic paradoxes and set-theoretic paradoxes. If both of them have the problem of indefinitely extensible concepts, and for ‘sets’ we replace the na ïve concept with a definite one (e.g. by axiomatization), then shouldn’t we do the same thing for semantic notions? What is the justification for advocating a functional-deflationary conception for semantics in previous chapters, which leaves semantic notions totally intensional and indefinitely extensible?

7.3.4. Same Cause, Different Solutions

The answer to this question is that mathematicians and philosophers have different aims for their research on these two kinds of paradoxes. For the mathematicians, their aim is to block set-theoretic paradoxes efficiently, while the system is still strong enough to serve as a foundation of mathematics. That is why they are content with axiomatic set theory, which blocks paradox by a definite and extensional understanding of ‘set’. On the other hand, semantic notions like ‘denote’, ‘heterological’, and ‘true’ are essential for semantic paradoxes, but these are not terms about which mathematicians feel obliged to think very hard. Furthermore, what mathematicians do is to clarify a certain notion whose meaning is not clear in natural language, and make it a fruitful mathematical model for analysis. They do not have to capture every intuition associated with a notion understood in natural language. That is to say, objects and constructions in mathematical discussion are always some idealization of our thinking. Therefore, we should not accuse mathematicians of being unable to capture some intuitions associated with the naïve concept ‘set’ or ‘class’.

On the other hand, philosophers are not just concerned with an ideal model. They cannot simply solve semantic paradoxes by ignoring the intuitions associated with semantic notions understood in natural language. Philosophers cannot accept an externalization solution as a successful one for semantic paradoxes, because it simply circumvents or ignores the problem with natural language. A good theory about semantic notions should explain why we have paradox, rather than just show how to construct a theory so that the contradiction cannot emerge. Since semantic notions are essentially intensional and indefinitely extensible, because of their functional-deflationary role in our use of language, we cannot 'make' them extensional and definite.

In an explanation of his regimentation treatment of the concept 'true', Tarski says that 'if in consequence semantic concepts lose philosophical interest, they will only share the fate of many other concepts of science, and this need give rise to no regret.' (Tarski 1944: 364) Nowadays, it is generally acknowledged that the Tarskian solution for the Liar Paradox is unworkable. That is because, besides the problem of the strengthened Liar, the loss of philosophical interest means ignoring the real problem with natural language, and thus it is indeed a serious problem for a proposal for the semantic paradoxes. As argued in the previous two chapters, the definite and extensional understanding of semantic notions amounts to understanding them as representational, while semantic notions are essentially non-representational.

By the axiomatization of 'set', mathematicians have made this concept scientific. In this case, 'this need give rise to no regret', even though it loses philosophical interest. Whether a theory is ‘philosophically interesting’ is not a requirement in the mathematical realm. Mathematicians concern themselves with mathematical fruitfulness, rather than philosophical interest. Mathematicians need a scientific theory with useful, consistent concepts. If a theory can provide a consistent and new tool for their analysis, then they will accept it as a good theory. However, for philosophers, when they deal with semantic paradoxes, they want a theory which can explain the intuitions associated with natural language, a theory which can promote our understanding of the mechanisms of natural language. That is why, though both of the two groups of paradoxes have the same cause – indefinitely extensible concepts, they end up with different solutions.

7.4. Concluding Remarks about the Set-theoretic Paradoxes

In this chapter, I have discussed the philosophical issues concerning the set-theoretic paradoxes. I argue that the domain principle, which Cantor employed to justify the existence of transfinite sets, fails to achieve this goal. This principle cannot distinguish the transfinite from the absolute infinite. The same problem can be found in the limitation of size theory as well. Instead, the real reason to justify the transfinite is its usefulness in mathematical constructions. Then, I also examined Dummett’s argument about indefinitely extensible concepts, and argued against his thesis that there is no distinction between the set of real numbers and the absolute infinite. Following Cantor’s finitism about the transfinite, I argued that, since transfinite sets are restricted by some larger totality, they can be treated as single objects. Furthermore, the axiomatization of the theory of sets removes the indefiniteness in this concept ‘set’, and makes it totally definite and extensional. This treatment is required by mathematical research, which finally justifies why there should be different solutions to set-theoretic paradoxes and semantic paradoxes.

Chapter 8: Conclusion

In this thesis, I provided a philosophical treatment for two groups of logical paradoxes: semantic paradoxes and set-theoretic paradoxes. My treatment of the semantic paradoxes is based on the analysis of a kind of argument: the diagonal arguments, which I discussed intensively in Chapter 2. A diagonal argument contains the following components: a side, a top, an array, a diagonal, a value, and a countervalue. I argued that the diagonal on a diagonal array is a function which governs every element in the totality (i.e. it passes through every row of the array, no matter existent or potential). As a function, it is dynamic and should not be confused with the value of the diagonal, which could be represented as a row on the array. Also, the diagonal is important in achieving the feature of self-reference in diagonal arguments. These features play an essential role in my treatment of the semantic paradoxes.

Chapter 3 concerns some preliminary work for the discussion of the most important semantic paradox, the liar paradox. Firstly, I made a distinction between sentences and propositions. I argued that propositions, as the content of sentences, should not be confused with sentences, or the meaning of sentences. A meaningful sentence can fail to express a proposition, and this observation is one of the bases for my treatment of the semantic paradoxes. After that, I went through four most important contemporary theories for the liar paradox and their problems: the Tarskian hierarchy approach, the truth gap approach, contextualism, and the paraconsistent approach. All of these theories are flawed because they fail to meet one or more of the following criteria for an adequate solution to the liar paradox in natural language. First, a proposed solution should accord as much as possible with natural 'pre-theoretic' semantic intuitions. Second, an adequate analysis of a paradox must diagnose the source of the problem in the paradox, and thereby help us refine the concepts involved, making them truly coherent. To design some apparatus which simply circumvents the problem is not a good solution according to this standard. Finally, an adequate account of the liar must provide a proper treatment of the problem called 'the revenge of the liar'.

Although the truth gap approach has flaws too, it can be fixed by providing a philosophical interpretation for the nature of truth value gaps. In Chapter 4, I examined the merits and flaws in Kripke's truth gap theory. The most serious problem for Kripke's theory is that it suffered from the problem called 'the revenge of the liar'. Moreover, the language he constructed cannot contain the predicate 'either false or undefined', therefore this language is not semantically closed.

In the second part of Chapter 4, I examined Soames' theory on the liar paradox, which is an attempt to fix the two prominent problems in Kripke's theory. I argued that his explanation still has some intrinsic flaws. Firstly, Soames' argument that a liar sentence can express a proposition is based on examples of contingent liar sentences, while he does not explain how an 'essential' liar sentence also can express a proposition. Secondly, it is not clear whether there is any explicit, artificial linguistic convention for our usage of the truth predicate. Thirdly, the definition that he provides for the truth predicate is essentially circular, therefore cannot be a proper definition.

In Chapter 5, I provided a philosophical justification for truth gaps associated with the heterological paradox and the liar paradox. This treatment is based on Kripke's truth gap theory. Through a simplified model for natural language, I analyzed the heterological paradox as a diagonal argument, and argued that the heterological predicate *Het* is a dynamic notion and thus cannot be fixed by any row of cells in the diagonal array. By recognizing the functional role of *Het*, the heterological paradox is solved without resorting to a hierarchy of heterological predicates, nor need we abandon the intuitive idea that natural language is semantically universal. For the liar paradox, I advocated a functional-deflationary conception of truth, with the result that the truth predicate *T* should not be treated as a fixed set of cells in the diagonal array either. Their functional role shows that semantic notions such as *Het* and *T* are not representational, and this explains the nature of truth gaps. That is, truth gaps associated with semantic notions are not caused by some artificial linguistic rules, but are caused by the systematic features of natural language. The heterological sentence and the liar sentence are not appropriate candidate for truth bearers, because there are no cells corresponding to them in the diagonal array. Also, there is no problem like the revenge of the liar for this interpretation, because it is impossible to apply the truth predicate to the liar sentence. At the end of this chapter, I compared my treatment with contextualism, and showed that the latter violates some important intuitions associated with natural language. Therefore, I concluded that the functional-deflationary conception of truth can deal with our semantic intuitions in a better way and thus can be an adequate treatment of the liar paradox.

In Chapter 6, I extended the functional-deflationary interpretation to another kind of semantic paradox: paradoxes of definability. To defend this interpretation, I argued against a form of physicalism, and concluded that semantic concepts cannot be reduced to physical concepts, since that would involve contradiction. I also argued against another leading approach to the semantic paradoxes: Priest's dialetheism. Through a discussion of Berry's paradox and the semantic notion 'definable', I argued that (i) the Inclosure Schema that Priest proposed is not fine-grained enough to capture the underlying cause of the semantic paradoxes, i.e. the 'indefiniteness' of semantic notions; and (ii) the traditional separation of the two groups of logical paradoxes should be retained. Based on the analysis in Chapter 5 and Chapter 6, I concluded that semantic notions are not representational. Semantic paradoxes, such as the liar, the heterological paradox, and paradoxes of definability are all caused by confusing non-representational terms with representational ones. Thus, they all can be solved by clarifying the relevant confusion: the liar sentence and the heterological sentence do not have truth values, and phrases used to generate paradoxes of definability (such as Berry's paradox) do not denote an object.

After the arguments for the proper separation between the semantic paradoxes and the set-theoretic paradoxes, in Chapter 7, I argued that the axiomatic solution is an adequate solution to the set-theoretic paradoxes. Firstly, I argued that Cantor’s domain principle fails to justify the existence of transfinite sets. This principle cannot distinguish the transfinite from the absolute infinite. The same problem can be found in the limitation of size theory as well. Instead, I argued that the real reason to justify the transfinite is its usefulness in mathematical constructions. After that, I examined Dummett’s argument about indefinitely extensible concepts, and argued against his thesis that there is no distinction between the set of real numbers and the absolute infinite. The axiomatization of sets is to remove the indefiniteness in the concept ‘set’, and make it totally definite and extensional. This treatment is required by the aim of mathematical research, since mathematicians concern themselves with mathematical fruitfulness, rather than philosophical interest. However, for philosophers, when they deal with the semantic paradoxes, they want a theory which can explain the intuitions associated with natural language, a theory which can promote our understanding of the mechanisms of natural language. That is why though both of the two groups of paradoxes have the same cause, i.e. indefinitely extensible concepts, they end up with different solutions.

References

Abad, J. V. 2008. The inclosure scheme and the solution to the paradoxes of self-reference. *Synthese* 160 (2): 183 - 202.

Armour-Garb, B., and J. C. Beall, eds. 2005. *Deflationary Truth*: Open Court Press.

Armour-Garb, B., and J.A. Woodbridge. 2006. Dialetheism, semantic pathology, and the open pair. *Australasian Journal of Philosophy* 84 (3): 395-416.

Arthur, R. T. W., ed. 2001. *The Labyrinth of the Continuum: Writings on the Continuum Problem, 1672-1686*: Yale University Press.

Arthur, R. T. W. 2001. Leibniz on infinite number, infinite wholes and the whole world: a reply to Gregory Brown. *The Leibniz Review* 11: 103-116.

Arthur, R. T. W. forthcoming. Leibniz's actual infinite in relation to his analysis of matter.

Austin, J. L., P. F. Strawson, and D. R. Cousin. 1950. Symposium: Truth. *Aristotelian Society Supplementary Volume* 24: 111 - 172.

Badici, E. 2008. The Liar Paradox and the Inclosure Schema. *Australasian Journal of Philosophy* 86 (4): 583-596.

Baldwin, T. 1991. The Identity Theory of Truth. *Mind* 100 (1): 35-52.

Barnes, J., ed. 1985. *The Complete Works of Aristotle*. 2 vols. Princeton: Princeton University Press.

Barwise, J., and J. Etchemendy. 1989. *The liar: An essay on truth and circularity*: Oxford University Press.

Beall, J. C. 2000. On truthmakers for negative truths. *Australasian Journal of Philosophy* 78 (2): 264-268.

Beall, J. C. 2001. A neglected deflationist approach to the liar (Paradox). *Analysis* 61 (2): 126-129.

Beall, J.C., ed. 2003. *Liars and Heaps: New Essays on Paradox*. Oxford: Oxford University Press.

Beall, J. C., ed. 2007. *Revenge of the liar : new essays on the paradox*. New York: Oxford University Press.

Beall, J. C. 2009. *Spandrels of Truth*: Oxford University Press.

Beall, J. C., and B. Armour-Garb, eds. 2006. *Deflationism and Paradox*: Oxford University Press.

Benacerraf, P., and H. Putnam, eds. 1983. *Philosophy of mathematics: Selected readings*: Cambridge University Press.

Blackburn, S. and K. Simmons, eds. 1999. *Truth*. Oxford: Oxford University Press.

Boghossian, P.A. 1990. The status of content. *The Philosophical Review* 99 (2): 157-184.

Boolos, G. 1989. A new proof of the Gödel incompleteness theorem. *Notices of the American Mathematical Society* 36 (4): 388–390.

Boolos, G. 1998. *Logic, Logic, and Logic*: Harvard University Press.

Boolos, G, J. Burgess, and R. Jeffrey. 2007. *Computability and Logic*. Cambridge: Cambridge University Press.

Bostock, D. 1972. Aristotle, Zeno, and the Potential Infinite. *Proceedings of the Aristotelian Society* 73: 37-51.

Burali-Forti, C. 1897. A Question on Transfinite Numbers. In *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, ed. J. van Heijenoort, 104-111: Harvard University Press.

Burge, T. 1979. Semantical paradox. *Journal of Philosophy* 76 (4): 169-198.

Burge, T. 1982. The Liar Paradox Tangles and Chains. *Philosophical Studies* 41 (3): 353-366.

Burge, T. 1992. Philosophy of language and mind: 1950-1990. *The Philosophical Review* 101 (1): 3-51.

Cantor, G. 1874. On a Property of the Set of Real algebraic Numbers. In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, ed. W. B. Ewald, 839-843: Oxford University Press.

Cantor, G. 1883. Foundations of a general theory of manifolds: a mathematico-philosophical investigation into the theory of the infinite. In *From Kant to Hilbert: a source book in the foundations of mathematics*, ed. W. Ewald, 878-920: Oxford University Press.

Cantor, G. 1886. Uber die verschiedenen Ansichten in Bezug auf die actualun-endlichen Zahlen. . *Bihang Till Koniglen Svenska Vetenskaps Akademiens Handligar* 11 (19): 1-10.

Cantor, G. 1891. On an Elementary Question in the Theory of Manifolds. In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, ed. W. B. Ewald, 920-922: Oxford University Press.

Cantor, G. 1899. Letter to Dedekind. In *From Frege to Godel: A Source Book in Mathematical Logic, 1879-1931*, ed. J. van Heijenoort, 113-117: Harvard University Press.

Cantor, G. 1932. *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Edited by E. Zermelo. Berlin: Springer.

Cantor, G. 1955. *Contributions to the Founding of the Theory of Transfinite Numbers*. Edited by P. E. B. Jourdain. New York: Dover Publications.

Chaitin, G.J. 1975. Randomness and Mathematical Proof. *Scientific American* 232 (5): 47-52.

Charlton, W., and L. Judson. 1991. Aristotle's Potential Infinites. *Aristotle's Physics: A Collection of Essays*: 129-49.

Chihara, C.S. 1979. The semantic paradoxes: A diagnostic investigation. *The philosophical review* 88 (4): 590-618.

Chihara, C.S. 1984a. Priest, the Liar, and Gödel. *Journal of Philosophical Logic* 13 (2): 117-124.

Chihara, C.S. 1984b. The semantic paradoxes: Some second thoughts. *Philosophical Studies* 45 (2): 223-229.

Dauben, J.W. 1979. *Georg Cantor: His mathematics and philosophy of the infinite*: Princeton University Press.

Davidson, D. 1967. Truth and meaning. *Synthese* 17 (1): 304-323.

Davidson, D. 1969. True to the Facts. *The Journal of Philosophy* 66 (21): 748-764.

Davidson, D. 1970. Mental events. In *Experience and theory*, eds. L. Foster and J. W. Swanson, 79-102: University of Massachusetts Press.

Dedekind, R. 1872. Continuity and irrational numbers. In Dedekind 1901, 1-13.

Dedekind, R. 1901. *Essays on the theory of numbers*. Trans. by W. W. Beman: The Open court publishing company. Available online at: http://www.gutenberg.org/files/21016/21016-pdf.pdf

Donnellan, K.S. 1966. Reference and definite descriptions. *The Philosophical Review* 75 (3): 281-304.

Dummett, M.A.E. 1959. Truth. *Proceedings of the Aristotelian Society* 59 (1): 141-62.

Dummett, M.A.E. 1963. The philosophical significance of Gödel's theorem. *Ratio* 5: 140-155.

Dummett, M.A.E. 1978. *Truth and other enigmas*: Harvard University Press.

Dummett, M.A.E. 1991. *Frege: Philosophy of Mathematics*: Harvard University Press.

Dummett, M.A.E. 1993. *The seas of language*: Oxford University Press.

Etchemendy, J. 1988. Tarski on Truth and Logical Consequence. *The Journal of Symbolic Logic* 53 (1): 51-79.

Ewald, W. B. 2007. *From Kant to Hilbert: a source book in the foundations of mathematics*: Oxford University Press.

Field, H. 1972. Tarski's Theory of Truth. *The Journal of Philosophy* 69 (13): 347-375.

Field, H. 1992. Truth by Paul Horwich. *Philosophy of Science* 59 (2): 321-330.

Field, H. 1994. Deflationist views of meaning and content. *Mind* 103 (411): 249-285.

Field, H. 2003. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic* 32 (2): 139-177.

Field, H. 2008. *Saving truth from paradox*: Oxford University Press.

Fraenkel, A.A., Y. Bar-Hillel, and A. Levy. 1973. *Foundations of set theory*. Vol. 67: North Holland.

Frege, G. 1918-19. The Thought: A Logical Inquiry. In *Collected Papers on Mathematics, Logic, and Philosophy*, 351-372. Oxford: Basil Blackwell.

Frege, G. 1984. *Collected Papers on Mathematics, Logic, and Philosophy*. Translated by P. Geach and R. H. Stoothoff. Oxford: Basil Blackwell.

Friedman, M. 1975. Physicalism and the Indeterminacy of Translation. *Noûs* 9 (4): 353-374.

Frost-Arnold, G. 2004. Was Tarski's theory of truth motivated by physicalism? *History and Philosophy of Logic* 25 (4): 265-280.

Gibson, A. 2003. *Metaphysics and transcendence*. Vol. 5: Psychology Press.

Glanzberg, M. 2001. The Liar in context. *Philosophical Studies* 103 (3): 217-251.

Glanzberg, M. 2003a. Against truth-value gaps. In *Liars and Heaps: New Essays on Paradox*, ed. J. C. Beall, 151–194.

Glanzberg, M. 2003b. Against truth-value gaps. *Liars and Heaps: New Essays on Paradox*: 159-194.

Glanzberg, M. 2003c. Minimalism and paradoxes. *Synthese* 135 (1): 13-36.

Glanzberg, M. 2004. A contextual-hierarchical approach to truth and the Liar Paradox. *Journal of Philosophical Logic* 33 (1): 27-88.

Glanzberg, M. 2005a. Minimalism, deflationism, and paradoxes. In *Deflationism and Paradox*, eds. J. C. Beall and B. Armour-Garb, 107-132: Oxford University Press.

Glanzberg, M. 2005b. Presuppositions, truth values, and expressing propositions. In *Contextualism in philosophy: knowledge, meaning, and truth*, eds. G. Preyer and G. Peter, 349–396: Clarendon Press.

Glanzberg, M. 2005c. Truth, reflection, and hierarchies. *Synthese* 142 (3): 289-315.

Glanzberg, M. 2007. Context, content, and relativism. *Philosophical Studies* 136 (1): 1-29.

Glanzberg, M. 2009. Semantics and truth relative to a world. *Synthese* 166 (2): 281-307.

G ödel, K. 1944. Russell's mathematical logic. In *The Philosophy of Bertrand Russell*, eds. S. Feferman, J. Dawson and S. Kleene, 119--141: Northwestern University Press.

Goldstein, L. 2000. A unified solution to some paradoxes.

Grattan-Guinness, I. 1998. Discussion. Structural similarity of structuralism? Comments on Priest's analysis of the paradoxes of self-reference. *Mind* 107 (428): 823-834.

Gupta, A. 1982. Truth and Paradox. *Journal of Philosophical Logic* 11 (1): 1-60.

Gupta, A., and R.L. Martin. 1984. A fixed point theorem for the weak Kleene valuation scheme. *Journal of Philosophical Logic* 13 (2): 131-135.

Haack, S. 1976. The Pragmatist Theory of Truth. *The British Journal for the Philosophy of Science* 27 (3): 231-249.

Halbach, V. 1999. Disquotationalism and Infinite Conjunctions. *Mind* 108 (429): 1-22.

Hallett, M. 1984. *Cantorian set theory and the limitation of size*. Oxford: Clarendon Press.

Herzberger, H.G. 1967. The truth-conditional consistency of natural languages. *The journal of philosophy* 64 (2): 29-35.

Herzberger, H.G. 1970. Paradoxes of grounding in semantics. *The journal of philosophy* 67 (6): 145-167.

Herzberger, H.G. 1973. Dimensions of truth. *Journal of Philosophical Logic* 2 (4): 535-556.

Herzberger, H.G. 1981. New paradoxes for old. *Proceedings of the Aristotelian Society* 81: 109-123.

Herzberger, H.G. 1982a. Naive Semantics and the Liar Paradox. *The Journal of Philosophy* 79 (9): 479-497.

Herzberger, H.G. 1982b. Notes on naive semantics. *Journal of Philosophical Logic* 11 (1): 61-102.

Hessenberg, G. 1906. *Grundbegriffe der mengenlehre*: Vandenhoeck & Ruprecht.

Hintikka, J. 1957. Necessity, universality, and time in Aristotle. *Ajatus* 20: 65-90.

Hintikka, J. 1966. Aristotelian infinity. *The Philosophical Review* 75 (2): 197-218.

Hofstadter, D.R. 1999. *Gödel Escher Bach: An Eternal Golden Braid*: Basic Books.

Holton, R. 2000. Minimalism and truth-value gaps. *Philosophical Studies* 97 (2): 135-165.

Horwich, P. 1998. *Truth*. 2nd ed. Oxford: Clarendon Press.

Horwich, P. 2001. A Defense of Minimalism. *Synthese* 126 (1/2): 149-165.

Horwich, P. 2006. The Value of Truth. *Noûs* 40 (2): 347-360.

Hussey, E. 1993. *Aristotle's physics: Books III and IV*, Clarendon Aristotle series, trans. by Edward Hussey, Oxford University Press.

Jacquette, D., ed. 2002. *A companion to philosophical logic*: Blackwell Publishing.

King, P. J. 1994. Reconciling Austinian and Russellian Accounts of the Liar Paradox. *Journal of Philosophical Logic* 23 (5): 451-494.

Kirkham, R.L. 1993. Tarski's physicalism. *Erkenntnis* 38 (3): 289-302.

Kripke, S. 1975. Outline of a theory of truth. *The Journal of Philosophy* 72 (19): 690-716.

Kripke, S. 1980. *Naming and necessity*: Harvard University Press.

Kripke, S. 2011. *Philosophical troubles: collected papers*. New York: Oxford University Press.

Landini, G. 2009. Russell's Schema, Not Priest's Inclosure. *History and Philosophy of Logic* 30 (2): 105-139.

Lavine, S. 1994. *Understanding the Infinite*. Cambridge, MA: Harvard University Press.

Lear, J. 1979. Aristotelian Infinity. *Proceedings of the Aristotelian Society* 80: 187-210.

Lear, J. 1982. Aristotle's philosophy of mathematics. *The Philosophical Review* 91 (2): 161-192.

Lear, J. 1988. *Aristotle: the desire to understand*. Cambridge University Press.

Makin, S. 2006. *Aristotle: Metaphysics Theta: Translated with an Introduction and Commentary*, Oxford: Clarendon Press.

Martin, R.L. 1967. Toward a solution to the Liar paradox. *The philosophical review* 76 (3): 279-311.

Martin, R.L. 1968. On Grelling's paradox. *The philosophical review* 77 (3): 321-331.

Martin, R.L., ed. 1970. *The paradox of the Liar*: Yale University Press.

Martin, R.L. 1976. Are natural languages universal? *Synthese* 32 (3): 271-291.

Martin, R.L. 1977. On a Puzzling Classical Validity. *The philosophical review* 86 (4): 454-473.

Martin, R.L., ed. 1984. *Recent essays on truth and the liar paradox*. Oxford: Clarendon Press.

Martin, R.L., and P.W. Woodruff. 1975. On representing ‘true-in-L’in L. *Philosophia* 5 (3): 213-217.

Martinich, A. P. 1983. A Pragmatic Solution to the Liar Paradox. *Philosophical Studies* 43 (1): 63-67.

McDowell, J. 1978. Physicalism and primitive denotation: Field on Tarski. *Erkenntnis* 13 (1): 131-152.

McGee, V. 1991. *Truth, Vagueness and Paradox*: Hackett.

McKeon, R., ed. 2009. *The Basic Works of Aristotle*: Random House Publishing Group.

Meager, R. 1956. Heterologicality and the Liar. *Analysis* 16 (6): 131-138.

Mendelson, E. 1997. *Introduction to mathematical logic*: Chapman & Hall/CRC.

Moore, A.W. 2001. *The infinite*: Psychology Press.

Moore, G. H. 1978. The origins of Zermelo's axiomatization of set theory. *Journal of Philosophical Logic* 7 (1): 307 - 329.

Moore, G. H. 1982. *Zermelo's axiom of choice: Its origins, development, and influence*. New York/Heidelberg/Berlin: Springer-Verlag.

Moore, G. H. 1995. The origins of Russell's paradox: Russell, Couturat, and the antinomy of infinite number. *Synthese Library*: 215-240.

Moore, G. H., and A. Garciadiego. 1981. Burali-Forti's paradox: a reappraisal of its origins. *Historia Mathematica* 8 (3): 319-350.

Nagel, E., and J.R. Newman. 1958. *Gödel's proof*. New York: New York University Press.

Parsons, C. 1974. The liar paradox. *Journal of Philosophical Logic* 3 (4): 381-412.

Parsons, T. 1984. Assertion, Denial, and the Liar Paradox. *Journal of Philosophical Logic* 13 (2): 137-152.

Parsons, T. 1990. True contradictions. *Canadian journal of philosophy* 20 (3): 335-353.

Priest, G. 1979. The logic of paradox. *Journal of Philosophical Logic* 8 (1): 219 - 241.

Priest, G. 1983. The logical paradoxes and the law of excluded middle. *Philosophical Quarterly* 33 (131): 160-165.

Priest, G. 1991a. Intensional paradoxes. *Notre Dame Journal of Formal Logic* 32 (2): 193-211.

Priest, G. 1991b. The limits of thought—and beyond. *Mind* 100 (399): 361.

Priest, G. 1993. Another disguise of the same fundamental problems: Barwise and Etchemendy on the liar. *Australasian Journal of Philosophy* 71 (1): 60 – 69.

Priest, G. 1994a. Derrida and Self-reference. *Australasian Journal of Philosophy* 72 (1): 103-111.

Priest, G. 1994b. The structure of the paradoxes of self-reference. *Mind* 103 (409): 25.

Priest, G. 1998. The import of inclosure: Some comments on Grattan-guinness. *Mind* 107 (428): 835-840.

Priest, G. 2000. On the principle of uniform solution: a reply to Smith. *Mind* 109 (433): 123-126.

Priest, G. 2002. *Beyond the Limits of Thought*. Oxford: Oxford University Press.

Priest, G. 2006a. *Doubt Truth to Be a Liar*: Oxford University Press.

Priest, G. 2006b. *In Contradiction: A Study of the Transconsistent*: Oxford University Press.

Priest, G. 2010a. Badici on Inclosures and the Liar Paradox. *Australasian Journal of Philosophy* 88 (2): 359-366.

Priest, G. 2010b. Inclosures, Vagueness, and Self-Reference. *Notre Dame Journal of Formal Logic* 51 (1): 69-84.

Priest, G. 2012. Definition Inclosed: A Reply to Zhong. *Australasian Journal of Philosophy*: 90 (4): 789-795.

Quine, W.V.O. 1960. *Word and object*. Vol. 4: The MIT Press.

Quine, W.V.O. 1969. *Set theory and its logic*: Harvard University Press.

Quine, W.V.O. 1976. *The ways of paradox, and other essays*: Harvard University Press.

Ramsey, F. P. 1926. The Foundations of Mathematics. *Proceedings of the London mathematical society* s2-25 (1): 338-384.

Rayo, A., and G. Uzquiano, eds. 2006. *Absolute generality*: Oxford University Press, USA.

Rescher, N. 2001. *Paradoxes: Their roots, range, and resolution*: Carus Publishing.

Restall, G. 1993. Deviant logic and the paradoxes of self reference. *Philosophical Studies* 70 (3): 279-303.

Rieger, A. 2011. Paradox, ZF, and the Axiom of Foundation. *Logic, Mathematics, Philosophy, Vintage Enthusiasms*: 171-187.

Rojszczak, A. 2002. Philosophical background and philosophical content of the semantic definition of truth. *Erkenntnis* 56 (1): 29-62.

Ross, D. 1995. *Aristotle*. 6th ed: London: Routledge.

Rucker, R. 1983. *Infinity and the Mind*: New York: Bantam Books.

Russell, B. 1902. Letter to Frege. In *From Frege to Gödel: a source book in mathematical logic, 1879-1931*, ed. J. Van Heijenoort, 124-125: Harvard University Press.

Russell, B. 1903. *Principles of Mathematics*: Routledge.

Russell, B. 1905. On denoting. *Mind* 14 (56): 479-493.

Russell, B. 1907. On Some Difficulties in the Theory of Transfinite Numbers and Order Types. *Proceedings of the London mathematical society* s2-4 (1): 29-53.

Russell, B. 1908. Mathematical logic as based on the theory of types. *American journal of mathematics* 30 (3): 222-262.

Russell, B. 1957. Mr. Strawson on referring. *Mind* 66 (263): 385.

Ryle, G. 1951. Heterologicality. *Analysis* 11 (3): 61-69.

Sainsbury, R.M. 2009. *Paradoxes*. 3rd edition ed: Cambridge University Press.

Schantz, R. 2001. Truth and reference. *Synthese* 126 (1): 261-281.

Shapiro, S., and C. Wright. 2006. All things indefinitely extensible. In *Absolute generality*, eds. A. Rayo and G. Uzquiano, 253-304: Oxford University Press

Simmons, K. 1990. The diagonal argument and the Liar. *Journal of Philosophical Logic* 19 (3): 277-303.

Simmons, K. 1993. *Universality and the liar: an essay on truth and the diagonal argument*: Cambridge University Press.

Simmons, K. 1994. Paradoxes of denotation. *Philosophical Studies* 76 (1): 71-106.

Simmons, K. 2002. Semantical and logical paradox. In *A Companion to Philosophical Logic*, ed. D. Jacquette, 115-130.

Simmons, K. 2003. Reference and paradox. In *Liars and Heaps: New Essays on Paradox*, ed. J. C. Beall, 230–252.

Simmons, K. 2005. A Berry and a Russell without self-reference. *Philosophical Studies* 126 (2): 253-261.

Simmons, K. 2007. Revenge and Context. In *Revenge of the liar: new essays on the paradox*, ed. J. C. Beall, 345-367.

Skolem, T. 1922. Some Remarks on Axiomatized Set Theory. In *From Frege to G ödel: a source book in mathematical logic, 1879-1931*, ed. J. Van Heijenoort, 290-301: Harvard University Press.

Skyrms, B. 1970. Return of the liar: three-valued logic and the concept of truth. *American Philosophical Quarterly*: 153-161.

Smith, N. J. J. 2000. The principle of uniform solution (of the paradoxes of self-reference). *Mind*: 117-122.

Smith, N. J. J. 2005. Vagueness as closeness. *Australasian Journal of Philosophy* 83 (2): 157-183.

Smith, N. J. J. 2006. Semantic regularity and the liar paradox. *Monist* 89 (1): 178-202.

Smith, N. J. J. 2008. *Vagueness and degrees of truth*: Oxford University Press.

Smullyan, R. M. 1992. *G ödel's Incompleteness Theorems*: Oxford University Press.

Smullyan, R. M. 1994. *Diagonalization and Self-Reference*: Oxford University Press.

Soames, S. 1984. What is a Theory of Truth? *The Journal of Philosophy* 81 (8): 411-429.

Soames, S. 1999. *Understanding truth*: Oxford University Press, USA.

Sorensen, R. A. 2003. *A Brief History of the Paradox: Philosophy and the Labyrinths of the Mind*: Oxford University Press.

Stalnaker, R. 1973. Presuppositions. *Journal of Philosophical Logic* 2 (4): 447-457.

Strawson, P. F. 1949. Truth. *Analysis* 9 (6): 83-97.

Strawson, P. F. 1950. On referring. *Mind* 59 (235): 320-344.

Suppes, P. 1972. *Axiomatic Set Theory*: Dover Publications.

Tarski, A. 1936. The Establishment of Scientific Semantic. In *Logic, Semantzcs, Metamathematics*, 401-408. New York: Oxford.

Tarski, A. 1939. On Undecidable Statements in Enlarged Systems of Logic and the Concept of Truth. *The Journal of Symbolic Logic* 4 (3): 105-112.

Tarski, A. 1944. The semantic conception of truth: and the foundations of semantics. *Philosophy and phenomenological research* 4 (3): 341-376.

Tarski, A. 1956. The concept of truth in formalized languages. In *Logic, Semantics, Metamathematics*, 152–278.

Tarski, A. 1969. Truth and Proof. *Scientific American* 220 (6): 63-77.

Tarski, A. 1983. *Logic, semantics, metamathematics: papers from 1923 to 1938*. Edited by J. Corcoran: Hackett publishing company.

Thomson, J. F. 1962. On some paradoxes. In *Analytical Philosophy*, ed. R. Butler, 104-19. London: Blackwell.

Tucker, D. 2010. Intensionality and Paradoxes in Ramsey's 'The Foundations of Mathematics'. *The Review of Symbolic Logic* 3 (01): 1-25.

van Fraassen, B.C. 1968. Presupposition, implication, and self-reference. *The journal of philosophy* 65 (5): 136-152.

van Heijenoort, J. 1967. *From Frege to Gödel: a source book in mathematical logic, 1879-1931*: Harvard University Press.

von Fintel, K. 2004. Would you believe it? The king of France is back! Presuppositions and truth-value intuitions. In *Descriptions and Beyond* eds. M. Reimer and A. Bezuidenhout, 315-341. Oxford: Oxford University Press.

Wang, H. 1963. *A survey of mathematical logic*: Science Press.

Wang, H. 1974. *From Mathematics to Philosophy*: Routledge & Kegan Paul.

Wang, H. 1990. *Reflections on Kurt Gödel*: MIT Press.

Wang, H. 1996. *A Logical Journey: From Gödel to Philosophy*: MIT Press.

Weber, Z. 2010. Explanation and solution in the inclosure argument. *Australasian Journal of Philosophy* 88 (2): 353-357.

Williamson, T. 1994. *Vagueness*: Burns & Oates.

Wittgenstein, L. 1953. *Philosophical Investigations.* Translated by G. E. M. Anscombe: Blackwell.

Wittgenstein, L. 2001. *Tractatus Logico-Philosophicus*. Edited by D. F. Pears and B. McGuinness: Routledge.

Wright, C. 1992. *Truth and Objectivity*: Harvard University Press.

Wright, J. 1986. Field and McDowell on reference. *Australasian Journal of Philosophy* 64 (3): 298-307.

Yablo, S. 1989. Truth, definite truth, and paradox. *Journal of Philosophy* 86 (10): 539-541.

Yablo, S. 1993. Paradox without Self--Reference. *Analysis* 53 (4): 251-252.

Zermelo, E. 1908. Investigations in the foundations of set theory I. *Math. Annal* 65: 261–281.

Zhong, H. 2012. Definability and the Structure of Logical Paradoxes. *Australasian Journal of Philosophy*: 90 (4): 779-88.