Chapter 2 Truth and Paradox

I am concerned in this chapter to introduce the T-schema, which traditionally associates truth with paradox, and to establish some semantic elements of my classification of Liar-like paradoxes to be given in the next chapter. Some of these elements will cut across paradoxes of satisfaction and set-theory in ways explained in Chapter 4. There is a need for a way of individuating types of paradoxes, that is, to be able to say what constitutes a different paradox and not just a novel instance of a paradox.

I assume a naive semantics and assume a version of the T-schema appropriate to this semantics. Moreover, I usually treat 'is false' as meaning the same as 'is not true'. I am not working *in* a first-order theory, though I do naively assume natural language has some classical rules of inference. It clearly includes its own truth-predicate, for better or worse. The rules of inference used are ones commonly used to represent the paradoxes; so they are no impost on this classification (with one notable exception to be introduced in Chapter 5).

In Section 1, I consider one motivating belief for the T-schema. With respect to truth, I begin with validity in section 2. As validity preserves truth, one might think the definition of validity would suggest rules for the truth predicate; but it does not establish the T-schema. In section 3, I discuss the bearers of truth in so far as I want to represent the Liar without prejudicing what these are. In section 4, I consider another set of beliefs that entail the T-schema. In section 5, I deal with truth inferences and the T-schema, explaining the canonical form I shall favour. In section 6, I investigate semantic valuations using the T-schema or inferences. In section 7, I briefly attempt to step above my naive viewpoint and consider metaphors for truth, as a way of explaining some of the seemingly incommensurable differences between the intuitions motivating some theories of truth. In section 8, I apply methods of valuation in relation to the semantic valuations given in section 6. In section 9, I briefly discuss strengthening and other ways in which attempts to resolve the Liar paradox encounter further paradoxes. In section 10, I discuss the relationship between paradoxical sentences (or expressions) and paradoxical arguments. I explore a number of misconceptions, consider the concept of paradox itself, and put forward some elements

I use to differentiate paradoxes. The section begins an ongoing theme of how to individuate variations of paradoxes, or how to categorize paradoxes (within a family). In section 11, I discuss the concept of *hypodox*.

2.1 The Substitution Thesis

A naive theory of truth may be motivated by the following tenet.

> Substitution Thesis: T⟨@⟩ and @ are intersubstitutable in any non-opaque context.

Historically, the primary argument for the Substitution Thesis has been that these expressions intuitively have the same meaning; to say that ⟨@⟩ is true is just to say that @.[1] Dummett [1959, pp. 144-145], attributes a similar view to Frege, cast in terms of Frege's theory of sense and reference:

> A popular account of the meaning of the word " true", also deriving from Frege, is that ⌜It is true that P⌝ has the same sense as the sentence P. If we then ask why it is any use to have the word "true" in the language, the answer is that we often refer to propositions indirectly, i.e., without expressing them, as when we say "Goldbach's conjecture" or "what the witness said". We also generalise about propositions without referring to any particular one, e.g., in "Everything he says is true".

The Substitution Thesis together with Identity (@ → @) warrants a biconditional connection between @ and T⟨@⟩, i.e. The T-schema:

> T-schema: T⟨@⟩ iff @

Here, as in the previous chapter, 'T' represents the truth-predicate, '@' is used as a meta-variable, and the angle brackets are used to represent a name-forming functor. However, the Liar paradox has appeared as a contradiction following from the T-schema and classical logic in a language with these facilities, i.e. the ability to name its own expressions and talk about its semantics in the same language [Tarski 1935].

---

[1] Dummett argues this equivalence is inconsistent with another view of Frege's that a sentence could have a sense but no reference. (On Frege's view sentences refer to truth values.)

The substitution thesis does not entail the T-schema in all logics. As Field [2003, pp. 139-40] points out, in Kripke's theory of truth [1975] using (just) the strong Kleene valuation scheme T⟨@⟩ and @ are intersubstitutable, but there is no biconditional for which @ ↔ @ is a logical law, let alone the T-schema. Field defines *the naive theory of truth* as having the T-schema with the Substitution Thesis. His theory depends on a non-material interpretation of the biconditional in the T-schema.[2]

Nevertheless, the Substitution Thesis is one motivation for a naive theory of truth. I will give another motivation for a naive theory in a subsequent section on *Truth Inferences and the T-schema*, but first I will discuss validity and the bearers of truth.

2.2 Validity, Truth and Paradox

As validity preserves truth, one might expect to learn something about inferences involving truth from definitions of validity. I accept, perhaps naively, the following inference as valid:

@ ⊢ @

And conclude from this being a valid inference that both the following are theorems:

⊢ @ → @

⊢ T⟨@⟩ → T⟨@⟩

What the validity of this inference has not determined is whether or not the following are theorems:

(T-elim) T⟨@⟩ → @

(T-intro) @ → T⟨@⟩[3]

The Substitution thesis would warrant T-intro and T-elim. It is tempting to argue for the Substitution Thesis on the basis that if validity preserves truth, then if @ is the case, then ⟨@⟩ is true; but this is not necessarily what preserving truth means. It would be a question begging argument.

---

[2] Again, Field [2005, pp. 27-28] defines the Naive theory of truth with reference to two principles, the T-schema and Intersubstitutivity (the *Substitution Thesis* above). He explains that these are classically equivalent, but not for some non-classical logics.

[3] These are not of course the rules of inference TI and TE, from which these could be derived as theorems were those rules given.

Nevertheless, if T-intro and T-elim are not theorems, then our logic begins to look strange. Consider a version of Moore's paradox, an epistemic paradox. It is strange to assert sentences like the following:

(2.1) It is raining, but I do not believe it.

Then it is certainly strange to assert something like:

(2.2) It is raining, but 'it is raining' is not true.

It is generally considered wrong in some way to assert (2.1); but it seems strange to think so unless we also think it is wrong to assert (2.2). Yet without affirming something like TI and TE, or T-intro and T-elim, we have given no reason to find it strange to sometimes assert sentences of the following forms:

$T\langle @\rangle \ \& \sim @$

$@ \ \& \sim T\langle @\rangle$

If we affirm the Substitution Thesis, rules like TI and TE, or schemata like T-intro and T-elim, then assertions of instances of $T\langle @\rangle \ \& \sim @$ or $@ \ \& \sim T\langle @\rangle$ would be inconsistent. Naively, there are some extremely strange things in our language, as the Liar provides such an expression, as in line 4 below. (Let '*a*' abbreviate 'my favourite sentence'.)

1. $a = \langle \sim Ta\rangle$	premise
2. $\sim Ta$ iff $\sim T\langle \sim Ta\rangle$	Identity ($@ \rightarrow @$), SL, 1 =E (Leibniz's law)
3. A iff $\sim T\langle A\rangle$	2 with 'A' abbreviating '~Ta'.
4. $A \ \& \sim T\langle A\rangle$ or $\sim A \ \& \ T\langle A\rangle$	3 SL

In combination with intuitions about truth we have paradoxes like the Liar. Validity is pair-wise consistent with either our intuitions about truth, or with the Liar sentence; but it is not consistent with both. The initial result of this section is negative: The T-Schema has not been derived from the concept of validity.

There are other intuitions that we could nevertheless derive from the concept of validity.

Firstly, as truth is preserved by validity, there are valid distribution rules for truth over connectives corresponding to valid inferences such as &I, &E, *modus ponens*:

$T\langle A\rangle, T\langle B\rangle \vdash T\langle A\&B\rangle$

$T\langle A\&B\rangle \vdash T\langle A\rangle$

$T\langle A \rightarrow B\rangle, T\langle A\rangle \vdash T\langle B\rangle$

Similar *T-rules* are justified by the concept of validity for each form of valid inference.

Secondly, my modification to Bradwardine's principle which I used in modifying Buridan's additional criterion for truth in Chapter 1 can be derived from the concept of entailment. This was given in Chapter 2 as:

Nothing entailed by assuming the truth of the proposition is false

Since entailment preserves truth, if a sentence (on its own or together with other truths) entails something not true, then the sentence is not true. So, if a sentence is true, then it does not entail something not true.

2.3 Truth-bearers

In representing forms for the Liar and Liar-like paradoxes in the next chapter, I will want to remain ecumenical about the bearers of truth. There are well-known arguments against sentences as the primary bearers of truth.[4] A solution may well involve paradoxical sentences or tokens not expressing a bearer of truth, be it a proposition or statement. Nevertheless, I am concerned with representing and classifying the forms of the paradoxes prior to such a resolution. So this section aims at warranting the use of sentences in representing the form of Liar-like paradoxes as they occur, prior to any transformations aimed at resolving them.

Grammatically, truth is a predicate with a name of a sentence as its subject term; however, truth is attributed to other objects as well. When we want to state exactly what is true or false, we either mention a sentence or use an expression like a that-clause; expressions such as the following are the norm.

(2.3) 'Minos commissioned the Labyrinth' is true.

(2.4) It is true that Minos commissioned the Labyrinth.

(2.5) That Minos commissioned the Labyrinth is true.

The first of the above grammatically attributes truth to a sentence. The second uses a truth operator, not a predicate. Its use of 'that' would generally be parsed as part of the operator. For present purposes I treat the truth operator as redundant and equate its

---

[4] The same truth can be expressed by different sentences in different languages, or even different sentences in the same language. From the converse point of view, the same sentence can be used to express different truths, perhaps simply because of indexicality, but nevertheless, the point is still valid. Cf. Cartwright [1962, pp. 86-88].

negation with the external negation of the sentence to which it attaches. The third sentence is of more interest at present. Although the third might be parsed as using an operator, *prima facie*, it predicates truth to the reference of a noun clause, 'that Minos commissioned the Labyrinth'. Grammar assures us of the existence of the sentence in quotes as the reference for the quotation name of the sentence, at least grammar provides (almost) effective procedures for verifying it is grammatical. Whether there is such an effective procedure to assures us of the existence of the reference of the that-clause depends on what sort of entity it is. Say the 'that'-clause refers to a semantic metaphysical entity such as a proposition. Then we can mitigate our exposure to an existential presupposition with a transformation like this:

(2.6) 'Minos commissioned the Labyrinth' expresses a true proposition.

Moreover, whatever is entailed by (2.3) is entailed by (2.5).

When we want to use truth in a general way, we might use expressions like the one below.

(2.7) Everything Epimenides said about Minos is not true.

We do not need to list, or even be able to list, everything Epimenides said about Minos in order to understand what this means. Nevertheless, my concern with such statements in representing the paradoxes is to be able to use universal instantiation (∀E) to infer expressions like the following.

(2.8) If Epimenides said that Minos had two brothers, then that Minos had two brothers is not true.

(2.9) If 'Minos had two brothers' was said by Epimenides, then 'Minos had two brothers' is not true.

(2.10) If Epimenides said that Minos had two brothers, then 'Minos had two brothers' is not true.

I have already suggested that the consequent of the second and third of these expressions is or can be treated as a logical consequence of the consequent in the first. I note that the antecedent of the second conditional is clearly false, given that the quote expression names a sentence. So, in fact, it would be incorrect to instantiate the general claim about things that Epimenides said about Minos with the second conditional

above.[5] Nevertheless, the first or third could be validly inferred, and whatever follows from the first would follow from the third, given the antecedent.

Truth is a semantic concept; so, its attribution to a sentence means that in some way truth relates to the meaning of the sentence; it would be a category mistake to *attribute* truth to a purely syntactic sentence. Nevertheless, if we interpret quotes as naming a syntactic expression, as Tarski did and as grammar suggests, then truth is *predicated* of a name of a syntactic sentence – any string of symbols equiform with the physical marks on the page. This poses a conundrum, as the most straight-forward transitions from predication to attribution are to say that the predicate or associated property is attributed to the subject of the predicate.

Quine's [1970, p.12] analysis was that truth is a device for disquotation: saying a sentence is true is a way of making a claim about the world. My own intuition is that predication of truth is to be interpreted as being attributed to the interpretation of the sentence named. However, the audience need not understand German to understand the basic claim made by the speaker who asserts:

'Schnee ist weiss' is true.

What we understand by the basic claim is that the interpretation of 'Schnee ist weiss' is true. However, if truth is attributed to the physical sentence, we should correctly say:

'Schnee ist weiss' is a true sentence of German.

Although we would need semantic information about the quoted sentence to determine the truth of the exhibited sentence, we do not need to understand the expression in quotes to know what the above sentence means. On Tarski's account, the first of the following two sentences is grammatically ill-formed or cursive for the second.

'Daedalus designed the Labyrinth' is true.
'Daedalus designed the Labyrinth' is a true sentence of English.

I do not think the first is ill-formed or abbreviated; the second entails the first; but it is arguable whether the first entails the second. Unless we can formally recognise

---

[5] This is an issue for the representation of the Epimenides using objective quantification over sentences in combination with multiple languages. Epimenides presumably wrote his poem in some archaic from of Greek. The derivation I gave in Chapter 1 is intended to represent the Epimenides within a language. A remedy would take us deeper into issues of translation than is warranted for representing Liar-like paradoxes however. I note that the use of a strict object- meta-language distinction, at least in a Tarskian way, would preclude representing the derivation – some expressions would be ill-formed.

the content of the quotation string as an English sentence, then the first does not entail the second.

A classification needs to be as ecumenical as possible and apply to whatever the bearers of truth are. I do not need to solve the conundrum associated with predicating truth to sentences, but I do need to know that representing this predication is not prejudicing differing interpretations relating to differing, orthodox theories of the bearers of truth. To this end, although Visser [2004, pp. 151-4] makes the distinction between sentential and propositional versions of the Liar a primary distinction, he introduces a rephrasing.[6]

This sentence is not true.

can be rephrased as:

This sentence does not express a true proposition.

The two of these can be represented with the same form. This generality is achieved by allowing flexible interpretation of a T-predicate and a name-forming function. A standard interpretation of the predicate represented by 'T' will be 'is true'. On other interpretations, 'T' may represent 'expresses a true proposition', 'expresses a true belief', 'makes a true statement', or the like.[7] I think this generalises pretty well, but I do not think that this will represent variations involving sentence tokens.[8]

---

[6] Actually, Visser [2004] implicitly accepts a primary distinction between semantic and set-theoretic paradoxes; so that he secondarily distinguishes sentential and propositional versions of the Liar among the semantic paradoxes. I wonder if he would extend this distinction to one between syntactic expressions and corresponding semantic entities like propositions and properties.

[7] Personally, I believe that a solution to the sentential versions will thus flow through to any other account of the bearers of truth. I do not believe the converse. I believe sentential versions are the harder versions. If truth is not born by sentences, we can invent a predicate that is attributed in this way. This may not be a problem for a theory of truth, but it is still a problem for logic. Indeed it is a bifurcate problem. Firstly, there is the T-predicate representing an attribute of syntactic sentences. Secondly, there is the T-predicate representing an attribute of interpreted sentences. Interpreted sentences are composite entities. A solution for the syntactic-T-predicate would flow through to the other, but not necessarily conversely.

Furthermore, It is well known that solutions that rely on varying the values or the bearers of truth are subject to strengthened or revenge forms of the Liar Paradox. Terminology is not standardised, but let us say: Strengthened versions are a response to varying the truth-values. Revenge Liars are a response to varying the bearers of truth.

[8] I believe Strengthened Liars are in some sense complemented by solutions to the Liar that attribute truth to sentence tokens. If 'The sentence token on the blackboard is not true' is written on the

Another interpretation of 'T' will be 'is a true sentence of English', or the relevant language, possibly one expanded with logical expressions. (I will generally not need to use the 'T' predicate in a way that would involve translation issues.)

There is another alternative. A standard interpretation of the function represented by angle brackets will be a name-forming functor, which forms canonical names, like quote-names, from which the expression named can be effectively recovered. Alternatively, or in addition, other interpretations may be used. According to the theory of truth, the use of '⟨@⟩' may be interpreted as forming a name of the appropriate truth-bearing thing such that the sentence that is or expresses that truth-bearing thing can be effectively recovered from the name. This may allow for grammatical variation, such as between a 'that'-clause and quote name. (Whether such a name is vacuous is presumably part of a solution.). I do not think this alternative is as effective as varying the interpretation of the 'T' predicate, however, and generally rely on the former.

We will need to account for the use of indexicals, arbitrary names and definite descriptions to pick out Liar-like sentences. For indexicals and arbitrary names, an interpretation is relative to a context, such as being interpreted in a particular way in the context of this text. Even what 'My favourite sentence is not true' refers to in this text may be different to what it would refer to in another by another author; so it seems natural enough to talk about these as different *statements*. By a *statement* I do not mean a sentence token, but rather a particular interpretation of a sentence, at least fixing the reference of all indexicals, names and definite descriptions.[9] Of course, it is

---

blackboard, then an equiform token of that sentence itself can be used to say truly that the sentence on the blackboard is not true. Regardless of the interpretation intended for 'not' in the sentence token on the blackboard, the interpretation of 'not' in the sentence token about it is intended to be complement-negation (not choice negation). Strengthened Liars do not distinguish tokens, and the second instance is intended to use complement negation to regain a paradox. Any restriction in the object-language to choice negation, still leaves open the possibility in the meta-language (often natural language) to use complement negation to make the statement. There is a lot more to investigate about Strengthened Liars in this regard, in relation to strong and weak truth, and in relation to generating infinite hierarchies of ever more strengthened paradoxes, but not here.

[9] Personally, I think truth is a predicate appropriate to interpreted sentences or 'statements', which allow for varying the actual referents of a sentence. For me, statements are simply intersection entities between sentences and propositions. In this respect, they are not tokens, but have tokens themselves. A sentence may be used to express multiple propositions and a proposition might be expressed by many sentences. Statements are interposed in this relationship such that many statements may be made using

theoretically contentious to fix the reference of 'My favourite statement'. It may be argued with reference to Russell's famous representation of definite descriptions, that the correct representation of 'My favourite statement' involves an existential quantifier. This complication does not constrain what we can prove. But should we wish for a faithful representation, we can map these expressions to representations involving existential quantification. These would be like Geach's existential variation from the previous chapter; but the uniqueness of the expression falling under the condition in the definite description would be built in. So, for our purposes, either an interpretation is still given by a standard model, which gives the denotation of all the names, including 'my favourite sentence' or the extension of the predicate 'is a favourite sentence of Peter Eldridge-Smith' has just one member, 'There exists a sentence y, which is a favourite sentence of Peter Eldridge-Smith and if any sentence is a favourite sentence of Peter Eldridge-Smith then it is identical with y, and y is not true.'[10] For simplicity, I treat expressions like 'my favourite sentence' as names.

2.4 The Equivalence Thesis

'This sentence is true' takes the same truth value as the sentence to which its indexical phrase refers. In this respect, 'is true' behaves like an operator, but is predicated of a sentence, @; so the indexical term must name or in some way refer to @. In general, the following principle captures this naive intuition:

Principle T: An attribution of truth has the same truth value as the

---

the one sentence but only one sentence makes a particular statement, and likewise between statements and propositions. Nothing really rides on this in this here. The classification presented is relatively robust with respect to beliefs about the primary bearers of truth – appropriate modifications can be made.

Personally, I expect that a derivational system involving the truth predicate will be relative to a model, i.e. derivational rules for sentences involving the truth-predicate will not be simply syntactic. McGee [1991], for example, carries this out in practice but remarks on it only in relation to his own system.

'True' is also grammatically used as an operator.

[10] Cf. Glanzberg [2001, pp. 226-229].

sentence to which it is attributed.[11]

Then, in a list of sentences, 'The next sentence is true' simply takes the same truth value as the next sentence in a list.

One can use a negation operator; so that 'It is not true that this sentence is true' does not take the same truth value as the sentence to which its indexical phrase refers. We equate the sentence with 'This sentence is not true', and say that it also does not take the same truth value as the sentence to which it refers. Call this Principle F.

> Principle F: An attribution of falsity (or 'is not true') has the opposite truth value to the sentence to which it is attributed.

Then, in a list of sentences, 'The next sentence is not true' does not take the same truth value as the next sentence. If 'This sentence is not true' refers to itself, then it says of itself that it does not have the same truth value as itself, that is, it is true iff it is not true. Here, again, the biconditional should be interpreted materially based on this argument.

In this way, Principle T and Principle F together with bivalence warrant the Equivalence Thesis, which is the T-schema using a material biconditional:

> Equivalence Thesis: If α is a term referring to $\mathcal{Q}$, then:
>
> $$T\alpha \equiv T\langle\mathcal{Q}\rangle \equiv \mathcal{Q}$$

This argument should be distinguished form the Substitution Thesis which licences use of a biconditional statement, wherein the biconditional may be non-material. (I shall use 'iff' as ambiguously representing a material or non-material interpretation.)

---

[11] A similar maxim is referred to as *Aristotle's rule* by Buridan [Hughes, 1982, pp. 45-47]. Indeed, one can interpret Aristotle as affirming principles T and F (to be introduced above), as indeed Tarski found motivation for the T-schema in Aristotle:

> To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true. [Aristotle, *Metaphysica,* Γ, 7, 27; Works, vol. 8, English translation by W. D. Ross, Oxford, 1908, cited in Tarski 1935/1983, footnote p. 155].

2.5 Truth-inferences and the T-schema

Let me expand on the introduction given to the Liar in the previous Chapter. For convenience, here is a version of the Liar again.

1 My favourite sentence = 'My favourite sentence is not true'

Premise, by stipulation or observation

2 'My favourite sentence is not true' is true iff my favourite sentence is not true

T-biconditional

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

In the above argument, the vinculum appears under Line 1 as it is the only premise: the T-schema is included in the logic. The logic also includes substitution of co-referring terms and some sentential logic by which line 4 follows from line 3.[12] In general, in this chapter, I use only identity premises, but in this section I will briefly consider not including the T-schema in our logic. Perhaps a T-biconditional should be a second premise or an assumption. The argument might then be a *reductio* of lines 2 or 3. As our intuitions were not strongly attached to these particular T-biconditionals, a solution recasting the argument as a *reductio* of line 2 or 3 has appealed to many. Indeed, why look at arguments at all? The identity premise and the T-biconditional of line 2 form an incompatible set of sentences, at least one of which must be rejected to make the set consistent. The identity is contingently true; so the T-biconditional must be rejected. However, the intuition on which the T-schema is based appeals to two very intuitive rules of inference, namely Truth-introduction (TI) and Truth-elimination (TE):

TI: $@ \vdash T\langle @ \rangle$

[12] Here, I believe, restrictions on the T-schema face a challenge. It would need some independent justification to restrict Line 3 and not Line 2 (by restricting the T-schema), because Line 3 follows from 2 by substitution of identicals. However, Line 2 only requires a canonical T-schema (to be introduced in a subsequent section) and that canonical T-schema has the truth-predicate effectively behaving as an operator. Arguably, the intent of restricting the T-schema was to restrict the truth-predicate not a truth-operator.

TE: $T\langle @\rangle \vdash @$

These rules of inference together with conditional proofs are sufficient to derive the T-biconditional at line 2. Indeed, given *modus ponens* and conditional proof, TI and TE together are equivalent to a restricted form of the T-schema, which I will call the "canonical T-schema" because it is restricted to canonical names.

Canonical T-schema: $\vdash T\langle @\rangle$ iff $@$

This version of the schema is restricted to canonical names composed from the sentence replacing the meta-variable '$@$' in the schema and a name-forming functor. These form canonical names, like quote-names, from which the expression named can be effectively recovered. (Of course, they do not form a quote name of '$@$' in the schema. The T-schema is in the meta-language; so are the rules of inference. Being schematic, any quantification is in the meta-language. The sentential variables are meta-language variables to be replaced by object-language sentences. The angle brackets are a device that perhaps refers to quotation in the object-language. In the object-language we never quantify into quotation marks, and in the meta-language we do not use them. There is perhaps an alternative interpretation. If we did not use a meta-language, we would use substitutional quantification when we use a schema and the angle brackets represent a function that takes a sentence as an argument and yields a name for that sentence as a value. For the canonical T-schema to hold, we must replace 'P' in *both* instances by the same sentence, either way quantification does not bind a variable inside quotes. Whether the schema is interpreted substitutionally or in the meta-language, the angle brackets must be distinguished from quotes.)

Any instance of a more general T-schema, involving non-canonical or canonical names, is derivable using TI, TE, an identity and substitution of identicals.[13] Use of a rule equivalent to the T-schema would then look like this:

$\alpha = \langle @\rangle \vdash T(\alpha)$ iff $@$

So that 'α' is a name for a sentence represented by '$@$'.

Furthermore, we might instead use substitution of co-referring terms, and appeal to the semantics to seemingly avoid the need for a premise at all. That is, if the

[13] Observation 2: If we use TI and TE without substitution of identicals, then the T-predicate behaves essentially the same way as a truth-operator. The resultant logic is still subject to versions of the Epimenides using substitutional quantification, which concerned Prior [1958].

semantics includes the reference of 'my favourite sentence', then use of a rule equivalent of the T-schema allows:

(Tarski's T-schema) $\vdash$ T(α) iff @

where 'α' denotes a sentence @ in the interpretation.

So, given the reference of 'my favourite sentence', we have a one line derivation of the Liar:

$\vdash$ My favourite sentence is true iff my favourite sentence is not true.

A premise is avoided by having an equivalent truth in the semantics. That is, in the semantics 'my favourite sentence' refers to 'My favourite sentence is not true', so that between the semantics and sentential naming conventions about the use of quotes, my favourite sentence and 'My favourite sentence is not true' are co-referential. An identity premise is a good way to represent this.

Indeed, the premise can be reduce to a theorem using canonical naming and a substitution function, as will be proven in Chapter 5. Nevertheless, an identity remains a good way to represent this commitment. For such an identity, or the co-reference it represents, is necessary to deriving the Liar in any case, even if in one case it is a theorem.

Finally, I add two notes. Firstly, that Tarski's T-schema uses an explicit material biconditional. Secondly, that Tarski's T-schema is stronger still. For on Tarski's account 'α' is a meta-language variable to be substituted with a meta-language name for an object-language sentence which is translated in the meta-language by the sentence represented by '@'. So, the following instance of Tarski's T-schema where the object-language is German and the meta-language is English is not validated by either the naive T-schema or the canonical T-schema; but by Tarski's full T-schema:

$\vdash$ 'Schnee ist weiss' is true iff snow is white.

This instance of the T-schema is not derived from TI and TE. In this respect, the *full* T-Schema goes beyond intuitions stemming from TI and TE inference rules. Nevertheless, we do not need Tarski's full T-schema to derive the Liar, and I will not be using it. The canonical and naive, weakened versions of the T-schema are all we need for formulating the Liar paradox.

2.5.1 Truth of the Negation and the Negation of Truth

I add this brief note that, naively, '@ is false' is representable by either

$\sim T\langle @ \rangle$

or

$T\langle \sim @ \rangle$.

These are equivalent, given the T-schema. Both the following are instances of the T-schema.

$T\langle \sim @ \rangle$ iff $\sim @$

$T\langle @ \rangle$ iff $@$

It follows from these, by Sentential Logic, that:

$T\langle \sim @ \rangle$ iff $\sim T\langle @ \rangle$.

Given bivalence, it is natural to think of 'is false' as 'is not true'. Given the T-schema, '$@$ is false' is representable by either of the above mentioned equivalent expressions.

2.6 The Semantic Value of the Liar Sentence and the Truth-teller

In this subsection I am concerned with the semantic value of the Liar and Truth-teller sentences. I contrast two ways of naively establishing their semantic value. Firstly, consider the bottom-up valuation of these sentences based on the values of their components. Secondly, consider the top-down or "modal" valuation based on alternately assuming a sentence is true and assuming it is false. "Modal" has other connotations, so I will refer to this valuation method as *top-down* (adapting a term from Information Communication and Technology).

Although my favourite sentence is proved true and false, it does not have these values in virtue of its form and the truth values of its components. The form and content of my second favourite sentence are insufficient for it to have any particular truth value. Paradoxical sentences and truth-tellers seem to fall outside the scope of the principle of *compositionality*, which (in this context) is the principle that the semantic value of a complex sentence is determined by the semantic values of its components. (This *compositional interpretation* is in contrast with a more general interpretation of compositionality in the literature, which considers possible truth values.) Under this compositional interpretation, as I use it here, $@ \vee \sim @$ does not have a semantic value compositionally unless $@$ or $\sim @$ has one. One must de-compose a sentence first, in order to establish its truth value. The truth values of sentences that cannot be further de-

composed are settled by reference to a base model. The Liar sentence cannot be decomposed to a sentence that does not need further decomposition.

For a one place predicate, $\mathcal{P}$, the truth of a sentence $\mathcal{P}\alpha$, where 'α' is a name, is determined by whether or not the referent of '*a*' satisfies the criteria for falling under the predicate '$\mathcal{P}$'. So, where 'T' is the truth predicate:

> Tα iff the referent of 'α' satisfies (the criterion for falling under) the truth predicate.

It seems natural to extend compositionality to the truth-predicate using the Principle T. (The principle that an attribution of truth takes the same truth value as the sentence to which it is attributed.)

Thus, for simple sentences, like 'Socrates is mortal', we can substitute:

> Socrates is mortal iff 'Socrates' refers to an object that satisfies the predicate 'is mortal'.

into:

> 'Socrates is mortal' is true iff 'Socrates is mortal' is a well-formed sentence and Socrates is mortal.

to obtain:

> 'Socrates is mortal' is true iff 'Socrates is mortal' is a well-formed sentence and 'Socrates' refers to an object that satisfies the predicate 'is mortal'.

In general, where '$\mathcal{P}$' represents a one-place predicate and 'α' is a name:

> '$\mathcal{P}\alpha$' is true iff '$\mathcal{P}\alpha$' is a well-formed sentence and 'α' refers to an object that satisfies the predicate '$\mathcal{P}$'.

Take just a moment to notice the magic of the T-schema: while in the left-hand side of the above biconditional 'truth' appears to be a predicate of a syntactic expression, the right-hand side assures that it will only be correctly attributed to well-formed, interpreted sentences. At least, this suffices for sentences of subject-predicate form. This goes some measure to addressing the conundrum posed earlier about how truth as a semantic predicate could be appropriately attributed to a syntactic sentence.

Compositionally complex sentences have their values determined by the values of their components. So, for example:

'Socrates is alive and living in Thrace' is true iff 'Socrates is alive' is true and 'Socrates is living in Thrace' is true.

Notice again that when determining the truth value of a statement compositionally in this way, we do not make reference to the set of possible values of the components but use only their determinate or given values – values of T-free atomic sentences given, say, by a standard model. (The *T-free* sentences are those that do not *use* the truth predicate. A standard model used in this way is referred to as the *base* model.[14])

To determine whether 'Tα' is true, we will have to determine the truth value of the sentence to which 'α' refers, and to do that we may need to further decompose that sentence and so on until we arrive at T-free atomic sentences.[15] If the reference to another sentence in 'Tα' can be decomposed to T-free atomic sentences so that the values of these are sufficient to compositionally determine the value of 'Tα', then 'Tα' is compositional and has that determinate truth value; otherwise, 'Tα' is non-compositional. Compositional truths and falsehoods are a subset of *grounded* sentences, for *grounded* sentences as Kripke [1975] defines them supervene in similar and additional ways on the base model.[16]

Compositionality requires eliminability of the expressions referring to sentences, whereas Kripke's definition of groundedness requires eliminability of the T-predicate. The Liar statement and the Truth-teller are non-compositional. The references to my

---

[14] A *ground* model may extend the *base* model with more sentences. In the way I will represent Kripke's theory in Chapter 3, there are sentences that can be consistently added, like truth-tellers but not paradoxes; in Gupta & Belnap's theory, these sentences are an initial guess.

[15] It follows from our criterion for the truth-predicate that, like the identity predicate, the truth-predicate is not given an extension in the base model. This is consistent with our intuition that it is a predicate of interpreted sentences. Kripke's theory of truth allows for both the recursive evaluation of sentences and an initial extension for Truth in the ground model. Gupta & Belnap's theory of truth does the same in these respects. Kripke's recursive evaluation is such that the extension of truth is monotonic. Gupta and Belnap's recursion is not monotonic. Kripke's recursion reaches a fixed point for the extension of truth. Gupta and Belnap's reaches a pattern that one observes is a sort of generalisation of a fixed point. Cf. Herzberger's 'grand loop' [1982 / 1984, pp. 150-153].

[16] Or, perhaps, I should say truth-functionally compositional sentences *should be* a subset of grounded sentences. The valuations used in constructing a fixed point are usually ones that preserve truth-functional compositional valuations, such as Kleene's strong three-valued matrices and van Fraassen's supervaluation technique. The framing of Kripke's definition in relation to supervenience is due to Kremer [1988, p. 237-9].

favourite sentences can never be used to *drill down* to T-free sentences whose truth values are determined in a ground model. Being non-compositional, we have no basis for giving either of them any valuation in this way. Compositionally, we can determine no semantic value that distinguishes a Liar statement from a Truth-teller statement. Compositionally, a Liar statement and a Truth-teller are *indeterminate* (or lack a *determinate* value) as follows:

> My favourite sentence is true iff 'my favourite sentence' refers to a well-formed sentence and it is not the case that 'my favourite sentence' refers to an object that satisfies the predicate 'is true'.

'My favourite sentence' is a name that refers to the well-formed sentence 'My favourite sentence is not true'. But to determine whether or not my favourite sentence satisfies the predicate 'is true', compositionally, all we can use is this very biconditional again.

The situation for my second favourite sentence, a Truth-teller, is essentially the same.

> My second favourite sentence is true iff 'my second favourite sentence' refers to a well-formed sentence and 'my second favourite sentence' refers to an object that satisfies the predicate 'is true'.

Likewise, 'My second favourite sentence' is a name that refers to the well-formed sentence 'My second favourite sentence is true'. But to determine whether or not my second favourite sentence satisfies the predicate 'is true', compositionally, all we can use is this very biconditional again.

We can theoretically solve the Liar paradox by restricting truth to compositionally determined truth values; but such a solution has three significant drawbacks. Firstly, it is a significant restriction on logic that excludes consideration of possible truth values; secondly it is incompatible with belief in bivalence; and, thirdly, it is natural to think that then a Liar statement is *not* true.[17]

My third favourite sentence is 'Either my third favourite sentence is true or it is not true'. It is an *intrinsic* truth, but we cannot determine this on a compositional basis.[18] My fourth favourite sentence is 'My fourth favourite sentence is true and not

---

[17] It is equally rational to think that then a Liar statement is *not* false. Rieger [2001] explains that this also leads to a strengthened Liar using an F-schema in place of a T-schema.

[18] This notion of an *intrinsic* truth is introduced by Kripke [1975, p. 74].

true'. Although it is an *intrinsic* falsehood, its truth value cannot be determined compositionally. My third and fourth favourite sentences have the forms of what are classically a tautology and a contradiction respectively, yet their truth values are compositionally indeterminate. Thus, not all instances of what would be usually tautologous schemata or their negations receive a truth value compositionally.

The restriction of logic to compositional valuations in natural language is not without interest. It is consistent. It is a logic of contingency. Not every instance of schemata for common logical principles is evaluated. As long as there are indeterminate sentences, there are no logically valid schemata. The truth values of all sentences determined as true or false in this way are determined by contingencies. If a sentence receives a truth value compositionally, it does so because it is a complex sentence or composite of sentences that receive truth values in the ground model.

Among the indeterminacies are my first, second, third and fourth favourite sentences. It is because *intrinsic* truth values cannot be determined, that there are no purely formal truths, no tautological forms.

The moment we move beyond purely compositional evaluation to consider possible truth values (restricting these to true and false), we quickly find that the Liar is *over-determined* – whichever value it takes, the above biconditional will also give it the other, and the Truth-teller is *under-determined* – it can consistently take one of either value but there is no basis for choosing which it has.

In summary, my favourite sentence and my second favourite sentence are *non-compositional* and *indeterminate.* (In Chapter 3, I will compare both these valuations to Kripke's theory of *ungroundedness.*[19]) These valuations are not simply a matter of self-reference, that my favourite sentence refers to itself, but that its truth value can never be reduced to a composite of truth values of T-free atomic sentences. Nevertheless, when one considers the alternative, given bivalence, my favourite

---

[19] *Ungroundedness* was named by Herzberger [1970] and first given formal definition by Kripke [1975]. I have introduced *groundedness* as an extension to compositional semantics, but Kripke's definition is more flexible. That it is contingent whether some sentences are problematic was noted by Prior [1958] and brought to the foreground by Kripke in the above work. I think its referential analogue is Ryle's [1954] requirement of a *namely-rider.* If one makes a claim that a statement is true using a name or definite description to refer to a statement, then one should be able to provide that statement in a way that ultimately eliminates references to further statements. Ryle's restriction is thus related to Gupta & Belnap's distinction between circular and non-circular reference, not that Gupta & Belnap would put it that way.

sentence is *over-determined.* My second favourite sentence is *under-determined.* We will use these terms in assessing our categorisation.

2.7 Metaphors for Truth

Indeterminacy is not a gap, just a lack of a way of deciding. Naively, one talks about a Truth-teller as having an indeterminate truth value. Gap theorists act as though truth were a determinable, like 'is red'. Anything red is coloured; but not everything is coloured. For a gap theorist, it does not make sense to talk about a paradoxical sentence as *having* a truth value.

Glut theorists, from a naive point of view, act as though they have a rule for truth and a rule for falsehood, which do not always exclude each other.[20] From their own point of view, glut theorists are *paraconsistent.* It makes sense to talk about a paradoxical sentence as having both truth values. Some glut theorists value the Truth-teller as having both truth values; some value the truth-teller as being neither true nor false. However, this is not necessarily a gap. With respect to the metaphor of having a rule of truth and rule of falsehood, the resultant extensions might overlap and not be exhaustive. In this way, it would make sense to talk about paradoxical sentences having both truth values and hypodoxical sentences having neither. Alternatively, one might adopt a mixed metaphor so that paradoxical sentences have both truth values, and perhaps hypodoxical sentences are in a truth-value gap.

2.8 Naive Valuations

While a naive theory of truth may be a misconceived, I have attempted in previous sections to say something about naive valuations, which I will now summarise and generalise for future use, and comparison with more sophisticated theories.

The naive position is driven by bivalence. There are only two truth values, true and false. These are naively thought to apply exclusively and exhaustively to a class of sentences, to which the Liar and the Truth-teller and their relatives are naively, at least on first encounter, thought to belong.

Contingencies settle the truth values of a large subset of sentences compositionally. I have assumed that such compositional evaluation is primary. It uses

---

[20] Cf. Priest [2004, pp. 26-27].

naive valuations that might be made explicit by the following rules (or very similar ones):

- An atomic contingent sentence has a truth value compositionally settled by reality. (How these are settled is a theoretical matter. One has in mind a standard model which settles contingencies in the standard way, the following valuations then need to be accommodated in the theory.)
- A negated sentence $\sim\mathcal{A}$ is compositionally true (false) iff $\mathcal{A}$ is compositionally false (true).
- A *T-sentence* of the form $T\langle\mathcal{A}\rangle$, is compositionally true (false) iff $\mathcal{A}$ is compositionally true (false).
- A *T-sentence* of the form $T\alpha$ where α denotes the expression canonically named by $\langle\mathcal{A}\rangle$ is compositionally true (false) iff $T\langle\mathcal{A}\rangle$ is compositionally true (false).
- A disjunction, $A \vee B$, is compositionally true iff at least one disjunct is compositionally true.
- A disjunction, $A \vee B$, is compositionally false iff both disjuncts are compositionally false.
- A conjunction, $A \& B$, is compositionally true iff both conjuncts are compositionally true.
- A conjunction, $A \& B$, is compositionally false iff at least one conjunct is compositionally false.
- A universal generalization, $\forall x\mathcal{P}x$, is compositionally true iff everything compositionally satisfies $\mathcal{P}$.
- A universal generalization, $\forall x\mathcal{P}x$, is compositionally false iff something compositionally satisfies $\sim\mathcal{P}$.
- An existential generalization, $\exists x\mathcal{P}x$, is compositionally true iff something compositionally satisfies $\mathcal{P}$.
- An existential generalization, $\exists x\mathcal{P}x$, is compositionally false iff nothing compositionally satisfies $\mathcal{P}$.

The Liar and the Truth-teller are compositionally indeterminate. Being compositionally indeterminate is not another truth value; it is a compositional valuation in the sense that it describes the result of evaluating them compositionally.

Naively, being compositionally indeterminate is not thought of as a truth-value gap, because the naive position still upholds bivalence and simply has not been able to determine a truth value compositionally.

Compositional valuation can be thought of in terms of the strong Kleene valuation scheme *with the following provisos*: that this must neither be thought to imply three values nor truth-value gaps.

Naively, a sentence whose truth value is not determined compositionally is still presumed to have a truth value. A subsequent attempt to settle which truth value is then made by considering possible truth values, based on the naive principle that it should be one of two values. Something like a supervaluation can be used to capture the consideration of possible truth values. Indeed, this will settle the truth values of some indeterminate sentences, in particular remaining instances of tautologies and contradictions that were not able to be determined compositionally. It does not settle the truth value of the Liar or the Truth-teller of course; but valuations of 'over-determined' and 'under-determined' are used to describe the results and to distinguish paradox from hypodox.

This second step in evaluating some sentences will be better exemplified in the next chapter, but I give the following brief description. Each indeterminate sentence is considered individually, it is assumed first to be true and then to be false. For each possible combination, a classical valuation is performed supplemented with a valuation corresponding to the canonical T-schema. (T-sentences involving non-canonical names can have their valuations related to the valuation of canonical T-sentences along the lines of the valuation rules given above for compositional valuations. Non-canonical names for sentences can also be related to canonical names in the semantics by the classical valuation rule for identities). If in each case, the sentence is allocated both truth values as a result of classically valuing the extended compositional semantics, then its valuation is over-determined. This was exemplified with the Liar in this chapter, but will become more interesting when it is exemplified by other expressions involving contingencies in the next chapter. If the sentence can consistently be allocated both truth values by allocating any combination of truth values to its compositionally indeterminate components it is under-determined. The Truth-teller is a case in point; I will give another example in the next chapter involving contingent and indeterminate components. If a sentence can consistently be true but cannot be consistently allocated the value false, it is true. If a sentence can consistently be allocated the value false but allocating it the value true results in inconsistency, then

the sentence is false. In this way, every compositionally indeterminate sentence will be valued as true or false, or receive a valuation of over-determined or under-determined.

2.9 Strengthened, vengeful, negative Liars

Barwise and Etchemendy [1987, p. 9] think that the Liar involves truth, reference and negation; and not much else. I have stressed that paradoxical arguments also need to be considered. One common consideration is the tendency of any proposed solution to encounter what seem to be new forms of the Liar. The foremost of these was called the *Strengthened Liar* by van Fraassen [1970, p.16]. In its simplest form it is exemplified by 'This statement is not true' (as opposed to the *Simple Liar* 'This statement is false'). The idea is that an attempt to address the Liar by saying it is neither true nor false might solve the Simple Liar and yet need ramification to solve the Strengthened Liar. The literature also talks about the Liar's *revenge*. Let us say that *strengthened* Liars are versions designed, as above, to pose problems or complications for a strategy that deals with the Liar by varying truth values (or by introducing truth-value gaps). Let us say that *revenge* Liars, such as 'This statement does not express a true proposition', are versions designed to pose problems or complications for a strategy that deals with the Liar by varying the bearers of truth (or denying that such statements express or otherwise relate to a bearer of truth). These two strategies are not exclusive. A third strategy involves trying to stipulate a restricted form of negation, but negation will not stay fixed for anyone in natural language. Issues with the interpretation of negation are involved in most strengthened Liars, and perhaps also in revenge Liars.

Say the Liar does not have a truth value. Then reasoning about '$a = \langle{\sim}\mathrm{T}a\rangle$' ought to continue '*a* does not have a truth value; therefore, *a* is not true: but that is what *a* says; so, *a* is true; but if *a* is true, then it is not true.'

I note that there are also 'Truth-teller paradoxes' that arise in response to theoretical moves to address the semantic value of the Liar [Mortensen & Priest 1981]. I observe how these parallel strengthened Liar paradoxes, in particular, from the assumption that the Truth-teller is neither true nor false.

Since I assume bivalence to classify the Liar-like paradoxes as they originate, I will not generally be classifying strengthened and revenge Liars that arise in response to particular sophisticated theories.

2.10 The Concept of Paradox

Naturally, in classifying paradoxes, one wants a way of individuating types of paradox. The schemata involved have seemed a relatively intuitive way of individuating families; it will be convenient to talk this way without prejudicing questions about whether these paradoxes have a common form or whether the paradoxes of truth and satisfaction are of a different type from paradoxes of membership. What concerns me in this section is how to individuate paradoxes within a family; such as, if they should be distinguished, the Liar from the Epimenides. I want to answer this question *via* a long detour aimed at clarifying the relationship between paradoxical sentences and arguments, and the definition of 'paradox'.

2.10.1 Premise and Paradox

I now make a particular observation that I conjecture holds generally.

Observation 1: The Liar sentence is not a premise of the Liar argument.

The identity at line 1 in my presentation of the Liar is *compositional.* Its truth value is determined in the normal way by the co-reference of two terms, being the definite description 'my favourite sentence' on the left-hand side and the quotation name of that sentence on the right-hand side. My favourite sentence is not the premise at line 1; rather, it is mentioned in the premise. In general, the Liar sentence is neither a premise nor is it asserted prior to reaching a conclusion. A derivation of the Liar argument merely requires the Liar sentence to be a grammatically well-formed sentence. That is, the Liar sentence is not relied upon in naive arguments to satisfy any other semantic or pragmatic criterion in addition to being a meaningful, syntactically well-formed, indicative sentence. Given that it meets these latter criteria, it is naively suitable for being assumed true or false. Premises are *asserted* (when an argument is used), assumptions are not. Assumptions do not equate to assertions any more than antecedents.[21] Some theorists think that effective assertions must use sentences that are

[21] This distinction may affect the validity of some inferences; indeed, some inferences such as Necessitation can be made only from theorems and similar restrictions on TI and TE feature in the ratiocinations of some solutions. In related work by Kaplan & Montague [1963] and McGee [1991], a set of sentences are proved inconsistent under certain conditions that relate to valid inferences. The set of sentences that is closed under logical consequence is intended as a theory. Are premises to be treated as part of the theory? Basically, if the set of sentences is the set of truths (or known truths), we can count the identity premise about my favourite sentence as a theorem. However, if the set of

or can be true or false (or express truths or falsehoods), and that the Liar sentence cannot be effectively asserted. The Liar sentence is *asserted* in a conclusion, at least the conjunction of the Liar sentence and its negation is asserted in the conclusion. Moreover, its earlier assertion would not be required even if the vinculum were moved down a line or two. Consider the two conditionals that make up the biconditional at line 3. One conditional, 'If my favourite sentence is true, my favourite sentence is not true,' is derivable from a conditional proof assuming the antecedent. The Liar sentence is only *asserted* under an assumption in the conclusion of that conditional proof. The Liar sentence itself is at most *conditionally asserted* in the first conditional. This might be a basis for restricting the use of this conditional in the Liar argument, but it needs more of a justification. The conditional proof of the converse conditional merely assumes the Liar sentence, it does not *assert* it. In this conditional, it is an *unasserted* antecedent[22] An important corollary of this observation is that the question of the truth value of a Liar sentence or statement could be resolved independently of resolving the paradox.[23]

> Corollary 1: Resolving the truth value of the Liar sentence does not make the Liar argument unsound.

Generally, Liar-like paradoxes *embed* or refer to certain potentially problematic sentences in their derivations, and both arguments and sentences are called *paradoxical. Embedding* a paradoxical sentence is assuming it or using it in a complex formula, as in the right-hand side of the T-biconditionals in my derivation of the Liar in the previous chapter.

---

sentences is the set of necessary truths, our premise is not a necessary truth, and though true, is not a theorem of a theory of necessary truths.

[22] I *am* sympathetic to the point of view that asserting my favourite sentence is self-defeating. When put forward as being the case, its content says it is not the case. Buridan criticises a similar view in his *Sophismata*, though he describes it as an approach that he once held [Hughes 1982, p. 67]. My own criticism is that this in itself does not resolve the Liar paradox, because my favourite sentence is not used as a premise or a theorem, and so, it is not asserted in the proof. Therefore, even if this approach tells us something about the truth-value of the Liar sentence, something more is needed to resolve the paradox.

[23] One could take Kripke [1975] and others this way – as a semantics that needs to be supplemented with a proof theory and also with a semantic account of validity / invalidity. The two notions of consequence, proof-theoretic and semantic, should match, but we need the semantic notion to tell us what the project is about.

2.10.2 Conclusion and Paradox

I now make a particular observation intended to refute a view (implicit in some literature) that a paradoxical sentence is the conclusion of a paradoxical argument.

> Observation 2: An Epimenides sentence is not necessarily in the conclusion of an Epimenides argument.

The conclusion of the Liar is a compound of the Liar sentence of the form @ & ~@. However, a proof of @ iff ~@ does not necessarily mean that @ is the paradoxical sentence involved in the proof. For example, take a variant of the Epimenides paradox, which proves a contingently false but arbitrary ~Q from an identity like my favourite conjunction. My favourite conjunction just happens to be 'My favourite conjunction is not true and Q'. One could add the negation of the conclusion, Q, as a premise and derive the conclusion Q iff ~Q. However, Q was some contingent falsehood, not the paradoxical conjunction.

> Corollary 2: The conclusion of a paradoxical argument cannot effectively be used to single out the paradoxical statement involved in the argument.

I note that this is not just because the argument goes too far. Although the paradoxical conditional involved in Curry's paradox is generally derived on the way to the conclusion, this is not the case in the derivation of the above variation of the Epimenides.

2.10.3 The Concept of Paradoxical Argument

In this section, I briefly discuss the concept of a *paradoxical argument* with particular reference to the argument about my favourite sentence. An analysis of *paradoxical argument* is wanting. We do not yet have the classification that I wish to use to analyse various definitions in the literature. But I will introduce the subject here, and put forward my working definition.

At first though, there is a popular definition, I wish to acknowledge and set aside:

> A paradox consists of a set of sentences – a paradox set – each sentence of which is plausible but which, together, yield a contradiction.
>
> [Armour-Garb 2005, p. 87]

Actually, not all paradoxical arguments yield a contradiction. Also, this definition seems to ignore the role of inferences. Not all inferences can or need be translated into premises.[24] It seems to me that when we encounter paradoxes, we are not concerned whether we use schemata or rules of inference; in this example, we use an instance of the T-schema for a derivation. We could instead use rules such as T-Introduction and T-Elimination to derive the same conclusion. Furthermore, from a naive point of view, both the singleton set consisting of the identity statement of my favourite sentence and the union of this set with the T-biconditional for the Liar are paradoxical. These are distinct sets. They are associated with the one paradox, but the above definition implies they are distinct paradoxes. Nevertheless, these days a paradox is often presented as a set of incompatible premises. This is considered more economical. Thus our Liar is the set:

$$\{\text{‘}a = \langle\sim\mathrm{T}a\rangle\text{’}, \text{‘}\mathrm{T}\langle\sim\mathrm{T}a\rangle \text{ iff } \sim\mathrm{T}a\text{’}\}$$

Notice that none of the members of this set is the paradoxical Liar sentence, ‘$\sim$Ta’. (It is merely mentioned in the identity statement that is a member of this set, and it is embedded in the biconditional.) Therefore this definition throws no light on the relationship between the paradoxical Liar sentence and these paradoxical sets. When we encounter the Liar *naively*, the instance of the T-schema is not used or at least not identified as a premise – yet this *economical* presentation of a paradox presumably compels us to add a premise or something to the premise set for the Liar. Drawing on these last two points, the set with the identity premise and the T-schema looks like a different paradox from the set that distinguishes the Liar above, and both of those look like a different type of paradox from the set consisting of the identity premise and two principles: the principle of Truth Introduction and the principle of Truth-Elimination.

It is folklore that a *paradox* is or relates to an argument from seemingly true premises using apparently valid reasoning to a false conclusion, usually a contradiction. Therefore, it is held, the solution to a paradox is to show that the premises are actually not all true or an inference is actually invalid or that the conclusion is actually true.[25]

---

[24] Cf. Sorensen [2003, p. 366-368] on this point.

[25] Quine [1962 / 1976] distinguished *veridical* paradoxes, like the one that they sing about in Gilbert and Sullivan's *The Pirates of Penzance*, where there is some consternation over when someone born on 29th February will turn 21; *falsidical* paradoxes, like de Morgan's "proof" that 2=1; and antinomies,

The definition is vague because of the use of words like 'seemingly' or 'apparently'. If we thought a premise was true but it turns out to be false, we may recognise the argument as a valid *reductio ad absurdum*. If we thought the conclusion was false but it turns out to be true, it may have been a matter of disambiguation, or we may have to change some theory, even a physical theory.

A *logical* paradox is not an epistemic confusion about the truth value of the premises or the conclusion however. Whenever the premise of a *logical* paradox turns out to be false, some corresponding hitherto-thought-to-be-valid inference turns out to be invalid. In particular, if an instance of the T-schema is falsified by the argument of the Liar taken as a *reductio*, then an instance of TI or TE must also be invalid. On the other hand, if schemata like the T-schema are treated as axioms or derived using rules of inference, a logical paradox is intuitively semantically invalid but proof-theoretically valid. There are two competing intuitions for developing this thought, a modal and a material one.

On the modal account, for a logical paradox, we can imagine a possible world in which the premise or premises are true and the conclusion is false, but there is a seemingly proof-theoretically valid derivation of the conclusion from the premise. On this account, the Self-referential Curry, which proves some arbitrary Q, and the Unquantified Epimenides, which can prove ~Q, are both paradoxical.

On the material account, for a logical paradox, in this actual world, the premise or premises are true and the conclusion is false, but there is a seemingly proof-theoretically valid derivation of the conclusion from the premise.

Kripke convinced us, perhaps too well, that it is often an empirical matter whether a statement is paradoxical or not. Nevertheless, Kripke's account is material, not modal. A sentence either is paradoxical or it is not. Whether it is may depend on the empirical facts.

2.10.4 Types of Paradoxes

Paradoxes are not necessarily premises or conclusions, or even interim steps in paradoxical arguments or their derivations. The concept of paradoxical argument may be intractable. Is there an effective way of identifying the paradoxical sentence involved in a paradoxical argument? How does one distinguish types of paradox? My

---

which '[establish] that some tacit and trusted pattern of reasoning must be made explicit and henceforward be avoided or revised' [p. 5]

interim conclusions are these. Firstly, paradoxes are among the sentences whose semantic value is compositionally indeterminate. Among these, paradoxes are those whose semantic value is over-determined using a top-down valuation; some are contingently so. Also among these, the hypodoxes include the truth teller which is under-determined using a top-down valuation whatever the circumstances and other expressions that are contingently so. The paradoxes are distinguished by their possible semantic valuations, in line with their consequences. A Liar sentence is provably both true and false. An Epimenides sentence is provably false. A Curry sentence is provably true. Associated with each of these are paradoxical arguments, which for the latter involve contingencies that would make the semantic values of the paradoxical sentence over-determined. These elements of classification apply to hypodoxes as well. The truth-teller is compositionally indeterminate and *under-determined* by a top-down valuation, whatever the contingent truths may be. Nevertheless, in the next chapter I shall show there are expressions that may be *under-determined* depending on the contingent circumstances.

Let me prepare to end this chapter by exemplifying the application of these criteria with an additional paradox of my own. This is a distinct paradox from the others because it can be true or false but still used to prove an arbitrary contingent statement. Consider my favourite biconditional, which happens to be

'My favourite biconditional is not true iff Q'.

Note first that my favourite biconditional is compositionally *indeterminate*. The contingent semantic value of Q is not sufficient to construct the value of my favourite biconditional. Note secondly, given the identity of my favourite biconditional, it follows that ~Q is the case. For if my favourite biconditional is true, then the biconditional is true and ~Q follows by sentential logic using *modus tollens*. On the other hand, if my favourite biconditional is not true, then by sentential logic 'My favourite biconditional is not true iff ~Q' is true: ~Q then follows by sentential logic using *modus ponens*. Whether my favourite biconditional is true or not, ~Q is the case. Note thirdly, that if Q is true then, now using a top-down valuation about the possible semantic values of my favourite biconditional, it is *over-determined*. I note also that if Q is false, then my favourite biconditional is *under-determined* based on such a top-down valuation.

2.10.5 Individuating Paradoxes

What individuates my favourite sentence and an associated argument as an example of the Liar paradox? If it were just the use of the T-schema or associated rules of inference, then there would be no reason to distinguish the Epimenides or Curry's from the Liar. However, as discussed in Chapter 1, the Epimenides and Curry's in some sense cut across paradoxes of truth and membership, and neither the Epimenides nor Curry's need result in a contradiction but may still prove a falsehood or arbitrarily many of them. So, there are good reasons to distinguish the Epimenides and Curry's paradoxes from the Liar.

The Epimenides and Curry's sentences are, in a sense, contingently paradoxical. However, the Liar is, in a sense, only *intrinsically* paradoxical given the identity premise (or something equivalent).

Thus, the identity statement is characteristic of the Liar, given the canonical T-schema. This is to say that the Liar has a certain form. It is a sentence which attributes falsity to the reference of a term that refers to the sentence itself. The term in the sentence and a canonical name of the sentence are co-referential. This co-referential relationship is distinctly represented by the following identity statement.

$$a = \langle \sim \mathrm{T}a \rangle$$

This statement and the relevant instance of the canonical T-schema are sufficient to derive a contradiction. That is, they are sufficient to derive the Liar.

Initially, I treat this identity statement as if it were sufficient to individuate the Liar. Therefore, I use identity statements like these to individuate paradoxes. When I say that I use this identity and the T-schema to individuate the Liar paradox, I do not mean that I *give* the identity as a definition. Rather *given* an identity of this form and the T-schema, a version of the Liar can be derived.[26]

---

[26] I use an identity statement, because, as I will argue, I think the use or implicit use of Leibniz's law (or something like it) in the proof of Liar-like paradoxes is distinctive of a type of paradox. There is a slightly simpler idea for individuating paradoxes in the literature. The sentence used and mentioned on either side of the biconditional instance of the T-schema may be used to individuate the Liar, (semantic) Curry's, and other paradoxes. Associated with this idea, Tennant [1982] uses loops in (attempts) to normalize proofs to identify paradoxes. Rogerson [2006, p. 172] uses a derivation of Curry's paradox to provide a counter-example to this latter means of identifying paradoxes through proof structures.

There are issues about identifying the Liar sentence with the sentences that instantiate such an identity. As a strictly formal criterion for individuating Liar sentences, it is too narrow. I would not want to consider

$$a = \langle \sim\sim\sim \mathrm{T}a \rangle$$

to represent a different form of paradox, nor any of the following:

$$a = \langle \sim \mathrm{T}a \ \& \ \sim \mathrm{T}a \rangle$$

$$a = \langle \sim \mathrm{T}a \vee \sim \mathrm{T}a \rangle$$

Simply appealing to the formal equivalence of these embedded expressions is suspect when dealing with paradoxical sentences. Therefore, ultimately I will use another semantic criterion to assure the individuation of the variations I identify.

Finding which other identities together with the T-schema may result in paradox will be a theme for the next chapter. These identities relate to statements that have a possible over-determined valuation. 'If this sentence is true, Sydney is the capital of France' is over-determined when Sydney is not the capital of France. If we do not add the fact as a premise, the argument simply concludes with a false conclusion, nevertheless, the sentence 'If this sentence is true, Sydney is the capital of France' is over-determined under these circumstances.

2.11 The Concept of Hypodox

The Truth-teller was described as *hypodoxical* in Chapter 1. Consider again my second favourite sentence.

My second favourite sentence = 'My second favourite sentence is true'.

Characteristic of a Truth-teller, my second favourite sentence can consistently be assumed true or false, but there is the lack of a good reason to say whether it is true or false. That very lack of a reason is associated with a lack of an argument too, unlike the paradoxes. Associated with paradoxes of satisfaction, denotation and set-theory are predicates, definitions and sets used in expressions with similar characteristics. It is useful to have a general term, and I have coined the term *hypodox*.[27] (It is convenient

[27] I first coined the word 'hypodox' in my 'Paradoxes of Truth, Satisfaction and Membership' paper presented at the 2005 *Australasian Association of Philosophy* (AAP) Conference. I used it again in my 'Paradoxes and Hypodoxes of Time Travel' paper presentation at the 2006 AAP Conference. It

also to have an adjective *hypodoxical*.) The Truth-teller is a hypodox. Whether the predicate 'is true of itself' is true of itself is a hypodox. Mackie [1973, p. 298] gives an interesting list of what he calls *Truth-teller Counterparts*, including:

> Obey this order.
> 'Yields a truth when appended to its own quotation' yields a truth when appended to its own quotation.
> The largest number named in this book.
> Is the class of all classes that are members of themselves a member of itself?

I have argued above that the semantic value of the Liar and the Truth-teller is compositionally indeterminate. However, when one considers possible truth values, in the case of Liar-like paradoxes, their semantic valuation is over-determined, in contrast, the semantic valuation of Truth-teller-like hypodoxes is under-determined. (The prefix 'hypo-' means under). 'Over-determined' and 'under-determined' are not truth values, but describe the result of trying to determine the semantic values of these expressions in this way. More generally, the semantic valuation of other expressions, such as predicates, may be said to be over-determined, for paradoxical expressions, and under-determined for hypodoxical expressions. So that the truth value of

'is not true of itself' is not true of itself

is over-determined and paradoxical, and so, the semantic value, i.e. the extension of

is true of itself

is over-determined in this case. Similarly, the truth value of

'is true of itself' is true of itself

is under-determined and hypodoxical, and so, the semantic value, i.e. the extension of

is true of itself

is under-determined in this case.

Moreover, in *Paradoxes and Hypodoxes of Time Travel* [2007], I argue that the concept extends to classifying a similar phenomenon related to paradoxes of time travel.

---

appeared on the internet in my abstracts for the 2005 and 2006 conferences. I first used it in a publication in Eldridge-Smith [2007].

2.12 Conclusion

In this chapter, I have characterised the naive evaluations of the Liar and Truth-teller, discussed the version of the T-schema that I will predominately use, argued that I can remain ecumenical about the bearers of truth while classifying the paradoxes and hypodoxes, and progressed the individuation of Liar-like paradoxes and hypodoxes.

I am working in the same spirit in which Herzberger [1982] began 'Notes on Naive Semantics'. Herzberger, however, was concerned to argue towards a theory and made recourse to levels in his valuations to address 'semantic instability without contradiction or incoherence' [Herzberger 1982/1984, p. 134]. I am classifying, not resolving. I have deliberately sought to remain agnostic about the *correct* valuations of paradoxical and hypodoxical sentences, and instead classified the situations one reaches in trying to evaluate them naively as *compositionally indeterminate*, and, when considering their possible truth values, *over-determined* and *under-determined.* As these are not truth values but descriptions of the ways in which our naive methods of evaluating are consternated, I do not proceed to use Explosion or Strengthened reasoning.

I have taken some steps towards individuating the paradoxes and hypodoxes. This will be an ongoing theme as it correlates with their classification, which is the subject of my investigation.