Chapter 3 Relatives of the Liar Paradox and the Truth-teller including a New One

As part of an investigation of the problem space surrounding the Liar Paradox, I revisit the relationships and classification of Liar-like paradoxes using the T-schema (or associated inferences). These include the Epimenides, Curry’s, Yablo’s, Sorensen’s Queue paradox, and a new paradox. I make and investigate the conjecture that every Liar-like sentence has a Truth-teller-like dual. In doing so, I make use of the concept of a hypodox, and list pairs of sentences that are paradoxical and hypodoxical under some circumstances. I compare this classification with the classifications in Kripke [1975], and Gupta & Belnap [1993].

I relate all variations either via list structures or collections. In relation to these structures, the paradoxes of the truth predicate are sliced into self-referential, quantificational, (proper-)circular and infinite versions. I believe my scope is substantive enough. Excluded are: sub-sentential variations such as Quine’s ‘yields a falsehood when appended to its own quotation’, which I classify in the next chapter; a modal slice, typified by the Possible Liar [Post 1970]; and intensional variants, where lying is given an intensional definition rather than the same extension as ‘is false’. Excluded also are theory-relative variants, such as strengthened and revenge Liars, which arise in response to particular theories, as well as Truth-teller paradoxes [Mortensen & Priest, 1981], which I compare to strengthened Liars; variants attributing truth to tokens for various theoretical reasons, such as restrictions on the principle of extensionality [Skyrms 1984], singularities [Simmons 1993]; and context-dependence [Glanzberg 2001]; and variations appealing to situational semantics [Barwise and Etchemendy 1987]. On the edge are: derivations using weakenings of the T-schema, as I will briefly explain in section 2; truth-operator forms of the Liar; and forms of the Liar that do not use the truth-predicate [Visser 2004 p. 153].

Section 2 focuses on unquantified, self-referential variations. It contains discussion of the completeness of the classification of unquantified self-referential variations. Section 3 compares the naive classification (as given so far in section 2) with some theory-relative classifications. Section 3 makes a comparison, firstly, with Kripke’s theory of ungroundedness, and secondly, with Gupta and Belnap’s revision
theory of truth. In Section 4, I resume my development of a classification using naive semantics, and bring in quantification. Section 5 focuses on circular versions. Section 6 is about infinite versions, like Yablo’s paradox and Sorensen’s Queue paradox. Now, the introduction continues with a discussion of some relevant concepts.

### 3.1 Introduction to Classification, Relationships, Family and Reference

What is the collective noun for paradoxes? The generally accepted term is ‘family’, for better or worse. Arthur Prior [1961] wrote *On a Family of Paradoxes* using the term, I believe, because paradoxes appear to have certain relationships to each other. Indeed, more recently, Priest [1994, p. 25] argued ‘the paradoxes of self-reference … all appear to be members of a single family’ and Roy Sorensen [1998] published a collection under the title *Yablo’s Paradox and Kindred Infinite Liars*.

There appear to be an unbounded assortment of examples within the Liar family – it is this matter with which I will be mostly concerned in this chapter. Simmons [1993, pp. 2-7] provides a list of 22 variations of the Liar, interspersed with general observations on their similarities and differences. Some are ‘simple’, some ‘strengthened’, some are examples of ‘revenge Liars’, some involve sentences referring to each other in referential loops, some combine this with quantification over the proportion of a set of sentences that are true or false, some involve ‘chains’ and others are said to be ‘empirical’ counterparts. The diversity of observations on their distinguishing features is at once fascinating and frustrating. Wanted is a thorough and systematic classification of members of *the Liar family* and their relationships.

#### 3.1.1 Introduction to the Classification Project

Gupta & Belnap sought to construct an account of truth that:

> Yields a classification of the sentences of [language] $L$ into
true/false/paradoxical/etc. – a classification that conforms to our
ordinary intuitions and uses of ‘true’.

[Gupta & Belnap 1993, p. 32]

With respect to ‘paradoxical/etc.’, one should ideally know what our ‘ordinary intuitions’ are. How can we be sure the theory covers all the ways in which Liar-like sentences can turn out to be paradoxical without some classification of the paradoxes
as they arise (before remediation by inoculating against inconsistency or quarantining their consequences)? Gupta & Belnap may have been just referring to our ordinary intuitions about truth; but the correctness of their classification ought to reflect on intuitions about differences between variations of Liar-like paradoxes. If all that were at issue were our intuitions concerning the Liar and the Truth-teller, our intuitions about these are relatively straightforward. The Liar, ‘This sentence is not true’, said with reference to itself, is true if it is not true, and vice versa. It is paradoxical in this way. The Truth-teller, ‘This sentence is true’, said with reference to itself, could be consistently true or false, but we lack a decisive reason for it to be either. But use of a seemingly infinite variety of sentences can turn out to be paradoxical depending on the empirical circumstances, as Kripke [1975] established. I present a detailed but intuitive classification against which such accounts can be compared. This relatively theory-independent classification has some commonalities and differences with theory-relative classifications in Kripke [1975], and Gupta & Belnap [1993]; so, it provides some means of confirmation.

Formal languages based on resolutions to the paradoxes introduce changes that make them inappropriate for capturing our ‘ordinary intuitions’ about paradoxes. In a domestic context such as a formal language, the paradoxical sentences may be discernable, but they do not come with an associated argument that brings inconsistency on the whole system. In a formal non-bivalent language, van Fraassen [1970, p. 16] distinguished the strengthened Liar, and the strengthened, strengthened Liar, and the strengthened, strengthened, strengthened Liar, etc. Such distinctions hardly make sense in natural language. Hybrids, rare in the wild, loom large in the classification of domesticated exhibits. Discussions of disjunctions or conjunctions of Liars and Truth-tellers fascinate like ligers or tigons, but examples of these have historically been almost non-existent in natural usage prior to modern formal languages. What is wanted is a classification of the paradoxes in natural language. My classification will be conducted in natural language extended by relevant distinctions. The extended language is a meta-language. However, a presumption that I could rely on a strict object-meta-language distinction would be specious and disingenuous.

In seeking to classify the Liar and its relatives as they occur in their native state, natural language, there is an apparent methodological issue: meaningful classification in the wild, within natural language, is suspect. As Tarski [1935/1983] demonstrated, the Liar naively proves natural language contains an inconsistency, and if rules like Explosion are valid, then it seems that natural language is riddled with inconsistency.
If my meta-language is not separate, then it is **provably** inconsistent. Nevertheless, in these circumstances, it is obviously not separate. Nevertheless, I am in the same situation as anyone who writes in natural language referring to the Liar. A theory of truth must show how we can use natural language consistently or at least coherently. It is irrational to expect me to have such a remedy in order to classify the strands of various pathological specimens. There is risk of infection. It is reasonable to expect our present project to provide principles that can be **consistently applied** to classify paradoxes. Furthermore, my methodological position is comparable with mathematicians using naive set-theory to prove paradoxes. Herzberger [1982 /1984, p.133] points out the historical disanalogy that naive set theory was a well-worked out theory prior to the discovery of paradoxes, and there is doubt that a theory of naive truth could be so worked out. Yet, we will not need much more than was already discussed in the previous chapter to evaluate paradoxes and hypodoxes.

This chapter gives a classification of the nuclear family of the Liar. I will investigate mapping the classification of Liar-like paradoxes and hypodoxes into other families such as the set-theoretic paradoxes and paradoxes of satisfaction in the next chapter.

### 3.1.2 Collections and List Structures

As well as classification, I am interested in the relationship between variations. My proposal is that they are all related **via** referential list structures or collections. An analysis in terms of collections dates back to Russell [1908], but I think the use I make of list structures contributes to relating some of the variations.

For example, I give an analysis of the relationship between the Infinite Liars (Yablo’s and Sorensen’s queue paradoxes) and the circular Liars. There was some debate between Sorensen [1998] and Priest [1997] about whether the infinite liars were really self-referential. Beall [2001] still disagrees with Sorensen, as does Goldstein [1999], who thinks infinite Liars relate (directly) to circular Liars. If that were correct, then the Sorensen – Priest debate should be settled in Priest’s favour, as self-reference is just a very tight circular reference. So then the relationship is still an open question. My analysis shows that considerations about lists of sentences referring to subsequent members relate to both the infinite and circular Liars.

In the end, I use collections and list structures as a factor in individuating variations.
3.1.3 Paradox and Hypodox Duality

My classification extends to Truth-teller-like sentences; and I investigate the following conjecture:

Dual conjecture: Every Liar-like sentence has a Truth-teller-like dual.

For example, the dual of ‘All Cretan statements are false’ is ‘Some Cretan statement is true’. Imagine, for sake of example, that all other sentences stated by Cretans are false. Then, if Epimenides the Cretan states the former sentence, it is paradoxical; on the other hand, if he states the latter existential sentence instead, it is truth-teller-like in that there is insufficient reason to determine whether it is true or false, but it could consistently be either. In investigating the Dual conjecture, I will use the concept of hypodox, which I introduced in the previous chapter. As discussed in the last chapter, the compositional evaluation of a paradox or hypodox is undetermined. When one considers its possible semantic values, given bivalence, a paradox is overdetermined and a hypodox is under-determined. I will then show with respect to the dual conjecture that:

Every Liar-like paradox in my classification has a hypodoxical dual.¹

It is a virtue of such a classification that an unbounded number of Liar-like paradoxes and hypodoxes can be reduced to a finite classification. I only achieve this for subclasses; but I believe the classification of paradoxes is progressed, and hope this classification is robust, extensible and relatively independent of particular theories of truth.

3.1.4 Elements of Reference

It will be useful to abstract away from various grammatical techniques for referring to sentences. Consider my favourite sentence once again, which is still ‘My favourite sentence is not true’, and an associated paradoxical argument:

1. My favourite sentence = ‘My favourite sentence is not true’
   Premise, by stipulation or observation

¹ I conjecture that this duality also holds at least for a subset of set-theoretic and semantic paradoxes. Russell’s set has a hypodoxical dual: the set of sets which are members of themselves. Even ‘yields a truth when appended to its own quotation’ seems to be a Truth-telling dual of Quine’s Liar. I believe the conjecture even holds, in at least some cases, for paradoxes of definition. I think it may apply to epistemic paradoxes. I do not know whether it applies to modal Liars, but I expect it does.
2 ‘My favourite sentence is not true’ is true iff my favourite sentence is not true

T-schema

3 My favourite sentence is true iff my favourite sentence is not true
   1, 2 Substitution of Identicals

4 My favourite sentence is true & not true
   3 Sentential Logic (SL)

The identity at line 1 can be achieved in a variety of ways, most straightforwardly associated with a way of referring, and in particular achieving co-reference. The following list of devices for achieving co-reference may not be exhaustive – in any case, new ways may be found or constructed, and its members may be used in conjunction with each other:

1. non-canonical names
2. canonical names
3. definite descriptions used as names
4. indexical expressions

I have been using angle brackets to represent a canonical naming device, such as quotes, or arithmetization, or some other canonical naming function. In some variations, no premise seems to be required at all. An identity is achieved by means of canonical naming and some function; for example, using arithmetization and a substitution function.

‘My favourite sentence’ is both a definite description and an indexical expression. I have treated the definite description as if it were a name in the above argument. Had I treated it in a Russelian way, then I would have used the Geach Quantified variant of Curry’s paradox towards a contradiction. However, I use it above to represent a version of the Eubulidean Liar.

Yet another way of achieving co-reference of an expression and a sentence is to use an indexical expression as in ‘This sentence is false’, which I have used in my exposition sometimes.

In my formal exposition, I use non-canonical names for the Liar-like sentences. I could have used any of the other means of achieving co-reference. This means that for each category of Liar-like paradoxes I individuate using a non-canonical name, there are also all the above versions. Each method of referring has its distinguishing features. Thus the Liar of the Unquantified Epimenides, which was introduced in Chapter 1, has
versions using each of the above methods of reference. However, its semantic
valuation is the same, which ever version is used; so, I do not count these as separate
categories of Liar-like paradoxes.

Still other variations and versions are associated with quantification, which may use:

Indefinite descriptions (or existentially quantified expressions)
Definite descriptions (analysed in the Russelian way)
Universal descriptions (or universally quantified expressions)

This list is neither exhaustive nor exclusive – expressions quantifying over the
majority of some sentences may interact with existentially quantified expressions in
paradoxical ways. The situation is considerably more complex also because
instantiation (in the derivation) might involve any member of the previous list of
devices used to achieve co-reference. It may help to explain the relationship between
these two groupings of devices for directly or indirectly achieving co-reference to
think of quantificational sentences as involving plural reference.

3.1.5 What does it mean to be a Member of this Family?

There is some fundamental contention as to which paradoxes are in the same family as
the Liar. On the one hand the truth predicate might be used, or self-reference, or
Ramsey’s [1925] division between set-theoretic and semantic paradoxes. There have
certainly been proposals for what all variations have in common: variations on Tarski’s
[1935] T-schema, or commonalities proposed for even broader groupings such as
Priest [1994] and Goldstein [2000], which go well beyond the Liar family.

My working assumption is that inferences or axioms relating to the truth
predicate are a common feature of members of the Liar family, or at least the nuclear
family, while other paradoxes depend on different axiom schemata. (Multiple
schemata may be used in hybrids.) We can articulate the difficulties in evaluating the
Liar without having a sophisticated theory or definition of truth. So, rather than
exploring the consequences of a sophisticated theory of truth, I revisit the relationships
and classification of Liar-like paradoxes using naive truth.

To begin with, I use identity sentences to individuate members of the Liar family,
as discussed in the last Chapter. I take the associated T-biconditional for granted. Both
sentences (or what they express) and associated arguments are described as
paradoxical. Context usually makes it clear which is meant.
I suggest some form of membership should also be extended to the hypodoxical duals of those paradoxical sentences (or sentences). Hypodoxes are not naively associated with arguments, and so, perhaps gain affiliated membership through their relationship to paradoxes. I suggest that relationship is one of duality.

3.2 Unquantified self-referential Liar-like Paradoxes and Hypodoxes

In this section, I present, categorise and discuss paradoxes that involve self-referential sentences without quantification. Variations are at first individuated by identity statements. In the first section, these sentences are chosen systematically in a way that uses truth functions. Among them is a new variation of these paradoxes. I will present and include their dual Truth-tellers in the categorisation. The logical principles are classical principles of sentential logic with identity extended with the truth predicate and the use of names of sentences. The biconditional in the T-schema may or may not be material, but is assumed to satisfy the Substitution Thesis that was introduced in Chapter 2. In the second subsection, I discuss some of the interesting characteristics of this subset of variations. There is still another method for individuating such variations, and in the third section, I show that this other criterion extends the set. This result drives a change in the means by which I individuate paradoxes. One can view the extended set as a closed set of valuations for these sentences; and thus, the possible valuations can be used to individuate categories of paradox or hypodox. In the fourth subsection, I consider the relationship between the Liar and these variations. In the fifth subsection, I discuss compounds of these sentences. In the sixth subsection, I consider non-truth-functional variations of some of these paradoxes.

3.2.1 Unquantified self-referential, truth-functional Variations

In this section, I investigate truth-functional variations of unquantified self-referential members of the Liar family. I begin by expanding on the individuation of such paradoxes. I investigate duality between these variations, in doing so I use valuations of 'over-determined' and 'under-determined', as discussed in the previous chapter, and I detail my use of a method of supervaluation to implement consideration of their possible truth values.

Mindful of what has been discussed in the previous chapter about individuating paradoxes and in this chapter about reference, one can replace the definite description 'my favourite sentence' with a term 'a' and present a form for a Liar sentence:
(1) \( a = \langle \neg Ta \rangle \)

for which there are associated a large number of paradoxical arguments, such as:

1. \( a = \langle \neg Ta \rangle \)  \hspace{1cm} \text{Premise}
2. \( Ta \iff \neg Ta \)  \hspace{1cm} \text{line 1, (general) T-schema}
3. \( Ta \& \neg Ta \)  \hspace{1cm} \text{Contradiction}

As discussed in Chapter 2, in the section on *Individuating Paradoxes*, I use the identity to individuate the Liar paradox. Under the standard interpretation, the identity in line 1 represents a co-reference. The identity is characteristic of large class of versions of the Liar wherein a term is co-referential with the canonical name of a sentence that denies truth of the sentence itself. I intend the identity to be *typical* with respect to any of the grammatical techniques for achieving self-reference listed previously. Possible interpretations of \( a \) include my favourite sentence or the first displayed sentence in §2 of JC Beall [2005, p. 8]. The co-reference of ‘my favourite sentence’ and “‘My favourite sentence is not true’” is an empirical fact. An equivalent first premise might be given by stipulation. The identity need not be a premise; it may be proven as a theorem. Indexicals might be used as in an expression like ‘This sentence is false’ when the reference of the indexical expression, ‘this sentence’, is the sentence itself. The indexical expression in this case and ‘This sentence is false’ are co-referential. As ‘\( a \)’ does not directly represent an indexical expression, some translation is required; nevertheless, I use the identity statement for \( a \) as typical of the Liar.

Provided such an identity, and given the T-schema, there will be an associated Liar argument. The converse will be topical in Chapter 5; but in this chapter, in order that a Liar argument will explicitly use co-referential terms, I restrict my formal proofs to using the canonical T-schema as defined in Chapter 2 and require any diagonal lemma be proven rather than assumed. As a result, Leibniz’s law, or the substitution of identicals, or the like will be used together with an identity and the Canonical T-schema or the like in the derivation. The argument above used the general T-Schema; here it is with the Canonical T-schema.

1. \( a = \langle \neg Ta \rangle \)  \hspace{1cm} \text{Premise}
2. \( T(\neg Ta) \iff \neg Ta \)  \hspace{1cm} \text{(Canonical) T-schema}
3. \( Ta \iff \neg Ta \)  \hspace{1cm} \text{2, 1 Substitution of identicals (=E)}
By varying the interpretation of ‘T’ as explained in Chapter 2, this formal identity is typical of an even larger class of versions of the Liar with respect to different bearers of truth.

I will not be distinguishing paradoxes based on argument types. Different derivations using the same identity do not yield different variations of the paradox. There are a large number of derivations that use weaker principles than the T-schema. Friedman and Sheard [1987] distinguish 20 combinations of such weaker principles that could be used to derive the above conclusion. (These weaker principles are entailed by the T-schema, but do not themselves entail the T-schema.) Nevertheless, these derivations also depend on an identity of the form in line 1 above (when using the canonical versions of the weaker principles). Although classifying derivations is of some formal interest, these derivations merely confirm that the sentence embedded in the identity is paradoxical.

The conclusion of the last derivation is a contradiction, in the form of a contradictory biconditional. The conclusion of the previous derivation was a contradiction in the form of (있다 & ~있다). One could deny that (있다 iff ~있다) is a contradiction, and deny that it entails (있다 & ~있다). Indeed, Priest [2006, p. 12] lists the validity of this step as an assumption of Tarski’s [1935]. There are logics on which this inference does not follow. Nevertheless, the biconditional is derived from the T-schema, and if one accepts the T-schema on the basis of either the Equivalence Thesis or the Substitution Thesis, which were presented in Chapter 2, then (있다 iff ~있다) either entails a material biconditional or supports substitution of equivalents. A contradiction follows from the material interpretation, or given Excluded Middle from substitution of equivalents, in which case the argument could continue as:

4. Ta ∨ ~Ta Excluded Middle
5. Ta ∨ Ta 4, 3 Substitution of Equivalents
6. Ta 5 Sentential Logic (SL)
7. Ta & Ta 6 SL
8. Ta & ~Ta 6, 3 Substitution of Equivalents

There is another candidate for individuating the Liar. The identity may not need to be made explicit, as it may be taken as proven by a theorem rather than given as a premise. Indeed, it is quite common to rely on Gödel’s diagonal lemma or the like to provide a proof for an equivalence of the form:
@ \leftrightarrow \neg T(@)

Nevertheless, the proof of an instance of such an equivalence – indeed, the proof of the lemma – relied on first proving an identity, of which line 1 above is typical. This matter will be discussed at length in Chapter 5.

Here is a proof of the equivalence, towards a proof of the Liar

1. a = \langle \neg Ta \rangle 
   premise
2. \neg Ta \iff \neg T(\neg Ta) 
   Identity (@ \rightarrow @), SL, 1 =E (Leibniz’s law)
3. A \iff \neg T(A) 
   2 with ‘A’ abbreviating ‘\neg Ta’.
4. T(A) \iff A 
   T-schema
5. A \iff \neg A 
   3, 4 SL

If the biconditional A \iff \neg T(A) was used to characterise the Liar; that would omit the case of A_2, such that A_2 \iff T(\neg A_2), as in line 3 in the argument below.

1. a = \langle \neg Ta \rangle 
   premise
2. Ta \iff T(\neg Ta) 
   Identity, SL, 1 =E (Leibniz’s law)
3. A_2 \iff T(\neg A_2) 
   2 with ‘A_2’ abbreviating ‘Ta’, SL.
4. T(\neg A_2) \iff \neg A_2 
   T-schema
5. A_2 \iff \neg A_2 
   3, 4 SL

As derived in Chapter 2, \neg T(A) \iff T(\neg A) is a theorem for a naive theory of truth anyway; even so, if one wants an expression to uniquely represent the Liar sentences, one cannot choose between the two equivalences, A \iff \neg T(A) and A \iff T(\neg A); it is the identity a = \langle \neg Ta \rangle that one wants. (I have acknowledge in Chapter 2 that there are formal issues with individuating by identities, and will ultimately depend upon a semantic categorization in a subsequent subsection.)

I will, therefore, continue to present identity sentences rather than equivalences as representative and typical of the various paradoxes. The paradoxical sentence is not of course the identity, a = \langle \neg Ta \rangle, but \neg Ta, mentioned in the identity. The Liar is paradoxical given the identity, whatever the other empirical circumstances.

Whether some other sentence is paradoxical may depend on the empirical circumstances. Sentences stating these circumstances may be represented by Q.
Along these lines, to form an *Unquantified Epimenides* sentence, roughly speaking, add an arbitrary conjunct:

\[(2) \quad b = (\neg Tb \& Q)\]

This is a way of naming a representation of my favourite conjunction that just happens to be ‘My favourite conjunction is false and Q’. If this conjunction is true, then it is false; so it must be false. A conjunction is only false if at least one of the conjuncts is false. We have already proven the first conjunct; so Q must be false. Therefore, \(\neg Q\).

Here is a formal derivation of this argument:

1. \(b = (\neg Tb \& Q)\) \hspace{1cm} \text{Premise}
2. T(\neg Tb \& Q) iff. \neg Tb \& Q \hspace{1cm} \text{T-schema}
3. Tb iff. \neg Tb \& Q \hspace{1cm} 2, 1 \text{ =E}
4. \(\neg Tb\) \hspace{1cm} \text{Assumption}
5. \(\neg Tb \& Q\) \hspace{1cm} 4, 3 \text{ SL}
6. \(\neg Tb\) \hspace{1cm} 5 \text{ &E}
7. \(\neg Tb\) \hspace{1cm} 4-6 \text{ =I}
8. \(\neg(Tb \& Q)\) \hspace{1cm} 7, 3 \text{ SL}
9. Tb ∨ \neg Q \hspace{1cm} 8 \text{ DeM, DN}
10. \(\neg Q\) \hspace{1cm} 9, 7 \text{ DS}

While this does not result in a direct contradiction, it is paradoxical enough. We can imagine (a possible world in which) I favour this conjunction and that Q is true. That is, we can imagine circumstances in which the premise is true and the conclusion false. The argument is intuitively semantically invalid, yet the above derivation seems proof-theoretically valid. In any case, the above argument form has paradoxical instances. To deny this, would commit one to denying the same claim that is regularly made with respect to Curry’s paradox. In any case, if Q is a contingent truth, we can add it as a premise and derive a contradiction. (If Q is a necessary truth, we do not need to add it as a premise; \(\neg Q\) itself is a contradiction.)

Semantically, if Q is true, then \(b\) is over-determined. If \(b\) is true, the conjunct must be true and so \(b\) is false; moreover, if \(b\) is false then the conjunction is true, so \(b\) is true. Hence, given Q is true, \(b\) is naively over-determined.
To form an identity related to the Self-referential Curry paradox, use a disjunct as so:

\[ c = \langle \neg Tc \lor Q \rangle^2 \]

This could represent my favourite disjunction that just happens to be ‘Either my favourite disjunction is not true or Q’. From this identity, it follows that Q. For if c is false then both disjuncts must be false, but the first disjunct says c is false; so c must be true. Then by disjunctive syllogism, Q follows. The Curry sentence is usually presented as a conditional, Tc \rightarrow Q, but I am distinguishing the material version and representing it using disjunction for a later generalisation of the paradox. Here is a formal derivation of Curry’s paradox:

1. \( c = \langle \neg Tc \lor Q \rangle \) 
   \hspace{1cm} \text{Premise}
2. \( T \langle \neg Tc \lor Q \rangle \text{ iff. } \neg Tc \lor Q \) 
   \hspace{1cm} \text{T-Schema}
3. \( Tc \text{ iff. } \neg Tc \lor Q \) 
   \hspace{1cm} \text{2, 1 =E}
4. \( \neg Tc \) 
   \hspace{1cm} \text{Assumption}
5. \( \neg \neg Tc \lor Q \) 
   \hspace{1cm} \text{4, 3 SL}
6. \( Tc \& \neg Q \) 
   \hspace{1cm} \text{5 DeM, DN}
7. \( Tc \) 
   \hspace{1cm} \text{&E}
8. \( Tc \) 
   \hspace{1cm} \text{4-7 =E}
9. \( \neg Tc \lor Q \) 
   \hspace{1cm} \text{8, 3 SL}
10. \( Q \) 
    \hspace{1cm} \text{9, 8 DS}

There is another related paradox that embeds my favourite biconditional, ‘My favourite biconditional is not true if and only if Q’,\(^3\) which we shall represent using ‘d’:

\[ d = \langle \neg Td \leftrightarrow Q \rangle \]

\(^2\) While I try to respect existing nomenclature, I find I have to adapt nomenclature towards my classification from this point. This is the material variation of Curry’s paradox using self-reference in the way Geach and Löb rendered it, as described in Chapter 1. I treat non-material interpretations in a later section. Graham Priest also distinguishes the material and non-material interpretations of Curry’s paradox in representing the structure of paradoxes [1994, p. 33]. Yet Priest [1994] seeks to maximise the uniform treatment of the paradoxes. So, this distinction in my classification is not contentious.

\(^3\) As for the other paradoxes in this section, there is also a version of this paradox using a demonstrative self-referentially as in ‘This sentence is not true if and only if Q’ when said of itself.
Sentence $d$ is logically distinct by truth-tables from $a$, $b$, or $c$; and the associated paradox has some interesting properties. For the Liar proves that sentence $a$ is both true and false, whereas sentence $b$ is provably false and therefore $\neg Q$, and sentence $c$ is provably true, so $Q$ follows; but $d$ can be true or false and $\neg Q$ still follows! I note, though, that the proof requires that at least one of the conditionals in the biconditional is interpreted materially. To see this result notice that the biconditional in $d$, taken as a material biconditional, is logically equivalent to:

$$T_d \leftrightarrow \neg Q$$

Now, if $d$ is true, $\neg Q$ follows from this biconditional using modus ponens. On the other hand, notice that the negation of the material biconditional named by $d$ is logically equivalent to:

$$\neg T_d \leftrightarrow \neg Q$$

So, if $d$ is false, then we use this latest biconditional to once again derive $\neg Q$. Whether $d$ is true or false, $\neg Q$ follows. Here is a more formal derivation of the New Paradox.

1. $d = (\neg T_d \leftrightarrow Q)$  \hspace{1cm} \text{Premise}
2. $T(\neg T_d \leftrightarrow Q)$ iff. $\neg T_d \leftrightarrow Q$  \hspace{1cm} \text{T-schema}
3. $T_d$ iff. $\neg T_d \leftrightarrow Q$  \hspace{1cm} 2, 1 =E
4. $| T_d$  \hspace{1cm} \text{Assumption}
5. $| \neg T_d \leftrightarrow Q$  \hspace{1cm} 3, 4 SL
6. $| \neg Q$  \hspace{1cm} 5, 4 SL
7. $T_d \rightarrow \neg Q$  \hspace{1cm} 4-6 CP
8. $| Q$  \hspace{1cm} \text{Assumption}
9. $| \neg T_d$  \hspace{1cm} 7, 8 MT
10. $| (\neg T_d \leftrightarrow Q)$  \hspace{1cm} 9, 3 SL
11. $| (\neg T_d \rightarrow Q) \lor (\neg Q \rightarrow \neg T_d)$  \hspace{1cm} 10, Equiv, DeM
12. $| (\neg T_d \rightarrow Q)$  \hspace{1cm} 11, 7 SL
13. $| \neg T_d \& \neg Q$  \hspace{1cm} 12, material conditional defn, DeM
14. $| \neg Q$  \hspace{1cm} 13E
15. $\neg Q$  \hspace{1cm} 8-14 \neg I

The definition of the material conditional was used in going from line 12 to 13.
I first introduced this paradox in a paper to the ANU Philosophical Society in October 2002, and in a paper at the Australasian Association of Philosophy Conference in July 2004. I am hoping it might be called the Eldridge-Smith paradox (or ESP for short). It requires a truth functional interpretation of \( d \). Nevertheless, the biconditional of the T-schema, the main biconditional of Line 2, need not be a material conditional for the paradoxical argument to follow.

I turn now to the question of whether sentences \( a, b, c \) and \( d \) have Truth-teller-like duals. Syntactically, our representation of the Truth-teller is just the negation of \( a \) with re-lettering as \( e \):

\[
(5) \quad e = \langle Te \rangle
\]

So, to form the other corresponding Truth-teller-like sentences, take the external negation of the expression and re-letter the self-referential names as in the identity sentences below.

\[
(6) \quad f = \langle T_f \& Q \rangle \\
(7) \quad g = \langle T_g \lor Q \rangle \\
(8) \quad h = \langle T_h \leftrightarrow Q \rangle
\]

The other alternative to form candidate Truth-tellers would be an 'internal negation' just negating the self-referential component in the expression using \( a, b, c \) and \( d \) and re-lettering to form the new self-referential expressions. I use the external negation, because in the quantified cases the result of the external negation with re-lettering is hypodoxical under the same conditions that the original was paradoxical. (If Epimenides says instead 'Some Cretan sentence is true' when in fact all other Cretan sentences are false, then there is a lack of a reason to determine whether his sentence is true or false. It is hypodoxical under these circumstances.)

Let me tabulate the paradoxes and hypodoxes identified so far, as they form an intuitive subset. In the table below, the first column provides a common name, while the second lists an identity sentence that individuates the sentences identified by that common name (and mentions them in a canonical name on the right-hand side of the identity sign). The third column lists some significant consequence which seems to be validly derivable from the identity and the T-schema. The fourth and fifth columns give the semantic value of the expression given \( Q \) or \( \sim Q \) respectively. For some of them, their valuations differ depending on the status of contingencies represented by \( Q \).
<table>
<thead>
<tr>
<th>Common name</th>
<th>Identity premise</th>
<th>Consequences</th>
<th>Given Q</th>
<th>Given ~Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liar</td>
<td>( a = (\sim T a) )</td>
<td>( T a \ \text{iff} \ \sim T a )</td>
<td>Over-determined</td>
<td>Over-determined</td>
</tr>
<tr>
<td>Unquantified Epimenides</td>
<td>( b = (\sim T b \land Q) )</td>
<td>( \sim T b \land \sim Q )</td>
<td>Over-determined</td>
<td>False</td>
</tr>
<tr>
<td>Self-referential Curry</td>
<td>( c = (\sim T c \lor Q) )</td>
<td>( T c \land Q )</td>
<td>True</td>
<td>Over-determined</td>
</tr>
<tr>
<td>ESP paradox / hypodox</td>
<td>( d = (\sim T d \leftrightarrow Q) )</td>
<td>( \sim Q )</td>
<td>Over-determined</td>
<td>Under-determined</td>
</tr>
<tr>
<td>Truth-teller</td>
<td>( e = \langle T e \rangle )</td>
<td></td>
<td>Under-determined</td>
<td>Under-determined</td>
</tr>
<tr>
<td>Epimenidean Truth-teller</td>
<td>( f = \langle T f \lor \sim Q \rangle )</td>
<td>( T f \lor Q )</td>
<td>Under-determined</td>
<td>True</td>
</tr>
<tr>
<td>Curried Truth-teller</td>
<td>( g = \langle T g \land \sim Q \rangle )</td>
<td>( \sim T g \lor \sim Q )</td>
<td>False</td>
<td>Under-determined</td>
</tr>
<tr>
<td>ESP hypodox / paradox again</td>
<td>( h = \langle T h \leftrightarrow Q \rangle )</td>
<td>( Q )</td>
<td>Under-determined</td>
<td>Over-determined</td>
</tr>
</tbody>
</table>

In considering the tables, the identities should be held fixed, and the semantic status of Q varied. This could be thought of in two ways. Firstly by varying the model, and thereby the semantic value of Q; or, secondly, by varying the sentence that Q represents (and using the one model). In general, I mean the former.

The valuations in the last two columns are reached in one or two steps. As described in Chapter 2, one may first consider the compositional valuation of the sentence (using essentially a strong Kleene valuation scheme but without giving away the naive tenet of bivalence), then if the result is indeterminate, one may use a supervaluation (with the evaluation of a T-sentence defined in an analogous way to the T-schema). So, the valuations in the table are produced by first using compositional reasoning, in the sense defined in Chapter 2. If a valuation is reached in this first way, the sentence is said to be 'compositionally true' or 'compositionally false' under the circumstances. If a valuation is not reached this way, the sentence's truth value is compositionally indeterminate. Rather than settle on this semantic valuation, the second method is then used in these cases. Naively, their possible values are true or false. Each is considered in turn. If only one valuation is consistent, then the sentence
takes that value. However, in addition, a valuation describing the result may be returned. In the cases tabulated, the sentences are, in some circumstances represented by \( Q \) or \( \neg Q \), either over-determined, as they would receive both truth values if evaluated this way; or under-determined, as they could take either value consistently, but which value is not settled by supervaluation. The supervaluation method is used only to evaluate individual sentences that have not received a compositional valuation, not collectively.

Take the valuation of \( b \), where \( b = (\neg Tb \land Q) \), for example. Assume that the value of \( Q \) is settled. So, the value of \( \neg Q \) is settled compositional. In the first case, assume \( Q \) is false. Then, compositionally the conjunction \( \neg Tb \land Q \) is false. In the second case, assume \( Q \) is true. Now, the conjunction \( \neg Tb \land Q \) is compositionally indeterminate. The supervaluation of the conjunction holds the given value of \( Q \) fixed and varies the value of \( b \) as per the table below. In this table, the valuations in column 3 are obtained by the classical valuation rule for conjunction using columns 2 and 3, and the valuations in column 4 are obtained from 3 using the classical valuation of negation; whereas, the valuation in column 5 is obtained by using the valuation rule corresponding to the T-schema with column 1. As \( 'b' \) and \( '(\neg Tb \land Q)' \) are co-referring terms, the sentence evaluated in column 5 is the same as column 4. Thus \( b \) has an over-determined valuation.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \neg Tb \land Q )</td>
<td>( Q )</td>
<td>( \neg Tb )</td>
<td>( Tb )</td>
<td>( T(\neg Tb \land Q) )</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
</tbody>
</table>

Take the valuation of \( f \), where \( f = (Tf \lor \neg Q) \), as another example. In the first case, \( Q \) is false; so, \( \neg Q \) is compositionally true, and so is the disjunction. In the second case, \( Q \) is true and \( f \) is compositionally indeterminate. The following table summarises the subsequent supervaluation. It demonstrates that \( f \) can consistently be allocated either truth value under this circumstance. Column 4 is the valuation reached from columns 1 and 3 using the disjunction valuation rule. Column 5 is obtained from column 1 using the valuation rule corresponding to the T-schema. Column 5 is a rewrite of column 1, as \( 'f' \) and \( '(Tf \lor \neg Q)' \) are co-referring terms. The valuations in column 6 are obtained by using the valuation rule corresponding to the T-schema.
based on column 5. The table shows how supervaluation does not settle the truth value of \( f \), and the resulting valuation is under-determined.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T \lor \neg Q )</td>
<td>Q</td>
<td>( \neg Q )</td>
<td>( T )</td>
<td>( T \lor \neg Q )</td>
<td>( T \lor \neg Q )</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

Given the previous result for \( b \) and the duality conjecture, this result could be predicted for \( f \).

Consider, in contrast, the valuation of \( i = \langle T \lor \neg T \rangle \). It is compositionally indeterminate whether or not \( Q \); nevertheless, its supervaluation settles on a truth value, as per the following table.

<table>
<thead>
<tr>
<th>1</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T \lor \neg T )</td>
<td>( T )</td>
<td>( \neg T )</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

Applying a valuation top-down, columns on the left still determine columns on the right. The table sets out the truth-table for disjunction in the reverse of the normal order, but still allowing for all possible ways in which a disjunction might be true or false. However, the last two rows are inconsistent with the valuation rule for negation. In each of the consistent first two rows, \( i \) is true. So, a top-down use of supervaluation in this way settles that \( i \) is true.

Consider also, \( j = \langle T \land \neg T \rangle \). It is also compositionally indeterminate whether or not \( Q \); nevertheless, supervaluation settles that it is false, as per the following table. Only the last two rows of this table are consistent, and in both these cases \( j \) is false.

<table>
<thead>
<tr>
<th>1</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T \land \neg T )</td>
<td>( T )</td>
<td>( \neg T )</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>
Neither 'over-determined' nor 'under-determined' refers to a value, as the naive position is that the Liar and the Truth-teller ought to have a semantic value which is one member of \{true, false\}. Thus, for these sentences, in certain circumstances, there is no material basis to compositionally determine their truth values. They are \textit{indeterminate} in this first way. Subsequently consideration of their possible valuations is confounded. For a sentence which is paradoxical under the circumstances, being true would determine that it is false, yet being false would determine that the paradoxical sentence is true. For a sentence which is hypodoxical under the circumstances, being true determines that it is true and being false determines that it is false; but there is a lack of any reason to choose which value it should take.

'Over-determined' is not a truth-value glut and 'under-determined' is not a truth-value gap. A truth-value glut is a value, whereas 'over-determined' is a description of how our naive attempt to arrive at a single value is confounded. A truth-value gap is not a valuation but the absence or lack of a truth value, whereas 'under-determined' is a description of how our naive valuation methods provide no means to determine whether the sentence is true or false.

In the case of a sentence like \(a\), it would follow by supervaluation that it is both true and false. Naively, it has just one truth value; so, naively it is over-determined. In the case of a sentence like \(e\), nothing more follows, it is either true or false; so, its valuation is under-determined. (Note also that I am speaking of being logically under-determined rather than any epistemic constraint, although I dare say that the first entails the second). The Liar and the Truth-teller cannot be semantically distinguished by purely compositional reasoning. They are both \textit{indeterminate} on that basis. In general, over-determined and under-determined sentences are \textit{materially indeterminate} in the given circumstances.\(^4\) The Liar and Truth-teller can be semantically distinguished by supervaluation (using the evaluation of T-sentences). They are respectively over-determined and under-determined on that basis. Other sentences, \(b\) through \(d\) and \(f\) through \(h\), are over-determined or under-determined in given circumstances (represented in the table by \(Q\) and \(\neg Q\)).

\(^4\) The expressions 'determinately true' and 'definitely true' are used in the literature in similar ways. They are typically attributed to true grounded sentences under some definition of \textit{groundedness}. I have used 'indeterminate' for the semantic valuation of sentences that do not receive a truth value compositionally; but I consider sentences that receive a truth value either compositionally or through subsequent supervaluation to be grounded; so that a compositionally indeterminate sentence may yet turn out to be grounded (based on supervaluation).
There are a couple of intuitive symmetries between these sentences and their valuations. Firstly, there is the paradox-hypodox duality evident in these tabulated cases. The Liar, $a$, and the Truth-teller, $e$, are always paradoxical and hypodoxical respectively. The Unquantified Epimenides, $b$, and the Epimenidean Truth-teller, $f$, are paradoxical and hypodoxical respectively, when $Q$ is true. The Self-referential Curry paradox and Curried Truth-teller are paradoxical and hypodoxical respectively when $Q$ is false (or $\neg Q$ is true). An interesting feature of the ESP is that it is, in a sense, its own hypodox. Sentence $d$ is paradoxical and sentence $h$ is hypodoxical when $Q$ is true, whereas sentence $h$ is paradoxical and $d$ is hypodoxical when $Q$ is not true. The Dual Conjecture is thus supported in the eight cases in the table above. At least for these cases, a Liar-like sentence is paradoxical iff its external negation with relettering is hypodoxical.

The second intuitive symmetry suggests this set is in some sense complete. Given the T-schema, $a = \langle \neg Ta \rangle$ entails a contradiction; $b = \langle \neg Tb \& Q \rangle$ entails $\neg Tb$ and $\neg Q$; while $c = \langle \neg Tc \lor Q \rangle$ entails $Tc$ and $Q$; $d = \langle \negTd \leftrightarrow Q \rangle$ entails $\neg Q$, and $h = \langle Th \leftrightarrow Q \rangle$ entails $Q$. The intuitive symmetry is between the Liar argument proving a contradiction, while the ESP argument entails some arbitrary sentence; and the same sort of intuitive symmetry is evident between the Unquantified Epimenides argument entailing the sentence $b$ is (at least) false and an arbitrary sentence $\neg Q$, while the Self-referential Curry paradox argument entails the sentence $c$ is (at least) true and proves $Q$. In the next two sections, I try to capture a sense in which this set of sentences could be said to be a complete set.

3.2.2 An interesting Subset of unquantified self-referential Sentences

One would hardly claim soundness for this collection; nevertheless, there may be some sense in which they form a complete subset of unquantified self-referential sentences that are paradoxical or hypodoxical under some circumstances. These sentences were
obtained by systematically generating self-referential sentences involving truth-functions. My approach in this section will be to investigate all possible unary truth-functional operators and binary truth functions with a view to proving these are the only ones that are paradoxical or hypodoxical under some circumstances. (Here and throughout I will take the identities $a$ through $h$ as given.)

Unlike sentence $a$, sentences $b$, $c$, $d$, and $h$, are not always paradoxical. For example, if $Q$, then $c$ is true. Unlike sentence $e$, sentences $d$, $f$, $g$, and $h$ are not always under-determined. For example, if $Q$, then $g$ is simply true. Sentences $b$ and $c$ could be said to be *contingently paradoxical*, but that awkward phrase gives the wrong impression. A sentence which is *contingently true* is none the less true. What is intended by ‘contingently paradoxical’ is that there are possible circumstances under which the sentence is paradoxical. So I have been careful to say that these sentences are paradoxical or hypodoxical *under some circumstances*. One could say that these sentences are not necessarily grounded; under some circumstances, neither compositional valuation nor supervaluation returns a truth value for them.

The subset I have in mind is obtained by considering the ways in which truth functional operators and connectives can be combined with self-referential $T$-sentences. The Liar and the Truth-teller represent two of four possible truth-functional operators on such a $T$-sentence. Let ‘$\heartsuit$’ represent the operator that returns the truth value true. When $@$ is instantiated with a sentence that is compositionally indeterminate, however, then $\heartsuit@$ is compositionally indeterminate; nevertheless, when one supervaluates $\heartsuit@$ is always true, and the expression is said to be *intrinsically true*. The expression originates with Kripke [1975, p74] who talks about similar expressions as ‘intrinsic truths’. Just to be clear, the semantic value of an ‘intrinsic truth’ is true, prefixing the value with ‘intrinsically’ says something about how the valuation was reached, as opposed to ‘compositionally’.

Then the four possibilities are:

<table>
<thead>
<tr>
<th>Common name</th>
<th>Identity premise</th>
<th>Consequences</th>
<th>Semantic valuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liar</td>
<td>$a = \langle \neg T a \rangle$</td>
<td>$T_0 \wedge \neg T_0$</td>
<td>Over-determined</td>
</tr>
<tr>
<td>Truth-teller</td>
<td>$e = \langle T e \rangle$</td>
<td>$T_0 \vee \neg T_0$</td>
<td>Under-determined</td>
</tr>
<tr>
<td>a tautology</td>
<td>$i_0 = \langle \heartsuit T_{i_0} \rangle$</td>
<td>$T_{i_0}$</td>
<td>(Intrinsically) true</td>
</tr>
<tr>
<td>a contradiction</td>
<td>$j_0 = \langle \neg \heartsuit T_{j_0} \rangle$</td>
<td>$\neg T_{j_0}$</td>
<td>(Intrinsically) false</td>
</tr>
</tbody>
</table>
In addition to these operators, there are of course sixteen possible binary truth-functions. I will deal with them in groups. In the first are the two connectives one of which always returns ‘true’ and the other always returns ‘false’. In the second are four other trivial truth-functions. The third are the group which I will map to \( b, c, d, f, g, \) and \( h \). (It will turn out that \( h \) can be mapped back to \( d \) in the same way.)

The connective that always returns true whatever the values of its arguments can be reduced to the \( \heartsuit \)-operator and conjunction. Alternatively, it might be equated with the sentence \( i, i = \langle Ti \lor \neg Ti \rangle \); a compositionally indeterminate sentence that does have its truth value settled by supervaluation. The sentence is another intrinsic truth.

In any case, this connective can yield no new paradoxes or hypodoxes because it always evaluates as a tautology. Likewise the connective that always returns false can be reduced to negation, the \( \heartsuit \)-operator and conjunction. Alternatively, it might be equated with sentence \( j \), which is individuated by the identity \( j = \langle Tj \land \neg Tj \rangle \) and is naively, intrinsically false.\(^7\) In any case this connective can yield no new paradoxes or hypodoxes because it always evaluates as false. This leaves only fourteen of the sixteen possible binary truth-functions.

Of these remaining fourteen, there are four further trivial truth-functional connectives: one returns the semantic value of the first component, another returns the truth value of the negation of the first component, and two others return like values for the second component. Consider the connective that returns the semantic value of the negation of the first component, represented by ‘\( \blacklozenge \)’. The evaluation of \( a_i = \langle Ta_i \blacklozenge Q \rangle \) is always over-determined. \( Q \) is irrelevant in the evaluation of this expression. Its evaluation reduces to the Liar sentence. Similarly, the connective that returns the same truth value as that of the first component can be used in this way to generate another representation for the Truth-teller. (If I use a contingent truth in the first component, the result is a contingent sentence.) The functions favouring the second argument can be dealt with similarly. Thus, these connectives yield no new variations of the paradoxes or hypodoxes.

We are left with ten truth functions. These may be represented:

\(^7\) Sentence \( i \) is discussed in Kripke [1975]. The truth values of \( i \) and \( j \) are proven in Kukla [1985] using an analogous proof to Mortensen and Priest’s [1981] proof that the truth-teller (sentence \( e \)) is true or false. Mortensen’s and Priest’s [1981] Truth-teller Paradox shows that although one might think \( e \) is neither true nor false, this assumption gives rise to a reductio.
(i) \( P \& R \)
(ii) \( P \lor R \)
(iii) \( P \equiv R \)
(iv) \( \neg P \& R \)
(v) \( \neg P \lor R \)
(vi) \( \neg P \equiv R \)
(vii) \( P \& \neg R \)
(viii) \( P \lor \neg R \)
(ix) \( \neg P \& \neg R \)
(x) \( \neg P \lor \neg R \)

First consider the self-referential sentences obtained by mapping \( P \) to a self-referential expression \( \langle T\alpha \rangle \), such that \( \alpha = \langle \{\sim\} T\alpha \Delta Q \rangle \), where alpha is replaced by a name that refers to the whole expression in angle brackets, tilde is optional, delta is replaced by a truth-functional connective and \( Q \) is equivalent to \( R \) or \( \neg R \). Since \( Q \) was arbitrary, these are not really new variations. Moreover, if we interpret the biconditional in the ESP as a material biconditional, then there seems little difference between \( d \) and \( h \). Then there are the following mappings.

(iv) and (ix) map to \( b \).
(v) and (x) map to \( c \).
(i) and (vii) map to \( g \).
(ii) and (viii) map to \( f \).
(vi) maps to \( d \).
(iii) maps to \( h \).

Since (vi) is truth-functionally equivalent to \( P \equiv \neg R \) and \( Q \) is arbitrary, then (vi) also maps to \( h \). Thus, the ten remaining truth functions map to five categories of variations that are paradoxical or hypodoxical under some circumstances.

We have a second type of case to consider; the self-referential sentences obtained by mapping \( P \) and \( R \) to a self-referential expression \( \langle T\alpha \rangle \), such that \( \alpha = \langle \{\sim\} T\alpha \Delta \{\neg\} T\alpha \rangle \), where alpha is replaced by a name that refers to the whole expression in angle brackets, tilde is optional, delta is replaced by a truth-functional connective. Now we have these possibilities:

(i) \( P \& P \)
(ii) \( P \equiv P \)
(iii) $\neg P \& P$
(iv) $\neg P \lor P$
(v) $\neg P \equiv P$
(vi) $\neg P \& \neg P$

For this set of six,

(i) maps to $e$.
(ii) and (iv) are tautologies that map to *intrinsic* truths. In particular, (iv) maps to $i = (T_i \lor \neg T_i)$.
(iii) and (v) are contradictions that map to *intrinsic* falsehoods. In particular, (v) maps to $j = (T_j \& \neg T_j)$.
(vi) maps to $a$.

No new variations of possibly paradoxical or hypodoxical sentences were introduced in this way. The exercise is complete. The variations $a$ through to $h$ are all those that can be obtained by consideration of operators or binary truth-functional connectives. Thus, there appear to be only seven unquantified self-referential sentences that are paradoxical or hypodoxical under some circumstances, counting the ESP once instead of twice.

Life would be simpler if all n-ary truth-functional connectives could be reduced to binary and unary ones. While this is the case for classical sentential logic; one can hardly rely on that claim for the self-referential extension that admits paradoxes and hypodoxes.

### 3.2.3 A complete Subset of Variations

In this section, I complete the set of self-referential variations that is truth-functionally closed. My strategy now is to examine possible combinations of *supervaluations*. The subset of interest consists of those involving a valuation of ‘under-determined’ or ‘over-determined’.

If the ternary and greater truth-functional connectives can be reduced to binary truth functions, these other n-ary connectives will introduce no new paradoxes or hypodoxes. However, this assumes that any other truth-functional combination will reduce in the normal way to an “equivalence” with one of the sentences. It will be hard to make sense of this “equivalence”. One way in which a sort of equivalence can be warranted is with respect to the possible combination of valuations and indeterminacy under different circumstances. In this respect, the set of seven valuations above are just
an interesting subset of the combinations available. What is wanted are sentences with possible valuations corresponding to each of the subsets of \{ true, false, over-determined, under-determined \} that contain over-determined or under-determined.

Unquantified sentences taking the same *supervaluations* under the same circumstances will then be equivalent in a sense to any other unquantified sentence taking these *supervaluations*. The results of supervaluations are considered for each set of contingencies while holding the identity of various self-referential sentences fixed. As already explained, 'over-determined' and 'under-determined' are not new valuations in addition to 'true' and 'false'. The former are descriptions of why a truth value has not been settled. While these valuation descriptions are not valuations in themselves, they can be treated as distinguishable and used to individuate Liar-like variations. One begins to generalize from these “equivalences” by abstracting away from the circumstances by letting P and Q be arbitrary. Indeed, the categories are made more general still by disregarding the order in which the valuations are reached and simply grouping each subset of the valuations described in the previous paradox as a category. I should add that the categories referred to are for self-referential, unquantified sentences.

This set of possible valuations (and valuation descriptions) will be closed under truth-functional operations because of the normal truth-functional equivalences, as P and Q can be used to group any of the components with compositionally determinate truth values. In this way, the seven variations considered so far represent a complete subset that is the union of the set of variations with one valuation and the set of variations with two valuations, which differ depending on the circumstances represented by Q as follows.

Those variations involving just one supervaluation from \{ true, false, over-determined, under-determined \} are just the Liar and the Truth-teller:

<table>
<thead>
<tr>
<th>Common name</th>
<th>Identity premise</th>
<th>Semantic valuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liar</td>
<td>( a = \langle \neg T a \rangle )</td>
<td>Over-determined</td>
</tr>
<tr>
<td>Truth-teller</td>
<td>( e = \langle T e \rangle )</td>
<td>Under-determined</td>
</tr>
</tbody>
</table>

Those variations involving just two supervaluations are the remaining members of the set of seven.
I now have made tacit use of another method of individuating these paradoxes, their possible semantic valuations. Curry’s paradox is certainly distinct from its dual hypodox. Curry’s paradox is never under-determined, whereas the Curried Truth-teller is; so, the two are clearly distinct variations. On its possible semantic supervaluations, however, the ESP is one variation. It is one, which may be paradoxical or hypodoxical depending on the circumstances. However, using the identity sentences the ESP is two variations:  
\( d = \langle \neg Td \leftrightarrow Q \rangle \) and \( h = \langle Th \leftrightarrow Q \rangle \). The semantic difference between these two is that one will be paradoxical in circumstances when the other is hypodoxical. So, they are distinguishable. A purely formal distinction could be made between \( d = \langle \neg Td \leftrightarrow Q \rangle \) and \( d_2 = \langle Td_2 \leftrightarrow \neg Q \rangle \); but as such a distinction makes no naive semantic difference, there is insufficient reason to make this distinction. One is guided by semantic differences in individuating the variations.

Those variations involving just three or four distinct supervaluation results are represented by the following:
<table>
<thead>
<tr>
<th>Identity premise with sentence embedded</th>
<th>Given Q &amp; P</th>
<th>Given P &amp; ¬Q</th>
<th>Given ¬P &amp; Q</th>
<th>Given ¬P &amp; ¬Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k = (P &amp; (Tk \lor Q)))</td>
<td>T</td>
<td>Under-determined</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>(l = (P \lor (~Tl \land Q)))</td>
<td>T</td>
<td>T</td>
<td>Over-determined</td>
<td>F</td>
</tr>
<tr>
<td>(m = (P \lor (Tm \leftrightarrow Q)))</td>
<td>T</td>
<td>T</td>
<td>Under-determined</td>
<td>Over-determined</td>
</tr>
<tr>
<td>(n = (P &amp; (Tn \leftrightarrow Q)))</td>
<td>Under-determined</td>
<td>Over-determined</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>(o = ((\neg P &amp; Q) \lor [(To \leftrightarrow (Q &amp; P)) \land (Q \lor \neg P)]))</td>
<td>Under-determined</td>
<td>F</td>
<td>T</td>
<td>Over-determined</td>
</tr>
</tbody>
</table>

Thus there are twelve purely self-referential, unquantified variations that are distinguished by their different supervaluations.

With respect to sentences \(k\) through \(o\), one should like to know whether the dual conjecture is born out. The table above provides the explicit duals for these specific sentences; but no new cases are really added to our set of twelve in doing so. As \(P\) and \(Q\) are arbitrary, \(k_2\) below has the same set of possible valuations as \(l\). Sentence \(l_2\) below has the same set of possible valuations as \(k, m_2\) as \(n\), and \(n_2\) as \(m\). Like the ESP, \(o\) is in a sense its own dual, as \(o_2\) has the same set of valuations as \(o\).
<table>
<thead>
<tr>
<th>Identity premise with sentence embedded</th>
<th>Given Q &amp; P</th>
<th>Given P &amp; ~Q</th>
<th>Given ~P &amp; Q</th>
<th>Given ~P &amp; ~Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_2 = \langle \neg P \lor (\neg T_k \land \neg Q) \rangle$</td>
<td>F</td>
<td>Over-determined</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>$l_2 = \langle \neg P \land (T_{l_2} \lor \neg Q) \rangle$</td>
<td>F</td>
<td>F</td>
<td>Under-determined</td>
<td>T</td>
</tr>
<tr>
<td>$m_2 = \langle \neg P \land (T_{m_2} \leftrightarrow \neg Q) \rangle$</td>
<td>F</td>
<td>F</td>
<td>Over-determined</td>
<td>Under-determined</td>
</tr>
<tr>
<td>$n_2 = \langle \neg P \lor (T_{n_2} \leftrightarrow \neg Q) \rangle$</td>
<td>Over-determined</td>
<td>Under-determined</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>$o_2 = \langle \neg Q \land [(T_{o_2} \leftrightarrow (\neg Q \lor \neg P)) \lor (\neg Q \land P)] \rangle$</td>
<td>Over-determined</td>
<td>T</td>
<td>F</td>
<td>Under-determined</td>
</tr>
</tbody>
</table>

I have taken self-reference strictly speaking to apply to sentences that refer to the sentence as a whole; but there are two other possibilities, either of which may generate more self-referential paradoxes. The first involves self-referential components, where a disjunct refers to itself for example, rather than the whole sentence. The second involves truth-functional compounds of the purely self-referential sentences given so far. I will consider each of these in turn.

### 3.2.4 Unquantified truth-functional Compounds

So far we have restricted our expressions to ones that are themselves self-referential, whereas we could have referred to suitable self-referential expressions. Sentences $k_3$, $l_3$, $m_3$, and $n_3$ are not themselves self-referential, and yet will have the same possible valuations as $k$, $l$, $m$, and $n$ respectively.

\[
k_3 = \langle P \lor Th \rangle
\]

\[
l_3 = \langle P \land Tf_3 \rangle
\]

\[
m_3 = \langle P \lor Th \rangle
\]

\[
n_3 = \langle P \land Th \rangle
\]
Thus, in a sense, at least four of our extra five cases can be built from our existing eight. We have also cases where the eight can be combined, such as

\[
Tb \lor Tc \\
Ta \lor Te \\
Td \& Te
\]

Combinations of Liars and Truth-tellers have been of particular interest in theoretical classifications, as we shall see.

### 3.2.5 Relationships of the Liar, Epimenides, Curry and ESP

A too simple relationship between the Liar and its unquantified relatives suggests itself: that the others are obtained from the Liar by adding a conjunct, disjunct, or combining with another sentence using a biconditional. Similar views to this effect have been advocated by Mackie [1973, pp. 276-290] and Goldstein [1986, passim]. I shall now begin to argue that the relationship is not quite so simple.

When we consider sentences \( b_2, c_2, \) and \( d_2 \) below (that embed a reference to my sentence \( a \)), they look like my Epimenides, Curry and ESP respectively, that is, sentences \( b, c \) and \( d \).

\[
\begin{align*}
b_2 &= \langle \neg Ta \& Q \rangle \\
c_2 &= \langle \neg Ta \lor Q \rangle \\
d_2 &= \langle \neg Ta \equiv Q \rangle 
\end{align*}
\]

From the identity of \( b_2 \), we can derive \( \neg Tb \& \neg Q \) by using an argument similar to the Epimenides above. Likewise, from the identity of \( c_2 \) we can derive \( T c_2 \& Q \) using an argument similar to the Curry argument above. And from the identity of \( d_2 \) one can prove \( \neg Q \). However, given the identity of \( a \) one can prove anything (classically).

The folklore belief is that the Liar paradox itself is explicitly or tacitly used in the reasoning for the Unquantified Epimenides and the Self-referential Curry.\(^8\)

Consider in comparison the following alternative versions:

This conjunct is false \& Q.

This disjunct is false \lor Q.

---

\(^8\) It now becomes necessary to distinguish ‘the Liar’ as in the Liar paradox from the Liar family, which includes the Liar Paradox, the Epimenides, Curry’s and the ESP as members.
This component is false $\leftrightarrow Q$.

Consider first the conjunction. Is the Liar contained in this conjunction, or the self-referential sentence $b$? I believe it is the former, because the Liar does not contain anything else in the scope of its self-reference – this relation is preserved by the self-referential conjunct. Furthermore, in sentential logic, both conjuncts and sentences are treated as sentential expressions. The truth value of the self-referential conjunct is the same as the Liar sentence, it is paradoxical. And, as a result, the whole conjunction is paradoxical whether or not the second conjunct is true or false. This is not the case with the Unquantified Epimenides, $b$. In this latter case, where the second conjunct is contingent, then it is an empirical matter whether or not the expression is paradoxical. The point is that the possible semantic values of the above expressions are distinct from the corresponding semantic values of the expression below for some semantic values of $Q$.

This conjunction is false & $Q$.
This disjunction is false $\lor Q$.
This biconditional is false $\leftrightarrow Q$.

I conclude that the folklore belief properly relates to the former Liar versions, and not to the self-referential variations that I have been individuating.

3.2.6 *Unquantified non-truth-functional Interpretations*

The Equivalence Thesis warrants any of the truth-functional proofs given beforehand; in addition, the Substitution Thesis can be used in derivations for self-referential sentences involving the entailment relation or non-truth-functional connectives. To obtain these proofs, first convert the sentences to ones using negation, if required, and an arrow instead of other truth functions. Then cease to rely on the arrow having a truth-functional interpretation and seek a proof not reliant on the material conditional or biconditional. (One does not have in mind a particular logic but just one with a suitable non-material biconditional.) I have done this for the Epimenides at the end of Chapter 1. It can be done for Curry's paradox, of course.

I do not believe it can be done for the ESP. Nevertheless, this does not mean that the biconditional in the T-schema must be material for the ESP to obtain. That biconditional, represented by 'iff', might be given a non-material interpretation consistent (so to speak) with the proof below.
3.3 Theoretical Comparisons

In this section, I take an interlude from developing the naive classification, to make comparisons between the classification given so far and some theory-relative classifications, in particular those of Kripke [1975], and Gupta & Belnap [1993].

3.3.1 Kripke’s Theory of Truth-value Gaps

This subsection compares my classification of sentences $a$ through $j$ with that entailed by Kripke’s theory of Truth, a milestone in the development of the philosophy of truth. Prior to Kripke’s [1975] paper the dependency of many paradoxes on contingencies was not widely or perhaps even properly appreciated, after his paper it should not be ignored. In this respect, Kripke’s work is a watershed in the development of our understanding of the Liar-like paradoxes. I have made reference and use of this point of Kripke’s already. In this section though, my objective is to make a comparison with his theory. The comparison is generally favourable; there are differences though, and there is one open issue that I will highlight towards the end. First, let me take a step back and begin with the intuition behind Kripke’s theory, then give an introduction to Kripke’s theory, and finally draw comparisons.

There are more conservative intuitions about the use of the truth-predicate than the general T-schema, such as restricting truth values to grounded sentences. In Kripke’s theory, ungrounded sentences are not always allocated a truth value.
Chapter 3

Relatives of the Liar Paradox

Sentences not allocated a truth value are said to belong in a *truth-value gap*, (which has no complement or other functions ranging over it, but has those functions introduced only in the philosophical meta-language). Kripke introduces the concept of *groundedness* informally in a *top-down way* (*without the presumption of bivalence*). Loosely speaking, if one can decompose a sentence into component sentences, and decompose any sentences referred to into their component sentences until one reaches atomic components that are sufficient to settle the truth value of the original sentence, the original sentence is *grounded*. He attributes the formulation of the concept to Herzberger [1970], who gave it this name, and Kripke notes that similar ideas have occurred previously.

Kripke’s formal definition for *grounded sentence* is from the *bottom-up*, as it were, for a given model.9 Kripke points out that we can learn to use the truth-predicate in a paradox-free way if we just classify as true or false those sentences that are or are not the case based on facts. We then proceed iteratively to attribute truth or falsity to sentences containing the truth-predicate that we were prepared to assert or deny at the previous step. Such inferences are safe. He also points out that we can safely quantify over those sentences we are prepared to assert or deny in this process, and attribute truth or falsity to those. Kripke gives a formal inductive definition of the sentences that can be determined true or not true in this way. These are the grounded sentences, the truth or falsity of which ultimately depends on the truth or falsity of basic sentences that did not contain the truth-predicate. Grounded sentences are assured of a truth value. Among the ungrounded sentences are paradoxical sentences and Truth-tellers.

Kripke’s induction starts with a T-free model, $M_0$, which is itself a standard model, but is then extended with a truth-predicate, which has a partial interpretation captured in an extension, $X_1$, and anti-extension, $A_1$, for truth. The extended model is the *base* model, $M_1$. If Cerberus is in the extension of ‘barks a lot’ in $M_0$, then ‘Cerberus barks a lot’ is in $X_1$. For anything, $\alpha$, in the domain that is not in the extension of ‘barks a lot’ in $M_0$, ‘$\alpha$ barks a lot’ is added to truth’s anti-extension, $A_1$. (Indeed, all non-sentences are also added into $A_1$ at this stage.) The extension and anti-extension of truth are further populated by an induction. At the first step every sentence made true by $M_1$ is added into $X_2$, and every sentence made false by $M_1$ is added into $A_2$. Say, for example, “‘Cerberus barks a lot’” is now in $X_2$, and “‘Cerberus

---

9 Kremmer [1988, p. 227] makes this point, contrasting the direction of an intuitive explanation with that of the formal definition using the terms ‘downwards’ and ‘upwards’.
does not bark a lot'” is now in $A_2$. $M_2$ makes ‘‘Cerberus barks a lot is true’” true, because “‘Cerberus barks a lot’” is in $X_2$. At the next step, the sentences made true or false by $M_2$ are added to $X_3$ or $A_3$ respectively. At each step of the induction, sentences made true or false by the current model are added to the extension of the truth predicate or its anti-extension respectively.

Which sentences are made true by a model depends on the valuation method used. The T-free atomic sentences are settled by the extensions of the predicates in the T-free model. Then a method for valuing sentences that are truth-functional or quantificational compounds of atomic sentences made true by the model is wanted, such as Kleene’s strong three-valued logic or van Fraassen’s supervaluation technique. Those complex sentences given a valuation by this method are also included in the extension and anti-extension of truth at each step.

The results are monotonic. At each step we add, to the extension and anti-extension of truth all those sentences that have now become the case or not in the interpretation at the previous step. For example, ‘Socrates is president’ was determined to be in the anti-extension of truth in the initial step; so ‘‘Socrates is president’ is not true” is determined to be in the extension of truth at the next step, (and ‘‘Socrates is president’ is true’ is likewise determined to be in the anti-extension). When a sentence is added to the extension or anti-extension of truth, it remains in the extension or anti-extension and never switches between these. The induction continues ad infinitum … and beyond. For, Kripke extends the induction into the transfinite, the details of which do not concern us, but it leads to Kripke’s important result that eventually there are no more sentences to add, having reached a construction known as a fixed point.

If ‘$\sigma$’ represents the function taking one step to the next in the monotonic construction of the extension and anti-extension of truth, then at a fixed point:

$$\sigma(X_n, A_n) = (X_{n+1}, A_{n+1})$$

That is, $X_n = X_{n+1}$ and $A_n = A_{n+1}$.

This definition of truth results in an extension and anti-extension for a truth predicate expressible in the object-language.

The basic construction described above defines ‘grounded sentence’. In this basic construction (the lowest fixed point), Truth-tellers and paradoxical sentences will be among the ungrounded sentences in the truth-value gap. So far no part of the construction admits Liars or Truth-tellers. However, as I will discuss, fixed points can be extended with Truth-tellers. The T-schema is not generally valid at any fixed point.
Nevertheless, in line with the Substitution Thesis mentioned in Chapter 2, any \( T(\alpha) \)-sentence, made true or false in the lowest fixed point has the same truth value as \( \alpha \). Indeed, Kripke [1975, p. 80] makes a general claim for the fixed points constructed monotonically in this way (but allows for other approaches). In any case, fundamentally, such fixed points correspond to consistent extensions for the truth-predicate under such valuation schemes.

There is a comparison between paradoxical and hypodoxical sentences not being in the lowest fixed point, and the indeterminacy one finds when naively attempting to evaluate them compositionally. The composition of one of these sentences is such (under circumstances when it is paradoxical or hypodoxical) that one never reaches a set of grounded sentences sufficient to determine its truth value. Nevertheless, Kripke’s theory allows for the use of a wide variety of valuation methods in reaching the lowest fixed point; so, in making such a comparison I would want to distinguish between compositional valuation methods, and methods that consider possible truth values. Kripke does not make such a distinction; nevertheless, on Kripke’s theory both paradoxical and hypodoxical sentences are indeterminate at the lowest fixed point, in the sense of being in the truth-value gap.

Among the ungrounded sentences (not in the lowest fixed point), Kripke [1975, p. 73] distinguishes paradoxical and hypodoxical sentences in the following way. Truth-tellers, like \( e \), are in the extension of truth in some fixed point and in the anti-extension of truth in another fixed point.

This suggests the following definition: a sentence is paradoxical if it has no truth value in any fixed point. ... [A Truth-teller] is ungrounded, but not paradoxical. This means that we could consistently use the predicate ‘true’ so as to give [it] a truth value.

Kripke [1975, p. 73]

Kripke does not specify how these fixed points that include hypodoxes are constructed, I make some assumption about this below. I assume they are included in a monotonic construction and that once a Truth teller is allocated in the construction to the extension or anti-extension of truth, it so remains, and is included in the resultant fixed point. Paradoxical sentences, like \( \alpha \), cannot be added consistently, that is, they cannot have a truth value in any fixed point [Kripke 1975, p. 73]. I interpret this to mean that if they were added to the extension, they would result in a sentence and its negation in the extension of truth [Kripke 1975, p. 76]. I will return to this point.
shortly, because I want to contrast it with an open issue about implementing this
definition using the monotonic constructions that Kripke has outlined. In any case,
Kripke distinguishes paradoxes and hypodoxes (of the Liar family) in this way:
relative to a given model, among the sentences ungrounded in the lowest fixed point,
the paradoxes are not in any higher fixed point while the hypodoxes are both in some
fixed point in the extension of truth and in some fixed point in the anti-extension of
truth.

Paradoxes are not over-determined on Kripke’s theory; they lack a truth value.
The Liar does not have a truth-value gap; it is in the truth-value gap. It is natural to
think of the truth-value gap as complementing the Range of Applicability (RA) of the
truth predicate.\textsuperscript{10} The truth value gap between the extension of the truth predicate and
its anti-extension can only be talked about in the meta-language; otherwise, a
Strengthened Liar paradox (in the sense of being over-determined) would be provable.

Depending on the valuation method, sentences like $i$ and $j$ may or may not be in
the truth-value gap in the lowest fixed point. Even so, sentences like $i$, that is, $(Ti \lor
\neg Ti)$, can be added to the extension of truth but not the anti-extension; whereas
sentences like $j$, that is, $(Tj \land \neg Tj)$, can be added to the anti-extension of truth but not
the extension. Kripke [1975/1984, p. 74] classifies these sentences under valuations
that do not include them in the lowest fixed point as intrinsic truths and intrinsic
falsehoods respectively.

If the valuation method uses van Fraassen’s supervaluations, then there are only
paradoxes and Truth-tellers remain in the truth-value gap, as contingent sentences are
among the grounded sentences, and any classical tautologies and contradictions will
receive a truth value as a result of supervaluation. Furthermore, in the lowest fixed
point, obtained using supervaluations from an unextended T-free model, my conjecture
about paradox and hypodox duality predicts that members of the Liar family of
paradoxes and hypodoxes in the truth-value gap can be put into one-to-one
 correspondence.

Returning to Liar paradoxes not being in any fixed point, I draw an analogy
between a fixed point and a theory, and use a theorem essentially related to one McGee
but for a modification to (3) and (4) below:

\textsuperscript{10} Cf. Martin and Woodruff [1975]
The Liar is not in a fixed point theorem: Consider \( \Gamma \), the set of sentences in the extension of truth at a fixed point, which

1. contains the identity, \( a = \langle \neg Ta \rangle \);
2. is closed under first order consequence;
3. contains \( \alpha \) whenever it contains \( T(\alpha) \); and
4. contains all instances of the schema
\[ \alpha \rightarrow T(\alpha). \]

Then \( \Gamma \) is inconsistent.

Proof:

1. \( \neg Ta \rightarrow \neg T(\neg Ta) \) SL, (1), =E
2. \( \neg Ta \rightarrow T(\neg Ta) \) An instance of (4)
3. \( Ta \) 1, 2, by (2),
4. \( \neg Ta \) 3, 1 =E, (3)

For a fixed point, if every sentence in the fixed point satisfies the T-schema, then assumptions (3) and (4) are satisfied for any sentence included in a fixed point; therefore no Liar-like paradoxical sentence can be consistently included in such a fixed point.

The matter is different for a hypodox. Intuitively, a sentence that is hypodoxical under the given circumstances could be consistently added to the extension or anti-extension of truth of a fixed point. In any case, the above argument will not rule a hypodoxical sentence out, because if it did the same argument would prove that the sentence was paradoxical, not hypodoxical.

A distinctive feature of Kripke’s outline is a monotonic construction [1975 / 1984, p. 68]. I think there is an open issue about how Kripke’s distinction between paradoxical and hypodoxical sentences can be implemented while preserving monotonicity. That is, if the set of fixed points for a given model is not assumed, but each needs to be able to be constructed as per Kripke’s outline, can a purely monotonic construction exclude all paradoxical sentences and assure that every sentence in the fixed point satisfies assumptions (3) and (4) above?

Certainly, if the Liar sentence or its negation is added to the extension of truth, a contradiction would result as a fixed point so constructed would satisfies all these assumptions. (A similar claim can be made for other identities of the forms \( b, c, d, \) and \( h, \) if the sentences they identify are paradoxical under the circumstances.) The base
model contains all identities for Liar sentences in the language, so assumption (1) is satisfied. The use of supervaluation in constructing the fixed point assures us assumption (2) is satisfied. If the Liar sentence or its negation were added, (3) would be satisfied; although one begins to wonder if (3) would be satisfied in any such case. Nevertheless, the function that takes an extension and anti-extension to the next level, \( \sigma(X_n, A_n) = (X_{n+1}, A_{n+1}) \), assures (4) is satisfied. Intuitively, assumption (4) aligns with T-introduction implemented in a monotonic construction.

Kripke allows for considerable variation in the way in which supervaluations are used, and does not specify the mechanism for adding Truth-tellers to the extension or anti-extension of truth. Let me sketch one possible procedure and investigate some of its implications for the classification of paradoxes and hypodoxes.

1. The extension (X) and anti-extension (A) of truth are initially ‘loaded’ with all the sentences made respectively true or false by the base model, \( M_0 \), which is a complete classical model that settles the truth values of all T-free sentences, and only those sentences. (All non-sentences are also ‘loaded’ into \( A_1 \) at this stage.)

2. Any additions to the extension or anti-extension may be attempted for sentences that have not already been allocated to X or A. The attempt is successful, provided a consistent evaluation can be given. If not, the sentence cannot be included in a fixed point.

3. Use a valuation method for the partial model, such as supervaluation. Let me use supervaluation as follows:
   Distribute all sentences that do not yet belong to either \( X_n \) or \( A_n \) to the extension and anti-extension, then, disregarding the anti-extension, evaluate all sentences using first-order semantics with identity for every such distribution.

4. Any sentence that is evaluated as true for each such distribution is added to \( X_{n+1} \).
   Any sentence that is evaluated as false under each distribution is added to \( A_{n+1} \).

5. \( X_{n+1} \) contains every sentence that is true at stage \( n+1 \), \( X_n \) is contained in \( X_{n+1} \), and similarly for \( A_{n+1} \).

6. Repeat steps 3 to 5 until a fixed point is reached.
There are alternate ways of using supervaluations at step 3 [Kripke 1975, p. 76]. In the above way, the supervaluations are *totally defined*, as Kripke says, the union of the extension and anti-extension of truth is the whole domain. Many such distributions will be inconsistent, in that way, being classical, every sentence in them will be both true and false, so such inconsistent valuations do not effect the result of the overall supervaluation, so long as there are consistent distributions. If ‘$\neg T\alpha$’ and ‘$T(\neg T\alpha)$’ are added to the extension in a distribution, then the latter is identical to ‘$T\alpha$’ and that distribution will be inconsistent. Nevertheless, even the Liar can be distributed consistently. If ‘$\neg T\alpha$’ is added to the extension of truth, and so is ‘$\neg T(\neg T\alpha)$’, the resulting distribution may be consistent, as there is no evaluation rule corresponding to the T-schema. If, however, the Liar is added at step 2 to either the extension or anti-extension, then inconsistency will emerge at step 4. This is because at step 4, ‘$\neg T(\neg T\alpha)$’ will be added to whichever of X or A already contains ‘$\neg T\alpha$’.

Consider first the Epimenidean Truth-teller, $f = \langle T\neg Q \lor Q \rangle$. If $\neg Q$ is made true by the base model, then $f$ will be in $X_1$ and the lowest fixed point (with the valuation methods I am considering). If $Q$ is made true by the base model, then $f$ will not be in the lowest fixed point. As a means of adding $f$, let me add it into $X_1$.

$$X_1 = \{ \langle Q \rangle, f, \langle P\alpha \rangle, \ldots \}$$
$$A_1 = \{ \langle \neg Q \rangle, \langle \neg (Tf \lor \neg Q) \rangle, \langle \neg P\alpha \rangle, \ldots \}$$

Distributions are considered which take $M_1$ to a classical model, $M_1'$, $M_1 \rightarrow M_1'$. $M_1'$ contains a truth predicate with an extension that is the union of $X_1$ and some sentences that are in the truth value gap, i.e. not members of $X_1$ and $A_1$.

$$M_1' \models Tf, \neg Q, T\langle Q \rangle, P\alpha, T\langle P\alpha \rangle, Tf \lor \neg Tf, Tf \lor Q, T(\neg Tf \lor \neg Q),$$
$$T(\neg (Tf \lor \neg Q)) \ldots$$

$Tf$ is forced by every distribution as $f$ was in the extension. Then $\langle Tf \rangle$ is added to the extension $X_2$, and $\langle \neg Tf \rangle$ to the anti-extension. (Notice that as $\langle \neg (Tf \lor \neg Q) \rangle$ is in $A_1$ it could not be in the extension of truth in any distribution, so $M_1'$ forces $\neg T(\neg (Tf \lor \neg Q))$, which goes into $X_2$, and, so its negation is added to $A_2$.)

$$X_2 = \{ \langle Q \rangle, f, Tf, \langle P\alpha \rangle, T\langle Q \rangle, T\langle P\alpha \rangle, \langle \neg T(\neg (Tf \lor \neg Q)) \rangle, \langle T(\neg (Tf \lor \neg Q)) \rangle, \ldots \}$$
$$A_2 = \{ \langle \neg Q \rangle, \langle \neg (Tf \lor \neg Q) \rangle, \neg Tf, \langle \neg P\alpha \rangle, \langle T(\neg (Tf \lor \neg Q)) \rangle, \ldots \}$$
One can discern a general pattern in the sentences added at subsequent levels, the
Eubulidean Truth-teller can be consistently added to the extension of truth, it will
remain in the extension, and be included in a fixed point.

Consider, however, if $T\langle Ta \rangle$ is added to the extension of truth, where $a = \langle \neg Ta \rangle$.
Then $T\langle T\langle Ta \rangle \rangle$ is forced by the model, and $a$ might take either value in a
supervaluation.

$$X_1 = \{ \langle T\langle Ta \rangle \rangle, \ldots \}$$
$$A_1 = \{ \langle \neg T\langle Ta \rangle \rangle, \ldots \}$$

As $a$ is neither in the extension nor anti-extension, it can be distributed to either; so
neither $Ta$ nor $\neg Ta$ is ever added to the extension or anti-extension.
1. If $a$ is distributed to the extension, no contradiction need occur:

$$M_1' \models T\langle T\langle Ta \rangle \rangle, \neg T\langle \neg T\langle Ta \rangle \rangle, Ta, T\langle \neg Ta \rangle, \text{etc.}$$

2. If $Ta$ is added to the extension no contradiction need occur:

$$M_1' \models T\langle T\langle Ta \rangle \rangle, \neg T\langle \neg T\langle Ta \rangle \rangle, T\langle Ta \rangle, \neg T\langle \neg Ta \rangle$$

The extension and anti-extension increment monotonically; but no contradiction is
foreseeable.

$$X_2 = \{ \langle T\langle Ta \rangle \rangle, \langle T\langle T\langle Ta \rangle \rangle \rangle, \langle \neg T\langle \neg T\langle Ta \rangle \rangle \rangle, \langle Ta \lor \neg Ta \rangle, \text{etc.} \}$$
$$A_2 = \{ \langle \neg T\langle Ta \rangle \rangle, \langle \neg T\langle T\langle Ta \rangle \rangle \rangle, \langle T\langle \neg T\langle Ta \rangle \rangle \rangle, \text{etc.} \}$$

Similar results obtain by adding $T\langle Th \rangle$ to the extension of truth in the
circumstance where $Q$ is not true. Intuitively, this is paradoxical, and one cannot
consistently add $h$ or $\neg Th$ under these circumstances, but the procedure that I have
outlined does not result in inconsistency if $T\langle Th \rangle$ is added.

Such a monotonic construction results in a consistent set of sentences that is
closed under classical consequence but assumption (3) above is not satisfied. This set
satisfies the definition of a fixed point, relative to this construction, but does not satisfy
the Substitution Thesis for every sentence. It is quite likely I have missed a relevant
feature in Kripke’s outline, but, as I say, this is a monotonic construction that results in
a consistent fixed point. It contains a sentence that is intuitively paradoxical, and yet it
appears it would be theoretically classified as hypodoxical in this way.

Another valuation method implementing Kripke’s theory might correct these
counter-intuitive results. In which case, these examples would discriminate between
valuation methods. Alternatively, another procedure implementing Kripke’s theory
might correct these results. These results pose an open question about how best to carry out Kripke’s suggestion for discriminating between paradoxes and hypodoxes while preserving the monotonic character of the definition of truth.

I can now compare my classification with that of Kripke’s theory. Given the identities of our tables, sentence $a$ is paradoxical, but whether $b$, $c$, $d$ and $h$ are paradoxical depends on the semantic value of $Q$. Likewise, sentence $e$ is a Truth-teller, but whether $d$, $g$, $f$, and $h$ can consistently take any truth value depends on the semantic value of $Q$. If $Q$ is a contingent sentence, it is an empirical matter whether sentences $b$, $c$, $f$, and $g$ are grounded or ungrounded. Whether $Q$ or $\neg Q$, sentences $d$ and $h$ are always ungrounded in the lowest fixed point, it is a contingent matter whether they are paradoxical sentences or Truth-tellers.

Kripke’s account is material. It was Kripke’s point that it is often an empirical matter whether a sentence is paradoxical or not. As Kripke put it:

...many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavourable, to exhibit paradoxical features.

[Kripke 1975/1984, p. 54]

Kripke’s theory is relative to a particular T-free model which settles whether $Q$ or $\neg Q$. Presumably this favoured T-free model is intended to represent how things actually are. So, given a contingent sentence $Q$, $b$ is not contingently paradoxical, $b$ is either paradoxical or grounded (and false). Although it is a contingent matter whether $b$ is paradoxical, $b$ is either paradoxical or it is not, we just do not know which until we know the facts.

In giving my tabulated classification, I have been holding the interpretation of ‘$Q$’ fixed and varying models; one could vary the interpretation of $Q$ to find examples of the different possible valuations that my classification refers to. Nevertheless, these would not be the same sentences – obviously, as the interpretation of ‘$Q$’ has been varied. So, Kripke’s theory does not make discriminations between $b$, $c$, $d$, $f$, and $g$ as variations of Liar-like paradoxes and hypodoxes. If under the circumstances they are paradoxical, they are just versions of the Liar paradox; and there is no reason in Kripke’s semantics to distinguish their valuations from each other.

With reference to the valuations in my tables, all my ‘under-determined’ valuations are in Kripke’s theory ‘determined in higher fixed points only’ and my ‘over-determined’ valuations are now cases where the sentence is in a truth-value gap.
(Of these sentences, the only determinations that can be made in the object-language will be when a sentence has a determinate truth value for a particular fixed point. If truth-tellers included in a fixed point could be said in the object-language to be true, then this suggests that a maximal fixed point should be the one giving the extension and anti-extension for the language; but there are multiple maximal fixed points for a language.\textsuperscript{11} In any case, Truth-tellers cannot be said in the object-language to be hypodoxical, because that would require expressing the thought that they are neither true nor false for some fixed point – a thought that cannot be expressed in the object-language [Cf. Kripke 1975, p. 80].)

Thus, the supervaluation method with Kripke's semantics seems to yield a similar classification to our intuitive results. Although Kripke's semantics is only material and does not discriminate between sentences \(a\) to \(h\) as \textit{variations}, one can, by varying the sentence that \(Q\) represents for each pair of paradox and its hypodoxical dual, find instances of the different categories of paradox and hypodox \(a\) through \(h\), and even \(k\) through \(o\). In this way, Kripke's theory would support paradox and hypodox duality, at least for this subset of such sentences, were these variations discerned by the theory. However, if the open issue I raised cannot be resolved in line with intuitive results, then intuitive paradoxes will be classified by Kripke's theory as hypodoxical, and duality is out of the question.

\textbf{3.3.2 Gupta and Belnap's Revision theory of truth}

This subsection compares my classification with that entailed by Gupta and Belnap's revision theory. I begin with an introduction to Gupta and Belnap's [1993] circular conception of truth; and then draw comparisons between their theory-relative classification and the classification of self-referential sentences that I have given. I will also consider whether their theory supports the dual conjecture. There is little to say about paradoxical \textit{arguments}, except for how the theory avoids them.

Gupta and Belnap think of truth as a circular concept. They motivate their approach by analogy with circular definitions:

\textsuperscript{11} This would not be a conundrum were there an independent reason to favour the minimal fixed point, perhaps in line with the supervenience conception of truth. Krenner [1988, pp. 240-242] contrasts the \textit{supervenience} conception of truth, which favours the smallest fixed point, with the \textit{fixed point conception of truth}, which does not discriminate among fixed points as suitable interpretations for a truth predicate.
The logics of circular definitions and of truth, in our view, illuminate and support one another.

[Gupta & Belnap 1993, p. 117].

Each T-biconditional provides a partial definition of truth, and is interpreted in a special way that supports iterative revisions of the extension of truth.

A circular definition, though it may not determine the extension of the definiendum, does provide a rule that can be used to calculate what the extension should be once we make a hypothesis concerning the extension of the definiendum. ... a circular definition ... does not determine the conditions of applicability absolutely, but only hypothetically. We cannot pick a set and say that it is the extension of the definiendum. We can say only that it should be the extension if such-and-such other set is supposed to be the extension.

[Gupta & Belnap 1993, p. 119].

Gupta and Belnap interpret the T-schema as a circular definition of truth. They use a specific interpretation of the biconditional. They find that the T-biconditionals, as partial definitions, are analytically true. They represent an intensional equivalence, not necessarily an extensional one. The biconditionals in these sentences are definition biconditionals [p. 138]. Thus, Gupta and Belnap [1993, p. 138, & p. 254] support their Definition Thesis. They contrast it with the Substitution Thesis, which I introduced in Chapter 2. Gupta and Belnap [1993, p. 138] contrast the following equivalences, where $s_i = \langle A_i \rangle$:

- $s_i$ is true $\equiv_D A_i$
- $s_i$ is true $\leftrightarrow A_i$

The former represents their definitional equivalence, the latter the conventional T-schema. (I note that both of these represent what I would call a general, as opposed to a canonical T-schema.) In virtue of acting as a definition, the revision T-schema is asymmetrical. Subscripts, denoting revision levels, can be added on either side of a corresponding equivalence, such that the left-hand side is one revision level above the right-hand side. This undermines the Substitution Thesis. Presumably Principle T, which was used to motivate the Equivalence Thesis in Chapter 2, would need to be modified to accommodate an asymmetry based on the definitional use of truth. The derivation of the Liar is unsound because it uses a T-biconditional with the same
revision level for both components. This T-biconditional is not true. Were the argument put forward using the definitional biconditional with correct revision levels, it would be invalid [Gupta and Belnap 1993, p. 254].

Let me outline how the theory works and give an example. In the revision theory of truth, one adds to a T-free model, $M_0$, an initial extension for the truth-predicate — an empty set, a guess, our best theory, or every sentence. This forms a base model, $M_1$. In this theory, the base model is a standard and complete model. The truth predicate has an extension in the model, not an anti-extension. All sentences (indeed, everything) that is not in the extension of truth in the model are not true (at that revision level). One iteratively revises the extension of truth using a variation on the T-schema that distinguishes revision levels.

Revisions are not monotonic. Indeed, every sentence receives a truth value at every revision, and some flip-flop between values. The extension that results from each revision completely replaces the previous extension for the truth predicate; so, that at each step, the union of the new extension and the T-free model is taken as an argument, as it were, for the next revision of the extension of truth. Any paradoxical consequences are avoided in this way, by distinguishing what is true at each revision.

A sentence that becomes and stays true through a sequence of revisions is said to be ‘stably true’ for that initial guess. Likewise sentences that become and remain false are ‘stably false’. All sentences in the T-free model will be stably true. If the Truth-teller is in the initial guess, it will be stably true; but will be stably false if it is not in the initial guess. Still others may be neither stably true nor stably false, oscillating between being in and out of the extension of truth. These are paradoxical. It is those whose truth values stabilise to the same value for every initial guess that are true or false. Truth-tellers will have a stable value for any initial guess, but a different one depending whether they’re included as true in the initial guess or not. Paradoxical sentences will flip-flop, oscillating in and out of the extension of truth.

There is an intuitive analogy between Kripke’s theory and theories like Gupta and Belnap’s, or Herzberger’s. The analogue of the T-schema in Kripke’s theory is a fixed point. Gupta and Belnap’s theory will not reach a fixed point if it contains paradoxes; but it may settle down to a pattern which loops with some periodicity, what Herzberger [1982, p150] has described as a ‘grand loop’. Such a loop is the analogue of a fixed point in the theory, perhaps it even bears some analogy to a generalization of which a fixed point is a special case.
Let me exemplify revision with a simplified example, in the T-free model, $M_0$.

Socrates is among the things that are mortal. The T-free model also includes our favourite identities, in particular our Truth-teller, $e$, and our Liar, $a$. However, say that "Socrates is not mortal", the truth teller and the liar are in our initial guess for the set of true sentences. In performing the first revision, we apply the following T-biconditionals (wherein ‘(Socrates is mortal)$_{\text{level}}$’ says that the sentence in parentheses has truth value $T$ at that level):

\begin{align*}
(T1) \quad & (T\langle \text{Socrates is mortal} \rangle)_{\text{level} + 1} \iff (\text{Socrates is mortal})_{\text{level}}. \\
(T2) \quad & (T\langle T\langle \text{Socrates is not mortal} \rangle \rangle)_{\text{level} + 1} \iff \ \\
& (T\langle \text{Socrates is not mortal} \rangle)_{\text{level}}. \\
(T3) \quad & (T\langle e \rangle)_{\text{level} + 1} \iff (e)_{\text{level}} \\
(T4) \quad & (T\langle a \rangle)_{\text{level} + 1} \iff (a)_{\text{level}} \\
(T5) \quad & (e)_{\text{level} + 1} \iff (e)_{\text{level}} \\
(T6) \quad & (a)_{\text{level} + 1} \iff (\neg a)_{\text{level}} \\
\end{align*}

Note that the revision levels, represented by subscripting, are in the semantics, not the syntax of the language. ‘$T$’ represents the truth predicate in the language. And angle brackets are a name-forming functor in the language. (T5) and (T6) arise from using all biconditionals of the general T-schema. (T6) is not contradictory because of the subscripting. It might be clearer to say that given T(7) and T(8) and related identities (T5) and (T6) are derivable:

\begin{align*}
(T7) \quad & (T\langle e \rangle)_{\text{level} + 1} \iff (e)_{\text{level}} \\
(T8) \quad & (T\langle \neg a \rangle)_{\text{level} + 1} \iff (\neg a)_{\text{level}} \\
\end{align*}

(T1) is applied with respect to the interpretation of ‘is mortal’, and consequently ‘Socrates is mortal’ is added to the extension of ‘is true’ in the model after the first revision, $M_2$. (T2) is applied with respect to the interpretation of ‘is true’ in the base model with the result that ‘‘Socrates is not mortal’ is true’ is also placed in the extension of ‘is true’ in $M_2$. Revising in this way had the effect of removing a sentence from the extension of truth – something that could not happen using Kripke’s induction. The initial extension of ‘is true’ (the initial guess) in the base model, $M_1$, was:

Socrates is not mortal

$e$

$a$
After the first revision, the extension of ‘is true’ in the model, M₂, is:

\[
\begin{align*}
&\text{Socrates is mortal} \\
&T\langle\text{Socrates is not mortal}\rangle \\
&T\neg T e \\
&T\langle T e \rangle \\
&T\neg T a \\
&T\langle T a \rangle \\
&\ldots
\end{align*}
\]

‘Socrates is not mortal’ has not remained in the revised extension of truth. It has had a ripple effect, and will continue to ripple through the revisions. ‘Socrates is mortal’ has been added based on the T-free model, and will continue to be so included in every subsequent revision. Te remains included. It was in the initial guess, and so application of T(5) preserves its truth at each subsequent revision. If the initial guess had not contained e, it would not have been added by any subsequent revision, and would have remained false throughout the revisions. The Liar, however, will alternate in and out of the extension of truth. If the Liar is in the initial guess, as above, it will be out of the extension in M₂ as above (and back in the extension in M₃, etc.). If the Liar were not in the initial guess, it would be in the extension of truth after the first revision in M₂.

I will now discuss a comparison between the naive classification of sentences a through o and their classification under Gupta and Belnap’s theory. Gupta and Belnap’s proposed intuitive classification of Liar sentences is extremely generic and simple. There are Contingent Liars like the Epimenides and Simple Liars like my favourite sentence [Gupta and Belnap 1993, p6].\(^{12}\) A Simple Liar ...

is intrinsically paradoxical in the sense that it is paradoxical no matter what the contingent facts may be. It is paradoxical in all possible worlds. But the Epimenides version is not intrinsically paradoxical. It is paradoxical if facts are one way (all other Cretan utterances are false, etc.), but not paradoxical otherwise. We shall call sentences that exhibit this behaviour Contingent Liars. Self-referential sentences such as The

---

\(^{12}\) As Gupta and Belnap note the alternative between the Epimenides sentence being contingent or paradoxical has been discussed by some authors; see Prior [1958] and Kripke [1975].
Liar [= ‘The Liar is not true’] above we shall call *Simple Liars*. All
Simple Liars are intrinsically paradoxical’. ...[A Simple Liar] is a
limiting case of a Contingent Liar’

[Gupta and Belnap 1993, p. 6 ...p. 9].

However, ‘Contingent Liars’ are not actually distinguished by the theory.
Whether some sentences, like my favourite conjunction, $b$, are paradoxical will depend
on the T-free model, and to that extent at least some sentences are ‘Contingent Liars’.
(Indeed, some are ‘Contingent Truth-tellers’.) However, the theory is concerned with
patterns that emerge from varying the initial guess, not the T-free model. Like Kripke’s
theory, there is just one T-free model. Varying the initial guess will not discriminate
between the Liar and ‘Contingent Liars’. Assume my favourite disjunction, $c$, and my
favourite sentence, $a$, are in the domain of $M_0$, and their identities, $c = \langle \neg Tc \vee Q \rangle$ and $a
= \langle \neg Ta \rangle$ are in the model $M_0$. Then, if $\neg Q$ is in the T-free model, then my favourite
disjunction, $c$, and the Liar, $a$, are both paradoxical, whatever the initial guess may be.
Neither is distinguished in this way as ‘intrinsic’ or ‘contingent’. \(^{13}\) Nevertheless, the
correct valuations will be reached. So, the revision theory can give the corresponding
valuations to sentences $a$ to $o$. However, it will give only the valuation from one
column of my tables, depending on the semantic value of $Q$ and $P$ in the T-free model.

Because the revision process is non-monotonic, unlike the monotonic
construction outlined in Kripke’s theory, there is no issue arising from having $T(Th)$ in
the initial guess. If $\neg Q$ and the identity $h = \langle Th \leftrightarrow Q \rangle$ are in the T-free model, and the
sentence $h$ is in the domain, $T(Th)$ will become false, and then oscillate in and out of
the extension of truth, yielding a typically paradoxical pattern.

The patterns that emerge through a sequence of revisions for various initial
guesses are summarised by Vann McGee [1991, p. 129] as in the table below, where $\lambda$
is a Liar, and $\tau_1$ and $\tau_2$ are Truth-tellers.

\(^{13}\) Gupta and Belnap [1993 pp.136-7] evaluate ‘Something said by Jones is not true’ against a specific T-
free model, and only vary the initial guess. In the circumstances they specify it is paradoxical and its
valuation oscillates; yet they do not remark on how if they varied the T-free model it would be stably
false. They, of course, know this; but it is not in the scope of their project to provide a theory that
varies the T-free model. However, such a theory is required to distinguish what they call ‘Contingent
Liars’.
<table>
<thead>
<tr>
<th>McGee’s classification</th>
<th>Stably true for ... initial guesses</th>
<th>Stably false for ... initial guesses</th>
<th>Oscillating between valuations for ... initial guesses</th>
<th>Example</th>
<th>Gupta &amp; Belnap’s description</th>
<th>Common name</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>No</td>
<td>No</td>
<td>Stably true</td>
<td></td>
<td>Contingently or Necessarily True</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>All</td>
<td>No</td>
<td>Stably false</td>
<td></td>
<td>Contingently or Necessarily False</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>All</td>
<td>Paradoxical</td>
<td>Liar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some</td>
<td>Some</td>
<td>No</td>
<td>Weakly stable</td>
<td>Truth-teller</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Some</td>
<td>Some</td>
<td>$\tau_1 \land \lambda$</td>
<td>(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some</td>
<td>No</td>
<td>Some</td>
<td>$\tau_1 \lor \lambda$</td>
<td>(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some</td>
<td>Some</td>
<td>Some</td>
<td>$\tau_1 \land (\tau_2 \lor \lambda)$</td>
<td>(3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As mentioned, the Liar does not have a stable truth value for any initial guess. The Truth-teller is stably true when included in the initial guess, and stably false where it is not. The classification of the last three rows of the table above is curious. The valuations of (1), (2) and (3) in the table above partially align with our naive value for these expressions. The way in which the table valuations are reached depends on the valuation of $\tau_1$. (1) is stably false iff $\tau_1$ is (not in the initial guess and thus) stably false. Otherwise, (1) is neither stably true nor stably false. (2) is stably true iff $\tau_1$ is (in the initial guess and thus) stably true. Otherwise, (2) is neither stably true nor stably false. (3) is stably false under similar conditions as (1); but it may be stably true in a way similar to (2) as well. If $\tau_1$ is in the initial guess and $\tau_2$ is not in the initial guess, then (3) is neither stably true nor stably false.
The valuation of (3) curiously involves giving different values to two truth-tellers. There certainly are different Truth-teller sentences; but naively, by Principle T in chapter 2, each Truth-teller should take the same truth value as itself. Not only do we have no reason to decide the truth value of a Truth-teller, we have no reason to decide whether or not different Truth-tellers might take different truth values. Allowing this, has surprising consequences. One might think that if a Liar is paradoxical, then the conjunction of two such paradoxical Liars is paradoxical. However, using revisions and admitting that two simple Liars could take different initial values, their disjunction could sometimes be stably true while their conjunction is sometimes stably false.

The above classification does not support the Dual conjecture; or, more to the point, the revision theory does not support the Dual conjecture. A sentence is paradoxical for a particular initial guess; but one must examine all initial guesses to find out if a sentence is a Truth-teller. In other words, according to the revision theory, a sentence either is a Truth-teller or it is not; but a sentence can be paradoxical for some initial guesses but not for others.

Once again, (1), (2) and (3) are not ‘Contingent Liars’. Whether they are paradoxical does not vary with the contingent value of $\tau_1$ because $\tau_1$ is not contingent – it is a Truth-teller. I have discussed examples like (1), (2) and (3) briefly in the section on compounds for truth-functional variations. Naive intuitions are grey in this area. Because of the way I have implemented naive valuations using a particular variation of supervaluations, the results are dissimilar to the valuations given by Gupta & Belnap. In contrast to a conjunction like $j$, i.e. $Tj \& \sim Tj$, the conjunction of a Truth-teller and a Liar is over-determined, for any row in its supvaluation table will be inconsistent. Unlike a disjunction like $i$, i.e. $Ti \lor \sim Ti$, the disjunction of a Truth-teller and a Liar would be over-determined, because all rows in its supvaluation table contain an inconsistency. Given that the two Truth-tellers in (3) take the same truth value, then (3) is just a Truth-teller. In any case, my main point is that these compound cases do not involve contingency – neither Gupta & Belnap nor McGee claimed they did – however, Gupta & Belnap set out to provide a theory of our intuitive classification and

---

14 In contrast, Sorensen [2001] expresses consternation in connection with his ‘No no’ paradox that two equiform tokens of a sentence should take different values where he perceives a lack of any reason for them to do so.
gave that classification in terms of 'Contingent Liars', but these can only be
distinguished if the base model is varied.

3.4 The quantificational 'self-referential' Variations

Resuming my classification with quantified variations, the Epimenides typically
involves universal quantification, and a similar paradox that Prior attributes to Geach
typically involves existential quantification. Nevertheless, these paradoxes are
ultimately distinguished by their set of possible valuations. These are each related to
the Unquantified Epimenides and Self-referential Curry paradox respectively.

3.4.1 Distinguishing the Epimenides

Distinctive of the historical Epimenides is a Cretan saying something about Cretan
sayings. Thus, I will argue, the following sentence may be used to distinguish the
(quantified) Epimenides paradox.

\[ \exists \forall x (\exists x \supset \neg Tx) \]

If something represented by this expression holds, then a derivation may assume
the embedded universal,

\[ \forall x (\exists x \supset \neg Tx) \]

instantiate with the universal sentence replacing \( x \) by the sentence itself, and use the
premise to apply MP to obtain the following expression under the assumption:

\[ \neg T(\forall x (\exists x \supset \neg Tx)) \]

This sentence, (3), refers to the sentence (2) from which it is derived. Truth
inferences can then be used to obtain a contradiction from the assumption. Therefore,
(2) is not the case.

Self-application is not sufficient for this reasoning. (2) is in the range of
applicability of the open sentence (4) below.

\[ \exists x \supset \neg Tx \]

So, it is fair to say that (4) is applied to itself in a way when (2) is instantiated to:

\[ \exists (\forall x (\exists x \supset \neg Tx)) \supset \neg T(\forall x (\exists x \supset \neg Tx)) \]
However, the premise (1) is needed for *modus ponens*. The open sentence (4) applies to anything in the domain (truly or falsely) and is not sufficient to derive the diagonalization, (3).

A premise of form (1) is necessary for the Epimenides paradox. The premise provides that (2) is a sentence that satisfies its own condition. The Epimenides' premise refers to a quantified sentence as being in the scope of some predicate _domains, and that quantified sentence says something about sentences in the extension of the predicate D. (An explanation in terms of plural reference seems fair, although not forced, one might say that the quantified sentence uses that very same predicate to thereby refer to itself among others.) The distinctive premise (1) and common principles of logic are sufficient to derive (3).

On the one hand an analysis in terms of self-application is not sufficient to explain how paradox arises in this case; and on the other, an analysis of the required premise shows that it refers to a sentence that talks about the premise among other things. In either case, the analysis needs to be supplemented to explain why self-application or plural reference is problematic in this case.

Nevertheless, there is an analogy between the Liar sentence's use of self-reference and the Epimenides sentence's semantics. The identity

\[ a = \langle \neg Ta \rangle \]

makes evident that the Liar sentence contains a term that refers to the sentence in which it occurs. Analogously, (1)

\[ \exists \langle \forall x (\exists x \supset \neg Tx) \rangle \]

makes it evident that an Epimenides sentence contains a predicate which includes the sentence in which the predicate occurs in the predicate's extension. The analogy is closest in a limit case like (6) below, where _domains_ has been substituted with 's ='.

\[ s = \langle \forall x (s = x \supset \neg Tx) \rangle \]

(6) contains a term that refers to the sentence in which it occurs. The general form of (1) is obtained by abstracting away from the use of identity in (6) to other predicates that take a sentence as an argument and in particular this one by using the same predicate in the expression.
3.4.2 Deriving the Epimenides

In this section, I show the similarity between deriving the Epimenides and deriving the Liar paradox. At the heart of all variations of the (quantified) Epimenides paradox is the Church [1946] conundrum: from a Cretan saying that all Cretan sentences are false, it follows that some Cretan sentence is true. If circumstances are such that no other Cretan sentence is true, then a paradox obtains.

Let us adapt the usual Tarski-style example. Let D abbreviate the predicate ‘is a sentence with tokens printed in this section’. Then sentence (7) below is itself provably false. It itself falls under its own scope so if it is true, it is (at least) false:

\[ (7) \quad \forall x \ (Dx \supset \neg Tx) \]

Here is a proof:

1. \[ D(\forall x \ (Dx \supset \neg Tx)) \]  \hspace{1cm} \text{premise}
2. \[ \forall x \ (Dx \supset \neg Tx) \]  \hspace{1cm} \text{assumption}
3. \[ D(\forall x \ (Dx \supset \neg Tx)) \supset \neg T(\forall x \ (Dx \supset \neg Tx)) \] \hspace{1cm} 2 \forall E
4. \[ \neg T(\forall x \ (Dx \supset \neg Tx)) \] \hspace{1cm} 3, 1 MP
5. \[ \forall x \ (Dx \supset \neg Tx) \supset \neg T(\forall x \ (Dx \supset \neg Tx)) \] \hspace{1cm} 2-4 CP
6. \[ T(\forall x \ (Dx \supset \neg Tx)) \leftrightarrow \forall x \ (Dx \supset \neg Tx) \] \hspace{1cm} T-biconditional
7. \[ \neg \forall x \ (Dx \supset \neg Tx) \] \hspace{1cm} 5, 6 SL
8. \[ \neg T(\forall x \ (Dx \supset \neg Tx)) \] \hspace{1cm} 7, 6 SL
9. \[ \exists x \ (Dx \& Tx) \] \hspace{1cm} 7, QN, defn, DeM

As per line 8, simply from sentence (7) being in this section we can prove there is some true sentence in this section. If every sentence in this section other than (7) were false, paradox would result.

In developing this derivation, I first reconstructed Prior’s [1958] axiomatic proof, and then shuffled steps around to clearly separate the derivation of the semi-diagonalization at line 5 from the subsequent use of the T-schema. This more closely parallels the proof the Liar I gave using the canonical T-schema, which showed that the derivation of a diagonalization can be separated from the subsequent use of the T-schemata. This separability of diagonalization from use of schemata will be of some significance in discussion in later chapters.

Here is what one might think is a counter-example to my claim that Church’s conundrum is characteristic of the Epimenides:
Chapter 3  Relatives of the Liar Paradox

(8) \( \forall x (a = x \supset \neg Tx) \)

The idea can be generalized by abstracting away from the predicate ‘\( a = \)’ to \( \varnothing \), provided the sentence is in the extension of \( \varnothing \), but that is why I characterised the Epimenides using (1), wherein the quantified statement is referred to. If (8) relies on the identity \( a = (\neg Ta) \), then the Liar follows, and can be used as a reductio of (8), proving the same sort of conclusion as above, (9):

(9) \( \exists x (a = x \& Tx) \),

from which a contradiction follows. Consider, by comparison, (10):

(10) \( \forall x ((10) = x \supset \neg Tx) \)

An identity statement for (10) has the form of (1), and from the identity of (10), we can obtain the result for Church’s conundrum. As (10) is the only sentence to satisfy ‘(10) =’, then a paradox follows. The difference between (8) and (10) is that (8) refers to a Liar sentence, whereas (10) depends on an identity that is of the same general form as (1).

3.4.3 Evaluating the Epimenides

The possible valuations of the Epimenides sentence are the same as those of the Unquantified variation, identified by \( b = (\neg T b \& Q) \). The unquantified variation is identified by an identity premise (or theorem), whereas Epimenides’ sentence is identified in a premise. If any other sentence said by a Cretan is true, then Epimenides’ sentence is false, and compositionally so. If no other sentence said by a Cretan were true, then Epimenides’ sentence is compositionally indeterminate. Supervaluation, under such circumstances, will result in an over-determined valuation. If it is assumed true, then it is false; and if it is false that assumption together with the circumstance that every other Cretan statement is false, will entail it is true. Then Epimenides’ sentence is false or over-valued, just like \( b \).

Sentence (8) from the previous section does not have all these possible valuations because it refers to a Liar sentence. Based on its valuation and the identity statement it uses, (8) is a form of the Liar.

Now, in a case like (10), (10) happens to refer to itself; and the nature of identity is such that it will be the only thing that satisfies the predicate ‘(10) =’, so it is over-determined, yet it uses a statement of the same general form as (1). It is an example of a sort of limit case of the Epimenides, which has the same evaluation as the Liar.
3.4.4 Relating the Epimenides to List Variations

When the $\mathcal{D}$-predicate has a finite extension (and it is given that it has at least one member),

\[(\forall x) (\mathcal{D} x \supset \neg T x)\]

is materially equivalent to:
\[\neg T(\text{first } \mathcal{D}-\text{sentence}) \& \neg T(\text{second } \mathcal{D}-\text{sentence}) \& \ldots \& \neg T(\text{last } \mathcal{D}-\text{sentence}).\]

So, instead of using a quantified sentence, each member of the extension of $\mathcal{D}$ can be referred to from each conjunct in a conjunction of $\neg T$-sentences.\(^{15}\) For example, consider the brief list:

\[(11) \quad \neg T(11) \& \neg T(12) \& \neg T(13)\]
\[(12) \quad 1 = 2.\]
\[(13) \quad \text{I am rich.}\]

(11) cannot be true; so its first conjunct is true. (12) is false; so (11)'s second conjunct is true. So the only way (11) can be false is for (13) to be true. However, if (13) is as a matter of fact false, then we have a contradiction (when the fact is included as a premise).

\(^{15}\) Prior's [1958] representation of the Epimenides argument uses substitutional quantification and an operator (he represents by '\(\&\)'), but no truth predicate. The truth-predicate can be added in easily enough, as I have done, and then the quantification appears to be objective. But I note the essential role played by $T$-sentences and particularly $T$-biconditionals in showing the relationship of Prior's Epimenides to a form more amenable to truth theories. I note, moreover, that the corresponding equivalence to that above, using a Priorian delta-operator instead of the $D$-predicate, would fail as a definition, for circularity. Interestingly, one has to invoke the $T$-schema using the $D$-predicate version to get from the version quantifying over propositions to an equivalent finite conjunction. These translations from one representation to another are fine provided access to $T$-biconditionals is unrestricted and unproblematic; but that is not usually given. That is, Prior's delta-operator representation is problematic. To get from the delta variation to the $D$-variation one must accept the $T$-biconditionals; but the $T$-biconditionals are problematic, so use of them must be justified, which is going to be hard for a theorist who eventually seeks to find fault with some of these very $T$-biconditionals. Should these truth theories advocate a purge of such $T$-biconditionals, the relationship is lost and the natural language rendition of Prior's proof becomes a potentially vengeful anachronism lurking with no clear link to other Liar-like paradoxes, let alone theories of truth.
If the extension of D is just (11), (12) and (13), then (11) is materially equivalent to a sentence with the form of (2); indeed, use of ⊢ can be replaced in the English generalization by its extension as in:

\[(11')\] (11'), (12) and (13) are all false.

Paradoxes like this have been known since Medieval times and can be wrapped up into the one sentence simply consisting of two or more conjuncts.

\[(14)\] ¬T(14) & 1 ≠ 2 & I am not rich.

And this is how sentences of the form of our original b are related to the Epimenides sentence. Each member of the previous list now appears as a negated conjunct, by substituting identicals and using truth inferences. In this way, an Epimenides-like sentence can be reduced to a finite conjunction, if the extension of the predicate represented by ⊢ is finite. The converse need not be the case; not all conjunctions are represented by such universal sentences.

So, the quantified Epimenides relates in this way to a conjunction, and in this way the unquantified Epimenides; both are about a group of sentences. Notice the relation to the Eubulidean Liar, which is a sort of trivial limit. If (6) is the only sentence to fall under its own scope, like (9), a contradiction can be proven; and, a is trivially a conjunction with only one conjunct.

Despite the neatness of this semi-formal relationship between the Epimenides and its unquantified material equivalent conjunction for D-predicates with finite extensions, there are another two relationships to consider.

The first alternate account is related to the self-application interpretation of the Epimenides. The instantiation of the quantified Epimenides yields a conjunction of disjunctions. So, for a finite domain of quantification (not just the extension of the D-predicate),

\[(2)\] (∀x) (Dx ⊃ ¬Tx)

is also materially equivalent to:

\[(¬D(2) \lor ¬D(2)) \land (¬D(1^\text{st} \text{ sentence}) \lor ¬D(1^\text{st} \text{ sentence})) \land (¬D(2^\text{nd} \text{ sentence}) \lor ¬D(2^\text{nd} \text{ sentence})) \land ... \land (¬D(1^\text{st} \text{ sentence}) \lor ¬D(1^\text{st} \text{ sentence})).\]

It is a contingent matter whether or not the universal sentence is in the scope of the D-predicate. In the case given, if (2) were true, the first conjunct would be false,
but then (2) would be false; so (2) must be false, which makes the first conjunct true. So, some other conjunct must be false. This can only be the case if some other sentence in the scope of the D-predicate is true.

The conjunction of disjunctions itself is materially equivalent to our original conjunction; so this relationship does not disturb the neat semi-formal relationship between the Epimenides and its unquantified relative, b. Furthermore, it brings out that the relationship relies on the reasoning that the sentence itself cannot be true, therefore something else must be true.

The second alternative account establishes the relationship relying on an equivalence. Goldstein [1986] relates the Epimenides to a single conjunction, such that (2) is to be rewritten as:

\[(15) \quad \sim T(15) \& \forall x (\exists x \& x \neq (15). \supset \sim Tx)\]

and then replaces the second conjunct with Q.

There are two issues with this account. One is that (2) is not equivalent to (15). It is not necessary that (2) is in the extension of the \(\exists\)-predicate. (This contingent fact was a premise in the proof above.) Whether this is a real issue depends on one’s interpretation of ‘rewritten as’.

The second issue is that replacing the second conjunct with any Q may preserve paradoxicality, but does not show that the two paradoxes are of the same type. This second issue can be somewhat overcome by letting ‘Q’ abbreviate the second conjunct. But even then this equivalence does not show how the relationship is naturally extended to the multiple conjunction version above, unless it be via some material equivalence. If that is being assumed, then this equivalence is materially equivalent to the list of conjunctions first given above; so, this account can be reduced to my account.

One correction can be made to Goldstein [1986, p. 119]. As I have mentioned, he does not distinguish the first conjunct of the unquantified Epimenides from the Eubulidean Liar. But the first component is not the Liar paradox, not in b anyway. Let me illustrate this. The corresponding version incorporating the Liar is:

\[(16) \quad \text{This conjunct is false} \& \forall x (Dx \& x \neq (15). \supset \sim Tx).\]

Here, the first conjunct is always paradoxical, whether \(\forall x (Dx \& x \neq (15). \supset \sim Tx)\) is the case or not. In (16), if the first conjunct is true, it is false; and if the first conjunct is false, then what it says is true. By contrast, the first conjunct of (15) can be
true. In this way, the Epimenides is a distinct variation from the Eubulidean Liar, and not merely a version formed by incorporating the Eubulidean Liar in a conjunction.

3.4.5 The Epimenides Hypodox

On the face of it there are two candidates for the dual hypodox of Epimenides paradox, ‘Cretans always tell the truth’ and ‘Some Cretan sentence is true’. The former is an obvious contrary of Epimenides’ sentence, the latter is its external negation. Ideally, we should want a sentence that is hypodoxical just when Epimenides’ sentence is paradoxical. If all other Cretan sentences are false, then the former is just false, but the latter is hypodoxical. So, it is ‘Some Cretan sentence is true’ (or ‘Some Cretan tells a truth’) that I believe is the hypodox of the Epimenides.

I note that in a Kripkean theory, where ‘Some D-sentence is true’ is the only sentence in the extension of the D-predicate, that sentence is hypodoxical.

3.4.6 Geach’s Conundrum as a Quantified variation of Curry’s Paradox

Had Epimenides taken Peter Geach’s good counsel, he might have said ‘Some Cretan sentence is not true’, then it would follow that his sentence is true and some Cretan sentence is false, some other one on pain of paradox. As mentioned in Chapter 1, this variation is attributed to Peter Geach in Prior [1961, p. 18]. In this section, I want to show the form of sentences that individuates Geach’s variation, that it too relates to collections that can be represented with quantification or disjunctions, and thus, that it is a quantified variation of Curry’s paradox.

Like the Epimenides, it is not the quantified sentence that individuates this paradox, rather it is the fact that it is a member of the extension of some predicate. It has the following form:

\[ \mathcal{D}(\exists x(\mathcal{D}(x) \& \neg T(x))) \]

Take the case of ‘\( \exists x(\mathcal{D}(x) \& \neg T(x)) \)’ using the earlier interpretation of ‘\( \mathcal{D} \)’. Given that it is among the sentences with tokens printed in this section and thus in the extension of the D-predicate, the sentence,

\[ \exists x(\mathcal{D}(x) \& \neg T(x)) \]

---

16 The actual case given is ‘A Cretan says that something said by a Cretan is not the case’. This can be represented without the truth predicate using substitutional quantification.
is itself provably true. So there must be some other false sentence with a token printed in this section.

This is essentially a quantified form of the Truth-functional, Self-referential Curry paradox. One can extend the self-referential variation of Curry’s paradox to talk about a collection. So, that (19) below is equivalent to (20) or (23):

\[
\begin{align*}
(19) & \quad \text{Either (19) is not true or Q or P.} \\
(20) & \quad \text{Either (20) is not true or (21) is not true or (22) is not true.} \\
(21) & \quad \neg Q \\
(22) & \quad \neg P \\
(23) & \quad \text{At least one of (23), (22) and (21) is not true.}
\end{align*}
\]

Given a predicate, C, whose extension is just (24) below, and (21) and (22) above, then (23) is equivalent to

\[
(24) \quad \exists x (C(x) \& \neg T(x))
\]

Clearly, the list of disjuncts of a sentence like (20) can be extended in appropriate ways.\(^{17}\)

If a sentence says that some member of a collection is not true and is itself a member of the collection, Geach’s conundrum becomes paradoxical if no other member of the collection is true. The Eubulidean Liar is a trivial limit case. In cases of more than one-membered collections like this, Geach’s conundrum has the set of possible semantic values of \{true, over-determined\}, the same set as pertains to Curry’s paradox.

In summarizing this section, I note the following. The Epimenides, \(p\), takes the same set of possible valuations as the Unquantified Epimenides, \(b\), where \(Q\), in these cases, is:

\[
\neg \exists x(\Box x \& x \neq p \& T x)
\]

Similarly, Geach’s example,

\[
q = \Box (\exists x(\Box x \& \neg T x))
\]

takes the same set of possible valuations as the Self-referential Curry, \(c\), where \(Q\), in these cases, is:

\(^{17}\text{Cf. Sorensen [1998].}\)
\[ \neg \forall x ((\exists x \land x \neq q) \Rightarrow \neg Tx) \]

These quantificational variations relate to collections; however, they may be specifiable independently of the collection; i.e. whether or not the relevant quantified expression are themselves members of the collections they say something about may be a contingent matter. Their pathology is not always predicted by the Vicious Circle Principle. The collections that these sentences are about, if finite and individually specifiable, can be used to translated these expressions into conjunctions or disjunctions, which confirms their relationship with the unquantified variations of the Epimenides and Self-referential Curry paradoxes.

### 3.5 The purely-circular Variations of the Liar-like Paradoxes and Hypodoxes

In this section, I focus on circular forms of the Liar and truth teller, particularly the unquantified versions. I discuss Sorensen’s so-called “no-no” paradox. I briefly outline truth-functional extensions of the circular paradoxes and quantificational versions. I relate circular variations to other categories in terms of lists.

#### 3.5.1 Circular Liars and Truth-tellers

Remember that our naive position is that “This sentence is true” takes the same truth value as the sentence to which its indexical phrase refers. This position was encapsulated in Principle T, the principle, which was introduced in Chapter 2, that a sentence attributing truth to another takes the same truth value as the other. In the following list, (1) below takes the same truth value as (3).

\[
(1) \quad (2) \text{ is true.} \\
(2) \quad (3) \text{ is true.} \\
(3) \quad \text{Ariadne taught Theseus a dance.}
\]

If a list loops, as the following one does, there is a lack of a basis to say whether the sentences are true or false.

\[
(4) \quad (5) \text{ is true.} \\
(5) \quad (6) \text{ is true.} \\
(6) \quad (4) \text{ is true.}
\]
A self-referential sentence is trivially such a loopy list, a circle of one. So, if ‘This sentence is true’ refers directly to itself, it is a Truth-teller, as its truth value is underdetermined. Likewise, when two such sentences attribute truth to each other, such a loop is also a Truth-teller. Clearly, we can form circles of sentences each saying the next sentence is true, and each of these sentences will be a Truth-teller in the sense that its truth value is underdetermined. Its truth value is not independent of the truth values of the other sentences in the series, however. Since each takes the same truth value as the sentence to which it refers, they must all be true or all be false. For example, consider the circular sequence above. (4), (5) and (6) can all be true or all be false. They are Truth-tellers, and yet they must all have the same truth value, rather than independent truth values.

Remember also that according to naive principles ‘This sentence is not true’ does not take the same truth value as the sentence to which its indexical phrase refers, which was incorporated in Principle F, the principle, which was also introduced in Chapter 2, that a sentence attributing falsity to another does not takes the same truth value as the other. So that in the loop below, (7) takes the same truth value as (8), and (8) does not take the same truth value as (7). This is a circular Liar. The semantic value of each sentence in this loop is over-determined.

(7) Sentence (8) is true.
(8) Sentence (7) is not true.

There is also what Sorensen (2001, pp. 165-166) calls ‘the no-no paradox’:

(9) Sentence (10) is not true.
(10) Sentence (9) is not true.

Here, while we cannot consistently say (9) and (10) are both true or both false, we could arbitrarily say (9) is true and (10) is not true or vice versa! There are a number of logical principles appealed to. Sorensen says ‘the no-no paradox poses the further problem of assigning asymmetrical truth values to symmetrical sentences’ [2001, p. 167]. Sorensen seems to have in mind some parsimonious principle, like if there is no relevant contextual, semantic or syntactic difference, then sentences have the same truth value. Some such principle seems to suggest that (9) and (10) have the same truth value, in which case a contradiction is provable. Non-contradiction compels us to say one is true if the other is false but we cannot logically decide which is which.

Nevertheless, these sentences are hypodoxes, as their truth values are underdetermined. They have the typical Truth-teller characteristic that it is possible to
allocate truth values to these sentences but there is a lack of reason to decide which truth values. We cannot allocate them truth values independently of each other; but that was equally true of (4), (5) and (6). There was nothing "paradoxical" about that.

The supposed ‘further problem’ of having to assign asymmetrical truth values is, I take it, suggesting that this is unprincipled and unparsimonious; but Principle F is a relevant principle and explains why each should take a different truth value to the other. We are surprised when the Liar arises in this way; but that we have to do it here where it does not result in inconsistency does not cause enough surprise to warrant being called a paradox, at least in the sense I am using it throughout. (I am not using the same definition of ‘paradox’ as Sorensen [2003], who includes a wide variety of conundrums as ‘paradoxes’.) Our consternation at (naively) proving (9) and (10) have different truth values, when they appear to be so close to being tokens of the one sentence is explained by Principle F. (Furthermore, they cannot actually be tokens of the one sentence, provided indexicality is paraphrased out.) Even allowing equiform sentences using indexicals, there is an adequate explanation for their taking opposite truth values. Given two equiform sentences in a list, one after the other, each saying ‘The next sentence is not true’, then the first should take an opposite truth value to the second, in accordance with Principle F. Therefore, the no-no phenomenon is not a paradox.

The no-no phenomenon is restricted to circles with an even number of ‘is not true’ sentences. If we have an odd number of such sentences, the result is that each is Liar-paradoxical. (A circle of one such sentence is of course the Liar itself.) Consider:

(11) (12) is not true.
(12) (13) is not true.
(13) (11) is not true.

Now, if (11) is true, (12) is not true; so (13) is true, but then (11) is not true. Also if (11) is not true, (12) is true; so (13) is not true, but then (11) is true. And we have a contradiction. None of these sentences can be consistently allocated a truth value.

Any sentence that simply says a sentence is not true must take the opposite truth value to that sentence to which it refers. This principle explains why an even number of ‘is not true’ sentences in a referential circle can consistently take alternating truth values but no sequential pair can consistently take the same truth value. It also explains why an odd number of ‘is not true’ sentences cannot consistently take the same truth value or different truth values. (This also applies to the single self-referential case.)
The insertion of ‘is true’ sentences into such referential circles makes no difference to whether truth values can be consistently allocated or not. Thus, a Liar-paradoxical circle can be converted to a Truth-teller by negating one of the sentences and *vice versa*. (In addition, inserting an odd number of ‘is not true’ sentences will also have the net effect of toggling from a Liar-paradoxical circle of sentences to a Truth-teller circle of sentences and *vice versa*.)

Nevertheless, I should not want to give up easily the idea that a dual hypodox can be paired with each paradox, even if the use of external negation proves limited.

Consider what is involved in converting the list above to a conjunction, and then negating it:

(14) The 2nd conjunct is not true & the 1st conjunct is not true.

The candidate dual for the above using external negation is:

(15) The 2nd disjunct is true ∨ the 1st disjunct is true.

This, however is another hypodox. Alternatively, one could simply negate a conjunct to get a paradox:

(16) The 2nd conjunct is true & the 1st conjunct is not true.

Then we have a dual paradox, but there are two such alternatives. Here is the second:

(17) The 2nd conjunct is not true & the 1st conjunct is true.

Conversely, both of these have two hypodoxical duals obtained by negating a single conjunct.

So, external negation with relettering is not an unexceptional rule for producing a dual. Still the one-to-one conjecture seems appropriate. Given two sentences of this type, there are 4 possible combinations: 2 of which are paradoxical and 2 hypodoxical. For 3 sentences, there are 8 possibilities: 4 paradoxical and 4 hypodoxical; etc. There just are no firm guidelines as to which should be paired. In previous cases, contingencies were a guide. For we should want (or at least I want) the dual to be hypodoxical when the other is paradoxical.
3.5.2 *Circular Variations of the Epimenides, Curry and the New paradox*

What about versions that are more Epimenidean, or Curry-like, or even like the ESP?

Here is an example like the Self-referential Curry:

(18) Either sentence (19) is not true or Q.

(19) Sentence (18) is true.

Here are the possible cases, three of which are *reductios* and where ‘#’ represents a contradiction is derivable:

\[
\begin{align*}
T(18) \& T(19) & \vdash Q \\
T(18) \& \neg T(19) & \vdash \# \\
\neg T(18) \& T(19) & \vdash \# \\
\neg T(18) \& \neg T(19) & \vdash \#
\end{align*}
\]

We note that (18) and (19) must both be true and that therefore Q.

Here is an example like the Epimenides:

(20) Sentence (21) is not true and Q.

(21) Sentence (20) is true.

\[
\begin{align*}
T(20) \& T(21) & \vdash \# \\
T(20) \& \neg T(21) & \vdash \# \\
\neg T(20) \& T(21) & \vdash \# \\
\neg T(20) \& \neg T(21) & \vdash \neg Q
\end{align*}
\]

Note that (20) and (21) must both be false and that therefore $\neg Q$

Here is our new paradox in two sentences:

(22) Sentence (23) is not true iff Q.

(23) Sentence (22) is true.

\[
\begin{align*}
T(22) \& T(23) & \vdash \neg Q \\
T(22) \& \neg T(23) & \vdash \# \\
\neg T(22) \& T(23) & \vdash \# \\
\neg T(22) \& \neg T(23) & \vdash \neg Q
\end{align*}
\]

Note that (22) and (23) can both be true or both false, either way $\neg Q$ follows.

There are clearly other combinations, but this is sufficient to demonstrate the semantic values obtainable from two sentence forms include every possibility identified in the section on unquantified self-referential paradoxes.
3.5.3 Circular Liar-like Paradoxes involving Quantification

These combine circularity with quantificational variants; their classification is perhaps open-ended, but I want to acknowledge their existence. There are well-known quantified versions of these types. Consider Cohen's [1957] example involving the policeman who simply testifies nothing the accused says is true and the accused who only says something the policeman says is true. And then there is Kripke's memorable example about Nixon saying everything Jones says about Watergate is true, when Nixon's other Watergate-related sentences are evenly balanced between true ones and lies, and Jones only statement about Watergate was to say that the majority of Nixon's sentences about Watergate are false.

3.6 The Infinite Variations of the Liar-like Paradoxes and their Truth-tellers

In this section I explain the relationship between infinite Liar-like paradoxes and circular ones in terms of list structures.

3.6.1 Infinite Liars and Truth-tellers

Circular sequences of 'is true' and 'is not true' sentences are a type of lists of such sentences. Finite lists have some interesting properties. Here is a list variation on the self-referential Epimenides:

(1) All members of this list are false
(2) P
(3) Q

(1) can be disproved by reductio for familiar reasons; so some member of the list must be (at least) true. We would lose this effect if sentence (1) were not self-referential. The following list does not license the familiar reasoning:

(4) All subsequent members of this list are false
(5) P
(6) Q

However, we regain this effect with two sentences that claim all subsequent members are false.
(7) All subsequent members of this list are false.
(8) All subsequent members of this list are false.
(9) P
(10) Q

Now, (7) is non-self-referential but cannot be true for if it was, then all members of the list below (8) would be false, in which case (8) would be true, making (7) false. So (7) must be false, and therefore some subsequent sentence is true.

Consider the list:

(11) All subsequent members of this list are false.
(12) All subsequent members of this list are false.
(13) All subsequent members of this list are false.
(14) All subsequent members of this list are false.

The first sentence in such a list cannot be true, because if it was then all the rest are false but that would make the second sentence true. So, the first sentence must be false, but that means a subsequent sentence must be true. In this respect the argument is very similar to the Church version of the Epimenides. No member of such a list can be true if it has two such sentences following it. Furthermore, the last sentence is vacuously true, thus it is the only sentence in such a list that can be true.

An infinite list of such sentences has no last sentence and Yablo's paradox results.

Likewise, we can move from a list variation of the Geach quantified variation of Curry's paradox to Sorensen's Queue paradox. Here is a list variation on Geach's paradox:

(15) Some member of this list is false
(16) P
(17) Q

(15) is provable for by now familiar reasons; so some member of the list must be true. We also achieve this effect with two sentences in the following list:

(18) Some subsequent member of this list is false.
(19) Some subsequent member of this list is false.
(20) P
(21) Q
Now, (18) must be true for if it was false, then (19) would have to be true, but if 
(19) is true then (18) is true. So (18) must be true, and therefore some subsequent 
sentence is false.

Consider the list:

(22) Some subsequent member of this list is false.
(23) Some subsequent member of this list is false.
(24) Some subsequent member of this list is false.
(25) Some subsequent member of this list is false.

By the above reasoning, any member of such a list must be true if it has another 
such sentence following it.

An infinite list of such sentences has no last sentence and Sorensen’s Queue 
paradox results.

Infinite Truth-tellers result from using ‘is true’ instead of ‘is false’. The members 
of the following infinite list can all be true or all false.

(26) Some subsequent member of this list is true.
(27) Some subsequent member of this list is true.
(28) Some subsequent member of this list is true.
...
(n) Some subsequent member of this list is true.
...

The same is true of a list of universally quantified ‘is true’ sentences. I take the 
above to be the hypodoxical correlate of the paradox using universally quantified 
expressions.

Goldstein [1999] claims that the behaviour of Yablo’s paradox can be modelled 
by circular paradoxes. This is true enough for a circle of ‘Each clockwise subsequent 
sentence is false’ sentences. However, a circle of ‘Some clockwise subsequent 
sentence is false’ sentences is not a model for Sorensen’s. If it has an even number of 
members, then each second one can be false and each alternate one true. There is no 
reason to choose between which half of the members should be true and which false. It 
is a hypodox of sorts. It is not a model for Sorensen’s queue paradox. The appropriate 
model is a list. The circular list is a different variation to the infinite list.
3.6.2 Infinite Epimenides and Curry’s Paradoxes

Variations from Yablo’s and Sorensen’s can be created by adding a sentential connective and a sentence in, say, the first line. Consider the sequence:

(Y1) For all k greater than 1, Yk is not true & Q.
(Y2) For all k greater than 2, Yk is not true.
...
(Yn-1) For all k greater than n-1, Yk is not true.
(Yn) For all k greater than n, Yk is not true.
...

If (Y1) was true, then (2) would be false, but if (2) is false then for some k greater than 2, Yk is true. So (Y1) must be false, and therefore either some subsequent sentence is true or ~Q. We know from our discussion of such lists that the only subsequent sentence that could be true is the last sentence but the list is infinite so there is no last sentence; therefore ~Q.

Likewise, consider the sequence:

(S1) For some k greater than 1, Sk is not true v Q.
(S2) For some k greater than 2, Sk is not true.
...
(Sn-1) For some k greater than n-1, Sk is not true.
(Sn) For some k greater than n, Sk is not true.
...

As we know, (S1) must be true for if it was not, then (2) would be true, but if (2) is true then (1) is true. So (1) must be true, and therefore either some subsequent sentence is false or Q. We know that the only subsequent sentence that could be false is the last sentence but the list is infinite so there is no last sentence; therefore Q.\(^\text{18}\)

\(^{18}\) This argument can also be run against a list like this:

(S1) For some k greater than 1, Sk is not true
(S2) Q.
(S3) For some k greater than 3, Sk is not true.
... (Sn-1) For some k greater than n-1, Sk is not true.
(Sn) For some k greater than n, Sk is not true. ...
3.7 Conclusions

The relation between some problematic collections of sentences and lists is important. Self-referential and circular sentences are all variations of lists of sentences referring to each other in an ungrounded way. Indeed, self-referential sentences are trivially very tight circular sentences, in a circle of one. So, self-referential sentences are a subset of circular-referential sentences, which are themselves a subset of lists of sentences that refer to themselves or other members of the list. So the relationship between self-referential, circular and infinite variations in the Liar family can be given in terms of lists.

List structures or unordered collections together with identity statements have been used to classify paradoxes within the Liar family; and these have been used with the dual conjecture to so classify dual hypodoxes. Individuation by using identities has encountered some limitations, and the possible valuations of sentences has been used as a complementary way of individuating paradoxes (within this family and within list structures).

The Liar, Unquantified Epimenides, Self-referential Curry and ESP form an interesting subset of four variations of sentential paradoxes of self-reference and there are four associated hypodoxes. The ESP paradox is, in a sense, its own Truth-teller. This subset can be extended by considering possible combinations of valuations to a complete subset of unquantified self-referential categories of sentences that are possibly paradoxical or hypodoxical. The Epimenides, Curry, ESP variations, and other members of this set can be used in combination with circular and infinite Liar-like paradoxes and hypodoxes.

I have shown how the quantificational variations of the Epimenides and Geach's paradox relate to the unquantified variations of the Epimenides and the self-referential Curry paradox. I have briefly discussed how such quantificational sentences relate to collections, not surprisingly. The membership of some of these collections is a contingent matter, in contrast to intuitions about the Vicious Circle Principle.

Although it seemed innocuous, the dual conjecture would encounter some anomalies. (Of course it never would have been considered were it not for naively equating being not true with being false.) Nevertheless, in the circular cases, the conjecture still seems to hold in terms of being able to give a one-to-one correspondence between circular paradoxes and hypodoxes for a circle of \( n \) sentences, because the "no-no" cases are actually Truth-tellers; but this was not obvious to begin
with. However, the neat semi-formal production of a specific hypodox from a specific proper-circular paradox is compromised by a choice of which sentence to negate (and "re-letter") in circular cases.

The impact of Kripke’s outline of a theory of truth on research into the Liar was like a Gödel theorem – it at once provided a long sought result, a way of sifting out of Liar-like paradoxes, at the same time as proving the impossibility of a successful conclusion to a research project, that of sifting them all out purely syntactically or semantically. Kripke’s definition of *paradoxical sentence is material* in the sense that it is relative to one fixed model. Kripke’s point constrains the present project: Kripke settles it that there can be no purely syntactic or semantic sieve for all paradoxes.

Nevertheless, I hope to have shown that the classification of paradoxes can be complete in at least parts, the classification of truth-functional unquantified self-referential paradoxes and hypodoxes being my prime example. In that case, although I had to use possible valuations in order to obtain a set of categories that was truth-functionally closed; nevertheless, both the subset of seven identities and the full set of twelve possible sets of valuations are a useful classification. At the other extreme, I own that circular quantified paradoxes, such as the one involving the majority of Nixon’s statements about Watergate, have an open-ended classification and all I aim to do is provide some elements that would be used to classify them.

While the classification of circular, quantified variations seems open-ended, the prospects for the classification of paradoxes as a project seem positive. As an immediate result, it helps scope and structure the problem space. I have focused on the Liar’s immediate relatives. I have explored these case-wise, being content here with observation. Some subclasses are complete in a sense, such as the unquantified self-referential Liar-like paradoxes. This is satisfying. Other portions of the classification are open-ended. Here I have aimed at a robust, but extensible list of elements for classification of Liar-like paradoxes and hypodoxes. The dual conjecture has been somewhat confirmed by this semi-formal, semi-case by case exposition. Further work along the present lines involves varying the predicate and the schema used. In this way, we can map parts of this classification onto other families, such as paradoxes of satisfaction and membership – I explore this in the next chapter: the mapping is not as straightforward as might be expected.19 Although calling a collection of paradoxes a

---

19 I have also researched mapping to some other families with positive results, such as Knowler paradoxes, and even paradoxes of time travel [Peter Eldridge-Smith 2007].
with. However, the neat semi-formal production of a specific hypodox from a specific proper-circular paradox is compromised by a choice of which sentence to negate (and "re-letter") in circular cases.

The impact of Kripke’s outline of a theory of truth on research into the Liar was like a Gödel theorem – it at once provided a long sought result, a way of sifting out of Liar-like paradoxes, at the same time as proving the impossibility of a successful conclusion to a research project, that of sifting them all out purely syntactically or semantically. Kripke’s definition of paradoxical sentence is material in the sense that it is relative to one fixed model. Kripke’s point constrains the present project: Kripke settles it that there can be no purely syntactic or semantic sieve for all paradoxes.

Nevertheless, I hope to have shown that the classification of paradoxes can be complete in at least parts, the classification of truth-functional unquantified self-referential paradoxes and hypodoxes being my prime example. In that case, although I had to use possible valuations in order to obtain a set of categories that was truth-functionally closed; nevertheless, both the subset of seven identities and the full set of twelve possible sets of valuations are a useful classification. At the other extreme, I own that circular quantified paradoxes, such as the one involving the majority of Nixon’s statements about Watergate, have an open-ended classification and all I aim to do is provide some elements that would be used to classify them.

While the classification of circular, quantified variations seems open-ended, the prospects for the classification of paradoxes as a project seem positive. As an immediate result, it helps scope and structure the problem space. I have focused on the Liar’s immediate relatives. I have explored these case-wise, being content here with observation. Some subclasses are complete in a sense, such as the unquantified self-referential Liar-like paradoxes. This is satisfying. Other portions of the classification are open-ended. Here I have aimed at a robust, but extensible list of elements for classification of Liar-like paradoxes and hypodoxes. The dual conjecture has been somewhat confirmed by this semi-formal, semi-case by case exposition. Further work along the present lines involves varying the predicate and the schema used. In this way, we can map parts of this classification onto other families, such as paradoxes of satisfaction and membership – I explore this in the next chapter: the mapping is not as straightforward as might be expected.19 Although calling a collection of paradoxes a

---

19 I have also researched mapping to some other families with positive results, such as Knower paradoxes, and even paradoxes of time travel [Peter Eldridge-Smith 2007].
family of paradoxes suggests some unpredictability about the number and characteristics of its members, at least parts of the collection have a complete classification and the collection's total membership is more predictable than one might expect of members of a family.