Chapter 5: The Liar and Grelling's Paradoxes as related to the Inconsistency of Naive Truth and Satisfaction

The Liar and Grelling's paradox have been presumed to be closely related. As the inconsistency of naive truth was proved by adapting the Liar into a *reductio*, the inconsistency of naive satisfaction is proved by adapting Grelling's paradox into a *reductio*. Although guided by the proof of the former, the proof of the latter is distinct in two ways. Firstly, less logical machinery is used. Secondly, I show reflexivization is required. These differences arise in the logical principles necessary for the paradoxes themselves. I reflect therefore on the relationship between the two paradoxes. It was plausible that the Liar stands to the T-schema as Grelling's paradox stands to the satisfaction schema. I prove that this is not so. Moreover, I present a new Liar-like paradox of satisfaction. I conclude that the Liar and Grelling's paradox are of different types.

5.1 Introduction

Antinomies can be construed as dual *reductios*. If 'This sentence itself is not true' is assumed to be true, it is false; but if it is false, it is true. So runs the Liar, backed by logical principles, grammar, and intuitions relating to the use of 'true'. So too, 'does not satisfy itself' satisfies itself if, and only if, it does not. So runs Grelling's paradox, based on: logical principles, grammar for talking about the predicate itself, and principles relating to the use of 'satisfies'. When deriving the paradoxes in natural language, we are free of some systematic constraints to explicitly exhibit whether we use inference rules or axiom schemata. Treating a result of certain rules or an instance of a certain schema as an assumption, we may attempt to convert the paradoxes to *reductio ad absurdum* proofs of the invalidity of at least one of those rules or of that schema. So, a Liar-like proof may be applied to invalidate the T-schema, every instance of which is a theorem for a naive theory of truth: thus Tarski [1935, pp. 247ff.] argued that any formal system that attempted to encapsulate the naive notion of truth would be inconsistent. Naive truth is thereupon
considered an inconsistent or incoherent notion [Quine 1962, pp. 5-7]. So too, a Grelling-like argument may prove that any formal system that attempted to encapsulate the naive notion of satisfaction would be inconsistent, and naive satisfaction may thereupon be considered an inconsistent notion. Comparing these two proofs facilitates comparing the ingredients for the two paradoxes.

The derivation of a contradiction from expressions such as ‘This sentence itself is not true’ is typical of the Liar; whereas using expressions such as ‘does not satisfy itself’ and ‘is not true of itself’ are typical of Grelling’s paradox. The truth relation, ‘is true of’ is of course the converse of satisfaction. It might seem then that these paradoxes of truth are simply related, the Liar as a paradox of the truth predicate and Grelling’s as a paradox of the truth relation. Two ways this might be formalized are either if the Liar and Grelling’s can be mapped to each other using the relationship between the truth predicate and the truth relation, or if the Liar is simply a paradox of sentential expressions involving truth and Grelling’s a paradox of predicates involving truth. However, I will show the natural relationship between the truth predicate and truth relation does not map the Liar into Grelling’s or vice versa. Moreover, I will present a new Liar-like paradox of the truth relation which uses distinct expressions from those typical of Grelling’s.

Furthermore, I find differences in the logical principles involved in the proofs. If the reductio proofs were after all invalid, such invalidity would resolve the paradoxes. One or more commonly accepted principles of logic would themselves then be the subject of valid reductios that use rather than disprove the schemata or inferences that the so-called naive notions of truth and satisfaction license. One should like to know which principles are involved. My comparison focuses on whether they are the same for the Liar and

---

1 One might reject the application of truth to sentences; but then ipso facto one ought to reject the application of satisfaction to pairs of objects and predicates. But the burden of proof is on those that would do so, for if objects have properties, then prima facie they satisfy the predicate that expresses that property. (If properties exist but predicates do not in some way express properties, how should we talk about them?) Thus, if an object does not satisfy a predicate, then the object does not have the property expressed by the predicate or the predicate does not express a property. Either way the predicate is not true of the object. If this is correct, the derivation of Grelling’s paradox is unaffected by the failure of Grelling’s predicate, ‘does not satisfy itself’, to express a property. I rely then on a naive grammatical point of view, truth is applied to sentences (at least in a secondary sense) and objects are taken to satisfy predicates; and this is how the paradoxes are presented in natural language.
Grelling's paradoxes. I focus on those ingredients required to prove a contradictory biconditional, $\theta \iff \neg \theta$. I will ignore other logical principles involved in deriving a contradiction of the form $(\theta \& \neg \theta)$ from $(\theta \iff \neg \theta)$ as I think they are common to the Liar and Grelling's paradoxes.

A symptomatic difference between the Liar and Grelling's paradoxes is that the Liar has a premise. But this can be circumvented. Through the use of Gödel's diagonal lemma and arithmetization, we can prove a language has a suitable sentence.\(^2\) An apparent difference is the Liar's reliance on ungrounded reference and, in special, paradigm cases, self-reference, where Grelling's paradox uses self-application. But, it has seemed to some that ungrounded application was the more general form, perhaps because Grelling's paradox has seemed to use a subset of the logical principles used to derive the Liar.\(^3\) Alternatively, self-application and self-reference have seemed like unimportant variants of each other.\(^4\) However, another difference between the two paradoxes will emerge on investigation. This difference concerns the logical principles necessary to derive any version of the Liar as compared to those required to derive any version of Grelling's paradox.

In the second section, I sketch the intuitive relationship between the Liar and the Grelling. As paradoxes of truth, they are intuitively related via schemata for the truth predicate and the truth relation respectively. In the third section, I present Tarski's theorem, and Gödel's diagonal lemma in relation to the Liar. In the fourth section, I prove the inconsistency of naive satisfaction reliant on the reasoning of Grelling's paradox, and compare the two inconsistency results, in particular over the logical resources used. I relate this comparison to a comparison of the two paradoxes. I propose a candidate solution of Grelling's paradox that does not resolve the Liar. It arises as another ad hoc

\(^2\) Neither Tarski [1935] nor Quine [1962] appeal to arithmetization when discounting this difference. Tarski [1935 / 1983, p.165, footnote 1] discounts it by reference to Grelling's, thereby assuming a relationship, one which is under examination herein. Quine [1962, p. 7] gives a version of the Liar, which strikes him as on a par with the antimony of Grelling's. Quine's Liar is derived by considering whether:

'Yields a falsehood when appended to its own quotation' yields a falsehood when appended to its own quotation.

\(^3\) Cf. Quine [1962, p. 7]

\(^4\) Cf. Priest [1994] passim, and Russell [1908, p. 224], who refers to 'a common characteristic, which we may describe as self-reference or reflexiveness'.

proposal, and has at this time merely Machiavellian virtue to recommend it. In the fifth section, I explore the possibilities of deriving the Liar using the truth relation and the Grelling using the truth predicate. This can be done. It counts against a folk intuition that the relationship is basically that the Liar involves the truth predicate where Grelling’s involves the truth (or satisfaction) relation. In the sixth section, I give some particular attention to Quine’s explication of the relation between the Liar and proving the inconsistency of naive truth. In the seventh section, I discuss how the inconsistency proofs are tied to the use of the truth predicate or truth relation. Thus there is a Liar-style inconsistency proof for both the notions of truth and satisfaction, and a Grelling-style inconsistency proof for each notion. In the eighth section, I explore the relation between substitution and the truth relation. In the ninth section, I briefly consider the relevant aspects of quantified forms of the Liar paradox. I conclude that the necessary conditions for the Liar are not the same as Grelling’s. There are two paradoxes: Grelling’s is not just another version of the Liar. (There are two appendices related to this chapter. Appendix A extrapolates the conceivable solution to Grelling’s towards a solution to Russell’s paradox; and Appendix B adds a further comment on diagonalization in relation to these two paradoxes.)

5.2 The T-schemata

The Liar and Grelling paradoxes may be related in their use of $T^n$-schemata. We have:

$$T^n((\phi_1, x_2, x_3, \ldots, x_n), t_1, t_2, t_3, \ldots, t_n) \iff \langle t_1, t_2, t_3, \ldots, t_n \rangle$$

where ‘$T^n$’ represents an n-ary truth relation, ‘$\phi$’ represents an (n-1)-place relation, ‘$t_1$’, ‘$t_2$’, ‘$t_3$’, ‘$t_n$’ represent closed terms, and the angle brackets represent a canonical naming device, like quotes or arithmetization.

Now, take a zero-place predicate, i.e., a sentence, then let ‘$T$’ represent the truth predicate, and the T-schema falls out as an instance:

$$T(\langle \phi \rangle) \iff \phi$$

e.g., ‘Snow is white’ is true iff snow is white.

Actually, this is a restricted form of the T-schema, which in its more general form may use non-canonical names:
Where \( n \) is a name for a sentence \( P \),
\[
T(n) \iff P
\]
The sentence in quotes four lines above is true
iff snow is white.

I will usually use the canonical T-schema unless I explicitly refer to the general T-schema. The T-schema is naively thought of as true of any well-formed sentence, or at least any meaningful, well-formed sentence. However, the Liar is derivable using an instance of the T-schema.

Then, take a monadic predicate, and the \( T^2 \)-schema falls out as an instance:
\[
T^2((Px_1), t_1) \iff Pt_1
\]
‘\( x_1 \) is white’ is true of snow iff snow is white.

Wherein, \( T^2 \) represents the truth relation, more specifically, the binary truth-relation, for we can continue to produce schemata for the \( T^3, T^4 \), and so on truth-relations. The notation ‘\( Px_1 \)’ suggests by convention that \( P \) contains \( x_1 \). While the schema is intended to apply to monadic predicates, the schema could be trivially satisfied by a sentential expression; but the schema does not apply to relations. The \( T^2 \)-schema is intuitively true for every monadic predicate for all terms for objects in the range of applicability of the predicate. However, Grelling’s paradox is derivable using an instance of the \( T^2 \)-schema (as I’ll show in a subsequent section).

We could go on to the \( T^3 \)-schema, etc., for which there are associated Liar-like and Grelling-like paradoxes. We might think of the T-predicates and \( T^a \)-relations as devices for disquotation, as Quine might put it; but we also need to take care that the T-predicate only disquotes sentences, the \( T^2 \)-relation only disquotes monadic formulas or sentences, the \( T^3 \)-relation only disquotes dyadic or monadic formulas or sentences, etc.

I note again that the \( T^2 \)-relation is the converse of the dyadic satisfaction relation. I gave an example of Grelling’s paradox using the satisfaction relation in the introductory paragraph. There is a Sat-schema:

---

5 With respect to Grelling-like paradoxes, see Simmons [1993, pp. 40-42]. With respect to Liar-like paradoxes, I’ll present one using the \( T^3 \)-schema in due course.

6 Cf. Tarski [1935 / 1983, p. 190]. My so-called Sat schema is an instance of the more general:
\[
(y) \ (Sat(y, (Px_1)) \iff Py
\]
for which read: for all \( y \), \( y \) satisfies ‘\( Px_1 \)’ iff \( Py \)

Cf. also McGee [1991, p. 31].
Chapter 5  The Liar and Grelling’s Paradoxes  192

Sat(t₁, ⟨Px₁⟩) iff Pr₁  e.g., Snow satisfies ‘x₁ is white’ iff snow is white.

I also note the following equivalences:

T²(⟨Px₁⟩, t₁) iff Pr₁ iff T⟨Pr₁⟩  e.g., ‘x₁ is white’ is true of snow iff snow is white iff ‘Snow is white’ is true.

From these equivalences, I isolate and label a schematic equivalence that will feature in later discussion, the T2T-schema:

T²(⟨Px₁⟩, t₁) iff T⟨Pr₁⟩  e.g., ‘x₁ is white’ is true of snow iff ‘Snow is white’ is true.

The T2T-schema can be derived from the T-schema and the T²-schema; and I think it represents the intuitive relationship between the truth predicate and truth relation, at least as far as subject-predicate sentential form is concerned. On the basis that it does represent this relationship, any restriction of the T-schema needs to be balanced by a restriction of the truth-relation. Indeed, it is plausible that the T2T-schema might provide a means of relating the Liar and Grelling’s. It would be interesting if the Liar could be derived using just the T²-schema without using the T-schema, or if Grelling’s paradox could be derived using the T-schema without using the T²-schema. Since these schemata are prima facie characteristic of the Liar and Grelling’s paradoxes, which otherwise seem very similar, the T2T-schema would seem sufficient to relate the two paradoxes. However, we shall find that there are other ways of characterising the Liar and Grelling’s paradoxes.

5.3  The Liar, Gödel’s Diagonal Lemma, and Tarski’s Theorem

We can get around the need for a premise in the derivation of the Liar using techniques introduced by Gödel. Tarski used these techniques and adapted the Liar argument as a reductio of (an instance of) the T-schema. In the next two subsections, I revisit Gödel’s Lemma and Tarski’s theorem. I subsection 3, I investigate alternatives to using Leibniz’s law. I conclude in subsection 4 that something like Leibniz’s law is necessary for the Liar. In the last two sections, I investigate whether or not diagonalization is necessary for the Liar.
5.3.1 Gödel's Diagonal Lemma

A language that can canonically refer to its own expressions and describe its own syntax cannot consistently contain all instances of the T-schema. The proof is often given using Gödel's diagonal lemma, perhaps more aptly referred to by Vann McGee as Gödel's "self-referential lemma" [McGee 1991, p. 24]. Using canonical names or grammatical encoding of expressions we can substitute a description which describes in fact the resulting sentence itself. To see how, consider a substitution function \( s(y, z) \), defined for expressions as arguments, which yields the result of substituting the canonical name of \( z \) for \( 'x' \) in \( y \); in other words, the result of substituting the canonical name of the value of the variable \( 'z' \) for \( 'x' \) in what \( 'y' \) stands for. If the value of \( 'z' \) is an expression \( e \), then

\[
s((\text{P}x), z) = \langle \text{P}(\langle e \rangle / x) \rangle
\]

where \( 'P' \) represents some predicate or relation, and as in the previous section, angle bracket expressions represent canonical names.

Let me first give some examples, for \( P \) representing \( 'x' \) barks', and \( z \) instantiated by \( 'Fido' \), then:

\[
s('x \text{ barks}', 'Fido') = 'Fido' \text{ barks}'
\]

The function may seem a little awkward. There may seem at first to be too many quotes; but this is the correct value of the function, although the value is a false sentence. The function is undefined in many instances, for example, because \( 'Fido' \) is not a canonical name for Fido:

\[
s('x \text{ barks}', \text{Fido})
\]

is not defined.\(^7\)

Nevertheless, \( s \) is tailored to talk about expressions. Consider:

\[
s('x \text{ has four words}', 'Cerberus \text{ barks a lot}') = 'Cerberus \text{ barks a lot has four words}'.
\]

---

\(^7\) For brevity's sake, I assume it is always \( 'x' \) that is replaced. A fuller substitution function would be a 3-place function, which would have the variable to be replaced as an additional argument.

\(^8\) Had I used a two-sorted logic, with variables used to define the function ranging just over expressions (and other variables over everything), then this instance of the function would be ill-formed.
In this instance, \( z \) is instantiated with a term which is the name of an expression, i.e.:

‘Cerberus barks a lot’

which is represented by:

\( \langle e \rangle \)

and canonically names:

Cerberus barks a lot

and which syntactically replaces ‘\( x \)’ in ‘\( x \) has four words’ to form the value of the function.

In subsequent sections it will be sufficient to restrict our attention to terms, where ‘\( t \)’ is used to represent a term, if \( n \) refers to \( t \),

\[
s(\langle P_\alpha \rangle, n) = \langle P[\langle t \rangle / x] \rangle
\]

In particular, ‘\( \langle t \rangle \)’ represents the canonical name of a term, \( t \), and we can use the following schema:

\[
s(\langle P_\alpha \rangle, \langle t \rangle) = \langle P[\langle t \rangle / x] \rangle
\]

I note in passing a similarity with the T-schema. The former schema with ‘\( n \)’ is the more general and allows for non-canonical names; but after this section we shall usually have in mind the less general schema with a canonical name, ‘\( \langle t \rangle \)’, as the second argument on the left-hand side of the identity.

It is assumed that the language under discussion contains this function, or at least that the function is expressible in the language. In particular, it is expressible in natural language, but also in more formal languages.

The substitution function’s arguments can name the same expression; so that the name of an expression is substituted into itself, i.e. free occurrences of ‘\( x \)’ in an expression are replaced with a name of the expression itself. Thus:

\[
s(\langle \text{‘x has four words’} \rangle, \langle \text{‘x has four words’} \rangle) = \langle \text{‘x has four words’ has four words} \rangle
\]
What is wanted now is a way of doing this so that the value of the substitution function is a name for the sentence in which it occurs. Here is a derivation of Gödel’s diagonal lemma:9

Take any one-place predicate, \( \phi(x) \).

In \( \phi(s(\langle \phi(s(x, x)) \rangle, \langle \phi(s(x, x)) \rangle)) \).

\[ s(\langle \phi(s(x, x)) \rangle, \langle \phi(s(x, x)) \rangle) = \langle \phi(s(\langle \phi(s(x, x)) \rangle, \langle \phi(s(x, x)) \rangle)) \rangle \]

So, by Sentential Logic (SL) and Leibniz’s law:

\[ \phi(s(\langle \phi(s(x, x)) \rangle, \langle \phi(s(x, x)) \rangle)) \text{ iff } \phi(s(\langle \phi(s(x, x)) \rangle, \langle \phi(s(x, x)) \rangle)) \]

If we abbreviate \( \phi(s(\langle \phi(s(x, x)) \rangle, \langle \phi(s(x, x)) \rangle)) \) using ‘C’ in the meta-language, we have proved:

For any predicate represented by ‘\( \phi(x) \)’, we can form a sentence abbreviated here by ‘C’, such that:

\[ C \text{ iff } \phi(\langle C \rangle) \]

As foreshadowed, this result can be used to derive the Liar. It can also be used to prove the inconsistency of naive truth.

5.3.2 Tarski’s Theorem and the Inconsistency of Naive Truth

A contradiction now follows if we assume every instance of the T-schema is a theorem. For by Gödel’s diagonal lemma there is a formula \( A \) such that:

\[ A \text{ iff } \neg T(\langle A \rangle) \]

---

9 This derivation of Gödel’s lemma takes the substitution function as primitive. I required a derivation that used canonical names without the complexity of Gödel numbering to facilitate comparison of my derivations of the Liar and Grelling’s paradoxes, and the inconsistency of satisfaction as related to Grelling’s. Peter Roeper showed me this particular derivation of Gödel’s lemma, which, like Quine [1995, pp. 238-241], does not require Gödel numbering. In hindsight, Gödel numbering provides at once both a reference function mapping names of expressions to their references and a canonical naming function mapping expressions to names of those expressions. Canonical naming is in some sense built into this substitution function. This is sufficient for Gödel’s lemma.
And, as an instance of the T-schema, we have:

\[ T(A) \text{ iff } A \]

A contradiction is derived by an immediate inference.\(^\text{10}\)

This proof relied on the substitution function, the T-schema and Leibniz's law.

Compare its derivation with the Liar:

<table>
<thead>
<tr>
<th>The Liar</th>
<th>Reductio of the Consistency of Naive Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (a = \langle \neg Ta \rangle)</td>
<td>(s(\langle \neg T(s(x, x)) \rangle, \langle \neg T(s(x, x)) \rangle) = \langle \neg T(s(\langle \neg T(s(x, x)) \rangle, \langle \neg T(s(x, x)) \rangle) \rangle)</td>
</tr>
<tr>
<td>Premise</td>
<td>Theorem</td>
</tr>
<tr>
<td>A term is co-referential with the name of a sentence in which it occurs</td>
<td>The substitution function provides a term co-referential with the name of a sentential expression in which the very same term occurs, or in any case, the identity statement is provable</td>
</tr>
<tr>
<td>2. (\neg Ta \text{ iff } T(\neg Ta))</td>
<td>(\neg T(s(\langle \neg T(s(x, x)) \rangle, \langle \neg T(s(x, x)) \rangle) \rangle) \text{ iff } T(s(\langle \neg T(s(x, x)) \rangle, \langle \neg T(s(x, x)) \rangle) \rangle))</td>
</tr>
<tr>
<td>SL, 1, Leibniz's law</td>
<td>Instance of Diagonal Theorem; or SL, 1, Leibniz’s law</td>
</tr>
<tr>
<td>3. (T(\neg Ta) \text{ iff } \neg Ta)</td>
<td>(T(s(\langle \neg T(s(x, x)) \rangle, \langle \neg T(s(x, x)) \rangle) \rangle) \text{ iff } \neg T(s(\langle \neg T(s(x, x)) \rangle, \langle \neg T(s(x, x)) \rangle) \rangle))</td>
</tr>
<tr>
<td>Instance of the T-schema (The T-schema is naively valid.)</td>
<td>Instance of the T-schema (All instances of the T-schema are assumed true for reductio.)</td>
</tr>
<tr>
<td>4. (T(\neg Ta) \text{ iff } T(\neg Ta))</td>
<td>(T(s(\langle \neg T(s(x, x)) \rangle, \langle \neg T(s(x, x)) \rangle) \rangle) \text{ iff } \neg T(s(\langle \neg T(s(x, x)) \rangle, \langle \neg T(s(x, x)) \rangle) \rangle))</td>
</tr>
<tr>
<td>2, 3 Sentential Logic</td>
<td>2, 3 Sentential Logic</td>
</tr>
</tbody>
</table>

In proving the Liar paradox, the instance of the T-schema is taken as given; whereas it is assumed for the reductio. The table shows that the Liar can be derived in analogy with the reductio. The proof of the Liar is made a little longer than necessary in order to align

\(^{10}\) Cf. McGee [1991, p. 25].
Chapter 5  
The Liar and Grelling’s Paradoxes

with use of an instance of the Diagonal Theorem. Conversely, the contradiction for a
*reductio* can also be derived more succinctly in direct analogy with the Liar, as follows:

<table>
<thead>
<tr>
<th>The Liar</th>
<th><em>Reductio</em> of the Consistency of Naive Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ( a = \langle \neg a \rangle )</td>
<td>( s(\langle \neg s(s(x, x)) \rangle, \langle \neg s(s(x, x)) \rangle) = \langle \neg s(s(\langle \neg s(s(x, x)) \rangle, \langle \neg s(s(x, x)) \rangle) \rangle )</td>
</tr>
<tr>
<td>Premise</td>
<td>Theorem</td>
</tr>
<tr>
<td>A term is co-referential with the name of a sentence in which it occurs</td>
<td>The substitution function provides a term co-referential with the name of a sentential expression in which the very same term occurs, or in any case, the identity statement is provable</td>
</tr>
<tr>
<td>2. ( T(\neg a) ) iff ( \neg a )</td>
<td>( T(\langle \neg s(s(s(x, x)) \rangle, \langle \neg s(s(x, x)) \rangle) \rangle \equiv \langle \neg s(s(s(x, x)) \rangle, \langle \neg s(s(x, x)) \rangle \rangle) )</td>
</tr>
<tr>
<td>Instance of the T-schema (The T-schema is naively valid.)</td>
<td>Instance of the T-schema</td>
</tr>
<tr>
<td></td>
<td>(All instances of the T-schema are assumed true for <em>reductio.</em>)</td>
</tr>
<tr>
<td>3. ( a ) iff ( \neg a )</td>
<td>( T(\langle \neg s(s(s(x, x)) \rangle, \langle \neg s(s(s(x, x)) \rangle) \rangle \rangle \equiv \langle \neg s(s(s(x, x)) \rangle, \langle \neg s(s(s(x, x)) \rangle) \rangle) )</td>
</tr>
<tr>
<td>1, 2 Leibniz’s law</td>
<td>1, 2 Leibniz’s law</td>
</tr>
</tbody>
</table>

There are trivial variations. For example, instead of substituting for identicals into the left hand side of the T-biconditional of line 2 in the last table, I could have substituted into the right hand side. This would have yielded exactly the same conclusion as the first table.

Alternatively, I could commute ‘\( T \)’ and tilde, which the T-schema licenses, as a variation on either of the above derivations, to obtain:

<table>
<thead>
<tr>
<th>( \neg T(\langle a \rangle) ) iff ( T(\langle a \rangle) )</th>
<th>( \neg T(\langle T(s(\langle \neg s(s(x, x)) \rangle, \langle \neg s(s(x, x)) \rangle) \rangle) \rangle \equiv \langle \neg T(s(s(\langle \neg s(s(x, x)) \rangle, \langle \neg s(s(x, x)) \rangle) \rangle) \rangle) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>using commutation of ‘( T )’ and ‘( \neg )’</td>
<td>using commutation of ‘( T )’ and ‘( \neg )’</td>
</tr>
</tbody>
</table>
Chapter 5  The Liar and Grelling's Paradoxes

So, we have variations on how to derive a contradiction. For each variation, a direct analogy can be drawn between the proofs. This is because a derivation of the Liar is used as part of the proof of the theorem of the inconsistency of naive truth, the part that proves a contradiction follows from the assumption. So, the proofs in the right hand side of the tables double as proofs of the Liar paradox. The tabulated variations down the right-hand side use an identity theorem, which is an instance of the substitution function. Possibly the most common variation is that which I presented at the beginning of the section, which uses an instance of the Diagonal Theorem. This variation may appear more succinct than the derivation of the Liar in the left-hand side of the first table; be that as it may, the Diagonal Theorem is derived using an identity theorem. I think the analogy with other proofs using empirical or stipulated identity premises is appropriate and evident. Thus, the derivation of Tarski's theorem parallels the derivation of the Liar paradox. And indeed, the derivation of a contradiction based on the substitution function or Diagonal Theorem can be used as a proof of the Liar rather than a reductio of the T-schema.

I dare suggest that one could prove the inconsistency of naive truth using an empirically true premise, as per the left-hand side. That is, given the identity premise, a proof could proceed as for the proof of the Liar except that the T-schema is assumed for a reductio. The result then is that either the premise or the instance of the T-schema line is not the case. If the plausibility of the premise is greater or stronger than one's intuition that all instances of the T-schema are true, one may be persuaded by reductio that an instance of the T-schema is not the case. It is difficult to assess the plausibility of an empirical premise relative to intuitions about an axiom schema.

The matter seems conclusive when the identity is given as a theorem in the right hand side; yet it is still a case of trading off intuitions about the T-schema against intuitions about the validity of Leibniz's law. We must uphold Leibniz's law to conclude that the T-schema is inconsistent. These derivations both depend on Leibniz's law, and the right-hand side depends on the substitution function, whereas the left-hand side depends on a given premise, perhaps empirically true or true by stipulation. Conjoined, these are sufficient logical principles for deriving the contradictory biconditional of the Liar. I am immediately interested in the logical principles necessary to derive the Liar, and subsequently to prove the inconsistency of naive truth. (Despite the analogy between the above proofs, these necessary conditions may turn out not to be the same if non-Liar-like
proofs can be given for the inconsistency of naive truth.) I am particularly interested in Leibniz's law. I will also comment on diagonalization in this connection.

5.3.3 Liars using Alternatives to Leibniz's Law

It may be objected that the identity is not required, either that it is not required as an explicit premise, or that even when it is used as a premise, Leibniz's law is not formally required.

Firstly, if in the semantics $a$ is a name for $\sim Ta$, the full power of the T-schema (as explained in section 1) would enable the derivation of the contradiction in one step.

1. $Ta = \sim Ta$ 
   T-schema, semantics

This is not a purely formal proof; nevertheless, it is akin to proofs in natural language. Something like this was perhaps what some people had in mind when they considered the proof of the Liar prior to Tarski's more formal analysis. This version of the Liar supports its being referred to as a "semantic paradox" as opposed to a purely formal paradox. Nevertheless, a proof in the right-hand side of either of the above tables taken as a proof of the Liar paradox, rather than a reductio, somewhat belies the title "semantic paradox" in that it does not rely on semantics at all. It is a formal proof of a contradiction.\(^{11}\)

I am considering whether Leibniz's law is a necessary logical principle in deriving the Liar; purely formal derivations are designed to exhibit the principles that are being used. Semi-formal or informal derivations may tacitly rely on certain logical principles. We should at least make the identity premise explicit, which leads to the second variant of this objection.

The objection need not discount the identity premise. The proofs on either side of the second table can do without using Leibniz's law by using the full power of the T-schema. Given the identity in line 1, the third line can be derived directly using the (general) T-schema as so:

---

\(^{11}\) "T" is intended to be interpreted as the truth predicate, just as "e" is intended to be interpreted as the membership relation for Russell's paradox. If we talk about semantic paradoxes as opposed to set-theoretic paradoxes, such labelling reflects the concepts represented, such as truth and membership. However, if we talk about semantic as opposed to logical paradoxes, the proof in the right-hand side of the table belies the distinction suggested by those labels.
Chapter 5  The Liar and Grelling’s Paradoxes

1. \( a = \langle \neg T a \rangle \)  Premise
2. \( T a \equiv \neg T a \)  Instance of the (general) \( T \)-schema

No formal use of Leibniz’s law is made here. Nevertheless, it seems to me that these “general” uses of the \( T \)-schema depend for their plausibility on something like Leibniz’s law in that they may be said to rely on co-reference of terms. I suggest that when Leibniz’s law is used explicitly, both it and the weaker “canonical” \( T \)-schema seem more plausible. Use of the general schema can be broken down or justified by use of Leibniz’s law and the canonical form of the \( T \)-schema.

If truth is introduced as a grammatical device represented by ‘\( T \)’, it may be defined by Truth Introduction and Truth Elimination (using canonical names). These two rules can be used to derive the canonical \( T \)-schema. Leibniz’s law can be used in combination with the canonical \( T \)-schema to derive the “general” naive \( T \)-schema. All this is purely formal.

If moreover truth is introduced as an extensional predicate (whether interpreted metaphysically as a property or not), i.e. truth is taken to be a predicate of an object, then in an interpreted language it does not matter how one refers to that object, which would warrant the general \( T \)-schema. While co-reference is not necessarily represented by an identity in the object-language what I say in the meta-language uses co-reference which has to do with an object that can be referred to in various ways; so truth is a property (predicate) of objects.\(^{12}\) Other derivations rely on indexicality, and would not normally be said to use Leibniz’s law of the substitution of identicals. When ‘this sentence’ in ‘This sentence is false’ is used self-referentially, a part refers to the whole sentence in which it occurs; the co-reference of ‘this sentence’ and the canonical name of the sentence is not necessarily represented by an identity. Nevertheless, this co-reference is sufficient to warrant substitution.

A purely formal proof can eschew talk of co-reference. The formal use of Leibniz’s law does not require this interpretation; but if I am to use the general \( T \)-schema, I need to justify it by reference to an interpretation, such as attributing truth to co-referring terms; whereas, formally, the ‘\( T \)’ in the canonical \( T \)-schema can be introduced merely as a device for disquotation.

\(^{12}\) In this way the Identity premise may be analytic.
In summary, while the general T-schema or indexical constructions would seem to obviate the need for using Leibniz's law, I have drawn the derivation out this way because it makes most explicit a reliance on co-reference, whether through identity, indexicality, a name in the semantics, or definite descriptions that contingently self-refer. (With respect to the latter, consider my favourite sentence, which still just happens to be 'My favourite sentence is not true'; or consider an abbreviation for a definite description of the sentence on line 3 of page whatever of some prestigious journal that happens to refer to a sentence using that same definite description and saying of its referent that it is not true.) Co-reference is a semantic matter. It may be represented formally in a number of ways. It is a bit too strong to say Leibniz's law itself is necessary for all formal derivations of the Liar. Nevertheless, an identity statement is a perspicuous way of representing and formally abstracting co-reference. Furthermore, in the canonical T-schema, 'T' can be introduced merely as a device for disquotation.

5.3.4 Something like Leibniz's Law is necessary for the Liar

With the above in mind, I maintain that even if all instances of the canonical T-schema were axioms, the proof of the Liar would still require something more, in particular Leibniz’s law of the substitution of identicals. On the one hand, the canonical T-schema is in a sense more fundamental than the general T-schema; it is logically weaker than the general T-schema. The canonical T-schema is more plausible than the general T-schema. And Leibniz’s law is given independently of intuitions about truth. From these two we can derive the formal version of the general schema, which requires an identity premise. The canonical T-schema is more purely formal. The general T-schema may use semantic facts about naming. By using the canonical T-schema and Identity we can represent the Liar more formally.

Skyrms [1970] championed a failure of substitution of identicals to preserve truth as a solution to the Liar. Appealing to truth-value gaps, Skyrms makes a sophisticated distinction between Leibniz’s law preserving truth and its avoiding falsity. In the case of the Liar, Skyrms claims it does the latter but not the former. This solution, like others, is not without its issues. In particular, it needs to be extended to accommodate other versions of the Liar, such as those using indexicals. For present purposes I simply note
the existence of such a proposed solution, and that it would invalidate Tarski’s theorem, at least as derived above.

If we do not restrict Leibniz’s law, I think we should accept Tarski’s theorem of the inconsistency of naive truth, at least as a valid reductio. In saying this, I assume the biconditionals in the last row of the table are contradictory. (Some non-classical logics will avoid the derivation of \((\& \& \sim \&\)) from the biconditionals.)

It could be objected that a non-classical logic might apply Leibniz’s law but still reject this reductio. There are at least two possibilities. The first is Skyrms’ own. For as Skyrms [1970] has it, Leibniz’s law will always avoid inference to a falsehood, but will not always preserve truth. Thus, Skyrms’ approach is to modify the definition of validity, as avoiding falsehood but then not necessarily assuring truth, as we may infer a sentence in the truth value gap from true ones. The second is to deny that a reductio always follows from a contradiction, perhaps because of some variant of a paraconsistent logic. However, in either case, it is difficult to conceive how this could be decided in this case without involving a petitio principii, i.e. how would one know that this reductio was invalid unless one already knew that a naive theory of truth was correct.

Furthermore, the reductio in this case need not rely on the law of non-contradiction, but rather can simply rely on the bedrock principle that validity preserves truth (in that it avoids taking a truth to a falsehood). Given the truth values true simpliciter, false simpliciter, both true and false, and neither true nor false, validity assures that a falsehood cannot be derived from premises each of which is true simpliciter.

No bedrock principle should remain unturned in the search for the solution to the Liar paradox. Indeed, Skyrms flexes the bedrock principle that validity preserves truth in favour of the weaker principle that it avoids falsehood in order to have substitution of identicals valid while not always taking us from truth to truth, but sometimes from truth to a gap. So, if Skyrms were correct, the instance of the T-schema is true, the identity premise is true, but the contradictory biconditional derived by Leibniz’s law is a truth-value gap. In this way, a contradiction is not derived, and there is no reason to conclude that the instance of the T-schema, which was assumed true, is not true. I note the relationship here between being ad hoc and begging the question. How could we come to be assured that this bedrock principle failed in this case without first coming to believe that substitution of identicals might be invalid (in the normal sense) in this case.
Nevertheless, maintaining the bedrock principle and the T-schema will lead to either suspecting identity or that dialethias invalidate *reductio*. But without a very weak logic it could not invalidate this *reductio*. This instance of the T-schema is assumed true, which assumption leads by merely Leibniz’s law to a falsehood (in the form of a contradictory biconditional) – even if the contradictory biconditional should subsequently prove to be a true contradiction, it is still a falsehood. Since validity is truth-preserving (at least in that it avoids taking a truth *simpliciter* to a falsehood), then the assumption must be (at least) false. (To deny that this biconditional is contradictory would be extremely counter-intuitive, particularly if the T-schema has been maintained on the basis of the Substitution Thesis.) Considerations of relevance are not at issue here either, as all premises and assumptions are used in the proof and no use is made of explosion (*ex falso vel contradictoni quod libet*).\(^\text{13}\) So, given Leibniz’s law, we have a valid *reductio* of the assumption of the truth of an instance of the T-schema. Given a valid *reductio* of an instance of the T-schema, the only way of upholding a naive theory of truth is to maintain that instance is a true contradiction. In which case, the very criterion of a naive theory of truth is not isolable from falsehood; but then it is no longer a naive theory of truth, any more than a gap theory. If one holds, despite Tarski’s proof, to a naive theory of truth *simpliciter*, one needs find a failure of Leibniz’s law (but not Skyrms’ particular account). There is a sense in which to accept Tarski’s theorem is not only to accept the validity of the argument but also to conclude that it is impossible to give a definition for naive truth. However, to accept that the *reductio* is valid is sufficient to conclude that a naive theory of truth is inconsistent.\(^\text{14}\)

---

\(^\text{13}\) None of the sorts of issues discussed in Read’s [2003] analysis of entailment in terms of truth-preservation with respect to relevance are at issue here.

\(^\text{14}\) Tarski’s [1935 / 1983] Criterion T may need clarification as to whether it allows for or excludes *dialethics* intuitions. I myself take it for granted that for a naive theory of truth each instance of the T-schema is true *simpliciter*. Given this, glut theories are more sophisticated, as are gap theories.
5.3.5 Liars with and without Diagonalization

By the diagonalization of an expression, we mean the result of substituting the quotation of the expression for every occurrence of the variable "x" in the expression.

[Smullyan 1994, p. 3]

The first variation of proving a contradiction might be of significance because it is obtained by using the Diagonalization Theorem. However, the second variation in the right-hand side of the second table (p. 197) shows that use of the Diagonalization Theorem is not essential. Yet many do appear to regard diagonalization as necessary for the Liar and the proof of the inconsistency of naive truth. What then is diagonalization?

Jacquette [2004, p. 68] characterizes Smullyan's [1994] definition of diagonalization as involving "self-intra-substitution", substituting the expression for a part of itself. Jacquette seems, in particular, to have in mind substituting for a variable; but it would not need to be a variable, I think. Diagonalization is involved in all the right-hand side proofs, because that is how the substitution function is used. These derivations do substitute the expression for a variable in the expression. Self-intra-substitution is also involved in deriving line 2 in the left-hand side of the first table (p. 196). Here, a canonical name for the sentence replaces a non-canonical name for the sentence in the sentence.

While the derivation can be re-arranged to use self-intra substitution, the alternative where it does not is significant in eliminating diagonalization as a necessary principle to derive the Liar.

The alternative proof in the left-hand side of the second table was not obtained by diagonalization; so, the use of diagonalization to obtain a contradictory conclusion is redundant. And if diagonalization is significant for obtaining a conclusion simply because it is used in one variation, surely there is something equally significant to say about the second variant conclusion. (Diagonalization is involved in the theorem that gives the identity in the first row on the right-hand-side, and I will discuss that use of diagonalization shortly.)

While many of those who relate the Liar to Grelling's and give the Liar as a paradigm case, do so through truth schemata, there are also many who relate the Liar and Grelling's but use Grelling's as a paradigm case, and claim both paradoxes involve diagonalization. One advantage of setting the derivation out as per the second table is that
we can see that diagonalization occurs in the premise in the right-hand side, and not in the instance of the T-schema. (It is not so separable in the case of Grelling's, as we shall see.)

But diagonalization is not involved in the premise for the derivation in the left-hand side, even though that derivation involves a sentence containing a term referring to the sentence. Paradox is not obtained on the left-hand side of the table by diagonalization. This Liar sentence is certainly not substituted for a variable in itself, nor is it even substituted into itself. Use of diagonalization is optional — well, actually — redundant. At the expense of an extra step, a second line could derive $T a \iff T(\neg T a)$ from the identity in the first row as follows:

| $a = \langle \neg T a \rangle$ | Premise |
| $T a \iff T(\langle \neg T a \rangle)$ | Leibniz's law |
| $T(\langle \neg T a \rangle) \iff \neg T a$ | T-schema |
| $T a \iff \neg T a$ | 2, 3 Sentential Logic |

And then this line combined with the instance of the T-schema could be used to derive the contradictory biconditionals in the last row.\footnote{15} This deflates the view of diagonalization as providing a uniform approach to the paradoxes. The simple fact is that the Liar does not necessarily involve diagonalization!

5.3.6 Diagonalization's attempted Revenge

I might like to conclude from the above that diagonalization is not necessary for the Liar, and that Leibniz's law (or something like it) is. While the first part of this follows from the above, the second does not. We have in fact already seen in the proof of the Liar from the Diagonal theorem, that if the diagonal theorem were taken as primitive, then the Liar immediately follows without use of Leibniz's law.

So, we have a diagonal schema:

$$C \equiv \phi(\langle C \rangle)$$

and given such a schema as a primitive, the Liar may be derived:

\footnote{15} The indexical version of the Liar may well use self-intra-substitution, but the example in the left-hand side of the second table does not.
Chapter 5  The Liar and Grelling’s Paradoxes

1. L iff \( \sim T(L) \)  Instance of the diagonal schema
2. \( T(L) \) iff L  Instance of the T-schema
3. \( T(L) \) iff \( \sim T(L) \)  1, 2 SL

Diagonalization (or at least the diagonal theorem) can be taken as a primitive from which a Liar can be proven; so, it appears that Leibniz’s law is not strictly necessary for proving the Liar. Moreover, Leibniz’s law is not in this way necessary for proving the inconsistency of naive truth either. Strictly speaking, if we assume the T-schema and treat the Diagonal Theorem as a given primitive schema, we can derive a reductio of an instance of the T-schema. While technically correct, such a proof would be unconvincing, because the plausibility of the T-schema would, other things being equal, far outweigh the plausibility of the Diagonal Theorem. The proof of the inconsistency of naive satisfaction using the Diagonal Theorem is taken as sound because the Diagonal Theorem can be proven rather than merely taken as a primitive.

Likewise, I suggest, the proof of the Liar which takes the Diagonal Theorem as a primitive would be viewed as a reductio of that diagonalization were it not the case that the Diagonal Theorem can be proven. Again, the diagonalization, taken as primitive, is far less plausible than the T-schema. It is because the Diagonal Theorem can be proven that we treat this derivation as paradoxical rather than a reductio of the diagonal premise (or primitive presumption). In this sense, I suggest Leibniz’s law (or something like it) remains a candidate necessary condition for the Liar paradox, despite diagonalization’s attempted revenge.

### 5.4 Grelling’s Paradox and the Inconsistency of Naive Satisfaction

Grelling’s paradox is often presented using an adjective ‘heterological’. An expression is heterological if, and only if, it does not describe itself. ‘Long’ is heterological, as ‘long’ is not long. ‘Onomatopoeic’ is not onomatopoeic, and so is heterological too. But ‘English’ is English and is not heterological. ‘Blue’ (taken as naming a type not a token) is not blue, and is heterological; whereas ‘meaningful’ is meaningful and not heterological.
Chapter 5  The Liar and Grelling's Paradoxes 207

We could shorten the definition to the following schema, abbreviating "heterological" as "het" and defining it as a predicate for self-non-application: 16

het("\( \phi \)") iff \( \sim \phi ("\( \phi \)\)).

The contradiction follows immediately by taking "het" as the replacement for the schematic \( \phi \).

[Sainsbury 1995, p. 147]

Sainsbury's presentation of the paradox is interesting because it does not use the 'true of' relation. If Sainsbury's presentation is formally acceptable, then Grelling's paradox is \textit{prima facie} not a paradox of truth. While I personally think this presentation is formally acceptable, I suspect it may attract criticism as relying on substitutional quantification. However, the same can be said for the \( T^2 \)-schema. Consider the two schemata:

\[
\text{het}(\langle P \rangle) \text{ iff } \sim P(\langle P \rangle) \\
T(\langle P_1 \rangle, t) \text{ iff } P[t/x]
\]

In these, 'het' and 'T' are respectively predicates and relations of the object-language that are included in the meta-language, 'P' is used as a schematic letter in the meta-language for an arbitrary object-language predicate, and 't' is used in the latter as a schematic letter in the meta-language for an arbitrary object-language term. In the latter, the use of 't' could be replaced by a bound variable, 'y'; but 'P' cannot be replaced by a bound variable without recourse to second-order logic.

I personally do not see a problem with using substitutional quantification here, because one is intending to quantify over expressions rather than all (and possibly nameless) objects in the domain. 17

\[16\] One might think of self-non-application as the predicate obtained by reflexivizing the 'x does not apply to x'. My use of 'applies' here is different to use of 'applicability' in 'Range of applicability', as something is in the latter if it is true of or not true of a predicate. One needs an expression like 'is not the case' but for relations.

\[17\] I would make similar comments about Prior's [1958] presentation of the Epimenides using substitutional quantification without the truth predicate. In that case, Prior makes explicit use of substitutional quantification in the object-language.
Chapter 5  The Liar and Grelling's Paradoxes

Peter Roeper has pointed out to me that a paradox can be obtained without using truth by introducing a grammatical device:

Suppose one uses a grammatical device represented by 'τ' as a placeholder for the sentence in which it occurs. Thus

A & τ

is equivalent to

A & (A & τ)

Then the Liar is formed:

¬τ

¬τ ≡ ¬(¬τ)

This does not use a truth locution. This paradox is avoided by simply not admitting such a grammatical device into the language.

Truth enters the Liar and Grelling's paradoxes in a number of ways, as I list in Chapter 1. The cases above appear to be exceptions to the use of truth in the relevant schema, or any associated schema for that matter. I suggest they are associated with locutions like 'This sentence is the case', and an analogous sense of 'applies'. If they are acceptable (and I think they are) and if they do not use some sort of converse device to truth, then there existence undermines approaches to the Liar and Grelling's by restricting the range of applicability of the truth predicate and relations and pushes the case for failure of Leibniz's law and reflexivizations among candidate solutions by eliminating, undermining or limiting the scope of the mainstream approach.

It is historically common to define 'heterological' as applying truly to just those expressions that are not true of themselves. Take for example,

For any predicate x, x is heterological if and only if x applies falsely to itself.

[Martin 1968, p. 321]

and later, as a "stronger" version:

For any predicate x, x is heterological if and only if x does not apply truly to itself.

[Martin 1968, p. 330]
However, the definition of ‘heterological’ is redundant. The language already has the wherewithal to talk about whether an expression is true of itself or not; and the paradox arises over whether ‘is not true of itself’ is true of itself.

Assume for reductio that every instance of the ‘true of’ schema, a schema characteristic of the truth relation, i.e. the $T^2$-schema, is a theorem:

1. $T(\langle Px \rangle, t)$ iff $P[t/x]$ $\quad$ $T^2$-schema
2. $T(\langle \neg T(x, x) \rangle, t)$ iff $\neg T(t, t)$ $\quad$ 1, $\neg T(x, x)$ for $Px$
3. $T(\langle \neg T(x, x) \rangle, \langle \neg T(x, x) \rangle)$ iff $\neg T(\langle \neg T(x, x) \rangle, \langle \neg T(x, x) \rangle)$ $\quad$ 2, $\langle \neg T(x, x) \rangle$ for $t$

So, we have a contradiction as a consequence of our assumption that every instance of the $T^2$-schema is a theorem. This proof relied on the $T^2$-schema (and having canonical names for expressions of the language); it did not rely on assuming the substitution function or Leibniz' law, let alone the Diagonal Theorem. This derivation, which is given in the meta-language as it includes schematic letters in the body of the proof, does use an instance of a schema to derive lines 2 or 3, wherein ‘$\neg T(x, x)$’ was substituted for ‘$Px$’ and ‘$\langle \neg T(x, x) \rangle$’ instantiated $t$. Nevertheless, this is not using the substitution function or something like Leibniz’s law. This “proof” merely demonstrates how the third line is an instance of the $T^2$-schema. Assuming the theory contains all instances of the $T^2$-schema, a contradiction (or at least contradictory biconditional) can be obtained formally by a one step proof:

1. $T(\langle \neg T(x, x) \rangle, \langle \neg T(x, x) \rangle)$ iff $\neg T(\langle \neg T(x, x) \rangle, \langle \neg T(x, x) \rangle)$ $\quad$ $T^2$-schema

I want to highlight the austerity of proving the inconsistency of satisfaction (or the truth relation) relative to the truth predicate, consider the simple scenario where all instances of the $T$- and $T^2$-schemata are assumed as axioms. The Grelling-like proof above immediately provides a contradiction from an instance of the $T^2$-schema. Yet whether a contradiction follows from a Liar-like instance of the $T$-schema depends on the validity of the substitution function and something like Leibniz’s law. As a corollary, the proof of the Liar paradox has additional necessary logical principles that Grelling’s does not necessarily require.
5.4.1 Intuitions about the Relationship between the Liar and Grelling's

Yet Grelling's paradox is often considered not just another semantic paradox, but a
version of the Liar (Cf. Simmons [1993, p. 18]). When listing Liar versions, Visser
[2004] is a bit more hesitant, and introduces the paradox thus:

Of the following ‘versions’ it is disputable whether they are ‘really’
versions of the Liar rather than closely related paradoxes. I’m going to present
them anyway.

The next ‘version’ is known as Grelling’s Paradox. The German
mathematician Kurt Grelling found it in 1908. It dispenses with the use of a
self referential sentence altogether. Instead it employs self-application of words.
Let’s call a word “heterological” when it does not apply to itself, e.g.
“long” is heterological. Is “heterological” heterological? If it is, it
applies to itself, so it is not heterological. Contradiction. So
“heterological” is not heterological, hence it is heterological.
Contradiction on no assumptions.

[Visser 2004, p. 152]

The variations on Grelling’s paradox that I have been using strip away the unnecessary
definition and use just canonical naming of predicates (represented by open formulae):

‘Does not satisfy itself’ satisfies itself iff it does not satisfy itself.

Alternatively, using the converse relation, the truth relation, there is an instance of the T²-
schema:

‘Is not true of itself’ is true of itself iff ‘is not true of itself’ is not true of
itself.

Visser [2004, p. 152] anticipates the use of this paradox towards an inconsistency
result but does not comment on its austerity:

It is worthwhile to observe that Grelling’s Paradox can be converted in a
“meta-mathematical” version for languages containing their own
satisfaction predicate. First note that the relevant ‘words’ for Grelling’s
definition of ‘heterological’ are predicate words. These may be replaced or
Chapter 5  The Liar and Grelling’s Paradoxes

mimicked by formulae having just \(x_0\) free. Now define a formula standing for “heterological(\(x_0\))” by:

\[ \neg \text{Sat}(x_0, \langle x_0 \rangle). \]

Then Grelling’s sentence ““heterological” is heterological” becomes:

\[ \neg \text{Sat} \left( \langle \neg \text{Sat}(x_0, \langle x_0 \rangle) \rangle, \langle \langle \neg \text{Sat}(x_0, \langle x_0 \rangle) \rangle \rangle \right). \]

[Visser 2004, p. 152]

By the usual properties of Sat, this quickly leads to Paradox. Our rephrasing of Grelling’s sentence is closely related to Gödel’s construction of his celebrated sentence: what corresponds to ‘\(\neg \text{Sat}(x_0, \langle x_0 \rangle)\)’ there, is ‘\(\neg \text{Prov} (\text{sub}(x_0, x_0))\)’, where sub(\(a, b\)) is the function computing the Gödel number of the result of substituting the numeral of the number \(b\) for \(x_0\) in the formula with Gödel number \(a\).

I have shown the proof of the inconsistency of naive satisfaction within a language that has canonical names for its own expressions need not use the substitution function and Leibniz’s law or the like. These were not required for my proof about the truth relation. The same applies to the schema for the converse relation, the satisfaction relation.

I add this note in respect to Visser’s intuition: ‘\(\neg T(x, x)\)’ looks like ‘\(\neg T(s(x, x))\)’, which we used in the proof of the Liar in the previous section, but it is this latter expression that bears the closest analogy with ‘\(\neg \text{Prov} (\text{sub}(x_0, x_0))\)’. So, despite Visser’s [2004, p152] insight, it is the Liar that bears a more direct analogy to Gödel’s construction, rather than Grelling’s paradox. Visser’s insight re-emerges, when in a subsequent section we consider equivalences between the Truth relation on the one hand and expressions involving the truth predicate and substitution. It will appear at first that the two expressions above are equivalent; but there are constraints.

5.4.2 Grelling’s does use Diagonalization

Unlike the Liar, to the best of my knowledge, Grelling’s always uses diagonalization. In the proof above, diagonalization is involved in instantiating the \(T^2\)-schema such that a negated expression is substituted into the expression itself. This is clearly within the intent of diagonalization. Diagonalization was involved sometimes in the premise or corresponding theorem for the Liar, not the instantiation of the canonical T-schema. While this difference in the use of diagonalization is a difference between the Liar and the
Chapter 5 The Liar and Grelling’s Paradoxes

Grelling’s, as diagonalization is optional for the Liar and apparently necessary for Grelling’s, there is another more significant difference between the two paradoxes.

5.4.3 A ‘Fallacy’ in this Proof of Grelling’s Paradox

Remember that it is important for the $T^2$-schema that $P_{x}$ is (at most) monadic. So, the leftmost ‘$T$’ in the below biconditional represents a relation, whereas the ‘$T$’ in its argument represents a predicate. It is this latter $T$ predicate that appears outmost on the right-hand side of the biconditional. So, this is not a purely formal contradiction of the form, $\forall a \forall b$ $\neg \exists c$.

1. $T(\langle \neg T(x, x) \rangle, \langle \neg T(x, x) \rangle) \iff \neg T(\langle \neg T(x, x) \rangle, \langle \neg T(x, x) \rangle)$
   $T^2$-schema

Yet the above is a formal contradiction, as far as first-order logic is concerned. An ambiguity is not discerned in first-order logic: $T(a, a)$ may be an instance of $T(x, y)$ or $T(x, x)$. This ambiguity is immaterial if reflexivization is always semantically valid. Reflexivization assures us of the mutual entailment of a dyadic relation with identical arguments and a monadic predicate. First-order logic does not represent reflexivization as a formal rule, but indicates it by having identical arguments. (In which case, it seems that both the LHS and the RHS above are themselves monadic, and not representing the $T$-relation, but a $T$-predicate that formally represents the ambiguous natural language expression ‘true of itself’ as a predicate.)

It is impossible to adequately represent this formal issue in first-order logic. Let me then ill-use it just for this subsection for pedagogic purposes. We represent the distinction between a predicate and relation by a superscripted numeral. $F^1$ is distinct from $F^2$.

Although it is strictly speaking ill-formed to represent a dyadic relation as $F^2(x, x)$; let me use this non-standard representation for my stated purpose. I do promise a formally correct representation will be given in the next subsection. Let $F^1(x, x)$ represent ‘$x$ bears $F$ to itself’, and allow for the moment $F^2(x, x)$ to represent a restricted relation ‘$x$ bears $F$ to $x$’ so as to highlight the formal fallacy in the above derivation of Grelling’s paradox.

‘$F^1(x, x)$’ is formally distinct from ‘$F^2(x, x)$’, i.e. ‘$x$ bears $F$ to $x$’. The extension of $F^1(x, x)$ is a set of individuals that bear $F$ to themselves. The extension of $F^2(x, x)$ by contrast is a set of ordered pairs, a subset of the extension of $F^2(x, y)$. In particular, I use ‘$T^1(x, x)$’ to represent the monadic predicate ‘$x$ is true of itself’, which is distinct from ‘$T^2(x, x)$’ that represents a restriction of the relation ‘$x$ is true of $y$’, namely, ‘$x$ is true of $x$’. The
extension of $T^1(x, x)$ is a set of individuals. (Note that it is not the same set of individuals as the extension of $T^1(x)$. $T^1(x)$, which is the truth predicate used in the T-schema, is another distinct monadic predicate.) $T^1(x, x)$ is distinct from the extension of $T^2(x, x)$, which is a set of ordered pairs. Syntactically and semantically, ‘$T^1(x, x)$’ is a monadic predicate, just as ‘$T^2(x, x)$’ is a dyadic relation.

We cannot validly replace ‘$P x$’ in the $T^2$-schema with $T^2(x, x)$ as the latter is a two place relation. We can replace it with $T^1(x, x)$. This representation is sufficiently clear to enable us to make this distinction. Now, it seems to me that a contradiction is narrowly avoided in line 3 of the derivation below:

1. $T^2(P^1x), t$ iff $P^1[t/x]$  
   T-schema
2. $T^2(\langle \neg T^1(x, x) \rangle, t)$ iff $\neg T^1(t, t)$  
   1, $\neg T^1(x, x)$ for $P^1x$
3. $T^2(\langle \neg T^1(x, x) \rangle, \langle \neg T^1(x, x) \rangle)$ iff $\neg T^1(\langle \neg T^1(x, x) \rangle, \langle \neg T^1(x, x) \rangle)$  
   2, $\langle \neg T^1(x, x) \rangle$ for $t$

The outer-most ‘$T$’s on either side of the biconditional in line 3 are not the same: one is a relation, the other a predicate. So, line 3 is not a formal contradiction. 3 is read as:

‘is not true of itself’ is true of ‘is not true of itself’ iff ‘is not true of itself’ is true of itself.

The left-hand side uses a binary relation where the right-hand side uses the reflexivized predicate of that relation.

In line 2, the schema requires a monadic predicate and line 2 is justified only if one can assume that such a monadic predicate is derived by reflexivization from the converse of the satisfaction relation; but that inference has not been formally represented. Again in line 3, the same assumption must hold for the left-hand side of the biconditional to be the contradiction of the right-hand side.

### 5.4.4 Grelling’s Paradox quickly regained

Yet in first-order logic a contradiction does follow, but only because of its cursive representation. First-order logic does not represent reflexivization as a formal rule, but indicates it by having identical arguments. For our purposes, using first-order logic with superscripts, we might represent reflexivization by adding the schema:

$$(R^2F^1)(x) \iff F^2(x, x)$$

read ‘$x$ Fs itself iff $x$ Fs $x$’
Now paradox is quickly regained. Given reflexivization, a system is formally proved inconsistent if it contains all instances of the $T^2$-schema. For by reflexivization and substitution of equivalences, we have:

1. $T^2((P^1x), t) \text{ iff } P^1[t/x]$
2. $T^2((\neg R(T^2)^1(x)), t) \text{ iff } \neg R(T^2)^1(t)$
3. $T^2((\neg R(T^2)^1(x)), (\neg R(T^2)^1(x))) \text{ iff } \neg R(T^2)^1((\neg R(T^2)^1(x)))$
4. $R(T^2)^1((\neg R(T^2)^1(x))) \text{ iff } R(T^2)^1((\neg R(T^2)^1(x)))$

OR, substituting equivalents based on reflexivization into the right-hand side of line 3:

4a. $T^2((\neg R(T^2)^1(x)), (\neg R(T^2)^1(x))) \text{ iff } T^2((\neg R(T^2)^1(x)), (\neg R(T^2)^1(x)))$

I note that at line 2, reflexivization was used as an operator to take the 2-place truth relation to a one place predicate. (This is clearly not the one place predicate in the $T$-schema; it represents the reflexive predicate ‘is not true of itself’.) Line 2 and indeed line 3 are thus well-formed instances of the $T^2$-schema. Our derivation needs only three logical principles, the $T^2$-schema, Reflexivization and substitution of equivalents, as in the three-step proof below:

1. $T^2((\neg R(T^2)^1(x)), (\neg R(T^2)^1(x))) \text{ iff } \neg R(T^2)^1((\neg R(T^2)^1(x)))$

$T^2$-schema

2. $R(T^2)^1((\neg R(T^2)^1(x))) \text{ iff } T^2((\neg R(T^2)^1(x)), (\neg R(T^2)^1(x)))$

Reflexivization

3. $T^2((\neg R(T^2)^1(x)), (\neg R(T^2)^1(x))) \text{ iff } T^2((\neg R(T^2)^1(x)), (\neg R(T^2)^1(x)))$

1, 2 Substitution of equivalences

Reflexivization is formally necessary here, one might wonder if it could ever fail to be valid.
Chapter 5  The Liar and Grelling’s Paradoxes  215

5.4.5  Grelling’s Paradox resolved independently of the Liar

It is at least conceivable that reflexivization could fail. It is ad hoc to deny reflexivization of cases where there is no fact of the matter; but it is nevertheless possible to do so (perhaps based in some way on ungroundedness), and it is no more ad hoc than denying Leibniz’s law in the case of the Liar. Note that failure for reflexivization involving ungrounded expressions like Grelling’s predicate provides a solution to Grelling’s paradox, but not a solution to the Liar. Although this is a semantic proposal and the above proofs of paradox were purely formal, we should charitably allow that some formal rule might capture this semantic restriction. Conversely, failure of Leibniz’s law involving ungrounded sentences like the Liar sentence provides a solution to the Liar paradox, but not a solution to the Grelling’s paradox. Denying such instances of reflexivization is no more ad hoc than denying some instances of Leibniz’s law as a solution to the Liar. This provides some confirmation that there are two distinct – or at least distinguishable – paradoxes. Grelling’s paradox is not a version of the Liar. They may be related, but they are not the same paradox.

5.4.6  Two Paradoxes

Even if reflexivization never fails, there appear to be two formally distinct paradoxes in any case: The Liar for which something like Leibniz’s law seems necessary, and Grelling’s for which reflexivization seems necessary. We shall need to investigate these claims with respect to more radical variations on the Liar and Grelling’s in the next section.

5.5  Satisfied Liars, and Truth-predicated Grelling’s Paradoxes

There is a folk belief that the Liar and Grelling are distinguished by the T-schema and the T²-schema respectively, and a tacit belief that this is little distinction as these are just T-schemata applying to sentences, predicates, etc. If this were so, one would expect to be able to map Liar paradoxes into Grelling paradoxes using the relationship between the T-schema and the T²-schema, which is represented by the T2T-schema. However, when one uses the T2T-schema with the Liar sentence and the T²-schema, one does not get a Grelling-like paradox. Neither does one get a Liar-like paradox when one uses the
Chapter 5  The Liar and Grelling’s Paradoxes

Grelling predicate together with the T2T-schema and the T-schema to derive a contradiction. I demonstrate in this section that this is so; the Liar and Grelling’s are not related to each other by the relationship between the T-schema and the T²-schema.

Furthermore, I present a new paradox of satisfaction (the truth relation). It is characterised by a predicate expression; but is not a variation of Grelling’s paradox. It seems more Liar-like. So, there are two paradoxes of satisfaction, the Grelling’s and, of a distinct type from it, a Liar-like paradox.

The Liar is intuitively characterised by the sentences it uses, and Grelling’s is intuitively characterised by the predicates it uses. One might think that the Liar is to sentences as the Grelling’s to predicates. However, I have already shown that the Liar and Grelling’s have some distinct necessary logical principles. They have some in common, most notably the use of their respective T-schemata. As a corollary to the results of this section then, the Liar and Grelling’s cannot be respectively characterized and related as truth paradoxes of sentences and predicates.

5.5.1 The Liar using the T2T-schema

Because there is a relationship between the T-schemata (at least over sentences with suitable subject predicate form – and the Liar has such a grammatical form), it is reasonable to expect to be able to map the Liar to a proof involving the truth relation. The T2T-schema was introduced in the section on T-schemata as representing just the relationship between the T-schema and the T²-schema.

\[ T^{2}(\langle Px, a \rangle_{T}) \equiv T(Px[t/x]) \]

Consider then a proof of the Liar itself (or a reductio argument) using the T2T-schema and the T²-schema.

1. \( a = \langle \sim Ta \rangle \)  
   given premise (or theorem as above)

2. \( T^{2}(\langle \sim Tx, a \rangle_{T}) \equiv \sim Ta \)
   T²-schema (as a theorem, or assumed for reductio)

3. \( T^{2}(\langle \sim Tx, a \rangle_{T}) \equiv T(\sim Tx[a/x]) \)
   T2T-schema

4. \( T^{2}(\langle \sim Tx, a \rangle_{T}) \equiv T(\sim Ta) \)
   syntactic variant of the previous line

5. \( T^{2}(\langle \sim Tx, a \rangle_{T}) \equiv Ta \)
   1, 4 Substitution of Identicals

6. \( \sim Ta \equiv Ta \)
   2, 5 Substitution of Equivs
Chapter 5  The Liar and Grelling’s Paradoxes

Thus the Liar can be derived without the T-schema, using instead the $T^2$-schema, and furthermore the Liar itself can be used to prove the inconsistency of the truth-relation, given that the T2T-schema represents the relationship between the T- and $T^2$-schemata. For the proof taken as a *reductio* now derives a contradiction from the assumption of related instances of each of the T2T-schema and the $T^2$-schema. Suppose the instance of the T2T-schema is not true, then either the corresponding instance of the $T^2$-schema is not true or the T2T-schema fails to represent the relationship between the T-schema and the $T^2$-schema. And giving up the relationship is something more than giving up on some instances of the schemata. Better to give up some instances of the schemata than the relationship. Given the significance of the T2T-schema, which I will discuss next, neither of the T-schemata can fail independently.

5.5.2 *The Significance of the T2T-schema*

The T2T-schema cannot be restricted in isolation from the T-schemata it relates. If each of the biconditionals are or entail material equivalences, then this is conclusive. Even if they do not, the relationship between the T-schemata is more plausible than every instance of the T-schemata holding. Given that the T2T-schema represents the relationship between the first two of the T-schemata, The T-schema ought not to be modified without a corresponding modification of the $T^2$-schema, and *vice versa*. This means that even if the T-schema and $T^2$-shema are modified; provided they are modified consistently, the T2T-schema will hold (of all unrestricted cases).

5.5.3 *The New Liar-like Paradox of Satisfaction*

Here again is the new Liar-like paradox of satisfaction, which I introduced in Chapter 4:

My favourite predicate just is ‘does not satisfy my favourite predicate’.

Anything satisfies my favourite predicate iff it does not satisfy it.

Using the converse ‘true of’ relation:

My second favourite predicate just is ‘my second favourite predicate is not true of’. My second favourite predicate is true of anything iff it does not satisfy it.

Here is a formal representation of a particular case.
Chapter 5 The Liar and Grelling’s Paradoxes

\[ a = \langle \neg T(a, x) \rangle \]  
Premise

\[ T(\langle \neg T(a, x) \rangle, a) \iff \neg T(a, a) \]  
\( T^2 \) schema

\[ T(a, a) \iff \neg T(a, a) \]  
Leibniz’s law

And, this is just a special case of

\[ p = \langle \neg T(p, x) \rangle \]  
Premise

\[ T(\langle \neg T(p, x) \rangle, z) \iff \neg T(p, z) \]  
\( T^2 \) schema

\[ T(p, z) \iff \neg T(p, z) \]  
Leibniz’s law

The above argument uses a predicate, not a sentence; but it is distinct from Grelling’s paradox. My favourite predicate is paradoxical whatever it is said to be true of. It refers to itself and does not have to apply to itself to be paradoxical. Its self reference appears at first irresolvable in the sense that its non-canonical name cannot be eliminated by replacement by a canonical one. Nevertheless, like the Liar I believe the substitution function could be used to address this. The ubiquity of its paradoxicality distinguishes it from Grelling’s paradox. It is still a paradox of truth. Either it is a new variant of the Liar or some totally new paradox. In my opinion, it is a new variant of the Liar as it requires Leibniz’s law (or something like it).

In this way, we can prove the inconsistency of naive satisfaction in a Liar-like way using Leibniz’s law but not reflexivization.

Take any two-place relation, \( \phi(x, y) \).

In ‘\( \phi[s(\langle \phi[s(x, x), y] \rangle, \langle \phi[s(x, x), y] \rangle), y] \)’,

\( s(\langle \phi[s(x, x), y] \rangle, \langle \phi[s(x, x), y] \rangle) = \langle \phi[s(\langle \phi[s(x, x), y] \rangle, \langle \phi[s(x, x), y] \rangle), y] \rangle \)

So, by Leibniz’s law:

\( \phi[s(\langle \phi[s(x, x), y] \rangle, \langle \phi[s(x, x), y] \rangle), y] \iff \phi[s(\langle \phi[s(x, x), y] \rangle, \langle \phi[s(x, x), y] \rangle), y] \)

If we abbreviate ‘\( \phi[s(\langle \phi[s(x, x), y] \rangle, \langle \phi[s(x, x), y] \rangle), y] \)’ using ‘Cy’ in the meta-language, we have proved:

For any relation represented by ‘\( \phi(x, y) \)’, we can form a predicate abbreviated here by ‘Cy’, such that:

\[ \forall y (Cy \iff \phi(\langle Cy \rangle, y)) \]
In particular,
\[
\forall x \, (Ax \iff \neg T(\langle Ax \rangle, x)) \quad \text{Diagonal Theorem / Self-referential Lemma}
\]
\[
Az \iff \neg T(\langle Ax \rangle, z) \quad \forall E \text{ (Universal instantiation)}
\]
\[
T(\langle Ax \rangle, z) \iff Az \quad T^2\text{-schema}
\]
\[
T(\langle Ax \rangle, z) \iff \neg T(\langle Ax \rangle, z) \quad SL
\]

This can be used to prove the inconsistency of naive satisfaction (or its converse relation, the naive truth relation); but it is not based on Grelling’s paradox.

Once again, other variations need not use the theorem but just the identity and Leibniz’s law. For corresponding to:

\[
p = \langle \neg T(p,y) \rangle \quad \text{Premise}
\]
\[
T(\langle \neg T(p,y) \rangle, z) \iff \neg T(p,z) \quad T^2\text{-schema}
\]
\[
T(p,z) \iff \neg T(p,z) \quad 1,2 \text{ Leibniz’s law}
\]

there is:

1. \[
s(\langle \neg T(s(x,x),y) \rangle, \langle \neg T(s(x,x),y) \rangle) = \langle \neg T(s(\langle \neg T(s(x,x),y) \rangle, \langle \neg T(s(x,x),y) \rangle), y) \rangle
\] \quad \text{Theorem}

2. \[
T(\langle \neg T(\langle \neg T(s(x,x),y) \rangle, \langle \neg T(s(x,x),y) \rangle, y) \rangle, z) \iff \neg T(s(\langle \neg T(s(x,x),y) \rangle, z), \langle \neg T(s(x,x),y) \rangle, y) \rangle
\] \quad T^2\text{-schema}

3. \[
T(s(\langle \neg T(s(x,x),y) \rangle, \langle \neg T(s(x,x),y) \rangle), z) \iff \neg T(s(\langle \neg T(s(x,x),y) \rangle, \langle \neg T(s(x,x),y) \rangle), z)
\] \quad 1,2 \text{ Leibniz’s law}

This paradox can be extended to variation for any number of places. For example, my favourite relation paradox is a variant for three places:

For any relation represented by ‘ϕ(x, y, z)’, we can form a relation abbreviated here by ‘C(y, z)’, such that:
\[
\forall y \forall z \, [C(y, z) \iff ϕ(\langle C(y, z) \rangle, y, z)]
\]

In particular,
\[
\forall x \forall y \, (Axy \iff \neg T(\langle Axy \rangle, x, y)) \quad \text{Diagonal Lemma}
\]
\[
Aab \iff T(\langle Axy \rangle, a, b)) \quad \forall E \text{ (Universal instantiation)} \times 2
\]
\[
T(\langle Axy \rangle, a, b) \iff Aab \quad T^3\text{-schema}
\]
\[
T(\langle Axy \rangle, a, b) \iff \neg T(\langle Axy \rangle, a, b)) \quad SL
\]
Chapter 5  The Liar and Grelling’s Paradoxes  220

Here is the proof of the theorem (for this number of places):

In ‘\( \phi[s(\phi[s(x, x), y, z]), \langle \phi[s(x, x), y, z]\rangle, y, z] \)’,

\( s(\phi[s(x, x), y, z]), \langle \phi[s(x, x), y, z]\rangle = \phi[s(\phi[s(x, x), y, z]), \langle \phi[s(x, x), y, z]\rangle, y, z] \)

So, by Leibniz’s law:

\( \phi[s(\phi[s(x, x), y, z]), \langle \phi[s(x, x), y, z]\rangle, y, z] \text{ iff } \phi[s(\phi[s(x, x), y, z]), \langle \phi[s(x, x), y, z]\rangle, y, z] \)

If we abbreviate ‘\( \phi[s(\phi[s(x, x), y, z]), \langle \phi[s(x, x), y, z]\rangle, y, z] \)’ using ‘Cyz’ in the metalinguage, we have proved:

For any relation represented by ‘\( \phi(x, y, z) \)’, we can form a relation abbreviated here by ‘\( \text{C}(y, z) \)’, such that:

\( \forall y \forall z [\text{C}(y, z) \text{ iff } \phi((\text{C}(y, z)), y, z)] \)

5.5.4 Grelling’s using the T-Schema

There is a Grelling-like contradiction using the T-schema and the following equivalence, the T2T-Schema:

\( T^2(\langle Px, t \rangle) \equiv T(\langle Px[t/x] \rangle) \)

The intuitive appeal of this schema might be justified this way: the truth predicate applies to sentences of subject-predicate logical form just when the predicate is true of the object referred to by the grammatical subject term.

Using the T2T Schema, it is possible to construct examples that look initially like they do not use reflexivization. Consider:

1. \( T(\langle \sim T(x, x) \rangle, \langle \sim T(x, x) \rangle) \equiv T(\langle \sim T(x, x) \rangle [\langle \sim T(x, x) \rangle / x]) \quad \text{T2T-schema} \)
2. \( T(\langle \sim T(x, x) \rangle, \langle \sim T(x, x) \rangle) \equiv T(\langle \sim T(\langle \sim T(x, x) \rangle, \langle \sim T(x, x) \rangle) \quad \text{syntactic variant of above line} \)
3. \( T(\langle \sim T(x, x) \rangle, \langle \sim T(x, x) \rangle) \equiv T(\langle \sim T(x, x) \rangle, \langle \sim T(x, x) \rangle) \quad \text{T-schema} \)

Once again, however, there is an ambiguity that needs to be addressed with reflexivization. (Indeed, if the ambiguity is not valid, i.e. does not preserve truth, the argument is fallacious.) For line one to be a valid instance of the T2T-schema, the first
argument in the left hand-side must be a monadic predicate. We can make this clearer, and correct it, by adding addicity superscripts and using Reflexivization:

1. $T^2(\langle \neg \exists \neg (T^2)^1(x) \rangle, \langle \neg \exists \neg (T^2)^1(x) \rangle) \text{ iff } T^2(\neg \exists \neg (T^2)^1(x), \langle \neg \exists \neg (T^2)^1(x) \rangle / x)\]
   
   $T2T$-schema

2. $T^2(\langle \neg \exists \neg (T^2)^1(x) \rangle, \langle \neg \exists \neg (T^2)^1(x) \rangle) \text{ iff } T(\neg \exists \neg (T^2)^1(\langle \neg \exists \neg (T^2)^1(x) \rangle))$
   
   syntactic variant of above line

3. $T^2(\langle \neg \exists \neg (T^2)^1(x) \rangle, \langle \neg \exists \neg (T^2)^1(x) \rangle) \text{ iff } \neg \exists \neg (T^2)^1(\langle \neg \exists \neg (T^2)^1(x) \rangle)$
   
   T-schema, substitution of Equiv

4. $T^2(\langle \neg \exists \neg (T^2)^1(x) \rangle, \langle \neg \exists \neg (T^2)^1(x) \rangle) \text{ iff } \neg T^2(\langle \neg \exists \neg (T^2)^1(x) \rangle, \langle \neg \exists \neg (T^2)^1(x) \rangle)$
   
   Reflexivization, substitution of equivalents

In my opinion, this is a variant of Grelling’s paradox as reflexivization is needed. It relies upon it in two places. Firstly, in the instance of the $T2T$-schema. Without there being a one place predicate $\langle \neg \exists \neg (T^2)^1(x) \rangle$, this would not be a valid instance of the $T^2$-schema. (It is hard to doubt the existence of the predicate, but its extension or the existence of a corresponding property may be called into doubt). Secondly, reflexivization is essential in inferring line 4 from line 3. In the above derivation, the right-hand side has been replaced using an equivalence based on reflexivization. (Reflexivization could instead have been used to replace the left-hand side.)

### 5.5.5 Grelling’s using Leibniz’s Law

Use of Leibniz’s law does not preclude use of reflexivization. I am merely arguing that use of Leibniz’s law (or something like it) is necessary to derive a Liar paradox; and reflexivization is necessary to derive a Grelling’s paradox. Conceivably, there are hybrids of the Liar and Grelling’s which use perhaps some complex expression and both Leibniz’s law and reflexivization. Here is a trivial case where use of identity is clearly redundant:

\[
\begin{align*}
  a &= \langle \neg \exists \neg (T^2)^1(x) \rangle & \text{Premise} \\
  T^2(\langle \neg \exists \neg (T^2)^1(x) \rangle, a) & \text{iff } \neg \exists \neg (T^2)^1(a) & T^2\text{-schema} \\
  T^2(\langle \neg \exists \neg (T^2)^1(x) \rangle, \langle \neg \exists \neg (T^2)^1(x) \rangle) & \text{iff } \neg \exists \neg (T^2)^1(\langle \neg \exists \neg (T^2)^1(x) \rangle)) & \text{Leibniz’s law} \\
  \neg T^2(\langle \neg \exists \neg (T^2)^1(x) \rangle, \langle \neg \exists \neg (T^2)^1(x) \rangle) & \text{iff } \neg T^2(\langle \neg \exists \neg (T^2)^1(x) \rangle, \langle \neg \exists \neg (T^2)^1(x) \rangle) & \text{Reflexivization, substitution of equivalents}
\end{align*}
\]
Chapter 5  The Liar and Grelling’s Paradoxes

Now this is basically Grelling’s, but redundantly using an identity; it is not the Liar. We can see that the use of Grelling’s predicate and reflexivization are essential, whereas the use of Leibniz’s law could be eliminated using a canonical name for ‘~\(\beta(T^2)^1(x)\)’.

5.6 Quine’s View

Quine’s view is that self-predication may be employed to derive the Liar or Grelling’s paradox. Thus Quine [1995, pp. 238-241] gives a proof of the inconsistency of naive truth using the Liar without appealing to Gödel numbering. Initially, he uses a self-predication function, SP,

> Given any open sentence in one free variable ‘\(x\)’, let us substitute for ‘\(x\)’, in it, a name of that sentence itself. The result is the self-predication of the original sentence.  

[Quine 1995, p. 239]

Thus,

\[
SP(\ ‘x\ \text{is long}\ ’ ) = ‘ ‘x is long’ is long’
\]

and towards a Liar

\[
SP( ‘SP(\ x\ )\ is\ false’ ) = ‘SP( ‘SP(x) is false’ ) is false’
\]

or, slightly more formally:

1. \[
SP( ‘\sim T( SP(x) )’ ) = ‘\sim T( SP( ‘\sim T( SP(x) )’ )’
\]

(‘SP’ in the latter might even be read as self-prefixing its own quotation, in which case an alternate formal similitude of self-intra-substitution is achieved by exploiting grammatical machinery or formation rules. This type of alternative to self-intra-substitution using self-suffixing to its own quotation was used by Quine in [1962 / 1976] towards his original version of his Liar using ‘yields a falsehood when appended to its own quotation’ to this end. Smullyan [1957] exploited self-prefixing to its own quotation.)

So, from 1 by Leibniz’s law:

2. \[
T(‘\sim T( SP( ‘\sim T( SP(x) )’ )’ )’ ) \iff T( SP( ‘\sim T( SP(x) )’ )’ )
\]

By the T-schema (even if for sake of reductio)
3. $T(\sim T(SP(\sim T(SP(x)''))) \iff \sim T(SP(\sim T(SP(x)'))))$

Then from 2 and 3 by substitution of equivalents:

4. $T(SP(\sim T(SP(x)''))) \iff \sim T(SP(\sim T(SP(x)'))))$

Alternatively, substituting into the RHS of 3 yields:

4a. $T(\sim T(SP(\sim T(SP(x)'')))') \iff T(\sim T(SP(\sim T(SP(x)')')))''$

Quine goes on to reduce self-predication to substitution and naming. So, it is probably fair to represent him as relating the Liar and Grelling’s in terms of diagonalization, taken as self-intra-substitution or a formal alternative, as this accords with how we get line 1. However, as I have already noted, not all versions of the Liar use self-intra-substitution to obtain their premise. Diagonalization involves the replacement of a canonical or non-canonical name with a canonical name; but there are examples of the Liar, as I have presented, where a canonical name is replaced by a non-canonical name. This cannot be self-intra-substitution or its grammatical similitude.

Quine’s derivation of the Liar uses Leibniz’s law to obtain line 4 above. It is tempting to wonder what Quine’s derivation of Grelling’s might be:

1. $SP('x$ is not true of itself') = ‘‘x is not true of itself’ is not true of itself’

does not have an obvious relation to an instance of the $T^2$-schema:

2. ‘‘x is not true of itself’ is true of itself iff ‘x is not true of itself’ is not true of itself

And in any case, the single instance of the $T^2$-schema is sufficient for paradox on its own. This seems off the mark. Quine may be trying to use (self-)application and the Grelling expression to form a Liar-like version of Grelling’s (and so relate them) using, perhaps the following in some way:

\[ \sim T(x, x) \iff \text{het}(x) \iff \sim T(SP(x)) \]
5.7 Two Paradoxes and two Proofs of Inconsistency

There are then two paradoxes and thus two derivations of inconsistency. As the Liar and Grelling’s as paradoxes may each use either the T-schema or the T^2-schema, either paradoxical proof may be used as the basis of an inconsistency proof for either the naive truth predicate or the naive truth relation.

5.7.1 Two Paradoxes as two Paradoxes of Truth

Previously, the paradoxes were distinguished by the logical principles necessary for their proof; now, the relationship between the T-schemata can also be used to distinguish them. The Liar can be derived using the T-schema, or using the T2T-schema and the T^2-schema. Likewise, Grelling’s can be derived using either the T^2-schema, or the T-schema and the T2T-schema. The T2T-schema represents the relationship between the T-schema and the T^2-schema; but when it is used to map the proof of the Liar into a proof using the truth relation, the result is not Grelling’s paradox; nor is Grelling’s able to be mapped into a proof of a Liar paradox by using the T2T-schema. Thus the relationship between the T-schemata can be used to distinguish the paradoxes as two different types of paradoxes of truth.

If the T-schemata are restricted to avoid expressions that refer to expressions in which they occur, like a or an expression involving the substitution function and canonical naming as in our earlier proof of the Liar, this will avoid the self-referential Liar, the self-referential version of my favourite predicate paradox and some derivations of Grelling’s, but not all derivations of Grelling’s. (Of course, such a restriction would have to be extended to circular and infinite variations of the Liar and my favourite predicate paradox even to cover those; but there are some variations of Grelling’s that such a principle misses entirely.) Here is an example of Grelling’s that would be avoided by such a restriction.

1. \( b = \langle \neg \text{T}(x, x) \rangle \)  
   Premise (presumably by stipulation)

2. \( \text{T}(\langle \langle \neg \text{T}(x, x), b \rangle \rangle) \iff \neg \text{T}(b, b) \)  
   \( \text{T}^2 \)-schema

3. \( \text{T}(\langle \langle \neg \text{T}(x, x), (\neg \text{T}(x, x)) \rangle \rangle) \iff \neg \text{T}(\langle \neg \text{T}(x, x), (\neg \text{T}(x, x)) \rangle) \)  
   1, 2 Leibniz’s law
However, the use of the identity is redundant, and the version that does without it will not be avoided by such a restriction.

If we restrict the T-schemata to avoid self-intra-substitution (diagonalization), then this will avoid Grelling's and some derivations of the Liar, but not all derivations of the Liar (or even my favourite predicate paradox).

A restriction on both sources of pathology may be too broad, or might simply be disjunctive, which makes it look like two restrictions really. While not conclusive, these are reasons to expect that to restrict logic to avoid all derivations of the Liar and Grelling's would require two restrictions.

In other words to avoid just the Liar and Grelling paradoxes by restricting T-schemata, it is necessary to avoid the T-biconditionals for expressions potentially involved in Liar-like identity statements and those that diagonalize some predicates obtained by reflexivization, in particular the predicate obtained by reflexivization from the truth relation. The restrictions are *prima facie* heterogeneous. The burden of proof is on claims to the contrary. Yet apparently some theories do avoid just them, in particular, because both are *ungrounded*. A theory of *ungroundedness*, can and should appeal to contingencies. It also uses a model of some sort to provide the semantic value for non-canonical names and, moreover, the extensions of predicates that may contingently be involved in quantificational variants of the Liar; it can effectively restrict the valuation of the application of T-predicates and T-relations from both canonical and non-canonical names of the expressions in question for both paradoxes.

**5.7.2 Two Proofs of each Inconsistency Result**

Since the Liar and Grelling's have been shown to have different necessary conditions, we know we have two proofs for inconsistency, one is Tarski's proof of the inconsistency of truth, paralleling the Liar, and another proof of the inconsistency of satisfaction, paralleling a derivation of Grelling's paradox. Moreover, since we can use the T2T-schema and the T2-schema to derive a version of the Liar paradox, we can use this to prove the inconsistency of naive truth, i.e. given the T2T-schema, substitution function and Leibniz's law we can derive a *reductio* of the T-schema. My favourite predicate paradox can also be adapted to prove the inconsistency of naive truth. (Given the common presumption that the Liar and Grelling's are closely related, all previous arguments to
these results may have been presumed to be Liar-like. Yet the proof of the inconsistency of naive satisfaction using Grelling’s paradox is more economical and distinct from that adapted from my favourite predicate paradox.) Finally, since we can use the T2T-schema and the T-schema to derive a version of Grelling’s paradox, we can use this to prove the inconsistency of naive truth, i.e. given the T2T-schema and reflexivation we can derive a \textit{reductio} of the T-schema. In some respects, this proof is more economical that the proof that uses the Liar, in any case, it is a distinct proof.

\\[5.8 \ \textbf{The Truth Relation and the Substitution Function}\\]

Interest in the inter-definability of the concepts we have been dealing with may lead us to wonder whether the Liar and the Grelling may yet have a common form or pathology when related in this way. In particular, what is the relationship between the substitution function and the $T^2$-schema, if we took one as primitive and redefined the other in its terms would this show a deeper relationship between the paradoxes? The substitution function cannot however be used to disquote, in a general sense, like the $T^2$-schema; so more than the substitution function is needed to define the T-schemata. Nevertheless, the $T^2$-schema could perhaps be defined in terms of the T-schema and the substitution function – this is reminiscent of Quine’s construction using self-predication with the truth predicate. For now, with the resources we have, we can prove an equivalence relating the $T^2$-schema to the T-predicate and the substitution function.

The TST-schema: $T^2(\langle Px \rangle, \langle t \rangle) \equiv T(s(\langle Px \rangle, \langle t \rangle))$, for monadic or sentential $P$

1. $s(\langle Px \rangle, \langle t \rangle) = \langle P(\langle t \rangle/x) \rangle$ \hspace{1cm} Definition of $s^2(\ )$
2. $T(\langle A \rangle$ iff $T(\langle A \rangle$ \hspace{1cm} Tautology
3. $T(s(\langle Px \rangle, \langle t \rangle)) = T(\langle P(\langle t \rangle/x) \rangle)$ \hspace{1cm} 1, 2 Leibniz’s law, for monadic or sentential $P$
4. $T(\langle P(\langle t \rangle/x) \rangle \equiv P(\langle t \rangle/x)$ \hspace{1cm} T-schema, for monadic or sentential $P$
5. $T^2(\langle Px \rangle, \langle t \rangle) \equiv P(\langle t \rangle/x)$ \hspace{1cm} $T^2$-schema, for monadic or sentential $P$
6. $T^2(\langle Px \rangle, \langle t \rangle) \equiv T(s(\langle Px \rangle, \langle t \rangle))$ \hspace{1cm} 2, 3, 4 SL, for monadic or sentential $P$
Chapter 5

The Liar and Grelling’s Paradoxes

Note, however, the equivalence at line 6 can only be used to eliminate $T^2$ under two conditions. Firstly, it only applies to terms, rather than things generally. The following is an instance of the $T^2$-schema, but not of (5) or (6):

$$T^2(\text{‘x barks’}, \text{Fido}) \iff \text{Fido barks.}$$

Secondly, $P_x$ must be a monadic predicate (or a sentence), i.e., not a dyadic relation or any $n$-adic relation with adicity greater than 1. The following is an instance of the substitution schema, but it cannot be substituted into (6):

$$s(\langle x \text{ has more letters than y} \rangle, \text{‘Cerberus’}) = \text{‘Cerberus’ has more letters than y’}$$

The TST-schema, like the T2T-schema considered earlier, is reasonably intuitive, and could be taken as primitive. In this section, my object in doing so is to show that Leibniz’s law is still necessary for deriving the Liar, as is reflexivization for deriving Grelling’s.

Firstly, let us consider Grelling’s again. The following is an instance of the TST-schema:

$$T^2(\langle \neg T(x,x) \rangle, \langle \neg T(x,x) \rangle) \equiv T(s(\langle \neg T(x,x) \rangle, \langle \neg T(x,x) \rangle))$$

It is ambiguous. Either the ‘$T$’ in the arguments is intended to be dyadic, in which case it does not satisfy the restriction on the TST-schema, or it is intended to be monadic and reflexivized. To have the reflexive predicate represented we shall use the reflexivity-operator, (and a slightly non-standard representation, over-taxing the superscripts:)

1. $T^2(\langle \neg R(T^2)^1(x) \rangle, \langle \neg T^2(x,x) \rangle) \equiv T(s(\langle \neg R(T^2)^1(x) \rangle, \langle \neg T^2(x,x) \rangle))$

   TST-schema

2. $T^2(\langle \neg R(T^2)^1(x) \rangle, \langle \neg T^2(x,x) \rangle) \equiv T(\neg R(T^2)^1(\langle \neg T^2(x,x) \rangle))$

   s-function, Leibniz’s law

3. $R(T^2)^1(x) \iff \neg T^2(x,x)$

   Reflexivization

4. $T^2(\langle \neg T^2(x,x) \rangle, \langle \neg T^2(x,x) \rangle) \equiv T(\neg T^2(\langle \neg T^2(x,x) \rangle, \langle \neg T^2(x,x) \rangle))$

   2, 3 2, 3, substitution of equivalents

5. $T^2(\langle \neg T^2(x,x) \rangle, \langle \neg T^2(x,x) \rangle) \equiv -T^2(\langle \neg T^2(x,x) \rangle, \langle \neg T^2(x,x) \rangle)$

   T-schema, substitution of Equiv
This proof used Reflexivization, and Leibniz’s law and the T-schema. Nevertheless, reflexivization is still necessary to derive a contradiction using Grelling’s predicate. Even when the T²-schema was avoided by reducing the truth relation to the T-predicate and the substitution function, the same sort of restrictions applied and compelled the need for Reflexivization.

Secondly, let us take the canonical Liar and translate it using the definition of T²:

1. \( s(\neg T(s(x, x)), \neg T(s(x, x))) = \neg T(s(\neg T(s(x, x)), \neg T(s(x, x)))) \)
   Definition of the s-function

2. \( s(\neg T^2(x, x)), \neg T^2(x, x)) = \neg T(s(\neg T^2(x, x)), \neg T^2(x, x)) \)
   1, TST-schema, substitution of equivalents

3. \( T(s(\neg T^2(x, x)), \neg T^2(x, x)) \) iff \( T(s(\neg T^2(x, x)), \neg T^2(x, x)) \)
   2, Leibniz’s law

4. \( T(s(\neg T^2(x, x)), \neg T^2(x, x)) \) iff \( T(s(\neg T^2(x, x)), \neg T^2(x, x)) \)
   T-schema, 5 substitution of equivalents

This is a Liar because it is still using an identity of the form \( a = \neg T a \), it is just a bit more complex, that’s all. I hesitated to replace any more instances of ‘T(s( )’ in line 2, although it would have transformed this identity, because of the restriction on the TST-schema; however, the definition of the substitution function itself will give:

1. \( s(\neg T^2(x, x)), \neg T^2(x, x)) = \neg T^2(s(\neg T^2(x, x)), \neg T^2(x, x)) \)
   Definition of the s-function

Now we have a premise of a quite different form, the form \( s(b, b) = \neg T^2(b, b) \), and then, we have:

2. \( T(s(\neg T^2(x, x)), \neg T^2(x, x)) \) iff \( T(s(\neg T^2(x, x)), \neg T^2(x, x)) \)
   1, Leibniz’s law

but, we cannot now proceed, as we might wish, to:

3. \( T^2(s(\neg T^2(x, x)), \neg T^2(x, x)) \) iff \( T^2(s(\neg T^2(x, x)), \neg T^2(x, x)) \)
   because of the restriction on the TST-schema. The valid way to work within this restriction is to use reflexivization.

3b. \( T(s(\neg R(T^2) x)), \neg T^2(x, x)) \) iff \( T(s(\neg R(T^2) x)), \neg T^2(x, x)) \)
Chapter 5  The Liar and Grelling’s Paradoxes

Reflexivization, 2 substitution of equivalents

4. $T^2(\langle \neg\beta(T^2)^1(x) \rangle, \langle \neg T^2(x, x) \rangle)$ iff $T(\neg T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle)$

3b. TST-schema, substitution of equivalents

5. $T^2(\langle \neg\beta(T^2)^1(x) \rangle, \langle \neg T^2(x, x) \rangle)$ iff $\neg T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle)$

T-schema, 4, substitution of equivalents

6. $T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle)$ iff $\neg T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle)$

Reflexivization, 5, substitution of equivalents

Perhaps this is a hybrid paradox. It is at least a version of Grelling’s as it uses a reflexive predicate and reflexivization.

Suppose we had used instead a combination of the T-predicate and the substitution function, unrestricted.

1. $s(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle) = \langle \neg T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle)$

Definition of the s-function

2. $T(s(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle))$ iff $T(\neg T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle))$

1, Leibniz’s law

3. $T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle)$ iff $T(\neg T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle))$

2 unrestricted TST-schema, substitution of equivalents

4. $T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle)$ iff $\neg T^2(\langle \neg T^2(x, x) \rangle, \langle \neg T^2(x, x) \rangle)$

T-schema, 3 substitution of equivalents

This is invalid; because it violates the restriction on the $T^2$-schema. This was not just a restriction where the value of the $T^2$ relation was undefined, as though it were only partially defined and could be completed by fiat. This is a restriction on the range of applicability of the $T^2$ relation. It is invalid to apply it to dyadic relations. Given the validity of reflexivization, the relation can apply to monadic reflexive predicates. While these predicates just are the reflexivized relation, the relation cannot apply to the relation, but does apply to the monadic reflexive predicate.

5.9 Quantified Liars

We might consider the Epimenides; but let us consider a related paradox, one whose history is distinguished by its being undistinguished. As related in Chapter 1, Prior [1961,
p. 18] attributes it to Peter Geach. A Cretan says that something a Cretan says is not true. It follows that this statement must be true, and therefore there must be some other Cretan statement which is false. Surprise would turn to paradox were this the only Cretan utterance, or if all others were true. The proof of the initially surprising conclusion, Geach’s conundrum, uses identity only late in the proof – to show it is not this statement which is false. (If we followed Prior in using substitutional quantification, it would not involve the truth predicate either; but we will stick with objective quantification.) Nevertheless, given the requisite additional premises, proof of the contradictory conclusion would, at least usually, use identity or indexicality.

Take ‘D’ to represent some predicate like ‘A Cretan says ...’ or ‘is said by a Cretan’

\[
\begin{align*}
1 & : D(\exists x(D(x) \land \neg T(x))) & \text{Premise} \\
2 & : \neg \exists x(D(x) \land T(x)) & \text{Assumption} \\
3 & : \forall x (Dx \supset Tx) & 3 \text{ QN, defn} \\
4 & : D(\exists x(D(x) \land \neg T(x))) \supset T(\exists x(D(x) \land \neg T(x))) & 4 \text{ \forall E} \\
5 & : T(\exists x(D(x) \land \neg T(x))) & 5, 1 \text{ MP} \\
6 & : \exists x(D(x) \land \neg T(x)) & 5 \text{ TE} \\
7 & : \exists x(D(x) \land \neg T(x)) & 2-6 \text{ \neg E} \\
8 & : D(\exists x(D(x) \land \neg T(x))) \land T(\exists x(D(x) \land \neg T(x))) & 7 \text{ TI, 1 \&I} \\
9 & : \exists x(D(x) \land T(x)) & 8 \exists I \\
\end{align*}
\]

Towards a contradiction, we might add either (10) or (11) as an additional premise.

\[
\begin{align*}
10 & : \forall x (Dx \supset x = (\exists x(D(x) \land \neg T(x)))) \\
11 & : \forall x ((Dx \& x \neq (\exists x(D(x) \land \neg T(x)))) \supset Tx) \\
\end{align*}
\]

Although identity becomes involved in proving a contradiction, the pathology seemed to come into the argument beforehand, specifically in the use of quantifier elimination. This bears some semblance with the proof of Grelling’s paradox. Even though the proof of Grelling’s paradox did not use quantifier elimination – rather it used a single instance of the T²-schema – it bears at least an intuitive relationship to quantifier elimination. Nevertheless, the proof of Grelling’s paradox has turned out to require reflexivization; something not required for the proof of the Liar, the Epimenides proper, or the above paradox.
As discussed in Chapter 3, whereas the Identity premise in the Liar establishes coreference of terms, the premise in this paradox uses a more general relation. It says of an expression that it is among those that satisfy a particular predicate. It says of at least one of those sentences that satisfy that predicate that that sentence is not true. So, the sentence embedded in the premise is merely potentially self-referential, whether it is will be a contingent matter. Now, rather than using Leibniz’s law to effect a substitution, this proof uses quantifier instantiation (elimination) to substitute the expression into a related expression at line 4, which is at least the similitude of diagonalization if not diagonalization. It also, conversely, uses generalization (quantifier introduction) to replace occurrences of the expression to form a related expression, which is at best the simulacrum of diagonalization. These are separate inferential moves to the use of T-Introduction (TI) and T-Elimination (TE), which are inferences related by modus ponens to the two conditionals in the canonical T-schema. This parallels the proof of the Liar, not Grelling’s. The parallel with the proof of the Liar is given in the table below.

<table>
<thead>
<tr>
<th>Towards Geach’s conundrum</th>
<th>The Liar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 D(∃x(D(x) ∧ ¬T(x)))</td>
<td>premise</td>
</tr>
<tr>
<td>2 ¬∃x(D(x) ∧ T(x))</td>
<td>assumption</td>
</tr>
<tr>
<td>3 ∀x (Dx ⊃ Tx)</td>
<td>2, QN</td>
</tr>
<tr>
<td>4 D(∃x(D(x) ∧ ¬T(x))) ⊃</td>
<td></td>
</tr>
<tr>
<td>T(∃x(D(x) ∧ ¬T(x)))</td>
<td>3 ∀E</td>
</tr>
<tr>
<td>5 T(∃x(D(x) ∧ ¬T(x)))</td>
<td>2, 4 MP</td>
</tr>
<tr>
<td>6 ∃x(D(x) ∧ ¬T(x))</td>
<td>5, TE</td>
</tr>
<tr>
<td>7 ∃x(D(x) ∧ ¬T(x))</td>
<td>2-6 ¬I, DN</td>
</tr>
<tr>
<td>8 T(∃x(D(x) ∧ ¬T(x)))</td>
<td>7, TI</td>
</tr>
<tr>
<td>9 D(∃x(D(x) ∧ ¬T(x))) &amp;</td>
<td></td>
</tr>
<tr>
<td>T(∃x(D(x) ∧ ¬T(x)))</td>
<td>7, 8 &amp;I</td>
</tr>
<tr>
<td>10 ∃x(D(x) ∧ T(x))</td>
<td>9 ∃I</td>
</tr>
</tbody>
</table>

Of course, Geach’s derivation requires an additional premise and further steps before a contradiction is derivable. Nevertheless, where the Liar uses Leibniz’s law, this proof uses
quantificational rules of instantiation and conversely generalization, and neither proof uses reflexivization.

5.10 Discussion, Summation and Conclusion

History is inept at logically rationalizing itself, but it has more time to perfect its rationalizations than the rest of us. Prior to Tarski’s proof of the inconsistency of truth, some, like Russell, had treated the paradoxes of set-theory and semantics as uniform; while others, like Peano and Ramsey, had made a distinction between the then formal paradoxes of set-theory and the semi-formal paradoxes of semantics, grouping the Liar and Grelling’s in the latter. I wonder why after Tarski’s work and result, the issue of whether or not the paradoxes are all of the same formal type was not settled. Tarski gave a formally adequate definition of truth. (He even established the adequacy condition.) Tarski identified the conditions for the Liar. (Such conditions should be individually necessary and collectively sufficient.) So, why is it not a simple matter to say whether the Liar is formally of the same type as Russell’s? The Liar appears to require more logical machinery, more logical principles than Russell’s. But Quine [1962] discounts this, and even acknowledges a common form, yet Quine maintains a distinction between paradoxes of semantic concepts and paradoxes of membership. A better, or at least more deferential, rationalisation would have been to argue that because they have the same form, both concepts are inconsistent, so alternate consistent concepts must be motivated, such as the iterative conception of sets on the one hand, and grounded truth, satisfaction and perhaps denotation on the other. However, historically responses have diverged.

One says, the Liar and Grelling’s are antinomies of truth, in some deep sense – that it is really the naive concept that is the issue (whereas our naive concept of set was the wrong conception of set). Another response focuses on general forms – paradoxes of diagonalization, or paradoxes of truth, denotation and membership (taken as a common group, with common formal but paradoxical attributes.) In either case, the Liar and Grelling’s are grouped together. Still other responses concern the bearers of truth. These too generally group the Liar and Grelling’s paradoxes. Some do take issue with the T-schema as attributing truth to sentences even in a secondary way; and would, I assume, take issue with satisfaction as a relation between objects and expressions, in which cases
the formal representations of the semantic paradoxes need to be changed in ways that may no longer align with the paradoxes of set-theory.

The persistence with naive truth as the concept of truth, even in the face of Tarski’s result and despite set-theory’s success with an alternate conception of set, may be symptomatic of a semantic difference between the Liar and Russell’s (and Grelling’s). It is natural to think that the Liar expressions are in the range of applicability of the truth predicate, given that truth applies to meaningful sentences (even if in some secondary sense). If this intuition is contravened by an alternate conception of truth, then that alternate conception does not respect our intuitions about truth. If this intuition is validated by an alternate conception of truth, then something more needs to be said about how the Liar and strengthened Liar are to be avoided. On the other hand, whether Russell’s set is in its own range of applicability (so to speak) is something we have no natural strong intuition about. Likewise, I add, whether Grelling’s predicate is in the range of its own applicability is not something we have strong intuitions about. We are typically led to ask whether the predicate applies to itself or not via a preceding step which itself suggests that any predicate can be distinguished as one which is true of itself or is not. We are typically led to ask whether Russell’s set is a member of itself by a preceding step which suggests any set can be distinguished as self-membered or non-self-membered.

The usual proof of Tarski’s theorem that naive truth is inconsistent parallels the proof of the Liar. Leibniz’s law (or something like it) is necessary for both. The inconsistency of naive Satisfaction can be proved by paralleling the proof of Grelling’s paradox, and seems more economical in terms of the logical principles it uses than the Liar-like proof of Tarski’s theorem. A Grelling-like proof does not rely on Gödel’s diagonal lemma, or Leibniz’s law or even the substitution function. It seemed to depend on just the T²-schema. No sooner had I made this claim than the derivation was found to also require reflexivization and – well, actually, not much else (except substitution of equivalences).

The substitution of identicals, Leibniz’s law (or something like it) is a necessary condition to derive a contradiction for the Liar sentence. (Semantically, this may be thought of as in some way employing the substitution of co-referring terms.) Even when using Gödel’s diagonal lemma to prove the Liar, the derivation of the lemma itself relies on Leibniz’s law.
Leibniz’s law is not necessary to derive a Grelling’s paradox. This accords with the intuition that the Liar requires more logical machinery than Grelling’s. Yet what I have not seen made explicit in the literature is that reflexivization is necessary to derive Grelling’s paradox (and not the Liar).

Skyrms [1970] championed failure of substitution of identicals as a solution to the Liar; but such a solution will not solve Grelling’s paradox. It seems possible to independently solve Grelling’s by restricting Reflexivization without solving the Liar. The two paradoxes are therefore distinguishable – they are not the same paradox.

Notice that this is not just a case of different empirical facts affecting which statements are paradoxical. This is a distinction made on the basis of different paradoxes depending on different logical principles. We can conceive of restricting these principles in different ways that might solve one paradox but not the other. If, nevertheless, such restrictions are merely conceivable and the correct answer is to restrict the T-schemata, are they not the same or so very closely related that distinguishing them is trivial? On the one hand, if the restriction on the T-schema is a restriction involving reference whereas the restriction of the $T^2$-schema is on its range of applicability, then the two paradoxes still seem distinguishable. On the other hand, if the restriction in both cases is a restriction on range of applicability, then the paradoxes might seem very closely related. Nevertheless, I show there are two types of paradox of satisfaction and that the Grelling does not map to the Liar in an intuitive way.

My new relational Liar, which was introduced in terms of my favourite predicate in Chapter 4, can be used to prove the inconsistency of naive satisfaction in a Liar-like fashion. Such a proof would rely on Leibniz’s law but not reflexivization. The new paradox is of some interest in its own right as it is ubiquitously paradoxical. It contrasts nicely with Grelling’s as they are both paradoxes of satisfaction (or the truth relation), yet the new paradox necessarily uses Leibniz’s law in its derivation, like the Liar, whereas Grelling’s depends on reflexivization.

There appear to be cases, at least for ‘heterological’, where none of the T-schemata are required. This would be iconoclastic. A great deal of labour has been motivated by the idea that these paradoxes are paradoxes of truth. They can be used to prove naive truth and satisfaction are inconsistent; but, if the paradoxes can be formulated without truth or satisfaction, perhaps there is something else that gives rise to the inconsistency
inherent in the paradoxes, which would then bring into question their use to prove naive truth and satisfaction are inconsistent.

In any case, each paradox is not tied to a particular one of the T-schemata. The T-schema is perhaps the default schema for the Liar, while Grelling’s more naturally uses the $T^2$-schema. Nevertheless, the T2T-schema can be used in combination with another schema. Moreover, a contradiction can be derived from a Liar-like sentence using the $T^2$-relation without the T2T-schema. Given that my favourite formula just is ‘my favourite formula is not true of x’, the truth relational schema (the $T^2$-schema) can be used to derive a contradiction. Using a T-schemata, nevertheless, seems to remain a necessary condition to derive a contradiction for a Liar-like expression.

The $T^2$-schema is demonstrably not a necessary condition for deriving a contradiction from Grelling’s predicate. For, given that the T2T-schema relates the truth predicate (the T-predicate) to the truth relation (the T-relation), we can use the schema for the truth predicate (the T-schema, not the $T^2$-schema) to derive a contradiction from an equivalence in which Grelling’s predicate is embedded. Diagonalization, as self-intra-substitution, appears a necessary condition for Grelling’s; however, no substitution function is necessary to derive a Grelling’s paradox. So, it is difficult to be formally explicit about the sense in which diagonalization is necessary.

Using the T2T-schema we can give a Grelling-like proof of the inconsistency of naive truth. (As the T2T-schema relates the truth predicate and truth relation in a fundamental way, we would not attempt to avoid contradiction by rejecting the T2T-schema and keeping the T-schema.)

Diagonalization is not a necessary condition for the Liar: we can derive the Liar using an empirically true premise, the T-schema and Leibniz’s law. Likewise, a substitution function is not necessary to derive a Liar paradox for the same reason.

Given the intuitive relationship between the schemata, inconsistency of naive truth can also be derived using a Grelling’s predicate, the T-schema and reflexivization. Furthermore, given the relationship between the schemata, inconsistency of naive satisfaction can also be derived using a Liar sentence, the $T^2$-schema, the substitution function and identity.

In any case, I have shown the two paradoxes have different necessary conditions, and are distinguished this way. I have also shown that when one maps the Liar to a truth
relational paradox one does not get Grelling’s and *vice versa*. This mapping used the natural relationship between the T-schema and $T^2$-schema; so, even as paradoxes of truth, they are distinct types. They are distinct types, that is, except if $(\phi \& \sim \phi)$ does not necessarily follow from $(\phi \iff \sim \phi)$. For all I have said, they may have this in common; but were it even so, they still have distinct necessary conditions. These types cut across paradoxes of other concepts. I show in an appendix that the logical principle of reflexivization is also a necessary condition for Russell’s paradox: Russell’s is formally the same as Grelling’s, but not the Liar.